UNIVERSIDADE DE SÃO PAULO INSTITUTO DE QUÍMICA

Programa de Pós-Graduação em Química

ANDRÉ MARCELO DE SOUZA

Análise exploratória multivariada empregando o perfil cromatográfico de compostos carbonílicos na atmosfera da cidade de São Paulo

São Paulo

Data de depósito na SPG: 21/02/2010

ANDRÉ MARCELO DE SOUZA

Análise exploratória multivariada empregando o perfil cromatográfico de compostos carbonílicos na atmosfera da cidade de São Paulo

Dissertação apresentada ao Instituto de Química da Universidade de São Paulo para Obtenção do Título de Mestre em Química Analítica

Orientadora: Profa. Dra. Lilian Rothschild

São Paulo 2010 Aluno: André Marcelo de Souza

Título da dissertação: "Análise exploratória multivariada empregando o perfil cromatográfico de compostos carbonílicos na atmosfera da cidade de São Paulo".

Dissertação apresentada ao Instituto de Química da Universidade de São Paulo para obtenção do Título de Mestre em Química Analítica

Aprovado em: _____

Banca Examinadora

Prof. Dr.	
Instituição:	
Assinatura:	
Prof. Dr.	
Instituição:	
Assinatura:	
Prof. Dr.	
Instituição:	
Assinatura:	

Dedicatória

Ao meu querido irmão Aparecido Marciano Rodrigues, que me incentivou a estudar.

AGRADECIMENTOS

Ao Instituto de Química da Universidade de São Paulo (IQ-USP), e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ) pela bolsa, que foi fundamental para o desenvolvimento do trabalho.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) por ter financiado pesquisas que permitiram de forma direta e indireta a execução deste trabalho.

À minha querida orientadora Professora Dra. Lilian Rothschild por ter me aceitado como orientando, pela imensa paciência que teve comigo durante este período e pela e presente orientadora que foi para mim.

Ao querido Professor Dr. Ronei Jesus Poppi do Instituto de Química da Universidade Estadual de Campinas (IQ-UNICAMP), pelo imenso apoio que me ofereceu, pela sua presença nos trabalhos e paciência que teve comigo durante este período.

Aos queridos colegas do LEMA, Adriano Hasegawa, Eduardo Dias, MSc. Kely Ferreira dos, MSc. Ivan Petroni, Luciana da Silva Cunha, Marcelo Luiz Araujo Lopes, Renato Vieira do Nascimento Junior, e Silvana Pisani.

Ao grupo do Prof. Dr. João Vicente de Assunção da Faculdade de Saúde Pública da (FSP-USP), que foi fundamental para a execução da campanha de amostragem realizada na FSP.

Ao grupo da Professora Maria de Fátima Andrade do Instituto de Astronomia, Geofísica e Ciências Atmosféricas (IAG-USP), que foi fundamental para que execução das campanhas realizadas no Instituto de Ciências Biomédicas (ICB-USP) e no IAG.

Aos queridos MSc. Aline Klassen e Juan Claudio Mancila pelo constante apoio, principalmente nas prévias das apresentações de seminários.

Aos técnicos da área de Química Analítica do IQ-USP, meus ex. colegas de trabalho, Maria Cristina Rodrigues Machado, Fernando Silva Lopes, Maria Perpétua Batista M. de Araújo, Roberto Rosim Bertoza, José Vinicius, Simone Tessarini Estevão, que participaram, de forma positiva, da minha vida acadêmica.

Aos professores com os quais trabalhei direta e indiretamente como técnico de laboratório no IQ-USP: Dr. Fábio Rodrigo Piovizani Rocha, Dr. Ivano Gebhardt Rolf Gutz, Dra. Nina Coichev e Dr. Pedro Vitoriano de Oliveira, onde comecei a desenvolver alguns trabalhos que me levaram a tomar a feliz decisão de fazer mestrado.

À querida MSc. Eni Cardoso Tolle pelo constante apoio e participações nos trabalhos de campo realizados durante a campanha do IAG.

Ao querido Prof. Nilton R. Fiorotto do Senai "Fundação Zerrener".

Ao querido Prof. Jivaldo Rosário dos Santos Matos, por ter participado muito da minha vida acadêmica, pelos bons momentos nas aulas de laboratório e pelo constante apoio.

À MSc. Márcia Cristina Breitkreitz sempre presente na minha vida desde os tempos de cursinho pré-vestibular, de graduação na UNICAMP e que colaborou na realização do trabalho.

À Amelina Rodrigues dos Santos: "mãe é quem cria".

A todas as pessoas que conviveram comigo durante o período em que estive no IQ-USP. Gostaria que vocês tivessem a certeza de que para sempre serão lembrados com muito carinho.

SÚMULA CURRICULAR

1. Dados pessoais

- André Marcelo de Souza
- Data de Nascimento: 09/07/1977, Jacarezinho PR
- 2. Educação
- Química Bacharelado (2001-2005) UNICAMP
- Licenciatura- (2001-2006) UNICAMP

3. Ocupação

Bolsista, SAE-UNICAMP (bolsa trabalho):

01/05/2001 a 28/02/2002 480 hs 01/03/2002 a 28/02/2003 720 hs 01/03/2003 a 28/02/2004 720 hs 01/09/2004 a 28/02/2005 360 hs 01/03/2005 a 28/02/2006 720 hs

- Técnico em Química, Fevereiro/2006 a Março/2008, IQ-USP/SP
- 2º Semestre/2008, QFL 230, IQ-USP, Princípios de análises químicas, Química Analítica.
- 1º Semestre/2009, QFL 3200, IQ-USP, Princípios de análises químicas, Química Analítica.
- Bolsista de Mestrado, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ),

4. Produção científica

4.1. Resumo dos trabalhos científicos apresentados em congressos

- Propriedades mecânicas de blendas de polietileno de alta densidade e polietileno pós-consumo com poliamida 6, A. M. Souza; M. A. S. Spinacé, M. R. Vallin; M. A. De Paoli, 26^a Reunião Anual da Sociedade Brasileira de Química, Poços de Caldas MG/Brasil, Painel (QM 004) 20 a 23 de Maio de 2002.
- Preparação e caracterização de blendas de polietileno de alta densidade e polietileno pós-consumo com poliamida 6, A. M. Souza; M. A. S. Spinacé, M. R. Vallin; M. A. De Paoli, 7º Congresso brasileiro de polímeros, Belo Horizonte MG/Brasil, Painel (1047-1048), 09 a 13 de Novembro de 2003.
- Avaliação do teor de As em peixes do canal de São Sebastião por espectrometria de absorção atômica com atomização eletrotermica, R. M. Carvalho; A. M. Souza; C. S. Nomura; P. V. Oliveira, 30^a Reunião Anual da Sociedade Brasileira de Química, Águas de Lindóia – SP/Brasil, Painel (QA 029), 31 de Maio a 3 de Junho de 2007.
- Análise estatística das concentrações atmosféricas de compostos carbonilicos em regiões afetadas pela queima da cana-de-açúcar, M. L. A. Lopes; A. M. Souza; R. J. Poppi; L. Rothschild, 14° Encontro Nacional de Química Analítica, João Pessoa – PB/Brasil, Painel (QB099), 7 a 11 de Outubro de 2007.
- Comparação entre as concentrações atmosféricas de compostos voláteis medidos antes e durante a queima da cana-de-açúcar, M. L. A. Lopes; A. M. Souza; R. J. Poppi; L. Rothschild, 4º Encontro Nacional de Química Ambiental, Aracaju – CE/Brasil, Painel (070), 11 a 14 de Março de 2008.
- Determinação de ânions solúveis em água no MP10 e a caracterização de três sítios em São Paulo por meio de análise multivariada, D. Z. Souza; M. Perpétua; B. M. Araújo, P. C. Vasconcellos, A. M. Souza; R. J. Poppi, 31^a Reunião Anual da Sociedade Brasileira de Química, Águas de Lindóia SP/Brasil, Painel (AB 113), 26 a 29 de Maio de 2008.
- Análise multivariada das concentrações de HPA medidas em indústria de cimento que co processa resíduos, R. P. Silva; J. C. Masini; A. M. Souza; R. J.

Poppi, 31^a Reunião Anual da Sociedade Brasileira de Química, Águas de Lindóia – SP/Brasil, Painel (AB 115), 26 a 29 de Maio de 2008.

- Análise quimiométrica empregando perfil cromatográfico de compostos carbonílicos na atmosfera da cidade de São Paulo A. M., Souza; E. C., Tolle; K. F., Souza; M. L. A. Lopes; R. J., Poppi; L., Rothschild, (Submetido), 5º Encontro Nacional de Química Ambiental, São Pedro –SP/Brasil, 14 a 17 de Maio de 2010.
- Relação entre formaldeído e acetaldeído na atmosfera de São Paulo frente à nova frota veicular, A. M., Souza; E. C., Tolle; K. F., Souza; M. L. A. Lopes; R. J., Poppi; L., Rothschild, (Submetido), 5º Encontro Nacional de Química Ambiental, São Pedro –SP/Brasil, 14 a 17 de Maio de 2010.
- Alinhamento de cromatogramas em CLAE para análise multivariada, A. M., Souza; L. A. F., Godoi; S. A. A. Souza; M. C. Breitkreitz; L., Rotschild; R. J., Poppi, (Submetido), 33a Reunião Anual da Sociedade Brasileira de Química, Águas de Lindóia – SP/Brasil, 28 a 31 de maio de 2010.
- Multivariate analysis of carbonyl compounds in the urban air of São Paulo City, Brazil, A. M., Souza; E. C., Tolle; K. F., Souza; M. L. A. Lopes; M. C. Breitkreitz; R. J., Poppi; L., Rothschild, (Submetido), Urban Environmental Polution, Boston/ USA, 20 a 23 de Junho de 2010

RESUMO

Souza, A.M., Análise exploratória multivariada empregando o perfil cromatográfico de compostos carbonílicos na atmosfera da cidade de São Paulo, 2010, 130 p., Dissertação de Mestrado, Programa de Pós-Graduação em Qímica, Instituto de Química, Universidade de São Paulo, São Paulo.

Na última década, a utilização de misturas de gasolina e etanol em qualquer proporção como combustível em veículos do tipo "flex" aumentou de forma significativa no Brasil. Compostos Carbonílicos (CC) pertencem a uma importante classe de poluentes emitidos pela frota veicular. A cidade de São Paulo, Brasil, tem 9,2 milhões de veículos, os quais afetam diretamente a qualidade do ar. Neste trabalho, foi determinada a concentração de quatorze CC na atmosfera, em três diferentes sítios de amostragem localizados na zona oeste da cidade de São Paulo, dois nos campus da Cidade Universitária: Instituto de Astronomia, Geofísica e Ciências Atmosféricas (IAG) e Instituto de Ciências Biomédicas (ICB) e um em Cerqueira César, Faculdade de Saúde Pública (FSP), no período de 2008/2009. Os CC foram coletados em cartucho de sílicagel, impregnada com 2,4-dinitrofenilhidrazina e analisados por Cromatografia Líguida de Alta Eficiência (CLAE) com detecção UV-visível. O conjunto de dados gerado foi caracterizado aplicando Análise de Componentes Principais (PCA) e Análise de Agrupamentos Hierárquicos (HCA). Foi utilizado o cromatograma completo, uma abordagem quimiométrica ainda não utilizada para avaliação de CC na atmosfera. Os cromatogramas das amostras foram alinhados utilizando os algoritmos Peakmatch e

Optim_COW, deslocamentos de linha base e ruído foram removidos empregando, respectivamente, a primeira derivada e o método de ajuste polinomial de Savitzky-Golay.

As amostras coletadas nos três sítios apresentaram contrações significativas de formaldeído (F), acetaldeído (A) e acetona e concentrações menores de butiraldeído, benzaldeído, valeraldeído e propionaldeído. A quantidade de acetaldeído foi superior à de formaldeído para a maioria das amostras, levando a uma razão F/A menor do que 1, fato que tem sido atribuído ao uso de etanol em larga escala como combustível veicular. Na análise univariada, foi observado um perfil semelhante nos gráficos de concentração dos CC em qualquer um dos sítios de amostragem. Na análise multivariada por PCA, a PC1 representou a quase totalidade da variação experimental (97,6%). Estes resultados podem sugerir a presença de uma única fonte de emissão. O gráfico de *scores* da PCA e o dendograma da HCA indicaram a formação de dois agrupamentos: um formado pelas amostras FSP e IAG e outro pelas amostras ICB. O gráfico de *loadings* da PCA indicou que as diferenças são devidas à concentração de todos os CC. A separação das amostras ICB das demais pode ser explicada pelas diferentes condições meteorológicas observadas no sítio ICB.

Palavras-chave: Análise Exploratória Multivariada, Compostos Carbonílicos, Atmosfera.

ABSTRACT

Souza, A.M., **Multivariate analysis using the entire chromatographic profile of carbonyl compounds in the air of São Paulo city,** 2010, 130 p., Masters Thesis -Graduate Program in Chemistry, Universidade de São Paulo, São Paulo.

In the last decade, the use of ethanol-blended-gasoline in any proportion as fuel in light vehicles (flex fuel) has significantly increased in Brazil. Carbonyl compounds (CC) are an important class of vehicular total hydrocarbons emission. São Paulo City, Brazil, has 9,2 million vehicles that directly affect the air quality. In this work, the concentration of fourteen CCs in the air of three different sites of São Paulo city was determined during 2008/2009. Two sites (IAG and ICB) were located inside São Paulo University whereas a third one was located in Cerqueira César (FSP). Compounds were collected in a silica-gel cartridge coated with 2,4-dinitro-phenylhydrazine and analyzed by CLAE with UV-Vis detection (365 nm). The sites were characterized by Principal Component Analysis (PCA) and Hierarquical Cluster Analysis (HCA). The entire chromatogram was used to multivariate calculations, which represents an innovative approach in CC studies. Chromatograms of the samples were aligned using the Optim_COW and Peakmatch algorithms, the baseline was corrected by the first derivative and instrumental noise was removed by Savitzky-Golay polynomial fit.

The samples collected at the three sites presented different amounts of formaldehyde (F), acetaldehyde (A) and ketone and minor concentrations of butiraldehyde, benzaldehyde, valeraldehyde and propionaldehyde. The acetaldehyde

content was observed to be superior to the formaldehyde, so that the F/A was lower than 1 for all sites. This fact can be related to the increasing use of ethanol as fuel. It was observed a similar profile of CC concentration in all sites studied. In PCA analysis, only one Principal Component could describe a large amount of experimental variability (97.57 %), indicating that a single emission source (vehicular) might have been present. The PCA scores graph and HCA dendogram pointed out that samples from FSP and IAG were similar whereas samples from ICB presented a different behavior. The loadings graph indicated that those differences were due to concentration of all CC. This fact could be due to different climatic conditions during sampling.

Keywords: Multivariate Analysis, Carbonyl Compounds, Atmosphere.

