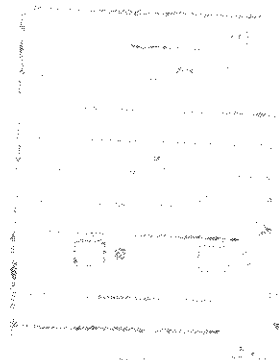


UNIVERSIDADE ESTADUAL DE CAMPINAS

Instituto de Química

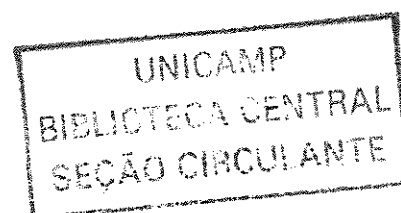
Departamento de Físico-Química



**QUIMIOMETRIA APLICADA À POLUIÇÃO ATMOSFÉRICA:
UMA METODOLOGIA PARA IDENTIFICAÇÃO DE FONTES**

Edilton de Souza Barcellos

Orientadora: Prof^a. Dra. Márcia M. C. Ferreira



Campinas, novembro/2003

UNIDADE	12
Nº CHAMADA	T/Unicamp
	B235q
V	EX
TOMBO BC	58647
PROC.	16-117-89
C	☐ ☒
PREÇO	2511,00
DATA	23/06/09
Nº CPD	

CM00198279-9

Bib id: 317480

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

B235q

Barcellos, Edilton de Souza.

Quimiometria aplicada à poluição atmosférica:
uma metodologia para identificação de fontes /
Edilton de Souza Barcellos. -- Campinas, SP:
[s.n], 2003.

Orientador: Márcia Miguel Castro Ferreira.

Tese (Doutorado) – Universidade
Estadual de Campinas, Instituto de Química.

1. Fontes poluição atmosférica. 2. PCA.
3. PARAFAC. 4. Tucker. I. Ferreira, Márcia Miguel
Castro. II. Universidade Estadual de Campinas.
III. Título.

AGRADECIMENTOS

A presente tese é fruto de um esforço do qual contribuíram várias pessoas, dentre as quais devo agradecer especialmente:

À Profa. Dra. Márcia Miguel Castro Ferreira por sua orientação, paciência e constante preocupação em proporcionar um ambiente de seriedade indispensável ao bom desempenho do trabalho acadêmico.

Aos colegas do Grupo de Quimiometria Teórica e Aplicada da UNICAMP (GQTA), em especial às amigas Luciana Sabino e Thais de Oliveira, pelo ambiente fraterno que puderam proporcionar mesmo nos momentos mais adversos;

Ao amigo Ciríaco Pinheiro (GQTA) pela acirrada, porém motivadora competição no uso da estrutura computacional à nossa disposição, e pelas discussões em temas extra acadêmicos;

À amiga Fabiana Alves Ribeiro de Lima (GQTA) pelo agradável convívio nos grupos de trabalhos acadêmicos em disciplinas cursadas, e pelas discussões e sugestões na parte ambiental;

Ao amigo Cristiano Gomes de Faria (GQTA) pelo suporte nas operações computacionais e discussão na construção de algoritmos;

Ao amigo Marlon Martins dos Reis (GQTA) pelas discussões e sugestões em Análise Multimodo e na Álgebra Linear Computacional;

Aos amigos Jesuíno Romano e Massayuki Kuromoto, da CETESB, pelo acesso ao Banco de Dados da companhia.

Agradeço de forma muito especial aos meus filhos Andressa, Guilherme e Fernando por concordarem mais uma vez em abdicar da minha presença.

Não poderia deixar de registrar meus agradecimentos a Maria das Graças M. Ribeiro, companheira para além de todas as horas nessa viagem.

Agradeço também ao Prof. Luis Henrique Mendes da Silva (DEQ/UFV) pelo seu constante apoio à concretização dessa tese.

Aos amigos Sílvia R. Silva, Andréa Moreno e sua filha Camila, pelos importantes momentos de descontração que nos permitimos usufruir.

À Isabel e ao André pela competência e presteza com que sempre me atenderam na secretaria do programa;

Aos funcionários da biblioteca do IQ/UNICAMP pelo zelo e atenção que sempre dedicaram às minhas solicitações;

Ao amigo Reinaldo Francisco Teófilo (GQTA), pela colaboração na preparação da apresentação final.

Agradeço ainda ao Prof. Benjamin Gonçalves Milagres, Chefe do Departamento de Química da UFV, aos professores das áreas de Físico-Química e Química Geral, que concordaram em assumir meus encargos durante o período de afastamento; assim como a Margarida Rezende e Maria Perpétuo Messias, pelo seu constante e competente acompanhamento na Assessoria de Assuntos Interinstitucionais da UFV.

Finalmente, agradeço à CAPES pela bolsa que tornou possível a realização dessa tese.

Edilton de Souza Barcellos

Nascimento: 18 de março de 1948

Nacionalidade: brasileira

Naturalidade: Rio de Janeiro – RJ

Filiação: Jacy Cerqueira Barcellos e Maria de Souza Barcellos

Participação em Congressos:

Barcellos, E. de S.; dos Reis, M. M.; Ferreira, M. M. C.. Chemometric Study on Atmospheric Pollution Sources. Seattle, WA, The United States, 22-26/09/2002: Eighth International Conference on Chemometrics in Analytical Chemistry (CAC VIII, CAC-2002). Program and Abstracts Book, 89, **2002**

Barcellos, E. de S.; dos Reis, M. M.; Ferreira, M. M. C.. Uma Metodologia Para Identificação de Fontes de Poluição Atmosférica. 24ª Reunião Anual da Sociedade Brasileira de Química, Poços de Caldas, MG, **2001**.

Barcellos, E. de S.; Faria de, C. G.; Sales, H. J. S.; Ferreira, M. M. C.. Quimiometria Aplicada a Espectros em NIR Com Correção Multiplicativa de Espectros. 23ª Reunião Anual da Sociedade Brasileira de Química, Poços de Caldas, MG, **2000**.

Barcellos, E. de S.; dos Reis, M. M.; Ferreira, M. M. C.. Análise Multivariada: Instrumento Para Análise e Interpretação Ótima dos Resultados de Uma Pesquisa Científica. 21ª Reunião Anual da Sociedade Brasileira de Química, Poços de Caldas, MG, **1999**.

Barcellos, E. de S.; Ferreira, M. M. C.; Dias, C. A.; Oliveira de, D. M.. Construção de Modelos Para Enquadramento de Cursos de Água. 21ª Reunião Anual da Sociedade Brasileira de Química, Poços de Caldas, MG, **1998**.

Resumo

Essa tese tem por objetivo a identificação das principais fontes de poluição atmosférica na região metropolitana de São Paulo (RMSP) no período de 1996 a 1999. Os dados para a sua realização foram fornecidos pela Companhia de Saneamento Básico de São Paulo (CETESB) registrados pelas estações de medição de sua rede telemétrica. Esses dados constituem-se principalmente de espécies químicas emitidas por fontes automotivas e industriais. Para identificar tais fontes de emissão buscou-se encontrar perfis capazes de "denunciar" a presença constante de poluentes na atmosfera com certa regularidade de emissão. Com isto em mente procurou-se concentrar a atenção nas variações sistemáticas presentes nos dados, de modo que estas pudessem emergir. Assim a metodologia desenvolvida envolveu a construção de estruturas que pudessem ser submetidas a adequados métodos e modelos de análise multivariada criteriosamente selecionados. Tais arranjos, quando em dois modos (matrizes) submetidas à análise de componentes principais (PCA), forneceram as variações sistemáticas para os poluentes individualmente, com periodicidade diária, enquanto aquelas de ordem superior (três modos ou mais) revelaram as variações sistemáticas diárias e semanais. Foi possível identificar, através do modelo PARAFAC, os poluentes primários e os secundários, enquanto o modelo Tucker3 forneceu, através da combinação linear de fatores abstratos obtidos por orientação apropriada do seu núcleo, o perfil de acumulação de poluentes na baixa troposfera. Adicionalmente, alguns poluentes-chave foram acomodados três ciclos, com reflexos na orientação de vias preferenciais para os mecanismos, e na estimativa da quantidade de ozônio formado. O centro da tese reside, todavia, no desenvolvimento da metodologia, a qual foi capaz de orientar adequadamente a construção de estruturas e a seleção de modelos apropriados de ajuste aos dados, permitindo identificar as principais fontes de emissão de poluentes, emprestando visibilidade a todo um conjunto de reações químicas que ocorrem na atmosfera.

ABSTRACT

The aim of this thesis is the identification of the main sources of atmospheric pollution in the São Paulo metropolitan area since 1996 to 1999. The data used was kindly provided by the Sanitation Department of the State of São Paulo – CETESB. The data consists of concentration measurements made by the company's telemetric network. Based on the fact that the pollutants are emitted with regularity, and have a characteristic profile, we have searched for such systematic variations. Thus the methodology encompasses the construction of structures to be submitted to adequate and critically selected multivariate analysis methods and models. Such arrays, when in two modes, were submitted to a principal component analysis (PCA), giving the daily systematic variations for each pollutant individually, while the multimode arrays (three modes or more) showed the daily and weekly systematic variations. Primary and secondary pollutants were identified by PARAFAC model, while the Tucker model gave us the accumulation profiles of the pollutants in the low troposphere, through the linear combination of their abstract factors by adequate orientation of its core. Additionally, the key pollutants were accommodated into three cycles with repercussion in the orientation of the preferential ways for the mechanisms, and in the ozone quantity formed. The central edge of the thesis is, however, the development of the methodology, which was able to guide the construction of the structures and selection of adequate models to fit the data, and identify the sources of pollutants emission, giving light to set of chemical reactions occurring in the atmosphere.

SUMÁRIO

LISTA DE TABELAS	i
LISTA DE QUADROS	i
LISTA DE FIGURAS	ii
INTRODUÇÃO	1
CAPÍTULO I	
O HOMEM, A CIÊNCIA E O AMBIENTE	6
A Química Ambiental	9
CAPÍTULO II	
A ATMOSFERA E A QUÍMICA AMBIENTAL	16
A Atmosfera	16
Regiões da Atmosfera	17
A Química da Troposfera	22
Considerações Sobre a Química da Troposfera	23
Revisão da Reatividade Química na Troposfera	24
Oxidação do Metano na Troposfera	26
A Química da Baixa Troposfera e a Poluição do Ar	28
Origem e Ocorrência de “Smog”	29
Oxidação de Hidrocarbonetos no “Smog” Fotoquímico	31
Oxidação do SO ₂ na Atmosfera	36
A Poluição Atmosférica na RMSP	38
CAPÍTULO III	
METODOLOGIA	40
SEÇÃO III.1	
INTRODUÇÃO À ANÁLISE MULTIVARIADA	40
Notação e Terminologia	42
Análise Exploratória de Dados	42
A Organização dos Dados	43
A Matriz de Dados	44
Visualização dos Dados	44
Pretratamento de Dados I	44
Transformação nos Dados	45
Preprocessamento de Dados I	45
Análise de Componentes Principais (PCA)/Análise de Fatores (FA)	47
Análise Confirmatória de dados	54
SEÇÃO III.2	
BASE MATEMÁTICA EM ANÁLISE MULTIVARIADA	55
Base, Dimensão e Modo	55
Análise de Componentes Principais – PCA:	
Interpretações Filosófica, Algébrica e Geométrica	58
Componente Principal em Duas Dimensões:	
Cálculo por Meio do ângulo de Rotação	59

Distinção Entre PCA e FA	71
Métodos, Modelos, Algoritmos	73
Análise Multimodo	74
Uma Visão Geral	74
As Estruturas Multimodo ou de Ordem Superior	82
PCA: a identificação do método	82
SVD: origem e propriedades	84
A Estrutura multimodo	89
Modelos de Tucker, PARAFAC e PCA Multimodo (PCA–SUP)	91
Modelo PCA Multimodo (PCA–SUP)	94
Modelo de Tucker	97
Modelo PARAFAC	98
Estrutura e Hierarquia dos Modelos	99
Unicidade dos Modelos	104
SEÇÃO III.3	
AS BASES DA METODOLOGIA	108
A Construção da Metodologia	108
As Etapas da Construção da Metodologia	110
Pretratamento de Dados II	123
Os Dados I	124
Validação dos Dados	124
Dados Anômalos (“Outliers”)	125
Dados Faltantes (“Missing Data”)	130
Preprocessamento de Dados II	142
Restrições	142
Rotinas Desenvolvidas	143
Estrutura Computacional e Programas (“Softwares”)	147
Análises em Dois Modos (PCA)	147
Análises Multimodo (PARAFAC e Tucker)	147
PARAFAC	148
Tucker	149
CAPÍTULO IV	
ORIGEM DOS DADOS	153
A Origem dos Dados	153
As Estações de Medição	153
Os Dados II	155
Instrumentos de Medição	157
Estação de Trabalho	159
CAPÍTULO V	
A ANÁLISE DE DADOS	161
SEÇÃO V.1 – “Arquitetura” da Análise de Dados	161
Análise PCA (Pedro II – Nove Variáveis: 1999)	162
PCA (Pedro II – CO: 1999)	162
PCA (Pedro II – NO: 1999)	165
PCA (Pedro II – NO ₂ : 1999)	168
PCA (Pedro II – O ₃ : 1999)	171
PCA (Pedro II – SO ₂ : 1999; duas fontes)	174

PCA (Pedro II – SO ₂ : 1999; uma fonte)	177
PCA (Pedro II – CH ₄ : 1999)	180
PCA (Pedro II – NMHC: 1999)	183
PCA (Pedro II – Temp: 1999)	186
PCA (Pedro II – Umid: 1999)	189
Análise PCA (Monóxido de Carbono – Sete Estações: 1998)	193
PCA (CO – Pedro II: 1998)	193
PCA (CO – São Caetano do Sul: 1998)	197
PCA (CO – Ibirapuera: 1998)	200
PCA (CO – Cerqueira César: 1998)	203
PCA (CO – Centro: 1998)	206
PCA (CO – Lapa: 1998)	209
PCA (CO – Osasco: 1998)	212
Análise PCA (1999 – 1997)	215
PARAFAC (NO ₂ 99 – C. César)	216
PARAFAC (NO ₂ 98 – C. César)	217
PARAFAC (NO ₂ 97 – C. César)	218
PARAFAC (NO ₂ 99 – Centro)	219
PARAFAC (NO ₂ 97 – Centro)	220
PARAFAC (NO ₂ 99 – Lapa)	221
PARAFAC (NO ₂ 99 – Osasco)	222
PARAFAC (NO ₂ 99 – C. César)	223
PARAFAC (NO ₂ 98 – C. César)	224
PARAFAC (NO ₂ 99 – Centro)	225
PARAFAC (NO ₂ 99 – Lapa)	226
PARAFAC (NO ₂ 99 – Osasco)	227
PARAFAC (NO ₂ 98 – Osasco)	228
PARAFAC (O ₃ 97 – Lapa)	229
PARAFAC (O ₃ 97 – Osasco)	230
PARAFAC (SO ₂ 98 – Osasco)	231
PARAFAC (SO ₂ 99 – Centro)	232
CAPÍTULO VI	
RESULTADOS E DISCUSSÃO	235
SEÇÃO VI.1 – Principais Resultados e Discussão	235
SEÇÃO VI.2 – Resultados	236
SEÇÃO VI.3 – Análise em Dois Modos (PCA) e em Três Modos (PARAFAC)	237
PCA e PARAFAC: 1999	238
Análises PCA	238
Estação 1: P. D. Pedro II – 1999	241
Monóxido de Carbono – CO	241
Óxidos de Nitrogênio – NO e NO ₂	242
Ozônio – O ₃	244
Dióxido de Enxofre – SO ₂	245
Temperatura e Umidade – Temp. e Umid	251
Análises PARAFAC (P. D. Pedro II)	252
Modelo PARAFAC para o CO	252
Modelo PARAFAC para o NO	254

Modelo PARAFAC para o NO ₂	255
Modelo PARAFAC para o O ₃	256
Modelo PARAFAC para o SO ₂ (1 fonte)	257
Modelo PARAFAC para o SO ₂ (2 fontes)	258
Modelo PARAFAC para a Temp	261
Modelo PARAFAC para a Umid	262
Estação 2: São Caetano do Sul – 1999	263
Monóxido de Carbono – CO	263
Óxidos de Nitrogênio – NO e NO ₂	264
Ozônio – O ₃	264
Dióxido de Enxofre – SO ₂	265
Temperatura e Umidade – Temp. e Umid	271
Análises PARAFAC (São Caetano do Sul)	272
Modelo PARAFAC para o CO	272
Modelo PARAFAC para o NO	273
Modelo PARAFAC para o NO ₂	274
Modelo PARAFAC para o O ₃	275
Modelo PARAFAC para o SO ₂ (1 fonte)	277
Modelo PARAFAC para o SO ₂ (1 f, 2ª média)	278
Modelo PARAFAC para o SO ₂ (2 fontes)	279
Modelo PARAFAC para a Temp	282
Modelo PARAFAC para a Umid	283
Estação 3: Ibirapuera – 1999	284
Monóxido de Carbono – CO	284
Óxidos de Nitrogênio – NO e NO ₂	284
Ozônio – O ₃	285
Dióxido de Enxofre – SO ₂	285
Temperatura e Umidade – Temp. e Umid	296
Análises PARAFAC (Ibirapuera)	297
Modelo PARAFAC para o CO	297
Modelo PARAFAC para o NO	298
Modelo PARAFAC para o NO ₂	299
Modelo PARAFAC para o O ₃	300
Modelo PARAFAC para o SO ₂ (1 fonte)	301
Modelo PARAFAC para o SO ₂ (2 fontes)	302
Modelo PARAFAC para a Temp	305
Modelo PARAFAC para a Umid	306
Análise PCA para CH ₄ e NMHCs (Pedro II: 1999)	307
PCA – CH ₄	307
Análise PARAFAC para o CH ₄	308
Modelo para o CH ₄	308
PCA – NMHC	309
Análise PARAFAC para o NMHC	309
Modelo para o NMHC	309
Análise PCA para CH ₄ e NMHCs (S. Caetano: 1999)	311
PCA – CH ₄	311
Análise PARAFAC para o CH ₄	311

