



UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE QUÍMICA

DEPARTAMENTO DE QUÍMICA ANALÍTICA

**EMPREGO DE MÉTODOS QUIMIOMÉTRICOS EM
ANÁLISES POR CROMATOGRAFIA GASOSA
BIDIMENSIONAL ABRANGENTE**

Tese de Doutorado

Luiz Antonio Fonseca de Godoy Junior

Orientador: Prof. Dr. Ronei Jesus Poppi

Co-orientador: Prof. Dr. Fabio Augusto

Campinas, 30 de maio de 2011

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

G548e Godoy Junior, Luiz Antonio Fonseca de.
Emprego de métodos quimiométricos em análises
por cromatografia gasosa bidimensional abrangente /
Luiz Antonio Fonseca de Godoy Junior. -- Campinas, SP:
[s.n], 2011.

Orientador: Prof. Dr. Ronei Jesus Poppi.
Coorientador: Prof. Dr. Fabio Augusto.

Doutorado - Universidade Estadual de Campinas,
Instituto de Química.

1. Análise multivariada. 2. Cromatografia gasosa
bidimensional abrangente. 3. Gasolina. 4. Perfume.
I. Poppi, Ronei Jesus. II. Augusto, Fabio.
III. Universidade Estadual de Campinas. Instituto de
Química. IV. Título.

Título em inglês: Application of chemometric methods in comprehensive two-dimensional gas chromatography analysis

Palavras-chaves em inglês: Multivariate analysis, Comprehensive two-dimensional gas chromatography, Gasoline, Perfume

Área de concentração: Química Analítica

Titulação: Doutor em Ciências

Banca examinadora: Prof. Dr. Ronei Jesus Poppi (orientador), Prof. Dr. Edénir Rodrigues Pereira Filho (DQ-UFSCar), Prof. Dr. Marco Flores Ferrão (IQ-UFRGS), Profa. Dra. Carla Beatriz Grespan Bottoli (IQ-UNICAMP), Profa. Dra. Maria Izabel Maretti Silveira Bueno (IQ-UNICAMP)

Data de defesa: 30/05/2011

Você é assim:
Um sonho pra mim
E quando eu não te vejo
Eu penso em você
Desde o amanhecer
Até quando eu me deito

Eu gosto de você
E gosto de ficar com você
Meu riso é tão feliz contigo
O meu melhor amigo é o meu amor

E a gente canta
E a gente dança
E a gente não se cansa
De ser criança
A gente brinca
Na nossa velha infância

Seus olhos, meu clarão,
Me guiam dentro da escuridão
Seus pés me abrem o caminho
Eu sigo e nunca me sinto só

Você é assim:
Um sonho pra mim
Quero te encher de beijos
Eu penso em você
Desde o amanhecer
Até quando eu me deito

Eu gosto de você
E gosto de ficar com você
Meu riso é tão feliz contigo
O meu melhor amigo é o meu amor

E a gente canta
E a gente dança
E a gente não se cansa
De ser criança
A gente brinca
Na nossa velha infância

Seus olhos, meu clarão,
Me guiam dentro da escuridão
Seus pés me abrem um caminho
Eu sigo e nunca me sinto só

Você é assim: um sonho pra mim
Você é assim...
Você é assim: um sonho pra mim
Você é assim...

Velha Infância - Tribalistas
(Composição Arnaldo Antunes, Carlinhos Brown e Marisa Monte)

Dedico esta canção à minha esposa

AGRADECIMENTOS

- Aos Profs. Dr. Ronei Jesus Poppi e Dr. Fabio Augusto pela oportunidade de trabalho, pela confiança depositada em mim, pela orientação e aprendizado ao longo destes anos e pela excelente convivência.
- Às Profas. Dra. Carla Beatriz Grespan Bottoli, Dra. Susanne Rath e Dra. Ana Valéria Colnaghi Simionato Cantú pelas sugestões dadas durante o exame de qualificação.
- Aos Profs. Dr. Marco Flôres Ferrão e Dr. Edenir Rodrigues Pereira Filho e às Profas. Dra. Carla Beatriz Grespan Bottoli e Dra. Maria Izabel Maretti Silveira Bueno pelas correções desta tese de doutorado.
- À Central Analítica do IQ-UNICAMP, laboratório que realiza análise da qualidade de amostras de gasolina para ANP, por fornecer as amostras de gasolinas utilizadas neste trabalho. Ao meu amigo Betinho pelos esclarecimentos sobre as análises de gasolina.
- Aos professores e funcionários do IQ que de alguma maneira contribuíram para este trabalho.
- Ao Instituto de Química pelas excelentes condições de trabalho.
- Ao CNPq e a FAPESP pelo suporte financeiro dado ao laboratório e bolsa concedida (CNPq).
- Aos amigos do grupo LAQQA: Salim, Danilo, Patrícia, PH, Marcelo, Guilherme, Werickson, Mônica, Renato, Jez, Cristina, Gilmare, Genésio, Julio, Mariana, Luciana, Fabiana, Ingrid, Fernando, Luciana, Alessandra, André e Marcia. Aos amigos do LCG: Borges, Alex, Gago, Aline, Carlos, Cláudia, Bá, Élton, Lucília, Maria, Mayra, Raquel, Juliano e Sandra.
- Agradecimento especial aos amigos Marcio, China e Ernesto que contribuíram diretamente no desenvolvimento desta tese.
- À minha família: em especial aos meus pais Luiz e Tereza, minha irmã Adriana, meu sogro Luis e minha esposa Laura pelo apoio, compreensão, força e por estarem ao meu lado sempre, nos momentos mais alegres e mais difíceis também.

CURRICULUM VITAE

1. Formação Acadêmica

1.1. Doutorado em química (03/2007 – 05/2011)

Área de concentração: química analítica.

Título da Tese: “Emprego de Métodos Quimiométricos em Análises por Cromatografia Bidimensional Abrangente”.

Orientador: Prof. Dr. Ronei Jesus Poppi. Co-orientador: Prof. Dr. Fabio Augusto.

Agência financiadora: CNPq.

Instituto de Química – Universidade Estadual de Campinas (UNICAMP).

1.2. Mestrado em química (08/2004 a 11/2006)

Área de concentração: química orgânica.

Título da Dissertação de Mestrado: “Síntese da 5-monobromocavernicolina e do 3-bromoverongiaquinol. Reações de Adição de Nucleófilos de Carbono à Íons Oxônios Cíclicos em Sistema sem Solvente”.

Orientador: Prof. Dr. Ronaldo Aloise Pilli.

Agência financiadora: FAPESP.

Instituto de Química – Universidade Estadual de Campinas (UNICAMP).

1.2. Bacharelado em Química e em Química Tecnológica (03/2000 a 07/2004).

Instituto de Química – Universidade Estadual de Campinas (UNICAMP).

2. Atuação Profissional

Empresa: Nanotimize Tecnologia LTDA. (11/2010 – atual)

Cargo: Químico.

2.1. Atividades Didáticas

- 08/09 a 12/09 Programa de Estágio Docente B – UNICAMP, disciplina: QA416.
- 03/09 a 07/09 Programa de Estágio Docente C – UNICAMP, disciplina: QA313.
- 03/04 a 07/04 Programa de Auxiliar Didático (PAD) – UNICAMP, disciplinas QG100 e QG102.

3. Publicações em Periódicos

- de Godoy, L. A. F.; Wang, L.; Pedroso, M.; Poppi, R. J.; Augusto, F. *Anal. Chim. Acta* **2011**, 699, 120.
- de Godoy, L. A. F.; Pilli, R. A. *Quim. Nova* **2011**, 33, 2042.
- de Godoy, L. A. F.; Pedroso, M. P.; Wang, L.; Augusto, F.; Poppi, R. J. *Talanta*, **2011**, 83, 1302.
- de Godoy, L. A. F.; Pedroso, M. P.; Ferreira, E. C.; Augusto, F.; Poppi, R. J. *J. Chromatogr. A* **2011**, 1218, 1663.
- Augusto, F.; Poppi, R. J.; Pedroso, M. P.; de Godoy, L. A. F.; Wang, L. *LC-GC Europe* **2010**, 23, 1.
- Pedroso, M. P.; de Godoy, L. A. F.; Fidelis, C. H. D.; Ferreira, E. C.; Poppi, R. J.; Augusto, F. *Quim. Nova* **2009**, 32, 421.
- Pedroso, M. P.; de Godoy, L. A. F.; Ferreira, E. C.; Poppi, R. J.; Augusto, F. *J. Chromatogr. A* **2008**, 1201, 176.
- de Godoy, L. A. F.; Ferreira, E. C.; Pedroso, M. P.; Fidélis, C. H. V.; Augusto, F.; Poppi, R. J. *Anal. Lett.* **2008**, 41, 1603.
- de Godoy, L. A. F.; Camilo, N. S.; Pilli, R. A. *Tetrahedron Lett.* **2006**, 47, 7853.

4. Resumos em Congressos

- 6 trabalhos apresentados em congressos nacionais na forma de pôster.
- 12 trabalhos apresentados em congressos internacionais na forma de pôster.
- 2 apresentações orais em congressos internacionais

5. Premiação Acadêmica

Prêmio Lavoisier de melhor aluno entre os formados do Instituto de Química - UNICAMP - em 21/08/2004 nas modalidades: Bacharelado em Química e Bacharelado em Química Tecnológica.

Resumo

EMPREGO DE MÉTODOS QUIMIOMÉTRICOS EM ANÁLISES POR CROMATOGRAFIA BIDIMENSIONAL ABRAGENTE

Nesta tese, foram construídos modelos quimiométricos a partir de cromatogramas de diferentes amostras obtidos por cromatografia gasosa bidimensional abrangente com detecção por ionização em chama (GC×GC-FID). Numa primeira aplicação, modelos foram desenvolvidos para identificar adulteração em gasolina. Para isso, uma amostra de gasolina não adulterada fornecida pela Agência Natural do Petróleo (ANP) foi utilizada na preparação de um conjunto de calibração e de um conjunto de validação de amostras de gasolinas adulteradas. Em seguida, o modelo construído foi utilizado para avaliar outras amostras de gasolinas e os resultados obtidos foram comparados com os resultados obtidos pela ANP. Em um segundo estudo com gasolinas, 51 amostras foram fornecidas pela ANP juntamente com os resultados dos testes físico-químicos utilizados pela agência na avaliação da qualidade da gasolina. O objetivo deste estudo foi a elaboração de modelos quimiométricos utilizando dados de GC×GC-FID para a previsão de parâmetros físico-químicos de gasolina. Os métodos de seleção de variáveis siPLS e algoritmo genético foram utilizados na preparação de modelos para a previsão da temperatura de destilação a 10, 50 e 90 % v/v de destilado e no ponto final da destilação. Um modelo PLS foi construído para a previsão da densidade das amostras de gasolinas. Em um terceiro estudo, utilizando PARAFAC e GC×GC-FID, foi realizada a diferenciação entre amostras de gasolina do Brasil e da Venezuela, devido ao intenso contrabando de gasolina existente da Venezuela para o Brasil. Em uma quarta aplicação, foi desenvolvido um algoritmo para seleção de intervalos em dados de ordem superior, o qual foi avaliado através da quantificação de alergênicos em perfumes. Por último, foi relatado o primeiro uso de MCR-ALS na construção de modelos quimiométricos utilizando-se dados obtidos por GC×GC-FID para quantificação de óleo essencial de alecrim em amostras complexas preparadas em laboratório.

Abstract

APPLICATION OF CHEMOMETRIC METHODS IN COMPREHENSIVE TWO-DIMENSIONAL GAS CHROMATOGRAPHY ANALYSIS

In this thesis, some chemometric models were built using chromatograms of different kinds of samples obtained by comprehensive two-dimensional gas chromatography with flame ionization detection (GC×GC-FID). In the first application, models were developed to identify adulteration in commercial gasoline samples. A non-adulterated gasoline sample provided by the *Agência Nacional do Petróleo* (ANP) was used to build a calibration and a validation sample set of adulterated gasoline. Then, the built model was used to evaluate other gasoline samples and the obtained results were compared to the results obtained by ANP. In a second study with gasoline, 51 samples were supplied by ANP with the results of the physicochemical tests used by ANP to evaluate the quality of gasoline samples. The aim of this study was the elaboration of chemometric models using GC×GC-FID chromatograms to preview some of the physicochemical tests used by ANP. The variable selection siPLS and genetic algorithm methods were used to build multivariate models in order to preview the distillation temperatures of gasoline at 10, 50 and 90 % v/v of distilled and at the final point of the distillation. A PLS model was built to predict the density of the gasoline samples. In a third study, using PARAFAC and GC×GC-FID, it was performed the differentiation between Brazilian and Venezuelan gasoline, due to the intense smuggling of gasoline from Venezuela to Brazil. In a fourth application, an algorithm was developed for variable selection in multi-way data, which was evaluated by quantification of allergens in perfume. Finally, it was reported the first use of MCR-ALS algorithm to develop chemometric models using data obtained by GC×GC-FID to quantify rosemary essential oil in complex samples prepared in laboratory.

Lista de Tabelas

Tabela 1. Composição das amostras de calibração e das amostras de validação em % v/v. As amostras de validação estão identificadas pela letra V.	38
Tabela 2. RMSECV (%) obtido em função do número de variáveis latentes.	42
Tabela 3. Valores previstos pelo modelo e entre parênteses a quantidade adicionada em % v/v.	44
Tabela 4. Valores de RMSEP e coeficiente de correlação obtido em cada modelo.	45
Tabela 5. Comparação entre os resultados previstos pelo modelo quimiométrico e os obtidos pela ANP.	46
Tabela 6. Composição das amostras com cinco padrões utilizadas para testar o iNLPS. Concentrações em % v/v.	79
Tabela 7. Composição das amostras de perfume com alergênicos adicionados em laboratório. Concentrações em $\mu\text{L/L}$	81
Tabela 8. Número de intervalos e variáveis latentes utilizados em cada modelo de calibração. Valores de RMSEC, RMSECV e RMSEP obtidos.	84
Tabela 9. Número de intervalos e variáveis latentes utilizadas em cada modelo de calibração. Valores de RMSEC e RMSECV obtidos.	87
Tabela 10. Comparação dos resultados obtidos por iNPLS, iPLS e NPLS.	89
Tabela 11. Composição das amostras do conjunto de validação.	97
Tabela 12. Valores previstos de concentração de óleo essencial de alecrim nas amostras do conjunto de validação.	106
Tabela 13. Tipos de amostras analisadas por GC \times GC-FID e métodos quimiométricos utilizados.	113

Lista de Figuras

Figura 1. Esquema de funcionamento de um modulador criogênico de quatro jatos. O jato azul indica o gás frio e o jato vermelho o gás quente.....	6
Figura 2. Resultado do processo de modulação: série de picos modulados correspondentes ao pico não modulado (linha tracejada).....	7
Figura 3. Forma de representação dos cromatogramas em GC×GC.....	8
Figura 4. Comparação da capacidade de pico entre 1D-GC, GC-GC e GC×GC. ...	9
Figura 5. Seções de cromatogramas de gasolina: (a) GC-FID e (b) GC×GC-FID. Colunas: GC-FID: HP-5 (30 m × 0,25 mm × 0,25 μm); GC×GC-FID: HP-5 (30 m × 0,25 mm × 0,25 μm) + DBWax (1,0 m × 0,1 mm × 0,1 μm).....	9
Figura 6. Tipos de dados com a ordem de cada um entre parênteses.....	14
Figura 7. Representação gráfica do PARAFAC.....	18
Figura 8. Codificação de cada gene (variável) de um cromatograma em 0 ou 1. O conjunto de vários genes forma um cromossomo (indivíduo) e o conjunto de vários cromossomos forma uma população.	24
Figura 9. Etapa de cruzamento simples e duplo do GA.	25
Figura 10. Diagrama de cores do GC×GC-FID obtido para gasolina C. E= etanol, B = benzeno, B1 = tolueno, B2 = benzeno C2 substituído, B3 = benzeno C3 substituído, B4 = benzeno C4 substituído, B5 = benzeno C5 substituído, N = naftaleno, A = alcanos, cicloalcanos e alcenos. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 μm) + DB-Wax (1,0 m x 0,1 mm x 0,1 μm). Programação de temperatura do forno cromatográfico: 40 °C (3 min.) até 160 °C (4 °C/min.) e então aquecido até 240 °C (20 °C/min.) (3 min.). Razão de divisão do injetor 1:300, volume de injeção 0,6 μL e PM 6 s.	40
Figura 11. Cromatogramas GC×GC-FID obtidos para gasolina C, aguarrás, querosene e thinner. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 μm) + DB-Wax (1,0 m x 0,1 mm x 0,1 μm). Programação de temperatura do forno cromatográfico: 40 °C (3 min.) até 160 °C (4 °C/min.) e então aquecido até 240 °C (20 °C/min.) (3 min.). Razão de divisão do injetor 1:300, volume de injeção 0,6 μL e PM 6 s.	41

Figura 12. Concentrações previstas por validação cruzada versus concentrações das amostras de calibração para os adulterantes e para gasolina C.	43
Figura 13. Comparação dos cromatogramas obtidos para a amostra N3 e para a gasolina utilizada no modelo de calibração. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 µm) + DB-Wax (1,0 m x 0,1 mm x 0,1 µm). Programação de temperatura do forno cromatográfico: 40 °C (3 min.) até 160 °C (4 °C/min.) e então aquecido até 240 °C (20 °C/min.) (3 min.). Razão de divisão do injetor 1:300, volume de injeção 0,6 µL e PM 6 s.	47
Figura 14. Diagrama GC×GC-FID obtido para gasolina. E= etanol, B = benzeno, B1 = tolueno, B2 = benzeno C2 substituído, B3 = benzeno C3 substituído, B4 = benzeno C4 substituído, B5 = benzeno C5 substituído, N = naftaleno, A = alcanos, cicloalcanos e alcenos. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 µm) + DB-Wax (1,0 m x 0,1 mm x 0,1 µm). Programação de temperatura do forno cromatográfico: 40 °C até 120 °C (3 °C/min.) e então aquecido até 220 °C (20 °C/min.). Razão de divisão do injetor 1:250, volume de injeção 1,0 µL e PM 4 s..	51
Figura 15. Diagrama GC×GC-FID de gasolina mostrando os intervalos selecionados pelo siPLS em cinza para o modelo de previsão de temperatura a 10% de destilado. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 µm) + DB-Wax (1,0 m x 0,1 mm x 0,1 µm). Programação de temperatura do forno cromatográfico: 40 °C até 120 °C (3 °C/min.) e então aquecido até 220 °C (20 °C/min.). Razão de divisão do injetor 1:250, volume de injeção 1,0 µL e PM 4 s..	54
Figura 16. Valores da temperatura a 10% de destilado versus valores previstos pelo modelo quimiométrico para as amostras de validação.	56
Figura 17. Diagrama GC×GC-FID de gasolina mostrando os intervalos selecionados pelo siPLS em cinza para o modelo de previsão de temperatura a 50% de destilado. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 µm) + DB-Wax (1,0 m x 0,1 mm x 0,1 µm). Programação de temperatura do forno cromatográfico: 40 °C até 120 °C (3 °C/min.) e então aquecido até 220 °C (20 °C/min.). Razão de divisão do injetor 1:250, volume de injeção 1,0 µL e PM 4 s..	57
Figura 18. Valores de temperatura a 50% de destilado versus valores previstos pelo modelo quimiométrico para as amostras de validação.	58

Figura 19. Diagrama GC×GC-FID de gasolina mostrando os intervalos selecionados pelo siPLS em cinza para o modelo de previsão de temperatura a 90% de destilado. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 μm) + DB-Wax (1,0 m x 0,1 mm x 0,1 μm). Programação de temperatura do forno cromatográfico: 40 °C até 120 °C (3 °C/min.) e então aquecido até 220 °C (20 °C/min.). Razão de divisão do injetor 1:250, volume de injeção 1,0 μL e PM 4 s.. 60

Figura 20. Valores de temperatura a 90% de destilado versus valores previstos pelo modelo quimiométrico para as amostras de validação..... 61

Figura 21. Diagrama GC×GC-FID de gasolina mostrando os intervalos selecionados pelo siPLS em cinza para o modelo de previsão de temperatura no ponto final da destilação. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 μm) + DB-Wax (1,0 m x 0,1 mm x 0,1 μm). Programação de temperatura do forno cromatográfico: 40 °C até 120 °C (3 °C/min.) e então aquecido até 220 °C (20 °C/min.). Razão de divisão do injetor 1:250, volume de injeção 1,0 μL e PM 4 s.. 62

Figura 22. Valores de temperatura no ponto final de destilação versus valores previstos pelo modelo quimiométrico para as amostras de validação..... 63

Figura 23. Valores de massa específica das amostras de validação versus valores previstos pelo modelo quimiométrico. 64

Figura 24. Cromatograma obtido para (A) gasolina brasileira e (B) gasolina venezuelana. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 μm) + SupelcoWax (1,0 m x 0,1 mm x 0,1 μm). Programação de temperatura do forno cromatográfico: 60 °C até 223 °C (3 °C/min.). Razão de divisão do injetor 1:300, volume de injeção 1,0 μL e PM 6 s. 68

Figura 25. Escores obtidos para as amostras de gasolinas brasileira e venezuelana..... 70

Figura 26. Pesos obtidos nas duas dimensões da separação cromatográfica. Em azul fator 1, em verde fator 2 e em vermelho fator 3..... 71

Figura 27. Divisão da matriz de dados em intervalos nas duas dimensões dos dados. 76

Figura 28. Diagrama GC×GC obtido para amostra 25. A área marcada em cinza representa o intervalo selecionado pelo iNPLS em cada modelo. (A) hexano. (B)

tolueno. (C) ciclohexanona. (D) 2-octanona. (E) 1-octanol. (F) undecano. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 μ m) + BPX-50 (1,0 m x 0,1 mm x 0,1 μ m). Programação de temperatura do forno cromatográfico: 60 $^{\circ}$ C durante 3 min., sendo em seguida aquecido até 120 $^{\circ}$ C (3 $^{\circ}$ C/min.) e, então, aquecido até 240 $^{\circ}$ C (20 $^{\circ}$ C/min.). Razão de divisão do injetor 1:250, volume de injeção 1,0 μ L e PM 4 s.	84
Figura 29. Concentrações previstas pelo modelo iNPLS para cada analito.	85
Figura 30. Diagrama GC \times GC obtido para amostra contendo (A) álcool benzílico, (C) citronelol e (G) geraniol. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 μ m) + BPX-50 (1,0 m x 0,1 mm x 0,1 μ m). Programação de temperatura do forno cromatográfico: 60 $^{\circ}$ C a 240 $^{\circ}$ C (3 $^{\circ}$ C/min.). Razão de divisão do injetor 1:30, volume de injeção 1,0 μ L e PM 4 s.	86
Figura 31. Diagrama GC \times GC obtido para a amostra 19. O quadrado preto indica os intervalos selecionados pelo iNPLS para o modelo de (A) álcool benzílico, (B) citronelol e (C) geraniol. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 μ m) + SPwax (1,0 m x 0,1 mm x 0,1 μ m). Programação de temperatura do forno cromatográfico: 60 $^{\circ}$ C a 240 $^{\circ}$ C (3 $^{\circ}$ C/min.). Razão de divisão do injetor 1:30, volume de injeção 1,0 μ L e PM 4 s.	88
Figura 32. Concentrações previstas pelo modelo iNPLS para (A) álcool benzílico, (B) citronelol e (C) geraniol.....	89
Figura 33. (A) Exemplo de um suposto cromatograma GC \times GC-FID vetorizado. (B) Diagrama de contorno do mesmo cromatograma, no qual o retângulo representa o intervalo selecionado pelo iNPLS em um suposto modelo para quantificar o composto Y e a região compreendida entre as linha pontilhadas representa o intervalo selecionado na suposta quantificação de Y pelo iPLS.....	91
Figura 34. Esquematização para utilização de MCR-ALS em dados obtidos por GC \times GC-FID.	99
Figura 35. GC \times GC-FID obtido na análise da amostra contendo 15% v/v de óleo essencial de alecrim em álcool de cereais. Regiões que foram excluídas para a construção do modelo quimiométrico destacadas em cinza. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 μ m) + Supelcowax (1,0 m x 0,10 mm x 0,10 μ m).	

Programação de temperatura do forno cromatográfico: 60 °C até 250 °C (3 °C/min). Razão de divisão do injetor 1:200, volume de injeção 1,0 µL e PM 6 s. 100

Figura 36. GC×GC-FID obtido para (A) óleo essencial de alecrim, (B) óleo essencial de abacaxi e (C) perfume. Regiões do cromatograma excluídas para a construção do modelo quimiométrico não mostradas. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 µm) + Supelcowax (1,0 m x 0,10 mm x 0,10 µm).

Programação de temperatura do forno cromatográfico: 60 °C até 250 °C (3 °C/min). Razão de divisão do injetor 1:200, volume de injeção 1,0 µL e PM 6 s. 102

Figura 37. GC×GC-FID obtido para as amostras de validação (A) V5 e (B) V9. Regiões do cromatograma excluídas para a construção do modelo quimiométrico não mostradas. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 µm) + Supelcowax (1,0 m x 0,10 mm x 0,10 µm).

Programação de temperatura do forno cromatográfico: 60 °C até 250 °C (3 °C/min). Razão de divisão do injetor 1:200, volume de injeção 1,0 µL e PM 6 s. 103

Figura 38. GC×GC-FID obtido para (A) óleo essencial de alecrim, (B) óleo essencial de abacaxi e (C) perfume. GC×GC-FID recuperados do modelo MCR para (D) óleo essencial de alecrim, (E) óleo essencial de abacaxi e (F) perfume.

Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 µm) + Supelcowax (1,0 m x 0,10 mm x 0,10 µm). Programação de temperatura do forno cromatográfico: 60 °C até 250 °C (3 °C/min). Razão de divisão do injetor 1:200, volume de injeção 1,0 µL e PM 6 s. 105

Figura 39. Concentração prevista pelo modelo MCR de óleo essencial de alecrim nas amostras de validação..... 107

Índice

1. Cromatografia Gasosa Bidimensional Abrangente.....	3
2. Quimiometria.....	13
2.1. Classificação dos Dados.....	13
2.2. Mínimos Quadrados Parciais Multimodo (NPLS).....	16
2.3. Análise de Fatores Paralelos (PARAFAC).....	17
2.4. Resolução Multivariada de Curvas (MCR).....	20
2.5. Métodos para Seleção de Variáveis	22
2.5.1. Mínimos Quadrados Parciais por Sinergismo de Intervalos (siPLS) ...	22
2.5.2. Algoritmo Genético	23
3. Objetivo Geral	29
4. Análises de Amostras de Gasolina.....	33
4.1. Introdução.....	33
4.2. Desenvolvimento de um Modelo Quimiométrico para Identificar Adulteração em Gasolina.....	36
4.2.1. Objetivos	36
4.2.2. Parte Experimental.....	36
4.2.3. Resultados e Discussão	39
4.2.4. Conclusões.....	47
4.3. Previsão de Propriedades Físico-Químicas de Gasolina.....	49
4.3.1. Objetivos	49
4.3.2. Parte Experimental.....	49
4.3.3. Resultados e Discussão	50
4.3.3.1. Previsão das Temperaturas da Curva de Destilação de Gasolina. 51	
4.3.3.1.1. Previsão da Temperatura a 10 % de Destilado.....	53
4.3.3.1.2. Previsão da Temperatura a 50 % de Destilado.....	56
4.3.3.1.3. Previsão da Temperatura a 90 % de Destilado.....	58
4.3.3.1.4. Previsão da Temperatura no Ponto Final da Destilação	61
4.3.3.2. Previsão do Valor da Massa Específica de Gasolina	63
4.3.4. Conclusões.....	64

4.4. Diferenciação entre Amostras de Gasolinas do Brasil e da Venezuela	65
4.4.1. Objetivos	65
4.4.2. Parte Experimental	65
4.4.3. Resultados e Discussão	66
4.4.4. Conclusões.....	72
5. Desenvolvimento de um Algoritmo para Seleção de Variáveis por Intervalos para o Método de Mínimos Quadrados Parciais Multimodo.....	75
5.1. Introdução.....	75
5.2. Objetivos.....	78
5.3. Parte Experimental	78
5.3.1. Análise de padrões.....	78
5.3.2. Análise de alergênicos em perfume	80
5.4. Resultados e Discussões.....	82
5.4.1. Análise de padrões.....	82
5.4.2. Análise de alergênicos em perfume	85
5.5. Conclusões	91
6. Análise de Óleo Essencial de Alecrim Utilizando o Método de Resolução Multivariada de Curvas (MCR) em Dados Obtidos por GC×GC-FID.....	95
6.1. Introdução.....	95
6.2. Objetivos.....	96
6.3. Parte Experimental	97
6.4. Resultados e Discussões.....	98
6.5. Conclusões	107
7. Conclusões Gerais	111
8. Referências Bibliográficas	117

CAPÍTULO 1:

Cromatografía Gasosa Bidimensional Abrangente

1. Cromatografia Gasosa Bidimensional Abrangente

A cromatografia gasosa (GC) é uma técnica com um poder de resolução excelente, tornando possível, muitas vezes, a análise de dezenas de substâncias de uma mesma amostra. O uso bastante acentuado da GC se deve também aos baixos limites de detecção que podem ser alcançados. Essa característica faz com que haja necessidade de apenas pequenas quantidades da amostra em estudo, o que, em certos casos, é um fator crítico e limita a utilização de outras técnicas. É importante salientar ainda que a GC é excelente como técnica quantitativa¹.

A análise de algumas amostras por GC pode ser considerada um desafio, principalmente nas quais o número de compostos presentes é da ordem de centenas e, também, quando muitos dos analitos presentes estão em concentrações muito baixa. Mesmo em situações nas quais as condições de separação foram otimizadas, é comum que alguns compostos apresentem coeluição. Esse problema é geralmente encontrado em amostras de petróleo e derivados, ambientais, alimentos, óleos essenciais e perfumes².

O poder de separação da GC pode ser mensurado pela capacidade de pico do sistema cromatográfico, que é definida como o número máximo de constituintes que podem ser separados sob determinadas condições cromatográficas¹. Uma forma comum de se buscar uma melhor separação é empregar uma coluna cromatográfica mais longa; contudo, esta estratégia tem como consequência um aumento no tempo da análise, podendo resultar também em alargamento de picos, e, conseqüentemente, em limites de detecção menores. Outra estratégia empregada é a redução do diâmetro da coluna, o que aumenta a eficiência da separação, entretanto colunas com diâmetros inferiores aos menores atualmente disponíveis seriam inviáveis para aplicações rotineiras. Outra abordagem é a utilização de detecção seletiva, como a espectrometria de massas (MS)³.

Uma alternativa às abordagens anteriormente descritas é a utilização de separação multidimensional, no qual a amostra é submetida sequencialmente a diferentes processos de eluição. As separações multidimensionais apresentam uma maior capacidade de pico em relação às separações monodimensionais.

Para que uma separação seja realmente multidimensional, dois requisitos devem ser seguidos: os componentes da mistura devem ser submetidos a duas ou mais etapas ou mecanismos de separação e quando dois componentes são separados em uma etapa, eles devem permanecer separados durante toda a análise⁴.

A operacionalização de separações bidimensionais em GC é muito complexa, já que a coleta de frações do eluato gasoso de uma coluna capilar e transferência para uma segunda coluna sem perdas, contaminação ou alargamento excessivo da banda cromatográfica não é trivial. A primeira tentativa de implementação foi o cromatógrafo a gás de dois estágios de Simmons e Snyder⁵, no qual frações selecionadas do eluato de uma coluna, contendo analitos não separados, eram desviadas para uma segunda coluna instalada no mesmo forno através de um sistema complexo de válvulas. Esse procedimento em GC de coletar em uma ou mais frações de uma coluna e reinjetá-las em uma segunda coluna é chamado de cromatografia gasosa bidimensional de frações parciais (GC-GC)⁶.

Em 1991, Liu e Philips⁷ desenvolveram um novo sistema bidimensional consistindo de duas colunas capilares conectadas em série, sendo a primeira coluna (ou primeira dimensão, ¹D) de polietilenoglicol com dimensões de 21 m de comprimento e 0,25 mm de diâmetro interno e a segunda coluna (ou segunda dimensão, ²D) de polidimetilsiloxano com dimensões de 1 m de comprimento e 0,1 mm de diâmetro interno. Uma seção de 15 cm do início da segunda coluna, recoberto com um filme de ouro, era mantido fora do forno cromatográfico e conectado a uma fonte de 40 V. Esse setor metalizado da segunda dimensão é denominado modulador, que tem a função de coletar continuamente frações do eluato da primeira coluna, reconcentrá-las e transferi-las para a segunda coluna. Essa técnica, no qual em um cromatógrafo a gás duas colunas de fases estacionárias distintas são conectadas em série e todo o eluato é submetido a separação nas duas colunas é denominada cromatografia gasosa bidimensional abrangente (GC×GC)⁶.