SUMÁRIO		
CAPÍTULO 1 - INTRODUÇÃO	22	
1.1. Poluição atmosférica na Região Metropolitana de São Paulo	23	
1.2. Emissão de compostos carbonílicos na atmosfera: fonte veicular	24	
1.3. Tratamento estatístico das medidas experimentais	26	
1.3.1. Quantificação por Mínimos Quadrados em CLAE	28	
1.3.2. Notação utilizada em análise multivariada de cromatogramas	34	
1.3.3. Pré-tratamento dos cromatogramas	36	
1.3.3.1. Alisamento	37	
1.3.3.2. Correção de linha de base	38	
1.3.3.3. Alinhamento	44	
1.4. Pré-processamento das variáveis	48	
1.4.1. Centrar os dados na média	49	
1.4.2. Escalamento da variância	49	
1.4.3. Autoescalamento	50	
1.5. Métodos quimiométricos de análise exploratória	50	
1.5.1. Análise por Componentes Principais	51	
1.5.2. Análise de Agrupamentos Hierárquicos	53	
1.6. Objetivo do trabalho	55	
CAPÍTULO 2 – PROCEDIMENTO EXPERIMENTAL	56	
2.1. Padrões e solventes	57	
2.2. Limpeza do Material	57	
2.3. Instrumentação	57	
2.4. Método de análise de compostos carbonílicos no ar	58	
2.5. Caracterização dos sítios de amostragem	61	
2.5.1. Sítio IAG	61	
2.5.2. Sítio ICB	62	
2.5.3. Sítio FSP	63	
CAPÍTULO 3 – RESULTADOS E DISCUSSÕES	65	
3.1. Tratamento univariado dos resultados		
3.1.1. Curvas de calibração e avaliação do modelo linear		

14

3.1.2. Cromatogramas	75
3.1.3. Quantificação em ng/mL	78
3.1.3.1. Quantificação - FSP	78
3.1.3.2. Quantificação - IAG	79
3.1.3.3. Quantificação - ICB	82
3.1.3.4. Razão Formaldeído/Acetaldeído (F/A)	86
3.1.4. Quantificação em ppbv ou razão de mistura	
3.2. Tratamento multivariado dos resultados	93
3.2.1. Inspeção visual dos cromatogramas	93
3.2.2. Alinhamento dos cromatogramas	94
3.2.2.1. Alinhamento dos cromatogramas dos padrões	94
3.2.2.2. Alinhamento dos cromatogramas das amostras	97
3.2.3. Correção da linha base	104
3.2.4. Pré-processamento das variáveis	105
3.2.5. Aplicação de PCA e HCA na matriz final de dados	106
3.2.5.1. Sítios IAG, ICB e FSP	107
3.2.5.2. Sítios IAG e FSP	112
3.2.5.3. Sítio ICB	115
CAPÍTULO 4: CONCLUSÕES	118
Referências Bibliográficas	122

LISTA DE ILUSTRAÇÕES

Figura 1: Produção de veículos por tipo de combustível, em porcentagem, de 1978 a Figura 2: Construção da curva de calibração, ilustrando o resíduo presente no ajuste Figura 3: Resíduos de um modelo hipotético: A) resíduos com distribuição aleatória indicando modelo adequado e B) resíduos com tendência, indicando modelo com Figura 4: Gráfico de distribuição Normal dos resíduos para um modelo hipotético, Figura 5: Gráfico de valores previstos por um modelo hipotético vs valores *Figura 6*: A) Curva analítica, intervalo de confiança e ilustração do cálculo de LD; B) Curva analítica, intervalo de confiança e ilustração do cálculo de LQ (Adaptada de Figura 7: Ilustração do processo de transformação do cromatograma em uma matriz de dados para o tratamento multivariado......35 Figura 8: Cromatograma hipotético indicando as várias partes do sinal analítico: A) sinal bruto obtido do cromatógrafo; B) sinal analítico relevante; C) deslocamento de Figura 9: Ilustração do procedimento de alisamento pelo algoritmo de Savitzky-Golay. Figura 12: Ilustração do processo de derivação e alisamento de um cromatograma: A) cromatograma original, B) cromatograma na forma de primeira derivada com uma janela muito pequena e C) cromatograma na forma de derivada com uma janela ideal para o conjunto de dados......43 Figura 13: Cromatogramas hipotéticos que apresentam deslocamento no tempo de retenção de um mesmo composto em diversas injeções......44

Figura 14: Ilustração do processo de centrar um conjunto de dados na média49
Figura 15: Ilustração do direcionamento dos primeiros dois Componentes Principais no
método PCA53
Figura 16: Métodos de ligação de amostras que podem ser empregados na HCA: A)
ligação simples e B) ligação centróide54
Figura 17: Reação de adição nucleofílica do CC com o reagente derivatizante DNPH.59
Figura 18: Cromatograma dos padrões originais sem nenhum processamento
Figura 19: Cromatograma dos padrões originais após alinhamento com identificação
dos picos67
Figura 20: Curva de calibração para o formaldeído para quantificação das amostras das
campanhas na FSP e IAG68
Figura 21: Gráfico de probabilidade Normal dos resíduos deixados pelo modelo linear
para o formaldeído69
<i>Figura 22</i> : Gráfico de resíduos vs amostra para o formaldeído70
Figura 23: Gráfico de valores previstos pelo modelo linear vs valores experimentais70
Figura 24: Curvas analíticas para a quantificação dos compostos carbonílicos
formaldeído, acetaldeído, acetona, acroleína, propionaldeído, crotonaldeído,
butiraldeído, benzaldeído utilizadas para quantificação das amostras coletadas na FSP
e IAG71
Figura 25: Curvas analíticas para a quantificação dos compostos carbonílicos
isovaleraldeído, valeraldeído, o-tolualdeído, m,p-tolualdeido, hexaldeido e 2,5-
dimetilbenzaldeído utilizadas para quantificação das amostras coletadas na FSP e IAG.
Figura 26: Curvas analíticas para a quantificação dos compostos carbonílicos
formaldeído, acetaldeído, acetona, acroleína, propionaldeído, crotonaldeído,
butiraldeído e benzaldeído utilizadas para quantificação das amostras coletadas no ICB.
Figura 27: Curvas analíticas para a quantificação dos compostos carbonílicos

isovaleraldeído, valeraldeído, o-tolualdeído, m,p-tolualdeido, hexaldeido e 2,5dimetilbenzaldeído utilizadas para quantificação das amostras coletadas no ICB.......74

Figura 28: Cromatogramas das amostras originais provenientes da campanha na FSP:
A) Cromatograma inteiro, B) pico do formaldeído aumentado
Figura 29: Cromatogramas das amostras originais provenientes da campanha IAG: A)
Cromatograma inteiro, B) pico do formaldeído aumentado76
Figura 30: Cromatogramas das amostras originais provenientes da campanha no ICB:
A) Cromatograma inteiro e B) pico do formaldeído aumentado
Figura 31: Cromatogramas das amostras originais provenientes das três campanhas:
A) Cromatograma inteiro; B) pico do formaldeído aumentado
Figura 32: Concentração dos principais analitos encontrados nas amostras coletadas
na FSP
Figura 33: Concentração dos principais CC encontrados nas amostras coletadas no
IAG81
Figura 34: Concentração dos principais CC encontrados nas amostras coletadas no
ICB
Figura 35: Comparação das concentrações de formaldeído, acetaldeído, acetona,
butiraldeído e benzaldeído nos três sítios estudados85
Figura 36: Comparação das concentrações de formaldeído e acetaldeído nos três sítios
estudados
Figura 37: Gráficos da razão formaldeído/acetaldeído (F/A) nos três sítios estudados.
Figura 38: A) Gráfico de concentração dos principais CC (ppbv) encontrados nas
amostras coletadas na FSP e B) gráfico enfatizando o perfil de concentração do
formaldeído e do formaldeído (ppbv)91
Figura 39: A) Gráfico de concentração dos principais CC (ppbv) encontrados nas
amostras coletadas no IAG e B) gráfico enfatizando o perfil de concentração do
formaldeído e do formaldeído (ppbv)92
Figura 40: A) Gráfico de concentração dos principais CC (ppbv) encontrados nas
amostras coletadas no ICB e B) gráfico enfatizando o perfil de concentração do
formaldeído e do formaldeído (ppbv)
Figura 41: Cromatograma das amostras coletadas nos três sítios enfatizando
informações não relevantes presentes no inicio do cromatograma

Figura 42: Cromatograma dos padrões após alinhamento utilizando o algoritmo COW, com amostra 1 selecionada como referência: A) cromatograma inteiro e B) pico do Figura 43: Cromatograma dos padrões após alinhamento utilizando o algoritmo COW, com amostra 13 selecionada como referência. A) cromatograma inteiro e B) pico do Figura 44: Cromatograma dos padrões após alinhamento utilizando o algoritmo COW, com amostra 25 selecionada como referência. A) cromatograma inteiro e B) pico do Figura 45: Cromatograma dos padrões após alinhamento utilizando o algoritmo Optim-Figura 46: Cromatogramas das amostras provenientes das três campanhas após alinhamento empregando algoritmo COW e amostra 1 como referência: A) Figura 47: Cromatogramas das amostras provenientes das três campanhas após alinhamento empregando algoritmo COW e amostra 16 como referência: A) Figura 48: Cromatogramas das amostras provenientes das três campanhas após alinhamento empregando algoritmo COW e amostra 33 como referência: A) Figura 49: Cromatogramas das amostras provenientes das três campanhas após o alinhamento empregando o algoritmo Peakmatch e amostra 1 como referência A) Cromatograma inteiro; B) pico do formaldeído aumentado......100 Figura 50: Cromatogramas das amostras provenientes das três campanhas após o alinhamento empregando o algoritmo Peakmatch e amostra 16 como referência A) Cromatograma inteiro; B) pico do formaldeído aumentado......101 Figura 51: Cromatogramas das amostras provenientes das três campanhas após o alinhamento empregando o algoritmo Peakmatch e amostra 33 como referência A) Cromatograma inteiro; B) pico do formaldeído aumentado......101 Figura 52: Cromatogramas das amostras originais provenientes das três campanhas após o alinhamento empregando o algoritmo Optim-COW após pré-alinhamento

empregando o algoritmo Peakmatch (Ref 1): A) Cromatograma inteiro: B) pico do
formaldeído aumentado.
<i>Figura 53</i> : Cromatogramas das amostras originais provenientes das três campanhas
após o alinhamento empregando o algoritmo Optim-COW após pré-alinhamento
empregando o algoritmo Peakmatch (Ref. 16): A) Cromatograma inteiro: B) pico do
formaldeído aumentado
Figura 54: Cromatogramas das amostras originais provenientes das três campanhas
após o alinhamento empregando o algoritmo optim-COW após pré-alinhamento
empregando o algoritmo Peakmatch (Ref. 33): A) Cromatograma inteiro; B) pico do
formaldeído aumentado
Figura 55: Cromatogramas das amostras após alinhamento, destacando os
deslocamentos de linha base104
Figura 56: Cromatograma das amostras alinhadas na forma de primeira derivada105
Figura 57: Gráfico de variância explicada ou capturada vs número de Componentes
Principais107
Figura 58: Gráfico de scores das amostras dos três sítios na PC1108
Figura 59: Gráfico de scores, PC1 vs PC2 das amostras dos três sítios108
Figura 60: Loadings na PC1 (forma derivada)109
Figura 61: Loadings na PC1 (forma integrada)110
Figura 62: Dendograma das amostra dos sítios, IAG, ICB e FSP111
Figura 63: Gráfico de scores, PC1 vs PC2 das amostras dos sítios IAG e FSP112
Figura 64: Loadings na PC1 dos sítios IAG e FSP (forma integrada)113
Figura 65: Dendograma das amostras dos sítios IAG e FSP
Figura 66: Gráfico de scores, PC1 vs PC2, das amostras do sítio ICB115
Figura 67: Loadings na PC1 do sítio ICB (forma integrada)
Figura 68: Dendograma das amostras do sítio ICB117

LISTA DE TABELAS

Tabela 1 : Padrões e solventes utilizados
Tabela 2: Informações sobre a amostragem no sítio IAG.61
Tabela 3:Informações sobre a amostragem no sítio ICB.62
Tabela 4: Informações sobre a amostragem no sítio FSP.64
Tabela 5: Concentração (em ng/mL) dos compostos carbonílicos encontrados nas
amostras coletadas na FSP78
Tabela 6: Concentração (em ng/mL) dos compostos carbonílicos encontrados nas
amostras coletadas no IAG80
Tabela 7: Concentração (em ng/mL) dos principais CC encontrados nas amostras
coletadas no ICB (Amostras 1-10)82
Tabela 8: Concentração (em ng/mL) dos principais CC encontrados nas amostras
coletadas no ICB (Amostras 11-17)83
Tabela 9: Concentração (em ppbv) dos principais CC encontrados nas amostras
coletadas na FSP89
Tabela 10: Concentração (em ppbv) dos principais CC encontrados nas amostras
coletadas no IAG90
Tabela 11: Concentração (em ppbv) dos principais CC encontrados nas amostras
coletadas no ICB (Amostras 1-10)90
Tabela 12: Concentração (em ppbv) dos principais CC encontrados nas amostras
coletadas no ICB (Amostras 11-17)91
Tabela 13: Código e informação das amostras para a análise multivariada106

CAPÍTULO 1 - INTRODUÇÃO

1.1. Poluição atmosférica na Região Metropolitana de São Paulo

A deterioração da qualidade do ar na Região Metropolitana de São Paulo (RMSP) é decorrente principalmente das emissões atmosféricas de aproximadamente 9,2 milhões de veículos (DETRAN/PRODESP, 2008). Esta frota é composta por 7,4 milhões de veículos do ciclo Otto (gasolina e álcool), 490 mil veículos a diesel e 1,2 milhão de motos, que representam cerca de 1/5 do total nacional. De acordo com as estimativas de 2008, são lançados para a atmosfera os seguintes poluentes: 1,56 milhões de t/ano de monóxido de carbono (CO), 387 mil t/ano de hidrocarbonetos (HC), 367 mil t/ano de óxidos de nitrogênio (NO_X), 62,3 mil t/ano de material particulado (MP) total e 25,5 mil t/ano de óxidos de enxofre (SO_X), sendo que os veículos são responsáveis por 98% das emissões de CO, 97% de HC, 96% de NO_X, 40% de MP e 33% de SO_X (CETESB, 2009).

A partir de 2003, no que diz respeito ao tipo de combustível utilizado, é possível verificar uma tendência de crescimento no uso do etanol devido ao aumento no número de carros Flex (Figura 1). Esta notável mudança da característica da frota veicular pode sugerir que, futuramente, haja um aumento no emprego do bio-combustível proveniente de cana-de-açúcar em veículos de passeio. (ANFAVEA, 2009).



Figura 1: Produção de veículos por tipo de combustível, em porcentagem, de 1978 a 2008.

1.2. Emissão de compostos carbonílicos na atmosfera: fonte veicular

A combustão incompleta em veículos automotivos gera uma grande quantidade de compostos orgânicos voláteis (COV), dentre os quais se destacam os compostos carbonílicos (CC). Os CC podem ser emitidos para a atmosfera a partir de uma grande variedade de fontes naturais e antrópicas ou podem ser formados *in situ* em reações de fotólise e fotoxidações. A emissão veicular, entretanto, representa a fonte direta mais significativa de emissão desses compostos. Os CC liberados para a atmosfera em nível traço podem causar impacto na saúde humana e no ambiente por serem altamente reativos e importantes precursores de radicais livres, ozônio e peroxiacetilnitratos (Carter, 1995; Pang *et al.*, 2008). Dentre os CC atmosféricos, os mais abundantes são o formaldeído (HCHO) e o acetaldeído (CH₃CHO)₂, embora exista uma fração significativa desses compostos (cerca de 10%) sob forma de propionaldeído

(CH₃CH₂CHO), propanona (CH₃COCH₃), acroleína (CH₂CHCHO), benzaldeído (C₆H₅CHO), entre outros (De Andrade et al., 2002). Compostos como formaldeído, acetaldeído e acroleína, são tóxicos, mutagênicos e/ou carcinogênicos (WHO, 2000; Kean *et al.*, 2001).

Nas últimas décadas, o estudo das emissões veiculares de CC tem recebido mais atenção dos pesquisadores devido ao aumento na utilização de combustíveis alternativos derivados de biomassa (CETESB, 2009), como por exemplo, metanol e etanol. O uso desses combustíveis tem como principal objetivo a melhoria da qualidade do ar em áreas urbanas através da redução da emissão de monóxido de carbono (CO) e hidrocarbonetos, levando a concentração atmosférica destes poluentes para valores abaixo dos exigidos nos padrões de qualidade do ar. Entretanto, as emissões de alguns produtos de combustão incompleta, por exemplo os CC, não são levadas em consideração quando se avaliam parâmetros de qualidade do ar pois para esses compostos não existe legislação específica para ambientes urbanos (De Andrade *et al.*, 2002).

Em uma das primeiras pesquisas realizadas no Brasil, Grosjean *et al.*, 1990, determinaram os níveis de CC em três locais de Salvador - BA, e relacionaram os níveis de acetaldeído ao uso de etanol como combustível veicular. De Andrade *et al.*, 2002, fizeram uma extensa revisão dos principais CC, suas fontes de emissão, reatividade, níveis de concentrações em vários locais do Brasil e do mundo e seus efeitos toxicológicos. No mundo todo, diversos trabalhos de campo têm sido realizados em áreas urbanas com tráfego intenso de veículos com o objetivo de obter informações a respeito da natureza, da magnitude e do impacto que essas emissões de CC podem causar ao meio ambiente (Kean *et al.*, 2001; Montero et *al.*, 2001; Kim et *al.*, 2007).