Modelo para o CH ₄	311
PCA – NMHC	312
Análise PARAFAC para o NMHC	313
Modelo para o NMHC	313
SEÇÃO VI.4 – Análises Multimodo	314
Análises PARAFAC (P. D. Pedro II – 1999)	314
PARAFAC: CO, NO, NO ₂ e O ₃	316
PARAFAC: CH ₄ , NO, NO ₂ e O ₃	317
PARAFAC: NMHC, NO, NO ₂ e O ₃ (1)	319
PARAFAC: NMHC, NO, NO ₂ e O ₃ (2)	320
PARAFAC: SO ₂ , NO, NO ₂ e O ₃	321
PARAFAC: CO, NO, NO ₂ , O ₃ e CH ₄	322
PARAFAC: CO, NO, NO ₂ , O ₃ e NMHC	324
PARAFAC: NMHC, NO, NO ₂ , O ₃ e CH ₄	325
PARAFAC: CO, NO, NO ₂ , O ₃ e SO ₂	326
PARAFAC: CH ₄ , NO, NO ₂ , O ₃ e SO ₂	327
PARAFAC: NMHC, NO, NO ₂ , O ₃ e SO ₂	328
PARAFAC: CO, NO, NO ₂ , O ₃ , SO ₂ e CH ₄	329
PARAFAC: CO, NO, NO ₂ , O ₃ , SO ₂ e NMHC	330
PARAFAC: CO, NO, NO ₂ , O ₃ , CH ₄ e NMHC	331
PARAFAC: SO ₂ , NO, NO ₂ , O ₃ , CH ₄ e NMHC	332
PARAFAC: CO, NO, NO ₂ , O ₃ , SO ₂ , CH ₄ e NMHC	333
Análises Tucker (P. D. Pedro II – 1999)	334
Tucker: CO, NO, NO ₂ e O ₃	334
Tucker: CH ₄ , NO, NO ₂ e O ₃	336
Tucker: NMHC, NO, NO ₂ e O ₃	338
Tucker: SO ₂ , NO, NO ₂ e O ₃	339
Tucker: CO, NO, NO ₂ , O ₃ e CH ₄	341
Tucker: CO, NO, NO ₂ , O ₃ e NMHC	343
Tucker: CH ₄ , NO, NO ₂ , O ₃ e NMHC	344
Tucker: CO, NO, NO ₂ , O ₃ e SO ₂	345
Tucker: CH ₄ , NO, NO ₂ , O ₃ e SO ₂	346
Tucker: NMHC, NO, NO ₂ , O ₃ e SO ₂	347
Tucker: CO, NO, NO ₂ , O ₃ , CH ₄ e SO ₂	348
Tucker: CO, NO, NO ₂ , O ₃ , SO ₂ e NMHC	349
Tucker: CO, NO, NO ₂ , O ₃ , CH ₄ e NMHC	350
Tucker: SO ₂ , NO, NO ₂ , O ₃ , CH ₄ e NMHC	351
Tucker: CO, NO, NO ₂ , O ₃ , CH ₄ , SO ₂ e NMHC	352
Arranjos do Núcleo (“core”) no Modelo de Tucker	354
CAPÍTULO VII – DISCUSSÃO	355
Seção VII.1 – Breve Retrospecto	355
Seção VII.2 Ciclo do Monóxido de Carbono	369
Seção VII.3 – Ciclo do Metano	385
Seção VII.4 – Ciclo dos Hidrocarbonetos Não Metano	397
Seção VII.5 – Conexão dos Ciclos	414
Seção VII.6 – Discussão Final	420

CONCLUSÕES	427
APÊNDICES	429
Apêndice A1	430
Apêndice A2	437
Apêndice A3	438
Apêndice A4	440
Apêndice A5	441
Apêndice A6	443
Apêndice A7	445
Apêndice A8	446
Apêndice A8.1	447
Apêndice A8.2	449
Apêndice A8.3	451
Apêndice A8.4	452
Apêndice A8.5	454
Apêndice A8.6	455
Apêndice A8.7	455
1º caso: ciclo básico do CO	456
2º caso: ciclo básico do CH ₄	457
3º caso: ciclo básico dos NMHCs	458
4º caso: substituição do CO ou CH ₄ pelo SO ₂ no ciclo	458
5º caso: ciclos básicos do CO e CH ₄	459
6º caso: ciclos básicos dos NMHCs e CO	459
7º caso: ciclos básicos dos NMHCs e CH ₄	460
8º caso: inclusão do SO ₂ no ciclo do CO	460
9º caso: inclusão de SO ₂ no ciclo do CH ₄	461
10º caso: inclusão do SO ₂ no ciclo dos NMHCs	461
11º caso: inclusão do SO ₂ nos ciclos do CO e CH ₄	462
12º caso: inclusão do SO ₂ nos ciclos do CO e NMHCs	462
13º caso: ciclos do CO, CH ₄ e NMHCs	463
14º caso: inclusão do SO ₂ nos ciclos do CH ₄ e NMHCs	463
15º caso: inclusão do SO ₂ nos ciclos do CO, CH ₄ e NMHCs	464
Apêndice B1	470
Apêndice B2	472
Apêndice C1	475
Apêndice C2	478
REFERÊNCIAS BIBLIOGRÁFICAS	480

Lista de Tabelas

Tabela II.1 – Fontes naturais de emissão de gases	22
Tabela IV.1 – Variáveis e suas respectivas unidades	158
Tabela VI.1 – Variância capturada pelas três primeiras PCs (SO ₂ 99 – Pedro II	259
Tabela VI.2 – Variância capturada pelas três primeiras PCs (SO ₂ 99 – S. Caetano	281
Tabela VII.3 – Variância capturada pelas três primeiras PCs (SO ₂ 99 – Ibirapuera	293

Lista de Quadros

Quadro VI.1 – Hierarquia de apresentação de resultados (3 Estações)	239
Quadro VI.2 – Hierarquia de apresentação de resultados (2 Estações)	240
Quadro VI.3 – Parâmetros de saída de software (PCA e PARAFAC)	253
Quadro VI.4 – Análises multimodo (PARAFAC e Tucker)	315

Lista de Figuras

Figura 1.1 – Ambientes atmosférico, aquático e terrestre (seres vivos e energia)	7
Figura 2.1 – O ciclo hidrológico	17
Figura 2.2 – Principais regiões da atmosfera	19
Figura 2.3 – O espectro eletromagnético	21
Figura 3.1 – A matriz de dados	44
Figura 3.2 – Comparação usual entre PCA e FA	48
Figura 3.3 – Projeção de pontos nas bases original e rotacionada	51
Figura 3.4 – Eliminação do ruído experimental no modelo PC	53
Figura 3.5 – Variabilidade de dados e colinearidade entre variáveis	59
Figura 3.6 – Variabilidade de dados na direção da componente principal	60
Figura 3.7 – Centróide de um conjunto de dados	61
Figura 3.8 – Obtenção das direções dos novos eixos	63
Figura 3.9 – Coordenadas de um ponto nos sistemas original e rotacionado	67
Figura 3.10 – Estrutura dos dados em análise multimodo	80
Figura 3.11 – Esquema de uma decomposição SVD	88
Figura 3.12 – Desdobramento (“unfolding”) de um cubo de dados	90
Figura 3.13 – Conjuntos múltiplos de dados (multiple data sets)	94
Figura 3.14 – Estrutura de dados em PCA–SUP (“vetorização”)	96
Figura 3.15 – Estrutura de dados em PCA–SUP (obj. $\times$ var. $\times$ loc.)	96
Figura 3.16 – Esquema de uma decomposição para o modelo de Tucker	102
Figura 3.17 – Esquema de uma decomposição para o modelo PARAFAC	103
Figura 3.18 – Um vetor de dados (1×365)	111
Figura 3.19 – Decomposição da matriz de dados	114
Figura 3.20 – Representação esquemática de um vetor de dados	115
Figura 3.21 – Construção de uma matriz 365 (dias) $\times$ 24 (horas)	116
Figura 3.22 – Esquema das matrizes de dados e de “scores” e de “loadings”	117
Figura 3.23 – Construção de uma estrutura multimodo a partir de matrizes	119
Figura 3.24 – Cubo de dados e seus modos A, B e C	119
Figura 3.25 – Dados Anômalos (CO99 – Ibirapuera)	126

Figura 3.26 – Correção em Dados Anômalos (CO99 – Ibirapuera)	127
Figura 3.27 – Correção em Variações Súbitas (CO99 – Ibirapuera)	128
Figura 3.28 – Saída Típica PCA	128
Figura 3.29 – Dados Estimados (CO99 – Ibirapuera)	129
Figura 3.30 – Resíduos (CO99 – Ibirapuera)	129
Figura 3.31 – “Lacunaridade” (Temp99 – Ibirapuera)	130
Figura 3.32 – Dados Estimados e Resíduos (Temp99 – Ibirapuera)	132
Figura 3.33 – Dados Estimados e Resíduos (Temp99 – Ibirapuera: 1ª PC)	133
Figura 3.34 – Pesos (Temp99 – Ibirapuera: PC ₁ , PC ₂ e PC ₃)	135
Figura 3.35 – Pesos (Temp99 – Ibirapuera: PC ₁)	135
Figura 3.36 – Escores (Temp99 – Ibirapuera: PC ₁)	136
Figura 3.37 – Influência Dados Faltantes (Temp99 – Ibirapuera)	136
Figura 3.38 – PARAFAC (Temp99 – Ibirapuera: com “lacunaridade”)	137
Figura 3.39 – PARAFAC (Temp99 – Ibirapuera: sem “lacunaridade”)	137
Figura 3.40 – Teste Influência Dados Faltantes (CO e NO – Pedro II: 1999)	139
Figura 3.41 – Teste Influência Dados Faltantes (O ₃ e NO ₂ – Pedro II: 1999)	140
Figura 3.42 – Teste Influência Dados Faltantes (SO ₂ 1 e 2 fontes – Pedro II: 1999)	140
Figura 3.43 – Teste Influência Dados Faltantes (CH ₄ , NMHC, Temp e Umid – Pedro II: 1999)	141
Figura 3.44 – Influência “lacunaridade” em Dados Faltantes (Temp99 – Ibirapuera)	144
Figura 3.45 – Influência “lacunaridade” nos Escores (Temp99 – Ibirapuera)	145
Figura 3.44 – Influência “lacunaridade” em Dados Estimados (Temp99 – Ibirapuera)	145
Figura 4.1 – Estação de Medição da rede telemétrica da CETESB na RMSP	154
Figura 4.2 – Matrizes com grande quantidade de dados faltantes 1	156
Figura 4.3 – Matrizes com grande quantidade de dados faltantes 2	157
Figura 5.1 – Dados originais e sua transposta (CO99 – Pedro II)	162
Figura 5.2 – Pesos (“loadings”) para PC ₁ , PC ₂ e PC ₃ (CO99 – Pedro II)	163
Figura 5.3 – Escores (“scores”) para PC ₁ (CO99 – Pedro II)	163
Figura 5.4 – Dados estimados e resíduos: posto completo (CO99 – Pedro II)	164
Figura 5.5 – Dados estimados e resíduos: três PCs (CO99 – Pedro II)	164
Figura 5.6 – Dados originais e sua transposta (NO99 – Pedro II)	165

Figura 5.7 – Dados originais corrigidos e sua transposta (NO99 – Pedro II)	165
Figura 5.8 – Pesos (“loadings”) para PC ₁ , PC ₂ e PC ₃ (NO99 – Pedro II)	166
Figura 5.9 – Escores (“scores”) para PC ₁ (NO99 – Pedro II)	166
Figura 5.10 – Dados estimados e resíduos: posto completo (NO99 – Pedro II)	167
Figura 5.11 – Dados estimados e resíduos: três PCs (NO99 – Pedro II)	167
Figura 5.12 – Dados originais e sua transposta (NO ₂ 99 – Pedro II)	168
Figura 5.13 – Pesos (“loadings”) para PC ₁ , PC ₂ e PC ₃ (NO ₂ 99 – Pedro II)	168
Figura 5.14 – Escores (“scores”) para PC ₁ (NO ₂ 99 – Pedro II)	169
Figura 5.15 – Dados estimados e resíduos: posto completo (NO ₂ 99 – Pedro II)	169
Figura 5.16 – Dados estimados e resíduos: três PCs (NO ₂ 99 – Pedro II)	170
Figura 5.17 – Dados originais e sua transposta (O ₃ 99 – Pedro II)	171
Figura 5.18 – Dados originais corrigidos e sua transposta (O ₃ 99 – Pedro II)	171
Figura 5.19 – Pesos (“loadings”) para PC ₁ , PC ₂ e PC ₃ (O ₃ 99 – Pedro II)	172
Figura 5.20 – Escores (“scores”) para PC ₁ (O ₃ 99 – Pedro II)	172
Figura 5.21 – Dados estimados e resíduos: posto completo (O ₃ 99 – Pedro II)	173
Figura 5.22 – Dados estimados e resíduos: três PCs (O ₃ 99 – Pedro II)	173
Figura 5.23 – Dados originais e sua transposta (SO ₂ 99 – Pedro II: 2 fontes)	174
Figura 5.24 – Dados orig. corrigidos e sua transposta (SO ₂ 99 – Pedro II: 2 fontes)	174
Figura 5.25 – Pesos (“loadings”) para PC ₁ , PC ₂ e PC ₃ (SO ₂ 99 – Pedro II: 2 fontes)	175
Figura 5.26 – Escores (“scores”) para PC ₁ (SO ₂ 99 – Pedro II: 2 fontes)	175
Figura 5.27 – Dados estim. e resíduos: posto completo (SO ₂ 99 – Pedro II: 2 fontes)	176
Figura 5.28 – Dados estimados e resíduos: três PCs (SO ₂ 99 – Pedro II: 2 fontes)	176
Figura 5.29 – Dados originais e sua transposta (SO ₂ 99 – Pedro II: 1 fonte)	177
Figura 5.30 – Dados orig. corrigidos e sua transposta (SO ₂ 99 – Pedro II: 1 fonte)	177
Figura 5.31 – Pesos (“loadings”) para PC ₁ , PC ₂ e PC ₃ (SO ₂ 99 – Pedro II: 1 fonte)	178
Figura 5.32 – Escores (“scores”) para PC ₁ (SO ₂ 99 – Pedro II: 1 fonte)	178
Figura 5.33 – Dados estim. e resíduos: posto completo (SO ₂ 99 – Pedro II: 1 fonte)	179
Figura 5.34 – Dados estimados e resíduos: três PCs (SO ₂ 99 – Pedro II: 1 fonte)	179
Figura 5.35 – Dados originais e sua transposta (CH ₄ 99 – Pedro II)	180
Figura 5.36 – Pesos (“loadings”) para PC ₁ , PC ₂ e PC ₃ (CH ₄ 99 – Pedro II)	180
Figura 5.37 – Escores (“scores”) para PC ₁ (CH ₄ 99 – Pedro II)	181

Figura 5.38 – Dados estimados e resíduos: posto completo (CH ₄ 99 – Pedro II)	181
Figura 5.39 – Dados estimados e resíduos: três PCs (CH ₄ 99 – Pedro II)	182
Figura 5.40 – Dados originais e sua transposta (NMHC99 – Pedro II)	183
Figura 5.41 – Dados orig. corrigidos e sua transposta (NMHC99 – Pedro II)	183
Figura 5.42 – Pesos (“loadings”) para PC ₁ , PC ₂ e PC ₃ (NMHC99 – Pedro II)	184
Figura 5.43 – Escores (“scores”) para PC ₁ (NMHC99 – Pedro II)	184
Figura 5.44 – Dados estimados e resíduos: posto completo (NMHC99 – Pedro II)	185
Figura 5.45 – Dados estimados e resíduos: três PCs (NMHC99 – Pedro II)	185
Figura 5.46 – Dados originais e sua transposta (Temp99 – Pedro II)	186
Figura 5.47 – Dados orig. corrigidos e sua transposta (Temp99 – Pedro II)	186
Figura 5.48 – Pesos (“loadings”) para PC ₁ , PC ₂ e PC ₃ (Temp99 – Pedro II)	187
Figura 5.40 – Escores (“scores”) para PC ₁ (Temp99 – Pedro II)	187
Figura 5.50 – Dados estimados e resíduos: posto completo (Temp99 – Pedro II)	188
Figura 5.51 – Dados estimados e resíduos: três PCs (Temp99 – Pedro II)	188
Figura 5.52 – Dados originais e sua transposta (Umid99 – Pedro II)	189
Figura 5.53 – Dados orig. corrigidos e sua transposta (Umid99 – Pedro II)	189
Figura 5.54 – Pesos (“loadings”) para PC ₁ , PC ₂ e PC ₃ (Umid99 – Pedro II)	190
Figura 5.55 – Escores (“scores”) para PC ₁ (Umid99 – Pedro II)	190
Figura 5.56 – Dados estimados e resíduos: posto completo (Umid99 – Pedro II)	191
Figura 5.57 – Dados estimados e resíduos: três PCs (Umid99 – Pedro II)	191
Figura 5.58 – Dados originais e sua transposta (CO98 – Pedro II)	193
Figura 5.59 – Dados orig. corrigidos e sua transposta (CO98 – Pedro II)	194
Figura 5.60 – Pesos (“loadings”) para PC ₁ , PC ₂ e PC ₃ (CO98 – Pedro II)	194
Figura 5.61 – Escores (“scores”) para PC ₁ (CO98 – Pedro II)	195
Figura 5.62 – Dados estimados e resíduos: posto completo (CO98 – Pedro II)	195
Figura 5.63 – Dados estimados e resíduos: três PCs (CO98 – Pedro II)	196
Figura 5.64 – Dados originais e sua transposta (CO98 – São Caetano)	197
Figura 5.65 – Dados orig. corrigidos e sua transposta (CO98 – São Caetano)	197
Figura 5.66 – Pesos (“loadings”) para PC ₁ , PC ₂ e PC ₃ (CO98 – São Caetano)	198
Figura 5.67 – Escores (“scores”) para PC ₁ (CO98 – São Caetano)	198
Figura 5.68 – Dados estimados e resíduos: posto completo (CO98 – São Caetano)	199

Figura 5.69 – Dados estimados e resíduos: três PCs (CO98 – São Caetano)	199
Figura 5.70 – Dados originais e sua transposta (CO98 – Ibirapuera)	200
Figura 5.71 – Dados orig. corrigidos e sua transposta (CO98 – Ibirapuera)	200
Figura 5.72 – Pesos (“loadings”) para PC ₁ , PC ₂ e PC ₃ (CO98 – Ibirapuera)	201
Figura 5.73 – Escores (“scores”) para PC ₁ (CO98 – Ibirapuera)	201
Figura 5.74 – Dados estimados e resíduos: posto completo (CO98 – Ibirapuera)	202
Figura 5.75 – Dados estimados e resíduos: três PCs (CO98 – Ibirapuera)	202
Figura 5.76 – Dados originais e sua transposta (CO98 – C. César)	203
Figura 5.77 – Dados orig. corrigidos e sua transposta (CO98 – C. César)	203
Figura 5.78 – Pesos (“loadings”) para PC ₁ , PC ₂ e PC ₃ (CO98 – C. César)	204
Figura 5.79 – Escores (“scores”) para PC ₁ (CO98 – C. César)	204
Figura 5.80 – Dados estimados e resíduos: posto completo (CO98 – C. César)	205
Figura 5.81 – Dados estimados e resíduos: três PCs (CO98 – C. César)	205
Figura 5.82 – Dados originais e sua transposta (CO98 – Centro)	206
Figura 5.83 – Dados originais corrigidos e sua transposta (CO98 – Centro)	206
Figura 5.84 – Pesos (“loadings”) para PC ₁ , PC ₂ e PC ₃ (CO98 – Centro)	207
Figura 5.85 – Escores (“scores”) para PC ₁ (CO98 – Centro)	207
Figura 5.86 – Dados estimados e resíduos: posto completo (CO98 – Centro)	208
Figura 5.87 – Dados estimados e resíduos: três PCs (CO98 – Centro)	208
Figura 5.88 – Dados originais e sua transposta (CO99 – Lapa)	209
Figura 5.89 – Dados orig. corrigidos e sua transposta (CO99 – Lapa)	209
Figura 5.90 – Pesos (“loadings”) para PC ₁ , PC ₂ e PC ₃ (CO99 – Lapa)	210
Figura 5.91 – Escores (“scores”) para PC ₁ (CO99 – Lapa)	210
Figura 5.92 – Dados estimados e resíduos: posto completo (CO99 – Lapa)	211
Figura 5.93 – Dados estimados e resíduos: três PCs (CO99 – Lapa)	211
Figura 5.94 – Dados originais e sua transposta (CO99 – Osasco)	212
Figura 5.95 – Pesos (“loadings”) para PC ₁ , PC ₂ e PC ₃ (CO99 – Osasco)	212
Figura 5.96 – Escores (“scores”) para PC ₁ (CO99 – Osasco)	213
Figura 5.97 – Dados estimados e resíduos: posto completo (CO99 – Osasco)	213
Figura 5.98 – Dados estimados e resíduos: três PCs (CO99 – Osasco)	214
Figura 5.99 - % de Variância capturada por PC ₁ , PC ₂ e PC ₃ (7 estações: 1998)	214