Na GC×GC é fundamental que a segunda coluna seja curta e eficiente para que as frações coletadas ao final da ¹D possam ser separadas em alguns segundos e com alta eficiência na ²D. Os mecanismos de separação da ¹D e da

²D devem ser ortogonais, possibilitando assim que uma separação pouco eficiente na primeira coluna seja melhorada na segunda.

Como mencionado anteriormente, na GC×GC duas colunas são conectadas em série, através de um modulador, sendo a coluna da ¹D de dimensões convencionais e a da ²D mais curta (do tipo de coluna usada em cromatografia gasosa rápida). A função do modulador pode ser dividida em três etapas: *i*) coletar continuamente pequenas frações do efluente da ¹D, garantindo que a separação nessa dimensão seja mantida, *ii*) reconcentrar o efluente da ¹D em uma banda estreita e *iii*) transferir rapidamente para a ²D a fração coletada. O tempo necessário para realizar essas três etapas é denominado de período de modulação (PM), o qual geralmente varia de 2 a 10 s e está relacionado ao tempo necessário para que os compostos sejam eluídos na ²D. O PM deve ser o menor possível, a fim de que a separação obtida na ¹D seja mantida⁸.

Os moduladores podem ser classificados em dois tipos básicos de acordo com seu princípio operacional: os moduladores térmicos e os moduladores a base de válvulas. Os térmicos se dividem em dois grupos: os baseados em aquecimento e os criogênicos. Atualmente, os moduladores criogênicos são os mais empregados⁹. O modulador empregado ao longo desta tese foi um modulador criogênico de quatro jatos, dois jatos quentes e dois jatos frios, sendo N₂ gasoso resfriado em nitrogênio líquido como fluido criogênico e N₂ aquecido como gás quente, o qual foi desenvolvido em laboratório¹⁰. Neste modulador os jatos são posicionados dois a dois em duas posições diferentes no início da segunda coluna, sendo um jato quente e um jato frio em cada posição. A primeira posição no sentido do fluxo de gás é chamada de coleta e a segunda é chamada de focagem. Primeiramente, o gás frio é acionado na posição de coleta aprisionando todo o efluente da primeira coluna, em seguida, é acionado o gás quente na posição de coleta e o gás frio na posição de focagem, deste modo o efluente é liberado da posição de coleta e aprisionado na posição de focagem na forma de uma banda estreita, por último é acionado o gás quente na focagem liberando o efluente para a segunda coluna e o gás frio na posição de coleta

aprisionando novamente o efluente da 1^{D} , completando assim o ciclo de modulação (figura 1).

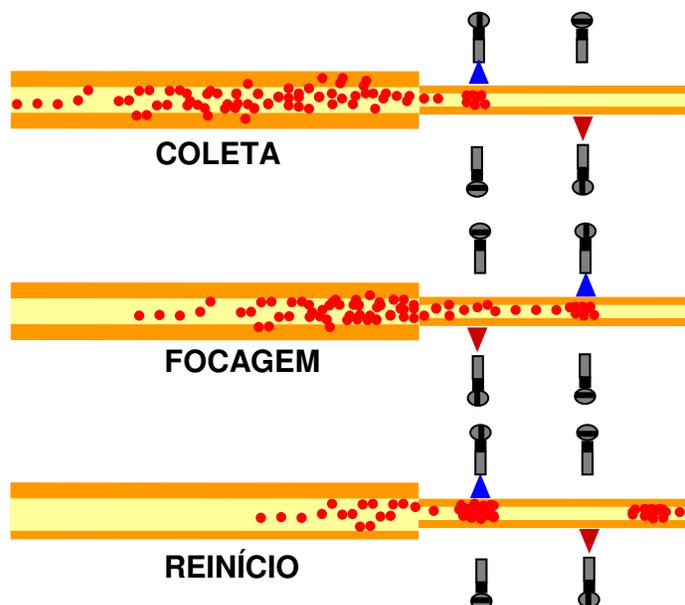


Figura 1. Esquema de funcionamento de um modulador criogênico de quatro jatos. O jato azul indica o gás frio e o jato vermelho o gás quente.

Uma importante característica da GC \times GC que deve ser destacada é a reconcentração do efluente da 1^{D} , que ocorre durante a modulação, ocasionando um aumento considerável na sensibilidade. Na figura 2 está ilustrada a comparação de um pico cromatográfico antes e após a modulação. Como pode ser visto, um pico largo e pouco intenso ao final da 1^{D} é transformado em uma série de picos estreitos e intensos, sendo mais facilmente detectados devido ao aumento na relação sinal/ruído¹¹. A área do pico ao final da 1^{D} é igual a soma das áreas de todos os picos após a modulação. Outra importante característica da GC \times GC é a estruturação cromatográfica, que é o ordenamento no cromatograma de compostos quimicamente relacionados. O agrupamento destes compostos resulta no “efeito telhado”, no qual cada conjunto de “telhas” corresponde a isômeros com propriedades químicas similares¹². A estruturação cromatográfica será exemplificada na seção de análise de amostras de gasolina (capítulo 4).

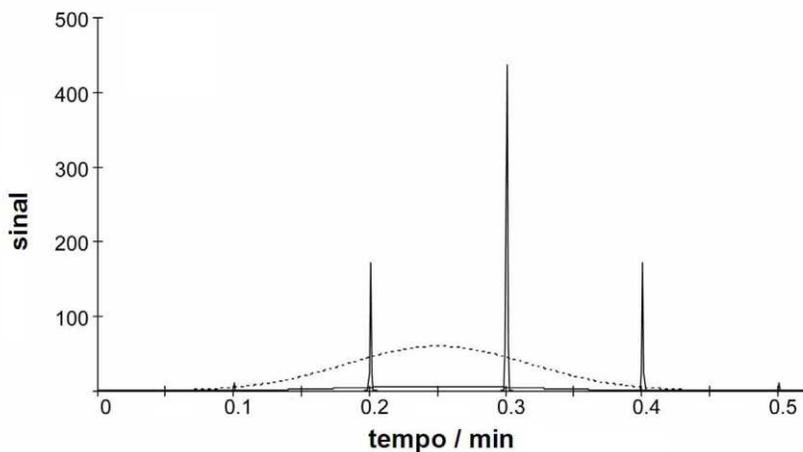


Figura 2. Resultado do processo de modulação: série de picos modulados correspondentes ao pico não modulado (linha tracejada).

Uma diferença da GC×GC em relação a 1D-GC é a forma de representação dos cromatogramas. O registro do sinal do detector em função do tempo em um sistema GC×GC é uma seqüência contínua e encadeada dos cromatogramas curtos para cada fração eluída na ²D. A partir desse registro digitalizado e conhecido o PM, pode-se construir gráficos tridimensionais sinal do detector × tempo de retenção na primeira coluna (¹t_R) × tempo de retenção na segunda coluna (²t_R). Na Figura 3, uma banda cromatográfica larga, consistindo de três analitos não separados na ¹D (A), é modulada e eluída na ²D, gerando um cromatograma modulado bruto (B). Cada um dos cromatogramas individuais das frações pode ser destacado do cromatograma bruto (C) e a partir deles pode-se gerar cromatogramas GC×GC do tipo sinal × ¹t_R × ²t_R como: diagrama tridimensional (“3D Plot”, D), como uma série de curvas de nível monocromáticas chamadas de diagrama de contorno (“Contour Plot”, E) ou usando escalas cromáticas de intensidade de sinal, o diagrama de cores (“Colour Plot”, F)¹³.

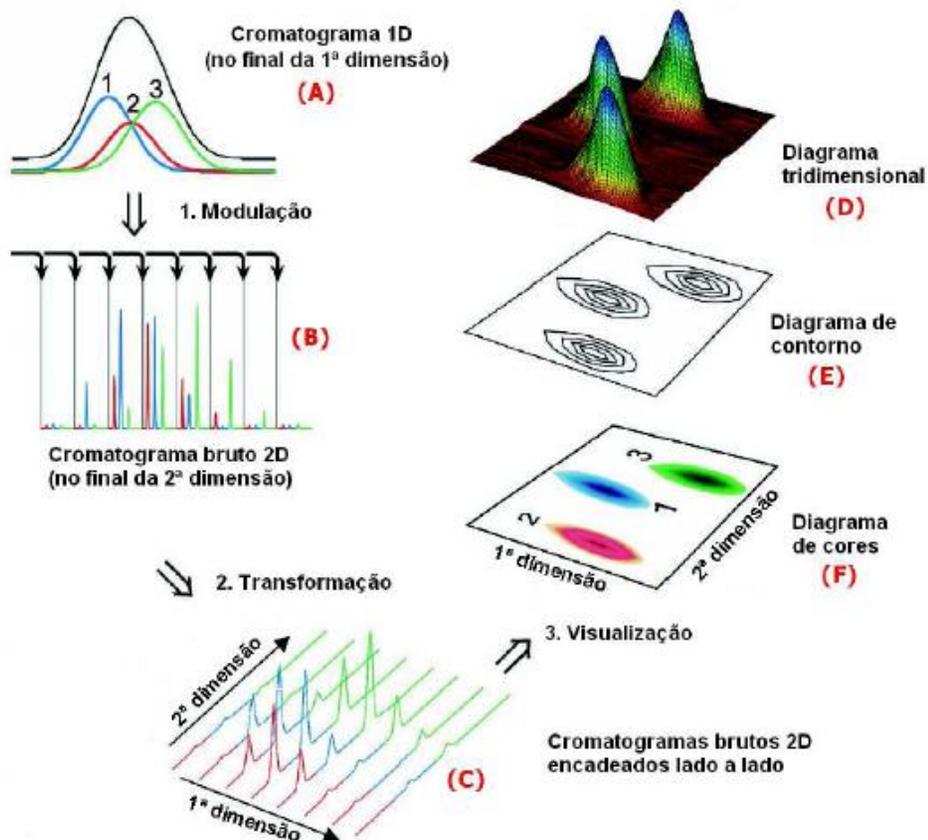


Figura 3. Forma de representação dos cromatogramas em GCxGC.

A principal vantagem da GCxGC frente à GC-GC e a 1D-GC é a elevada capacidade de pico. Por exemplo, assumindo que a capacidade de pico da primeira coluna seja n_1 e a capacidade de pico da segunda coluna seja n_2 , a capacidade de pico da GCxGC será $n_1 \times n_2$, pois toda a amostra é submetida à separação nas duas colunas, enquanto que a capacidade de pico da GC convencional será apenas n_1 e da GC-GC será $n_1 + y(n_2)$, sendo y o número de frações que foram submetidas à separação nas duas colunas (figura 4)¹⁴. A diferença no poder de separação entre a GCxGC e a 1D-GC está ilustrada na figura 5, na qual é mostrada uma parte de um cromatograma obtido para uma amostra de gasolina com ¹D-GC-FID e com GCxGC-FID.

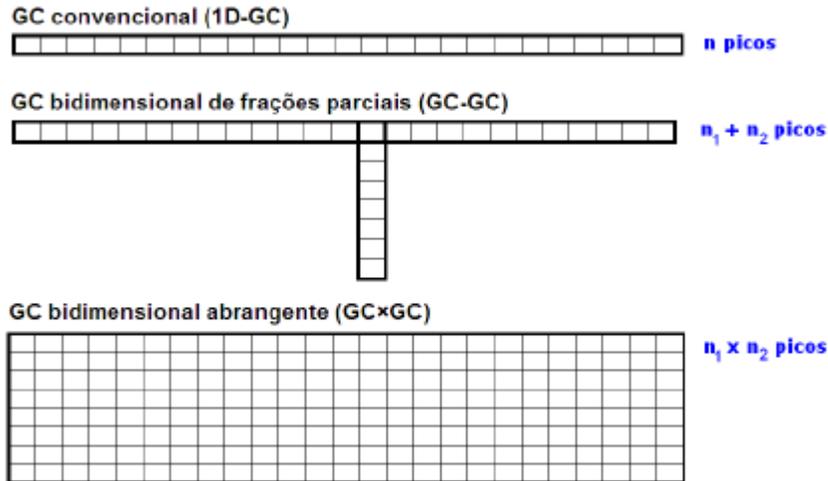


Figura 4. Comparação da capacidade de pico entre 1D-GC, GC-GC e GCxGC.

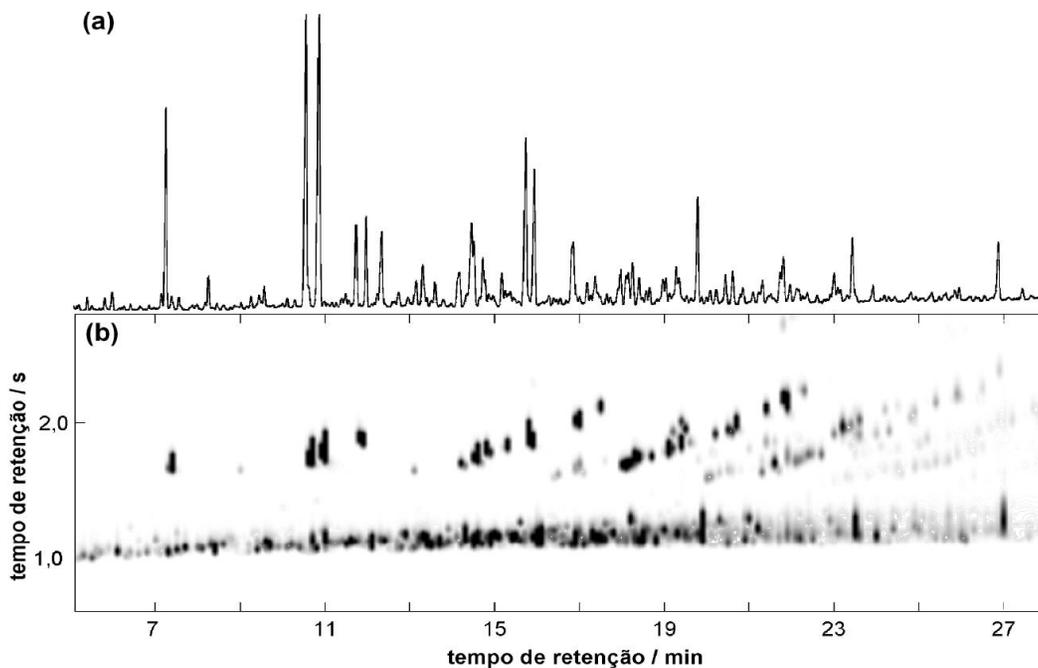


Figura 5. Seções de cromatogramas de gasolina: (a) GC-FID e (b) GCxGC-FID. Colunas: GC-FID: HP-5 (30 m × 0,25 mm × 0,25 μm); GCxGC-FID: HP-5 (30 m × 0,25 mm × 0,25 μm) + DBWax (1,0 m × 0,1 mm × 0,1 μm).

Apesar de a GCxGC ter sido desenvolvida há quase vinte anos⁷, pode-se dizer que ela é uma técnica nova, pois o número de publicações a partir de 2003 representa mais de 80% do total de publicações, segundo pesquisa realizada no

sítio *Web of Science*, buscando o termo *comprehensive two-dimensional gas chromatography* no título do artigo. Atualmente, existem três grupos de pesquisas que desenvolvem atividades de pesquisa em GC×GC no Brasil, um na UNICAMP, um na UFRGS e um na UFRJ.

CAPÍTULO 2:

Quimiometria

2. Quimiometria

O uso de computadores para analisar dados químicos apresentou um crescimento acentuado nos últimos trinta anos, devido aos avanços em informática na parte de *hardware* e *software* e também devido à diminuição dos custos. Em consequência disso, a aquisição de dados, principalmente na área de química analítica, atingiu um nível bastante sofisticado com o interfaceamento de instrumentos aos computadores. Um fato marcante desse desenvolvimento é que vem sendo possível obter uma quantidade cada vez maior de informações sobre o meio em estudo, o que torna possível a análise de espécies de interesse mesmo em sistemas complexos. Contudo, a análise dessa quantidade de dados muitas vezes não constitui uma tarefa simples, sendo freqüentemente necessária a utilização de procedimentos mais elaborados para a obtenção das informações desejadas. A dificuldade em lidar com dados complexos motivou o desenvolvimento da quimiometria, que é uma área destinada à análise de dados químicos de natureza multivariada¹⁵.

A quimiometria pode ser definida como a aplicação de métodos matemáticos e estatísticos, assim como de lógica, em química, visando extrair uma maior quantidade de informação química de um conjunto de dados e também para planejar experimentos de forma otimizada¹⁶. A maioria dos métodos quimiométricos é universal, isto é, podem ser aplicados em dados obtidos de diferentes instrumentos analíticos¹⁷.

2.1. Classificação dos Dados

Dependendo da instrumentação utilizada na aquisição dos dados, estes podem ser constituídos por apenas um número ou escalar, um vetor, uma matriz ou mesmo um cubo de dados por amostra, arranjos que, de acordo com a nomenclatura comumente utilizada em álgebra linear, são denominados de tensores de ordem zero, primeira, segunda e terceira ordens, respectivamente (figura 6)¹⁸.

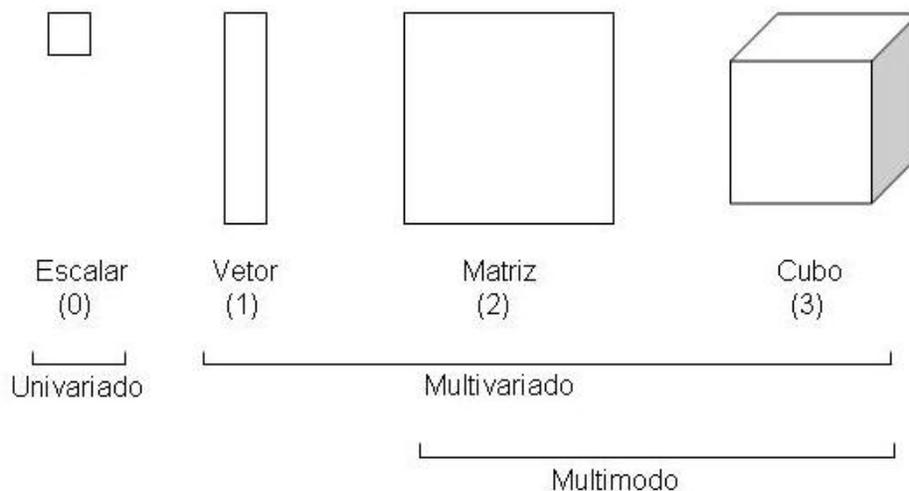


Figura 6. Tipos de dados com a ordem de cada um entre parênteses.

Dados de ordem zero são obtidos de instrumentos que produzem apenas uma resposta por amostra, por exemplo, absorvância em um único comprimento de onda e leitura em eletrodos íon-seletivos. Modelos de calibração univariada são construídos com este tipo de dados. Estes modelos são os mais aplicados em análises de rotina, devido à sua simplicidade, e dispõem de procedimentos de validação bem estabelecidos e descritos na literatura¹⁹. Entretanto, o uso destes modelos requer que o sinal instrumental medido não apresente contribuição de outras espécies além do analito em estudo, ou seja, esteja livre de interferentes²⁰.

Dados de primeira ordem são obtidos em equipamentos que geram um vetor de resposta por análise, caso da GC-FID e da espectroscopia no infravermelho. O conjunto de várias amostras forma uma matriz de dados. Modelos de calibração multivariada são construídos a partir deste tipo de dados e apresentam uma série de vantagens em relação à calibração univariada, entre elas, análises mesmo na presença de interferentes, desde que os interferentes estejam presentes nas amostras de calibração, determinações simultâneas de mais de uma espécie e determinações quando os sinais dos analitos em estudos apresentam-se sobrepostos entre si ou com os sinais dos interferentes calibrados. Existem vários métodos quimiométricos que utilizam dados de primeira ordem para construção de um modelo de calibração multivariada, entre eles, regressão

linear múltipla (MLR), regressão por componentes principais (PCR) e regressão por quadrados mínimos parciais (PLS). Esses modelos têm apresentado ótimos resultados com aplicações em diversas áreas, sendo um exemplo a sua utilização em dados de GC-FID para previsão de parâmetros de qualidade de café²¹. Porém, a utilização destes métodos possui a desvantagem de requerer um número relativamente grande de amostras de calibração, principalmente no caso de amostras complexas¹⁸.

A obtenção de uma matriz de dados por amostra representa dados de segunda ordem. O conjunto de várias amostras forma um cubo de dados. Existem duas maneiras distintas de obtenção deste tipo de dado: através de um único equipamento, por exemplo, espectrofluorímetro registrando matrizes de emissão e excitação, e através da hifenização de técnicas de primeira ordem, por exemplo, GC×GC. A principal vantagem de alguns dos métodos quimiométricos utilizados para construir modelos com dados de segunda ordem é a capacidade de determinação de espécies de interesse na presença de interferentes, mesmo que estes interferentes não estejam presentes na etapa de calibração, característica conhecida como “vantagem de segunda ordem”²². São exemplos de métodos quimiométricos de segunda ordem: análise de fatores paralelos (PARAFAC), resolução multivariada de curvas (MCR) e mínimos quadrados parciais multimodo (NPLS). Alguns destes métodos fornecem, após a otimização do modelo, os perfis característicos de cada composto puro presente na amostra fornecendo, dessa forma, informações qualitativas sobre o sistema. Outra vantagem destes métodos é o número de amostras requerido para a construção das regressões, que é menor que aquele necessário para modelos de primeira ordem. Uma alternativa para elaboração de modelos quimiométrico com dados de segunda ordem é o desdobramento da matriz de dados de cada amostra para um vetor, com isso o conjunto de várias amostras (vetores) formará uma matriz de dados, que pode ser analisada por métodos quimiométricos de primeira ordem²³.

Os métodos quimiométricos utilizados nesta tese no tratamento de dados obtidos por GC×GC-FID estão apresentados a seguir.

2.2. Mínimos Quadrados Parciais Multimodo (NPLS)

A regressão por mínimos quadrados parciais (PLS) é um método para construção de modelos quimiométricos de dados de primeira ordem entre variáveis independentes (\mathbf{X}), dados experimentais, e variáveis dependentes (\mathbf{Y}), que pode ser concentração de um ou mais compostos ou alguma outra propriedade que se deseja estudar. Estas matrizes são decompostas de acordo com as equações 1 e 2²⁴.

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (1)$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F} \quad (2)$$

onde \mathbf{X} ($i \times j$) é a matriz de dados obtida experimentalmente, \mathbf{Y} ($i \times n$) é a matriz da propriedade de interesse, \mathbf{T} ($i \times l$) e \mathbf{U} ($i \times l$) são matrizes de escores, \mathbf{P} ($j \times l$) e \mathbf{Q} ($n \times l$) são matrizes de pesos, \mathbf{E} ($i \times j$) e \mathbf{F} ($i \times n$) são matrizes de resíduos, i é o número de amostras, n é o número de compostos em estudo e l é o número de variáveis latentes.

O modelo PLS é obtido através de um processo iterativo, no qual as matrizes de escores \mathbf{T} e \mathbf{U} de cada componente são ajustadas para uma relação linear de modo que as matrizes \mathbf{X} e \mathbf{Y} apresentem a maior covariância possível.

O método dos mínimos quadrados parciais multimodo (NPLS) é uma extensão do PLS para dados de ordem dois ou mais e foi proposto por Bro²⁵ em 1996. Neste método o arranjo de dados \mathbf{X} ($i \times j \times k$) é decomposto em uma matriz de escores \mathbf{T} ($i \times l$) e duas matrizes de pesos \mathbf{W}_j ($j \times l$) e \mathbf{W}_k ($k \times l$) referentes ao segundo e terceiro modo de \mathbf{X} , respectivamente, enquanto \mathbf{Y} ($i \times n$) é decomposta em uma matriz de escores \mathbf{U} ($i \times l$) e uma matriz de pesos \mathbf{Q} ($n \times l$) (equações 3 e 4).

$$\mathbf{X} = \mathbf{TW}_j^T \mathbf{W}_k^T + \mathbf{E} \quad (3)$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F} \quad (4)$$

As matrizes de pesos e escores são ajustadas iterativamente de modo que \mathbf{X} e \mathbf{Y} apresentem a maior covariância possível. Uma relação linear entre \mathbf{T} e \mathbf{U} é estabelecida, onde os coeficientes de regressão \mathbf{b} são obtidos (equação 5).

Quando deseja-se determinar os valores da variável dependente de novas amostras, estas são decompostas segundo a equação 3, na qual a matriz **T** é obtida. Em seguida, **U** é obtida através da equação 5 e, finalmente, através da equação 4 obtêm-se os valores de **Y**²⁶.

$$\mathbf{U} = \mathbf{bT} + \mathbf{E} \quad (5)$$

Apesar do NPLS não apresentar vantagem de segunda ordem, ou seja, amostras desconhecidas não podem apresentar interferentes ausentes nas amostras de calibração, ele é muito utilizado para realizar quantificação em dados de segunda ordem, uma vez que a informação da variável dependente é incluída no ajuste do modelo e o método trata os arranjos multimodo sem a necessidade de desdobrá-los. Outra característica muito importante do NPLS é o fato do algoritmo conseguir trabalhar bem na presença de pequenos desvios na trilinearidade dos dados, o que é muito comum em GC×GC-FID, onde pequenas variações nos tempos de retenção são observadas entre as amostras.

2.3. Análise de Fatores Paralelos (PARAFAC)

No início dos anos 70, o pesquisador na área de lingüística Harshman²⁷ desenvolveu a análise de fatores paralelos (PARAFAC). Paralelamente Carrol e Chang²⁸ propuseram um modelo idêntico com o nome de decomposição canônica (CANDECOMP), mas este algoritmo só começou a ser largamente utilizado pelos químicos após 1997, quando Bro²⁹ publicou um trabalho descrevendo o PARAFAC e disponibilizou o algoritmo para download gratuito. A principal vantagem do PARAFAC é o fato de apresentar vantagem de segunda ordem.

O modelo PARAFAC efetua a decomposição do cubo de dados contendo *i* amostras **X** (*i* x *j* x *k*) utilizando *f* fatores em uma matriz de escores **A** (*i* x *f*), duas matrizes de pesos **B** (*j* x *f*) e **C** (*k* x *f*) e um tensor de resíduos **E** (*i* x *j* x *k*) (equação 6)²⁹. Na figura 7 está ilustrado graficamente o PARAFAC.

$$\mathbf{X} = \mathbf{A}\mathbf{B}^T\mathbf{C}^T + \mathbf{E} \quad (6)$$

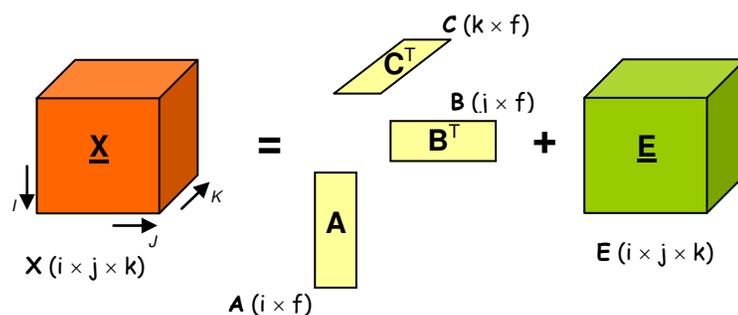


Figura 7. Representação gráfica do PARAFAC.

Os seguintes aspectos devem ser considerados na utilização do PARAFAC: o conjunto de dados deve apresentar uma estrutura trilinear; a inicialização do algoritmo; as restrições impostas em cada dimensão do modelo; o valor de convergência e o número de componentes ou fatores²⁹.

O algoritmo utilizado para resolução por PARAFAC é o dos mínimos quadrados alternantes (ALS), que inicializa assumindo uma estimativa inicial para os pesos nos modos \mathbf{B} e \mathbf{C} e usa-os para estimar os valores de \mathbf{A} . Essa estimativa inicial pode ser realizada por diferentes métodos, sendo os mais utilizados, decomposição em valores singulares (SVD) ou decomposição trilinear direta (DTLD). Uma vez que uma estimativa inicial de \mathbf{B} e \mathbf{C} é obtida, o algoritmo estima \mathbf{A} e prossegue até atingir um critério de convergência ou um número de iterações previamente definidos. Após cada iteração, o ALS promove uma melhora no ajuste do modelo aos dados, sendo essa sua principal vantagem, mas o tempo requerido para isso pode ser longo em alguns casos. Em algumas situações, as restrições de não-negatividade, unimodalidade e/ou ortogonalidade podem ser usadas em cada modo para obter uma solução com um significado químico mais consistente. Os passos do algoritmo para um modelo PARAFAC podem ser resumidos da seguinte maneira³⁰:

- (1) Inicialização de \mathbf{B} e \mathbf{C}
- (2) Estimativa de \mathbf{A} por um ajuste de mínimos quadrados a partir de \mathbf{X} , \mathbf{B} e \mathbf{C}
- (3) Estimativa de \mathbf{B} de maneira análoga.
- (4) Estimativa de \mathbf{C} de maneira análoga.

(5) Volta ao passo (2) até o critério de convergência ser atingido.

Uma etapa muito importante na utilização do PARAFAC é a escolha do número de fatores ou componentes, pois uma escolha errônea pode gerar interpretações falsas do sistema em estudo. Essa escolha pode ser feita de acordo com o conhecimento prévio do sistema, na variância explicada pelo modelo ou em validação cruzada³¹. Em 2003, Bro e Kiers³² propuseram o teste de consistência trilinear (CORCONDIA) para uma estimativa correta do número de fatores. O CORCONDIA é baseado na interpretação do PARAFAC como um modelo de Tucker3 restrito, no qual um modelo é considerado válido quando os elementos da superdiagonal do tensor **G** são próximos a 1 e os demais próximos a 0 (equação 7). O método quimiométrico de Tucker3 foi desenvolvido na década de 60 por L. Tucker³³ e foi o método a partir do qual o PARAFAC foi elaborado.

$$CORCONDIA(\%) = 100 \times \left(1 - \frac{\sum_{d=1}^F \sum_{e=1}^F \sum_{f=1}^F (g_{def} - t_{def})^2}{\sum_{d=1}^F \sum_{e=1}^F \sum_{f=1}^F (t_{def})^2} \right) \quad (7)$$

onde g_{def} é o elemento do tensor central **G** calculado com o Tucker3 a partir dos pesos do PARAFAC e t_{def} é o elemento de um tensor contendo 1 na superdiagonal e 0 nos demais elementos. O modelo PARAFAC ideal é válido se o valor de CORCONDIA for próximo a 100%. Dependendo do caso, valores de CORCONDIA até 50% são aceitáveis. O CORCONDIA é uma ferramenta para indicar sobreajuste em um modelo PARAFAC, mas não sub-ajuste. Um modelo construído com número inferior ao real de fatores apresentará um valor de CORCONDIA próximo a 100%. Portanto, não deve-se utilizar o CORCONDIA como única ferramenta na escolha do número correto de fatores. Ele pode ser utilizado em conjunto com a variância explicada para cada fator, para assim não gerar modelos PARAFAC sub ou sobreajustados.

Uma vez obtidos os perfis **A**, **B** e **C**, a identificação de cada analito presente nas amostras é feita pela comparação entre os perfis obtidos e perfis experimentais conhecidos obtidos com padrões. Para quantificação utilizando-se PARAFAC, uma curva analítica é construída pelo analista e é obtida através de uma regressão por mínimos quadrados entre as concentrações de referência da

espécie de interesse nas amostras de calibração e os valores de escores obtidos pelo modelo para esta espécie.

2.4. Resolução Multivariada de Curvas (MCR)

O método de resolução multivariada de curvas (MCR) possui vantagem de segunda ordem e foi desenvolvido por Tauler e colaboradores³⁴ em 1995, apresentado o pré-requisito de que em um conjunto de dados a resposta do sistema seja linear em relação à quantidade de analito. Os principais objetivos do MCR são o isolamento, a resolução e a quantificação das fontes de variação do conjunto de dados. O modelo bilinear MCR decompõe uma matriz de dados de uma amostra de acordo com a equação 8.

$$\mathbf{D} = \mathbf{CS}^T + \mathbf{E} \quad (8)$$

onde \mathbf{D} ($i \times j$) é a matriz de dados obtida experimentalmente, a matriz \mathbf{C} ($i \times f$) é geralmente associada com os perfis de concentração, \mathbf{S} ($j \times f$) é a matriz com os perfis experimentais puros recuperados para cada analito, \mathbf{E} é a matriz de resíduos e f o número de componentes. Vale ressaltar que um modelo MCR pode ser construído para a quantificação de uma mistura e não apenas de um analito, como no caso da quantificação de óleos essenciais, sendo assim, os perfis experimentais recuperados na matriz \mathbf{S} serão referentes a cada mistura pura presente no conjunto de amostras.