Resultados comparativos entre as emissões de veículos movidos a combustíveis oxigenados (metanol e etanol) e gasolina mostram que os veículos leves que utilizam álcool como combustível emitem mais aldeídos do que os veículos a gasolina. Medidas realizadas diretamente no escapamento dos veículos mostraram que veículos movidos a metanol e etanol aumentam, respectivamente, os níveis de emissões de formaldeído e acetaldeído (de Andrade, 1998; de Andrade, 2002; Corrêa *et al.*, 2003; Corrêa e Arbila; 2005; Pang *et al.*, 2006; Corrêa e Arbila; 2008; Pang *et al.*, 2008). As razões [HCHO]/[CH₃CHO] calculadas a partir de medidas feitas no exterior são sempre maiores que 1 (Guo *et al.*, 2007; Kim *et al.*, 2007;Pang *et al.*, 2009)., enquanto que para o Brasil, na maioria dos resultados apresentados recentemente, são menores que 1. Esse fato mostra um perfil distinto do Brasil em relação a outros países onde o etanol não é utilizado em larga escala como combustível automotivo, indicando que, a utilização deste pode afetar o perfil dos níveis de concentração atmosférica de CC (De Andrade *et al.*, 2002).

1.3. Tratamento estatístico das medidas experimentais

O método instrumental mais empregado para a análise de CC é a Cromatografia Líquida de Alta Eficiência (CLAE) com detecção espectrofotométrica na região do ultravioleta. Historicamente, a análise das medidas dos CC presentes no ar tem sido feita utilizando-se estatística univariada. Poucos autores empregaram análise estatística multivariada (Quimiometria) para interpretar resultados cromatográficos e, nestes casos, os valores discretos de concentração dos CC determinados por CLAE foram utilizados. Santarsiero e Fuselli, 2008, avaliaram os níveis de CC em ambientes fechados e abertos e processaram os dados empregando técnicas de análise estatística multivariada. A análise apontou as fontes de CC e intercorrelações entre as mesmas; Guo *et al.* 2007, aplicaram análise multivariada para identificar as fontes e quantificar COV, dentre eles CC, em diferentes sítios na atmosfera de Hong Kong, China.

A utilização de dados discretos de concentração de CC presentes em amostras de ar nem sempre permite uma avaliação global dos dados ambientais. A utilização do cromatograma completo entretanto, fornece mais informações sobre as amostras pois o mesmo pode ser utilizado como uma impressão digital de cada uma das amostras no que diz respeito ao perfil de CC naquele momento e nas condições utilizadas (Nielsen *et al.*, 1998; Skov *et al.*,2006). Este procedimento é de grande valia para análise de amostras de ar, considerando que os cromatogramas podem ser complexos e cada amostra possui um perfil único. Não tem sido descritos na literatura trabalhos envolvendo determinação de CC na atmosfera em centros urbanos empregando tratamento quimiométrico utilizando o cromatogramas completo. O objetivo principal do presente trabalho é utilizar os cromatogramas completos para estudos quimiométricos de reconhecimento de padrões. Para a utilização dos cromatogramas completos, é necessário, entretanto, que métodos de pré-processamento sejam aplicados às amostras e variáveis, os quais serão discutidos em tópicos subsequentes.

1.3.1. Quantificação por Mínimos Quadrados em CLAE

A quantificação por Mínimos Quadrados é feita através de uma curva de calibração contendo a concentração de padrões dos analitos como variáveis independentes (x) e os valores correspondentes de áreas como variáveis dependentes (y). Cada padrão é preparado em diversas concentrações, chamados de níveis de calibração (x_i). Os coeficientes de regressão são obtidos ajustando-se a equação do modelo aos pares (x_i,y_i). Considerando que existe um erro experimental embutido nas medidas, não é possível traçar uma reta que passe por todos os pontos, ou seja, a reta de regressão sempre deixará um resíduo, que é a diferença entre o valor experimental observado e aquele previsto pela equação do modelo (Figura 2). No método dos Mínimos Quadrados, a melhor curva é ajustada de modo a minimizar o somatório dos quadrados dos resíduos nos valores da variável dependente. Desta maneira, a análise dos resíduos tem papel fundamental na avaliação da qualidade do ajuste por Mínimos Quadrados.



Figura 2: Construção da curva de calibração, ilustrando o resíduo presente no ajuste pelo método dos Mínimos Quadrados (Adaptada de Ribeiro et al., 2008).

A solução geral para a obtenção dos coeficientes de um modelo por Mínimos Quadrados é dada pela Equação 1:

$$\mathbf{b} = (\mathbf{X}^{\mathsf{t}} \mathbf{X})^{-1} \mathbf{X}^{\mathsf{t}} \mathbf{y}$$
 Equação 1

onde b é o vetor contendo os coeficientes do modelo, X a matriz de valores da variável independente e y o vetor de valores obtidos experimentalmente para a variável dependente.

Um parâmetro muito utilizado em Química Analítica para a avaliação dos modelos de regressão é o coeficiente de determinação, R^2 , calculado como a razão entre a Soma Quadrática devido à Regressão e a Soma Quadrática Total (Bruns *et al.*, 2001). O valor de R^2 representa a porcentagem da variância experimental que é explicada pelo modelo em questão. Por este motivo, quanto maior o valor de R^2 , maior é a indicação de que o modelo é adequado para representar o conjunto de dados experimentais e normalmente o valor de R^2 é utilizado para decidir entre dois modelos. Por outro lado, o coeficiente de correlação, r, é uma medida da associação linear entre duas variáveis aleatórias (Bruns *et al.*, 2001). Para um ajuste linear, no caso hipotético de r = 1, todos os pontos da reta deverão estar em cima de uma reta com inclinação positiva. Uma relação não linear entre x e y resultará em um valor de r relativamente pequeno (Ribeiro *et al.*, 2008).

O valor de R² deve ser utilizado conjuntamente com ferramentas auxiliares, os gráficos de diagnóstico, para decidir sobre a qualidade do ajuste do modelo aos dados experimentais. Gráficos de diagnóstico incluem gráficos de distribuição dos resíduos

(Figuras 3 e 4) e gráficos de valores previstos *vs* valores experimentais (Figura 5). Considerando que o modelo esteja bem ajustado aos resultados experimentais, esperase que apenas flutuações aleatórias estejam presentes nos resíduos, as quais devem seguir uma distribuição Normal. Caso os resíduos, não se distribuam de forma aleatória (Figura 3A), apresentando um padrão de distribuição com regiões concentrando resíduos apenas positivos ou apenas negativos (Figura 3 B), existe a evidência de falta de ajuste e este modelo deve ser melhorado.



Figura 3: Resíduos de um modelo hipotético: A) resíduos com distribuição aleatória indicando modelo adequado e B) resíduos com tendência, indicando modelo com provável falta de ajuste.

Para melhor visualização do comportamento dos resíduos, muitas vezes, os mesmos são colocados em gráficos tendo no eixo y uma escala de distribuição Normal. Caso os resíduos sejam aleatórios, eles devem estar agrupados ao longo da linha deste gráfico, ou com uma leve ondulação ao redor desta linha, conforme mostrado na Figura 4. Desvios deste comportamento sugerem a não normalidade dos resíduos.



Figura 4: Gráfico de distribuição Normal dos resíduos para um modelo hipotético, indicando que não há evidência de falta de ajuste.

Na Figura 5 é mostrado o gráfico de valores previstos pelo modelo proposto *vs* valores experimentais. Se o modelo estiver bem ajustado, existirá uma boa concordância entre estes valores, ou seja, eles serão distribuídos ao longo de uma reta.



Figura 5: Gráfico de valores previstos por um modelo hipotético vs valores experimentais indicando boa concordância entre os dois valores.

Os limites de detecção (LD) e quantificação (LQ) podem ser obtidos através dos parâmetros da curva analítica que foi construída pelo método dos Mínimos Quadrados, levando-se em consideração o intervalo de confiança da regressão (Ribeiro *et al.*, 2008). A estimativa do sinal analítico a partir da equação de regressão apresenta um erro padrão, e o produto deste erro pelo valor apropriado de *t* da distribuição de Student permite calcular o intervalo de confiança da curva analítica (Figura 6A), que tem a forma de duas linhas hiperbólicas ao redor da curva. O intercepto do limite superior do intervalo de confiança é conhecido por *y* crítico (*yc*) e a sua projeção no limite inferior é uma estimativa da concentração mínima que pode ser medida com um determinado grau de confiança, ou seja, o limite de detecção do método. As Equações 2 e 3 descrevem, respectivamente, o cálculo de *yc* e LD.



Figura 6: A) Curva analítica, intervalo de confiança e ilustração do cálculo de LD; B) Curva analítica, intervalo de confiança e ilustração do cálculo de LQ (Adaptada de Ribeiro et al., 2008).

$$y_{c} = a_{0} + s_{y}t \sqrt{\left(\frac{1}{N}\right) + 1 + \frac{\overline{x}^{2}}{\sum_{i=1}^{n} (x_{i} - \overline{x})^{2}}}$$
 Equação 2

$$LD = 2\frac{s_{y}t}{a_{1}} \sqrt{\left(\frac{1}{N}\right) + 1 + \frac{(y_{c} - \overline{y})^{2}}{a_{1}^{2}\sum_{i=1}^{n}(x_{i} - \overline{x})^{2}}}$$
 Equação 3

O LQ também pode ser calculado a partir do intervalo de confiança da curva analítica, e sua estimativa pode ser observada na Figura 6 B, em que *xc* é o valor da concentração (*x*) no ponto em que o valor de a_0 intercepta a reta de regressão, e y_h é o valor de *y* para a projeção de x_c no limite superior do intervalo. As estimativas de y_h , x_c e *LQ* podem ser realizadas respectivamente, pela utilização das Equações 4, 5 e 6 (Ribeiro *et al.*, 2008).

$$y_{h} = a_{0} + s_{y}t_{1}\sqrt{\left(\frac{1}{N}\right) + 1 + \frac{(x_{c} - \overline{x})^{2}}{\sum_{i=1}^{n} (x_{i} - \overline{x})^{2}}}$$
Equação 4
$$x_{c} = \left(\frac{s_{y}t}{a_{1}}\right)_{1}\sqrt{\left(\frac{1}{N}\right) + 1 + \frac{(\overline{x})^{2}}{\sum_{i=1}^{n} (x_{i} - \overline{x})^{2}}}$$
Equação 5

$$LQ = \left(\frac{y_h - a_0}{a_1}\right) + \left(\frac{s_y t}{a_1}\right) \sqrt{\left(\frac{1}{N}\right) + 1 + \frac{(y_h - \overline{y})^2}{a_1^2 \sum_{i=1}^n (x_i - \overline{x})^2}}$$
Equação 6

1.3.2. Notação utilizada em análise multivariada de cromatogramas

Para a utilização dos cromatogramas completos em métodos multivariados de análise, os mesmos devem ser organizados na forma de matrizes de dados, contendo as amostras nas linhas e as variáveis nas colunas (Grung *et al.*, 1995; Nielsen *et al.*, 1998; Pravdova, 2002; Tomasi *et al.*, 2004; Skov *et al.*, 2006). As variáveis referem-se às intensidades de sinal a cada tempo de retenção adquirido pelo equipamento, conforme ilustrado na Figura 7.



Figura 7: Ilustração do processo de transformação do cromatograma em uma matriz de dados para o tratamento multivariado.

Desta maneira, o cromatograma passará a ser representado por uma matriz de

dados contendo n amostras e p variáveis.

1.3.3. Pré-tratamento dos cromatogramas

O sinal obtido em cromatografia, assim como em muitas outras técnicas instrumentais, pode ser dividido nas seguintes partes: sinal analítico de interesse, deslocamentos de linha base e ruído, conforme ilustrado na Figura 8.



Figura 8: Cromatograma hipotético indicando as várias partes do sinal analítico: A) sinal bruto obtido do cromatógrafo; B) sinal analítico relevante; C) deslocamento de linha base; D) ruído (adaptada de Skov et al., 2008).

Os deslocamentos de linha base e ruído não apresentam nenhuma relação com a informação química de interesse presente nos cromatogramas e por este motivo, devem ser eliminados ou minimizados antes do tratamento multivariado dos resultados. Deslocamentos de linha base podem ser corrigidos empregando diversos métodos, incluindo a derivação. Para a eliminação do ruído, normalmente são empregados filtros digitais, como o método de alisamento de Savitzky-Golay (Beebe et *al.* 1997). Estes métodos serão discutidos com mais detalhes nos próximos itens.
Além do deslocamento da linha base e ruído, é muito comum em cromatografia que os tempos de retenção dos analitos não sejam idênticos em injeções consecutivas, ou seja, os picos podem se deslocar em relação ao eixo x. Para corrigir este efeito são utilizados vários algoritmos de alinhamento de picos, sendo que o princípio de funcionamento de alguns também será discutido a seguir.

1.3.3.1. Alisamento

Métodos de alisamento são utilizados para reduzir matematicamente o ruído com o objetivo de aumentar a relação sinal/ruído. Nestes métodos, é selecionada uma janela, a qual contém certo número de variáveis. Os pontos na janela são utilizados para determinar o valor no centro da janela e assim, o tamanho da janela influencia diretamente o resultado do alisamento. Os métodos mais conhecidos para realizar o alisamento são: alisamento por média e média móvel, mediana, por polinômio móvel (método de Savitzky-Golay).

No método de Savitzky-Golay, um polinômio de ordem baixa é ajustado aos pontos da janela e utilizado para recalcular o ponto do meio (Beebe *et al.*, 1997). A Figura 9 ilustra o processo de suavização por polinômio móvel com janela de 13 pontos. Neste exemplo, um polinômio de segunda ordem é ajustado aos pontos da janela e o ponto do meio é substituído pelo valor calculado pelo polinômio (indicado pelo "x" nesta Figura). A janela se move, repetindo este procedimento até que todos os pontos sejam suavizados. Conforme o tamanho da janela aumenta, o ruído é continuamente reduzido, entretanto, quando a janela é muito grande, além da eliminação de ruído, o

sinal analítico pode começar a ser removido ou distorcido. O tamanho da janela depende do conjunto de dados e deve ser determinado experimentalmente.



Figura 9: Ilustração do procedimento de alisamento pelo algoritmo de Savitzky-Golay.

1.3.3.2. Correção de linha de base

De uma maneira geral, existem dois tipos de deslocamentos de linha base em cromatogramas: o deslocamento vertical (*offset*), (Figura 10) e de inclinação na linha base (Figura 11).



Figura 10: Cromatogramas com deslocamentos verticais na linha de base.



Figura 11: Cromatogramas com inclinação na linha de base.

Deslocamentos na linha base podem ser corrigidos por modelagem explícita ou derivação. De forma resumida, na modelagem explícita, o deslocamento da linha base é modelado por uma função matemática e subtraído do sinal analítico total. Para isto, é necessário identificar pontos no cromatograma que contenham apenas a informação de linha base e isto pode ser especialmente difícil caso existam muitas variações na linha base. A correção de linha base por derivação não requer esta seleção e por este motivo é útil em casos nos quais a linha base é difícil ou impossível de identificar (Beebe et *al.* 1997)

Para entender como as derivadas removem deslocamentos de linha base devese considerar que o sinal obtido experimentalmente em um cromatograma pode ser descrito pela soma dos sinais de interesse mais deslocamentos de linha-base. O deslocamento de linha base pode ser aproximado utilizando um polinômio, conforme mostrado na equação 7:

$$\mathbf{r} = \mathbf{s} + \alpha + \beta x + \gamma x^2 + \delta x^3 + \dots$$
 Equação 7

onde **r** representa o cromatograma bruto, **s** é o sinal analítico de interesse e o restante da equação caracteriza o deslocamento da linha base.

Desta maneira, um cromatograma que apresente um deslocamento vertical (*offset*) pode ser escrito pela Equação 8:

$$\mathbf{r} = \mathbf{s} + \alpha$$
 Equação 8

e deslocamentos lineares de inclinação podem ser representados pela Equação 9:

$$\mathbf{r} = \mathbf{s} + \alpha + \beta \mathbf{x}$$
 Equação 9

Tomando-se a primeira derivada da Equação 7, dada pela Equação 10,

$$\mathbf{r} = \mathbf{s}' + 0 + \beta + 2\gamma x + 3\delta x^2 + ...$$
 Equação 10

verifica-se que foram completamente removidos os deslocamentos verticais constantes (α). Vale lembrar que, se o deslocamento for apenas deste tipo, os demais coeficientes da Equação 7 serão iguais a zero. Se um deslocamento mais complexo for observado (β , γ , δ ... \neq 0), a aplicação repetida de derivadas irá sucessivamente remover os termos de ordem superior. Por exemplo, tomando-se a derivada da Equação 10, o que corresponde a segunda derivada da Equação 7, tem-se a Equação 11:

$$\mathbf{r} = \mathbf{s}'' + 0 + 0 + 2\gamma + 6\delta x + .$$
 Equação 11

Através da observação da Equação 11, pode-se verificar que a aplicação da segunda derivada removeu deslocamentos constantes (α) e deslocamentos lineares (β).