Figura 5.100 – VSD, VSS e VA (NO ₂ 99 – C. César)	216
Figura 5.101 – VSD, VSS e VA (NO ₂ 98 – C. César)	217
Figura 5.102 – VSD, VSS e VA (NO ₂ 97 – C. César)	218
Figura 5.103 – VSD, VSS e VA (NO ₂ 99 – Centro)	219
Figura 5.104 – VSD, VSS e VA (NO ₂ 97 – Centro)	220
Figura 5.105 – VSD, VSS e VA (NO ₂ 99 – Lapa)	221
Figura 5.106 – VSD, VSS e VA (NO ₂ 99 – Osasco)	222
Figura 5.107 – VSD, VSS e VA (NO99 – C. César)	223
Figura 5.108 – VSD, VSS e VA (NO98 – C. César)	224
Figura 5.109 – VSD, VSS e VA (NO99 – Centro)	225
Figura 5.110 – VSD, VSS e VA (NO99 – Lapa)	226
Figura 5.111 – VSD, VSS e VA (NO99 – Osasco)	227
Figura 5.112 – VSD, VSS e VA (NO98 – Osasco)	228
Figura 5.113 – VSD, VSS e VA (O ₃ 97 – Lapa)	229
Figura 5.114 – VSD, VSS e VA (O ₃ 97 – Osasco)	230
Figura 5.115 – VSD, VSS e VA (SO ₂ 98 – Osasco)	231
Figura 5.116 – VSD, VSS e VA (SO ₂ 98 – Centro)	232
Figura 6.1 – Pesos (“loadings”) para PC1 (CO: Pedro II – 1999)	241
Figura 6.2 – Pesos (“loadings”) para PC1 (NO e NO ₂ : Pedro II – 1999)	242
Figura 6.3 – Pesos (“loadings”) para PC1 (O ₃ : Pedro II – 1999)	244
Figura 6.4 – Pesos (“loadings”) para PC1 (SO ₂ : Pedro II – 1999)	245
Figura 6.4.1 – Dados originais e dados corrigidos e suas transpostas (SO ₂ 99: Pedro II – período afetado)	247
Figura 6.4.2 – “Zoom” nas Figs.6.4.1c e d (incluído o período afetado)	248
Figura 6.4.3 - “Zoom” nas Figs.6.4.1c e d (excluído o período afetado)	249
Figura 6.5 – Pesos (“loadings”) para PC1 (Temp e Umid: Pedro II – 1999)	251
Figura 6.6 – Modelo PARAFAC para o CO (Pedro II: 1999)	252
Figura 6.7 – Modelo PARAFAC para o NO (Pedro II: 1999)	254
Figura 6.8 – Modelo PARAFAC para o NO ₂ (Pedro II: 1999)	255
Figura 6.9 – Modelo PARAFAC para o O ₃ (Pedro II: 1999)	256
Figura 6.10.1 – Modelo PARAFAC para o SO ₂ (1 fonte: Pedro II – 1999)	257

Figura 6.10.2 – Modelo PARAFAC para o SO ₂ (2 fontes: Pedro II – 1999)	258
Figura 6.11 – Modelo PARAFAC para a Temp (Pedro II: 1999)	261
Figura 6.12 – Modelo PARAFAC para a Umid (Pedro II: 1999)	262
Figura 6.13 – Pesos (“loadings”) para PC1 (CO: S. Caetano – 1999)	263
Figura 6.14 – Pesos (“loadings”) para PC1 (NO e NO ₂ : S. Caetano – 1999)	264
Figura 6.15 – Pesos (“loadings”) para PC1 (O ₃ : S. Caetano – 1999)	264
Figura 6.16 – Pesos para (“loadings”) PC1 (SO ₂ : S. Caetano – 1999)	266
Figura 6.16.1 - Dados originais e dados originais corrigidos e suas transpostas (SO ₂ 99: São Caetano – período afetado)	267
Figura 6.16.2 – “Zoom” nas Figs.6.16.1c e d (incluído o período afetado)	268
Figura 6.16.3 - “Zoom” nas Figs.6.16.1c e d (excluído o período afetado)	269
Figura 6.17 – Pesos (“loadings”) para PC1 (Temp e Umid: S. Caetano – 1999)	271
Figura 6.18 – Modelo PARAFAC para o CO (S. Caetano: 1999)	272
Figura 6.19 – Modelo PARAFAC para o NO (S. Caetano: 1999)	273
Figura 6.20 – Modelo PARAFAC para o NO ₂ (S. Caetano: 1999)	274
Figura 6.21 – Modelo PARAFAC para o O ₃ (S. Caetano: 1999)	275
Figura 6.22.1 – Modelo PARAFAC para o SO ₂ (1 fonte: S. Caetano – 1999)	277
Figura 6.22.2 – Modelo PARAFAC para o SO ₂ (1 fonte, 2ª média: S. Caet. – 1999)	278
Figura 6.22.3 – Modelo PARAFAC para o SO ₂ (2 fontes: S. Caetano – 1999)	279
Figura 6.23 – Modelo PARAFAC para a Temp (S. Caetano: 1999)	282
Figura 6.24 – Modelo PARAFAC para a Umid (S. Caetano: 1999)	283
Figura 6.25 – Pesos (“loadings”) para PC1 (CO: Ibirapuera – 1999)	284
Figura 6.26 – Pesos (“loadings”) para PC1 (NO e NO ₂ : Ibirapuera – 1999)	284
Figura 6.27 – Pesos (“loadings”) para PC1 (O ₃ : Ibirapuera – 1999)	285
Figura 6.28.1 – Pesos (“loadings”) para PC1 (SO ₂ : Ibirapuera – 1999)	286
Figura 6.28.2 – Pesos (“loadings”) para PC1 e PC2 (SO ₂ : Ibirapuera – 1999)	287
Figura 6.28.3 – Dados originais e corrigidos (SO ₂ : Ibirapuera – 1999)	289
Figura 6.28.4 – “Zoom” para dados originais e corrigidos (SO ₂ : Ibirapuera – 1999)	290
Figura 6.28.5 – Escores (“scores”) para dados originais corrigidos (SO ₂ : Ibirapuera – 1999)	291
Figura 6.28.6 - Dados originais e dados originais corrigidos e suas transpostas	

(SO ₂ 99: Ibirapuera – período afetado)	292
Figura 6.28.7 – Escores (“scores”) para PC ₁ (incluído o período afetado)	295
Figura 6.28.8 - Escores (“scores”) para PC ₁ (excluído o período afetado)	295
Figura 6.29 – Pesos (“loadings”) para PC1 (Temp e Umid: Ibirapuera – 1999)	296
Figura 6.30 – Modelo PARAFAC para o CO (Ibirapuera: 1999)	297
Figura 6.31 – Modelo PARAFAC para o NO (Ibirapuera: 1999)	298
Figura 6.32 – Modelo PARAFAC para o NO ₂ (Ibirapuera: 1999)	299
Figura 6.33 – Modelo PARAFAC para o O ₃ (Ibirapuera: 1999)	300
Figura 6.34.1 – Modelo PARAFAC para o SO ₂ (1 fonte: Ibirapuera – 1999)	301
Figura 6.34.2 – Modelo PARAFAC para o SO ₂ (2 fontes: Ibirapuera – 1999)	302
Figura 6.34.3 – Modelo PARAFAC para o SO ₂ (período afetado: PC ₁ e PC ₂)	303
Figura 6.35 – Modelo PARAFAC para a Temp (Ibirapuera: 1999)	305
Figura 6.36 – Modelo PARAFAC para a Umid (Ibirapuera: 1999)	306
Figura 6.37 – Pesos (“loadings”) para PC1 (CH ₄ : Pedro II – 1999)	307
Figura 6.38 – Modelo PARAFAC para o CH ₄ (Pedro II: 1999)	308
Figura 6.39 – Pesos (“loadings”) para PC1 (NMHC: Pedro II – 1999)	309
Figura 6.40 – Modelo PARAFAC para o NMHC (Pedro II: 1999)	309-310
Figura 6.41 – Pesos (“loadings”) para PC1 (CH ₄ : São Caetano – 1999)	311
Figura 6.42 – Modelo PARAFAC para o CH ₄ (São Caetano: 1999)	311-312
Figura 6.43 – Pesos (“loadings”) para PC1 (NMHC: São Caetano – 1999)	312
Figura 6.44 – Modelo PARAFAC para o NMHC (São Caetano: 1999)	313
Figura 6.45 – PARAFAC (CO, NO, NO ₂ , O ₃ : Pedro II – 1999)	316
Figura 6.46 – PARAFAC (CH ₄ , NO, NO ₂ , O ₃ : Pedro II – 1999)	317
Figura 6.47.1 – PARAFAC 1 (NMHC, NO, NO ₂ , O ₃ : Pedro II – 1999)	319
Figura 6.47.2 – PARAFAC 2 (NMHC, NO, NO ₂ , O ₃ : Pedro II – 1999)	320
Figura 6.48 – PARAFAC (SO ₂ , NO, NO ₂ , O ₃ : Pedro II – 1999)	321
Figura 6.49 – PARAFAC (CO, NO, NO ₂ , O ₃ , CH ₄ : Pedro II – 1999)	322
Figura 6.50 – PARAFAC (CO, NO, NO ₂ , O ₃ , NMHC: Pedro II – 1999)	324
Figura 6.51 – PARAFAC (NMHC, NO, NO ₂ , O ₃ , CH ₄ : Pedro II – 1999)	325
Figura 6.52 – PARAFAC (CO, NO, NO ₂ , O ₃ , SO ₂ : Pedro II – 1999)	326

Figura 6.53 – PARAFAC (CH ₄ , NO, NO ₂ , O ₃ , SO ₂ : Pedro II – 1999)	327
Figura 6.54 – PARAFAC (NMHC, NO, NO ₂ , O ₃ , SO ₂ : Pedro II – 1999)	328
Figura 6.55 – PARAFAC (CO, NO, NO ₂ , O ₃ , SO ₂ , CH ₄ : Pedro II – 1999)	329
Figura 6.56 – PARAFAC (CO, NO, NO ₂ , O ₃ , SO ₂ , NMHC: Pedro II – 1999)	330
Figura 6.57 – PARAFAC (CO, NO, NO ₂ , O ₃ , CH ₄ , NMHC: Pedro II – 1999)	331
Figura 6.58 – PARAFAC (SO ₂ , NO, NO ₂ , O ₃ , SO ₂ , CH ₄ , NMHC: Pedro II – 1999)	332
Figura 6.59 – PARAFAC (CO, NO, NO ₂ , O ₃ , SO ₂ , CH ₄ , NMHC: Pedro II – 1999)	333
Figura 6.60 – Tucker (CO, NO, NO ₂ , O ₃ : Pedro II – 1999)	334
Figura 6.61 – Tucker (CH ₄ , NO, NO ₂ , O ₃ : Pedro II – 1999)	336
Figura 6.62 – Tucker (NMHC, NO, NO ₂ , O ₃ : Pedro II – 1999)	338
Figura 6.63 – Tucker (SO ₂ , NO, NO ₂ , O ₃ : Pedro II – 1999)	339
Figura 6.64 – Tucker (CO, NO, NO ₂ , O ₃ , CH ₄ : Pedro II – 1999)	341
Figura 6.65 – Tucker (CO, NO, NO ₂ , O ₃ , NMHC: Pedro II – 1999)	343
Figura 6.66 – Tucker (CH ₄ , NO, NO ₂ , O ₃ , NMHC: Pedro II – 1999)	344
Figura 6.67 – Tucker (CO, NO, NO ₂ , O ₃ , SO ₂ : Pedro II – 1999)	345
Figura 6.68 – Tucker (CH ₄ , NO, NO ₂ , O ₃ , SO ₂ : Pedro II – 1999)	346
Figura 6.69 – Tucker (NMHC, NO, NO ₂ , O ₃ , SO ₂ : Pedro II – 1999)	347
Figura 6.70 – Tucker (CO, NO, NO ₂ , O ₃ , CH ₄ , SO ₂ : Pedro II – 1999)	348
Figura 6.71 – Tucker (CO, NO, NO ₂ , O ₃ , SO ₂ , NMHC: Pedro II – 1999)	349
Figura 6.72 – Tucker (CO, NO, NO ₂ , O ₃ , CH ₄ , NMHC: Pedro II – 1999)	350
Figura 6.73 – Tucker (SO ₂ , NO, NO ₂ , O ₃ , CH ₄ , NMHC: Pedro II – 1999)	351
Figura 6.74 – Tucker (CO, NO, NO ₂ , O ₃ , CH ₄ , SO ₂ , NMHC: Pedro II – 1999)	352
Figura 7.1 – Densidade de tráfego (RMSP – 1999)	356
Figura 7.2 – Dados originais para o SO ₂ (Pedro II, S. Caet. e Ibirapuera (1999)	357
Figura 6.77 – Dados originais para o SO ₂ (Pedro II – 1999)	359
Figura 6.78 – Dados originais para o SO ₂ (S. Caetano – 1999)	359
Figura 6.79 – Dados originais para o SO ₂ (Ibirapuera – 1999)	360
Figura 6.80 – Frequência de precipitações pluviométricas (RMSP – 1996 a 1999)	361
Figura 6.81 – Fre. Inv. térm. e de condições desfavoráveis à dispersão (RMSP – 1999)	361
Figura 6.82 – Ciclo do monóxido de carbono	369
Figura 6.83 – Variação anual (CO, NO, NO ₂ , O ₃ : Pedro II – 1999)	370

Figura 6.84 – Perfis de acumulação (CO, NO, NO ₂ , O ₃ : Pedro II – 1999)	376
Figura 6.85 – Possíveis vias para destruição de O ₃ no ciclo do metano	386
Figura 6.86 – Mecanismo 1: decomposição aldeídos na atmosfera (baixa conc. NO _x)	403
Figura 6.87 – Mecanismo 2: decomposição aldeídos na atmosfera (baixa conc. NO _x)	406
Figura 6.88 – Mecanismo: decomposição aldeídos na atmosfera (alta conc. NO _x)	409
Figura 6.89 – Perfis: acumulação de poluentes (conexão dos ciclos)	415-416

INTRODUÇÃO

A rápida elevação, nas últimas décadas, dos valores de concentração de certas substâncias na atmosfera, tem sido motivo de preocupação em todo o mundo. Isto vem ocorrendo como consequência direta de atividades desenvolvidas pela espécie humana. Tais compostos têm um efeito deveras adverso principalmente em cidades com grande contingente populacional, nas quais a densidade de tráfego por veículos automotivos e a concentração de atividades industriais de vários tipos traduzem-se na emissão de compostos em grande escala, direta ou indiretamente, nocivos à vida e ao ambiente em geral.

Alguns desses compostos desempenham um papel fundamental nos mecanismos das reações que ocorrem na baixa troposfera. Independentemente do ciclo natural de formação e consumo de alguns deles no ambiente, a emissão adicional em grande quantidade de certos compostos chave nesse processo, confere-lhes o “status” de poluentes, seja pela elevada concentração alcançada, seja pela formação de espécies estranhas ao ambiente natural, também em concentrações altas o suficiente para afetar de algum modo o ambiente, ou ainda pelo aumento de complexidade nos mecanismos de reações de que participam.

Dentre os compostos que apresentam maior importância na química da baixa troposfera incluem-se o monóxido de carbono, CO, os óxidos de nitrogênio, NO e NO₂ (e a soma deles, NO_x, ou ainda incluindo seus derivados, NO_y), o metano, CH₄, outros hidrocarbonetos não metano, o ozônio, O₃, e o dióxido de enxofre, SO₂. Compostos que são emitidos diretamente para a atmosfera como o CO e o NO, por exemplo, são classificados como poluentes primários, e aqueles, como o O₃, que se formam por meio de reações químicas que ocorrem na baixa troposfera a partir dos primeiros, são os poluentes secundários. Somem-se ainda as poeiras, fuligem etc, classificados como partículas onde, a partícula inalável, PI, de tamanho inferior a 10 µm (PM10) capaz de penetrar nas vias

respiratórias, adquire importância fundamental. Adicionalmente, a evaporação de solventes, combustíveis líquidos e outros compostos orgânicos (pertencentes à classe dos compostos orgânicos voláteis, COVs) também é responsável pelo aumento da concentração de poluentes na atmosfera.

Algumas pesquisas avaliam as correlações entre os níveis desses compostos por meio de análise univariada, enquanto outros autores têm usado métodos quimiométricos. Estes se apresentam como mais adequados ao estudo de problemas que envolvem grande quantidade de dados. Nesse sentido, importantes resultados na identificação de fontes foram conseguidos com a aplicação de TTFA (Target Transformation Factor Analysis). Modelos receptores, como o CMB (Chemical Mass Balance) desenvolvidos no final dos anos 1960, também têm sido utilizados no estudo de problemas de resolução de misturas aplicadas à distribuição de fontes de partículas emitidas ao ar. Outros pesquisadores usaram FA (Análise de Fatores) com restrições (não negatividade), ajustando o modelo por quadrados mínimos alternantes (alternating least squares – ALS) para identificar fontes na costa da França e da Espanha.

Não obstante a ocorrência de estudos constatando a existência de correlação entre algumas das substâncias-chave nos mecanismos envolvendo a química da atmosfera, e mesmo a aplicação de análise multivariada de dados aos problemas ambientais, no que diz respeito ao estudo de poluição atmosférica e às metodologias aplicadas, muitas lacunas ainda devem ser exploradas.

O objetivo dessa tese é o de introduzir uma metodologia capaz de identificar sem ambigüidades (do ponto de vista algébrico), a principal (ou principais) fonte(s) de poluição atmosférica ao nível da baixa troposfera, por meio da aplicação de quimiometria a dados arranjados no modo bidimensional convencional (matrizes) e a dados em estruturas *multimodo*, alternativamente denominados *dados de ordem superior*.