Os parâmetros do modelo são estimados utilizando-se mínimos quadrados alternantes (ALS), que iterativamente ajusta as matrizes \mathbf{C} e \mathbf{S} ao conjunto de dados \mathbf{D} , utilizando um número de fatores f pré-definidos e uma estimativa inicial de \mathbf{C} ou \mathbf{S} , a qual pode ser obtida de diferentes maneiras, entre elas, análise de fatores evolucionários (EFA)³⁵, decomposição de valores singulares (SVD) e conhecimento prévio do sistema. O número de analitos pode ser estimado utilizando o conhecimento prévio do sistema em estudo ou a partir dos resultados de análise por SVD da matriz de dados constituída das amostras utilizadas na etapa de calibração³⁴.

Uma importante característica do MCR-ALS é a possibilidade de sua extensão para análises com múltiplas amostras. Neste caso, as matrizes são organizadas de uma forma adjunta, uma abaixo da outra ou uma ao lado da outra dependendo das estimativas iniciais serem os perfis de concentração ou experimentais. Quando as estimativas iniciais são os perfis experimentais, os dados são arranjados segundo a equação 9.

$$\mathbf{D}_k = \mathbf{C}_k \mathbf{S}^T + \mathbf{E} \iff \begin{bmatrix} \mathbf{D}_1 \\ \vdots \\ \mathbf{D}_k \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1 \\ \vdots \\ \mathbf{C}_k \end{bmatrix} \mathbf{S}^T + \begin{bmatrix} \mathbf{E}_1 \\ \vdots \\ \mathbf{E}_k \end{bmatrix} \quad (9)$$

onde \mathbf{D}_k corresponde a um conjunto de k amostras arranjadas de maneira bidimensional, \mathbf{C}_k são os perfis de concentração para cada k amostra e \mathbf{E}_k seus respectivos resíduos.

A imposição de uma ou mais restrições ao MCR, entre elas, não-negatividade, unimodalidade, posto local, trilinearidade e igualdade, pode ser utilizada para ajudar na convergência do algoritmo, resolver problemas de ambigüidade rotacional e garantir que os resultados apresentem significado químico. A restrição de trilinearidade consiste em forçar que os perfis instrumentais e de concentração não variem de amostra para amostra, ou seja, no caso de análise cromatográfica não poderiam ocorrer variações nos tempos de retenção. Na restrição de igualdade, o perfil instrumental ou de concentração é fixado durante a otimização do modelo, esta restrição pode ser aplicada no caso de conhecer previamente o perfil experimental ou de concentração do sistema em estudo. A restrição de posto local é utilizada quando se tem conhecimento prévio que determinado analito está ausente em uma ou mais amostras. Neste caso, esta informação é passada ao modelo fazendo com que o perfil de concentração deste analito seja igual a zero. Esta restrição é muito útil nos casos onde amostras contendo interferentes são decompostas conjuntamente como amostras de calibração, onde se requer que o perfil de concentração do interferente deva ser zero³⁶.

Comparativamente ao PARAFAC, o MCR permite a imposição de mais restrições, com isso o MCR pode ser otimizado mais adequadamente para o tipo de dados em questão. Entretanto, a otimização dessas restrições requer um maior conhecimento sobre o sistema e, dependendo do caso, pode requerer um longo tempo de estudo do analista.

Depois de realizada a decomposição do conjunto de dados, o processo de calibração é feito através de regressão linear entre os perfis recuperados que contém a informação relacionada à concentração da espécie de interesse e sua respectiva concentração.

2.5. Métodos para Seleção de Variáveis

Em alguns casos, não é toda a informação obtida experimentalmente que está relacionada com a variável dependente, mas somente uma parte dela. Para esses casos, existem métodos quimiométricos que foram desenvolvidos para realizar a seleção de variáveis ou intervalos da informação experimental para a construção do modelo de calibração. Nesta tese, foram utilizados o método dos mínimos quadrados parciais por sinergismo de intervalo (siPLS) e algoritmo genético (GA).

2.5.1. Mínimos Quadrados Parciais por Sinergismo de Intervalos (siPLS)

O método dos mínimos quadrados parciais por intervalos (iPLS) foi proposto por Norgaard e colaboradores³⁷ em 2000. No iPLS o conjunto de dados experimental é inicialmente dividido em um número de intervalos fornecidos pelo analista. Em seguida, é construído um modelo de calibração com cada um dos intervalos e o intervalo que forneça o modelo com menor erro é selecionado. Desta forma é avaliada a relevância da informação nos diferentes intervalos dos dados e as regiões, cujas variáveis apresentam menor importância e/ou estão mais sujeitas a ruídos, são removidas.

A seleção do melhor intervalo é realizada através do cálculo da raiz quadrada do erro quadrático médio de validação cruzada (RMSECV) (equação 10)

para cada intervalo, sendo selecionado o intervalo que apresenta o menor valor de RMSECV.

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n-1}} \quad (10)$$

onde que \hat{y}_i e y_i são os valores previstos e de referência para a variável dependente, respectivamente e n é o número de amostras de calibração.

Em 2001, Munck e colaboradores³⁸ propuseram o método dos mínimos quadrados por sinergismo de intervalos (siPLS), que consiste no mesmo princípio do iPLS, mas neste caso em vez de avaliar cada intervalo isoladamente, combinações de 2, 3 ou 4 intervalos, número definido pelo analista, são realizadas e a combinação com menor RMSECV é selecionada para a construção do modelo de calibração.

2.5.2. Algoritmo Genético

O algoritmo genético (GA) é um método para seleção de variáveis baseado nas teorias evolutivas de Darwin, na qual o mais adaptado sobrevive. Neste algoritmo, o primeiro passo é a criação de uma população randômica inicial, em que cada variável é codificada nos valores de 0 ou 1 e são chamadas de genes. Nesta codificação, 0 (zero) significa que a variável não foi incluída no modelo e 1 (um) significa que a variável foi incluída no modelo. O conjunto de vários genes forma um cromossomo e o conjunto de vários cromossomos forma uma população³⁹. Na figura 8 está exemplificado a formação de uma população, contendo 3 cromossomos, através da codificação de cada variável de um cromatograma, ou seja, se o cromatograma apresenta 50000 pontos (variáveis) o cromossomo apresentará 50000 genes com valor de 0 ou 1.

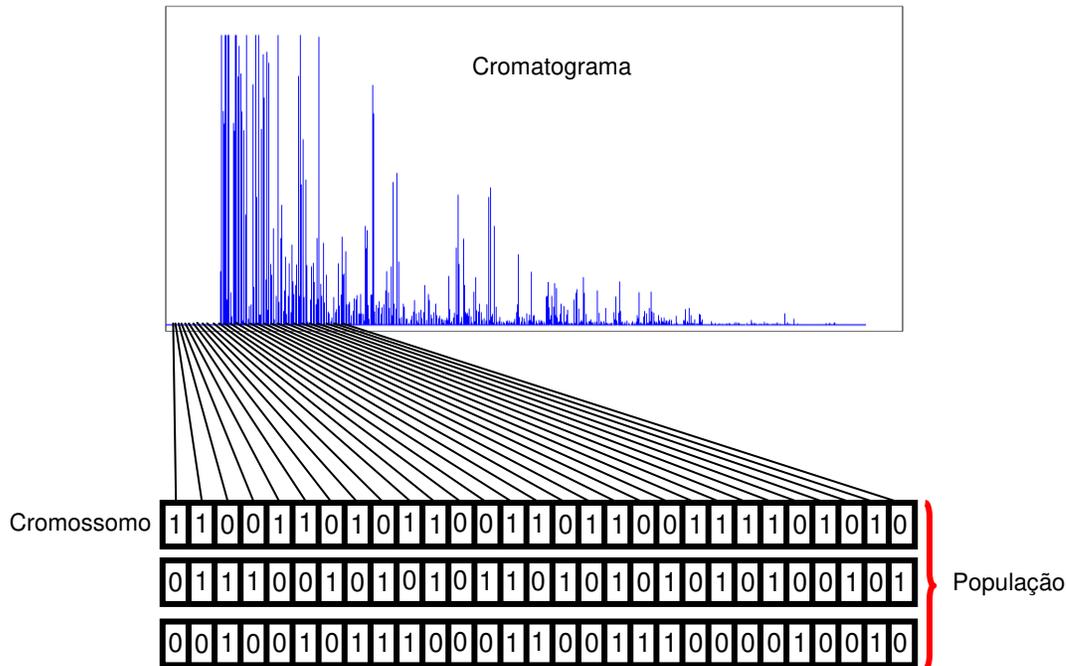


Figura 8. Codificação de cada gene (variável) de um cromatograma em 0 ou 1. O conjunto de vários genes forma um cromossomo (indivíduo) e o conjunto de vários cromossomos forma uma população.

Segundo a teoria da evolução, somente os indivíduos mais adaptados da população sobrevivem e geram descendentes, transmitindo assim sua herança genética para os próximos indivíduos. Isso é observado no GA, no qual indivíduos mais adaptados são gerados através de seleção, cruzamento e mutação⁴⁰.

A seleção consiste em selecionar os cromossomos mais adaptados da população inicial para as próximas etapas do algoritmo. Isso significa que os cromossomos que apresentam valores de erro menores que a média da população para RMSECV são selecionados para as próximas etapas do GA⁴¹.

Na etapa de cruzamento, dois cromossomos trocam parte de seu material genético para produzir novos cromossomos. O cruzamento pode ser simples ou duplo (figura 9). No cruzamento simples os genes dos indivíduos envolvidos são separados em um ponto aleatório em dois segmentos, em seguida, a primeira parte de cada um dos cromossomos é trocada com o outro cromossomo, gerando assim dois novos indivíduos. O cruzamento duplo ocorre de maneira similar ao

simples, mas, neste caso, dois pontos aleatórios são selecionados em cada cromossomo e o fragmento entre esses pontos é trocado⁴².

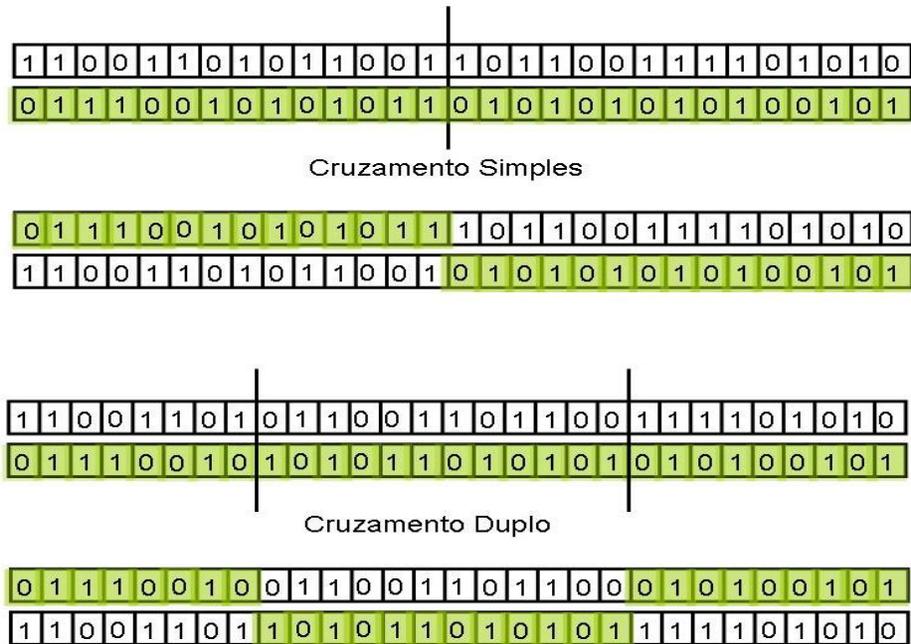


Figura 9. Etapa de cruzamento simples e duplo do GA.

Após a adição de novos indivíduos à população, ocorre a etapa de mutação, na qual qualquer gene pode ter seu valor de 0 ou 1 alterado. O objetivo da mutação é evitar a possibilidade de um gene com mesmo valor em todos os cromossomos, pois, neste caso, nenhum cruzamento possível poderia alterar esse gene.

O GA pode ser resumido nos seguintes passos:

1. Criação aleatória de uma população inicial de indivíduos;
2. Avaliação de cada indivíduo;
3. Descarte de metade dos indivíduos apresentando pior ajuste;
4. Cruzamento dos indivíduos remanescentes;
5. Mutação;
6. Repetição dos passos 2 a 5 até que o critério de parada seja alcançado;

CAPÍTULO 3:

Objetivo Geral

3. Objetivo Geral

Esta tese tem como objetivo principal o emprego de métodos quimiométricos para determinações quantitativas e qualitativas em diferentes tipos de matrizes realizadas em um GC×GC-FID construído em laboratório.

CAPÍTULO 4:

Análises de Amostras de Gasolina

4. Análises de Amostras de Gasolina

4.1. Introdução

A gasolina automotiva é um combustível derivado do petróleo constituído, principalmente, por uma complexa mistura de hidrocarbonetos na faixa de C_4 a C_{12} e com ponto de ebulição na faixa de 30 – 225 °C. Além disso, estão presentes na gasolina em baixas concentrações compostos oxigenados, compostos contendo enxofre, nitrogênio e metais⁴³. No Brasil, a gasolina utilizada como combustível em motores de combustão interna contém 25% de etanol anidro e é chamada de gasolina tipo C.

O monopólio estatal sobre o refino de petróleo e a distribuição de combustíveis foi extinto em 1995 no Brasil, desde então, muitas companhias de distribuição foram abertas, propiciando assim uma ampla concorrência neste mercado. Esta abertura do mercado não só aumentou a concorrência, provocando uma queda nos preços, mas também propiciou a oportunidade para algumas empresas de praticar a adulteração de combustíveis, principalmente de gasolina. A adição de solventes é a prática de adulteração mais comum encontrada no Brasil devido à grande diferença no preço da gasolina e destes solventes, principalmente por causa dos valores dos impostos de cada um. Os solventes mais utilizados na adulteração de gasolina são: etanol (em quantidade acima da estabelecida na lei), hidrocarbonetos alifáticos leves (C_4 - C_8), hidrocarbonetos alifáticos pesados (C_{13} - C_{20}), hidrocarbonetos aromáticos e em alguns casos até água⁴⁴. Essa adulteração na gasolina causa uma série de problemas, entre eles, aumento da emissão de poluentes, desempenho prejudicado e vida útil reduzida do motor e perdas na arrecadação dos impostos, pois alguns dos solventes utilizados são obtidos por meio de contrabando⁴⁵.

O controle de qualidade de combustíveis no Brasil é realizado pela Agência Nacional de Petróleo, Gás Natural e Biocombustíveis (ANP), que em associação com agências de inspeções e laboratórios de universidades tem tomado várias providências na tentativa de diminuir ou até eliminar o problema de adulteração de combustíveis. A ANP realiza mais de 20 testes físico-químicos baseados nos

procedimentos das Normas Brasileiras (NBR) e Métodos Brasileiros (MB) da Associação Brasileira de Normas Técnicas (ABNT) ou da *American Society for Testing Materials* (ASTM) para avaliar a qualidade da gasolina. Entre os testes realizados estão a determinação da massa específica, a temperatura de destilação a 10, 50, 90% v/v de destilado, a temperatura final de destilação, a octanagem MON e ROM e a quantidade de benzeno⁴⁶. A maioria dos testes requer grandes quantidades de amostras, profissionais qualificados e procedimentos que envolvem operações manuais, as quais são tediosas e podem levar a erros operacionais. Além disso, nem sempre é possível identificar adulteração em gasolina não conforme através destes métodos, pois a maioria dos solventes utilizados com esse propósito apresenta composição similar à gasolina e, conseqüentemente, não são flagrados por testes físico-químicos⁴⁷. Outro problema encontrado na fiscalização realizada pela ANP é o baixo número de amostras que são analisadas por ano. Por exemplo, em 2009 a ANP realizou fiscalização em 28565 postos, dos quais 824 foram flagrados vendendo gasolina adulterada, número baixo considerando que somente na grande São Paulo existem mais de 70000 postos autorizados a vender combustíveis. Embora o número de postos de gasolina fiscalizados ainda seja baixo, desde 2005 tem-se observado um aumento na quantidade dos postos fiscalizados e também uma queda na porcentagem de postos flagrados vendendo gasolina adulterada. Em 2005, essa porcentagem era de 7,2% e, em 2009, caiu para 2,9%⁴⁸.

Recentemente, a ANP introduziu marcadores em todos os solventes utilizados no Brasil, nacionais ou importados. Deste modo, a adição de solventes “marcados” à gasolina seria facilmente detectável por métodos analíticos específicos. Esta tecnologia tem-se mostrada eficiente, mas consome milhares de dólares anualmente e apenas alguns laboratórios são capazes de detectar a presença dos marcadores na gasolina e solventes adquiridos ilegalmente não apresentam os marcadores⁴⁹.

Devido aos problemas mencionados anteriormente, muitos estudos têm sido realizados para avaliação da qualidade de gasolina, sendo a cromatografia gasosa (GC) a principal técnica utilizada^{44,45,50,51,52,53}. Em 2008, Flumignan e

colaboradores⁵⁴ coletaram 2400 amostras de gasolinas em postos de combustíveis brasileiros e construíram modelos de calibração utilizando cromatografia gasosa (GC) e mínimos quadrados parciais (PLS) para previsão de alguns dos parâmetros físico-químicos utilizados pela ANP, dentre eles, massa específica e análise da temperatura de destilação da gasolina em 10%, 50%, 90% de destilado e ponto final. Os autores obtiveram valores de RMSEP de 4,3 kg/m³ para massa específica e 3,3, 5,8, 10,1 e 11,3 °C para temperatura a 10%, 50%, 90% de destilado e ponto final, respectivamente.

Em 2009, Ré-Poppi e colaboradores⁵⁵ realizaram análises por GC-FID de gasolinas obtidas em cinco distribuidores diferentes e através de análise hierárquica de agrupamentos (HCA) identificaram possíveis amostras adulteradas.

Em 2005, Micyus e colaboradores⁵⁶ utilizaram a GC×GC-FID para realizar a quantificação de compostos aromáticos em gasolina. A determinação destes compostos só foi possível devido ao maior poder de separação da GC×GC-FID em relação à GC-FID, uma vez que na GC-FID eles apresentaram sobreposição com outros compostos da amostras.

A polícia federal brasileira tem flagrado um intenso contrabando de gasolina venezuelana para o Brasil nos últimos anos. Muitos postos de combustíveis brasileiros localizados próximos à fronteira Brasil-Venezuela tem vendido gasolina venezuelana adquirida por esse contrabando, que é fortemente impulsionado devido à enorme diferença de preço e também devido à dificuldade de fiscalização na fronteira Brasil-Venezuela. Em abril de 2011, na Venezuela a gasolina era vendida por R\$0,17 o litro, enquanto que nas cidades brasileiras do estado de Roraima próximas à fronteira o preço médio do litro de gasolina saía por R\$3,00.

O objetivo nesta etapa do trabalho foi combinar o maior poder de separação de GC×GC-FID em relação à GC-FID com as vantagens dos métodos quimiométricos para análises de gasolinas. Construiu-se modelos quimiométricos para identificar adulteração em amostras de gasolina, para a previsão de alguns dos testes realizados pela ANP e também para diferenciar gasolinas brasileiras de gasolinas venezuelanas.

4.2. Desenvolvimento de um Modelo Quimiométrico para Identificar Adulteração em Gasolina

4.2.1. Objetivos

Elaboração de um modelo quimiométrico para identificar amostras adulteradas de gasolina, utilizando para isso os adulterantes: etanol, aguarrás, thinner e querosene.

4.2.2. Parte Experimental

Utilizou-se um sistema GC×GC-FID construído em laboratório¹⁰ a partir de um HP-5890 Series II GC-FID (Hewlett-Packard, Wilmington, DE, EUA) equipado com injetor do tipo split-splitless. Neste sistema usou-se um modulador criogênico de quatro jatos, sendo N₂ resfriado em nitrogênio líquido como fluido criogênico e N₂ aquecido como gás quente. Tanto o fluxo resfriante como o de aquecimento foram controlados por duas válvulas solenóides Asco (Asco Joucomatic, Rueil Malmaison, França). O controle das válvulas e a digitalização do sinal analógico do FID foram realizados pelo controlador A/D de 16 bits DAQPad-6015 (National Instruments Corporate Headquarters, Austin, Texas, EUA) conectado a um microcomputador AMD Athlon 64 X2 4600 Dual Core com 4Gb de memória RAM. O controle das válvulas e a coleta a 100 Hz do sinal do FID foram realizadas em LABView v.8.2 (National Instruments Corporate Headquarters, Austin, Texas, EUA).

O conjunto de colunas utilizado foi uma HP-5 (30 m x 0,25 mm x 0,25 µm) conectada a uma DB-Wax (1,0 m x 0,1 mm x 0,1 µm). As temperaturas do injetor e do detector foram 250 °C. A programação de temperatura do forno cromatográfico foi: 40 °C (3 min.) até 160 °C (4 °C/min.) e então aquecido até 240 °C (20 °C/min.) mantendo-se por 3 min. nesta temperatura. Como gás de arraste utilizou-se hidrogênio a uma vazão de 0,52 mL/min. Em todas as corridas cromatográficas o volume de injeção foi de 0,6 µL no modo *split* (*split ratio* 1:300). O período de modulação foi de 6 s.

A análise multivariada dos dados cromatográficos foi realizada em MATLAB 6.5 (MathWorks, Natick – MA, EUA) instalado em um microcomputador AMD Athlon 64 X2 4600 Dual Core e 2Gb de memória RAM. O algoritmo NPLS foi obtido do *N-way Toolbox 2.11*⁵⁷.

Uma gasolina certificada como não adulterada pela ANP foi utilizada na preparação das amostras de calibração e de validação. O conjunto de calibração foi preparado utilizando esta amostra, a qual foi adulterada em nosso laboratório através da adição dos seguintes solventes: thinner, aguarás e querosene comerciais. O planejamento experimental empregado na composição das amostras de calibração foi *Greco-Latin Squares*⁵⁸, o qual foi escolhido por apresentar cinco níveis de concentrações diferentes para cada solvente em apenas 25 amostras de calibração. Para avaliar o modelo, preparou-se também 14 amostras de validação (V1 a V14) de gasolina adulterada em laboratório utilizando a mesma gasolina da etapa de calibração. A composição em % v/v das amostras de calibração e das amostras de validação estão na tabela 1, onde vale ressaltar que para as amostras que apresentam etanol, esta concentração corresponde a quantidade de etanol adicionada além da já existente na gasolina C (25% v/v). Em seguida, foram analisadas diferentes amostras de gasolina, sendo 4 não adulteradas e 9 amostras adulteradas.

Tabela 1. Composição das amostras de calibração e das amostras de validação em % v/v. As amostras de validação estão identificadas pela letra V.

Amostra	Etanol	Thinner	Aguarrás	Querosene	Gasolina C
1	0	0	0	0	100
2	0	5	30	10	55
3	0	0	30	30	40
4	0	5	0	20	75
5	0	20	5	0	75
6	0	0	10	10	80
7	0	20	20	10	50
8	0	5	10	0	85
9	0	30	0	10	60
10	0	30	10	30	30
11	0	10	20	30	40
12	0	20	0	30	50
13	0	0	5	5	90
14	0	10	5	10	75
15	0	5	20	5	70
16	0	10	10	20	60
17	0	10	0	5	85
18	0	0	20	20	60
19	0	30	20	0	50
20	0	20	10	5	65
21	0	10	30	0	60
22	0	20	30	20	30
23	0	30	5	20	45
24	0	5	5	30	60
25	0	30	30	5	35
V1	0	5	10	0	85
V2	0	10	0	5	85
V3	0	0	5	10	85
V4	5	0	0	0	95
V5	15	0	0	0	85
V6	30	0	0	0	70
V7	0	5	12	6	77
V8	0	10	10	10	70
V9	0	15	20	10	55
V10	10	11	11	13	55
V11	50	0	0	0	50
V12	20	13	22	15	30
V13	0	21	10	7	62
V14	5	0	0	20	75

4.2.3. Resultados e Discussão

Para a construção do modelo de calibração para quantificação de gasolina foram utilizadas 25 amostras de calibração, 14 amostras de validação adulteradas com thinner, aguarrás, querosene e/ou etanol preparadas em laboratório e 13 amostras fornecidas pela ANP (9 adulteradas e 4 não adulteradas). Na figura 10 está representado o cromatograma obtido para gasolina C não adulterada, onde, através da estruturação cromatográfica da GC×GC, foi possível identificar compostos e classes de compostos. À direita da figura 10 está mostrada a barra de cores referente à intensidade dos sinais no cromatograma GC×GC, sendo a cor vermelha referente aos picos mais intensos. A barra de cores será omitida das figuras subseqüentes, porém o padrão de cores será mantido. A figura 11 apresenta os cromatogramas obtidos para os solventes utilizados e para gasolina C, na qual pode ser visto que a gasolina e os adulterantes apresentam sinais comuns, o que mostra a necessidade do uso de ferramentas quimiométricas para elaboração do modelo.

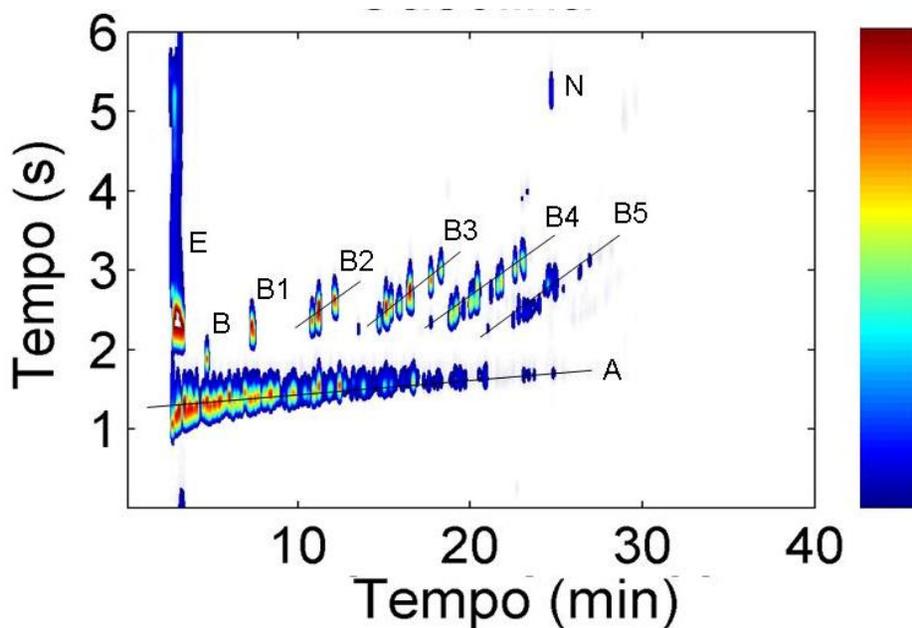


Figura 10. Diagrama de cores do GCxGC-FID obtido para gasolina C. E= etanol, B = benzeno, B1 = tolueno, B2 = benzeno C2 substituído, B3 = benzeno C3 substituído, B4 = benzeno C4 substituído, B5 = benzeno C5 substituído, N = naftaleno, A = alcanos, cicloalcanos e alcenos. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 μ m) + DB-Wax (1,0 m x 0,1 mm x 0,1 μ m). Programação de temperatura do forno cromatográfico: 40 °C (3 min.) até 160 °C (4 °C/min.) e então aquecido até 240 °C (20 °C/min.) (3 min.). Razão de divisão do injetor 1:300, volume de injeção 0,6 μ L e PM 6 s.

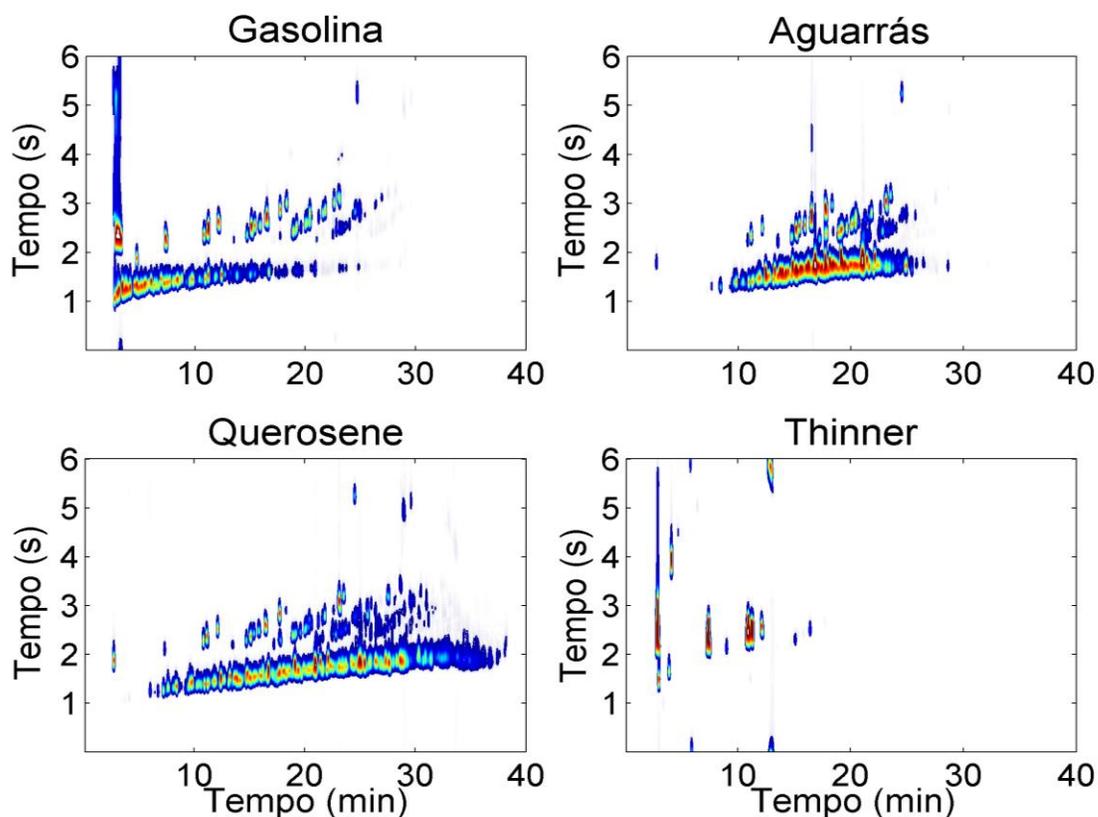


Figura 11. Cromatogramas GC×GC-FID obtidos para gasolina C, aguarrás, querosene e thinner. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 μ m) + DB-Wax (1,0 m x 0,1 mm x 0,1 μ m). Programação de temperatura do forno cromatográfico: 40 °C (3 min.) até 160 °C (4 °C/min.) e então aquecido até 240 °C (20 °C/min.) (3 min.). Razão de divisão do injetor 1:300, volume de injeção 0,6 μ L e PM 6 s.

Foram construídos modelos de calibração NPLS com diferentes números de variáveis latentes para cada adulterante e para a gasolina utilizando-se validação-cruzada “leave-one-out”, que consiste em retirar uma amostra do conjunto de calibração, construir o modelo com as restantes e prever o valor desta amostra que havia sido retirada. Este procedimento é repetido até que todas as amostras de calibração sejam retiradas uma vez. O número de variáveis latentes utilizado em cada modelo foi determinado pelo menor valor da raiz quadrada do erro quadrático médio de validação cruzada (RMSECV) em cada caso. A tabela 2

apresenta o valor de RMSECV obtido para cada modelo com os diferentes números de variáveis latentes; para o modelo de calibração de aguarás foram utilizadas três variáveis latentes, para thinner quatro e para gasolina e querosene cinco⁵⁹.

Tabela 2. RMSECV (%) obtido em função do número de variáveis latentes.

Variáveis Latente	1	2	3	4	5	6
Thinner	11,5	5,8	3,2	2,7	2,9	3,0
Aguarás	9,4	6,3	2,8	2,8	3,7	3,7
Querosene	10,3	4,4	3,1	3,0	2,8	3,0
Gasolina C	23,3	12,5	6,9	6,0	5,7	5,9

Na Figura 12 estão os gráficos das concentrações previstas pelo modelo das amostras de calibração versus o valor real de concentração para cada adulterante e para a gasolina, no qual bons coeficientes de regressão foram obtidos (thinner = 0,968, aguarás = 0,982, querosene = 0,991 e gasolina = 0,985).