As derivadas podem ser calculadas de diversas maneiras, tais como a simples diferença móvel (*Running Simple Difference*), média móvel (*Running Mean Difference*) e métodos baseados no alisamento de Savitzky-Golay. Neste último caso, seleciona-se

uma janela de determinado tamanho, ajusta-se um polinômio aos pontos da janela e o ponto do meio é substituído pela *derivada* do polinômio. As derivadas têm a propriedade de amplificar sinais pouco intensos, porém ruído instrumental também é amplificado. Por este motivo, é comum a suavização dos cromatogramas ou espectros juntamente com a aplicação das derivadas. No método de Savitzky-Golay, o alisamento é realizado junto com a derivação, bastando aumentar o tamanho da janela para aumentar o alisamento. A Figura 12 exemplifica o processo de derivação e alisamento de um cromatograma hipotético. Na Figura 12 A, é mostrado o cromatograma original, destacando-se picos pouco intensos. O processo de derivação do cromatograma original empregando uma janela pequena é mostrado na Figura 12 B, na qual é possível verificar que os deslocamentos de linha base foram removidos, porém o cromatograma permanece com ruído. Neste caso, é necessário aumentar a janela, entretanto, deve-se tomar cuidado para não eliminar os sinais pouco intensos. A Figura 12 C mostra um cromatograma alisado com um tamanho de janela adequado, o qual permitiu eliminar ruído sem eliminar os sinais pouco intensos.



Figura 12: Ilustração do processo de derivação e alisamento de um cromatograma: A) cromatograma original, B) cromatograma na forma de primeira derivada com uma janela muito pequena e C) cromatograma na forma de derivada com uma janela ideal para o conjunto de dados.

1.3.3.3. Alinhamento

Conforme mencionado anteriormente, os picos cromatográficos podem se deslocar em relação ao tempo de retenção (Figura 13) após injeções consecutivas. Este fenômeno é indesejado e acontece principalmente devido às flutuações instrumentais, tais como variação de pressão, temperatura, composição de fase móvel e estacionária, variações de fluxo etc (Grung *et al.*, 1995; Nielsen *et al.*, 1998; Pravdova, 2002; Tomasi *et al.*, 2004; Skov *et al.*, 2006)..



Figura 13: Cromatogramas hipotéticos que apresentam deslocamento no tempo de retenção de um mesmo composto em diversas injeções.

Em matrizes de dados cromatográficos, é importante que a resposta relativa a cada uma das variáveis esteja exatamente na coluna correspondente para todas as amostras para que a análise dos dados seja precisa. Este é um pré-requisito fundamental para a aplicação, tanto da estatística univariada, quanto de análise

estatística multivariada . A solução para estes problemas de deslocamentos de tempo de retenção não é trivial e, certamente soluções clássicas como considerar tempo de retenção relativo ou normalização dos cromatogramas, não são eficientes quando a matriz de dados é muito complexa (Skov *et al.*, 2006).

A essência de todo o método de alinhamento envolve uma transformação no eixo do tempo do cromatograma da amostra de forma a sobrepor o maior número possível de picos com um cromatograma selecionado como referência. Diversos métodos de alinhamento de cromatogramas estão disponíveis na literatura, dentre eles o COW (*Correlation Optimazed Warping*), Optim_COW e Peakmatch (Nielsen *et al.*, 1998; Tomasi *et al.*,2004).

O COW alinha os cromatogramas de interesse em relação a outro cromatograma escolhido como referência, preservando as informações originais de área e formato dos picos. Isto é feito "cortando" o cromatograma em diversos segmentos (pedaços), que sofrem deformações lineares (sendo esticados ou contraídos), chamadas de *warping*, em combinação com interpolações, otimizando os coeficientes de correlação entre os seguimentos correspondentes aos cromatogramas de referência e de amostra (Nielsen *et al.*, 1998). A magnitude do *warping*, ou seja, o quanto cada segmento pode ser esticado ou contraído, é chamada de *slack*. Um único cromatograma de referência é utilizado para alinhar toda a matriz de dados. Um fator crítico sobre a utilização do COW é justamente a escolha do cromatograma empregado como referência, uma vez que está baseada em critérios subjetivos. O cromatograma a ser escolhido como referência deve representar da melhor forma possível, toda a matriz de dados. No entanto, neste algoritmo, não existem parâmetros quantitativos para efetuar esta escolha. Esta

característica pode ser entendida como uma desvantagem do algoritmo, uma vez que a qualidade do alinhamento é muito depende da qualidade da referência.

O Optim_COW é uma otimização do COW feita justamente com o objetivo de se obter parâmetros quantitativos para a escolha do cromatograma de referência (Skov *et al.*, 2006). Estes parâmetros são: índice de similaridade, simplicidade, fator de pico e efeito de *warping*. O índice de similaridade está baseado no produto dos coeficientes de correlação entre todos os cromatogramas individuais. Para um dado cromatograma (\mathbf{x}_t), o índice de similaridade, $0 < índice de similaridade \le 1$, pode ser calculado pela Equação 12:

Indice de Similaridade =
$$\prod_{i=1}^{I} |r(x_t, x_i)|$$
 Equação 12

onde $r(x_t, x_i)$ é o coeficiente de correlação entre dois cromatogramas no conjunto de dados.

O cromatograma que é mais similiar aos outros possuirá o maior índice de similaridade (valor mais próximo ou igual a 1), e será selecionado como referência no alinhamento.

O parâmetro Simplicidade ($0 < simplicidade \le 1$), mostrado na Equação 13, está relacionada com as propriedades do algoritmo de decomposição em valores singulares (*Singular Value Decomposítiosn*, SVD), onde o tamanho dos valores singulares ao quadrado está diretamente relacionado com a soma dos quadrados da matriz de dados. Geralmente, cromatogramas bem alinhados irão resultar em um menor número de valores singulares significativos com valores maiores, os quais devem representar as informações químicas de interesse.

Simplicidade =
$$\sum_{r=1}^{R} \left(SVD\left(X / \sqrt{\sum_{i=1}^{I} \sum_{j=1}^{J} x(i, j)^2} \right) \right)^4$$
 Equação 83

O valor de simplicidade será muito próximo de 1 quando o alinhamento apresentar boa qualidade.

O fator de pico (Equações 14, 15 e 16) indica o quanto o conjunto total cromatogramas foi modificado após o pré-processamento:

Fator de pico =
$$\frac{\sum_{i=1}^{I} (1 - minimum(c(i), 1)^2)}{I}$$
 Equação 14

$$c(i) = \left| \frac{\|x_w(i)\| - \|x(i)\|}{\|x(i)\|} \right|$$
 Equação 15

$$\|x(i)\| = \sqrt{\sum_{j=1}^{J} x(i,j)^2}$$
 Equação 16

A Equação 16 é a distância Euclidiana ou norma para $\mathbf{x}(i)$; $\mathbf{x}(i)$ é o cromatograma antes do alinhamento enquanto $\mathbf{x}_w(i)$ é o mesmo depois do alinhamento.

O effeito de *warping* (Equação 17), ($0 < efeito de warping \le 2$) combina dois parâmetros quantitativos: simplicidade e fator de pico:

No algotimo Peakmatch, o cromatograma da amostra e o cromatograma de referência (alvo) são comparados pico a pico. Uma distancia fixa (janela) é estipulada para deslocamento no tempo de retenção da amostra, se dentro deste intervalo um pico for identificado como similar ao pico do alvo, então o pico é alinhado (Grung *et al.*, 1995; Pravdova *et al.*, 2001).

A maneira utilizada para determinar a qualidade do alinhamento de cromatogramas parecidos, nesse caso, é o coeficiente de correlação (r) entre eles. O coeficiente de correlação entre dois cromatogras (Equação 18), X e Y com N pontos cada, é calculado como:

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{N}\right)\left(\sum Y^2 - \frac{(\sum Y)^2}{N}\right)}}$$
Equação 18

Dois cromatogramas são idênticos se o coeficiente de correlação for igual a 1.

1.4. Pré-processamento das variáveis

Além dos métodos de pré-tratamento que podem ser aplicados às amostras, existem métodos de pré-processamento aplicados às variáveis que tornam as matrizes melhor condicionadas para o tratamento de dados. Estes métodos são aplicados a cada variável individualmente e simultaneamente sobre todas as amostras.

1.4.1. Centrar os dados na média

Para centrar na média, calcula-se o valor médio para cada variável (coluna) e subtrai-se este valor para cada elemento da variável. Desta maneira, cada variável passará a ter média zero, ou seja, as coordenadas são movidas para o centro dos dados, permitindo que diferenças nas intensidades relativas das variáveis sejam mais fáceis de perceber. Este processo é ilustrado na Figura 14.



Figura 14: Ilustração do processo de centrar um conjunto de dados na média.

Normalmente, este pré-processamento é aplicado à variáveis contínuas.

1.4.2. Escalamento da variância

No processo de escalamento da variância cada elemento de uma coluna é dividido pelo desvio padrão da coluna, fazendo com que as variáveis passem a ser expressas em unidades de desvio-padrão e que as influências relativas das diferentes variáveis nos cálculos tornem-se independentes das suas unidades.

1.4.3. Autoescalamento

O autoescalamento consiste na junção das operações de centrar as variáveis na média com o procedimento de escalamento da variância, de acordo com a equação 19:

$$x_{ij}(autoescalado) = \frac{x_{ij} - \overline{x}_j}{s_j}$$
 Equação 19

Normalmente este pré-processamento é aplicado à variáveis discretas.

1.5. Métodos quimiométricos de análise exploratória

O tratamento de resultados contendo várias amostras e variáveis pode ser extremamente trabalhoso e ainda não permitir a extração completa de informações do sistema. Para resolver este problema, existem métodos multivariados de análise, os quais permitem a visualização do conjunto de dados como um todo, permitindo um entendimento completo de sistemas complexos, extraindo informação útil é através da avaliação de padrões de comportamento no conjunto de dados. Métodos de Análise Exploratória ou Reconhecimento de Padrões são métodos de classificação baseados nas diferenças das propriedades de interesse das amostras (Brereton, 2009). De uma maneira geral, eles podem ser classificados como métodos supervisionados, como por exemplo, *K-Nearest Neighbor* (KNN) e *Self-Independent Modeling of Class Analogy* (SIMCA), ou não-supervisionados, como por exemplo, *Principal Component Analysis* (PCA) e *Hierarchical Cluster Analysis* (HCA) e permitem a interpretação multivariada de conjuntos de dados complexos por meio de gráficos bi ou tridimensionais. Nos métodos

supervisionados é necessário que exista alguma informação inicial sobre a identidade das amostras para a formação das classes e o objetivo é desenvolver um modelo baseado nas informações contidas nas amostras. Por outro lado, nos métodos não supervisionados, a separação de classes acontece sem a necessidade de informações iniciais sobre a natureza das amostras e o objetivo é encontrar agrupamentos naturais entre amostras semelhantes. Neste trabalho, serão discutidos os métodos PCA e HCA

1.5.1. Análise por Componentes Principais

A base fundamental da maioria dos métodos modernos para tratamento de dados multivariados é a PCA, um método que permite a redução da dimensionalidade através da representação do conjunto de dados em um novo sistema de eixos, denominados Componentes Principais (PC), permitindo a visualização da natureza multivariada dos dados em poucas dimensões. No espaço original, as amostras são pontos localizados em um espaço n- dimensional. Com a redução de dimensionalidade proporcionada pela PCA, as amostras passam a ser localizadas em espaços reduzidos por exemplo, bi ou tri dimensionais (Brereton, 2009).

Matematicamente, na Análise de Componentes Principais, a matriz X é decomposta em um produto de duas matrizes, denominadas *scores* (**T**) e *loadings* (**P**), mais uma matriz de erros (**E**), como mostrado na Equação 20:

$$\mathbf{X} = \mathbf{TP}^{\dagger} + \mathbf{E}$$
 Equação 20

Os *scores* representam as coordenadas das amostras no sistema de eixos formados pelos Componentes Principais. Cada PC é constituído pela combinação linear das variáveis originais e os coeficientes (pesos) da combinação são denominados *loadings*. Matematicamente, os *loadings* são os cossenos dos ângulos entre as variáveis originais e os Componentes Principais, representando, portanto, o quanto cada variável original contribui para uma determinada PC (Wold, 1987; Brereton, 2009).

A Primeira Componente Principal (PC1) é traçado no sentido da maior variação no conjunto de dados; a segunda (PC2) é traçado *perpendicularmente* a primeira, com o intuito de descrever a maior porcentagem da variação *não explicada* pela PC1 e assim por diante. Este procedimento é ilustrado na Figura 15. Enquanto os *scores* representam as relações de similaridade entre as amostras, os *loadings* indicam a contribuição de cada variável para a formação das PC. Através da análise conjunta do gráfico de *scores* e *loadings*, é possível verificar quais variáveis são responsáveis pelas diferenças observadas entre as amostras.



Figura 15: Ilustração do direcionamento dos primeiros dois Componentes Principais no método PCA.

O número de PC a ser utilizado no modelo PCA é determinado pela porcentagem de variância explicada por cada PC. Assim, seleciona-se um número de PC de tal maneira que a maior porcentagem da variação presente no conjunto de dados originais seja capturada.

1.5.2. Análise de Agrupamentos Hierárquicos

A HCA busca agrupar as amostras em classes, baseando-se na similaridade dos participantes de uma mesma classe e nas diferenças entre os membros de classes diferentes (Corrêa e Ferreira, 2007). A representação gráfica obtida é chamada de dendrograma, um gráfico bidimensional independente do número de variáveis do conjunto de dados. Para gerar o dendograma, o método de HCA forma agrupamentos de amostras baseados na sua proximidade no espaço n-dimensional, sendo n o número

de variáveis. A ligação entre as amostras é realizada através do valor da distância e o método de ligação pode ser simples ou centróide. Na ligação simples, cada amostra é conectada ao seu vizinho mais próximo. O método de ligação centróide é uma variação da ligação simples, no qual os centros dos agrupamentos são ligado, ao invés do vizinho mais próximo. Os dois métodos são ilustrados na Figura 16.



Figura 16: Métodos de ligação de amostras que podem ser empregados na HCA: A) ligação simples e B) ligação centróide.

A distância entre as amostras pode ser medida de várias maneiras, sendo as principais a distância Euclidiana (Equação 21), distância de Mahalanobis (Equação 22) e distância de Manhattan (Equação 23).

Distância Euclidiana:

$$D_{ij} = \left[\sum_{k=1}^{p} \left(X_{ik} - X_{jk} \right)^{2} \right]^{1/2}$$
 Equação 21

Distância de Mahalanobis:

$$D_{ij} = (X_i - X_k)^T \sum_{j=1}^{-1} (X_i - X_k)$$
 Equação 22

Distância de Manhattan:

$$D_{ij} = \sum_{k=1}^{p} \left| X_{ik} - X_{jk} \right|$$
Equação 23

onde os sub-índices *i* e *j* indicam as amostras e *k* as variáveis.

1.6. Objetivo do trabalho

Este trabalho teve como objetivo principal a utilização de métodos de análise exploratória multivariada para comparação de perfis cromatográficos de compostos carbonílicos na atmosfera em três sítios na cidade de SP impactados por emissão veicular.

CAPÍTULO 2 – PROCEDIMENTO EXPERIMENTAL

2.1. Padrões e solventes

Os padrões e solventes utilizados estão listados na Tabela 1.

Tabela 1: Padrões e solventes utilizados.

Substâncias	Pureza	Fornecedor
2,4-Dinitrofenilhidrazonas (<i>Aldehyde/Ketone-DNPH</i>) mistura padrão	99%	Supelco®
Acetonitrila (MeCN)	99,9% min. (grau HPLC)	J.T. Baker
Tetrahidrofurano (THF)	99,9% min. (grau HPLC)	J.T. Baker
H ₂ 0	Purificada	Ultrapure water system
Detergente Extran® neutro		Merck

2.2. Limpeza do Material

Todos os materiais e vidrarias usados foram limpos com detergente, enxaguados com água destilada, deionizada e, posteriormente, secos em estufa quando necessário. Os recipientes em contato direto com o reagente DNPH, como por exemplo, os frascos de estocagem das amostras, assim como as suas tampas e septos, foram limpos no ultra-som e finalmente enxaguados com acetonitrila.