Os dados analisados nessa tese são oriundos do banco de dados da CETESB (Companhia de Tecnologia de Saneamento Ambiental – Estado de São Paulo). Os arquivos contendo os dados possuem tamanho considerável, e o formato dos mesmos exigiu a construção de pequenos trechos de programas de modo a colocá-los em condições de análise. A natureza das análises e a interpretação dos resultados delas decorrente, implicou em uma estratégia de apresentação que propiciasse uma leitura isenta da necessidade de

consulta a textos básicos para suporte teórico aos capítulos que tratam do assunto sob o ponto de vista químico e matemático.

Assim, pretendeu-se dar ao texto, em alguns capítulos, a forma de tutorial, o qual permite uma interação mais constante entre as partes analíticas de um texto e os argumentos teóricos que lhes oferecem suporte, de modo que a leitura pudesse fluir mais naturalmente. Se isto facilita o processo de leitura em si mesmo, também torna o texto mais extenso, em função das informações adicionais que deve conter. Cabe acrescentar ainda que a tese tem como objetivo principal, o uso de quimiometria para o estabelecimento de uma *“metodologia para a identificação de fontes de poluição atmosférica”*, não obstante os resultados e conclusões para além dele. Do mesmo modo, a necessidade de enfatizar esse objetivo exigiu que ele fosse lembrado várias vezes no texto de formas diferentes, seja por meio de recorrências textuais, seja por meio de expressões algébricas, seja por meio de esquemas e/ou figuras. Com essa estratégia em mente, o texto foi distribuído em sete capítulos, acrescido das conclusões, conforme se expõe em seguida:

Capítulo I: Trata de forma introdutória e geral o problema da poluição. Destaca o papel das diversas disciplinas na questão ambiental, chama a atenção para a participação dos profissionais da área de química no estudo e soluções para os problemas ambientais, e enfatiza a necessidade de uma abordagem política conseqüente sobre os mesmos.

Capítulo II: Trata a questão específica da poluição atmosférica, discorrendo sobre as diversas camadas da mesma, porém com ênfase na camada mais próxima dos seres vivos na Terra, a *“Troposfera”*. Em seguida dá-se um tratamento mais completo às reações químicas ocorrendo nessa camada. Divide-se o estudo em atmosferas limpa e poluída, para que se tenha uma melhor idéia do papel desempenhado pelos óxidos de nitrogênio na baixa troposfera. Esse capítulo toma por base a abordagem de três autores importantes no cenário dos estudos sobre o meio ambiente. A forma de exposição, entretanto, segue a sugestão de Baird (1995), a qual, sob o ponto de vista didático, é excelente e engloba vários pontos de interesse da discussão que se pretendeu fazer nessa parte da tese.

Capítulo III: Esse capítulo foi dividido em três amplas seções: Seção 3.1, Seção 3.2 e Seção 3.3. Cada uma delas é apresentada a seguir.

Seção 3.1: Aborda os métodos quimiométricos mais usados para a decomposição de matrizes. Dá-se ênfase à Análise de Componentes Principais (PCA) e à Análise de Fatores

(FA), em primeiro lugar por suas posições pioneiras no cenário da análise multivariada e, em especial, em quimiometria. Em segundo lugar, porque são métodos usados por um expressivo número de autores, portanto, com ampla divulgação em diversas áreas do conhecimento científico. Nesse capítulo a explanação se situa em um plano mais introdutório, o qual serve de base para a abordagem mais profunda, do ponto de vista algébrico, que se faz na Seção 3.2. Vale acrescentar que a despeito desse capítulo explorar apenas de forma superficial as análises PCA e FA, seu conteúdo é suficiente para propósitos de sua simples aplicação. Assim, na leitura da Seção 3.2, a abordagem sobre PCA é uma contribuição para uma discussão algébrica mais detalhada da mesma, essencial para uma utilização que transcenda a aplicação pura e simples do método.

Seção 3.2: Aqui são discutidos os conceitos de base, dimensão e modo, essenciais para a compreensão da terminologia adotada. Dá-se um tratamento matemático um pouco mais elaborado à PCA. Inicia-se a discussão destacando-se as interpretações filosófica, algébrica e geométrica de PCA, e procura-se, a partir daí, mostrar as principais diferenças entre esta e a Análise de Fatores. Em seguida é dada uma visão geral sobre a análise multimodo (ou a dados de ordem superior), destacando-se os modelos PARAFAC e o Tucker3. Mostra-se como o modelo PARAFAC pode ser um tipo de generalização de PCA, e como o modelo Tucker3 pode ser uma outra generalização de PCA. Finalmente discute-se a unicidade dos modelos enfatizando-se as características de solução única para o PARAFAC, estruturalmente diferente do modelo de Tucker.

Seção 3.3: Pode-se considerá-la como o cérebro da tese. É nela que estão fundados os alicerces para a construção da metodologia. Os argumentos para o uso de cada método ou modelo são justificados em função de alguns pressupostos. Assume-se que as variações sistemáticas presentes nos dados podem emergir por meio da aplicação de métodos/modelos adequados de análise de dados. Dada a natureza dos dados, justifica-se todo um procedimento capaz de mostrar, por um lado, que modelos bilineares, embora necessários, são insuficientes para trazer à superfície o pleno comportamento das variáveis envolvidas, e por outro, que a análise multimodo fornece um caminho suficientemente seguro rumo aos perfis exibidos pelas variáveis submetidas a diferentes condições. A seção aborda também o pretratamento de dados. Discute-se o conjunto de dados na origem, sua validação, tratamento aplicado aos dados anômalos (“outliers”), e aos dados faltantes

(“missing data”), criando-se aqui o conceito de “lacunaridade”, útil porque de reflexos sobre os resultados, distintos daqueles causados por matrizes esparsas.

Capítulo IV: Trata da parte experimental. É um texto curto destacando as características da RMSP (Região Metropolitana de São Paulo), das Estações de Medição da CETESB e da instrumentação de sua rede telemétrica. Discorre rapidamente pelas dificuldades para colocar os dados em condições de análise.

Capítulo V: É um capítulo longo devido à grande quantidade de dados analisados. Nele são destacados aspectos importantes para efeito de comparações entre os poluentes e suas fontes. São lançados os resultados sobre nove variáveis para a Estação de Medição Pedro II, durante o ano de 1999; os resultados sobre o monóxido de carbono para as sete localidades investigadas, durante o ano de 1998; e alguns resultados adicionais para algumas variáveis, em algumas localidades no período 1997 a 1999. Tais resultados são acompanhados de breve discussão quando necessário, e servem como fonte introdutória de informações para subsidiar a discussão preliminar que se faz no Capítulo VI.

Capítulo VI: São lançados os resultados para três Estações de Medição (Parque D. Pedro II, São Caetano e Ibirapuera). Quando as variáveis metano e hidrocarbonetos não metano são incluídas, são lançados os resultados apenas para P. D. Pedro II e São Caetano, únicos locais nos quais as medições dessas variáveis foram realizadas. Tais resultados dão suporte à profunda discussão apresentada no Capítulo VII.

Capítulo VII: Se a Seção 3.3 pode ser considerada o cérebro, o Capítulo VII pode ser considerado o coração da tese. Da base fornecida pelo capítulo precedente estabelece-se associação entre os perfis encontrados e o comportamento dos poluentes, sugerindo-se vias preferenciais para mecanismos de algumas reações químicas na baixa troposfera. Discute-se a origem da ocorrência de pico noturno para o ozônio, sua acumulação nos finais de semana, e a relação de seus níveis de concentração com aqueles registrados para o metano e o monóxido de carbono, em presença de óxidos de nitrogênio.

Conclusões: São apresentadas em doze itens sucintos.

CAPÍTULO I

O HOMEM, A CIÊNCIA E O AMBIENTE

Uma abordagem sobre ambiente e as interações do homem com ele demandaria, dependendo do enfoque, um espaço que não se pode definir por antecipação. Essa abordagem deve englobar todos os aspectos que a ciência ambiental possa conter, tais como todas as disciplinas não só diretamente relacionadas com os aspectos técnicos da pesquisa ambiental, como também aquelas aparentemente "alheias" a esses problemas como as disciplinas do campo das ciências sociais.

Se, por um lado, nos últimos tempos, a ausência de consciência crítica de fato parece estar sendo estimulada pela política oficial e por formadores de opinião, mesmo naquelas áreas consideradas vitais para a produção e transferência de conhecimento científico-técnico, por outro, nas áreas mais relacionadas com as ciências sociais tem-se apontado para a extinção pura e simples da maioria dos cursos do seu âmbito, como se desse modo fosse possível finalmente, como tentou Fukuyama (Fukuyama, 1989, p.3-18), declarar o fim da história.

Não obstante a importância de todo esse complexo conjunto de eixos de discussões, não se pode negligenciar a importância e o papel da ciência e da tecnologia na construção de um futuro mais saudável à vida e aos materiais na Terra. Assim sendo, pretende-se iniciar a discussão por essa via, colocando em destaque a ciência ambiental.

Para que se tenha compreensão mais adequada da química ambiental é necessário uma visão da ciência ambiental como um todo. Esta é, num sentido mais amplo, a ciência que estuda as interações complexas ocorrendo entre os ambientes terrestres, aquáticos, atmosféricos, vivos e antropológicos. Ela inclui todas as disciplinas tais como, química,

biologia, ecologia, sociologia etc que afetam ou descrevem essas interações (Manahan, 1994, p.1).

Nesse sentido, o ar, a água, o solo, a vida, e a tecnologia estão fortemente interligados (Figura 1.1). Essa figura resume e delinea o tema abordado nesse capítulo.

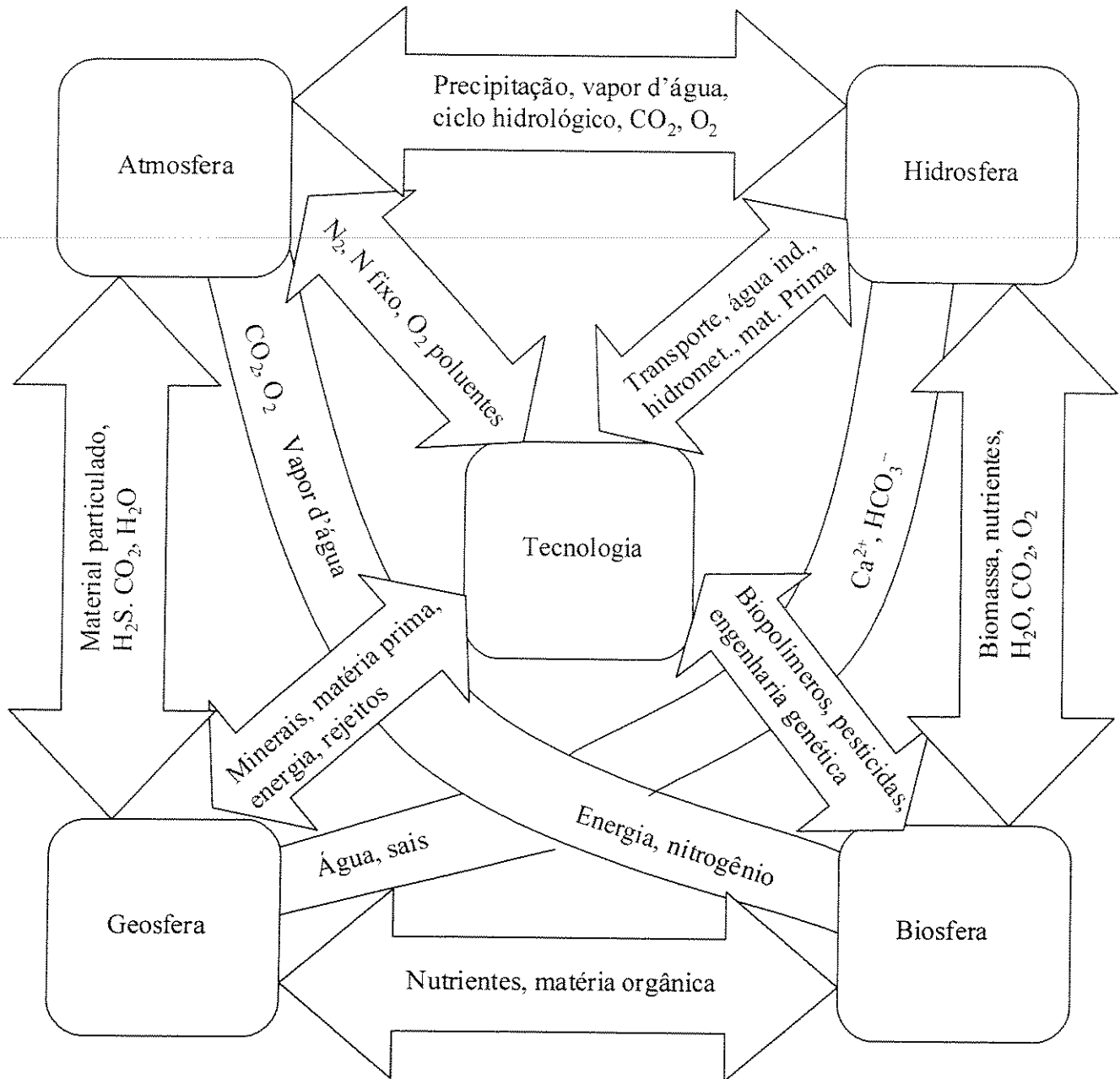


Fig.1.1: Ilustração das relações entre os ambientes atmosférico, aquático e terrestre, e suas inter-relações com os sistemas vivos e com a tecnologia. (Extraído de Manahan, 1994, p.2)

A ciência ambiental é, tradicionalmente dividida entre o estudo da atmosfera, da hidrosfera, da geosfera e da biosfera, de acordo com as definições de Manahan a seguir (Manahan, 1994, p.1-3).

A *atmosfera* é a fina camada de gases que cobre a superfície da Terra. Além do seu papel como reservatório de gases, ela modera a temperatura da Terra, absorve energia e radiação prejudicial ultravioleta do sol, transporta energia proveniente das regiões equatoriais, e serve como caminho para o movimento da fase vapor d'água no ciclo hidrológico.

A *hidrosfera* contém a água da Terra. Acima de 97% da água do planeta está nos oceanos, e a maioria da água doce está na forma de gelo. Então, somente uma porcentagem relativamente pequena do total da água na Terra está realmente envolvida com os processos terrestres, atmosféricos e biológicos. Exclusiva daquela dos mares, a água que circula através dos processos e ciclos ambientais ocorre na atmosfera, no subsolo como águas subterrâneas, e como águas de superfície em correntes, rios, lagos, açudes e reservatórios.

A *geosfera* consiste de terra sólida, incluindo o solo, o qual suporta a vida da maioria das plantas. A parte da geosfera que está diretamente envolvida com os processos ambientais através do contato com a atmosfera, a hidrosfera, e com os entes vivos, é a litosfera sólida. A litosfera varia de 50 a 100 km em espessura. A parte mais importante dela no que diz respeito às interações com as outras esferas do ambiente, é a sua fina camada externa composta grandemente de minerais leves à base de silicatos, chamada crosta.

Toda a entidade viva na Terra compõe a *biosfera*. Os organismos vivos e os aspectos do ambiente pertencentes diretamente a eles são chamados bióticos, e as outras porções do ambiente são abióticas.

Em grande medida, as fortes interações entre organismos vivos e as várias espécies do ambiente abiótico são melhor descritas por ciclos da matéria que envolvem processos e fenômenos biológicos, químicos, e geológicos. Tais ciclos são chamados de ciclos biogeoquímicos (Manahan, 1994, p.3).

A Química Ambiental

A Química Ambiental engloba diversos tópicos, podendo ser definida como “*o estudo das fontes, reações, transporte, efeitos, e destinos das espécies químicas na água, no solo, e no ar, e os efeitos da tecnologia sobre esses ambientes*” (Manahan, 1994, p.3).

Não obstante a ausência até bem pouco tempo de uma tentativa de "cultura ambientalista" nos cursos ligados mais diretamente à área de química, a química ambiental não é uma disciplina nova. Trabalhos excelentes foram produzidos nesse campo em boa parte de século XX, entretanto, até cerca de 1970 a maioria deles era desenvolvida em setores não necessariamente ligados diretamente com problemas químicos, isto é, cuja formação básica não era na área de química. Assim, quando os pesticidas foram sintetizados, os biólogos foram os primeiros a observar as conseqüências indesejáveis do seu uso. O mesmo aconteceu com os detergentes, quando profissionais de engenharia sanitária assistiam estupefatos o desaparecimento sob camadas de espuma, dos tanques de “aeração” das estações de tratamento de esgotos. Da mesma forma, limnologistas se surpreendiam porque lagos previamente normais tornavam-se “sufocados” com o odor desagradável de cianobactéria (Manahan, 1994, p.3).

É encorajador o fato de que nos últimos anos muitos químicos têm procurado se envolver profundamente com a investigação dos problemas ambientais. Um outro fato importante nesse aspecto é o esforço que tem sido feito na direção de se criar uma cultura para a preservação do ambiente. No Brasil, esse esforço parece confuso porque alguns pressupostos que devem ser considerados não o estão sendo. Por exemplo, uma verdadeira “enxurrada” de publicações sobre o tema parece estar ancorada muito mais no oportunismo e na visão utilitarista do momento do que, de fato, em conscientização da população em geral sobre o problema. Além disso, a própria propaganda oficial, coadjuvada por intervenções de boa parte dos meios de comunicação de massa, se dá sem uma discussão qualificada das origens dos problemas, acarretando apreensões diferenciadas e, via de regra, apenas superficiais, da questão ambiental pela população.

Nas universidades e institutos de pesquisas, um dos principais desafios da química ambiental é a determinação da natureza e da quantidade de poluentes específicos no ambiente. Assim, a análise química é uma etapa vital em pesquisa ambiental. As

dificuldades são de várias ordens. Não é possível estabelecer um mesmo padrão de qualidade ambiental em função dos níveis alcançados pelos poluentes, para os diferentes ambientes, terrestre, aquático e atmosférico. Assim, níveis significativos de poluentes do ar podem consistir de quantidades inferiores a cerca de um micrograma por metro cúbico (10^{-6} mg L⁻¹) de ar. Para poluentes aquáticos, valores relativamente fáceis de se detectar tais como uma parte por milhão em peso (essencialmente 1 miligrama por litro) são muito altos. Por outro lado, quantidades de apenas umas poucas partes por trilhão, podem representar níveis significativos, sob o ponto de vista ambiental. Assim, é obvio que as análises químicas usadas para estudar alguns sistemas ambientais exigem um limite de detecção muito baixo (Manahan, 1994, p.4).

A química ambiental envolve, todavia, um conjunto de outras disciplinas além da química analítica. Ela requer, de fato, uma visão mais completa da natureza e dos grandes problemas ambientais aos quais está sujeita, causados principalmente pela ação do homem na sua ânsia de exploração sem limites do planeta. Assim, para que a química possa oferecer uma contribuição adequada na solução dos problemas ambientais, o químico deve trabalhar na direção de um melhor entendimento da natureza, das reações, e do transporte de espécies químicas no ambiente (Manahan, 1994, p.5), sem, contudo, deixar de dotar-se dos conhecimentos adicionais necessários no que diz respeito aos recursos naturais, sua exploração, utilização e descarte dos resíduos gerados decorrentes daquela utilização.