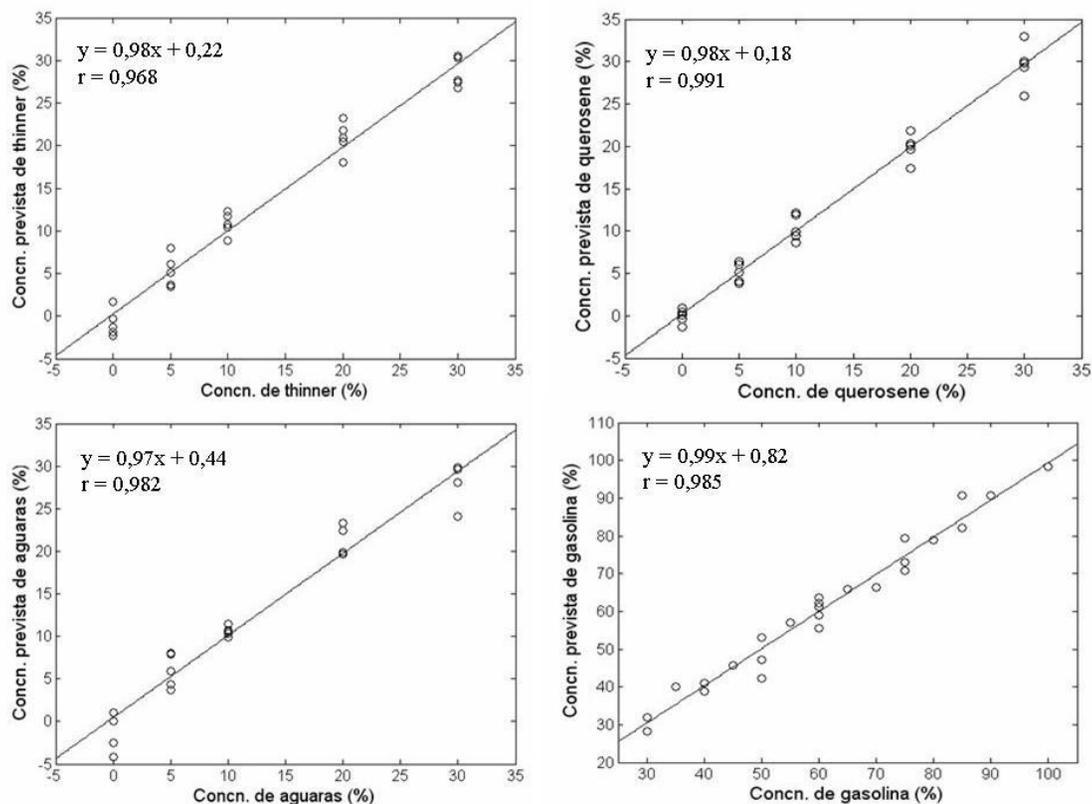


Figura 12. Concentrações previstas por validação cruzada versus concentrações das amostras de calibração para os adulterantes e para gasolina C.

Depois de construído o modelo de calibração NPLS para gasolina e para os adulterantes, realizou-se a previsão das amostras do conjunto de validação. Todas as concentrações previstas estão mostradas na tabela 3, onde é possível observar uma boa concordância entre os valores previstos e os valores reais.

Algumas das amostras de validação foram adulteradas com etanol para testar a resposta do modelo a este tipo de solvente, uma vez que a gasolina do tipo C apresenta até 25% de etanol em sua composição e a adulteração mais comum em nosso país é com etanol em quantidades acima da especificada na legislação. Quando uma amostra está adulterada com 5% v/v de etanol (amostra V4), além dos 25% v/v já presente na gasolina C, o modelo não conseguiu identificar a amostra como adulterada. Já para a amostra V6, contaminada com 30% v/v de etanol, além dos 25% v/v já presente, o modelo apresentou um erro positivo de 10% e para a amostra V11 com 50% v/v de etanol, além dos 25% v/v

já presente, o valor de gasolina previsto apresentou uma resposta com erro positivo de 18%, mas em ambos os casos o modelo identificou as amostras como adulteradas, pois valores muito distantes de 100% para concentração de gasolina foram obtidos. Testou-se também se a presença de etanol prejudicaria a análise dos outros adulterantes analisados, nos três casos (amostras V10, V12 e V14) a resposta do modelo foi concordante com a concentração das amostras, inclusive para gasolina. Portanto, pode-se concluir que na presença de etanol em até 5% v/v acima da quantidade estabelecida por lei, o modelo não conseguiu identificar a amostra como adulterada, para concentração de etanol acima de 5% e abaixo de 30% o modelo previu corretamente a concentração de gasolina e dos outros solventes, já para concentração acima de 30% o modelo apresentou uma resposta com erro positivo na concentração de gasolina. Como a adulteração de gasolina com etanol é facilmente identificada por uma extração líquido-líquido (norma NBR número 13992), os erros encontrados em algumas concentrações de etanol não são relevantes para o modelo de calibração em estudo⁵⁹.

Tabela 3. Valores previstos pelo modelo e entre parênteses a quantidade adicionada em % v/v.

Amostra	Thinner	Aguarrás	Querosene	Gasolina C
V1	5 (5)	12 (10)	1 (0)	91 (85)
V2	12 (10)	0 (0)	4 (5)	86 (85)
V3	3 (0)	8 (5)	9 (10)	88 (85)
V4	<0 (0)	<0 (0)	0 (0)	100 (95)
V5	<0 (0)	<0 (0)	<0 (0)	86 (85)
V6	<0 (0)	<0 (0)	<0 (0)	80 (70)
V7	8 (5)	12 (12)	6 (6)	79 (77)
V8	12 (10)	10 (10)	7 (10)	75 (70)
V9	17 (15)	21 (20)	10 (10)	54 (55)
V10	12 (11)	8 (11)	9 (13)	57 (55)
V11	1 (0)	0 (0)	<0 (0)	68 (50)
V12	8 (13)	18 (22)	13 (15)	30 (30)
V13	20 (21)	8 (10)	3 (7)	62 (62)
V14	1 (0)	1 (0)	22 (20)	79 (75)

Os valores da raiz quadrada do erro quadrático médio de previsão (RMSEP) e os coeficientes de correlação obtidos entre valores de concentração versus valores previstos obtidos para cada modelo estão na tabela 4. Os baixos valores obtidos de RMSECV e RMSEP e a boa correlação obtida entre os valores conhecidos versus os previstos validam os modelos de calibração construídos (tabelas 2, 3 e 4).

Tabela 4. Valores de RMSEP e coeficiente de correlação obtido em cada modelo.

Modelo	Variáveis Latentes	RMSEP^a (%)	r^a
Thinner	4	2,9	0,938
Aguarrás	3	3,1	0,936
Querosene	5	2,4	0,959
Gasolina C	5	3,5	0,996

^a Calculado retirando-se as amostras V6 e V11.

Após validado o modelo para previsão da concentração de gasolina, foi realizada a análise de novas amostras de gasolina certificadas pela ANP como adulteradas ou não adulteradas. Como pode ser observado na tabela 5, para todas as amostras classificadas como adulteradas, os valores previstos pelo modelo ficaram abaixo de 85% v/v de gasolina, concordando assim com os resultados obtidos pela ANP. No caso das amostras certificadas como não adulteradas pela ANP, obteve-se valores de previsão pelo modelo acima de 95% v/v de gasolina em três casos, concordando com os resultados obtidos pela ANP, mas para a amostra N3 obteve-se 44% v/v de gasolina, discordando assim do resultado obtido pela ANP. Os cromatogramas obtidos para a gasolina padrão (utilizada na calibração) e para a amostra N3 apresentam grandes diferenças (figura 13), uma vez que a amostra de gasolina padrão apresenta os hidrocarbonetos saturados voláteis em maior concentração. Esta diferença no resultado obtido pelo modelo e no obtido pela ANP pode ter sido causado simplesmente por problemas de estocagem da amostra ou podem revelar gasolinas adulteradas que são classificadas como não adulteradas pelos testes físico-químicos realizados pela ANP⁴⁷.

Tabela 5. Comparação entre os resultados previstos pelo modelo quimiométrico e os obtidos pela ANP.

Amostra	ANP	% v/v de gasolina C	NPLS
N1	não-adulterada	95	não-adulterada
N2	não-adulterada	97	não-adulterada
N3	não-adulterada	44	adulterada
N4	não-adulterada	98	não-adulterada
A1	Adulterada	66	adulterada
A2	Adulterada	48	adulterada
A3	Adulterada	55	adulterada
A4	Adulterada	64	adulterada
A5	Adulterada	81	adulterada
A6	Adulterada	74	adulterada
A7	Adulterada	82	adulterada
A8	Adulterada	82	adulterada
A9	Adulterada	85	adulterada

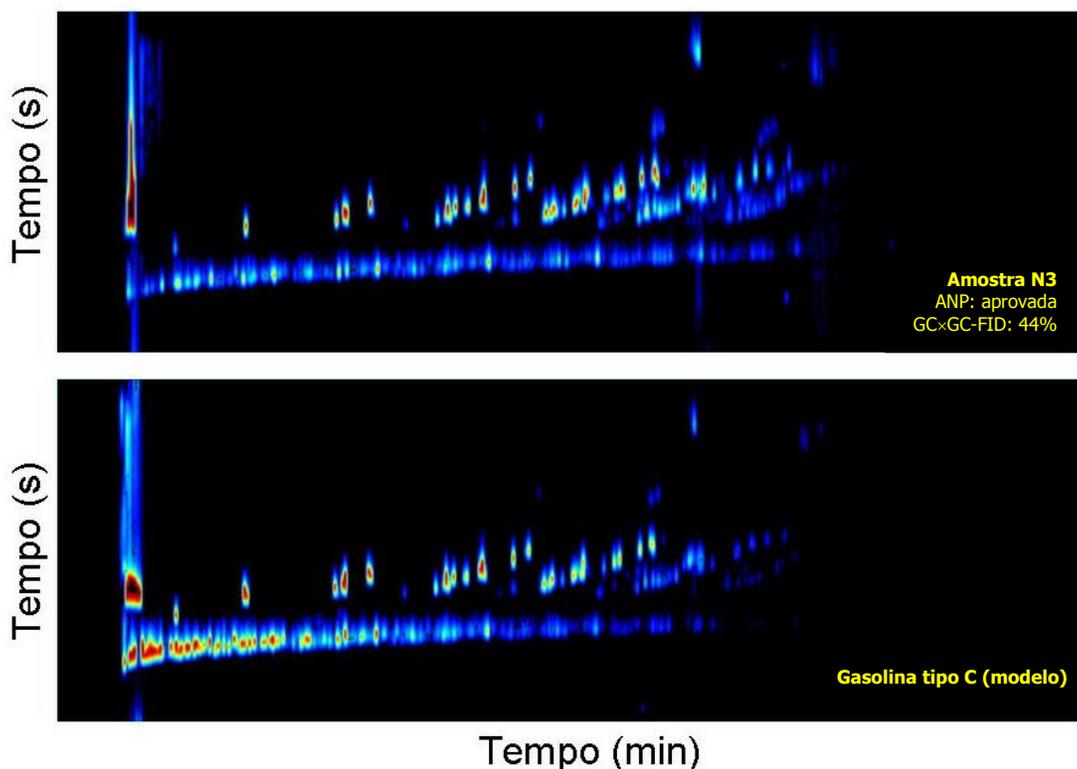


Figura 13. Comparação dos cromatogramas obtidos para a amostra N3 e para a gasolina utilizada no modelo de calibração. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 μ m) + DB-Wax (1,0 m x 0,1 mm x 0,1 μ m). Programação de temperatura do forno cromatográfico: 40 $^{\circ}$ C (3 min.) até 160 $^{\circ}$ C (4 $^{\circ}$ C/min.) e então aquecido até 240 $^{\circ}$ C (20 $^{\circ}$ C/min.) (3 min.). Razão de divisão do injetor 1:300, volume de injeção 0,6 μ L e PM 6 s.

4.2.4. Conclusões

Foi proposta uma metodologia para identificar amostras de gasolinas do tipo C adulteradas. Este é um problema muito comum no Brasil desde o fim do monopólio estatal em 1995. Atualmente, a ANP realiza testes físico-químicos baseados em normas ASTM e ABNT que requerem grandes volumes de amostras e procedimentos tediosos. Os solventes comumente utilizados na prática de adulteração apresentam composição química próxima a da gasolina e, por isso, nem sempre as amostras adulteradas são flagradas nos testes físico-químicos, com exceção para adulteração com etanol em quantidades acima da especificada

na legislação, que é facilmente identificada através de uma extração líquido-líquido.

Na metodologia proposta, as amostras de gasolina foram analisadas por GC×GC-FID e NPLS foi utilizado na construção do modelo quimiométrico. O modelo de calibração foi elaborado usando como adulterantes thinner, aguarrás e querosene e foi validado realizando a previsão de 14 amostras adulteradas preparadas em laboratório, as quais continham gasolina, os adulterantes utilizados na calibração e, em alguns casos, etanol, que foi incluído no conjunto de validação para avaliar a resposta do modelo frente a este tipo de adulteração. Observou-se que a partir de 5% v/v de etanol acima da quantidade estabelecida na lei, o modelo classificou a amostra como adulterada, mas abaixo desse valor o modelo não conseguiu flagrar a adulteração. Por último, o modelo foi utilizado na avaliação de novas amostras de gasolinas, 9 adulteradas e 4 não adulteradas segundo a ANP. Somente em um caso, o modelo quimiométrico apresentou resultado diferente do resultado fornecido pela ANP, que foi para uma gasolina classificada como não adulterada pela ANP e adulterada pelo modelo. Analisando os cromatogramas da gasolina utilizada na calibração e desta amostra é possível facilmente observar as diferenças. A diferença nos resultados pode ter sido causada por problemas de estocagem da amostra ou pode representar um exemplo de gasolina adulterada que não é flagrada nos testes físico-químicos.

No futuro, com o advento do cromatógrafo gasoso bidimensional abrangente portátil, estas análises poderiam ser feitas no próprio posto de combustíveis, após a corrida cromatográfica, as amostras adulteradas seriam identificadas na hora, eliminando assim, a necessidade de transporte e estocagem das amostras, e a comercialização de gasolina nestes postos poderia ser suspensa no mesmo instante.

4.3. Previsão de Propriedades Físico-Químicas de Gasolina

4.3.1. Objetivos

Elaboração de modelos quimiométricos para previsão de algumas das propriedades físico-químicas de gasolina utilizando GC×GC-FID, com o objetivo de propor uma alternativa aos testes realizados pela ANP na avaliação da qualidade da gasolina, os quais requerem um grande volume de amostra e procedimentos manuais trabalhosos.

4.3.2. Parte Experimental

Utilizou-se um sistema de GC×GC-FID construído em laboratório¹⁰ a partir de um HP-6890 GC-FID (Hewlett-Packard, Wilmington, DE, EUA) equipado com injetor do tipo split-splitless. Neste sistema usou-se um modulador criogênico de quatro jatos, sendo N₂ resfriado em nitrogênio líquido como fluído criogênico e N₂ aquecido como gás quente. Tanto o fluxo resfriante como o de aquecimento foram controlados por duas válvulas solenóides Asco (Asco Joucomatic, Rueil Malmaison, França) de três vias. O controle das válvulas e a digitalização do sinal analógico do FID foram realizados pelo controlador A/D de 16 bits DAQPad-6015 (National Instruments Corporate Headquarters, Austin, Texas, EUA) conectado a um microcomputador AMD Athlon 64 X2 4600 Dual Core com 4Gb de memória RAM. O controle das válvulas e a coleta a 100 Hz do sinal do FID foram realizadas em LABView v.8.2 (National Instruments Corporate Headquarters, Austin, Texas, EUA).

O conjunto de colunas utilizado foi uma HP-5 (30 m x 0,25 mm x 0,25 µm) conectada a uma DB-Wax (1,0 m x 0,10 mm x 0,10 µm). As temperaturas do injetor e do detector foram mantidas a 250 °C. A programação de temperatura do forno cromatográfico foi: 40 °C até 120 °C (3 °C/min.) e então aquecido até 220 °C (20 °C/min.). Como gás de arraste utilizou-se hidrogênio a uma vazão de 0,60 mL/min. Em todas as corridas cromatográficas o volume de injeção foi de 1,0 µL no modo split (*split ratio* 1:250). O período de modulação foi de 4 s.

A análise multivariada dos dados cromatográficos foi realizada em MATLAB 6.5 (MathWorks, Natick – MA, EUA) instalado em um microcomputador AMD Athlon 64 X2 4600 Dual Core e 2Gb de memória RAM. Os algoritmos utilizados nesta parte do trabalho foram obtidos do *iToolbox* 1.1 e do *PLS-Genetic Algorithm Toolbox*⁶⁰.

Foram analisadas 51 amostras de gasolinas fornecidas pela Central Analítica – UNICAMP, laboratório certificado pela ANP para análises de gasolina. Foram fornecidos também os resultados dos testes físico-químicos de cada amostra.

4.3.3. Resultados e Discussão

As 51 amostras de gasolina foram divididas em dois grupos utilizando o algoritmo de Kennard-Stone⁶¹, sendo 30 amostras para calibração e 21 amostras para validação.

Primeiramente, tentou-se, sem sucesso, utilizar métodos quimiométricos multimodo na construção dos modelos de calibração/previsão. Sendo assim, os cromatogramas foram “desdobrados” de matrizes para vetores e utilizou-se métodos quimiométricos de primeira ordem. Não foi utilizado nenhum tipo de pré-processamento.

Foram construídos modelos de calibração para as temperaturas de destilação de gasolina a 10%, 50%, 90% e 100% de destilado e para os valores de massa específica. Na figura 14 está representado o diagrama GC×GC-FID obtido para uma amostra de gasolina, onde é possível ver a estruturação cromatográfica.

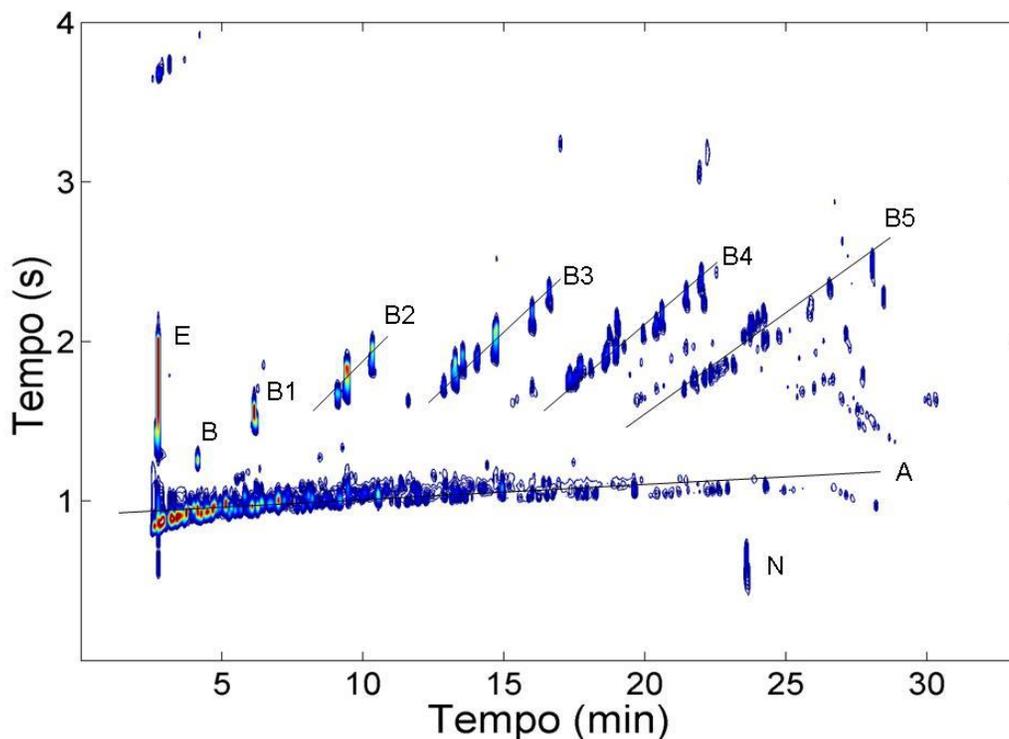


Figura 14. Diagrama GCxGC-FID obtido para gasolina. E= etanol, B = benzeno, B1 = tolueno, B2 = benzeno C2 substituído, B3 = benzeno C3 substituído, B4 = benzeno C4 substituído, B5 = benzeno C5 substituído, N = naftaleno, A = alcanos, cicloalcanos e alcenos. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 μ m) + DB-Wax (1,0 m x 0,1 mm x 0,1 μ m). Programação de temperatura do forno cromatográfico: 40 $^{\circ}$ C até 120 $^{\circ}$ C (3 $^{\circ}$ C/min.) e então aquecido até 220 $^{\circ}$ C (20 $^{\circ}$ C/min.). Razão de divisão do injetor 1:250, volume de injeção 1,0 μ L e PM 4 s.

4.3.3.1. Previsão das Temperaturas da Curva de Destilação de Gasolina

Um dos testes físico-químicos realizados pela ANP para análise de qualidade da gasolina é a medida da temperatura de destilação da amostra a 10, 50 e 90% v/v de destilado e a temperatura no ponto final de destilação. O procedimento utilizado na análise é o estabelecido na norma ASTM D86⁶², o qual utiliza 100 mL da amostra e compreende uma metodologia não muito simples.

Como critério de controle de qualidade, a ANP estabelece faixas de temperaturas correspondentes às frações volumétricas de 10, 50 e 90% de destilado e ao ponto final da destilação, que é a maior temperatura observada durante a destilação, caracterizada pelo aparecimento de fumos brancos no gargalo do balão que mostram o início do craqueamento. O ponto final de destilação não é associado à destilação de 100% de volume da amostra, pois normalmente existem resíduos que permanecem no fundo do balão. Estes pontos específicos da curva de destilação estão relacionados com a performance do motor segundo a descrição a seguir⁶³:

- O teor de 10% de gasolina destilada está relacionada à quantidade mínima que deve evaporar para dar a partida no veículo e aquecer o motor. Se a gasolina for muito volátil a temperatura para destilar os 10% estará muito abaixo da especificação, podendo ocorrer formação de bolhas de vapor no combustível líquido em dias quentes, interrompendo o fluxo de combustível, levando a uma parada do motor. Se o combustível for mais pesado, extrapolando a faixa de temperatura permitida, haverá dificuldades em dar a partida do motor em dias frios;
- O teor de 50% de destilado está relacionado com o aquecimento e desempenho do motor;
- A limitação da temperatura de 90% de destilado visa minimizar a formação de depósitos na câmara de combustão e nas velas de ignição;
- O ponto final da destilação limita a quantidade de produtos de alto ponto de ebulição que podem não queimar por completo na câmara de combustão.

Na elaboração dos modelos quimiométricos para previsão das temperaturas de destilação da gasolina foi utilizado o método quimiométrico siPLS para seleção de intervalos dos dados experimentais. Deste modo, selecionou-se apenas algumas regiões dos cromatogramas para cada ponto da destilação utilizado pela ANP na avaliação da gasolina, o que é coerente, pois a variação na temperatura de destilação entre as amostras para cada porcentagem de destilado está relacionada com a variação de amostra para amostra da concentração dos compostos que apresentam ebulição na respectiva faixa de temperatura ou de

alguma maneira a influenciam. Para os modelos a 10 e 50% de destilado foi utilizado o GA para fazer a seleção de variáveis dentro dos intervalos selecionados pelo siPLS.

4.3.3.1.1. Previsão da Temperatura a 10% de Destilado

Como mencionado anteriormente, o modelo quimiométrico para a previsão da temperatura de destilação a 10% de destilado de gasolina através de GC×GC foi elaborado utilizando-se, primeiro, siPLS e, depois, GA. O número de intervalos avaliados foi de 13 a 17 e de combinações foi 2 e 3. Através do valor de RMSECV, foram selecionados os valores de 15 para intervalos e de 2 para combinações. Na figura 15 estão ilustrados os intervalos selecionados.

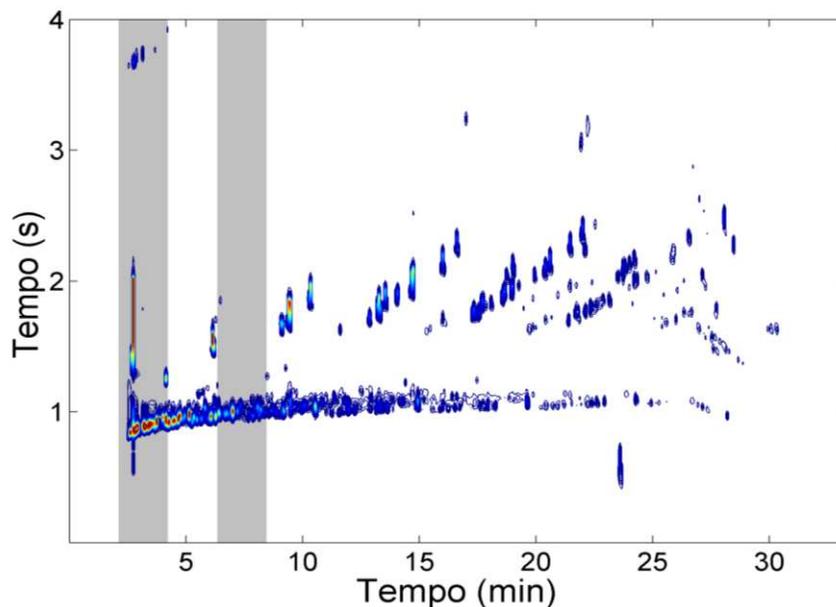


Figura 15. Diagrama GCxGC-FID de gasolina mostrando os intervalos selecionados pelo siPLS em cinza para o modelo de previsão de temperatura a 10% de destilado. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 μ m) + DB-Wax (1,0 m x 0,1 mm x 0,1 μ m). Programação de temperatura do forno cromatográfico: 40 $^{\circ}$ C até 120 $^{\circ}$ C (3 $^{\circ}$ C/min.) e então aquecido até 220 $^{\circ}$ C (20 $^{\circ}$ C/min.). Razão de divisão do injetor 1:250, volume de injeção 1,0 μ L e PM 4 s.

Após a seleção dos intervalos, utilizou-se GA para a seleção das variáveis dentro dos intervalos selecionados que melhor explicam a variância das temperaturas entre as amostras a 10% de destilado. As configurações do GA foram otimizadas, sendo a melhor configuração: 256 para tamanho da população; 200 gerações; 9 variáveis latentes; convergência a 50% e 1% de taxa de mutação. Em seguida, com as variáveis selecionadas pelo GA foi construído um modelo de calibração PLS. A faixa de temperatura das amostras utilizadas foi de 49,3 a 54,6 $^{\circ}$ C. O número de variáveis de cada amostra caiu de 190000 para 7528, após a seleção de variáveis realizada pelo siPLS e pelo GA.

Finalmente, foi realizada a previsão das amostras do conjunto de validação e obteve-se RMSEP de 0,44 $^{\circ}$ C, um coeficiente de correlação entre valores reais e previstos de 0,884 (figura 16) e um valor para a raiz quadrada do erro quadrático médio relativo da diferença de previsão percentual (RMSPD) (equação 11)⁶⁴ de

0,85%. Em um trabalho anterior na literatura, foi realizada a análise de amostras de gasolina por GC-FID e os cromatogramas obtidos foram utilizados para a elaboração de um modelo de calibração PLS para a previsão das temperaturas de destilação, no qual para 10% destilado foi obtido um RMSEP de 3,3 °C⁵⁴. Deste modo, ficou evidenciada que a maior capacidade de pico da GC×GC-FID em relação à GC-FID resulta em melhores modelos de calibração, pois picos que estariam sobrepostos são separados, e, também, foi evidenciada a vantagem de selecionar dentro do conjunto de dados experimental somente as variáveis, que, de fato, estão relacionadas com a variável dependente.

$$\text{RMSPD} = 100 \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{y_i^2} \times \frac{1}{n}} \quad (11)$$

onde que \hat{y}_i e y_i são os valores previstos e de referência para a variável dependente, respectivamente e n é o número de amostras de validação.

A norma ASTM D86 estabelece uma reprodutibilidade para a temperatura da curva de destilação de gasolina a 10% de destilado de 3,2 °C. Apesar de representarem tipos de erros diferentes, pode ser observado que o valor de RMSEP obtido no modelo quimiométrico proposto encontra-se abaixo do valor de reprodutibilidade da norma, o que mostra que a metodologia proposta tem potencial para ser uma alternativa a esta análise.

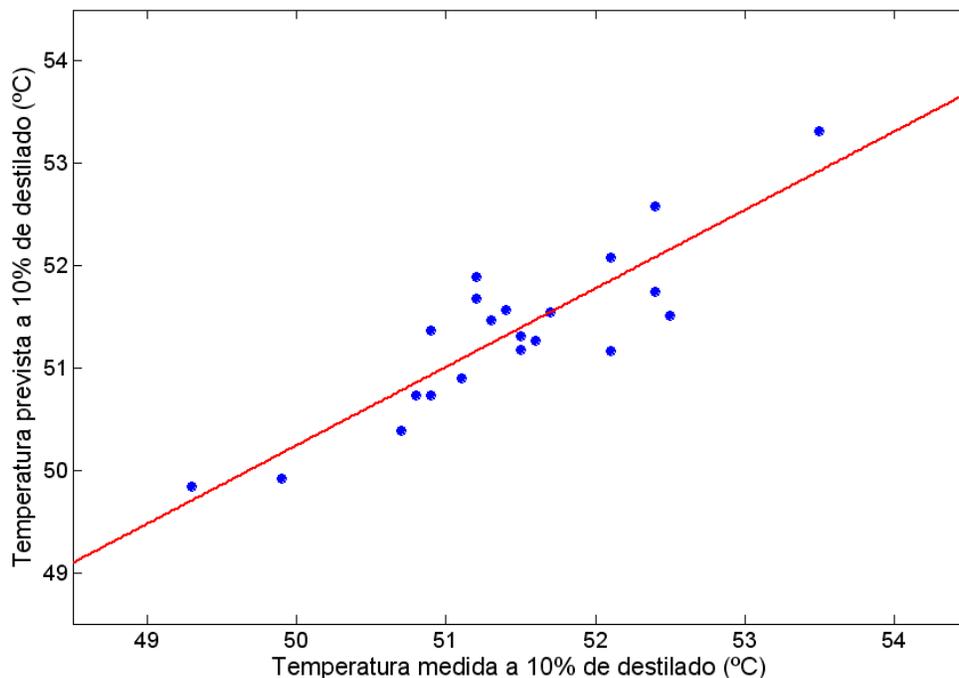


Figura 16. Valores da temperatura a 10% de destilado versus valores previstos pelo modelo quimiométrico para as amostras de validação.

4.3.3.1.2. Previsão da Temperatura a 50% de Destilado

Assim como para o modelo para temperatura a 10% destilado, no modelo de temperatura a 50% de destilado foi utilizado os métodos quimiométricos siPLS e GA.

No siPLS, o número de intervalos avaliados foi de 13 a 17 e de combinações foi 2 e 3. Através do valor de RMSECV, foram selecionados os valores de 15 para intervalos e de 2 para combinações. Os intervalos selecionados estão ilustrados na figura 17. Em seguida, o GA foi utilizado com os seguintes parâmetros: 256 para tamanho da população; 200 gerações; 8 variáveis latentes; convergência a 50% e 1% de taxa de mutação. A faixa de temperatura das amostras utilizadas foi de 70,1 a 74,9 °C. O número de variáveis de cada amostra caiu de 190000 para 7356, após a seleção de variáveis realizada pelo siPLS e pelo GA. Na previsão das amostras de validação obteve-se RMSEP de 0,34 °C, RMSPD de 0,48% e um coeficiente de correlação entre valores reais e

previstos de 0,902 (figura 18). O valor de RMSEP relatado anteriormente na literatura utilizando-se PLS para dados de 1D-GC foi de 5,8 °C⁵⁴. O valor de reprodutibilidade estabelecido na norma ASTM D86 para esta análise é 1,8 °C.

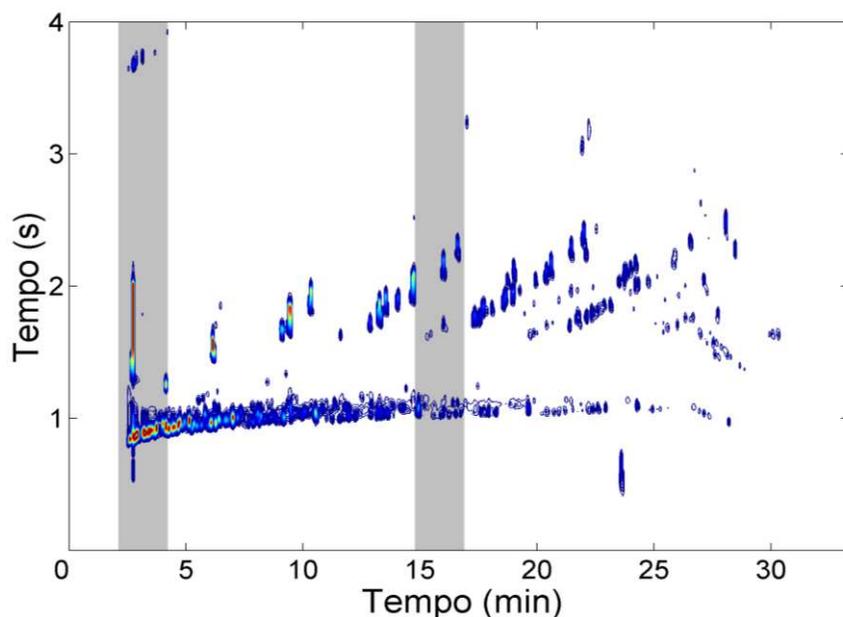


Figura 17. Diagrama GC×GC-FID de gasolina mostrando os intervalos selecionados pelo siPLS em cinza para o modelo de previsão de temperatura a 50% de destilado. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 μm) + DB-Wax (1,0 m x 0,1 mm x 0,1 μm). Programação de temperatura do forno cromatográfico: 40 °C até 120 °C (3 °C/min.) e então aquecido até 220 °C (20 °C/min.). Razão de divisão do injetor 1:250, volume de injeção 1,0 μL e PM 4 s.