2.3. Instrumentação

As amostras foram analisadas em Cromatógrafo Líquido de Alto Desempenho, (Shimadzu, modelo LC-9-A) composto por:

Duas bombas Shimadzu modelo LC-10AD, para programação de gradientes de concentração da fase móvel;

 Válvula de injeção Rheodyne com seis vias e alça de amostragem (*loop*) de 20 μL;

• Detector UV – Vis modelo SPD – 10AV;

• Sistema de controle modelo SCL – 10A;

 Programa Shimadzu Class–VP, para microprocessamento do equipamento e aquisição dos dados.

2.4. Método de análise de compostos carbonílicos no ar

Neste trabalho, foi utilizado o método TO-11A da *U.S. Environmental Protection Agency* (EPA). Este método consiste em três etapas:

I. Coleta de ar ambiente empregando-se cartuchos contendo partículas de sílica impregnada com o derivatizante 2,4 dinitrofenilhidrazina (DNPH);

II. Extração dos compostos carbonílicos retidos no cartucho com acetonitrila;

III. Análise do extrato orgânico por Cromatografia Liquida de Alta Eficiência(CLAE) com detecção por UV-Vis.

A CLAE permite separar, identificar e quantificar os carbonílicos desde os mais leves até os mais pesados. A reação de derivatização com DNPH se dá por adição nucleofílica do composto de nitrogênio básico ao carbono da carbonila, formando a 2,4dinitrofenilhidrazona correspondente, segundo a reação da Figura 17.



Figura 17: Reação de adição nucleofílica do CC com o reagente derivatizante DNPH.

Dois tipos de cartuchos foram utilizados durante o procedimento de coleta do ar ambiente:

I. Sep-Pack[®] - DNPH (*Waters*) para retenção dos CC: cartucho em polietileno preenchido com sílica gel impregnada com reagente derivatizante DNPH.
Eficiência de coleta > 95% para vazão de coleta de até 2 L/min e

II. Sep-Pack[®] *ozone scrubber* (*Waters*) para remoção de O₃ cartucho em polietileno, preenchido com iodeto de potássio sólido (KI). Capacidade teórica informada pelo fabricante igual a 4,2 mmol de O₃ (200 mg).

Após a coleta, as amostras foram transportadas em embalagens térmicas contendo gelo. A eluição dos cartuchos após as coletas das amostras foi efetuada no menor intervalo de tempo possível, que variou entre poucas horas até 12 h. Os

compostos carbonílicos retidos nos cartuchos foram eluídos pela passagem de 3 mL de acetonitrila, utilizando-se uma seringa de vidro, em balão volumétrico de 5 mL.

Os níveis de *background*, dos cartuchos adquiridos comercialmente, foram determinados pela "análise de brancos dos cartuchos".

A análise por CLAE foi realizada empregando a seguinte fase móvel: (A) mistura de água e MeCN nas proporções respectivas de 60% e 40% (B) - mistura de MeCN e THF nas proporções respectivas de 80% e 20%. As fases foram filtradas através da membrana Fluoropore[™] (Millipore) de politetrafluoroetileno (PTFE) laminado com polietileno, com dimensões de poro e diâmetro iguais a 0,5µm e 47 mm respectivamente. A concentração dos solventes A e B seguiram um gradiente de concentração com 30% de B por 5 min, rampa de 30% a 50% em 12 min, patamar em 50% de B por 8 min e rampa de 50 % até 30 % de B em 2 min. Os demais parâmetros de análise foram: vazão da fase móvel: 1,3 mL/min; tempo de análise 27 min; temperatura de análise: ambiente (25 °C em média); comprimento de onda fixo: 365 nm (lâmpada de deutério); pré-coluna Phenomenex[®] Gemini - preenchida com sílica quimicamente modificada com o grupo octadecila para separação em fase reversa (C18), 4 mm de comprimento e 3 mm de diâmetro interno; coluna Phenomenex[®], Gemini - preenchida com sílica quimicamente modificada com o grupo octadecila para separação em fase reversa (C18), com 4,6 mm de diâmetro interno, 250 mm de comprimento e 5 μ m de diâmetro da partícula.

As curvas de calibração foram construídas da seguinte maneira:

Preparo de solução estoque de concentração (3 μg/ml) a partir de solução de mistura-padrão comercial de hidrazonas (15 μg/ml), em acetonitrila. Preparo de soluções de trabalho nas concentrações 25 μg/ml, 50 μg/ml, 100 μg/ml, 150 μg/ml, 200 μg/ml, 250 μg/ml, 600 μg/ml, 1000 μg/ml, 3000 μg/ml a partir da solução estoque, em acetonitrila.

2.5. Caracterização dos sítios de amostragem

2.5.1. Sítio IAG

O sítio IAG está localizado no campus da Cidade Universitária, zona Oeste da cidade de São Paulo. O sítio IAG está localizado no campus da Cidade Universitária, zona Oeste da cidade de São Paulo, localizado a cerca de 6.338,29 m do cruzamento entre a Avenida Dr. Arnaldo e a Rua Teodoro Sampaio, 6.105,88 m do cruzamento entre a avenida Dr. Arnaldo e a Avenida Cardeal Arcoverde, 4.171,49 m do cruzamento entre a Avenida Faria Lima e a Rua Teodoro Sampaio, 5.350,86 m do cruzamento entre a Rua Teodoro Sampaio e a Rua Henrique Schaumann, 2.306,28 m do cruzamento entre a Avenida Alvarenga e a entrada do Portão 1 da Cidade Universitária, 1,121.86 m do cruzamento entre a Avenida Corifeu de Azevedo Marques (Portão 3). A campanha foi realizada no período entre 25 e 28 de maio de 2009, com amostragem de duas em duas horas a partir das 05 h até as 16 h. As informações de coleta das amostras neste sítio são mostradas na Tabela 2.

Tabela 2: Informações sobre a amostragem no sítio IAG.

Data	Período	Amostra
25/5/2009	05:00 - 07:00	IAG1
25/5/2009	8:00 - 10:00	IAG2
25/5/2009	10:00 - 12:00	IAG3
25/5/2009	12:00 - 14:00	IAG4
25/5/2009	14:00 - 16:00	IAG5
26/5/2009	05:00 - 07:00	IAG6
26/5/2009	8:00 - 10:00	IAG7
26/5/2009	12:00 - 14:00	IAG9
26/5/2009	14:00 - 16:00	IAG10
28/5/2009	10:00 - 12:00	IAG18

2.5.2. Sítio ICB

O sítio ICB está localizado no campus da Cidade Universitária, zona Oeste da cidade de São Paulo. As amostragens foram realizadas no ICB-USP sobre um trailer localizado na Avenida Lineu Prestes, no pátio do ICB II. Este ponto se situa a cerca de: 2.290 m do cruzamento entre a Avenida Alvarenga e a entrada do Portão 1 da Cidade Universitária, 2.048 m do cruzamento entre a Avenida Politécnica e a Marginal Pinheiros (próximo ao portão 2), 760 m da Avenida Corifeu de Azevedo Marques (Portão 3). A campanha foi realizada nos dias 18, 19, 20, 21 e 22 de agosto de 2008, com amostragem de duas em duas horas a partir das 08 h até as 16 h. As informações de coleta das amostras no ICB são descritas na Tabela 3.

Tabela 3: Informações sobre a amostragem no sítio ICB.

Data	Período	Amostra
18/8/2008	08:11 - 10:33	ICB 01
18/8/2008	14:55 - 16:55	ICB 02
19/8/2008	08:10 - 10:10	ICB 03
19/8/2008	10:12 - 12:17	ICB 04
19/8/2008	12:17- 14:17	ICB 05
19/8/2008	12:22 - 14:22	ICB 06
19/8/2008	14:25 - 16:00	ICB 07
20/8/2008	08:51 - 10:51	ICB 08
20/8/2008	10:01 - 12:01	ICB 09
20/8/2008	12:04 - 14:05	ICB 10
20/8/2008	14:10 - 16:11	ICB 11
21/8/2008	08:05 - 10:05	ICB 12
21/8/2008	10:06- 12:13	ICB 13
21/8/2008	12:11- 14:06	ICB 14
21/8/2008	14:09- 16:10	ICB 15
22/8/2008	10:01- 12:01	ICB 16
22/8/2008	11:56 - 14:03	ICB 17

2.5.3. Sítio FSP

O sítio FSP está localizado na região de Cerqueira César, zona oeste da cidade de São Paulo. As amostragens foram realizadas em um jardim em frente à entrada Principal da Faculdade de Saúde Pública localizado a cerca de 60 m da Avenida Dr. Arnaldo, 51 m da Rua Teodoro Sampaio, 191 m da Avenida Cardeal Arcoverde. A campanha foi realizada no período entre 25 e 28 de maio de 2009, com amostragem de duas em duas horas a partir das 05 h até as 16 h. As informações de coleta das amostras na FSP são descritas na Tabela 4.

Data	Período	Amostra
25/5/2009	05:00 - 07:00	FSP 1
25/5/2009	10:00 - 12:00	FSP 3
25/5/2009	12:00 - 14:00	FSP 4
25/5/2009	14:00 - 16:00	FSP 5
26/5/2009	05:00 - 07:00	FSP 6
26/5/2009	10:00 - 12:00	FSP 8
26/5/2009	12:00 - 14:00	FSP 9
26/5/2009	14:00 - 16:00	FSP 10
28/5/2009	08:00 - 10:00	FSP 17
28/5/2009	10:00 - 12:00	FSP 18
28/5/2009	12:00 - 14:00	FSP 19

Tabela 4: Informações sobre a amostragem no sítio FSP.

CAPÍTULO 3 – RESULTADOS E DISCUSSÕES

3.1. Tratamento univariado dos resultados

3.1.1. Curvas de calibração e avaliação do modelo linear

Os cromatogramas dos padrões originais são mostrados na Figura 18. Como se pode observar, o deslocamento em relação ao eixo x dificulta a visualização dos picos e identificação dos compostos individualmente. Por motivos de clareza, os padrões alinhados são mostrados na Figura 19, juntamente com a identificação de cada um dos picos.



Figura 18: Cromatograma dos padrões originais sem nenhum processamento.



Figura 19: Cromatograma dos padrões originais após alinhamento com identificação dos picos.

Empregando-se os resultados de área observados para os padrões, foram construídas curvas de calibração para todos os compostos carbonílicos estudados através de um modelo linear do tipo Área = a + b*concentração. Os modelos foram avaliados empregando-se os seguintes critérios:

I. Coeficiente de determinação (R²) = Soma Quadrática da Regressão/Soma
Quadrática Total e coeficiente de correlação (r);

II. Gráficos de diagnóstico: Gráfico de Probabilidade Normal dos resíduos, gráfico de resíduos *vs* amostra e gráfico de valores experimentais *vs* valores calculados pelo modelo linear.

Todos os modelos apresentaram R^2 e r > 0,99, o que pode ser observado pela linearidade observada entre o incremento de área com o incremento de concentração.

A mesma curva analítica foi utilizada para quantificação das amostras coletadas na FSP e IAG. Para as amostras coletadas no ICB, foi utilizada outra curva, pois o período de amostragem foi diferente. Para ilustrar o processo realizado na avaliação do modelo linear empregando os critérios mencionados acima, serão discutidos os resultados para o formaldeído. Para os demais analitos, será mostrada apenas a curva analítica e os valores de R², LQ e LD. A curva analítica para o formaldeído é mostrada na Figura 20, juntamente com os valores de R², LD e LQ, calculados a partir dos parâmetros da curva, conforme descrito por Ribeiro *et al*.



Figura 20: Curva de calibração para o formaldeído para quantificação das amostras das campanhas na FSP e IAG.

O gráfico de probabilidade Normal dos resíduos (Figura 21), resíduos vs amostra (Figura 22) e valores previstos vs valores experimentais (Figura 23), são mostrados a seguir. A Figura 21 indica que os resíduos seguem uma distribuição Normal, apresentando apenas variações aleatórias, devido ao erro experimental, as quais não podem ser reproduzidas pelo modelo, conforme esperado para um modelo bem ajustado. O gráfico de resíduos *vs* amostra (Figura 22) indica que os resíduos são homogeneamente e aleatoriamente distribuídos ao longo das amostras. O gráfico de valores previstos *vs* valores experimentais (Figura 23) apresenta excelente concordância, indicando que o modelo linear é capaz de fornecer estimativas bastante próximas daquelas obtidas experimentalmente.



Figura 21: Gráfico de probabilidade Normal dos resíduos deixados pelo modelo linear para o formaldeído.



Figura 22: Gráfico de resíduos vs amostra para o formaldeído.



Figura 23: Gráfico de valores previstos pelo modelo linear vs valores experimentais.

9.00E+05 y = 1289x + 8266,2 R² = 0,9993 7,00E+05 y = 1015,8x + 4491,3 8.00E+05 6,00E+05 LD = 19,05 ng/m LQ = 28,22 ng/m R² = 0,9995 LD = 16,58 ng/mL 7.00E+05 5.00F+05 6.00E+05 LQ = 24,54 ng/ml 4,00E+05 5.00E+05 4,00E+05 Área 3.00E+05 3,00E+05 2,00E+05 2,00E+05 1,00E+05 1,00E+05 0,00E+00 0,00E+00 0.0 100.0 200,0 300,0 400,0 500,0 600,0 700,0 400,0 0,0 100,0 200,0 300,0 500,0 600,0 700,0 Concentração (ng/mL) Concentração (ng/mL) Acetona Acroleína y = 530,5x - 1613,2 R² = 0,9974 LD = 21,22 ng/mL LQ =31,42ng/mL 1,60E+05 1,20E+05 y = 651,68x - 278,25 R² = 0,999 LD = 8,76 ng/mL 1,40E+05 1,00E+05 1,20E+05 8,00E+04 LQ = 12,88 ng/mL 1,00E+05 8,00E+04 Área 6,00E+04 érea 6,00E+04 4,00E+04 4,000+04 2.00E+04 2,00E+04 0.00E+00 0,00E+00 0.0 20.0 40.0 60.0 80.0 100.0 120.0 140.0 160.0 0,0 50.0 100.0 150.0 200.0 250.0 300.0 Concentração (ng/mL) ntração (ng/mL) Propinaldeido Crotonaldeído 1,20E+05 1,40E+05 y = 759,56x + 998,42 v = 676.71x + 964.021,20E+05 1,00E+05 R² = 0,9969 R² = 0,9973 LD = 12,44 ng/mL LQ = 18,42 ng/mL 1,00E+05 LD = 13,36 ng/mL 8,00E+04 LQ = 19.80 ng/mL 8,00E+04 Área Área 6,00E+04 6,00E+04 4,00E+04 4,00E+04 2,00E+04 2,00E+04 0,00E+00 0,00E+00 0.0 20.0 40.0 60.0 80,0 100,0 120,0 140,0 160,0 40,0 60,0 80,0 100,0 120,0 140,0 160,0 0,0 20,0 Concentração (ng/mL) Concentração (ng/mL) Benzaldeído Butiraldeído 1,00E+05 1.40E+05 y = 520,23x + 852,81 R² = 0,9983 LD = 14,69 ng/mL LQ = 21,79 ng/mL y = 590,67x + 646,36 R² = 0,9978 LD = 11,16 ng/mL LQ = 16,56 ng/mL 9,00E+04 1,20E+05 8.00E+04 7,00E+04 1.00E+05 6.00E+04 8,00E+04 Área 5,00E+04 4,00E+04 Área 6,00E+04 3,00E+04 2,00E+04 4,00E+04 2,00E+04 1,00E+04 0,00E+00 0,00E+00 0,0 20,0 40,0 60,0 80,0 100,0 120,0 140,0 160,0 0.0 50.0 100.0 150.0 200.0 250.0 300.0

As curvas analíticas para os demais compostos são mostradas nas Figuras 24 e



Concentração (ng/mL)

Formaldeído

25, respectivamente.

Figura 24: Curvas analíticas para a quantificação dos compostos carbonílicos formaldeído, acetaldeído, acetona, acroleína, propionaldeído, crotonaldeído, butiraldeído, benzaldeído utilizadas para quantificação das amostras coletadas na FSP e IAG.