A água com fórmula química extremamente simples, H₂O, tem importância vital em todas as partes do ambiente. Ela cobre cerca de 70% da superfície da Terra e ocorre em todas as esferas do ambiente, com sua maior quantidade nos oceanos, como um grande depósito de água salgada. Ocorre sobre o solo como águas de superfície nos rios e lagos; no subsolo como águas subterrâneas; na atmosfera como vapor d'água; e nas calotas polares no estado sólido ou gelo, como é conhecida (Manahan, 1994, p.5).

Além da água e do ar, as outras partes de importância para o estudo ambiental são a Terra propriamente dita, composta de sua parte sólida, a *geosfera*, seu manto exterior denominado *litosfera*, e a camada mais externa desta, a *crosta*. As duas últimas formam as porções em que a ciência ambiental está mais preocupada em estudar. A geologia é a ciência que se encarrega desse estudo. Uma outra parte desse elo é a vida existente no planeta. A ciência que se ocupa do estudo da vida é a biologia.

Em especial, já nesse início de século, a preocupação mundial estará voltada para os problemas de oferta de água potável que já encontra estrangulamento em alguns países. Aliado a isto, países não pertencentes ao G-7 (grupo dos países mais desenvolvidos) que possuam quantidade apreciável de recursos energéticos, deverão adotar medidas capazes de não comprometer esses recursos irreversivelmente, seja na venda de suas fontes ou mesmo na sua exploração e utilização irresponsável, de modo não só a preservá-las para uso futuro, como também para tornar mínimos os efeitos maléficos decorrentes da utilização em larga escala dos mesmos. As futuras gerações dependem dramaticamente dessa conscientização.

A atmosfera, camada que protege a Terra das radiações provenientes do sol e de eventuais objetos "hostis" ao seu ambiente, é a parte de que se ocupa mais especificamente essa tese e será abordada em detalhes em várias outras seções do mesmo.

Boa parte do conforto usufruído por uma pequena parte da população mundial, a mais favorecida pela divisão internacional do trabalho (e por outras medidas relacionadas à divisão geopolíticoeconômica), provém da energia e dos ciclos de energia que, de uma forma ou de outra, estão relacionados à energia que emana do sol. Entretanto, boa parte dos problemas ambientais pelos quais passa o mundo também se deve à fatores ligados diretamente à energia.

Durante os últimos dois séculos o impacto humano na utilização e conversão de energia tem sido enorme e tem resultado em muitos problemas ambientais com os quais se defronta hoje a espécie humana. Esse período é testemunha de uma transição com marco inicial no fim do uso quase exclusivo da energia capturada pela fotossíntese e utilizada como biomassa (alimento para prover força muscular, madeira para aquecimento), para o uso de combustíveis fósseis (Manahan, 1994, p.11) em cerca de 80%, e da energia nuclear em cerca de 6,9% de toda a energia empregada comercialmente (IEA, 2003, p.6).

O consumo de combustível fóssil é dividido primariamente entre petróleo, gás natural e carvão. Essas fontes são limitadas em sua quantidade e possuem alto potencial de poluição. A mineração de carvão e a extração de petróleo são destruidoras do ponto de vista do ambiente. A combustão de carvão com alto teor de enxofre libera dióxido de enxofre para a atmosfera, e todos os combustíveis fósseis produzem o dióxido de carbono, um gás de estufa. Então será necessário um movimento mais efetivo na direção de fontes alternativas de energia, particularmente aquelas que são renováveis. A energia solar e a

proveniente da biomassa estão entre as mais promissoras, e em seguida a eólica. A energia nuclear, com reatores nucleares mais seguros e eficientes, está ganhando cada vez mais atenção como uma fonte confiável e amigável do ponto de vista ambiental (Manahan, 1994, p.11), a despeito dos dois grandes acidentes nucleares do passado (Chernobil-URRS e Three Mile Islands-USA), e dos rejeitos que ela pode gerar. Correntes “ambientalistas” e ecologistas em todo o mundo são os principais entraves à sua utilização despreocupada. Deve ser enfatizado, entretanto, que enquanto não for encontrada uma solução viável para o problema dos rejeitos gerados pelas usinas nucleares, estas devem ser descartadas de programas de geração de energia.

Não obstante a grande preocupação com os diferentes tipos de fontes de energia e as alternativas àquelas não renováveis e simultaneamente mais poluentes, os problemas causados pela utilização irresponsável dessas últimas é grande o suficiente para demandar pesquisas em todo o mundo no sentido de se tornar mínimos seus efeitos maléficos. Nesse sentido, uma associação entre impacto humano e poluição ambiental é inevitável.

Com isto surgiram termos tais como:

contaminação – desvios da composição normal em um ambiente (Manahan, 1994, p.13),

poluição – ocorrência de substâncias em concentrações acima da natural, causada pela atividade humana, com conseqüente degradação sobre o ambiente (Manahan, 1994, p.13);

e outros como:

fonte – local de origem ou emissão de poluentes (Manahan. 1994, p.13; Rebouças, 1997, p.11),

receptor – seres vivos e materiais que removem e, portanto, são afetados por poluentes (Manahan. 1994, p.13; Rebouças, 1997, p.11) e

sumidouro – local depositário de poluentes de vida longa, os quais permanecem, não necessariamente de forma definitiva, mas por longo tempo (Manahan. 1994, p.13).

Assim, do impacto humano nos “front” primário (utilização direta) e secundário (rejeitos), surgiu a contrapartida ao frenesi irresponsável: *poluição dos sistemas terrestres, aquáticos e atmosféricos*.

A tecnologia tem um papel fundamental nesse processo e um pouco sobre ela deve ser comentado.

Manahan (1994) reserva um tópico (p.15-17) em um dos capítulos, e um capítulo inteiro (o 17, p.487) para se referir à tecnologia. O tópico ele intitula "Technology: the problems it poses and the solutions it offers". Aí ele argumenta que "... a tecnologia moderna tem fornecido os meios para uma alteração maciça e poluição do ambiente. Entretanto, a tecnologia, inteligentemente aplicada com uma forte consciência ambiental, também fornece os meios para lidar com os problemas de poluição e degradação ambiental" (Manahan, 1994, p.16). Em seguida ele cita alguns dos principais meios através dos quais a tecnologia tem contribuído para a alteração e poluição do ambiente. Finaliza o tópico citando outros meios através dos quais a tecnologia pode ser aplicada para "amenizar" o impacto ambiental.

Entretanto, independentemente do poder que ciência e tecnologia possuem para solucionar problemas ambientais, parece não existir um equilíbrio adequado na equação que liga a necessidade de produção de um bem qualquer e os benefícios que isto possa trazer para os entes envolvidos no processo. Parece que a ausência desse equilíbrio repousa na ânsia de se concentrar riqueza. Esse processo tem três conseqüências perversas: o desequilíbrio na divisão internacional do trabalho, o aumento de exclusão social e, finalmente, a degradação aparentemente sem limites dos recursos existentes no planeta.

Aqui os lucros, de qualquer espécie, concentram-se nas mãos de uns poucos. Quando dessa ação exploratória resulta um dano ao qual fica ou poderá ficar exposto o principal ou principais beneficiários (em geral uma minoria "empreendedora"), estes acionam recursos, sempre via poder público, para que a reparação ou atenuação dos efeitos maléficos possa ser implementada. Aqui também fica estabelecido pelos responsáveis pela ação predatória que os prejuízos, extra sua participação no ônus, sejam socializados. Então, ciência e tecnologia são "chamadas" a intervir. E sob essa ótica, todos os problemas gerados pela utilização do avanço tecnológico podem ser resolvidos com mais geração de conhecimento. Não parece sensato. Não parece sensato desenvolver tecnologia para oferecer conforto, causando simultaneamente desconforto, exigindo desenvolvimento de mais tecnologia para eliminar o desconforto o que, por sua vez, causa outro tipo de desconforto, exigindo mais tecnologia para eliminar o novo desconforto, causando um

terceiro tipo de problema, o qual exige mais tecnologia para eliminar ..., e assim sucessivamente. Assemelha-se ao conhecido problema da “bola de neve”.

Não obstante não seja contraditório, e a história é prova suficiente, é, no mínimo, paradoxal que seres humanos sejam capazes, por um lado, de desenvolver conhecimento para gerar tecnologia com finalidade específica para resolver problemas gerados por um estágio apenas anterior ou acessório da própria tecnologia, e, por outro, não sejam capazes de se antecipar ao processo, desenvolvendo tecnologia que evite ou contorne problemas ambientais. Mesmo que se leve em conta o desenvolvimento das chamadas tecnologias limpas, as peças não se encaixam. Está na ordem do dia, na agenda internacional, o tema da mudança climática para prová-lo. Isto remete a questão para o problema das emissões de gases com forte componente causadora de poluição.

Nesse sentido, enquanto muitos países estão comprometendo parcialmente sua tentativa de desenvolvimento para reduzir emissões, e outros mais desenvolvidos exigem dispêndio de tempo e argumentos muito além dos necessários para cumprir a meta, os EUA, o maior consumidor de combustíveis de fontes não renováveis e maior emissor de poluentes, em um primeiro momento recusa-se ao cumprimento do protocolo de Kyoto para a redução de emissões. Em um segundo momento retira-se do processo de discussão. A marca que o episódio deixa é a de que a nação mais poderosa do mundo há quase um século, não se predispõe a reduzir o conforto do qual goza boa parte de sua população. Conforto este, estimulado pelo próprio “american way of life”.

De uma maneira geral, as pessoas, adjetivadas como “clientes” ou “consumidores”, são estimuladas, com auxílio de propaganda maciça dos meios de comunicação de massa, a consumir muito além das necessidades mínimas para uma vida saudável. Como se exigir que a população em geral assuma hábitos de controle de consumo de bens se, contraditoriamente, a propaganda exerce uma postura fortemente agressiva de estímulo ao consumo de produtos que vão desde qualquer tipo de “quinquilharia” a luxuosos automóveis, que o mercado se encarrega de produzir? Alguma coisa no elo que liga as diversas interações na ciência ambiental parece não estar clara.

Se tais contradições não podem ser abordadas e discutidas abertamente, é porque interesses muito além do bem estar do planeta estão sendo privilegiados por forças contra as quais parece ser difícil se opor.

Nesse sentido, se é desejo da comunidade científica, de fato, se posicionar a favor de uma Terra mais saudável e preservada para as gerações futuras, esse prisma aparentemente proibido da discussão tem que ser trazido à baila. Ou seja, não basta que se multipliquem os esforços em pesquisas de toda a sorte, que se busquem recursos de todo o tipo, que se sugiram as mais diversas taxas ambientais "de segurança" na exploração e utilização de recursos ambientais, e que se adquira uma "cultura ambientalista" no sentido que está sendo imposto pela propaganda. É necessário, sobretudo, que se adquira uma consciência genuinamente ambiental, e ela deve passar necessariamente pela compreensão do que, em seus estudos sobre o capital, Marx denomina valor de uso e valor de troca (Marx, 1975, p.41-54).

CAPÍTULO II

A ATMOSFERA E A QUÍMICA AMBIENTAL

2.1 – A ATMOSFERA

A atmosfera é a camada protetora da superfície terrestre. Ela é a fonte de dióxido de carbono para a fotossíntese das plantas, e de oxigênio para a respiração. Ela fornece o nitrogênio, essencial em vários processos que ocorrem nos sistemas vivos. Os principais componentes (excluído o sempre presente mas variável vapor d'água) de uma versão não poluída da atmosfera da Terra são o nitrogênio diatômico (N_2 , cerca de 78% das moléculas), o oxigênio diatômico (O_2 , cerca de 21%), o argônio (Ar, cerca de 1%), dióxido de carbono (CO_2 , cerca de 0,03% e aumentando como um resultado da queima de combustíveis), e uma grande variedade de gases em níveis traço abaixo de 0,002% (Manahan, 1994, p.6,31; Baird, 1995, p.16). Essa mistura de compostos parece não reativa na atmosfera inferior mesmo à temperaturas ou intensidades de luz solar bem acima daquelas naturalmente encontradas na superfície da Terra. Entretanto, essa ausência manifesta de reatividade na atmosfera é enganadora. De fato, no ar, ocorrem muitos processos químicos importantes do ponto de vista ambiental, esteja ele poluído ou não (Baird, 1995, p.16).

Constituindo parte básica do ciclo hidrológico (Figura 2.1), a atmosfera transporta água dos oceanos para o solo, agindo como um condensador em imenso destilador à energia solar. Infelizmente a atmosfera também tem sido usada como depósito de lixo para muitos materiais poluentes, uma prática que causa danos à vegetação, aos animais e aos materiais, diminui a vida humana, e altera as características da própria atmosfera (Manahan, 1994, p.6,263).

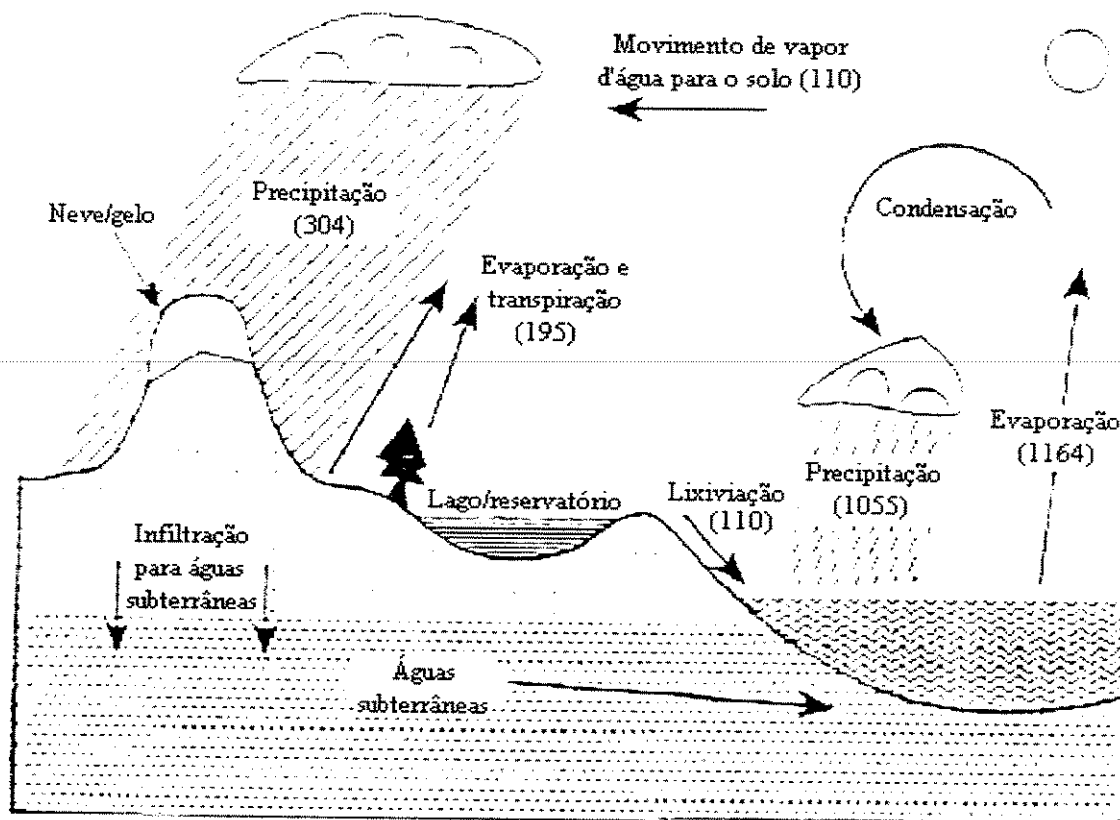


Fig.2.1: O ciclo hidrológico. Quantidade de água (valores entre parênteses) em trilhões de litros por dia (Extraído de Manahan, 1994, p.24).

2.1.1 – Regiões da Atmosfera

A ciência atmosférica lida com o movimento de massas de ar, com o balanço de calor, e com reações e composição químicas na atmosfera. Para entender a sua química e os problemas de poluição do ar, é necessário ter uma visão geral da atmosfera, de suas regiões, da sua composição, e principais características físicas (Manahan, 1994, p.6,263-264).

A estratosfera se estende da tropopausa com temperatura mínima de cerca de 190-215 K (−83 a −58 °C), próximo dos 15 km nos trópicos e 8 km a altas latitudes, até a

estratopausa com temperatura máxima de 240-280 K (-33 a -7 °C) , a aproximadamente 50 km (Brune, 1998, p.294, Manahan, 1994, p.268). Esse aumento de temperatura com a altitude deve-se à presença de ozônio, O_3 , o qual pode alcançar níveis da ordem de 10 ppmv na região média da estratosfera. A absorção de radiação ultravioleta pelo O_3 nessa região provoca o aquecimento nessa camada (Manahan, 1994, p.268). Embora as temperaturas na estratosfera em geral aumentem em qualquer lugar com a altura, os seus valores dependem da localização e da estação do ano, particularmente nas estratosfera inferior e superior (Brune, 1998, p.294).

A troposfera é a camada mais baixa da atmosfera, estendendo-se desde o nível do mar a altitudes de aproximadamente 10-16 km. O seu limite superior apresenta temperatura mínima de cerca de 217 K (-56 °C) e no limite inferior de aproximadamente 288 K (15 °C). Essa camada caracteriza-se por uma composição aproximadamente constante dos principais (exceto o vapor d'água) gases que a compõem, sendo a correlação negativa entre temperatura e altitude uma função do calor radiante da superfície. A composição homogênea da troposfera resulta de uma constante mistura causada por massas de ar circulante. Entretanto, o conteúdo de vapor d'água da troposfera é extremamente variável devido a formação de nuvens, precipitação e evaporação de água dos corpos de água terrestres (Manahan, 1994, p.268).

Outras regiões da atmosfera são a mesosfera, imediatamente acima da estratosfera, em altitudes de 50 a 85 km com temperaturas de 271 a 181 K (-2 a -92 °C), e a termosfera, acima desta, com temperaturas de até 1473 K (1200 °C) no seu limite superior de cerca de 500 km de altitude (Manahan, 1994, p.268).

A Figura 2.2 a seguir mostra as regiões da atmosfera.