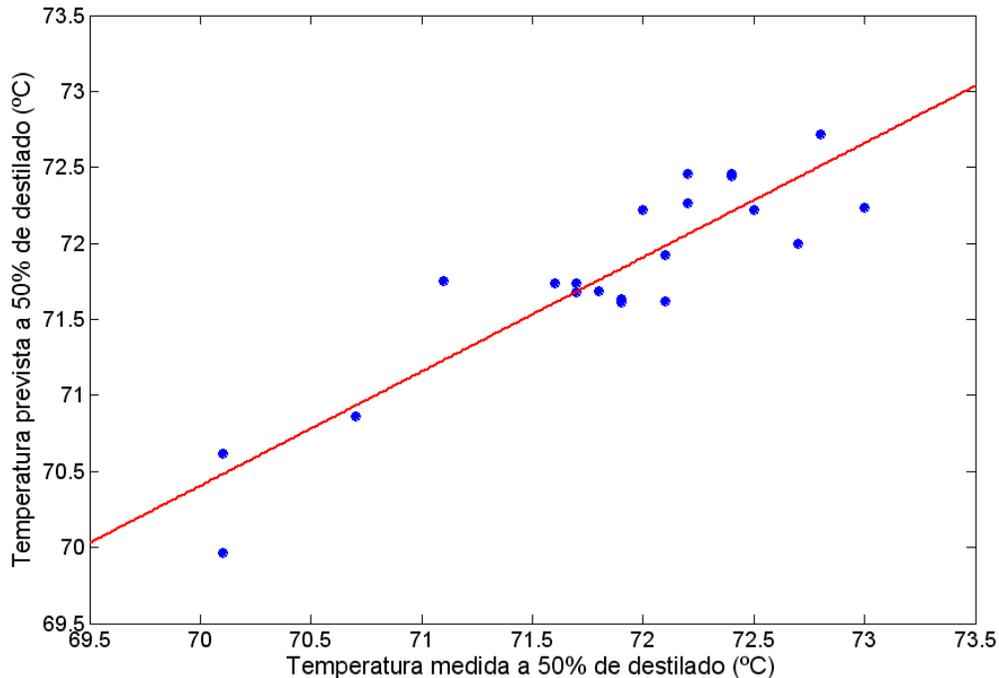


Figura 18. Valores de temperatura a 50% de destilado versus valores previstos pelo modelo quimiométrico para as amostras de validação.

4.3.3.1.3. Previsão da Temperatura a 90% de Destilado

Diferentemente dos modelos construídos para 10 e 50% de destilado de gasolina, na construção do modelo quimiométrico para 90% de destilado foi utilizado apenas o algoritmo siPLS usando 20 intervalos e 3 combinações. Foram utilizadas 6 variáveis latentes na construção do modelo. Nesse caso, não foi utilizado GA para seleção de variáveis dentro dos intervalos selecionados. Os intervalos selecionados estão ilustrados na figura 19. A faixa de temperatura das amostras utilizadas foi de 147,5 a 166,8 °C. O número de variáveis caiu de 190000 para 28500 após o siPLS. Comparando as figuras 15, 16 e 17, observa-se que dentro dos intervalos selecionados pelo siPLS para o modelo de temperatura a 90% de destilado há muito menos sinais cromatográficos que nos intervalos dos modelos para 10 e 50% de destilado, por isso, não foi necessária a utilização de GA para seleção de variáveis.

Depois de construído o modelo de calibração, as amostras do conjunto de validação tiveram sua temperatura de destilação a 90% de destilado prevista e obteve-se valores de RMSEP de 2,1 °C, RMSPD de 1,1% e coeficiente de correlação entre valores reais e previstos de 0,77 (figura 20). O valor de RMSEP relatado anteriormente na literatura utilizando-se PLS para dados de 1D-GC foi de 10,1 °C⁵⁴. O valor de reprodutibilidade estabelecido na norma ASTM D86 para esta análise é 5,8 °C.

Vale ressaltar que em um primeiro momento o valor do coeficiente de correlação obtido pode parecer baixo, mas considerando a faixa de temperatura das amostras utilizadas e o valor de reprodutibilidade estabelecido na norma ASTM D86 para esta análise, ou seja, a variação que pode ocorrer nos valores de referência obtidos para uma mesma amostra, este valor de coeficiente de correlação é coerente com a norma ASTM, o que é reforçado pelos valores de RMSEP e RMSPD obtidos. A mesma situação foi observada na determinação da temperatura no ponto final da destilação e na determinação da massa específica das amostras.

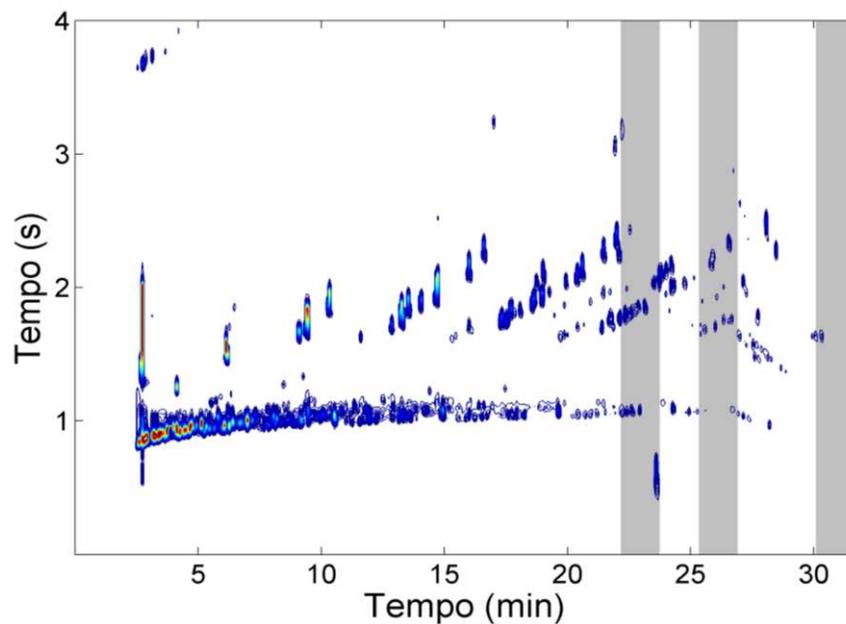


Figura 19. Diagrama GCxGC-FID de gasolina mostrando os intervalos selecionados pelo siPLS em cinza para o modelo de previsão de temperatura a 90% de destilado. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 μ m) + DB-Wax (1,0 m x 0,1 mm x 0,1 μ m). Programação de temperatura do forno cromatográfico: 40 $^{\circ}$ C até 120 $^{\circ}$ C (3 $^{\circ}$ C/min.) e então aquecido até 220 $^{\circ}$ C (20 $^{\circ}$ C/min.). Razão de divisão do injetor 1:250, volume de injeção 1,0 μ L e PM 4 s.

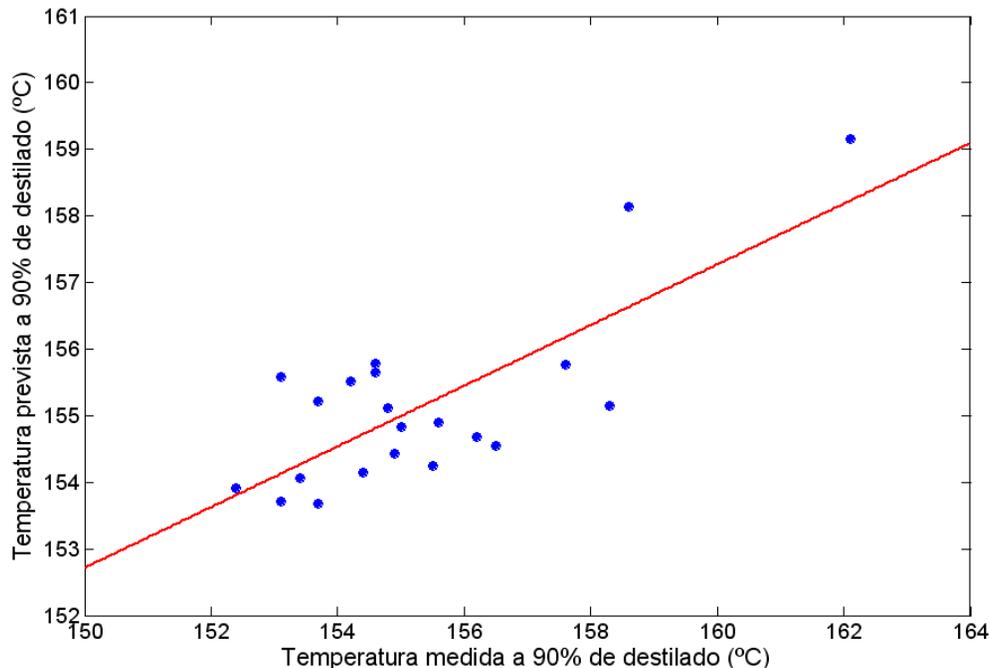


Figura 20. Valores de temperatura a 90% de destilado versus valores previstos pelo modelo quimiométrico para as amostras de validação.

4.3.3.1.4. Previsão da Temperatura no Ponto Final da Destilação

Assim como para 90% de destilado, na construção do modelo de calibração para a temperatura no ponto final da destilação utilizou-se somente o siPLS. A melhor condição obtida com o siPLS foi utilizando 15 intervalos e 3 combinações. Foram utilizadas 4 variáveis latentes na construção do modelo. Na figura 21 estão mostrados os intervalos selecionados para o melhor modelo. A faixa de temperatura das amostras utilizadas foi de 189,3 a 209,3 °C. O número de variáveis latentes caiu de 190000 para 38000 após a aplicação do siPLS. Na previsão das amostras do conjunto de validação obteve-se RMSEP de 2,08 °C, RMSPD de 1,7% e coeficiente de correlação para valores reais e previstos igual a 0,74 (figura 22). O valor de RMSEP relatado anteriormente na literatura utilizando-se PLS para dados de 1D-GC foi de 10,5 °C⁵⁴. O valor de reprodutibilidade estabelecido na norma ASTM D86 para esta análise é 6,8 °C.

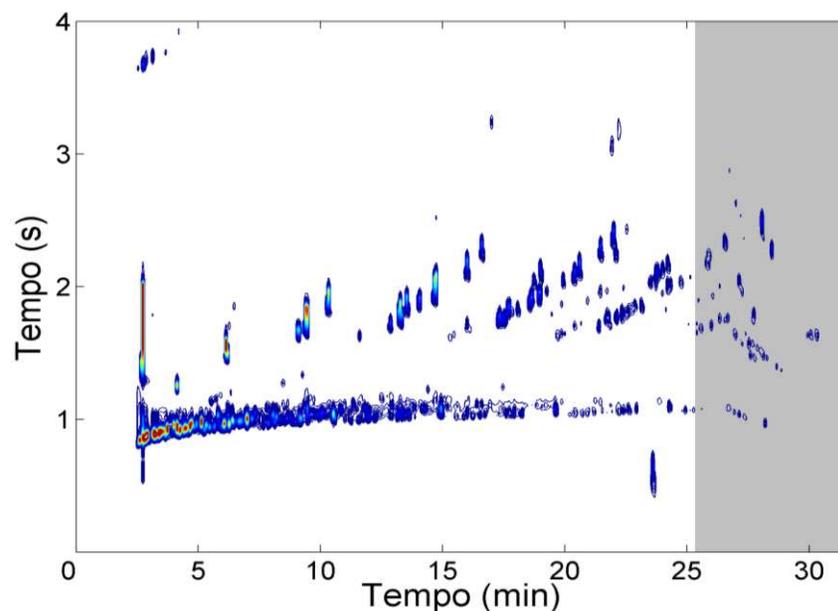


Figura 21. Diagrama GCxGC-FID de gasolina mostrando os intervalos selecionados pelo siPLS em cinza para o modelo de previsão de temperatura no ponto final da destilação. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 μ m) + DB-Wax (1,0 m x 0,1 mm x 0,1 μ m). Programação de temperatura do forno cromatográfico: 40 $^{\circ}$ C até 120 $^{\circ}$ C (3 $^{\circ}$ C/min.) e então aquecido até 220 $^{\circ}$ C (20 $^{\circ}$ C/min.). Razão de divisão do injetor 1:250, volume de injeção 1,0 μ L e PM 4 s.

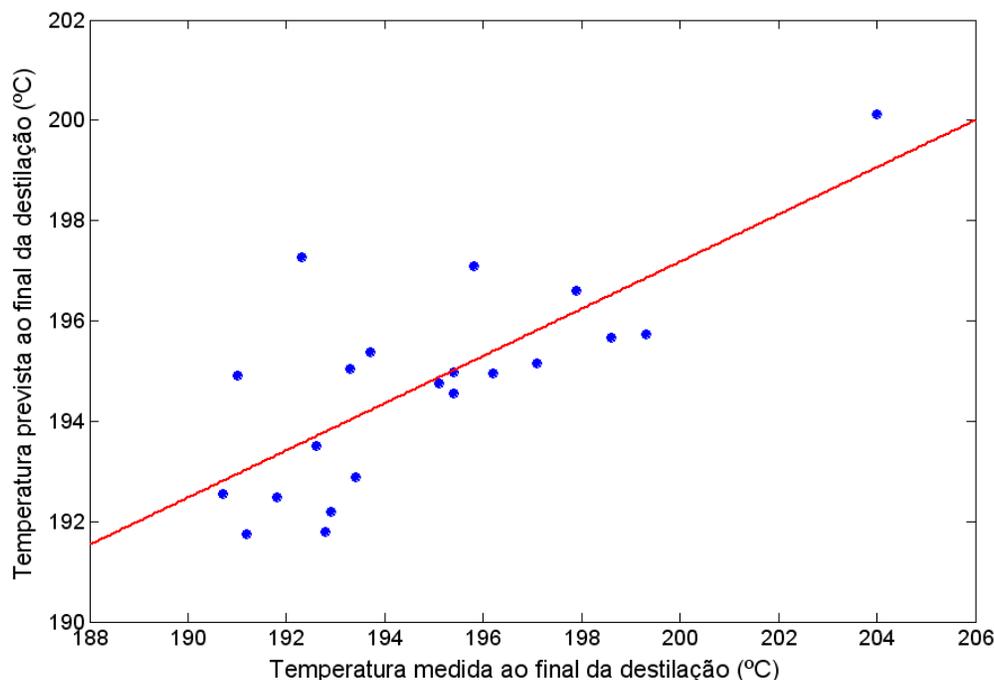


Figura 22. Valores de temperatura no ponto final de destilação versus valores previstos pelo modelo quimiométrico para as amostras de validação.

4.3.3.2. Previsão do Valor da Massa Específica de Gasolina

Outro teste físico-químico realizado pela ANP na análise de amostras de gasolinas brasileiras é a determinação da massa específica, a qual indica a adulteração com produtos mais leves ou mais pesados que a gasolina, para isso, a ANP segue o procedimento estabelecido na norma ASTM D4052⁶⁵. Diferentemente dos valores de temperatura de destilação, o valor da massa específica depende de toda a composição da amostra, por isso, na construção do modelo quimiométrico foi utilizado PLS, o qual utiliza todo o cromatograma. Os valores da massa específica das amostras utilizadas na construção do modelo estavam entre 738,2 e 761,3 kg/m³. Foram utilizadas 6 variáveis latentes para a construção do modelo. Na previsão das amostras de validação, obteve-se RMSEP igual a 1,8 kg/m³, RMSPD igual a 0,2% e coeficiente de correlação entre valores reais e previstos igual a 0,76 (figura 23). O valor de RMSEP relatado anteriormente na literatura utilizando PLS para dados de 1D-GC foi de 4,3 kg/m³

⁵⁴. O valor de reprodutibilidade estabelecido na norma ASTM D4052 para esta análise é $2,0 \text{ kg/m}^3$. Deste modo, assim como para os pontos da curva de destilação, o valor de RMSEP obtido na metodologia proposta encontra-se abaixo do valor de reprodutibilidade estabelecido na norma ASTM⁶⁶.

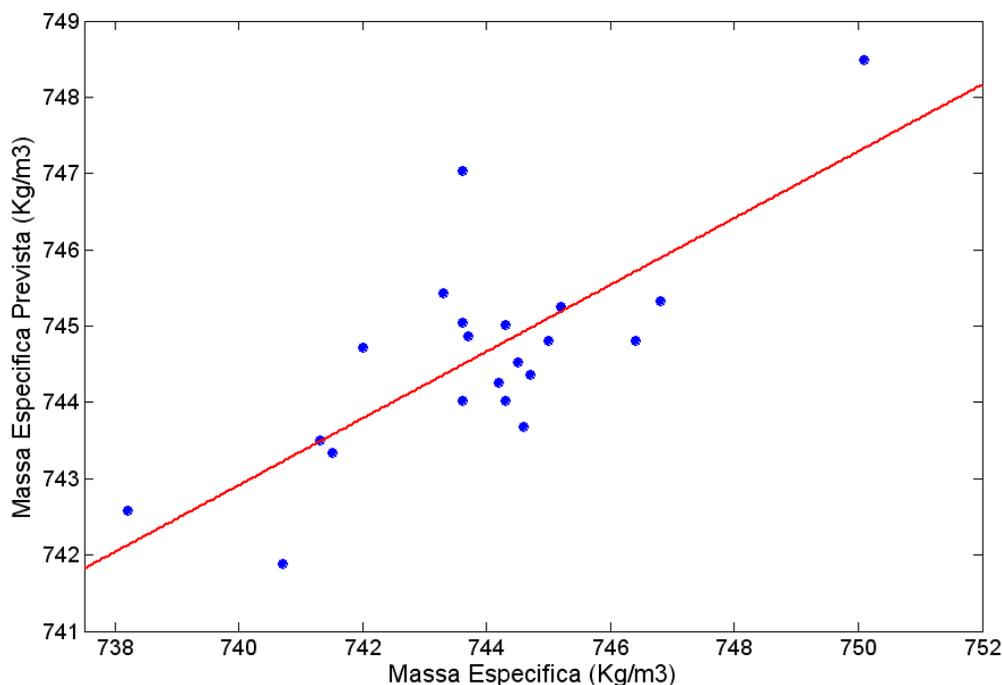


Figura 23. Valores de massa específica das amostras de validação versus valores previstos pelo modelo quimiométrico.

4.3.4. Conclusões

A previsão de alguns parâmetros físico-químicos utilizados pela ANP na avaliação de qualidade de gasolina foi realizada. Para isso, utilizou-se cromatografia gasosa bidimensional abrangente, na qual os cromatogramas bidimensionais foram vetorizados, combinada a métodos quimiométricos de primeira ordem devido à falta de métodos quimiométricos de ordem superior que realizem seleção de variáveis ou intervalos dos dados experimentais.

Os resultados obtidos apresentaram erros menores aos relatados anteriormente na literatura para GC convencional combinada a quimiometria. Para a previsão da temperatura de destilação a 10, 50 e 90% de destilado e no ponto

final da destilação, os valores de RMSEP obtidos ficaram abaixo dos valores da reprodutibilidade estabelecida na norma ASTM D86. Para a previsão dos valores de massa específica o valor de RMSEP ficou abaixo do valor de reprodutibilidade da norma ASTM D4052.

Deste modo, através dos modelos quimiométricos elaborados, utilizando uma única corrida cromatográfica é possível prever as temperaturas da curva de destilação e o valor da massa específica para a amostra. A metodologia proposta apresenta um procedimento experimental mais simples e rápido que as normas ASTM e utiliza um volume de amostra muito menor, apenas alguns microlitros contra 100 mL da norma para a curva de destilação, sendo assim, uma interessante alternativa aos métodos utilizados pela ANP.

4.4. Diferenciação entre Amostras de Gasolinas do Brasil e da Venezuela

4.4.1. Objetivos

Elaboração de um modelo quimiométrico utilizando dados obtidos por GC×GC-FID para diferenciar amostras de gasolinas brasileiras de venezuelanas, devido ao intenso contrabando existente da Venezuela para o Brasil. Nesta aplicação utilizou-se o algoritmo PARAFAC para a análise exploratória dos dados.

4.4.2. Parte Experimental

Utilizou-se um sistema de GC×GC-FID construído em laboratório¹⁰ a partir de um HP-6890 GC-FID (Hewlett-Packard, Wilmington, DE, EUA) equipado com injetor do tipo split-splitless. Neste sistema usou-se um modulador crioscópico de quatro jatos, sendo N₂ resfriado em nitrogênio líquido como fluído crioscópico e N₂ aquecido como gás quente. Tanto o fluxo resfriante como o de aquecimento foram controlados por duas válvulas solenóides Asco (Asco Joucomatic, Rueil Malmaison, França) de três vias. O controle das válvulas foi realizado por um controlador A/D de 16 bits DAQPad-6015 (National Instruments Corporate

Headquarters, Austin, Texas, EUA) conectado em um microcomputador AMD Athlon 64 X2 4600 Dual Core com 4Gb de memória RAM. O controle das válvulas foi realizado em LABView v.8.2 (National Instruments Corporate Headquarters, Austin, Texas, EUA). Os cromatogramas foram adquiridos e digitalizados em ChemStation Software (Agilent Technologies, Santa Clara, Califórnia, USA).

O conjunto de colunas utilizado foi uma HP-5 (30 m x 0,25 mm x 0,25 μ m) conectada a uma SupelcoWax (1,0 m x 0,10 mm x 0,10 μ m). As temperaturas do injetor e do detector foram mantidas a 250 °C. A programação de temperatura do forno cromatográfico foi: 60 °C até 226 °C (3 °C/min.). Como gás de arraste utilizou-se hidrogênio a uma vazão de 0,60 mL/min. Em todas as corridas cromatográficas o volume de injeção foi de 1,0 μ L no modo split (*split ratio* 1:300). O período de modulação foi de 6 s.

Os cromatogramas bidimensionais tiveram a região inicial e final descartadas, foram desdobrados para vetores e alinhados utilizando o algoritmo COW (Correlation Optimized Warping)⁶⁷.

A análise multivariada dos dados cromatográficos foi realizada em MATLAB 6.5 (MathWorks, Natick – MA, EUA) instalado em um microcomputador AMD Athlon 64 X2 4600 Dual Core e 4Gb de memória RAM. O algoritmo PARAFAC foi obtido do *N*-way Toolbox 2.11 (www.models.kvl.uk).

Foram analisadas 25 amostras de gasolinas brasileiras tipo A (sem adição de etanol) fornecidas pela PETROBRAS e 26 amostras de gasolinas venezuelanas fornecida pela Polícia Federal brasileira.

4.4.3. Resultados e Discussão

As gasolinas provenientes da Venezuela não apresentam etanol em sua composição, deste modo, foram utilizadas gasolinas brasileiras do tipo A, ou seja, sem etanol. Vale ressaltar que não foi realizada a adição de etanol à gasolina venezuelana para assim construir um modelo com gasolina brasileira tipo C, devido ao pequeno volume disponível (menos que 1,0 mL) de cada amostra de gasolina venezuelana. Na figura 24 estão representados os cromatogramas obtidos para gasolinas brasileira e venezuelana.

Devido aos desvios nos tempos de retenção de amostra para amostra, algo comum em cromatografia, os cromatogramas bidimensionais foram desdobrados para vetores e foram alinhados como se fossem 1D-GC utilizando o algoritmo COW⁶⁷. O início e o final dos cromatogramas, onde não existem compostos, foram descartados com a finalidade de diminuir o tempo necessário para o alinhamento. Também foi descartada a região na qual saem os primeiros compostos, devido a problemas de “espelhamento”, que consiste no não aprisionamento destes no modulador, devido à alta volatilidade. Esse alinhamento foi necessário, pois o modelo quimiométrico para classificação das amostras de gasolinas brasileira e venezuelana foi construído com PARAFAC e um dos pré-requisitos para utilização desse algoritmo é a trilinearidade dos dados.

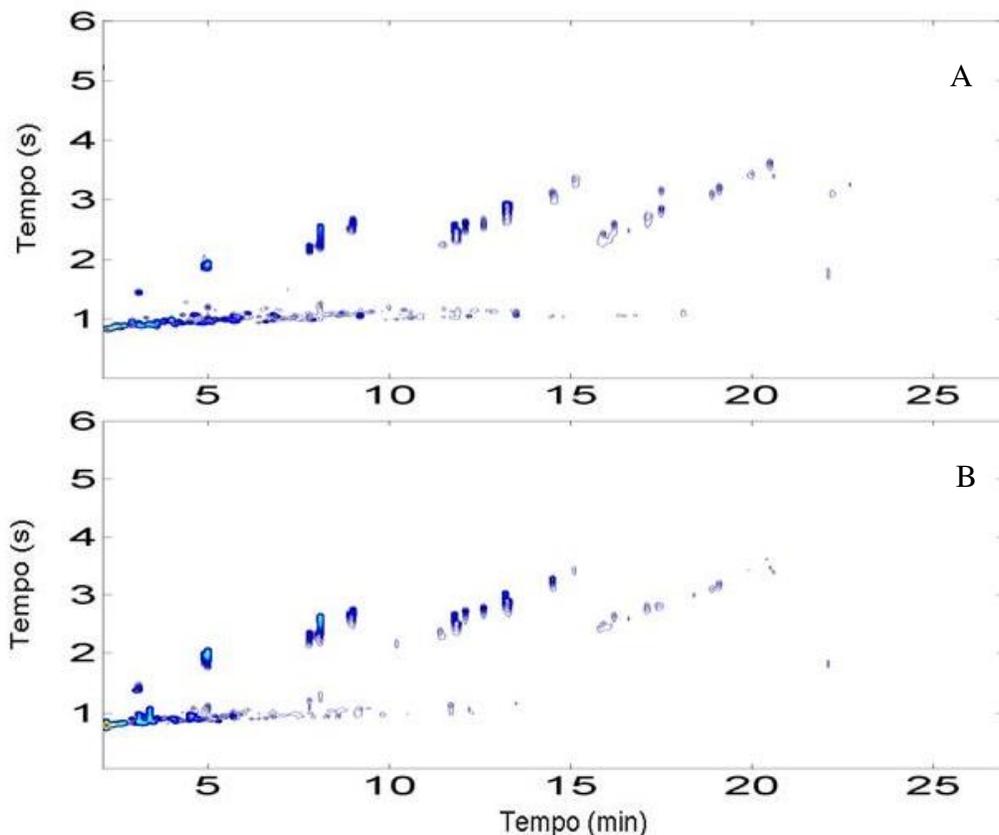


Figura 24. Cromatograma obtido para (A) gasolina brasileira e (B) gasolina venezuelana. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 μ m) + SupelcoWax (1,0 m x 0,1 mm x 0,1 μ m). Programação de temperatura do forno cromatográfico: 60 $^{\circ}$ C até 223 $^{\circ}$ C (3 $^{\circ}$ C/min.). Razão de divisão do injetor 1:300, volume de injeção 1,0 μ L e PM 6 s.

Como discutido na seção 2.3, a escolha do número de fatores é uma etapa crucial na elaboração de um modelo PARAFAC. Neste caso, utilizou-se o teste de CORCONDIA, no qual observou-se 3 ser o número ideal, pois 2 fatores não foi suficiente para separar as gasolinas, com 3 fatores as gasolinas puderam ser separadas e a partir de 4 fatores o valor do teste de CORCONDIA diminui para menos de 50%. Deste modo, o modelo PARAFAC foi construído com 3 fatores e aplicando não-negatividade como restrição nos dois modos da separação cromatográfica, pois não existem sinais negativos no detector FID.

Na figura 25 está representado o gráfico dos escores obtidos nos três fatores para cada amostra, onde pode ser visto que as gasolinas brasileiras e venezuelanas puderam ser separadas. Os pesos obtidos na decomposição dos dados estão mostrados na figura 26, na qual pode ser observado que o primeiro fator está relacionado aos alcanos, alcenos e cicloalcanos mais voláteis, o terceiro fator está relacionado aos alcanos, alcenos e cicloalcanos menos voláteis e, também, aos aromáticos e o segundo fator está relacionado aos compostos entre o primeiro e terceiro fator.

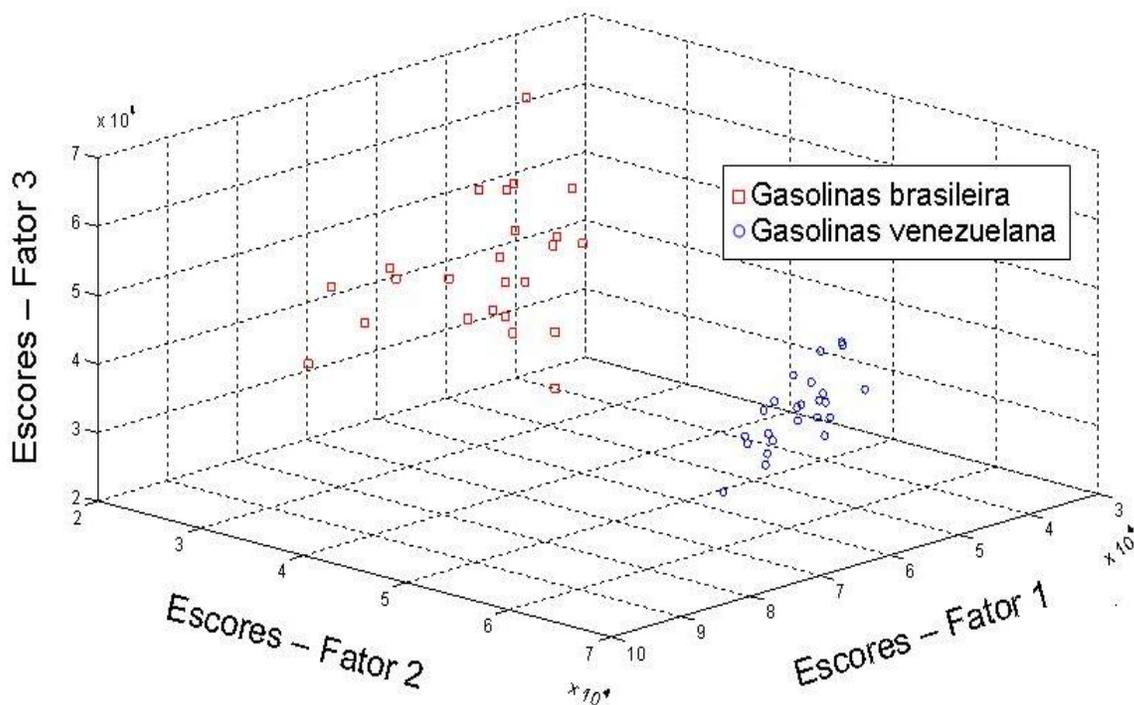
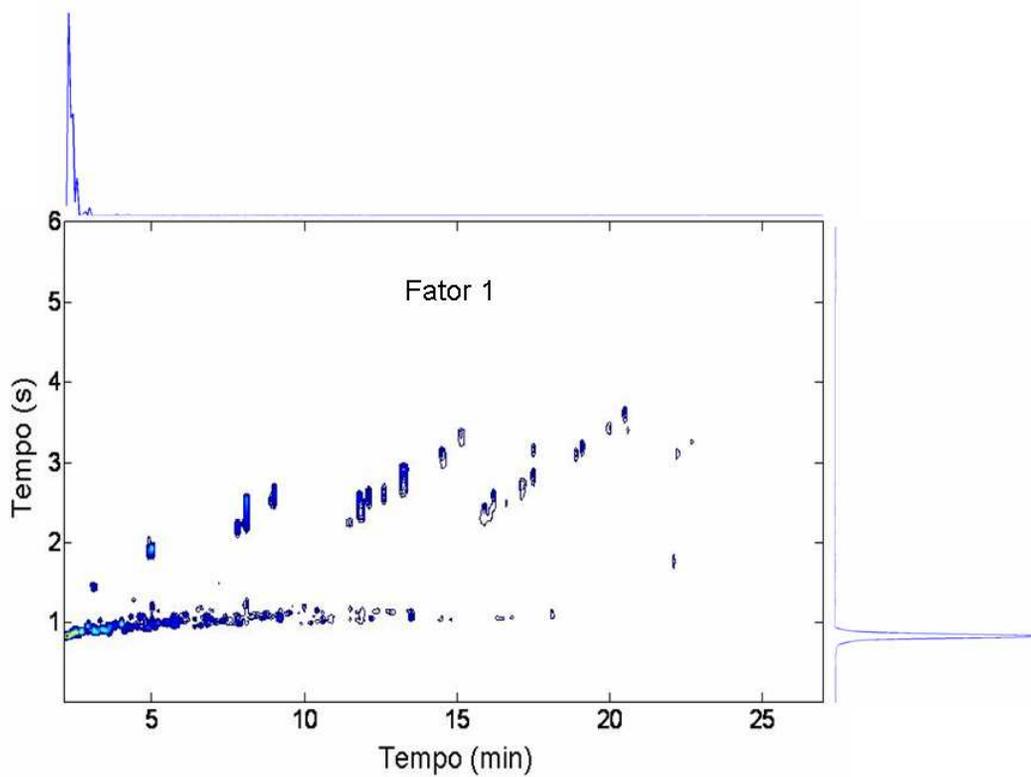


Figura 25. Escores obtidos para as amostras de gasolinas brasileira e venezuelana.