Acetaldeído

Concentração (ng/mL)



Figura 25: Curvas analíticas para a quantificação dos compostos carbonílicos isovaleraldeído, valeraldeído, o-tolualdeído, m,p-tolualdeido, hexaldeido e 2,5-dimetilbenzaldeído utilizadas para quantificação das amostras coletadas na FSP e IAG.
As curvas analíticas preparadas para a quantificação das amostras do ICB são mostradas nas Figuras 26 e 27, respectivamente.



Figura 26: Curvas analíticas para a quantificação dos compostos carbonílicos formaldeído, acetaldeído, acetona, acroleína, propionaldeído, crotonaldeído, butiraldeído e benzaldeído utilizadas para quantificação das amostras coletadas no ICB.



Figura 27: Curvas analíticas para a quantificação dos compostos carbonílicos isovaleraldeído, valeraldeído, o-tolualdeído, m,p-tolualdeido, hexaldeido e 2,5-dimetilbenzaldeído utilizadas para quantificação das amostras coletadas no ICB.

74

3.1.2. Cromatogramas

Os níveis de *background*, dos cartuchos adquiridos comercialmente não apresentaram valores quantificáveis de nenhum dos CC estudados.

O conjunto dos cromatogramas originais das amostras coletadas na FSP, IAG e ICB são mostrados nas Figuras 28, 29 e 30, respectivamente.



Figura 28: Cromatogramas das amostras originais provenientes da campanha na FSP: A) Cromatograma inteiro, B) pico do formaldeído aumentado.



Figura 29: Cromatogramas das amostras originais provenientes da campanha IAG: A) Cromatograma inteiro, B) pico do formaldeído aumentado.



Figura 30: Cromatogramas das amostras originais provenientes da campanha no ICB: A) Cromatograma inteiro e B) pico do formaldeído aumentado.

Os cromatogramas das amostras dos três sítios são mostrados em uma única matriz na Figura 31.



Figura 31: Cromatogramas das amostras originais provenientes das três campanhas: A) Cromatograma inteiro; B) pico do formaldeído aumentado.

3.1.3. Quantificação em ng/mL

3.1.3.1. Quantificação - FSP

As áreas dos cromatogramas das amostras coletadas na FSP foram utilizadas nas curvas de calibração mostradas nas Figuras 23-24, fornecendo as concentrações descritas na Tabela 5 e na Figura 32. Muitos analitos não foram observados por estarem abaixo do LD (< LD) do método e outros foram observados em concentração abaixo do LQ (< LQ).

Tabela 5: Concentração (em ng/mL) dos compostos carbonílicos encontrados nas amostras coletadas na FSP.

Concentração (ng/mL)	FSP 1	FSP 3	FSP 4	FSP 5	FSP 6	FSP 8	FSP 9	FSP 10	FSP 17	FSP 18	FSP 19
Formaldeído	115,49	98,80	104,74	98,28	31,75	88,27	90,54	118,68	207,94	111,79	69,19
Acetaldeído	281,61	167,38	135,94	141,86	78,37	123,12	98,28	194,95	356,34	191,74	124,48
Acetona	308,77	246,64	121,27	143,01	114,91	123,08	113,44	215,64	280,20	192,11	116,92
Acroleína	< LQ										
Propinaldeído	23,60	< LQ	< LQ	< LQ	< LD	< LQ	36,23	< LQ	36,23	< LQ	< LQ
Crotonaldeído	< LQ	< LQ	< LD	< LD	< LD	21,35	< LD	< LD	22,57	21,78	< LD
Butiraldeído	20,59	22,98	19,07	< LQ	< LD	17,54	14,81	16,55	30,92	18,74	< LQ
Benzaldeído	45,32	140,47	79,34	81,38	57,16	77,21	71,15	128,89	177,84	82,13	63,97
Isovaleraldeído	< LD	< LQ	< LD								
Valeraldeído	< LD	< LQ	< LD								
o-tolualdeído	< LD										
m,p-tolualdeído	< LQ	< LD	< LQ	< LD	< LD						
hexaldeído	< LD	19,32	< LD								
2,5-dimetilbenzaldeído	< LD										

< LD representa valores não detectáveis por estarem abaixo do LD do método e < LQ representa valores não quantificáveis, por estarem abaixo do LQ do método.</p>



Figura 32: Concentração dos principais analitos encontrados nas amostras coletadas na FSP.

Os CC encontrados neste sítio foram formaldeído, acetaldeído, acetona, butiraldeído e benzaldeído. De uma maneira geral, existe uma similaridade muito grande entre as concentrações dos CC, ou seja, o perfil das curvas é praticamente o mesmo para todos os analitos. Além disto, é possível observar a maior concentração de acetaldeído em relação ao formaldeído.

3.1.3.2. Quantificação - IAG

Os valores de concentração encontrados nas amostras coletadas no IAG estão mostrados na Tabela 6. Similarmente às amostras da FSP, nas amostras do IAG muitos analitos não foram observados e muitos foram observados em concentração não quantificável. Os principais analitos encontrados nestas amostras são mostrados de forma gráfica na Figura 33.

Tabela 6: Concentração (em ng/mL) dos compostos carbonílicos encontrados nas amostras coletadas no IAG.

Concentração (ng/mL)	IAG 1	IAG 2	IAG 3	IAG 4	IAG 5	IAG 6	IAG 7	IAG 9	IAG 10	IAG 18
Formaldeído	33,70	159,82	90,02	98,96	83,50	56,69	118,60	109,09	98,08	553,66
Acetaldeído	118,88	420,93	157,18	129,29	194,15	86,13	159,45	123,70	226,62	616,84
Acetona	218,06	295,90	229,59	366,66	340,08	156,34	196,93	220,85	216,01	468,41
Acroleína	< LD	< LD	< LQ	< LQ	< LD	< LQ				
Propinaldeído	< LQ	26,12	< LQ	79,30						
Crotonaldeído	< LD	40,13	< LD	< LQ	< LD	< LD	< LQ	< LD	< LD	80,66
Butiraldeído	< LQ	23,98	21,04	33,63	21,09	< LQ	20,88	20,45	18,58	60,08
Benzaldeído	< LQ	123,79	130,56	105,80	111,36	74,98	108,93	105,72	85,46	164,74
Isovaleraldeído	< LQ	< LD	< LQ	< LQ	< LD	< LD	< LQ	< LD	< LD	< LQ
Valeraldeído	< LQ	< LQ	< LD	37,35						
o-tolualdeído	< LD									
m,p-tolualdeído	< LD									
hexaldeído	< LD	30,74	< LD	< LD	27,47	< LQ	< LD	< LQ	18,99	29,50
2,5-dimetilbenzaldeído	< LD									

< LD representa valores não detectáveis por estarem abaixo do LD do método e < LQ representa valores não quantificáveis, por estarem abaixo do LQ do método.



Figura 33: Concentração dos principais CC encontrados nas amostras coletadas no IAG.

Similarmente à FSP, nas amostras do IAG foi observada uma alta similaridade entre os perfis de CC. A acetona, em algumas amostras, apresenta concentração maior do que acetaldeído e formaldeído. A predominância de maiores concentrações de acetaldeido em relação a formaldeído mais uma vez foi observada. As concentrações dos analitos encontrados no ICB estão mostradas nas Tabelas 7 e 8 os principais analitos encontrados são mostrados de forma gráfica na Figura 34.

Tabela 7: Concentração (em ng/mL) dos principais CC encontrados nas amostras coletadas no ICB (Amostras 1-10).

Concentração (ng/mL)	ICB 1	ICB 2	ICB 3	ICB 4	ICB 5	ICB 6	ICB 7	ICB 8	ICB 9	ICB 10
Formaldeído	798,78	265,42	599,66	599,66	336,45	336,45	247,67	652,95	475,84	256,92
Acetaldeído	1340,82	254,30	1164,50	1164,50	419,51	377,67	310,84	1019,18	685,11	349,18
Acetona	897,83	374,16	828,41	828,41	451,34	443,69	396,23	805,92	566,67	360,83
Acroleína	< LD	< LD	< LD	< LD	< LD	< LD	< LD	< LD	< LD	< LD
Propinaldeído	224,19	32,55	158,54	158,54	51,22	50,65	43,90	175,42	111,25	53,69
Crotonaldeído	22,25	NQ	48,62	48,62	< LD	< LD	< LD	34,30	21,60	< LD
Butiraldeído	138,89	34,73	99,97	99,97	39,38	36,52	37,63	102,43	68,80	42,17
Benzaldeído	103,34	81,88	96,22	96,22	73,14	83,07	49,30	309,36	282,96	76,88
Isovaleraldeído	NQ	< LD	NQ	NQ	< LD	< LD	< LD	< LD	< LD	< LD
Valeraldeído	75,91	18,19	60,70	60,70	24,77	24,53	20,91	84,21	39,06	19,97
o-tolualdeído	< LD	< LD	63,10	< LD	< LD	< LD	< LD	< LD	< LD	< LD
m,p-tolualdeído	< LD	< LD	< LD	< LD	< LD	< LD	< LD	40,18	NQ	< LD
hexaldeído	58,98	< LD	28,74	28,74	< LD	33,26	< LD	53,27	30,51	< LD
2,5-dimetilbenzaldeído	< LD	< LD	< LD	< LD	< LD	< LD	< LD	< LD	< LD	< LD

< LD representa valores não detectáveis por estarem abaixo do LD do método e < LQ representa valores não quantificáveis, por estarem abaixo do LQ do método.</p>

Concentração (ng/mL)	ICB 11	ICB 12	ICB 13	ICB 14	ICB 15	ICB 16	ICB 17
Formaldeído	268,44	286,62	190,92	163,25	268,44	320,46	256,36
Acetaldeído	349,90	397,29	195,34	158,39	349,90	331,13	299,26
Acetona	375,64	497,18	342,29	287,98	375,64	450,95	350,78
Acroleína	< LD						
Propinaldeído	56,05	69,43	29,99	21,70	56,05	47,66	48,18
Crotonaldeído	< LD	NQ	NQ				
Butiraldeído	42,73	48,48	40,80	28,94	42,73	44,26	41,53
Benzaldeído	61,11	46,37	53,51	55,11	61,11	241,73	217,13
Isovaleraldeído	< LD						
Valeraldeído	20,90	33,16	16,42	3,62	20,90	23,31	< LD
o-tolualdeído	< LD						
m,p-tolualdeído	NQ	< LD	< LD	< LD	NQ	< LD	NQ
hexaldeído	28,02	34,35	< LD	< LD	28,02	29,58	< LD
2,5-dimetilbenzaldeído	< LD						

Tabela 8: Concentração (em ng/mL) dos principais CC encontrados nas amostras coletadas no ICB (Amostras 11-17).



Figura 34: Concentração dos principais CC encontrados nas amostras coletadas no ICB.

Foi observado um número maior de CC em relação aos dois sítios anteriores, como por exemplo propionaldeído e valeraldeído. Além disto, pode-se verificar que há uma maior concentração dos CC formaldeído, acetaldeído, acetona, butiraldeído e benzaldeído. Finalmente, a similaridade entre os perfis de concentrações dos CC identificados e quantificados também foi observado para as amostras coletadas neste sítio. A Figura 35 apresenta de forma comparativa a concentração dos CC encontrados nos três sítioss de amostragem estudados.



Figura 35: Comparação das concentrações de formaldeído, acetaldeído, acetona, butiraldeído e benzaldeído nos três sítios estudados.

Como se pode observar na Figura 35, os sítioss FSP e IAG são similares entre si em relação à concentração dos CC, enquanto o ICB apresenta uma maior concentração de formaldeído, acetadeído, acetona e butiraldeído. Esta tendência apenas não foi observada para o benzaldeído.

85

3.1.3.4. Razão Formaldeído/Acetaldeído (F/A)

Para todos os sítios, os valores encontrados de acetaldeído foram maiores do que de formaldeído. Na Figura 36 é mostrado o perfil de concentrações de formaldeído e acetaldeído para cada um dos sítios.



Figura 36: Comparação das concentrações de formaldeído e acetaldeído nos três sítios estudados.

A razão formaldeído/acetaldeído (F/A) para cada um dos sítios é mostrada na Figura 37. É possível observar que a razão F/A é predominantemente menor do que 1. Este fato está em concordância com diversos estudos que indicam que a razão F/A < 1 está relacionada com a característica dos poluentes emitidos por veículos movidos a etanol (de Andrade, 1998).



ICB07

(CB))

Amostra

(CB13

(CB15

Figura 37: Gráficos da razão formaldeído/acetaldeído (F/A) nos três sítios estudados.

1(CB0)3

(CBO)

0,40 0,20 0,00

3.1.4. Quantificação em ppbv ou razão de mistura

Para que os dados de concentração de CC coletados no ar possam ser comparados com os dados da literatura, é necessária a conversão para a unidade ppbv ou razão de mistura. Esta conversão considera a Lei dos gases ideais, o volume de ar succionado através do cartucho de DNPH e a temperatura da coleta, através da Equação 24:

$$V_{CC} = \frac{M_{cc}}{MW} \times (R \times T_{amb}) \times \frac{760}{P_{amb}}$$
 Equação 24

em que:

 V_{CC} = volume do CC; M_{CC} = massa do CC no cartucho; MW = peso molecular do CC ; R = constante dos gases ideais; T_{amb} = temperatura ambiente; P_{CC} = pressão ambiente.

A concentração em ppbv do CC correspondente é dada pela Equação 25:

$$C_{CC} ppbv = \frac{V_{CC}}{V_{ar}}$$
 Equação 25

As Tabelas 9-12 apresentam os valores de concentração, em ppbv, para os CC nos sítios FSP, IAG e ICB, respectivamente.

Concentração (pphy)	EQD 1	EGD 2	ESD /	EGD 5	ESD 6	ECD 0	ESD 0	ESD 10	ESD 17	ECD 19	ESD 10
Concentração (ppbv)	rə r I	rə r 3	r ər 4	rə r 3	rə r 0	r ər 0	ror 9	F3P 10	F3F 1/	F3F 10	F3F 19
Formaldeído	2,18	1,86	1,97	1,85	0,60	1,66	1,71	2,24	3,92	2,11	1,30
Acetaldeído	3,47	2,17	1,76	1,84	1,02	1,58	1,26	2,42	4,58	2,44	1,57
Acetona	2,99	2,51	1,23	1,46	1,17	1,24	1,15	2,11	2,83	1,92	1,16
Acroleína	< LQ	< LQ	< LQ	< LQ	< LQ	< LQ	< LQ	< LQ	< LQ	< LQ	< LQ
Propinaldeído	0,22	< LQ	< LQ	< LQ	< LD	< LQ	0,35	< LQ	0,35	< LQ	< LQ
Crotonaldeído	< LQ	< LQ	< LD	< LD	< LD	0,17	< LD	< LD	0,18	0,17	< LD
Butiraldeído	0,16	0,18	0,15	< LQ	< LD	0,14	< LQ	0,13	0,24	0,15	< LQ
Benzaldeído	0,23	0,76	0,43	0,44	0,31	0,41	0,38	0,67	0,95	0,43	0,34
Isovaleraldeído	< LD	< LQ	< LD	< LD	< LD	< LD	< LD	< LD	< LD	< LD	< LD
Valeraldeído	< LD	< LQ	< LD	< LD	< LD	< LD	< LD	< LD	< LD	< LD	< LD
o-tolualdeído	< LD	< LD	< LD	< LD	< LD	< LD	< LD	< LD	< LD	< LD	< LD
m,p-tolualdeído	< LQ	< LD	< LD	< LD	< LQ	< LD	< LD				
hexaldeído	< LD	0,10	< LD	< LD	< LD	< LD	< LD	< LD	< LD	< LD	< LD
2,5-dimetilbenzaldeído	< LD	< LD	< LD	< LD	< LD	< LD	< LD	< LD	< LD	< LD	< LD

Tabela 9: Concentração (em ppbv) dos principais CC encontrados nas amostras coletadas na FSP.