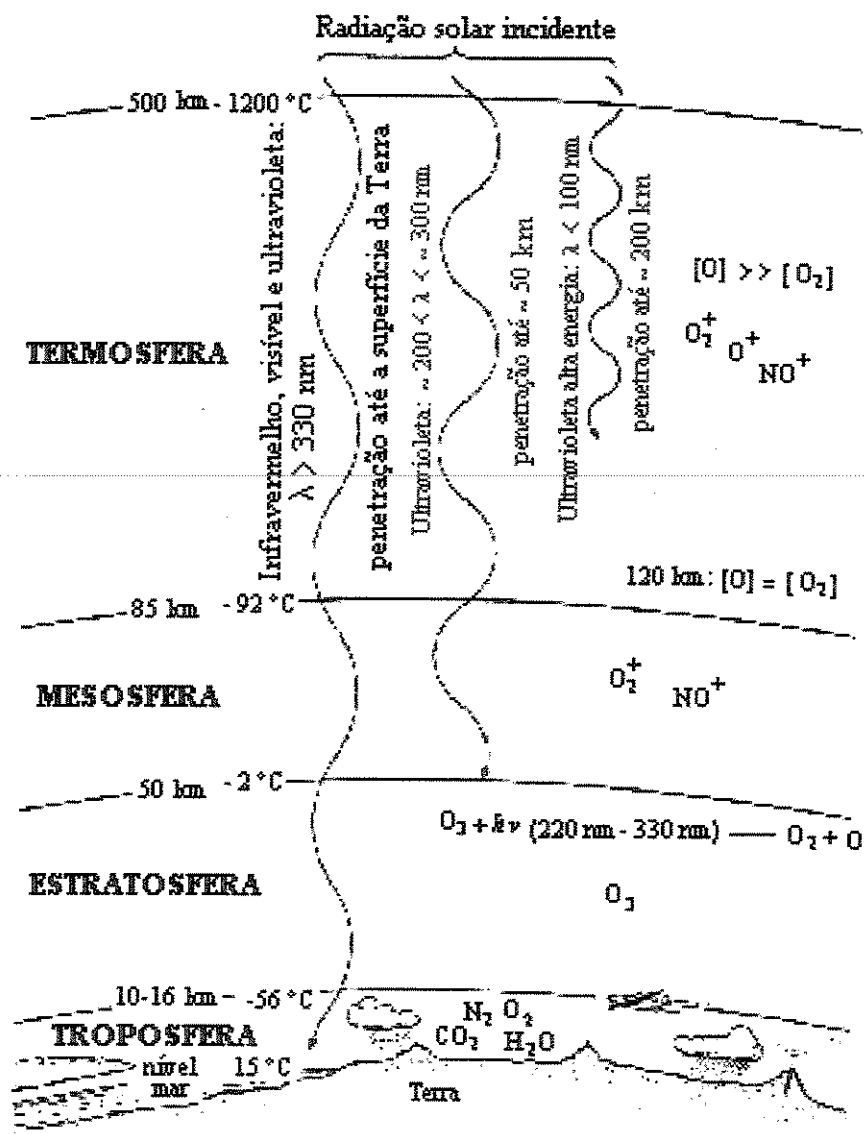


Fig.2.2: Principais regiões da atmosfera (Extraído de Manahan, 1994, p.267)

2.2 – O Espectro Eletromagnético

Boa parte dos processos que ocorrem na estratosfera são dirigidos pela energia associada à luz solar.

O espectro eletromagnético pode ser classificado segundo os comprimentos de onda da radiação emitida. A Figura 2.3 mostra a região de interesse do ponto de vista ambiental de tal espectro.

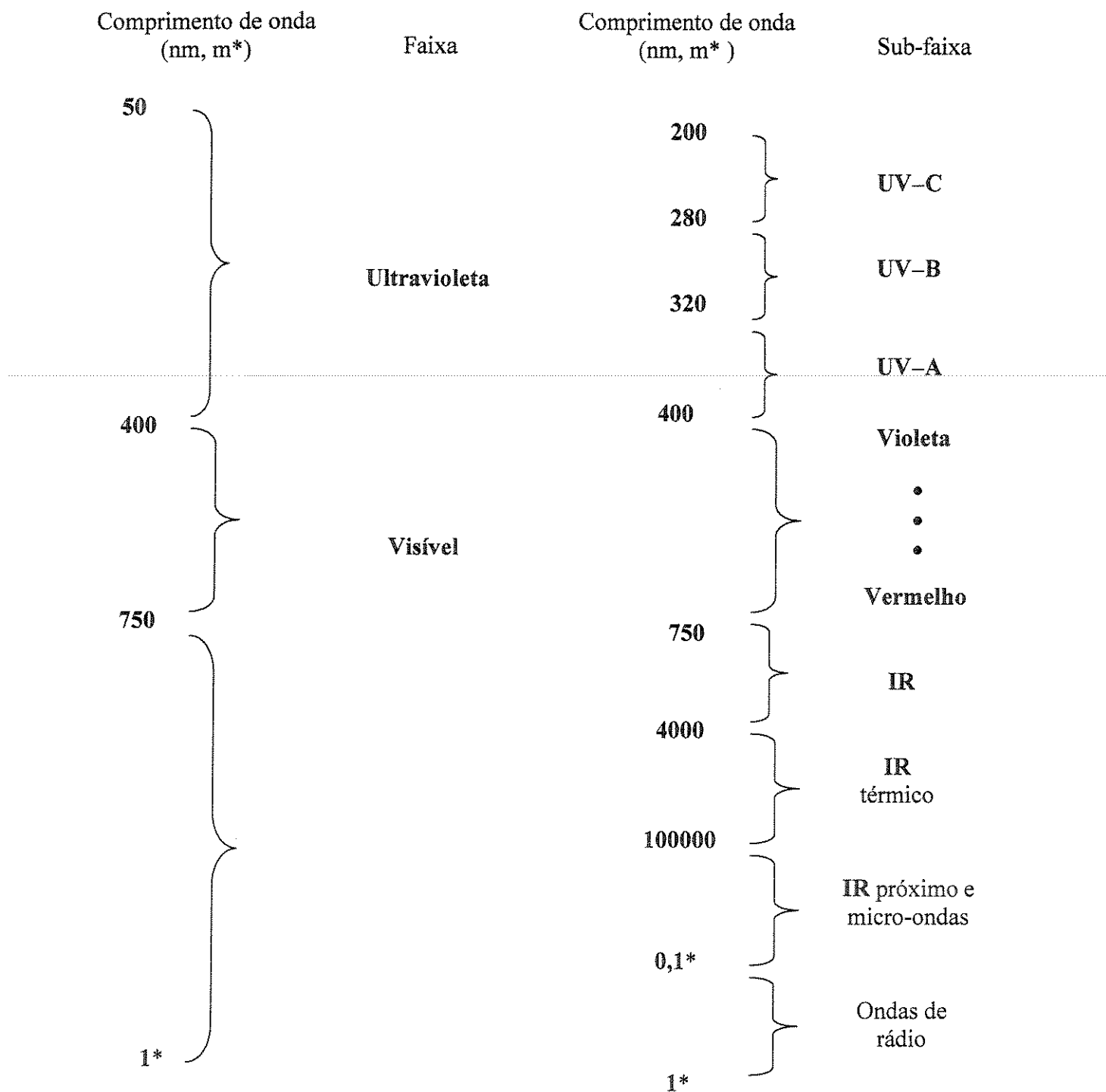


Fig.2.3: Espectro eletromagnético. Regiões de interesse ambiental

2.3 – A Química da Troposfera

A troposfera é a camada da atmosfera mais próxima da superfície da Terra. É em contato com ela, mais especificamente com a região mais baixa de sua parte inferior, a baixa troposfera, que os seres vivos e os materiais se encontram. Em razão da quantidade de emissões efetuadas pelo homem nessa camada da atmosfera, e em função das modificações no comportamento químico da mesma, decorrente daquelas emissões, faz-se necessário, cada vez mais, um profundo conhecimento da química da troposfera e de sua relação com os meios com os quais estabelece fronteira, tais como a estratosfera, o solo, e os ambientes aquáticos.

Uma das mais importantes características da atmosfera terrestre é a de que ela é um meio oxidante. Isto se deve à grande concentração de oxigênio diatômico que ela contém. Quase todos os gases liberados ao ar, sejam eles substâncias "naturais" ou "poluentes", são totalmente oxidados nele e seus produtos finais depositados na superfície da Terra. Assim, as reações de oxidação são vitais para a limpeza do ar (Baird, 1995, p.77-78), em que pese a potencial "vocaç  o" para transferir, nesse processo, a polui   o para locais al  m fronteira.

Um dos melhores exemplos de polui   o do ar s  o os "smogs" (se   o 2.3.4.1) que ocorrem em v  rias cidades atrav  s do mundo. Em geral os reagentes que produzem certo tipo de "smog" s  o principalmente as emiss  es de ve  culos automotores. Na verdade, a opera   o dos motores dos ve  culos produz mais polui   o do ar do que qualquer outra atividade humana isolada. O "smog"    um tipo, de polui   o que tem graves conseq   ncias. N  o apenas pelo odor desagrad  vel de alguns de seus componentes gasosos mas, principalmente devido a seus intermedi  rios e produtos finais das re    es no "smog", os quais podem afetar a sa  de humana, causar dano   s plantas, aos animais, e alguns materiais (Baird, 1995, p.77).

Nesse t  pico estuda-se o comportamento de alguns poluentes nessa camada mais baixa da troposfera.    oportuno, pois, discutir esse tipo de polui   o atmosf  rica bastante comum em cidades com apreci  vel tr  fego urbano.

Sob determinadas condi    es, a qu  mica da troposfera como um todo n  o difere daquela que ocorre na baixa troposfera. Assim, para se ter uma melhor compreens  o dos processos que ocorrem mais pr  ximos da superf  cie em que vivem plantas e animais, faz-se necess  rio um conhecimento sobre tais processos qu  micos nessa camada da atmosfera.

2.3.1 – Considerações Sobre a Química da Troposfera

Os processos subjacentes aos fenômenos atmosféricos visíveis (ou não) e aos quais toda a natureza é sensível, são complexos. Para entendê-los melhor é necessário que se faça um estudo mais detalhado da química da troposfera de modo que se possa compreender a reatividade por trás dos fenômenos e o potencial à reatividade das próprias espécies químicas envolvidas. Para tal é sempre conveniente abordar o problema primeiramente para os processos que ocorrem no ar limpo, ou seja, no ar não poluído. Entendidos pelo menos em parte, os mecanismos que ocorrem na ausência das espécies chave na formação da atmosfera poluída, pode-se partir para o estudo dos principais mecanismos de reação que explicam de forma satisfatória os fenômenos que caracterizam o ar poluído.

A atmosfera recebe regularmente, de fontes biológicas e vulcânicas, gases parcialmente oxidados como o monóxido de carbono, CO, e dióxido de enxofre, SO₂, e vários outros gases que são simples compostos de hidrogênio, alguns dos quais possuem seus átomos na forma altamente reduzida (por exemplo, H₂S, NH₃). Os compostos mais importantes dessas fontes naturais estão listados na Tabela II.1 (Baird, 1995, p.112-113).

Tabela II.1 – Gases emitidos na atmosfera a partir de fontes naturais

Fórmula	Nome	Fonte Natural Importante
CH ₄	metano	decaimento biológico anaeróbio
NH ₃	amônia	decaimento biológico anaeróbio
H ₂ S	sulfeto de hidrogênio	decaimento biológico anaeróbio
HCl	cloreto de hidrogênio	decaimento biológico anaeróbio, vulcões
CH ₃ Cl	cloreto de metila	oceanos
CH ₃ Br	brometo de metila	oceanos
CH ₃ I	iodeto de metila	oceanos
CO	monóxido de carbono	CH ₄ atmosférico, queimadas
N ₂ O	óxido nitroso	processos microbiológicos
N ₂ , NO _x	nitrogênio (e ox. nitr.)	relâmpagos
SO ₂	dióxido de enxofre	vulcões

Fonte: Adaptado de Baird, 1995, p.113 e Manahan, 1994, p.265

Embora a maioria desses gases seja gradualmente oxidado no ar, nenhum deles reage diretamente com oxigênio atômico. Ao contrário, todas as suas reações começam quando eles são atacados por um radical hidroxila livre, HO[•]. No ar troposférico limpo, o radical hidroxila é produzido quando uma fração de átomos de oxigênio excitado resultante da decomposição fotoquímica de quantidades em níveis traço de ozônio atmosférico reage com água no estado líquido para abstrair um átomo de hidrogênio de cada molécula de água (Manahan, 1994, p.288; Baird, 1995, p.113; Birks, 1998, p.236)



Deve-se observar que a reação envolvendo oxigênio atômico não excitado e água é endotérmica, tendo uma alta energia de ativação e, conseqüentemente ocorre muito lentamente para ser uma fonte expressiva de HO[•]. (Baird, 1995, p.113) Como será visto adiante, a oxidação de metano, CH₄, a dióxido de carbono, CO₂, gera radicais hidroxila adicionais na atmosfera aumentando o seu potencial oxidante.

2.3.2 – Revisão da Reatividade Química na Troposfera

A maioria dos gases na troposfera é gradualmente oxidado por uma seqüência de reações envolvendo radicais livres e, para um determinado gás, a seqüência pode ser prevista a partir de alguns princípios que norteiam tais reações, os quais podem ser considerados como generalizações.

Os radicais livres são espécies de fundamental importância na química da atmosfera. O radical hidroxila é bastante reativo com várias espécies químicas. Entretanto, devido a sua baixa concentração, sua presença na troposfera só foi confirmada recentemente (Baird, 1995, p.114). A importância do HO[•] decorre do fato de ser ele e não o O₂ que inicia a reação com todos os gases naturalmente presentes na atmosfera tais como aqueles emitidos pelas fontes naturais referidas acima (Tabela II.1). Sem o HO[•] e seus correlatos reativos como o HOO[•], esses gases não seriam eficientemente removidos da troposfera (Baird, 1995, p.114).

Para uma discussão sobre a reatividade química na troposfera seria interessante ter em mente o seguinte “quadro”:

"Gases são emitidos para a atmosfera e encontram outras espécies presentes (ou que se formam posteriormente) em condições de reagir com eles".

Nesse sentido, é importante conhecer as principais espécies químicas presentes, os tipos de ligações químicas existentes, os tipos de radiação solar predominantes e as condições a que estão submetidas esse conjunto de entidades na troposfera. Um breve levantamento indica a presença das seguintes espécies:

- ◆ Oxigênio atômico – O
- ◆ Oxigênio molecular – O₂
- ◆ Radicais livres – HO•
- ◆ Compostos com ligações múltiplas – SO₂, N₂, NO, NO₂, CO
- ◆ Compostos totalmente oxidados – CO₂, SO₃, N₂O₅
- ◆ Compostos sem ligações múltiplas, contendo hidrogênio – CH₄, NH₃, H₂S, CH₃Cl
- ◆ Luz solar,

e das espécies que podem se formar depois (como desdobramento das reações iniciais):

- ◆ Radicais peroxi do tipo HOO•
- ◆ Radicais peroxi do tipo H₃COO•
- ◆ Radicais não peroxi do tipo H₃-C=O•
- ◆ Radicais não peroxi do tipo R-C•=O

Uma abordagem completa sobre principais espécies químicas presentes, os tipos de ligações químicas existentes, os tipos de radiação solar predominantes pode ser encontrada em Baird (1995, p.130-149) ; Manahan (p.282-290, 293-298, 328-329, 338-345, 385-400) e Birks (p.235-238, 239, 242-246).

2.3.3 – Oxidação do Metano na Troposfera

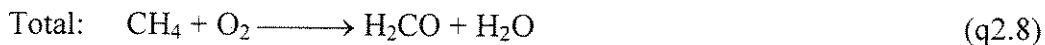
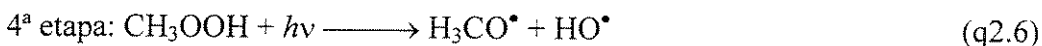
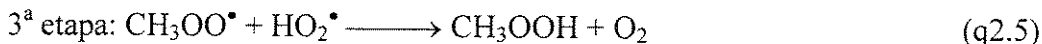
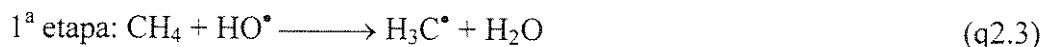
Algumas das principais reações envolvidas na oxidação de hidrocarbonetos no ar podem ser entendidas considerando-se a oxidação do metano. Este encontra-se bastante disperso e é o hidrocarboneto mais comum na atmosfera.

De acordo com Baird (1995, p.168) a concentração de metano atmosférico quase que dobrou em relação a era pré-industrial. E quase todo esse aumento ocorreu no século XX. Historicamente (antes de 1750) a concentração de metano era constante em 0,75 ppm, mas nos anos 80 do século passado (século XX) subiu cerca de 0,6% ao ano atingindo 1,7 ppm. Nos anos 70 do mesmo século, a taxa de aumento era quase duas vezes esse valor, e não se sabe com certeza porque ela decresceu, sendo hoje quase igual a zero (Baird, 1995 p.169).

Embora menos importante do que o CO₂ no aquecimento global, o metano também contribui para esse efeito porque suas moléculas absorvem uma fração maior dos fótons térmicos da região do infravermelho que passam por ele, do que o faz o CO₂ (Baird, 1995, p.168).

Para facilidade de compreensão da oxidação do metano na troposfera, é conveniente, primeiramente, uma abordagem do processo em uma atmosfera limpa. Em seguida estender-se-á o problema aos casos do ar poluído. No primeiro caso (ar limpo) o metano provém, em grandes quantidades, de processos de decaimento anaeróbico, e do uso de carvão, óleo e, especialmente, de gás natural (Baird, 1995, p.119).

Para o metano tem-se:



As reações q2.3 a q2.8 mostram que sob atmosfera limpa, a oxidação de metano conduz à produção de formaldeído. A decomposição fotoquímica deste leva à formação de CO, o qual por sua vez provoca destruição de O₃ (ver abordagem no Capítulo VII).

Supondo-se que as fontes citadas emitam o gás metano a uma taxa aproximadamente constante, v (moléculas $\text{cm}^{-3} \text{s}^{-1}$), é possível calcular a concentração do gás no ar (supondo ausência de reação) em função do tempo. Entretanto, o CH_4 é consumido pelos radicais livres, $\text{HO}^\bullet$, tal que a sua concentração no ar é função também da quantidade daqueles radicais na atmosfera. Então a variação da concentração de CH_4 com o tempo (a sua taxa de variação $v = (d/dt)[\text{CH}_4]$) é função, simultaneamente, de suas taxas de produção e de consumo. Isto pode ser representado pela equação do processo elementar (não global)



De acordo com a teoria das colisões para as velocidades de reação, a taxa de consumo de CH_4 é proporcional às concentrações de CH_4 e de $\text{HO}^\bullet$, sendo dada por:

$$v_1 = k_1 [\text{CH}_4][\text{HO}^\bullet] \quad (2.1)$$

onde k_1 é uma constante de proporcionalidade que engloba os fatores frequência de colisões, fração efetiva de colisões e o fator estérico (ou de geometria) (Russel, 1994, p.650-656).