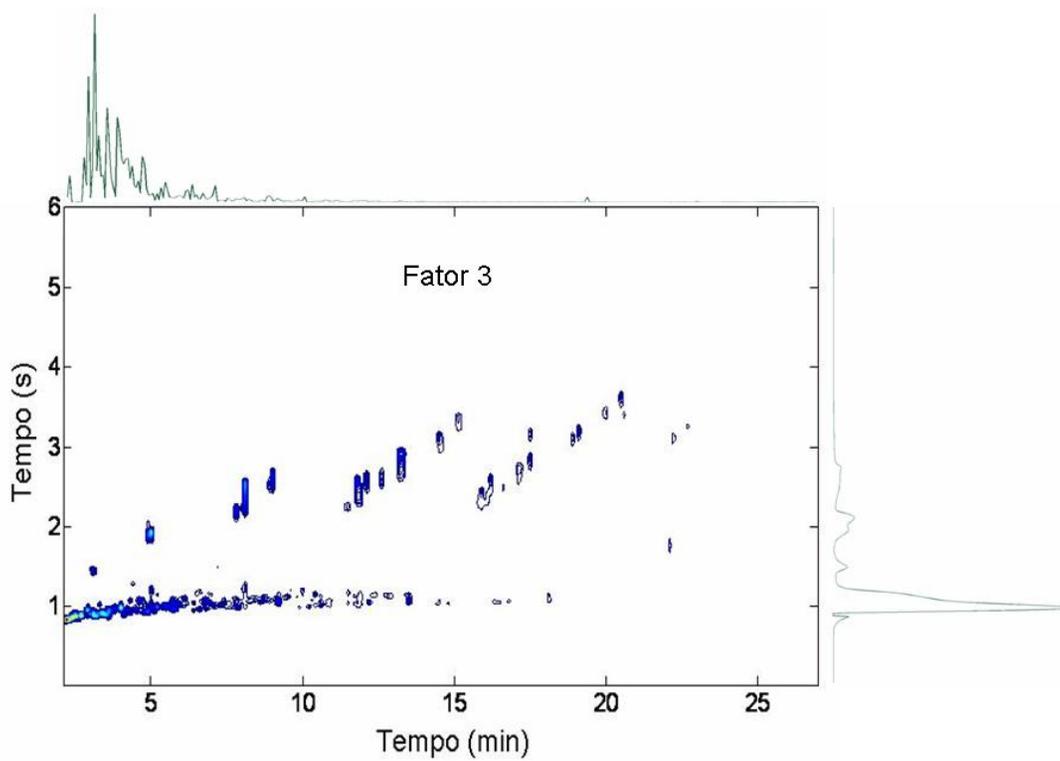
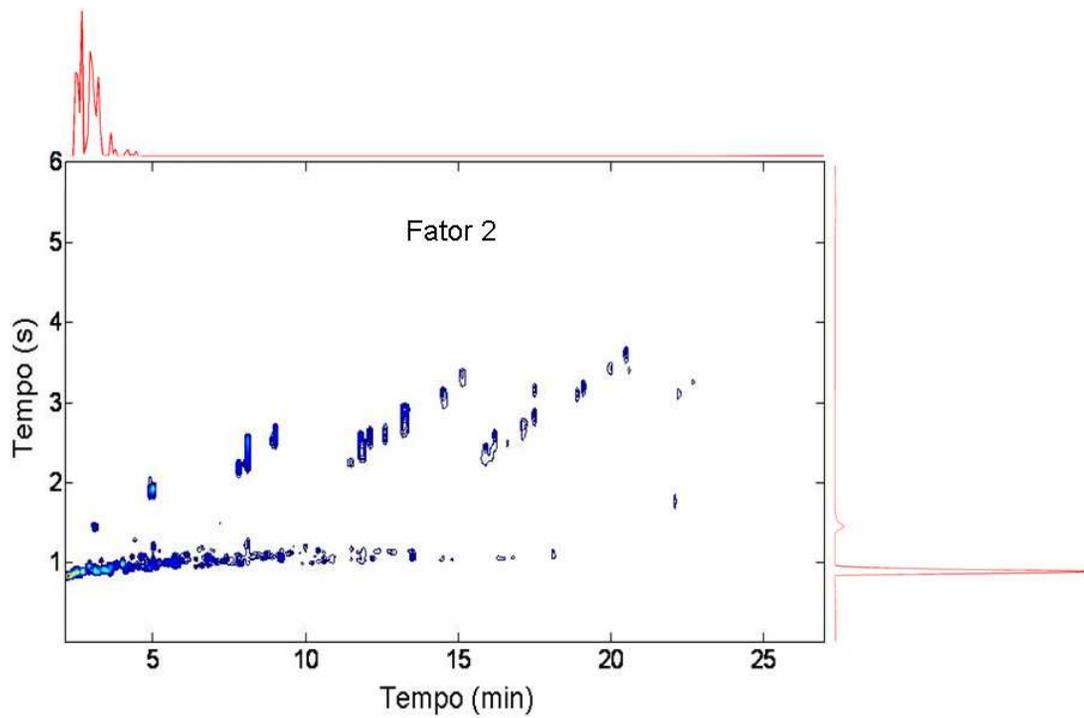


Figura 26. Pesos obtidos nas duas dimensões da separação cromatográfica. Em azul fator 1, em verde fator 2 e em vermelho fator 3.

4.4.4. Conclusões

Este trabalho permitiu realizar a diferenciação entre amostras de gasolinas brasileiras e venezuelanas, pois, atualmente, existe um intenso contrabando da Venezuela para o Brasil, devido à enorme diferença de preços existente.

Amostras de gasolinas venezuelanas foram fornecidas pela Polícia Federal brasileira e são resultados de apreensões realizadas em território nacional. Como a gasolina venezuelana não apresenta etanol em sua composição, foram utilizadas amostras de gasolina brasileira do tipo A, as quais foram fornecidas pela PETROBRAS. Todas as amostras foram analisadas por cromatografia gasosa bidimensional abrangente e PARAFAC foi utilizado na construção do modelo de classificação. Através do gráfico dos escores obtidos do PARAFAC para cada amostra foi possível realizar a distinção entre gasolinas brasileiras e venezuelanas.

A identificação de gasolina venezuelana sendo vendida ilegalmente no Brasil pode não ser uma tarefa simples. No caso de o receptor vender a gasolina venezuelana sem a adição de etanol, a amostra é facilmente identificada pela extração líquido-líquido realizada pela ANP para verificar a quantidade de etanol presente na amostra. Por outro lado, se o receptor adicionar etanol na quantidade correta, a identificação torna-se muito difícil, pois nos testes físico-químicos a amostra de gasolina venezuelana, provavelmente, apresentará resultados dentro dos limites estabelecidos pela ANP.

Neste trabalho, utilizou-se gasolina brasileira tipo A na construção do modelo devido ao pouco volume (menos que 1,0 mL) de amostra que possuíamos das gasolinas venezuelanas, ou seja, não foi possível adicionar 25 % v/v de etanol às amostras venezuelanas. Como foi possível a diferenciação entre as amostras de gasolinas sem etanol, um modelo PARAFAC utilizando gasolinas tipo C do Brasil e gasolinas da Venezuelana, nas quais se adicione etanol na porcentagem correta, provavelmente, resultará em resultados semelhantes. Deste modo, o trabalho por ora realizado abre a perspectiva para o desenvolvimento de um modelo de classificação entre amostras de gasolinas brasileiras e venezuelanas com etanol.

CAPÍTULO 5:

Desenvolvimento de um Algoritmo para Seleção
de Variáveis por Intervalos para o Método de
Mínimos Quadrados Parciais Multimodo

5. Desenvolvimento de um Algoritmo para Seleção de Variáveis por Intervalos para o Método de Mínimos Quadrados Parciais Multimodo

5.1. Introdução

Com o desenvolvimento da instrumentação analítica ocorrida nas últimas décadas, passou-se a obter muito mais informação de cada amostra analisada, deste modo um conjunto de várias amostras fornece uma quantidade de informação muito grande ao analista. Como mencionado no capítulo 1, o método dos mínimos quadrados parciais multimodo (NPLS) relaciona toda a informação obtida experimentalmente em instrumentos de ordem superior com a variável dependente em estudo, mas, em alguns casos, somente uma parte desta informação experimental está relacionada com as variáveis dependentes.

Para dados de primeira ordem, existem algoritmos como o iPLS e GA, que constroem modelos de calibração multivariada com seleção dos intervalos ou das variáveis que estão melhor relacionadas com a variável dependente, descartando a parte dos dados experimentais que não contribui para o modelo de calibração.

Os equipamentos de ordem superior têm sido muito utilizados para o estudo de amostras complexas, para as quais os equipamentos de primeira ordem, muitas vezes, não conseguem uma resolução mínima necessária dos compostos dessas amostras para a construção de modelos de calibração multivariada. Até o momento, não há relatos na literatura sobre um algoritmo que construa modelos de calibração multivariada com dados de ordem superior trabalhando por intervalos. Deste modo, foi proposto neste trabalho o método mínimos quadrados parciais por intervalos (iNPLS).

No iNPLS a matriz de dados é dividida em um número de intervalos selecionado pelo analista nas duas dimensões, esse número de intervalos pode ser diferente para cada dimensão (figura 27). Um conjunto de matrizes reduzidas são formadas a partir da matriz original e para cada uma destas matrizes um

modelo NPLS é construído. A matriz reduzida melhor relacionada com a variável dependente apresentará o menor RMSECV e, deste modo, será selecionada pelo iNPLS como a matriz a ser utilizada na construção do modelo de calibração.

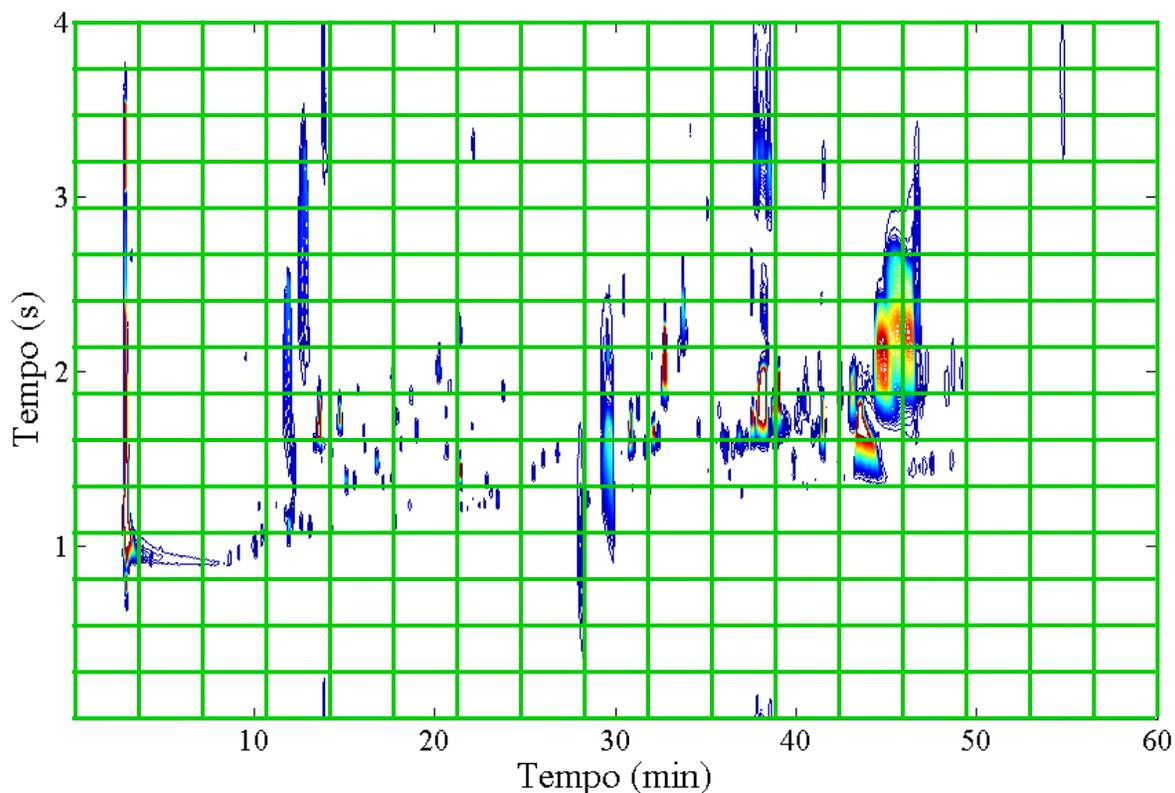


Figura 27. Divisão da matriz de dados em intervalos nas duas dimensões dos dados.

A escolha do número de intervalos em cada uma das dimensões dos dados pode ser um trabalho relativamente difícil, dependendo do conhecimento do analista sobre o conjunto de amostras sob análise. O iNPLS fornece, após a construção do modelo de calibração NPLS para todas as matrizes reduzidas, uma lista com as 10 matrizes que apresentaram os menores valores de RMSECV. Em uma situação, na qual o número de intervalos selecionado for o adequado, a matriz com menor RMSECV apresentará um valor muito menor em relação à segunda matriz com menor RMSECV. Para o caso de análise de composto-alvo, quando o número de intervalos escolhido pelo analista for maior que o adequado, o sinal deste composto estará dividido em 2 ou mais matrizes, sendo assim, na

lista com os 10 menores valores de RMSECV, estas matrizes apresentarão um valor muito menor em relação às demais e, através de uma análise visual do cromatograma, o analista poderá identificar o sinal referente ao composto-alvo e escolher o número adequado de intervalos para cada uma das dimensões. Além disso, quanto maior o número de intervalos maior é o tempo que o algoritmo demora na construção do modelo. Já no caso do analista escolher um número pequeno de intervalos, outros sinais além do sinal correspondente ao composto-alvo serão incluídos na matriz reduzida selecionada, o que pode levar a modelos com piores ajustes.

Apesar da escolha do número de intervalos poder não ser trivial, dependendo do conhecimento do analista sobre o conjunto de amostras sendo analisado, este procedimento ainda é muito mais simples que realizar a integração convencional na análise de compostos-alvos por GC×GC-FID, na qual os cromatogramas bidimensionais são desdobrados para vetores e todos os picos modulados referente ao composto-alvo são integrados e tem suas áreas somadas. Após isso, é construída uma curva analítica através de uma relação linear entre a concentração das amostras e a soma das áreas dos sinais modulados. Já para o iNPLS, além da escolha do número de intervalos, a outra tarefa do analista na elaboração do modelo é a escolha do número de variáveis latentes, algo simples de ser realizado através do procedimento de validação cruzada. Outro problema encontrado na integração convencional é a presença de interferentes que apresentem algum tipo de coeluição com os picos do composto-alvo, algo que pode tornar o desenvolvimento de um modelo de calibração por integração convencional inviável. Já para um modelo de calibração construído por iNPLS, o composto-alvo pode apresentar coeluição com algum interferente.

Como mencionado na seção 2.2, o método do NPLS não apresenta vantagem de segunda ordem, ou seja, as amostras desconhecidas que serão previstas pelo modelo não podem apresentar compostos diferentes daqueles presentes no conjunto de calibração. Entretanto, um resultado semelhante a vantagem de segunda ordem pode ser alcançada com o uso do iNPLS, desde que

os interferentes presentes em amostras desconhecidas e ausentes na calibração não estejam na mesma matriz reduzida que o composto-alvo.

O método iNPLS proposto foi testado, inicialmente, em uma mistura de cinco compostos preparada em laboratório. Os compostos foram escolhidos de maneira que não houvesse sobreposição entre eles e, deste modo, a seleção da matriz reduzida feita pelo iNPLS pudesse ser avaliada. Em seguida, o iNPLS foi testado na quantificação de três alergênicos em perfume, os quais não estavam presentes no perfume utilizado e foram adicionados em laboratório, para avaliar sua capacidade de seleção da matriz reduzida correta e de realizar a quantificação destes três compostos.

5.2. Objetivos

Desenvolvimento do algoritmo mínimos quadrados parciais multimodo por intervalos (iNPLS) para construção de modelos de calibração com seleção de intervalos em dados de ordem superior.

5.3. Parte Experimental

5.3.1. Análise de padrões

Utilizou-se o cromatógrafo descrito no item 4.4.2. para as análises.

O conjunto de colunas empregado foi uma HP-5 (30 m x 0,25 mm x 0,25 μ m) conectada a uma BPX-50 (1,0 m x 0,10 mm x 0,10 μ m). As temperaturas do injetor e do detector foram mantidas a 250 °C. Como gás de arraste utilizou-se hidrogênio a uma vazão de 0,60 mL/min. Em todas as corridas cromatográficas o volume de injeção foi de 1,0 μ L no modo *split* (*split ratio* 1:250). A programação de temperatura do forno cromatográfico foi: 60 °C durante 3 min., sendo em seguida aquecido até 120 °C (3 °C/min.) e, então, aquecido até 240 °C (20 °C/min.). O período de modulação utilizado foi 4 s e a taxa de aquisição foi 100 Hz.

A análise multivariada dos dados cromatográficos foi realizada em MATLAB 6.5 (MathWorks, Natick – MA, EUA) instalado em um microcomputador AMD

Athlon 64 X2 4600 Dual Core e 4Gb de memória RAM. O algoritmo iNPLS foi escrito em MATLAB 6.5 e o NPLS foi obtido do *N-way Toolbox 2.11*⁵⁷.

O algoritmo desenvolvido foi, inicialmente, avaliado na análise de um conjunto de amostras preparadas em laboratório contendo os seguintes compostos: undecano, 1-octanol, 2-octanona, cicloexanona e tolueno. Como solvente foi utilizado hexano. Foram preparadas 25 amostras seguindo planejamento fatorial *Greco-Latin Squares*⁵⁸ e concentrações variando entre 0,0 e 3,0% v/v (tabela 6).

Tabela 6. Composição das amostras com cinco padrões utilizadas para testar o iNPLS. Concentrações em % v/v.

Amostra	Undecano	1-octanol	2-octanona	Cicloexanona	Tolueno
1	1,0	1,0	3,0	0,0	2,0
2	0,5	0,0	0,5	3,0	2,0
3	0,0	0,0	0,0	0,0	0,0
4	1,0	3,0	0,5	1,0	0,0
5	2,0	0,0	2,0	1,0	3,0
6	0,5	0,5	1,0	0,0	3,0
7	0,0	3,0	3,0	3,0	3,0
8	1,0	0,5	2,0	3,0	1,0
9	0,5	2,0	3,0	1,0	0,5
10	3,0	0,5	0,0	1,0	2,0
11	0,5	3,0	0,0	2,0	1,0
12	1,0	2,0	0,0	0,5	3,0
13	0,5	1,0	2,0	0,5	0,0
14	3,0	2,0	1,0	3,0	0,0
15	3,0	1,0	0,5	2,0	3,0
16	2,0	0,5	3,0	2,0	0,0
17	3,0	3,0	2,0	0,0	0,5
18	0,0	2,0	2,0	2,0	2,0
19	3,0	0,0	3,0	0,5	1,0
20	1,0	0,0	1,0	2,0	0,5
21	0,0	1,0	1,0	1,0	1,0
22	2,0	1,0	0,0	3,0	0,5
23	2,0	2,0	0,5	0,0	1,0
24	0,0	0,5	0,5	0,5	0,5
25	2,0	3,0	1,0	0,5	2,0

5.3.2. Análise de alergênicos em perfume

Utilizou-se o cromatógrafo GC×GC-FID descrito no item 4.4.2. para as análises.

O conjunto de colunas utilizado foi uma HP-5 (30 m x 0,25 mm x 0,25 µm) conectada a uma SPwax (1,0 m x 0,10 mm x 0,10 µm). As temperaturas do injetor e do detector foram mantidas a 250 °C. Como gás de arraste utilizou-se hidrogênio a uma vazão de 0,60 mL/min. Em todas as corridas cromatográficas o volume de injeção foi de 1,0 µL no modo *split* (*split ratio* 1:30). A programação de temperatura do forno cromatográfico foi: 60 °C a 240 °C (3 °C/min.). O período de modulação utilizado foi de 4 s e a taxa de aquisição 100 Hz. Foram determinados os seguintes alergênicos: geraniol, citronelol e álcool benzílico.

As amostras foram preparadas utilizando uma marca comercial de perfume e etanol como solventes. Utilizou-se 0,5 mL do perfume em cada amostra e adicionou-se etanol para completar 1,0 mL de volume total. Foi escolhida uma marca de perfume que não apresentasse nenhum dos três alergênicos, os quais foram adicionados em laboratório. A faixa de concentração de cada alergênico foi de 0 a 100 µL/L. A composição das amostras está descrita na tabela 7 seguindo planejamento fatorial *Greco-Latin Squares*⁵⁸.

Tabela 7. Composição das amostras de perfume com alergênicos adicionados em laboratório. Concentrações em $\mu\text{L/L}$.

Amostra	Geraniol	Citronelol	Alcool Benzílico
1	50	0	50
2	25	75	100
3	75	50	0
4	0	50	50
5	0	25	25
6	75	25	100
7	0	75	75
8	25	25	50
9	25	50	75
10	75	75	25
11	100	100	75
12	100	25	0
13	50	100	25
14	100	75	50
15	100	0	100
16	75	100	50
17	50	25	75
18	25	100	0
19	50	50	100
20	0	0	0
21	25	0	25
22	50	75	0
23	100	50	25
24	0	100	100
25	75	0	75

A análise multivariada dos dados cromatográficos foi realizada em MATLAB 6.5 (MathWorks, Natick – MA, EUA) instalado em um microcomputador AMD Athlon 64 X2 4600 Dual Core e 4Gb de memória RAM. O algoritmo iNPLS foi escrito no MATLAB 6.5 e o NPLS foi obtido do *N-way Toolbox 2.11* (www.models.kvl.uk).

5.4. Resultados e Discussões

Inicialmente, o algoritmo proposto iNPLS foi testado na análise de um conjunto de amostras contendo cinco compostos (undecano, 1-octanol, 2-octanona, cicloexanona e tolueno) em hexano. As amostras foram preparadas como descrito na seção 5.3.1. Os objetivos deste estudo inicial foram avaliar se o algoritmo associaria cada composto ao intervalo em que ele aparece no conjunto experimental dos dados e os resultados da quantificação de cada composto.

Em seguida, o algoritmo foi testado na quantificação de três alergênicos (geraniol, citrionelol e álcool benzílico) em perfume. O objetivo deste estudo foi avaliar o desempenho do algoritmo na seleção correta do intervalo e, também, na quantificação de cada alergênico em um conjunto de amostras preparadas utilizando perfume como solvente, representado assim um conjunto de amostras complexas.

5.4.1. Análise de padrões

Para a avaliação inicial do algoritmo iNPLS foram escolhidos cinco compostos de modo que apresentassem total separação no espaço cromatográfico. Os compostos escolhidos foram undecano, 1-octanol, 2-octanona, cicloexanona e tolueno. O cromatograma bidimensional obtido para a amostra 25, a qual contém os cinco analitos, está ilustrado na figura 28, na qual pode ser observada uma total separação entre eles.

Para cada composto foi elaborado um modelo iNPLS, sendo que, em cada um, o conjunto das 25 amostras preparadas (seção 5.3.1) foi dividido em 17 amostras para calibração e 8 amostras para validação, as quais foram escolhidas para compor um conjunto de validação com uma amostra com concentração 0,0% v/v, duas com 1,0% v/v, duas com 2,0% v/v e uma com 3,0% v/v. Para o modelo de calibração do undecano as amostras números 2, 4, 13, 16, 17, 18, 20 e 25 formaram o conjunto de validação. No modelo do 1-octanol o conjunto de validação continha as amostras 1, 3, 10, 12, 15, 17, 23 e 24. O conjunto de validação para o modelo da 2-octanona foi composto pelas amostras 2, 3, 9, 13,

14, 17, 20 e 23. Para o modelo da cicloexanona as amostras do conjunto de validação foram as amostras 1, 10, 13, 14, 16, 20, 21 e 24. Já para o modelo de calibração do tolueno as amostras 2, 6, 8, 11, 14, 17, 18 e 22 compuseram o conjunto de validação. O número de intervalos utilizados em cada uma das dimensões do cromatograma GC×GC e o número de variáveis latentes usados na construção do modelo iNPLS para cada um dos compostos estão na tabela 8.

Na figura 28 está ilustrado o intervalo selecionado pelo iNPLS para construir o modelo de calibração de cada composto, onde observa-se que para todos os casos o algoritmo selecionou a região correta do cromatograma. Os valores de RMSEC e RMSECV estão na tabela 8. Após a construção do modelo de calibração, realizou-se a previsão das amostras de validação e os valores de RMSEP obtidos estão na tabela 8. Na figura 29 está ilustrado o gráfico dos valores de concentração das amostras de validação versus os valores previstos pelo modelo quimiométrico para cada composto. Através dos valores de RMSEP e dos gráficos na figura 29, observa-se que os valores previstos pelo modelo iNPLS foram concordantes com as concentrações das amostras.

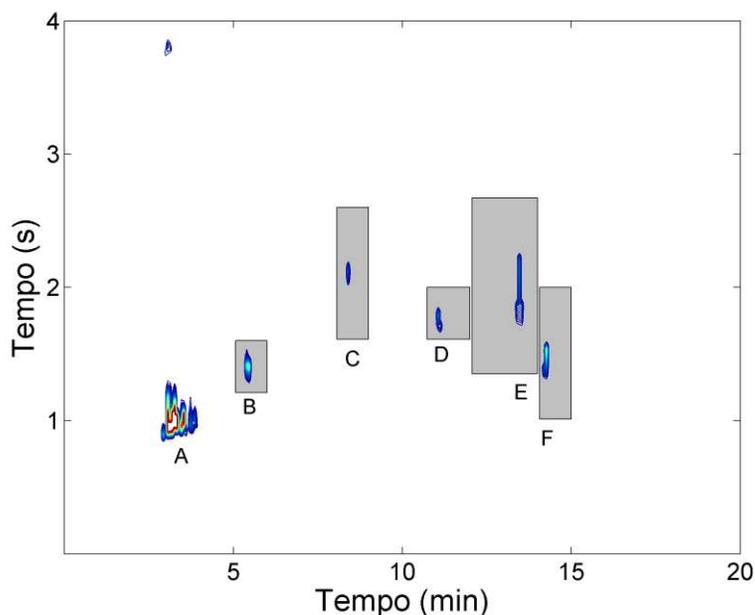


Figura 28. Diagrama GC×GC obtido para amostra 25. A área marcada em cinza representa o intervalo selecionado pelo iNPLS em cada modelo. (A) hexano. (B) tolueno. (C) ciclohexanona. (D) 2-octanona. (E) 1-octanol. (F) undecano. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 μ m) + BPX-50 (1,0 m x 0,1 mm x 0,1 μ m). Programação de temperatura do forno cromatográfico: 60 °C durante 3 min., sendo em seguida aquecido até 120 °C (3 °C/min.) e, então, aquecido até 240 °C (20 °C/min.). Razão de divisão do injetor 1:250, volume de injeção 1,0 μ L e PM 4 s.

Tabela 8. Número de intervalos e variáveis latentes utilizados em cada modelo de calibração. Valores de RMSEC, RMSECV e RMSEP obtidos.

	1D ^a	2D ^b	LV ^c	RMSEC (%)	RMSECV (%)	RMSEP (%)
Tolueno	20	10	2	0,12	0,16	0,11
Ciclohexanona	20	5	2	0,19	0,25	0,17
2-Octanona	15	10	2	0,20	0,24	0,17
1-Octanol	10	3	3	0,20	0,29	0,17
Undecano	20	4	3	0,09	0,13	0,10

^a número de intervalos na primeira dimensão.

^b número de intervalos na segunda dimensão.

^c variáveis latentes.

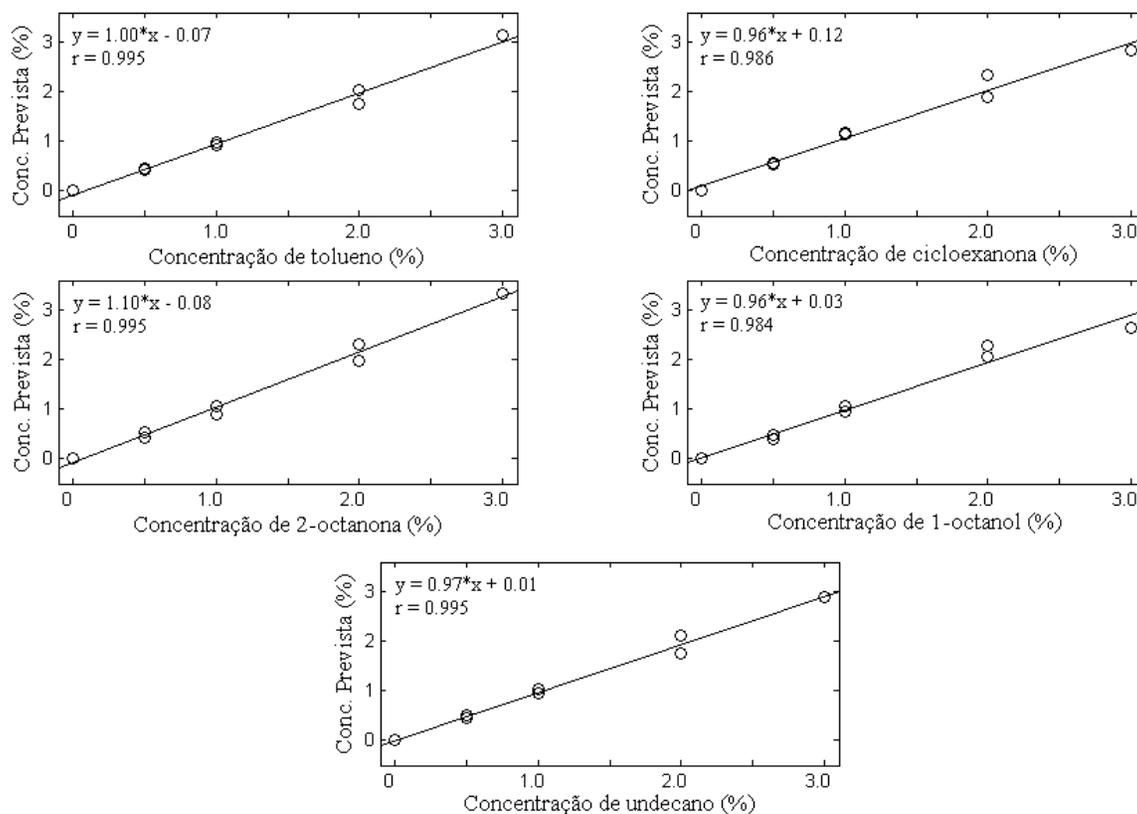


Figura 29. Concentrações previstas pelo modelo iNPLS para cada analito.

5.4.2. Análise de alergênicos em perfume

Após o iNPLS ter selecionado corretamente os intervalos e quantificado cada composto, o algoritmo foi testado em um conjunto de amostras mais complexo. Para isso, uma marca comercial de perfume, a qual não apresentava os alergênicos citrionelol, geraniol e álcool benzílico, foi escolhida e contaminada com os três alergênicos em laboratório. Foram preparadas 25 amostras (tabela 7), as quais foram divididas em 17 amostras para calibração e 8 amostras para validação. As amostras de validação foram escolhidas de modo a formar um conjunto de validação com uma amostra com 0 $\mu\text{L/L}$, duas com 25 $\mu\text{L/L}$, duas com 50 $\mu\text{L/L}$, duas com 75 $\mu\text{L/L}$ e uma com 100 $\mu\text{L/L}$. O conjunto de validação para o modelo do citrionelol continha as amostras 2, 6, 9, 12, 14, 15, 16 e 23, para o modelo do geraniol continha as amostras 2, 3, 10, 12, 13, 18, 22 e 24 e para o modelo do álcool benzílico continha as amostras 1, 5, 7, 13, 14, 15, 18 e 25.