Concentração (ppbv)	IAG 1	IAG 2	IAG 3	IAG 4	IAG 5	IAG 6	IAG 7	IAG 9	IAG 10	IAG 18
Formaldeído	0,55	2,50	1,44	1,56	1,34	0,93	1,63	1,68	1,52	10,43
Acetaldeído	1,32	4,49	1,72	1,39	2,12	0,96	1,49	1,30	2,40	7,93
Acetona	1,90	2,48	1,97	3,09	2,92	1,37	1,45	1,82	1,80	4,73
Acroleína	< LD	< LD	< LQ	< LQ	< LD	< LQ				
Propinaldeído	< LQ	0,21	< LQ	0,77						
Crotonaldeído	< LD	0,27	< LD	< LQ	< LD	< LD	< LQ	< LD	< LD	0,65
Butiraldeído	< LQ	0,16	0,14	0,22	0,14	< LQ	0,12	0,13	0,12	0,47
Benzaldeído	< LQ	0,55	0,59	0,47	0,51	0,35	0,42	0,46	0,38	0,88
Isovaleraldeído	< LQ	< LD	< LQ	< LQ	< LD	< LD	< LQ	< LD	< LD	< LQ
Valeraldeído	< LQ	< LQ	< LD	0,20						
o-tolualdeído	< LD	< LD								
m,p-tolualdeído	< LD	< LD								
hexaldeído	< LD	0,14	< LD	< LD	0,12	< LQ	< LD	< LQ	0,08	0,16
2,5-dimetilbenzaldeído	< LD	< LD								

Tabela 10: Concentração (em ppbv) dos principais CC encontrados nas amostras coletadas no IAG.

Tabela 11: Concentração (em ppbv) dos principais CC encontrados nas amostras coletadas no ICB (Amostras 1-10).

Concentração (ppbv)	ICB 1	ICB 2	ICB 3	ICB 4	ICB 5	ICB 6	ICB 7	ICB 8	ICB 9	ICB 10
Formaldeído	20,01	5,04	11,40	11,40	6,39	9,51	7,00	17,86	13,45	6,17
Acetaldeído	22,90	3,29	15,09	15,09	5,44	7,28	5,99	19,01	13,20	5,72
Acetona	12,05	3,81	8,43	8,43	4,59	6,72	6,00	11,81	8,58	4,64
Acroleína	< LD									
Propinaldeído	2,90	0,32	1,56	1,56	0,50	0,74	0,64	2,48	1,63	0,67
Crotonaldeído	0,24	< LQ	0,40	0,40	< LD	< LD	< LD	0,40	0,26	< LD
Butiraldeído	1,45	0,27	0,79	0,79	0,31	0,43	0,44	1,17	0,81	0,42
Benzaldeído	0.73	0,44	0.52	0,52	0.39	0,66	0,39	2,40	2,26	0.52
Isovaleraldeído	< LQ	< LD	< LQ	< LQ	< LD					
Valeraldeído	0.66	0.12	0.40	0.40	0.16	0.24	0.21	0.80	0.39	0.17
o-tolualdeído	< LD	< LD	0,30	< LD						
m,p-tolualdeído	< LD	0.27	< LQ	< LD						
hexaldeído	0.44	< LD	0.16	0.16	< LD	0.28	< LD	0.44	0.26	< LD
2,5-dimetilbenzaldeído	< LD									

< LD representa valores não detectáveis por estarem abaixo do LD do método e < LQ representa valores não quantificáveis, por estarem abaixo do LQ do método.</p>

Concentração (ppbv)	ICB 11	ICB 12	ICB 13	ICB 14	ICB 15	ICB 16	ICB 17
Formaldeído	9,95	10,80	6,12	4,83	10,03	6,04	4,47
Acetaldeído	8,84	10,21	4,27	3,20	8,92	4,25	3,55
Acetona	7,46	10,04	5,88	4,57	7,52	4,55	3,27
Acroleína	< LD						
Propinaldeído	1,07	1,35	0,50	0,33	1,08	0,46	0,43
Crotonaldeído	< LD	< LQ	< LQ				
Butiraldeído	0,66	0,76	0,54	0,36	0,67	0,35	0,30
Benzaldeído	0,64	0,49	0,49	0,46	0,65	1,29	1,07
Isovaleraldeído	< LD						
Valeraldeído	0,27	0,44	0,18	0,04	0,27	0,15	< LD
o-tolualdeído	< LD						
m,p-tolualdeído	< LQ	< LD	< LD	< LD	< LQ	< LD	< LQ
hexaldeído	0,31	0,39	< LD	< LD	0,31	29,58	< LD
2,5-dimetilbenzaldeído	< LD						

Tabela 12: Concentração (em ppbv) dos principais CC encontrados nas amostras coletadas no ICB (Amostras 11-17).

As Figuras 38 - 40 apresentam os gráficos em concentração (ppbv) vs amostras para os CC encontrados na FSP, IAG e ICB, respectivamente.



Figura 38: A) Gráfico de concentração dos principais CC (ppbv) encontrados nas amostras coletadas na FSP e B) gráfico enfatizando o perfil de concentração do formaldeído e do formaldeído (ppbv).



Figura 39: A) Gráfico de concentração dos principais CC (ppbv) encontrados nas amostras coletadas no IAG e B) gráfico enfatizando o perfil de concentração do formaldeído e do formaldeído (ppbv).



Figura 40: A) Gráfico de concentração dos principais CC (ppbv) encontrados nas amostras coletadas no ICB e B) gráfico enfatizando o perfil de concentração do formaldeído e do formaldeído (ppbv).

3.2. Tratamento multivariado dos resultados

3.2.1. Inspeção visual dos cromatogramas

Um passo muito importante na etapa de pré-processamento dos cromatogramas é a realização de uma inspeção visual em busca de picos que aparecem no início da corrida cromatográfica, os quais normalmente se referem a picos do solvente e, neste caso, também do DNPH. Foi verificado que até 5 min. de corrida, não há eluição de nenhum CC, conforme mostrado na Figura 41. Por este motivo, o cromatograma foi cortado a partir de 5 min. para o tratamento multivariado.



Figura 41: Cromatograma das amostras coletadas nos três sítios enfatizando informações não relevantes presentes no início do cromatograma.

3.2.2. Alinhamento dos cromatogramas

A seqüência de aquisição dos dados foi de 10 Hz, ou seja, 10 pontos a cada segundo. Portanto, sendo o tempo de corrida de 27 min., foram gerados 16201 pontos. Com os cromatogramas dos padrões de 25, 50, 75, 100 e 150 ppb foi montada uma matriz 25X13201 após a retirado dos picos do solvente e do DNPHD. Da mesma forma, com os cromatogramas das amostras foi montada uma matriz final 33x13201. Para fins exploratórios, os cromatogramas dos padrões foram alinhados com os algoritmos COW e Optim_COW. Os cromatogramas das amostras foram alinhados com os algoritmos COW, Optim_COW e Peakmatch.

A qualidade do alinhamento foi avaliada por inspeção visual nos casos mais simples e por PCA nos caso em que a simples inspeção visual não foi conclusiva. A avaliação da qualidade do alinhamento por PCA envolve a análise do gráfico dos *loadings*: quando a matriz não está completamente alinhada, surgem picos fantasmas nos *loadings*, geralmente para baixo

3.2.2.1. Alinhamento dos cromatogramas dos padrões

Para a aplicação do COW foram escolhidos três cromatogramas distintos como referência: o primeiro cromatograma da matriz (Ref. 1), o cromatograma do meio da matriz (Ref. 13) e o último (Ref. 25). Os resultados dos alinhamentos estão mostrados nas Figuras 42 a 44.



Figura 42: Cromatograma dos padrões após alinhamento utilizando o algoritmo COW, com amostra 1 selecionada como referência: A) cromatograma inteiro e B) pico do formaldeído aumentado.



Figura 43: Cromatograma dos padrões após alinhamento utilizando o algoritmo COW, com amostra 13 selecionada como referência. A) cromatograma inteiro e B) pico do formaldeído aumentado.



Figura 44: Cromatograma dos padrões após alinhamento utilizando o algoritmo COW, com amostra 25 selecionada como referência. A) cromatograma inteiro e B) pico do formaldeído aumentado.

Uma simples inspeção visual foi necessária para concluir que a qualidade do alinhamento não é satisfatória: nenhumas das três referências escolhidas forneceram um alinhamento adequado.

Foi observado neste trabalho, que o uso do COW (para as matrizes estudadas) requer aplicações sucessivas do algoritmo na mesma matriz e com a mesma referência para se conseguir melhorias na qualidade do alinhamento.

O Optim_COW foi aplicado para a matriz de dados contendo os cromatogramas dos padrões desalinhados (Figura 45).



Figura 45: Cromatograma dos padrões após alinhamento utilizando o algoritmo Optim-COW: A) cromatograma inteiro e B) pico do formaldeído aumentado.

Neste caso, uma simples inspeção visual é suficiente para notar a alta qualidade do alinhamento, entretanto, também foi avaliada através da PCA, e o resultado obtido confirmou a eficiência do alinhamento.

3.2.2.2. Alinhamento dos cromatogramas das amostras

Os cromatogramas das amostras dos três sítios foram organizados em uma matriz única e alinhados empregando os algoritmos COW, Optim_COW e PEAKMATCH, para fins comparativos.

O mesmo critério de escolha das referências aplicado no alinhamento das padrões foi utilizado no alinhamento das amostras, ou seja, foram escolhidos três cromatogramas distintos como referência: o primeiro cromatograma da seqüência (Ref.

1), o cromatograma do meio (Ref.16) e o último (Ref. 33). Os resultados dos alinhamentos estão mostrados nas Figuras 46 a 48.



Figura 46: Cromatogramas das amostras provenientes das três campanhas após alinhamento empregando algoritmo COW e amostra 1 como referência: A) Cromatograma inteiro; B) pico do formaldeído aumentado.



Figura 47: Cromatogramas das amostras provenientes das três campanhas após alinhamento empregando algoritmo COW e amostra 16 como referência: A) Cromatograma inteiro; B) pico do formaldeído aumentado.



Figura 48: Cromatogramas das amostras provenientes das três campanhas após alinhamento empregando algoritmo COW e amostra 33 como referência: A) Cromatograma inteiro; B) pico do formaldeído aumentado.

Através da inspeção visual das Figuras 46 - 48 pode ser verificado que não foi obtido um alinhamento satisfatório com nenhuma das referências. Além disto, ficou claro que a escolha da referência é um parâmetro importante para a qualidade do alinhamento: as referências 16 e 33 apresentaram resultados bastante parecidos, e superiores em relação ao alinhamento feito com a referência 1.

O Peakmatch, assim como o COW, necessita de uma escolha prévia do cromatograma a ser utilizado como referência no alinhamento. As mesmas referências escolhidas no alinhamento das amostras utilizando o COW foram utilizadas no alinhamento das amostras com o Peakmatch. Os resultados dos alinhamentos estão mostrados nas Figuras 49 - 51, empregando as referências 1, 16 e 33, respectivamente.



Figura 49: Cromatogramas das amostras provenientes das três campanhas após o alinhamento empregando o algoritmo Peakmatch e amostra 1 como referência A) Cromatograma inteiro; B) pico do formaldeído aumentado.



Tempo / minutos

Figura 50: Cromatogramas das amostras provenientes das três campanhas após o alinhamento empregando o algoritmo Peakmatch e amostra 16 como referência A) Cromatograma inteiro; B) pico do formaldeído aumentado.



Figura 51: Cromatogramas das amostras provenientes das três campanhas após o alinhamento empregando o algoritmo Peakmatch e amostra 33 como referência A) Cromatograma inteiro; B) pico do formaldeído aumentado.

Neste caso, a inspeção visual indica um alinhamento de boa qualidade, com qualquer uma das referências. Entretanto, quando a qualidade do alinhamento foi avaliada através da PCA, verificou-se que o alinhamento não foi eficiente para nenhuma das referências, pois apareceram picos para baixo no gráfico de *loadings*.

O Optim_COW, ao contrário do COW e do Peakmatch, não necessita de uma escolha prévia do cromatograma de referência. Entretanto, O Optim_COW não é recomendado para alinhamento de cromatogramas complexos, sendo necessário um pré-alinhamento antes da aplicação do mesmo (Skov *et al.*, 2006). Esta observação foi confirmada através da PCA para o conjunto de dados estudado. Sendo assim, foi

realizado o pré-alinhamento da matriz das amostras com o algoritmo PEAKMATCH com posterior alinhamento empregando o algoritmo Optim-COW. Os cromatogramas alinhados empregando as referências 1, 16 e 33 para o pré-alinhamento são mostrados respectivamente nas Figuras 52 a 54.



Figura 52: Cromatogramas das amostras originais provenientes das três campanhas após o alinhamento empregando o algoritmo Optim-COW após pré-alinhamento empregando o algoritmo Peakmatch (Ref.1): A) Cromatograma inteiro; B) pico do formaldeído aumentado.



Figura 53: Cromatogramas das amostras originais provenientes das três campanhas após o alinhamento empregando o algoritmo Optim-COW após pré-alinhamento empregando o algoritmo Peakmatch (Ref. 16): A) Cromatograma inteiro; B) pico do formaldeído aumentado.



Figura 54: Cromatogramas das amostras originais provenientes das três campanhas após o alinhamento empregando o algoritmo optim-COW após pré-alinhamento empregando o algoritmo Peakmatch (Ref. 33): A) Cromatograma inteiro; B) pico do formaldeído aumentado.

O alinhamento executado com as referências 1 e 33 apresentou ótima qualidade, confirmada através da PCA. Portanto, a combinação Peakmatch + Optim_COW foi a

melhor solução para o alinhamento dos cromatogramas das amostras. A matriz alinhada empregando a referência 1 para o pré-alinhamento seguida do Optim_COW foi utilizada para os pré-processamentos subseqüentes.

3.2.3. Correção da linha base

Após o alinhamento dos cromatogramas, os deslocamentos de linha base foram corrigidos empregando o método da derivação. A Figura 55 mostra os cromatogramas originais, destacando-se os deslocamentos de linha-base. Figura 56 mostra os cromatogramas das amostras alinhadas na forma de primeira derivada.



Figura 55: Cromatogramas das amostras após alinhamento, destacando os deslocamentos de linha base.



Figura 56: Cromatograma das amostras alinhadas na forma de primeira derivada.

Conforme é possível observar na Figura 56, a primeira derivada eliminou os deslocamentos de linha base presente nos cromatogramas originais. Com uma janela de 15, o cromatograma não apresenta ruído e também foi observado que não houve perda de informação analítica com esta suavização, o que foi constatado pela análise dos picos menores. Como não se observou inclinação da linha base, não foi necessária a utilização da segunda derivada.

3.2.4. Pré-processamento das variáveis

O cromatograma é considerado como um conjunto de variáveis contínuas e por este motivo, para a realização do tratamento quimiométrico, os mesmos foram apenas centrados na média. O autoescalamento é adequado para variáveis discretas, as quais podem ter diferenças de magnitude (Skov, 2008).

3.2.5. Aplicação de PCA e HCA na matriz final de dados

Para facilitar a visualização nos gráficos das análises de PCA e HCA, as amostras foram codificadas por números, conforme descrito na Tabela 13

ID amostra	Código	Data	Período	ID amostra	Código	Data	Período
IAG 1	1	25/5/2009	05:00-07:00	ICB 01	17	18/8/2008	8:11-10:33
IAG 2	2	25/5/2009	8:00-10:00	ICB 02	18	18/8/2008	14:55-16:55
IAG 4	3	25/5/2009	12:00-14:00	ICB 03	19	19/8/2008	8:10-10:10
IAG 5	4	25/5/2009	14:00-16:00	ICB 04	20	19/8/2008	10:12-12:17
IAG 6	5	26/5/2009	05:00-07:00	ICB 05	21	19/8/2008	12:17-14:17
IAG 7	6	26/5/2009	8:00-10:00	ICB 06	22	19/8/2008	12:22-14:22
IAG 9	7	26/5/2009	12:00-14:00	ICB 07	23	19/8/2008	14:25-16:00
IAG 18	8	28/5/2009	10:00-12:00	ICB 08	24	20/8/2008	8:51-10:51
FSP 3	9	25/5/2009	10:00-12:00	ICB 09	25	20/8/2008	10:01-12:01
FSP 4	10	25/5/2009	12:00-14:00	ICB 10	26	20/8/2008	12:04-14:05
FSP 5	11	25/5/2009	14:00-16:00	ICB 11	27	20/8/2008	14:10-16:11
FSP 9	12	26/5/2009	12:00-14:00	ICB 12	28	21/8/2008	8:05-10:05
FSP 10	13	26/5/2009	14:00-16:00	ICB 13	29	21/8/2008	10:06-12:13
FSP 17	14	28/5/2009	8:00-10:00	ICB 14	30	21/8/2008	12:11-14:06
FSP 18	15	28/5/2009	10:00-12:00	ICB 15	31	21/8/2008	14:09-16:10
FSP 19	16	28/5/2009	12:00-14:00	ICB 16	32	22/8/2008	10:01-12:01
				ICB 17	33	22/8/2008	11:56-14:00

Tabela 13: Código e informação das amostras para a análise multivariada.