Em condições de estado estacionário (taxa de produção de CH_4 , v , igual a taxa de consumo de CH_4 , v_1) a concentração de CH_4 será dada por:

$$[\text{CH}_4]_{\text{ee}} = v / (k_1 [\text{HO}^\bullet]) \quad (2.2)$$

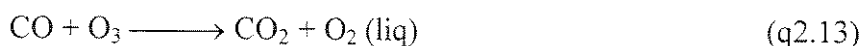
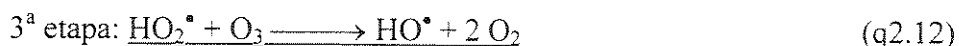
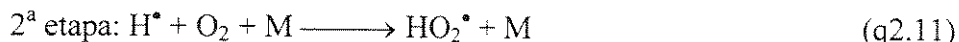
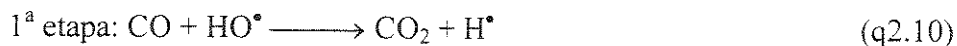
Essa concentração de CH_4 entretanto é apenas para o processo elementar que ocorre na 1ª etapa, equação (q2.9), na qual forma-se uma outra espécie reativa, o radical simples $\text{CH}_3^\bullet$.

Uma inspeção no esquema indica que há, na 1ª etapa, o consumo de um radical $\text{HO}^\bullet$ na reação com CH_4 , e na 4ª etapa, a produção de outro radical $\text{HO}^\bullet$. Estão dadas, portanto, as condições de se reiniciar o ciclo.

Visto dessa forma é possível entender porque a concentração de CH_4 no passado (antes de 1750) permanecia constante em um determinado valor (0,75 ppm).

Dentre as espécies traço atmosféricas que reagem com o radical $\text{HO}^\bullet$ encontram-se o monóxido de carbono (CO), o metano (CH_4), o dióxido de enxofre (SO_2), o sulfeto de hidrogênio (H_2S) e o óxido nítrico ($\text{NO}^\bullet$). A remoção de $\text{HO}^\bullet$ da troposfera é freqüentemente realizada através de reações com o metano ou com o monóxido de carbono (Manahan, 1994, p.288), o que reforça a adequação do mecanismo acima.

A oxidação do CO na troposfera limpa pode ser colocada em umas poucas etapas (Birks, 1998, p.235).



A principal diferença entre esses dois ciclos para as oxidações de CH₄ e de CO, na troposfera, reside no fato de que no ciclo do CH₄ não há destruição de ozônio enquanto no de CO o processo leva ao consumo do poluente na razão de um para um.

É importante enfatizar que os dois ciclos apresentados acima o são para o caso de uma atmosfera limpa. No caso de um ar impregnado de poluentes, como se verá adiante, ocorre formação de ozônio tanto no ciclo do metano quanto no do monóxido de carbono.

A lentidão da etapa inicial na oxidação do metano, e a quantidade crescente de gás liberado da superfície da Terra, têm levado a um aumento na concentração atmosférica de CH₄ na atmosfera nos últimos tempos (Baird, 1995, p.123). (Ver discussão sobre essa afirmação no Capítulo VII).

2.3.4 – A Química da Baixa Troposfera e a Poluição do Ar

Essa seção abordará, por meio de suas subseções, os principais aspectos relacionados com a química da baixa troposfera. Nessa tese o termo baixa troposfera é interpretado como a região imediatamente próxima à superfície da Terra. Assim os termos baixa troposfera, superfície e "nível solo" poderão ser usados com o mesmo significado dentro de um mesmo contexto. Outrossim, será reservado o uso do termo troposfera inferior quando se fizer referência à divisão da troposfera em suas três camadas convencionais (superior, média e inferior).

Devido a importância que o "smog" fotoquímico assume no problema da poluição do ar, ele será discutido detalhadamente.

2.3.4.1 – Origem e Ocorrência de "Smog"

O "smog" fotoquímico (ou "smog" oxidante) tem uma longa história. Para introduzir a discussão, talvez um bom começo seja tentar desvendar a origem e o significado do termo "smog". Ele foi usado originalmente para descrever uma desagradável combinação de "smoke" (fumaça) com "fog" (névoa) impregnada com dióxido de enxofre, que atingia Londres antigamente, quando carvão de alto teor de enxofre era o combustível primário usado na cidade. Tal mistura é marcada pela presença de dióxido de enxofre, um composto redutor, caracterizando-a como um "smog" redutor ("reducing smog" ou "sulfurous smog"), diferentemente do "smog" atual, oxidante (Manahan, 1994, p.381).

Embora Manahan (1994) refira-se a certas manifestações da população de Los Angeles sobre irritação nos olhos causada por poluição (de origem antropogênica) do ar em tempos tão remotos quanto 1868, parece que o fenômeno foi primeiro observado e caracterizado como "smog" fotoquímico tal como é hoje conhecido, a partir dos anos 40 do século XX naquela cidade (Manahan, 1994, p.381; Baird, 1995, p.82).

O "smog" fotoquímico é algumas vezes referido como "uma camada de ozônio no lugar errado" porque, ironicamente, a poluição da atmosfera pelo homem, tende a fazer decrescer a quantidade de ozônio da estratosfera onde altas concentrações são desejáveis e a aumentá-la na troposfera onde ele é indesejável devido a sua toxicidade às plantas e animais (Birks, 1998, p.234; Baird, 1995, p.82).

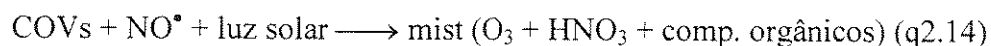
O ozônio, O_3 , é um dos mais importantes gases traço na atmosfera. Na estratosfera, como já visto, ele funciona como blindagem à radiação ultravioleta de comprimento de onda entre aproximadamente 200 e 320 nm. A radiação ultravioleta na região UV-B (280-320 nm) é particularmente maléfica devido a sua forte absorção por proteínas, ácidos nucleicos, e outras moléculas biológicas. Assim, existe, por um lado, uma grande preocupação pela diminuição dos níveis de ozônio na estratosfera, resultante de emissões antrópicas de halocarbonos e outros compostos. Por outro, a preocupação, no sentido de se evitar a formação de ozônio na troposfera, causada também por emissões de origem antrópica.

As emissões dos poluentes monóxido de carbono, hidrocarbonetos e óxidos de nitrogênio aumentam dramaticamente a produção oxidante na troposfera. Como será visto adiante, os principais reagentes originais em um episódio de "smog" fotoquímico são o

óxido nítrico, $\text{NO}^\bullet$, e hidrocarbonetos (compostos contendo somente átomos de carbono e hidrogênio) não queimados, emitidos ao ar como poluentes a partir de motores de combustão interna (Manahan, 1994, p.382-386; Baird, 1995, p.82-83; Birks, 1998, p.235).

Hoje há consenso de que a presença de hidrocarbonetos gasosos no ar urbano deve-se também à evaporação de compostos orgânicos voláteis (COVs).

O outro ingrediente fundamental no "smog" fotoquímico é a luz solar. Ela aumenta a concentração de radicais livres que participam dos processos químicos de formação de "smog" (Manahan, 1994, p.389; Baird, 1995, p.83; Birks, 1998, p.235). Os produtos finais são o ozônio, o ácido nítrico, e compostos orgânicos parcialmente oxidados e, em alguns casos, nitrogenados (Baird, 1995, p.83).



Quando um combustível é queimado na presença de ar, a altas temperaturas, parte do N_2 e O_2 no ar combina-se para formar óxido nítrico, $\text{NO}^\bullet$, o qual é gradualmente oxidado a $\text{NO}_2^\bullet$, em um período que varia de minutos a horas, dependendo da concentração desses gases (Baird, 1995, p.84)



Os pequenos níveis de NO_x no ar limpo (não poluído) resultam, em parte, da ocorrência da reação acima em ambientes muito energéticos causados por relâmpagos e, em parte da liberação de NO_x e NH_3 de fontes biológicas. A reação entre N_2 e O_2 possui alta energia de ativação, então é lenta, podendo ser considerada desprezível, exceto à temperaturas muito altas como as alcançadas nos modernos motores de combustão de veículos (particularmente à grandes velocidades) e em usinas geradoras de energia (Baird, 1995, p.84).

Em suma, a ocorrência de um episódio de "smog" fotoquímico está diretamente relacionada com a existência de algumas pré-condições (Manahan, 1994, p.389; Baird, 1995, p.84; Birks, 1998, p.235):

- a) Deve haver tráfego substancial de veículos para que ocorra a emissão suficiente de $\text{NO}^\bullet$, hidrocarbonetos, e outros COVs no ar.

- b) Deve existir calor e ampla luz solar para que as reações cruciais, algumas delas fotoquímicas, ocorram a uma taxa (velocidade) rápida.
- c) Finalmente, deve haver relativamente pouco movimento de massas de ar para não ocorrer dispersão dos reagentes.

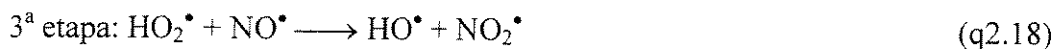
É interessante notar que devido ao transporte atmosférico de longo alcance de poluentes primários e secundários, muitas áreas não geradoras de emissões estão sujeitas a episódios regulares de altos índices de ozônio e outros oxidantes do "smog" no "nível solo". Dois fatores devem ser levados em conta aqui. O primeiro é que essas áreas (por exemplo rurais) podem ficar no caminho de tais massas de ar poluído. O segundo é que eles podem ironicamente registrar níveis de ozônio até mais altos do que o das áreas urbanas próximas, porque nestas últimas algum ozônio é eliminado por reação com o óxido nítrico proveniente das emissões dos veículos (Baird, 1995, p.85).

2.3.4.2 – A Oxidação de Hidrocarbonetos no "Smog" Fotoquímico

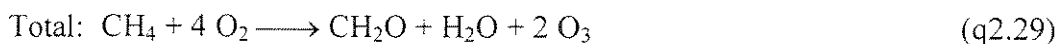
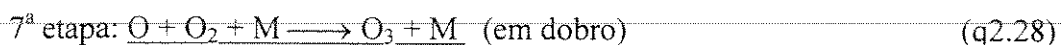
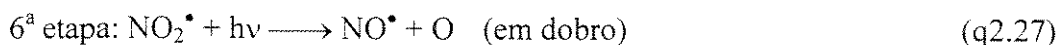
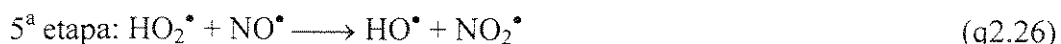
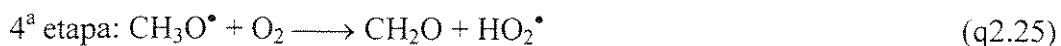
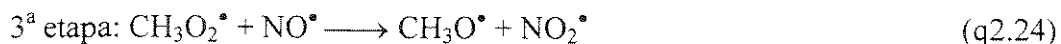
Não obstante a grande complexidade dos processos, as características mais importantes do fenômeno de "smog" fotoquímico podem ser entendidas considerando-se somente umas poucas categorias principais de reações. Elas não diferem muito daquelas que ocorrem no ar limpo como já discutidas para a oxidação do metano na troposfera limpa (Baird, 1995, p.125).

Inicialmente, para efeito de comparação com a seção 2.3.3, na qual se discutiu o processo de oxidação do metano e do monóxido de carbono na atmosfera limpa, é oportuno apresentar as etapas na oxidação do metano e do CO no ar poluído.

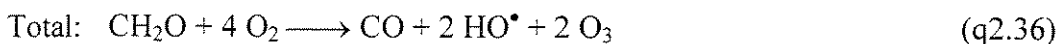
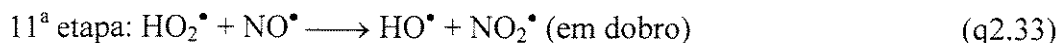
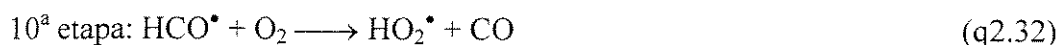
Em uma atmosfera poluída a sequência de reações para o monóxido de carbono resulta em produção de ozônio (Birks, 1998, p.235).



Nas mesmas condições, a sequência para o metano também resulta em produção de ozônio e formaldeído (Birks, 1998, p.236), como mostra a equação q2.29.



O aldeído, CH_2O , formado na 4ª etapa é decomposto fotoquimicamente ampliando a sequência de passos no mecanismo, o qual resultará, no final, em moléculas adicionais de ozônio (Birks, 1998, p.236).



O monóxido de carbono formado nesse ciclo reage com radicais $\text{HO}^\bullet$ para, em cinco etapas, formar uma molécula de ozônio, tal que no ciclo completo, há formação de cinco moléculas de ozônio para cada molécula de metano oxidada.

Resumindo, se existe grande emissão de poluentes como o CO e o CH_4 , em presença de alta concentração de óxidos de nitrogênio, para cada CO oxidado há formação de uma molécula de ozônio, e para cada CH_4 oxidado há formação de 5 moléculas de

ozônio, num total, ao final do ciclo, de 6 (seis) moléculas de O_3 quando existe emissão dos dois poluentes.

Uma análise imediata do problema poderia sugerir a diminuição das emissões de CH_4 porque é ele que produz mais ozônio por molécula que sofre oxidação. Porém, não obstante isto seja necessário, a redução de sua emissão apenas, não é o caminho que produz efeitos benéficos mais significativos. Conforme observa Birks (1998), no caso do CO, as reações da 3ª etapa para o ar limpo (equação q2.12) e para o ar poluído (equação q2.18), são diferentes. No caso de ar limpo, o radical $HO_2^\bullet$ reage com o ozônio destruindo-o, porém no caso do ar poluído, o radical $HO_2^\bullet$ reage com o $NO^\bullet$. Um dos produtos dessa última reação é o $NO_2^\bullet$ que se decompõe fotoquimicamente produzindo átomos de oxigênio, escassos na troposfera, mas necessários para, em reação com O_2 , gerar O_3 .

Então o ozônio e o óxido nítrico competem pelo radical $HO_2^\bullet$. O fator determinante dessa competição repousa nas emissões de NO_x . Para emissões em pequenas quantidades, a concentração de NO_x é baixa tal que ocorre destruição de ozônio. No caso de emissões em grandes quantidades, a concentração de NO_x é alta o bastante para acarretar produção de ozônio.

Por isso pode-se imaginar o óxido nítrico como uma "chave química" na formação oxidante. É possível calcular a concentração de $NO^\bullet$ necessária para que a oxidação de CO possa "chavear" o mecanismo do estado de destruição para o de produção de ozônio, o que ocorre quando as taxas de reação de $HO_2^\bullet + O_3$ e de $HO_2^\bullet + NO^\bullet$ forem iguais (Birks, 1998, p.236, Baird, 1995, p.433).

$$k_{NO+HO_2} [NO^\bullet][HO_2^\bullet] = k_{O_3+HO_2} [O_3][HO_2^\bullet] \quad (2.3)$$

$$[NO^\bullet] = (k_{O_3+HO_2} / k_{NO+HO_2}) [O_3] \quad (2.4)$$

A constante de taxa para a reação $HO_2^\bullet + NO^\bullet$ é cerca de 4000 vezes maior do que o valor para a reação $HO_2^\bullet + O_3$ (Howard e Evenson, 1977, apud Birks, 1998, p.236). Substituindo esse valor na equação 2.4 constata-se que a concentração crítica de $NO^\bullet$ para "chavear" da destruição para a produção de ozônio é de aproximadamente 1/4000 da concentração de O_3 (Birks, 1998, p.236).

Somente após uma quantidade expressiva de $\text{NO}^\bullet$ ter sido convertida a $\text{NO}_2^\bullet$, é que uma reação de "ozônio" urbano pode ser significativa (Baird, 1995, p.126; Birks, 1998 p.235). Se essa situação (grande quantidade de $\text{NO}_2^\bullet$) não for alcançada, o $\text{NO}^\bullet$ e o O_3 se auto destroem na reação entre eles, formando um ciclo nulo (Baird, 1995, p.127; Birks, 1998, p.236-237), já que não se formarão átomos de oxigênio em quantidade necessária.



.....Total.....ciclo nulo.....

Cabe lembrar, entretanto, que esse dióxido de nitrogênio, $\text{NO}_2^\bullet$, em atmosferas poluídas, é um produto da oxidação do óxido nítrico, $\text{NO}^\bullet$.

Portanto, o precursor do $\text{NO}_2^\bullet$ é fundamental no processo, tanto quanto é o próprio $\text{NO}_2^\bullet$ na equação do problema de formação do ozônio. Por essa razão, o $\text{NO}^\bullet$ é considerado a "chave química" para a formação oxidante (Birks, 1998, p.236).

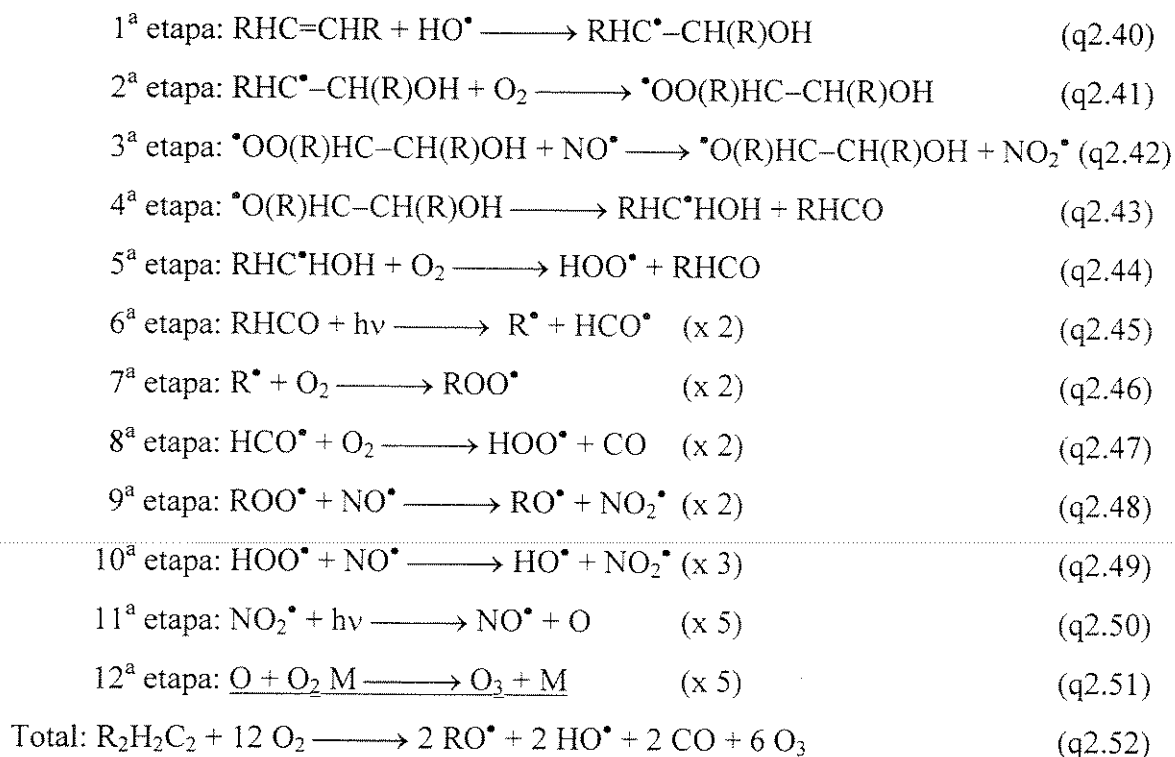
É importante observar que o dióxido de nitrogênio é a única fonte significativa de oxigênio atômico que pode dar origem ao ozônio (Baird, 1995, p.126).