Na figura 30 está mostrado o diagrama GC×GC obtido para uma amostra contendo apenas os três alergênicos em etanol, este cromatograma será utilizado posteriormente para avaliar se o algoritmo selecionou os intervalos corretos para cada alergênico. O cromatograma obtido para a amostra 19 está ilustrado na figura 31, na qual pode-se observar uma sobreposição parcial do sinal da matriz (perfume) com o álcool benzílico e, também, a grande quantidade de picos cromatográficos presentes na amostra.

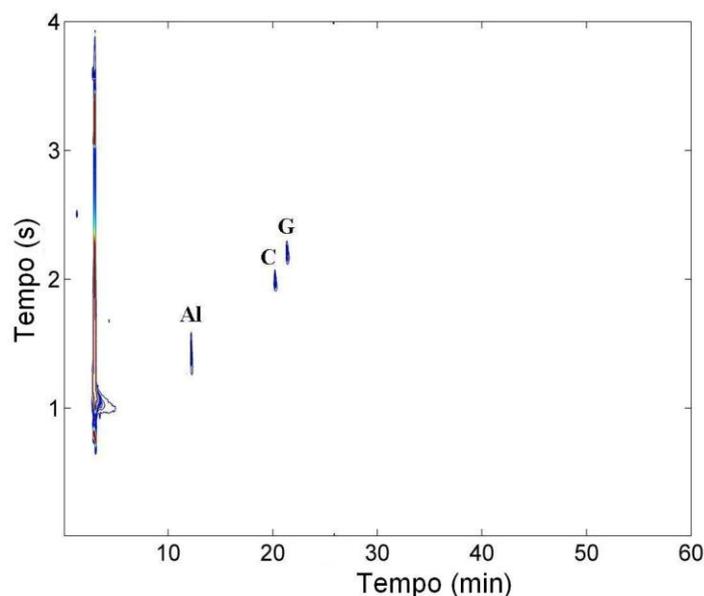


Figura 30. Diagrama GC×GC obtido para amostra contendo (Al) álcool benzílico, (C) citronelol e (G) geraniol. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 μ m) + BPX-50 (1,0 m x 0,1 mm x 0,1 μ m). Programação de temperatura do forno cromatográfico: 60 $^{\circ}$ C a 240 $^{\circ}$ C (3 $^{\circ}$ C/min.). Razão de divisão do injetor 1:30, volume de injeção 1,0 μ L e PM 4 s.

Um modelo iNPLS foi construído para cada alergênico. O número de intervalos e variáveis latentes utilizados estão na tabela 9. Comparando-se o número de intervalos utilizados para o estudo com padrões (tabela 8) com o número de intervalos utilizados para os alergênicos (tabela 9), observa-se um número consideravelmente maior para os alergênicos, isso se justifica devido ao fato que no caso dos alergênicos há muitos compostos ao redor do composto-alvo, o que não acontece no caso dos padrões. Na tabela 9 também estão

mostrados os valores obtidos para RMSEC e RMSECV. Pode ainda ser observado na tabela 9 que o número de variáveis latentes utilizado na construção dos modelos de calibração para álcool benzílico e para geraniol foi 2, enquanto que para o modelo para o citronelol foram utilizadas 5 variáveis latentes, essa diferença provavelmente foi devido a uma maior variação nos tempos de retenção de amostra para amostra no sinal cromatográfico correspondente ao citronelol em relação a variação observada para o sinal do álcool benzílico e do geraniol. Na figura 31 estão ilustrados os intervalos selecionados pelo iNPLS para o modelo de calibração de cada alergênico, na qual pode ser observado, por comparação com a figura 30, que o algoritmo novamente selecionou os intervalos corretos.

Tabela 9. Número de intervalos e variáveis latentes utilizadas em cada modelo de calibração. Valores de RMSEC e RMSECV obtidos.

Modelo	1D^a	2D^b	LV^c	RMSEC ($\mu\text{L/L}$)	RMSECV ($\mu\text{L/L}$)
Álcool Benzílico	180	4	2	11,4	15,2
Citronelol	90	11	5	6,9	9,3
Geraniol	60	10	2	8,9	12,3

^a número de intervalos na primeira dimensão.

^b número de intervalos na segunda dimensão.

^c variáveis latentes.

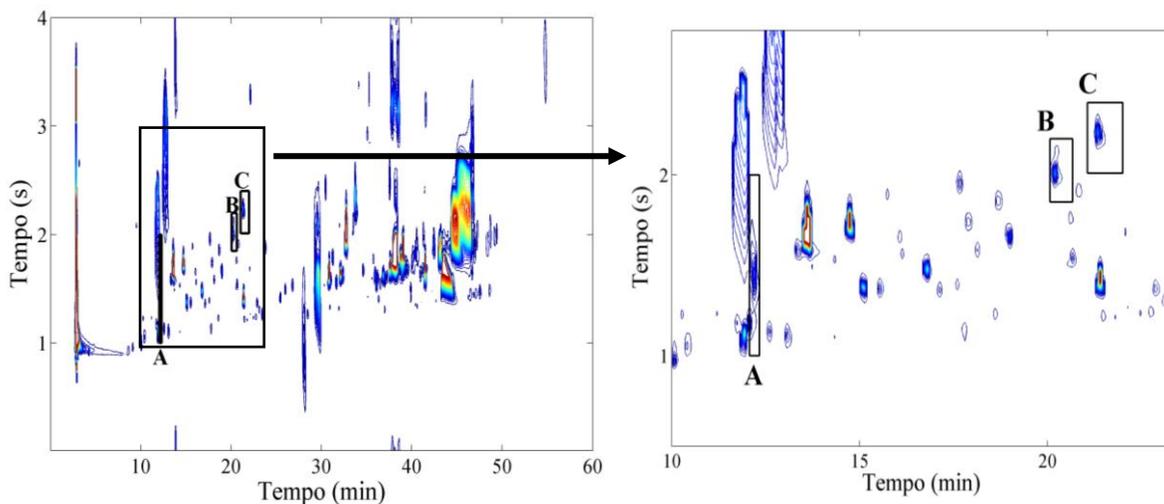


Figura 31. Diagrama GC×GC obtido para a amostra 19. O quadrado preto indica os intervalos selecionados pelo iNPLS para o modelo de (A) álcool benzílico, (B) citronelol e (C) geraniol. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 µm) + SPwax (1,0 m x 0,1 mm x 0,1 µm). Programação de temperatura do forno cromatográfico: 60 °C a 240 °C (3 °C/min.). Razão de divisão do injetor 1:30, volume de injeção 1,0 µL e PM 4 s.

Depois de construído os modelos iNPLS de calibração para cada alergênico realizou-se a previsão das amostras do conjunto de validação. Na tabela 10 estão mostrados os valores de RMSEP obtidos. Os gráficos da concentração das amostras de validação versus a concentração prevista pelo modelo estão ilustrados na figura 32. Analisando os valores de RMSEP e os gráficos da figura 32, observa-se que assim como para os padrões, o modelo previu valores de concentrações concordantes com as concentrações presentes nas amostras.

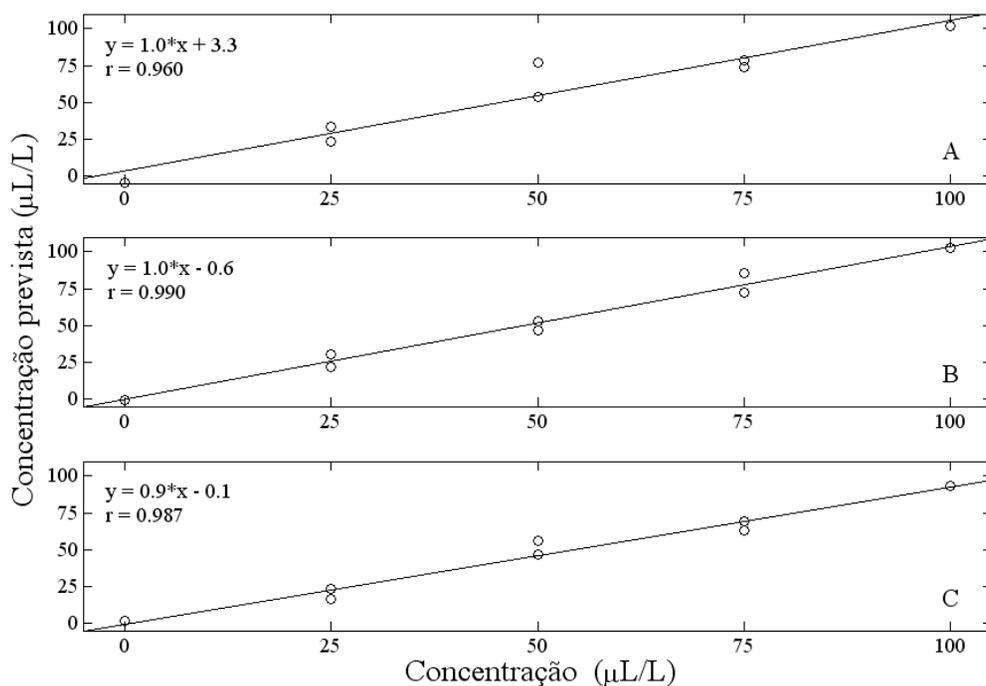


Figura 32. Concentrações previstas pelo modelo iNPLS para (A) álcool benzílico, (B) citronelol e (C) geraniol.

Tabela 10. Comparação dos resultados obtidos por iNPLS, iPLS e NPLS.

	RMSEP($\mu\text{L/L}$)		
	iNPLS	iPLS	NPLS
Álcool Benzílico	10,2	32,4	28,6
Citronelol	4,5	6,5	39,3
Geraniol	6,5	11,2	44,9

Modelos de calibração utilizando os algoritmos NPLS e iPLS foram construídos para comparar com o iNPLS. Os resultados obtidos estão na tabela 10, na qual pode ser observado que tanto o NPLS quanto o iPLS apresentaram valores de previsão com erros maiores que o iNPLS. Como discutido anteriormente (seção 2.2.), o método NPLS utiliza toda a informação experimental para correlacionar com a concentração das amostras e construir o modelo de calibração. Na figura 29 observa-se uma grande quantidade de compostos presentes em concentrações consideravelmente maiores que os compostos-alvos,

no iNPLS somente a região cromatográfica onde está presente o composto-alvo foi utilizada na construção do modelo, enquanto que para o modelo NPLS todos os compostos presentes na análise foram incluídos no modelo. Deste modo, foi mostrada a vantagem de se utilizar somente a parte dos dados experimental na qual está presente o composto de interesse, em vez de utilizar toda a informação experimental. Na seção 4.3.3. foi mencionado que métodos quimiométricos de primeira ordem, como o iPLS, podem ser utilizados para elaboração de modelos de calibração para dados de GC×GC-FID, desde que os cromatogramas bidimensionais sejam desdobrados de matrizes para vetores. Na seleção de intervalos em cromatogramas de GC×GC-FID vetorizados pelo iPLS todos os compostos separados na segunda coluna (dimensão) que apresentam eluição em tempos de retenção próximos irão ser parcialmente ou totalmente incluídos em um mesmo intervalo (figura 33). Como pode ser observado na tabela 10, o iPLS apresentou valores maiores de RMSEP que o iNPLS, principalmente para o álcool benzílico, pois este apresenta uma quantidade maior de compostos com retenção próxima na primeira dimensão, enquanto que o citronelol não apresenta outros compostos na sua faixa de retenção na primeira dimensão e, por isso, apresentou um RMSEP para o modelo iPLS próximo ao valor obtido para o iNPLS. Deste modo, foi mostrada a vantagem de utilizar somente a parte dos dados experimentais de interesse na construção do modelo de calibração e, também, a vantagem de não desdobrar os cromatogramas GC×GC-FID para seleção de intervalos⁶⁸.

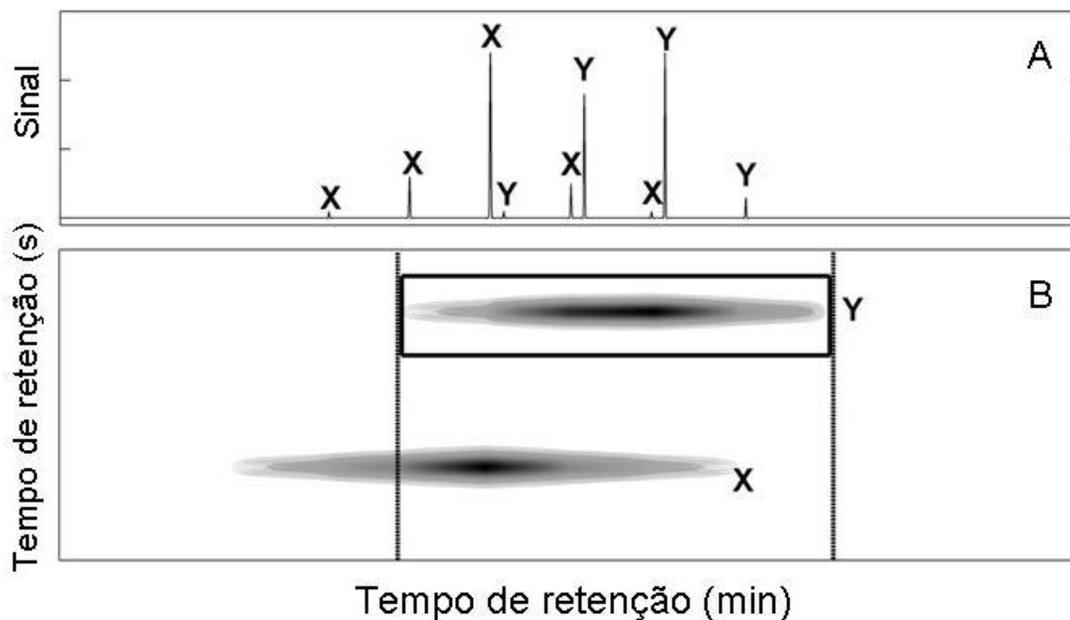


Figura 33. (A) Exemplo de um suposto cromatograma GCxGC-FID vetorizado. (B) Diagrama de contorno do mesmo cromatograma, no qual o retângulo representa o intervalo selecionado pelo iNPLS em um suposto modelo para quantificar o composto Y e a região compreendida entre as linha pontilhadas representa o intervalo selecionado na suposta quantificação de Y pelo iPLS.

5.5. Conclusões

Foi proposto o algoritmo chamado mínimos quadrados parciais múltimodo por intervalos (iNPLS), que consiste em um algoritmo para realizar a divisão de um conjunto de dados de ordem superior em intervalos, tanto na primeira como na segunda dimensão, formando assim novas matrizes de tamanho reduzido em relação à matriz inicial. Em seguida, a cada nova matriz reduzida formada, o algoritmo constrói um modelo de calibração NPLS e, após todas as matrizes reduzidas serem testadas, o iNPLS fornece uma lista com as matrizes com os dez menores valores de RMSECV. Deste modo, com essa lista e através de uma inspeção visual dos dados o analista pode decidir se o número de intervalos nas duas dimensões dos dados é adequado ou se precisa utilizar um número

diferente, de modo a isolar ao máximo a região dos dados experimentais que melhor se relaciona com a variável y .

Primeiramente, o iNPLS foi testado em um conjunto de amostras contendo cinco compostos bem separados no espaço cromatográfico. Neste teste, o algoritmo selecionou corretamente a região cromatográfica referente a cada analito e realizou a quantificação de cada um, na qual os valores previstos pelos modelos foram concordantes com a concentração das amostras.

Em seguida, o algoritmo foi avaliado frente a um conjunto de amostras complexas, que foi preparada a partir de uma marca comercial de perfume, a qual foram adicionados três alergênicos, que não estavam presentes inicialmente no perfume escolhido. Novamente, o iNPLS selecionou corretamente o intervalo correspondente a cada alergênico e foram obtidos valores previstos pelos modelos concordantes com a concentração das amostras na quantificação dos alergênicos. Realizou-se uma comparação dos resultados obtidos com o iNPLS com os resultados obtidos com NPLS e iPLS. Tanto o NPLS quanto o iPLS resultaram em modelos com maiores erros de previsão para as amostras de validação. Sendo assim, foi mostrada a vantagem do iNPLS em utilizar somente a região do cromatograma GC×GC-FID que está relacionada com a variável y para construir o modelo de calibração, em vez de utilizar todo o cromatograma como ocorre no NPLS. Também foi mostrada a vantagem de não desdobrar as matrizes cromatográficas para vetores, algo que acontece com o iPLS, na elaboração dos modelos, pois no iPLS compostos separados do composto-alvo na segunda dimensão acabam sendo incluídos no mesmo intervalo.

O método dos mínimos quadrados parciais (NPLS) não apresenta vantagem de segunda ordem, ou seja, as amostras de previsão ou desconhecidas não podem apresentar interferentes ausentes nas amostras de calibração, mas utilizando o iNPLS essa vantagem pode ser alcançada, desde que os novos interferentes não estejam no mesmo intervalo dos dados experimentais selecionado pelo iNPLS.

CAPÍTULO 6:

Análise de Óleo Essencial de Alecrim Utilizando
o Método de Resolução Multivariada de Curvas
(MCR) em Dados Obtidos por GC×GC-FID

6. Análise de Óleo Essencial de Alecrim Utilizando o Método de Resolução Multivariada de Curvas (MCR) em Dados Obtidos por GC×GC-FID

6.1. Introdução

Os estudos de quantificação de analitos em diferentes tipos de matrizes ao longo dos anos foram e ainda são objeto de intensa pesquisa na área da química analítica. O procedimento convencional de quantificação consiste em correlacionar um sinal instrumental, o qual pode ser a área de um pico cromatográfico ou o valor de absorvância em um determinado comprimento de onda, com a concentração de um composto alvo. Este procedimento univariado é o método mais tradicional utilizado em estudos de quantificação. Entretanto, em situações nas quais deseje-se quantificar um alvo que não seja composto por apenas um analito mas por uma mistura de analitos, o procedimento pode ser um tanto quanto tedioso quando não for impossível, este é o caso do estudos de óleos essenciais em perfumes ⁶⁹.

Uma alternativa para a quantificação univariada de óleos essenciais em perfumes é a busca por algum marcador que esteja presente somente no óleo em estudo. Entretanto, este marcador deve ser cuidadosamente escolhido após a análise de amostras do óleo essencial obtidas de diferentes locais e em diferentes épocas do ano ⁷⁰. Uma vez identificados, os marcadores são utilizados para a quantificação do óleo essencial em amostras simples ou complexas, como é o caso dos perfumes. Entretanto, este procedimento torna-se problemático quando o sinal instrumental do marcador do óleo essencial apresentar coeluição com componentes da matriz em análise ou quando o marcador apresentar concentração abaixo do limite de quantificação da técnica utilizada. Como mencionado anteriormente, introduzida em 1991 a cromatografia gasosa bidimensional abrangente (GC×GC) tem sido muito utilizada em estudos de amostras complexas, pois apresenta um aumento significativo na sensibilidade e detectabilidade, além de um maior poder de separação, em relação a GC

convencional⁷. Embora a GC×GC possa fornecer a capacidade de pico e a sensibilidade requerida na quantificação tradicional de óleos essenciais, ou seja, por marcadores químicos, este procedimento pode ser consideravelmente tedioso requerendo um longo tempo do analista. Nestes casos, a quantificação do marcador é realizada por integração convencional, na qual o sinal do marcador necessita estar totalmente resolvido dos outros componentes da amostra⁷¹.

Como mencionado anteriormente (capítulo 1), a quantidade de informação que pode ser obtida a partir de um cromatograma de GC×GC-FID é muito superior a informação obtida de um cromatograma de GC-FID. Sendo assim, em vez de realizar uma busca em um conjunto de amostras de um óleo essencial na tentativa de encontrar um marcador que o represente, com a utilização da quimiometria toda a informação obtida a partir de um cromatograma bidimensional pode ser utilizada em estudos de quantificação de óleos essenciais em amostras complexas.

Até o momento não há relatos na literatura do uso do método quimiométrico de resolução multivariada de curvas (MCR), cujas vantagens estão descritas no capítulo 2, em dados obtidos a partir de GC×GC-FID. Deste modo, o objetivo desta parte do trabalho foi a utilização de MCR para quantificação de óleo essencial de alecrim em amostras analisadas por GC×GC-FID e compostas de interferentes ausentes nas amostras de calibração para utilizar a vantagem de segunda ordem do algoritmo.

6.2. Objetivos

Elaboração de um modelo de calibração para quantificação de óleo essencial de alecrim em amostras complexas preparadas em laboratório utilizando o método da resolução de curvas multivariada (MCR) para construção do modelo multivariado de calibração e GC×GC-FID para análise das amostras.

6.3. Parte Experimental

Utilizou-se o cromatógrafo descrito no item 4.4.2. para as análises.

O conjunto de colunas utilizado foi uma HP-5 (30 m x 0,25 mm x 0,25 µm) conectada a uma Supelcowax (1,0 m x 0,10 mm x 0,10 µm). As temperaturas do injetor e do detector foram mantidas a 250 °C. Como gás de arraste utilizou-se hidrogênio a uma vazão de 0,60 mL/min. Em todas as corridas cromatográficas o volume de injeção foi de 1,0 µL no modo *split* (*split ratio* 1:200). A programação de temperatura do forno cromatográfico foi: 60 °C a 250 °C (3 °C/min.). O período de modulação utilizado foi de 6 s e a taxa de aquisição 100 Hz. Os cromatogramas foram adquiridos e digitalizados no software Chemstation (Agilent Technologies, Wilmington, DE).

Preparou-se cinco amostras de calibração nas concentrações de 2,5, 5,0, 7,5, 10,0 e 15,0% v/v de óleo essencial de alecrim em álcool de cereais. Para validação do modelo foram preparadas oito amostras utilizando álcool de cereais como solvente e óleo essencial de abacaxi e uma marca de perfume comercial como interferentes. A composição das amostras de validação está mostrada na tabela 11.

Tabela 11. Composição das amostras do conjunto de validação.

Amostra	Alecrim (% v/v)	Abacaxi (% v/v)	Perfume (% v/v)
V1	8,0	-	-
V2	12,0	-	-
V3	5,0	2,0	-
V4	4,0	4,0	-
V5	5,0	10,0	-
V6	5,0	6,0	-
V7	8,0	4,0	-
V8	4,0	-	6,0
V9	6,0	-	80,0
V10	5,0	-	95,0

Os cromatogramas foram adquiridos e digitalizados através do software Chemstation (Agilent Technologies, Wilmington, DE), Em seguida, foram exportados para o MATLAB 6.5 (MathWorks, Natick – MA, EUA), no qual foi

realizada a análise multivariada e o alinhamento dos dados cromatográficos utilizando o algoritmo peakmatch⁷². O algoritmo MCR foi obtido a partir do site do grupo de pesquisas do professor R. Tauler⁷³. O MATLAB estava instalado em um microcomputador AMD Athlon 64 X2 4600 Dual Core com 4Gb de memória RAM.

6.4. Resultados e Discussões

No método de resolução de curvas multivariado (MCR), a matriz de dados obtida experimentalmente deve ser organizada de maneira bidimensional. Deste modo, os cromatogramas bidimensionais obtidos através da GC×GC-FID para todas as amostras (conjunto de calibração e validação) foram desdobrados de matrizes para vetores. Em seguida, os vetores foram colocados um abaixo do outro formando a matriz bidimensional **D**, na qual o número de linhas corresponde ao número de amostras e o número de colunas corresponde ao número de variáveis obtidas em cada análise. Após a otimização por mínimos quadrados alternantes (ALS) do modelo MCR, foi obtida a matriz de concentrações **C** e a matriz com os perfis experimentais **S**, a qual corresponde aos cromatogramas vetorizados de cada componente puro presente no conjunto de dados. Por último, os cromatogramas vetorizados obtidos foram rearranjados para a forma bidimensional e comparados com o cromatograma bidimensional obtido para cada componente puro. Além disso, com as informações obtidas a partir da matriz **C** para as amostras de calibração construiu-se uma curva analítica e realizou-se a previsão das amostras de validação. A matriz **C** apresenta o número de linhas igual ao número de amostras e o número de colunas igual ao número de componentes presentes no conjunto de dados **D** (figura 34).

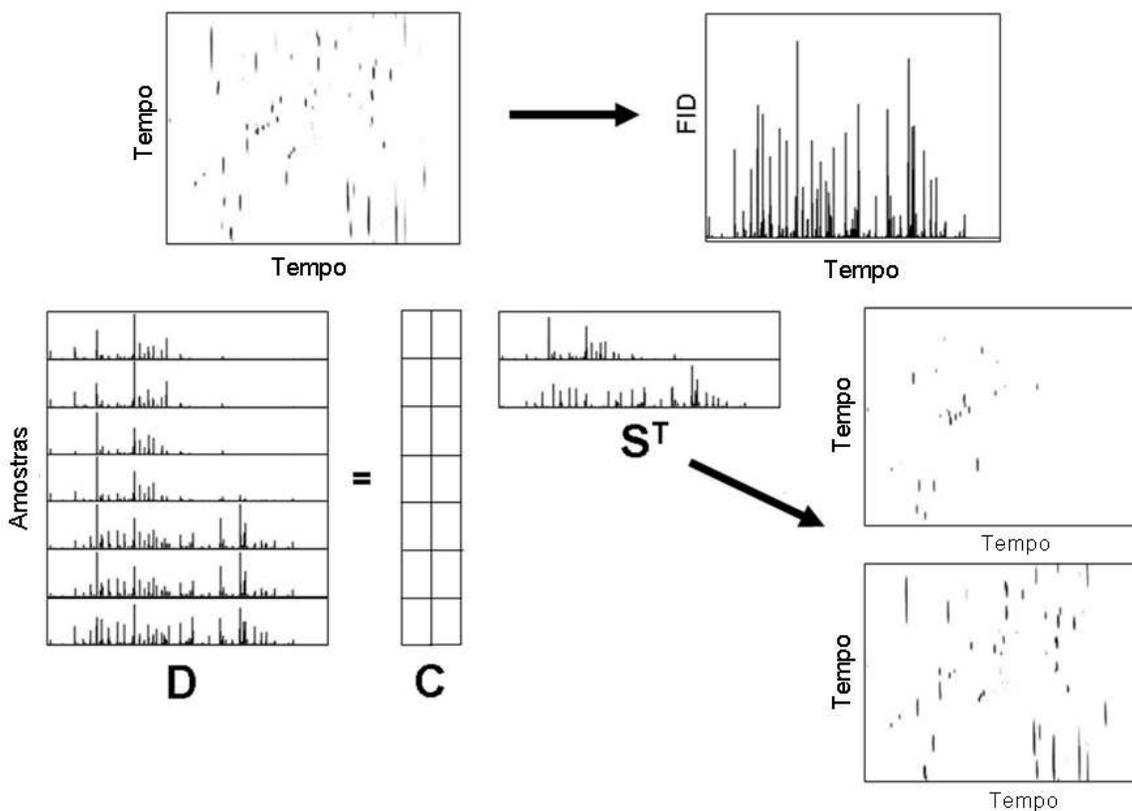


Figura 34. Esquemática para utilização de MCR-ALS em dados obtidos por GCxGC-FID.

O cromatograma obtido da análise da amostra de 15% v/v de óleo essencial de alecrim em álcool de cereais por GCxGC-FID está representado na figura 35, na qual pode ser observado um largo pico que espalha-se por toda a segunda dimensão em aproximadamente 10,5 min no tempo de retenção da primeira coluna. Este pico corresponde a coeluição de glicóis, os quais são utilizados em muitos produtos da indústria de perfumes como solventes e também com a função de fixar a fragrância. Devido à natureza altamente polar destes compostos e também da fase estacionária (polietilenoglicol) da segunda coluna, este pico apresenta forte interação com a fase estacionária e aparece arrastado por toda a segunda dimensão. Estes compostos não foram adicionados às amostras durante seu preparo, eles já estavam presentes nas fragrâncias (óleo essencial de alecrim e abacaxi) e no perfume utilizado. Na construção do modelo quimiométrico foi observado que esta região (tempo de retenção entre 10,2 a 14,4 min na primeira

dimensão) e a região correspondente ao pico cromatográfico do solvente (tempo de retenção na primeira coluna menor que 3,5 min) não acrescentavam nenhuma informação ao modelo e, sendo assim, foram excluídos dos dados cromatográficos para construir o modelo de calibração MCR.

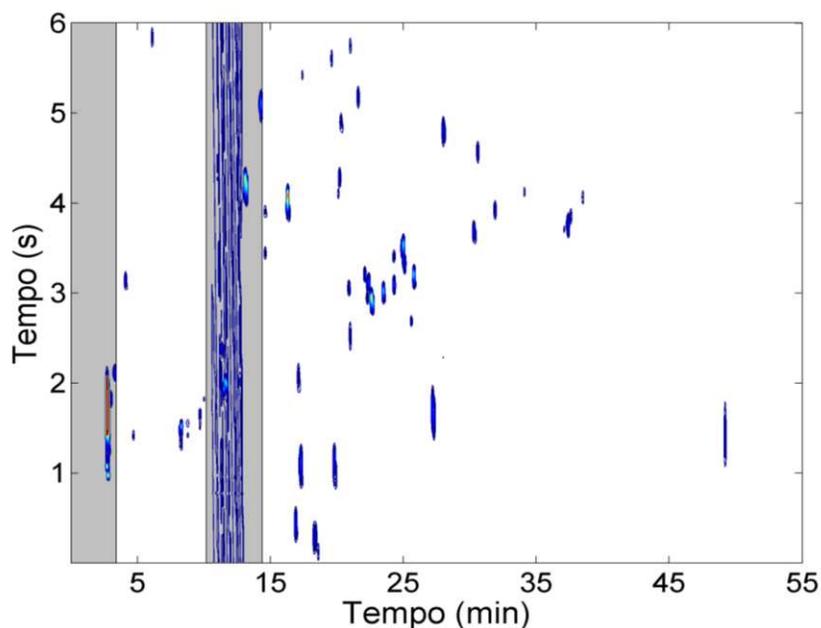
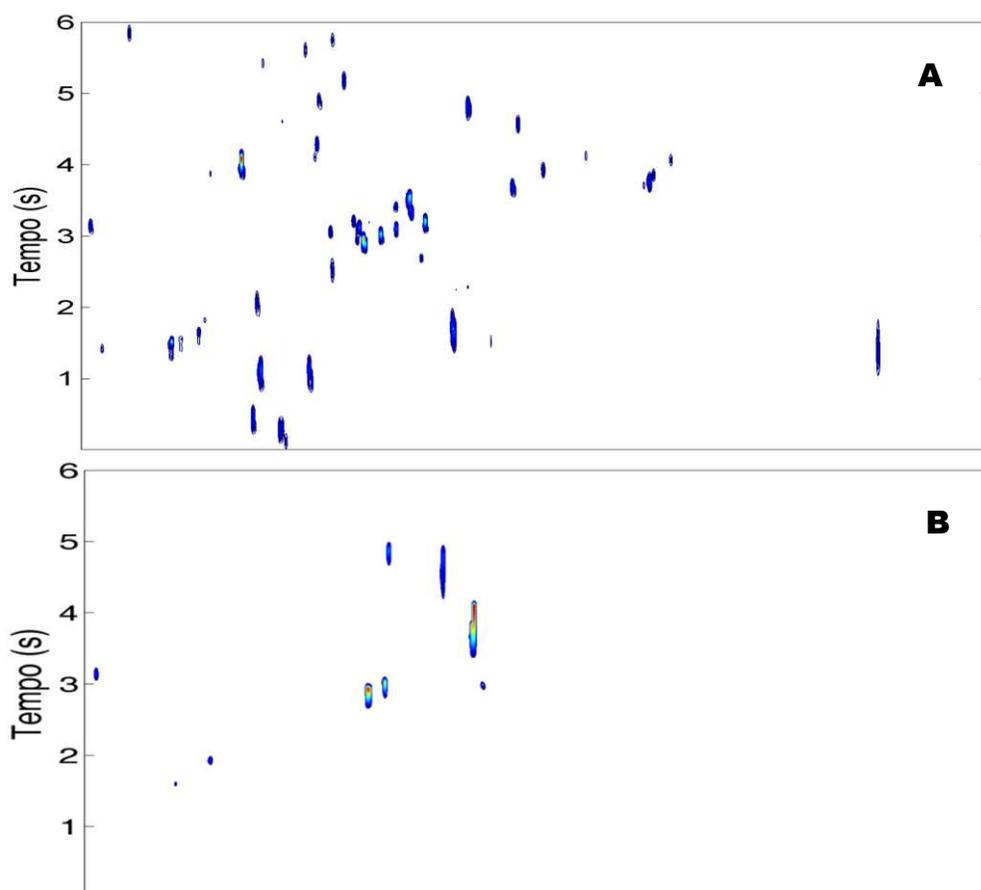


Figura 35. GCxGC-FID obtido na análise da amostra contendo 15% v/v de óleo essencial de alecrim em álcool de cereais. Regiões que foram excluídas para a construção do modelo quimiométrico destacadas em cinza. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 μ m) + Supelcowax (1,0 m x 0,10 mm x 0,10 μ m). Programação de temperatura do forno cromatográfico: 60 °C até 250 °C (3 °C/min). Razão de divisão do injetor 1:200, volume de injeção 1,0 μ L e PM 6 s.