3.2.5.1. Sítios IAG, ICB e FSP

Na aplicação da PCA, foi observado que o PC1 é capaz de descrever 97,57 % da variância experimental, conforme mostrado na Figura 57. Este resultado indica que todas as variáveis apresentam uma variação semelhante de tal forma que apenas uma PC representa a quase totalidade da variação. O perfil similar de concentrações dos CC também foi observado na análise univariada (Figuras 32, 33 e 34). A grande porcentagem de variação explicada pela PC1 pode sugerir que os CC foram lançados na atmosfera por uma única fonte de emissão nos três sítios estudados. Os gráficos de *scores* na PC 1 *vs* amostra e *scores* na PC 1 *vs* PC 2 são mostrados nas Figuras 58 e 59, respectivamente.



Figura 57: Gráfico de variância explicada ou capturada vs número de Componentes Principais.



Figura 58: Gráfico de scores das amostras dos três sítios na PC1.



Figura 59: Gráfico de scores, PC1 vs PC2 das amostras dos três sítios.
Através da análise do gráfico de *scores* na PC1, foi possível obter algumas informações a respeito da matriz de dados:

> As amostras do IAG e FSP são bastante similares entre si e diferentes das amostras do ICB.

> As amostras do ICB apresentam uma variabilidade maior entre si, pois estão mais dispersas que as demais.

O gráfico dos *loadings* na forma de primeira derivada é mostrado na Figura 60. Para melhor visualização, o cromatograma derivado foi integrado (Figura 61). Através da análise dos *loadings* foi observado que os compostos formaldeído, acetaldeído e acetona são responsáveis pela configuração das amostras no gráfico dos *scores*.



Figura 60: Loadings na PC1 (forma derivada).



Figura 61: Loadings na PC1 (forma integrada).

Na análise de HCA (dendograma mostrado na Figura 62), um grupo de amostras muito similares, formado pela maioria das amostras do IAG e da FSP, foi observado, além de uma maior dispersão entre as amostras do ICB, confirmando os resultados da PCA.



Figura 62: Dendograma das amostra dos sítios, IAG, ICB e FSP.

A PCA e a HCA foram eficientes em classificar as amostras, considerando que em ambos os métodos, as amostras do IAG e da FSP formaram um único agrupamento enquanto as amostras do ICB formaram um grupo separado, o que está de acordo com os resultados da quantificação univariada. Estes resultados indicam o potencial que os métodos de análise exploratória multivariada empregando o cromatograma completo apresentam para avaliar e classificar de forma qualitativa amostras ambientais. Estes métodos se tornam de grande valia quando um número maior de sítios for considerado, ou em casos de monitoramento ambiental, situações nas quais a análise univariada pode ser extremamente laboriosa.

No presente trabalho é importante considerar que as variáveis ligadas às condições meteorológicas podem ter sido determinantes na interpretação ambiental dos

dados. As amostras da FSP e IAG foram coletadas simultaneamente em maio de 2009, enquanto as amostras do ICB foram coletadas em outro período, agosto de 2008. O período de amostragem na FSP e IAG foi de dias chuvosos e tempo nublado. Por outro lado, o período de amostragem no ICB foi marcado por dias ensolarados. Em virtude dos períodos de amostragem dos três sítios terem sido diferentes, os resultados das amostras da FSP e IAG também foram tratados separadamente das amostras do ICB.

3.2.5.2. Sítios IAG e FSP

O gráfico dos scores na PC1 vs PC2 é mostrado na Figura 63.



Figura 63: Gráfico de scores, PC1 vs PC2 das amostras dos sítios IAG e FSP.

O resultado obtido anteriormente na PCA realizada com os três sítios em conjunto em que as amostras do IAG e da FSP formaram um único agrupamento foi confirmado pela PCA com os dois sítios (Figura 63). A PC1 descreveu aproximadamente 93% da variação, sendo que três amostras do IAG estão mais dispersas em relação ao conjunto: amostra 3 (IAG4, 25/5/2009,12:00-14:00), amostra 4 (IAG5, 25/5/2009,14:00-16:00) e amostra 8 (IAG18, 28/5/2009,10:00-12:00). Pela observação do gráfico de *loadings* é possível verificar que estas amostras possuem concentração diferenciada de formaldeído, acetaldeído e especialmente de acetona, o que foi de fato confirmado pela análise dos cromatogramas e resultados de quantificação.



Figura 64: Loadings na PC1 dos sítios IAG e FSP (forma integrada).

A HCA confirmou o resultado obtido na PCA, evidenciando a distância que as amostras 3, 4 e 8 possuem em relação ao agrupamento formado pelas demais amostras (Figura 65).



Figura 65: Dendograma das amostras dos sítios IAG e FSP.

O resultado obtido pela PCA e HCA através da análise conjunta dos sítios IAG e FSP ressalta a capacidade que a análise exploratória multivariada possui para evidenciar diferenças e semelhanças das amostras em uma análise global. A grande importância desta ferramenta pode ser evidenciada nos casos em que existe um número grande de amostras ou de análises de monitoramento contínuo, pois pode facilitar a interpretação ambiental do conjunto de dados através da indicação de amostras com comportamento diferenciado.

3.2.5.3. Sítio ICB

O gráfico dos *scores* na PC1 *vs* PC2 (Figura 66) evidencia que a PC1 separa a amostra 17 (ICB02, 18/8/2008,14:55-16:55), amostra 19 (ICB04, 19/8/2008, 10:12-12:17), amostra 20 (ICB 05, 19/8/2008, 12:17-14:17), amostra 21 (ICB06, 19/8/2008, 12:22-14:22) e amostra 28 (ICB13, 21/8/2008, 10:06-12:13), do agrupamento formado pelo restante das amostras. As amostras 17 e 19 são as mais distantes do grupo. O gráfico dos *loadings* (Figura 67) indica que os CC formaldeído, acetaldeído e acetona são responsáveis pela configuração no gráfico dos *scores*.



Figura 66: Gráfico de scores, PC1 vs PC2, das amostras do sítio ICB.



Figura 67: Loadings na PC1 do sítio ICB (forma integrada).

O dendograma proveniente da HCA é mostrado na Figura 68. A HCA confirma as conclusões observadas na PCA, ou seja, as amostras 17, 19, 20, 21 e 28 não estão agrupadas com o restante das amostras.



Figura 68: Dendograma das amostras do sítio ICB.

CAPÍTULO 4: CONCLUSÕES

Neste trabalho, foi observado que as amostras coletadas nos sítios IAG e FSP são similares quanto à natureza e concentração dos CC (formaldeído, acetaldeído, acetona, butiraldeído e benzaldeído). Por outro lado, as amostras coletadas no sítio ICB apresentaram concentrações superiores dos CC em relação aos outros dois sítios além de outros CC (propionaldeído e valeraldeído) não boservados nos mesmos. Outra observação importante foi que existe um perfil muito semelhante de concentração de praticamente todos os CC em qualquer um dos sítios de amostragem.

Foi observado que a razão F/A foi menor do que 1 para a maioria das amostras. Este resultado está de acordo com dados de diversos trabalhos que indicam que esta relação pode ser uma característica do uso de etanol em larga escala como combustível veicular.

Para a análise multivariada foram usados os cromatogramas completos e não os dados discretos de concentração. No alinhamento dos cromatogramas das amostras ficou evidente que a melhor estratégia de alinhamento inclui o pré-alinhamento com o algoritmo Peakmatch seguido do alinhamento com o algoritmo Optim_COW. Esta observação confirmou dados da literatura que mencionam o fato do Optim_COW não ser adequado para aplicações diretas em cromatogramas complexos, sendo necessário um pré-alinhamento da matriz. Com a matriz corrigida foi aplicada a PCA e observou-se que a PC1 descreveu a quase totalidade da variância experimental. Desta maneira, o conjunto de dados foi reduzido de 13201 variáveis para apenas uma Componente Principal, a qual é a combinação linear destas variáveis. Este resultado pode ser explicado pela alta correlação entre as variáveis sugerindo a existência de fontes de emissão similares ou idênticas. Essa interpretação concorda com os perfis semelhantes dos CC encontrados nas amostras dos três sítios estudados (análise univariada). O

gráfico de *scores* da PCA permitiu verificar a semelhança entre as amostras dos sítios IAG e FSP em relação às amostras do ICB. Também foi possível verificar que as amostras do IAG e FSP estavam menos dispersas, ou seja, apresentavam menor variabilidade em relação às amostras do ICB. O gráfico de *loadings* reproduziu o perfil do cromatograma indicando que todas as variáveis contribuem para a distinção das amostras do ICB, fato também observado na análise univariada. Salienta-se que a formação de agrupamentos distintos entre as amostras da FSP e IAG em relação às do ICB, observada no gráfico de *scores* e no dendograma, pode ter sido ocasionada devido às condições meteorológicas diferentes.

Com base nos resultados obtidos pela análise exploratória mutivariada é possível estabelecer um protocolo geral para utilização de cromatograma completo na análise de CC na atmosfera:

1. Construção de uma matriz com os cromatogramas das amostras;

2. Escolha das estratégias de pré-alinhamento e alinhamento;

3. Identificação dos picos nos cromatogramas das amostras, se necessário, utilizando os cromatogramas dos padrões como referência;

4. Aplicação da derivada ou outro algoritmo de correção, se existir deslocamentos verticais ou de inclinação de linha base;

5. Aplicação de alisamento e verificação do tamanho da janela;

6. Aplicação de PCA e HCA na matriz final.

Referências Bibliográficas

ANFAVEA, Anuário da Indústria Automobilística Brasileira. São Paulo, 2009.

BEEBE K. R.; PELL, R. J.; SEASHOLTZ, M. B.; Chemometrics: a practical guide. New York, John Wiley & Sons:,1997..

BRERETON, R. Chemometrics for Pattern Recognition. Chichester, UK, John Wiley & Sons, Ltd, 2009.

BRUNS, R.E.; SCARMINIO, I.S., de BARROS NETO, B. Como fazer experimentos. São Paulo. UNICAMP, 2001.

CARTER, W.P.L., Computer modeling of environmental clamber measurements of maximum incremental reactivities of volatile organic compounds. *Atmosphere Environment*, v.29, p.2513-2527, 1995.

CESTEB, Relatório da qualidade do ar no Estado de São Paulo 2008. São Paulo, 2009.

CORRÊA, P. R. M.; FERREIRA, M. M. C. Reconhecimento de padrões por métodos não supervisionados: explorando procedimentos quimiométricos para tratamento de dados analíticos, *Quimica Nova*, v. 30, p.481-487, 2007.

CORRÊA, S. M.; MARTINS, E.; ARBILLA, G. Formaldehyde and acetaldehyde in a high traffic street of Rio de Janeiro, Brazil. *Atmosferic. Environment*, v.37, p. 23-29, 2003.

CORRÊA, S.M.; ARBILLA, G. Carbonyl emissions in diesel and biodiesel exhaust. *Atmosferic. Environment*, v.42, p. 769-775, 2008.

CORRÊA, S.M.; ARBILLA, G. Carbonyl Formaldehyde and acetaldehyde associated with the use of natural gas as a fuel for light vehicles. *Atmosferic Environment*, v.39, p. 4513-4518, 2005.

De ANDRADE, J. B.; ANDRADE, M. V.; PINHEIRO, H. L.C. Atmospheric Levels of Formaldehyde and Acetaldehyde and their Relationship with the Vehicular Fleet Compositiosn in Salvador, Bahia, Brazil. *Journal of Brazilian Chemical Society*, v.9, p. 219-223, 1998.

De ANDRADE, M.V.A.S.; PINHEIRO, H.L.C.; PEREIRA, P.A.P.; DE ANDRADE, J.B. Compostos carbonílicos atmosféricos: fontes, reatividade, níveis de concentração e efeitos toxicológicos, *Química Nova*, v.25, p.1117-1131, 2002.

DETRAN/PRODESP, Arquivo: Frota Circulante- 2008. São Paulo, 2009.

EPA, Determination of Formaldeído in ambiente air using adsorbent cartridge followed by High Perfomace Liquid Cromatography: Compendium Method TO – 11A. second edition, 1999.

GROSJEAN, D.; MIGUEL, A.H.; TAVARES, T. Urban air pollution in Brazil: acetaldehyde and other carbonyls. *Atmosferic Environment*, v.24B, p.101-106. 1990.

GRUNG, B.; KVALHEIM, O. M. Retention time shift adjustments of two-way chromatograms using Bessel's inequality, *Analytica Chimica Acta*, v. 304, p. 57-66, 1995.

GUO, H.; SOB, K.L.; SIMPSONA, I.J.; BARLETTAA, B.; MEINARDIA, S.; BLAKEA, D.R. C1–C8 volatile organic compounds in the atmosphere of Hong Kong: Overview of atmospheric processing and source apportionment, *Atmospheric Environment*, v.41, p.1456–1472, 2007.

KEAN, A.J.; GROSJEAN, E.; GROSJEAN, D.; HARLEY, R.A. On-road measurement of carbonyls in California light-duty vehicle emissions. *Environmental Science and Technology*, v.35, p.4198-4204, 2001.

KIM, R. S.; DOMINICI, F.; BUCKLEY, T. J. Concentrations of vehicle-related air pollutants in an urban parking garage. Environmental Research, v.105, p.291-299, 2007.

MONTERO, L.; VASCONCELLOS, P. C.; SOUZA, S. R.; SANCHEZ-CCOYLLO, O. R; ANDRADE, M. F.; CARVALHO, L. R. F. Measurements of atmospheric carboxilic acids and carbonyl compounds in Sao Paulo city.; *Environmental Science Technology*, v. 35, p. 3071-3081, 2001.

NIELSEN, N. P. V.; CARTENSEN, J. M.; SMEDSGAARD, J. Aligning of single and multiple wavelength chromatographic profiles for chemometrics data analysis using correlation optimized warping, *Journal of chromatography A*, v. 805, p. 17-35, 1998.

PANG, X.; MU, Y. Characteristics of carbonyl compounds in public vehicles of Beijing city: Concentrations, sources, and personal exposures, *Atmospheric Environment*, v. 41, p. 1819-1824, 2007.

PANG, X.; MU, Y.; LEE, X.; ZHANG, Y.; Xu, ZHU. Influences of characteristic meteorological conditions on atmospheric carbonyls in Beijing, China, *Atmospheric Research*, v. 93, p. 913-919, 2009.

PANG, X.; MU, Y.; YUAN, J.; HE, H. Carbonyls emission from ethanol-blended gasoline and biodiesel-ethanol-diesel used in engines. *Atmospheric Environment*, v.42, p.1349-1358, 2008.

PANG, X.; SHI, X.; MU, Y.; HE, H.; SHUAI, S. Characteristics of carbonyl compounds emission from a diesel-engine using biodiesel-ethanol-diesel as fuel, *Atmospheric Environment*, v. 40, p. 7057-7065, 2006.

PRAVDOVA, V.; WALCZAK, B.; MASSART, D.L. A comparison of two algorithms for warping of analytical signals, *Analytica Chimica Acta*, v.456, p.77–92, 2002.

RIBEIRO, F.A.L.; MORANO S.C.; da SILVA, L.R.; Peter, R.S.; FERREIRA, M.M.C. Planilha de validação: uma nova ferramenta para estimar figuras de mérito na validação de métodos analíticos univariados, *Quimica Nova*, v. 31, n. 1, p.164-171, 2008.

SANTARSIERO, S. FUSELLI, Indoor and outdoor air carbonyl compounds correlation elucidated by principal component analysis, *Environmental Research*, v.106, p.139–147, 2008.

SKOV, T.; VAN DEN BERG, F.; TOMASI, G.; BRO, R. Automated alignment of chromatographic data. *Journal of chemometrics*, V. 20, P. 484–497, 2006.

SKOV, T. PhD Thesis, Mathematical resolution of complex chromatography measurements, Department of Food Science, Faculty of Life Sciences, University of Copenhagen, 2008.

TOMASI, G.; BERGAND, F. V. D.; ANDERSSON, C. Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data, *Journal of Chemometrics*, v. 18, p. 231-241, 2004.

WHO, Guidelines for Air Quality. Geneva, 2000.

WOLD, S. Principal Component Analysis, *Chemometrics and Intelligent Laboratory Systems*, v. 2; p. 37-52, 1987.