Na figura 36 estão ilustrados os cromatogramas bidimensionais, já sem as áreas descartadas para a construção do modelo, obtidos na análise de amostras puras de óleo essencial de alecrim, óleo essencial de abacaxi e do perfume utilizado neste estudo, na qual pode ser observado que alguns compostos estão presentes no óleo essencial de alecrim e no óleo de abacaxi ou no perfume e também alguns compostos que apresentam tempos de retenção próximos. Consequentemente é esperado que uma amostra contendo óleo essencial de

alecrim e óleo essencial de abacaxi ou perfume apresente regiões de coeluição, o que pode ser observado nos cromatogramas obtidos para as amostras de validação V5 e V9 (figura 37). As amostras de calibração foram preparadas somente com óleo essencial de alecrim em álcool de cereais, já nas amostras de validação foram adicionados os interferentes óleo essencial de abacaxi ou perfume, os quais estavam ausentes nas amostras de calibração, com o objetivo de utilizar a vantagem de segunda ordem característica do MCR.



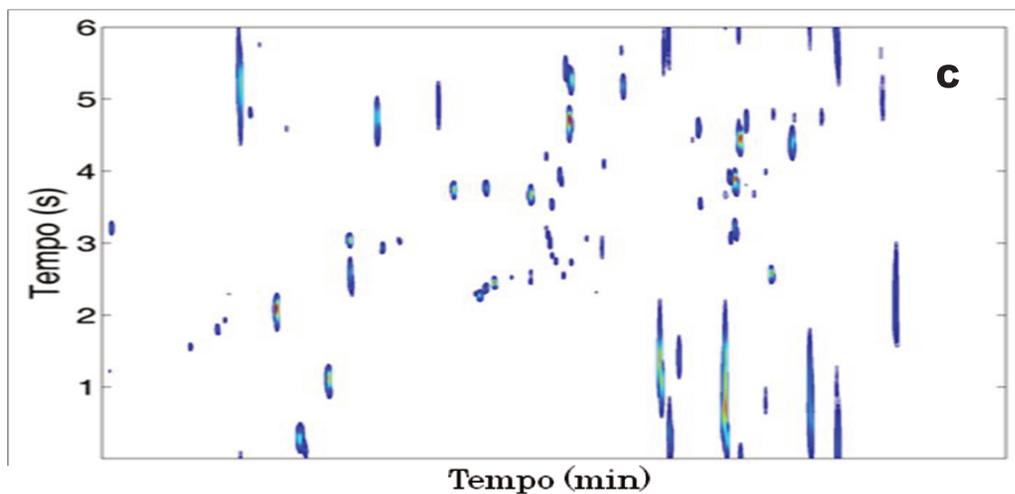


Figura 36. GC×GC-FID obtido para (A) óleo essencial de alecrim, (B) óleo essencial de abacaxi e (C) perfume. Regiões do cromatograma excluídas para a construção do modelo quimiométrico não mostradas. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 μm) + Supelcowax (1,0 m x 0,10 mm x 0,10 μm). Programação de temperatura do forno cromatográfico: 60 $^{\circ}\text{C}$ até 250 $^{\circ}\text{C}$ (3 $^{\circ}\text{C}/\text{min}$). Razão de divisão do injetor 1:200, volume de injeção 1,0 μL e PM 6 s.

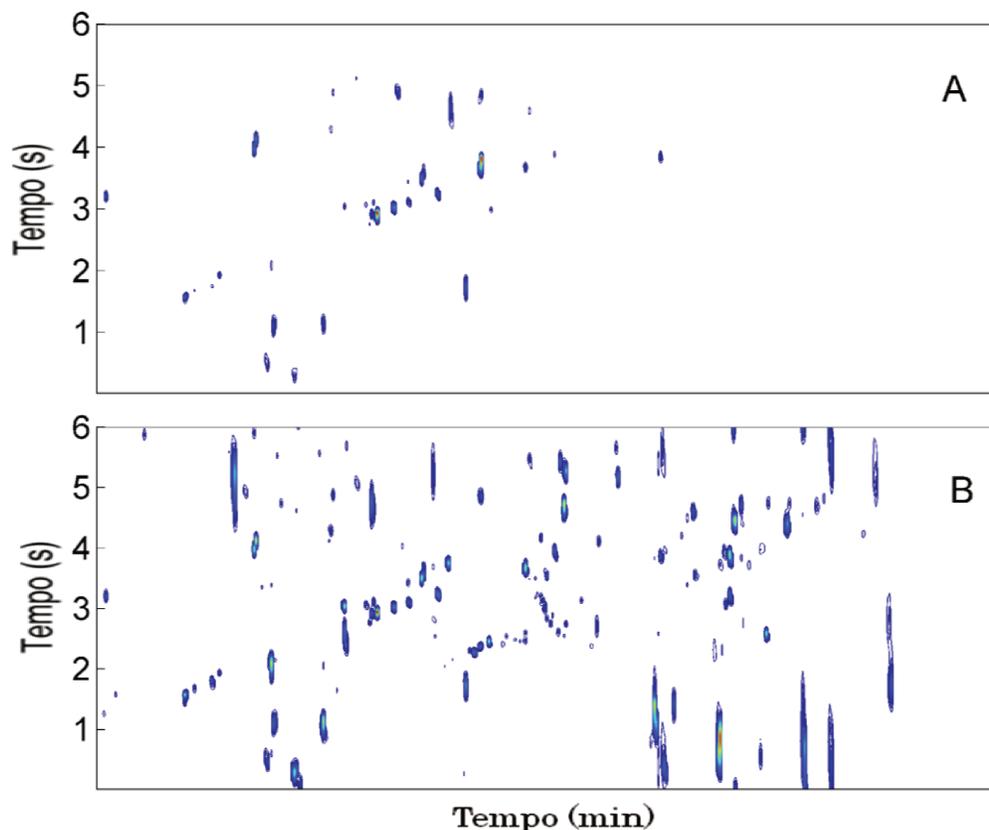


Figura 37. GC×GC-FID obtido para as amostras de validação (A) V5 e (B) V9. Regiões do cromatograma excluídas para a construção do modelo quimiométrico não mostradas. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 μ m) + Supelcowax (1,0 m x 0,10 mm x 0,10 μ m). Programação de temperatura do forno cromatográfico: 60 $^{\circ}$ C até 250 $^{\circ}$ C (3 $^{\circ}$ C/min). Razão de divisão do injetor 1:200, volume de injeção 1,0 μ L e PM 6 s.

Como estimativa inicial dos cromatogramas na construção do modelo MCR foram utilizadas as análises das amostras puras de óleo essencial de alecrim, óleo essencial de abacaxi e de perfume. Durante a otimização por ALS do modelo foram aplicadas as restrições de seletividade para as concentrações dos três produtos nas amostras e de não-negatividade para as concentrações e para os cromatogramas. A restrição de seletividade foi aplicada a fim de evitar que o algoritmo realizasse a otimização procurando os três produtos em todas as amostras e a restrição de não-negatividade foi para que as concentrações

encontradas fossem maior ou igual a zero e não fossem obtidos picos cromatográficos negativos na matriz **S**. Os cromatogramas obtidos da matriz **S** após a otimização do modelo estão ilustrados na figura 38, na qual pode ser observada a alta similaridade entre os cromatogramas das amostras puras de óleo essencial de alecrim, óleo essencial de abacaxi e do perfume e os cromatogramas recuperados do modelo MCR. Em seguida, utilizando as informações obtidas a partir da matriz **C** e a concentração de óleo de alecrim nas amostras de calibração, construiu-se uma curva analítica para determinação de óleo essencial de alecrim nas amostras de validação, na qual obteve-se um coeficiente de correlação de 0,996.

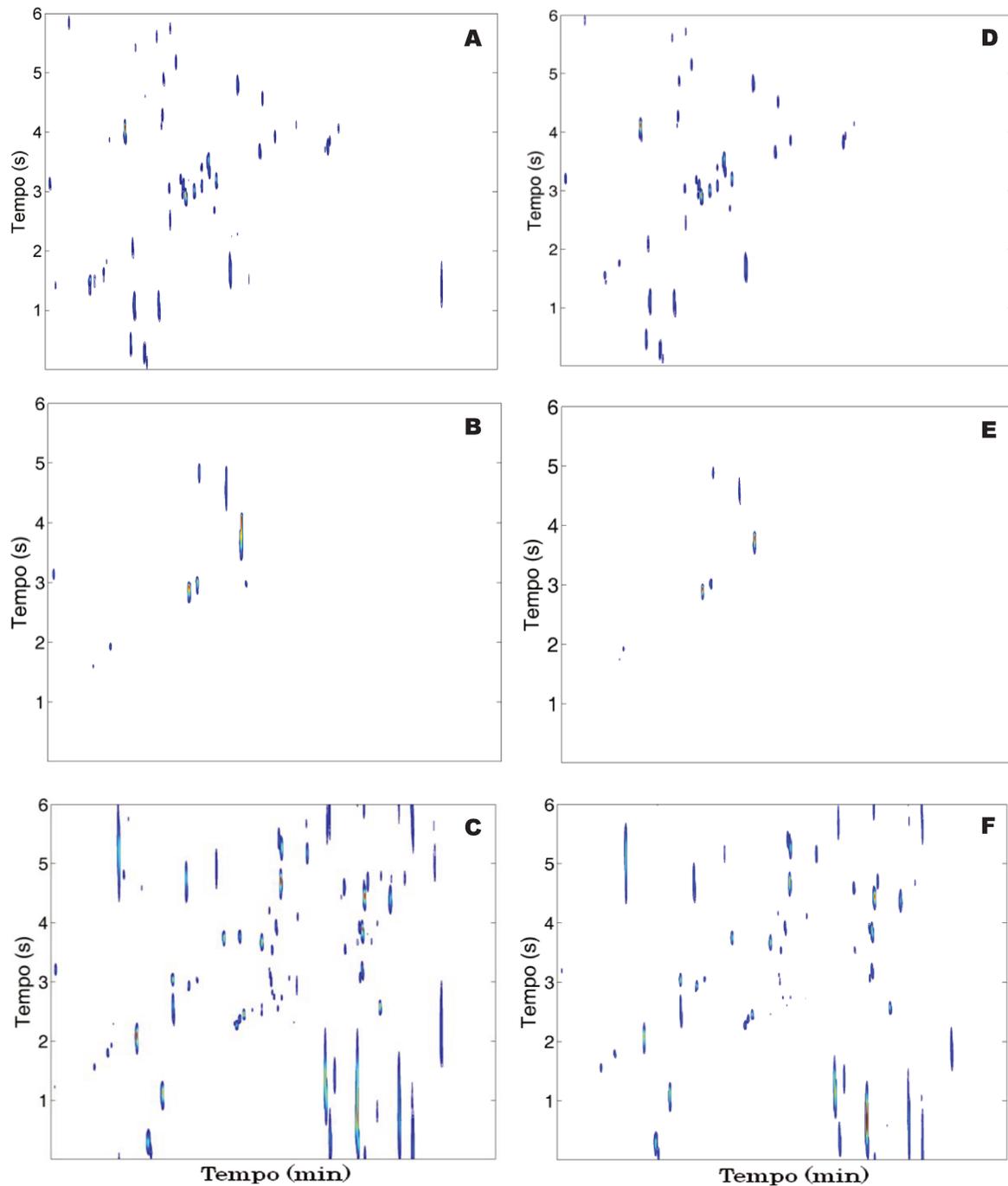


Figura 38. GC×GC-FID obtido para (A) óleo essencial de alecrim, (B) óleo essencial de abacaxi e (C) perfume. GC×GC-FID recuperados do modelo MCR para (D) óleo essencial de alecrim, (E) óleo essencial de abacaxi e (F) perfume. Conjunto de colunas: HP-5 (30 m x 0,25 mm x 0,25 μm) + Supelcowax (1,0 m x 0,10 mm x 0,10 μm). Programação de temperatura do forno cromatográfico: 60 $^{\circ}\text{C}$

até 250 °C (3 °C/min). Razão de divisão do injetor 1:200, volume de injeção 1,0 µL e PM 6 s.

Deste modo, após a determinação da curva analítica, foi realizada a previsão da concentração de óleo essencial de alecrim nas amostras do conjunto de validação. Na tabela 12 estão mostradas as concentrações previstas de óleo de alecrim para cada amostra de validação e, também, o erro obtido em cada previsão. O gráfico da concentração prevista versus a concentração de óleo essencial de alecrim nas amostras de validação está representado na figura 39, onde obteve-se um coeficiente de correlação de 0,988. Os valores de RMSPD e RMSEP obtidos foram 7,2% e 0,4% v/v, respectivamente.

Tabela 12. Valores previstos de concentração de óleo essencial de alecrim nas amostras do conjunto de validação.

Amostra	Real (% v/v)	Previsto (% v/v)	Erro (% v/v)
V1	8,0	7,3	-0,7
V2	12,0	11,4	-0,6
V3	5,0	5,5	0,5
V4	4,0	3,7	-0,3
V5	5,0	4,4	-0,6
V6	5,0	5,0	0,0
V7	8,0	7,9	0,1
V8	4,0	4,3	0,3
V9	6,0	6,2	0,2
V10	5,0	4,6	-0,4

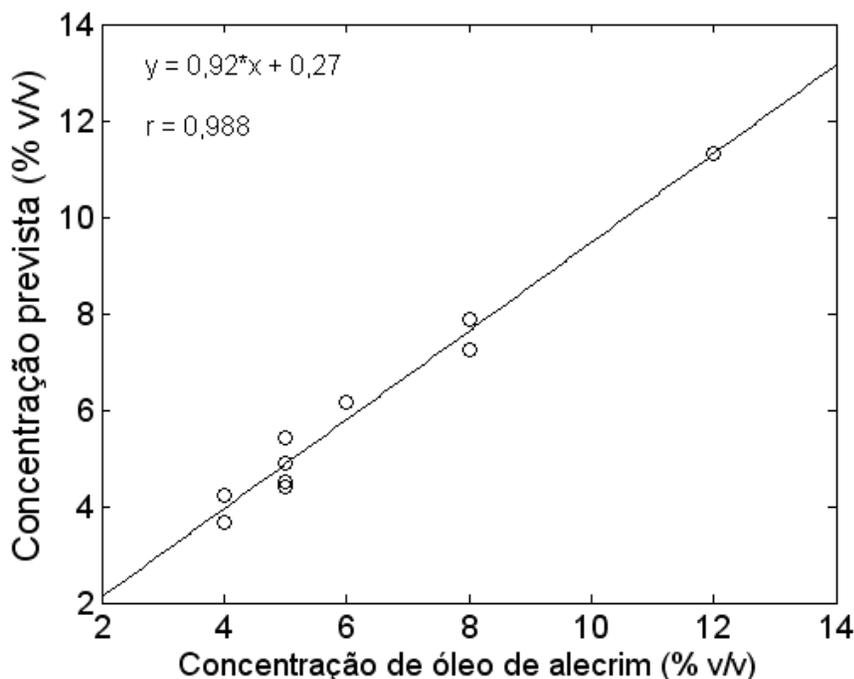


Figura 39. Concentração prevista pelo modelo MCR de óleo essencial de alecrim nas amostras de validação.

Os resultados obtidos a partir do modelo MCR-ALS construído para a determinação de óleo essencial de alecrim demonstram que este algoritmo pode ser utilizado na elaboração de modelos quimiométricos em dados obtidos a partir de GC×GC-FID. O cromatograma bidimensional recuperado do modelo para o óleo essencial de alecrim foi muito similar ao cromatograma bidimensional obtido na análise de uma amostra do óleo puro. Os valores de concentração previstos pelo modelo MCR foram concordantes com a concentração de óleo essencial de alecrim mesmo na presença de interferentes ausentes nas amostras de calibração, demonstrando assim a vantagem de segunda ordem do algoritmo.

6.5. Conclusões

Os resultados obtidos mostram que o algoritmo MCR-ALS pode ser utilizado na construção de modelos de calibração multivariada em análises realizadas por cromatografia gasosa bidimensional abrangente, a qual apresenta várias vantagens sobre a GC-FID convencional, como maior detectabilidade,

maior sensibilidade e maior poder de resolução. Sendo assim, estas vantagens podem ser combinadas com as propriedades do método MCR-ALS, entre elas, a vantagem de segunda ordem, na análise de amostras complexas. Este conjunto GC×GC-FID+MCR-ALS foi testado com sucesso na quantificação de óleo essencial de alecrim em amostras com interferentes ausentes nas amostras de calibração. Após a otimização do modelo quimiométrico elaborado, os cromatogramas GC×GC-FID recuperados foram muito similares aos cromatogramas GC×GC-FID das amostras puras de cada um das três misturas presentes no conjunto de amostras (óleo essencial de alecrim, óleo essencial de abacaxi e perfume). Por último, com as informações obtidas do modelo foi construída uma curva analítica e realizada a quantificação do óleo essencial de alecrim nas amostras de validação, na qual obteve-se valores de RMSEP e RMSPD iguais a 0,4% v/v e 7,2%, validando assim o modelo MCR-ALS construído.

Os resultados deste estudo indicam que a combinação GC×GC-FID+MCR-ALS pode ser uma ferramenta poderosa na resolução de sinais cromatográficos em amostras com interferentes desconhecidos, assim como, na quantificação de constituintes complexos em formulações como perfume.

CAPÍTULO 7:

Conclusões Gerais

7. Conclusões Gerais

Modelos quimiométricos foram construídos utilizando diferentes algoritmos na análise de diferentes tipos de amostras por cromatografia gasosa bidimensional abrangente (GC×GC-FID), mostrando que esta combinação pode ser utilizada em análises de alta complexidade abrindo possibilidade para o desenvolvimento de novas metodologias analíticas.

Primeiramente, um modelo de calibração por mínimos quadrados parciais multimodo (NPLS) foi elaborado com amostras de gasolina adulteradas em laboratório com querosene, aguarrás e thinner e analisadas por GC×GC-FID. Em seguida, realizou-se a previsão de outras amostras adulteradas preparadas em laboratório e de outras amostras de gasolina. Os resultados obtidos foram concordantes com a composição das amostras preparadas em laboratório e, no caso de novas amostras de gasolina, os resultados foram concordantes com os resultados obtidos pela ANP, exceto para uma amostra, a qual provavelmente não foi guardada em condições adequadas e teve parte de seus compostos mais leves volatilizados. Outro estudo com amostras de gasolinas foi o desenvolvimento de modelos de calibração multivariada utilizando mínimos quadrados parciais por sinergismo de intervalos (siPLS) e algoritmo genético para a previsão da densidade e da temperatura de destilação a 10, 50 e 90% de destilado e do ponto final de destilação, estes parâmetros são usados pela ANP na avaliação de qualidade de gasolina, os quais podem envolver procedimentos tediosos e requerem grande volumes de amostras. Os resultados obtidos para as amostras do conjunto de validação foram concordantes com àqueles fornecidos pela ANP. Por último, foi desenvolvido um modelo de classificação com análise de fatores paralelos (PARAFAC) para diferenciar amostras de gasolinas brasileiras de amostras de gasolinas venezuelanas devido ao intenso contrabando de gasolina da Venezuela para o Brasil existente próximo a fronteira entre os dois países.

O algoritmo dos mínimos quadrados parciais por intervalos (iPLS) é um método quimiométrico para construção de modelos de calibração multivariada em dados de primeira ordem, no qual o conjunto de dados experimental é dividido em um número de intervalos e o intervalo que apresenta a maior correlação com a

variável dependente é selecionado para construir o modelo. Até o momento não há na literatura nenhum método quimiométrico para desenvolvimento de modelos por seleção de intervalos para dados de segunda ordem, por isso, foi proposto o método dos mínimos quadrados parciais multimodo por intervalos (iNPLS), no qual a matriz de dados é dividida em ambas as dimensões em intervalos, formando assim sub-matrizes, sendo que a sub-matriz que melhor se relaciona com a variável dependente é selecionada. O iNPLS proposto foi, inicialmente, avaliado em um conjunto de amostras preparadas em laboratório utilizando cinco compostos. Em seguida, o iNPLS foi avaliado na determinação de três alergênicos em perfume, o qual não apresentava nenhum dos alergênicos que foram adicionados em laboratório. Em todos os casos, o algoritmo selecionou o intervalo correto e realizou com sucesso a quantificação de todos os compostos nas amostras de validação.

Por último, foi desenvolvido um modelo quimiométrico para a determinação de óleo essencial de alecrim em amostras com interferentes ausentes nas amostras de calibração. Para isso, as amostras foram analisadas por GC×GC-FID e o algoritmo utilizado foi o método de resolução de curvas multivariado com otimização por mínimos quadrados alternantes (MCR-ALS). O objetivo neste estudo foi a avaliação dos resultados quantitativos obtidos com o MCR-ALS em dados obtidos por GC×GC-FID, pois, até o momento, não há nenhum relato na literatura de uso de MCR-ALS na análise de dados obtidos por GC×GC-FID. O modelo MCR-ALS construído realizou com sucesso a recuperação dos cromatogramas bidimensionais de cada uma das três misturas (óleo essencial de alecrim, óleo essencial de abacaxi e perfume) presentes no conjunto de amostras e a quantificação do óleo de alecrim nas amostras de validação. Deste modo, foi demonstrado que a combinação GC×GC-FID+MCR-ALS pode ser empregada no desenvolvimento de modelos de calibração, combinando assim as vantagens da GC×GC-FID com as vantagens do MCR-ALS.

Tabela 13. Tipos de amostras analisadas por GC×GC-FID e métodos quimiométricos utilizados.

Amostra	Métodos Quimiométricos
Gasolina	NPLS
Óleo Essencial	iNPLS
Perfume	PARAFAC
	MCR
	PLS
	iPLS
	siPLS
	Algoritmo Genético

CAPÍTULO 8:

Referências Bibliográficas

8. Referências Bibliográficas

- ¹ Collins, C. H.; Braga, G. L.; Bonato, P. S. Fundamentos de Cromatografia, Editora da UNICAMP: Campinas, Brasil, **2009**.
- ² Bertsch, W. J. *High Resolut. Chromatogr.* **1999**, *22*, 647.
- ³ Rangunathan, N.; Krock, K. A.; Klawun, C.; Sasaki, T. A.; Wilkins, C. L. *J. Chromatogr. A* **1999**, *856*, 349.
- ⁴ Giddings J. C. *Anal. Chem.* **1984**, *56*, 1258A.
- ⁵ Simmons, M. C.; Snyder, L. R. *Anal. Chem.* **1958**, *30*, 32.
- ⁶ Pedroso, M. P.; de Godoy, L. A. F.; Fidélis, C. H. de V.; Ferreira, E. C.; Poppi, R. J.; Augusto F. *Quim. Nova* **2009**, *32*, 421.
- ⁷ Lui, Z.; Philips, J. B. *J. Chromatogr. Sci.* **1991**, *29*, 227.
- ⁸ Adahchour, M.; Beens, J.; Vreuls, R. J. J.; Brinkman, U. A. T. *Trends Anal. Chem.* **2006**, *25*, 438.
- ⁹ Zini, C. A. *Scientia Chromatographica* **2009**, *1*, 31.
- ¹⁰ Pedroso, M. P. Tese de Doutorado. UNICAMP. **2009**.
- ¹¹ Beens, J.; Boelens, H.; Tijssen, R.; Blomberg, J. *J. High Resolut. Chromatogr.* **1998**, *21*, 47.
- ¹² von Mühlen, C.; Zini, C. A.; Caramão, E. B.; Marriott, P. J. *Quim. Nova* **2007**, *30*, 682.
- ¹³ Dallüge, J.; Beens, J.; Brinkman, U. A. T. *J. Chromatogr. A* **2003**, *1000*, 69.
- ¹⁴ Davis, J. M.; Stoll, D.R.; Carr, P.W. *Anal. Chem.* **2008**, *80*, 8122.
- ¹⁵ Ferreira, M. M. C.; Antunes, A. M.; Melgo, M. S.; Volpe, P. L. O. *Quim. Nova* **1999**, *22*, 724.
- ¹⁶ Otto, M. *Chemometrics*. Weihem Wiley, **1999**.
- ¹⁷ Daszykowski, M.; Walczak, B. *Trends Anal. Chem.* **2006**, *25*, 1081.
- ¹⁸ Escandar, G. M.; Faber, N. M.; Goicoechea, H. C.; de la Peña, A. M.; Olivieri, A. C.; Poppi, R. J. *Trends Anal. Chem.* **2007**, *26*, 725.
- ¹⁹ J. N. Miller e J. C. Miller *Statistics and chemometrics for analytical chemistry*. 4^a ed., Prentice Hall, New York, **2000**.
- ²⁰ Bro, R. *Anal. Chim. Acta* **2003**, *500*, 185.
- ²¹ Ribeiro, J. S.; Augusto, F.; Salva, T. J. G.; Thomaziello, R.; Ferreira, M. M. C. *Anal. Chim. Acta* **2009**, *634*, 172.
- ²² Booksh, K. S.; Kowalski, B. R. *Anal. Chem.* **1994**, *66*, 782A.
- ²³ Rinnan, A.; Riu, J.; Bro, R. *J. Chemometr.* **2007**, *27*, 76.
- ²⁴ Breton, R. G. *Analyst* **2000**, *125*, 2125.
- ²⁵ Bro, R. *J. Chemometr.* **1996**, *10*, 47.
- ²⁶ Bro, R. Tese de Doutorado. Universidade de Amsterdam, Holanda, **1998**.
- ²⁷ Harshman, R. A. *UCLA Working Papers in Phonetics* **1970**, *16*, 1.
- ²⁸ Carroll, J. D.; Chang, J. J. *Psychometrika* **1970**, *35*, 283.
- ²⁹ Bro, R. *Chemom. Intell. Lab. Syst.* **1997**, *38*, 149.
- ³⁰ Faber, N. M.; Bro, R.; Hopke, P. K. *Chemom. Intell. Lab. Syst.* **2003**, *65*, 119.
- ³¹ Sena, M. N.; Trevisan, M. G.; Poppi, R. J. *Quim Nova* **2005**, *28*, 910.
- ³² Bro, R.; Kiers, H. A. I. *J. Chemometr.* **2003**, *17*, 274.

- ³³ Tucker, L. *Psychometrika* **1966**, *31*, 279.
- ³⁴ Tauler, R. Smilde, A.; Kowalski, B. *J. Chemometr.* **1995**, *9*, 31.
- ³⁵ Maeder, M.; Zuberbuehler, A. D. *Anal. Chim. Acta* **1986**, *181*, 287.
- ³⁶ Izquierdo-Ridorsa, A.; Saurina, J.; Hernhdez-Cassou, S.; Tauler, R. *Chemom. Intell. Lab. Syst.* **1997**, *38*, 183.
- ³⁷ Norgaard, L.; Saudland, A.; Wagner, J.; Nielsen, J. P.; Munck, L.; Engelsen, S. *B. Appl. Spectrosc.* **2000**, *54*, 413.
- ³⁸ Munck, L.; Nielsen, J. P.; Moller, B.; Jacobsen, S.; Sondergaard, I.; Engelsen, S. B.; Norgaard, L.; Bro, R. *Anal. Chim Acta* **2001**, *446*, 169.
- ³⁹ da Costa Filho, P. A.; Poppi, R. J. *Quim. Nova* **1999**, *22*, 405.
- ⁴⁰ Valderrama, P. Tese de Doutorado. UNICAMP. **2009**.
- ⁴¹ Shukla, K. K. *Inform. Software Tech.* **2000**, *42*, 701.
- ⁴² Carneiro, R. L.; Braga, J. W. B.; Bottoli, C. B. G.; Poppi, R. J. *Anal. Chim. Acta* **2007**, *595*, 51.
- ⁴³ Monteiro, M. R.; Ambrozin, A. R. P.; Lião, L. M.; Boffo, E. F.; Tavares, L. A.; Ferreira, M. M. C.; Ferreira, A. G. *Energy Fuels* **2009**, *23*, 272.
- ⁴⁴ Wiedemann, L. S. M.; d'Avila, L. A.; Azevedo, D. A. *J. Braz. Chem. Soc.* **2005**, *16*, 139
- ⁴⁵ Skrobot, V. L.; Castro, E. V. R.; Pereira, R. C. C.; Pasa, V. M. D.; Fortes, I. C. P. *Energy Fuels* **2005**, *19*, 2350.
- ⁴⁶ Portaria ANP número 309/2001.
- ⁴⁷ Flumignan, D. L.; Boralle, N.; de Oliveira, J. E. *Talanta* **2010**, *82*, 99.
- ⁴⁸ <http://www.anp.gov.br> acessado em 10/2010.
- ⁴⁹ Pereira, . C.; Skrobot, V. L.; Castro, E. V. R.; Fortes, I. C. P.; Pasa, V. M. D. *Energy Fuels* **2006**, *20*, 1097.
- ⁵⁰ Moreira, L. S.; d'Avila, L. A.; Azevedo, D. A. *Chromatographia* **2003**, *58*, 501.
- ⁵¹ Flumignan, D. L.; Anaia, G. C.; Ferreira, F. de O.; Tininis, A. G.; de Oliveira, J. E. *Chromatographia* **2007**, *65*, 617.
- ⁵² Flumignan, D. L.; Tininis, A. G.; Ferreira, F. de O.; de Oliveira, J. E. *Anal. Chim. Acta* **2007**, *595*, 128.
- ⁵³ Wiedemann, L. S. M.; d'Avila, L. A.; Azevedo, D. A. *Fuel* **2005**, *84*, 467.
- ⁵⁴ Flumignan, D. L.; Ferreira, F. de O.; Tininis, A. G.; de Oliveira, J. E. *Chemom. Intell. Lab. Syst.* **2008**, *92*, 53.
- ⁵⁵ Ré-Poppi, N.; Almeida, F. F. P.; Cardoso, C. A. L.; Raposo Jr.; J. L.; Viana, L. H.; Silva, T. Q.; Souza, J. L. C.; Ferreira, V. S. *Fuel* **2009**, *88*, 418.
- ⁵⁶ Micyus, N. J.; McCurry, J. D.; Seeley, J. V. *J. Chromatogr. A* **2005**, *1086*, 115.
- ⁵⁷ <http://www.models.kvl.dk>. Acessado em 10/2007.
- ⁵⁸ Morgan E.; *Chemometrics: Experimental Design*. Chichester, Inglaterra. John Wiley & Sons. **1997**.
- ⁵⁹ Pedroso, M. P.; de Godoy, L. A. F.; Ferreira, E. C.; Poppi, R. J.; Augusto, F. *J. Chromatogr. A* **2008**, *1201*, 176.
- ⁶⁰ <http://models.kvl.dk>. Acessado em 05/2008.
- ⁶¹ Kennard, R. W.; Stone, L. A. *Technom.* **1969**, *11*, 137.
- ⁶² ASTM D86. Standard test method for distillation of petroleum products at atmospheric pressure. ASTM International Standard Methods.

- ⁶³ Takeshita, E. V.; Rezende, R. V. P.; de Souza, S. M. A. G. U.; de Souza, A. A. U. *Fuel* **2008**, *87*, 2168.
- ⁶⁴ Trevisan, M. G.; Poppi, R. J. *Talanta* **2008**, *75*, 1021.
- ⁶⁵ ASTM D4052. Standard test method for density, relative density and API gravity of liquids by digital density meter. ASTM International Standard Methods.
- ⁶⁶ de Godoy, L. A. F.; Pedroso, M. P.; Ferreira, E. C.; Augusto, F.; Poppi, R. J. *J. Chromatogr. A* **2011**, *1218*, 1663.
- ⁶⁷ Nielsen, N. P. V.; Carstensen, J. M.; Smedsgaard, J. *J. Chromatogr. A* **1998**, *805*, 17.
- ⁶⁸ de Godoy, L. A. F.; Pedroso, M. P.; Hantao, L. W.; Poppi, R. J.; Augusto, F. *Talanta* **2011**, *83*, 1302.
- ⁶⁹ van Asten, A. *Trends Anal. Chem.* **2002**, *21*, 698.
- ⁷⁰ Bicchi, C.; Liberto, E.; Matteodo, M.; Sgorbini, B; Mondello, L.; Zellner, B. A.; Costa, R.; Rubiolo, P. *Flavour Fragr. J.* **2008**, *23*, 382.
- ⁷¹ Shellie, R.; Marriot, P.; Chaintreau, A. *Flavour Fragr. J.* **2004**, *19*, 91.
- ⁷² <http://synoveclab.chem.washington.edu/>. Acessado em 10/2009.
- ⁷³ <http://www.cid.csic.es/homes/rtaqam/>. Acessado em 10/2009.