Nesse ponto deve ser registrado que em 1992 uma pesquisa (Paulson and Seinfeld, 1992) apontou a formação de átomos de oxigênio a partir do 1-octeno.

Para os COVs mais reativos, isto é, aqueles hidrocarbonetos com uma ligação dupla $\text{C}=\text{C}$, o caso mais simples é o etileno, C_2H_4 .

Para se ter uma melhor idéia do resultado é oportuno escrever o mecanismo em suas diversas etapas como foi feito para o metano e para o monóxido de carbono.

O ponto de partida é uma molécula de hidrocarboneto geral de fórmula $\text{RHC}=\text{CHR}$ que reage com um radical hidroxila, $\text{HO}^\bullet$, na primeira etapa



As etapas de números 4 e 5 formam um aldeído cada uma. Então, na 6ª etapa leva-se em conta a soma das duas anteriores, 2 RHCO, devendo essa etapa ser multiplicada por dois. Na 5ª etapa também se forma um radical HOO•. Este somado com os outros dois produzidos na 8ª etapa totalizam 3 HOO•, os quais são consumidos por 3 NO• na 10ª etapa, gerando 3 NO₂•. Os outros resultados parciais podem ser encontrados sem dificuldades, tal que a reação global mostra a formação de 6 moléculas de ozônio para cada hidrocarboneto que reagiu com HO• no início do ciclo.

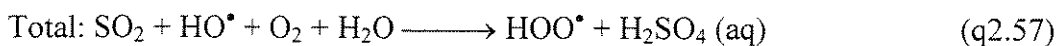
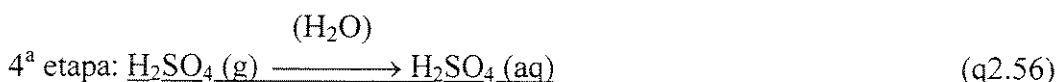
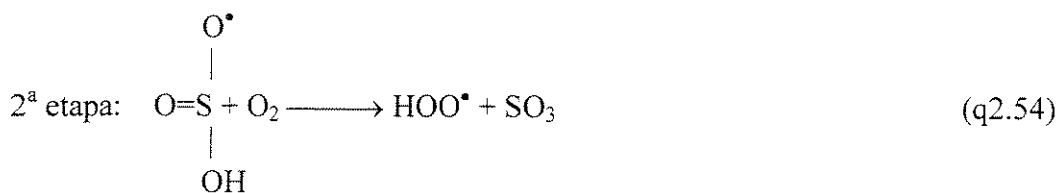
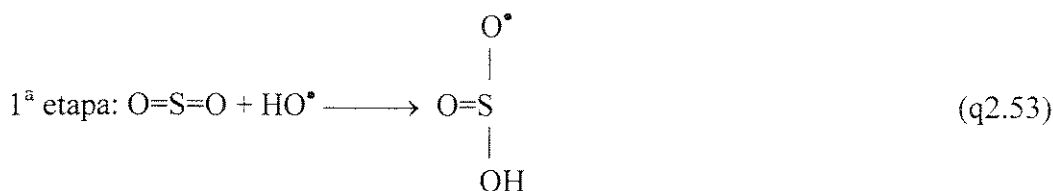
Considerando-se que duas moléculas de CO são produzidas durante o processo, e que elas irão gerar duas moléculas de ozônio (ver ciclo para o CO, o qual gera um ozônio por molécula de CO em atmosfera poluída), a oxidação de um hidrocarboneto como o mostrado no exemplo, produz 8 moléculas de ozônio durante um processo de "smog" fotoquímico.

Deve-se observar que no caso da oxidação do metano ocorre a formação de 5 moléculas de ozônio para cada molécula de metano que inicia o processo sob as mesmas condições. Então, pode-se concluir que o aldeído formado é o principal responsável pela produção de ozônio adicional (ver discussão no Capítulo VII).

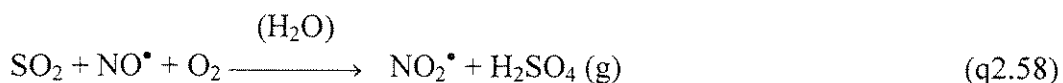
2.3.4.3 – Oxidação do SO₂ na Atmosfera

Em dias de céu claro ou com poucas nuvens, o mecanismo predominante para a conversão de SO₂ a H₂SO₄ é uma reação em fase gasosa homogênea. A sequência ocorre em várias etapas e tem início na reação com HO•.

Como o SO₂ possui ligações múltiplas mas não possui hidrogênio, espera-se que ocorra adição de HO• (1ª etapa), para formar uma molécula estável, o SO₃, por remoção do átomo de hidrogênio pelo O₂, originando um radical HOO• (2ª etapa). Finalmente, SO₃ se combina com moléculas de H₂O (g) dando origem a H₂SO₄ (g), o qual em presença de vapor d'água, forma um aerosol (3ª e 4ª etapas) cujas gotas são soluções de ácido sulfúrico (Baird, 1995, p.133-134).



O radical HOO• pode, entretanto, reagir com NO• formando NO₂• e recuperando HO•, sendo esta reação global considerada como uma co-oxidação de SO₂ e NO•, catalisada por HO• (Baird, 1995, p.134), como mostra a equação q2.127.



Mesmo sob concentrações representativas de HO• na atmosfera limpa, pouco SO₂ atmosférico é oxidado por esse mecanismo. A taxa é muito mais rápida sob “smog”

fotoquímico devido às concentrações significativamente mais altas de HO[•] (Baird, 1995, p.134).

Os mecanismos envolvendo o monóxido de carbono, o metano e os hidrocarbonetos não metano, acrescidos do dióxido de enxofre, são abordados no Capítulo VII (Discussões).

2.4 – A Poluição Atmosférica na Região Metropolitana de São Paulo (RMSP)

A RMSP possui área de aproximadamente 8000 Km² com uma topografia dominada por colinas que variam de 650 a 1200 m. Considerada um dos maiores conglomerados do mundo, com uma população de cerca de 17 milhões de pessoas, a RMSP possui, ainda, um grande parque industrial além de uma frota de veículos leves em torno de $7,2 \times 10^6$ (CETESB, 1998, p.2). Somem-se a estes um considerável número de veículos de grande porte e médios. Se esses números podem ser interessantes para determinadas atividades da vida humana, sob o ponto de vista ambiental representam fontes de emissão de poluentes que se traduzem em reflexos negativos para a “saúde” da RMSP. Outras fontes na RMSP constituem-se de queima de resíduos, movimentação e estocagem de combustíveis.

Considerando-se as fontes de emissões acima, a atmosfera da RMSP está sujeita a uma constante exposição a vários compostos com diferentes graus de agressividade ao homem e ao ambiente em geral. Assim, com base em uma legislação mais próxima da adotada pela EPA (Environmental Protection Agency) dos EUA, porém menos rigorosa do que a recomendada pela OMS (Organização Mundial de Saúde), a cidade de São Paulo tem, estabelecida por meio da Resolução CONAMA nº 3 de 28/06/1990 (âmbito nacional), padrões de qualidade do ar que servem de referência para avaliação da qualidade do ar respirado pela população local.

Os principais poluentes emitidos ao ar naquela área estão sujeitos a monitoração constante durante vinte e quatro horas por dia, por meio de rede telemétrica distribuída em 25 Estações de Medição na RMSP. Os resultados dessas medições são lançados em uma base de dados, a qual fornece suporte para os relatórios periódicos sobre a qualidade do ar na RMSP.

2.4.1 – Padrões de Qualidade do Ar

O estabelecimento de padrões de qualidade do ar constitui-se na forma legal para definir limites máximos para as concentrações de determinados componentes da mistura atmosférica de modo a garantir a proteção da saúde e o bem estar das pessoas (CETESB, 1998, p.9).

Tais padrões são divididos em primários e secundários. Padrões primários são aqueles cuja ultrapassagem dos valores limite possam afetar a saúde da população, e

secundários aqueles cuja ultrapassagem dos valores limite possam causar o mínimo efeito adverso sobre o bem estar da população, assim como o mínimo dano à fauna, à flora, aos materiais e ao meio ambiente em geral (CETESB, 1998, p.9).

Os parâmetros regulamentados são os seguintes: *partículas totais em suspensão, fumaça, partículas inaláveis, dióxido de enxofre, monóxido de carbono, dióxido de oxigênio e ozônio*. Os padrões nacionais de qualidade do ar fixados pela Resolução CONAMA nº 3 de 28/06/90 são apresentados na Tabela II.2.

Tabela II.2: Padrões nacionais de qualidade do ar (Res. CONAMA nº3 de 28/06/90)

Amostragem		Primário	Secundário	Medição
		µg/m ³	µg/m ³	
Partículas Totais em Suspensão	24 horas ¹	240	150	Amostrador de
	MGA ²	80	60	Grandes Volumes
Partículas	24 horas ¹	150	150	Separação
Inaláveis	MAA ³	50	50	Inercial/Filtração
Fumaça	24 horas ¹	150	100	Reflectância
	MAA ³	60	40	
Dióxido de Enxofre	24 horas ¹	365	100	Pararosanilina
	MAA ³	80	40	
Dióxido de Nitrogênio	1 hora	320	190	Quimiluminescência
	MAA ³	100	100	
Monóxido de Carbono	1 hora ¹	40000	40000	Infravermelho não
	8 horas ¹	10000	10000	Dispersivo
Ozônio	1 hora ¹	160	160	Quimiluminescência

1-Não deve ser excedido mais do que uma vez ao ano. 2-Média geométrica anual. 3-Média aritmética anual.
Fonte: CETESB, 1998, p.10.

A Tabela II.2 mostra a preocupação com o NO₂, o CO e o O₃ manifestada por meio da amostragem, isto é, enquanto para as outras variáveis os níveis para a amostragem no dia (24 horas) não devem ser ultrapassados mais do que uma vez no ano, para as três mencionadas, os níveis não devem ser ultrapassados no momento da amostragem (a cada hora) mais do que uma vez ao ano.

CAPÍTULO III

METODOLOGIA

3.1 – INTRODUÇÃO À ANÁLISE MULTIVARIADA

Os métodos estatísticos podem ser classificados num sentido mais amplo em dois grandes grupos. No primeiro encontram-se aqueles voltados para a Análise Univariada, ou seja, concentram a sua preocupação na análise de apenas uma variável aleatória. A análise univariada é, desse modo, bastante limitada quando se deseja observar a influência que determinadas variáveis têm simultaneamente sobre um conjunto de amostras representativas de um fenômeno em estudo.

No segundo grupo estão os métodos que se preocupam com a Análise de Dados Multivariados. Esses métodos são ferramentas poderosas quando se deseja analisar simultaneamente um grande número de variáveis medidas sobre um conjunto de amostras. Em espectroscopia, por exemplo, o número de variáveis é sempre muito maior que o de amostras.

Dentre os métodos de Análise de Dados Multivariados podem ser destacados os dois principais métodos para a redução de dimensão dos espaços das colunas e/ou linhas de uma matriz de dados: a Análise de Componentes Principais (PCA) e a Análise de Fatores (FA). Incluem-se aí outros métodos tais como a Análise Discriminante, a Análise de Agrupamentos, a Análise de Correlação Canônica e o Escalamento Multidimensional. Alguns autores se referem a análise de componentes principais e ao escalamento multidimensional como métodos de ordenamento, isto é, métodos que produzem novos

eixos contra os quais um conjunto de objetos de interesse pode ser projetado. Outros métodos de ordenamento são a Análise de Coordenadas Principais, a Análise de Correspondência (Manly, 1997, p.12-15), e a Análise de Fatores.

Os métodos de Análise de Dados Multivariados foram desenvolvidos por pesquisadores da área de psicometria. A gênese desses métodos remete ao século XIX, porém a sua crescente popularidade só se tornou possível com o advento do computador. Na verdade, o "núcleo duro" da disciplina álgebra linear, e mesmo boa parte da análise multivariada de dados, já se encontrava pronta à espera de máquinas capazes de processar tais dados a uma velocidade minimamente compatível com a parte mais rudimentar daquelas disciplinas. Porém, somente a partir dos anos sessenta do século XX é que os métodos de Análise de Dados Multivariados foram paulatinamente ganhando espaço nos diferentes campos de pesquisa que tradicionalmente não os usava, em uma razão que crescia em ritmo semelhante ao do desenvolvimento de máquinas de processamento de dados cada vez mais potentes.

Especificamente no campo da química, a Análise de Dados Multivariados foi introduzida através do método PCA. Na verdade, nos anos 1960, Malinowski introduziu o método sob o nome de Análise de Fatores Principais e, depois de 1970, um grande número de aplicações vem sendo publicado (Wold, 1987). O que deve ter motivado Malinowski a denominar o método de Análise de Fatores Principais foi o fato de ele tê-lo feito para aplicações à matrizes de covariância e de correlação, como ele mesmo explicita em sua introdução (Malinowski, 1991, p.1-4) para explicar a filosofia subjacente ao método.

Deve ser entendido, então, que, quando ele usa o termo "Fatores", está fazendo referência ao uso de matrizes de covariância e/ou de correlação. O termo "Principais" refere-se à semelhança com a análise PCA.

Como será visto adiante, PCA e FA, embora forneçam uma redução da dimensão do espaço ocupado pelas colunas e/ou linhas de uma matriz de dados, são métodos que apresentam diferenças sutis nem sempre percebidas durante a sua utilização.

O desenvolvimento da química analítica instrumental, paralelamente ao da ciência e tecnologia de computadores, permitiu aos pesquisadores da área de química a utilização de avançados métodos matemáticos e estatísticos em seu próprio campo de trabalho,

culminando com o surgimento do termo QUIMIOMETRIA, cunhado por Svante Wold no início dos anos 1970 (Einax et alli, 1997, p.1).

3.1.1 – Notação e Terminologia

Será usada a notação padrão adotada em boa parte das publicações sobre o assunto. Escalares são denotados por letras minúsculas em itálico ($a, a_1, x, x_1, c, \dots$), vetores por letras minúsculas em negrito ($\mathbf{x}, \mathbf{x}_1, \mathbf{v}, \mathbf{v}_1, \dots$), matrizes por letras maiúsculas em negrito ($\mathbf{X}, \mathbf{X}_1, \mathbf{A}, \mathbf{A}_2, \mathbf{B}, \mathbf{C}, \dots$) e arranjos multimodo por maiúsculas sublinhadas e em negrito ($\underline{\mathbf{X}}, \underline{\mathbf{X}}_1, \underline{\mathbf{A}}, \underline{\mathbf{A}}_2, \underline{\mathbf{B}}, \underline{\mathbf{C}}, \dots$). Os elementos de uma matriz e de uma estrutura multimodo são, respectivamente, indicados por x_{ij} e x_{ijk} , cujos índices variam desde 1 até a sua versão alfabética maiúscula, como a seguir: $i=1, \dots, I$ e $j=1, \dots, J$ para matrizes, e $i=1, \dots, I; j=1, \dots, J$ e $k=1, \dots, K$ para arranjos em três modos. Vetores linha serão representados como vetores coluna transpostos, $\mathbf{x}^T$ (T significa transposto). Quando os vetores coluna de uma matriz forem respectivamente tomados como vetores linha, a nova matriz será indicada como transposta, $\mathbf{X}^T$. Qualquer forma diferente de notação será assinalada. Quando não o for, poderá ser entendida a partir do contexto.

3.1.2 – Análise Exploratória de Dados

A caracterização de um determinado material é realizada através da geração de dados analíticos sobre amostras desse material*. A utilização de computadores de grande capacidade de armazenamento e de processamento de dados, aliada à informação extraída de sofisticada instrumentação analítica, especialmente sob a forma de dados espectrais, coloca o analista diante de um conjunto de dados cuja extensão e complexidade extrapolam a viabilidade de uso do grande e bem desenvolvido campo da estatística univariada. Esta preocupa-se apenas com a distribuição unidimensional de pontos, sendo útil, portanto, para fins confirmatórios, mas não para propósitos exploratórios, já que não se obtém nenhuma informação a respeito de como os diferentes tipos de medida estão relacionados (Sharaf *et alli*, 1986, p.181).

*A menos que explicitamente indicado de outra forma, esse texto considerará sempre que dados populacionais não estão disponíveis.

As técnicas estatísticas podem ser divididas em análises exploratórias e em análises confirmatórias. No primeiro caso, o objetivo é descrever o comportamento das variáveis da matriz de dados e procurar padrões e relações características do sistema, que não são, portanto, atribuíveis ao mero acaso. A análise exploratória está voltada para a compressão de dados e aproximação de matrizes (técnicas que são essencialmente o mesmo processo) pela troca das colunas ou linhas da matriz por um número muito menor destas que são representativas dos dados originais (Jobson, 1992, p.352). Em análise confirmatória, certas hipóteses ou modelos construídos são testados para determinar se os dados os suportam. A qualidade de um modelo é freqüentemente medida pela sua capacidade de fornecer boas previsões.

Em estudos exploratórios os pesquisadores incluem tantas variáveis quanto possível, de modo a assegurar que nenhuma variável relevante seja omitida. Na verdade, a natureza do estudo, ou seja, a necessidade de se analisar simultaneamente o maior número possível de fatores suspeitos de exercer influência sobre um dado fenômeno, é que acaba impondo a inclusão de um grande número de variáveis, cuja análise não é possível com o emprego de técnicas univariadas, já que a matriz resultante é geralmente grande e difícil de analisar, independentemente do nível de correlação entre as variáveis.

Nesse sentido, os métodos multivariados oferecem visibilidade na análise e interpretação dos dados oriundos de espaços multidimensionais.

3.1.2.1 – A Organização dos Dados

Antes de se iniciar a análise propriamente dita dos dados, deve-se ter em mente alguns procedimentos de modo a orientar adequadamente o trabalho em mãos, facilitando a realização das etapas sucessivas da análise. Dentre esses procedimentos destacam-se a construção da *matriz de dados*, a *visualização dos dados* e a *adequação dos dados* à análise, esse último levado a cabo pelo *pretratamento de dados*. Esse assunto é discutido em tutorial de Ferreira e colaboradores (Ferreira *et alli*, 1999).

3.1.2.1.1 – A Matriz de Dados

Os dados coletados, lançados em tabelas, são transportados para uma estrutura mais adequada à aplicação dos métodos de análise multivariada. Tal estrutura recebe o nome de “Matriz de Dados” e é, em geral, representada por $\mathbf{X}$. Nessa matriz, os objetos ou amostras são representados pelas linhas, sendo os valores das medições de cada variável sobre as amostras, representados pelas colunas.

Assim, para uma determinada matriz, $\mathbf{X}$, cada entrada, x_{ij} , corresponde à medição realizada para a j -ésima variável sobre o i -ésimo objeto (amostra) como mostra a Figura 3.1.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1J} \\ x_{21} & x_{22} & \cdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & x_{ij} & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots & \ddots & \vdots \\ x_{I1} & x_{I2} & \cdots & x_{Ij} & \cdots & x_{IJ} \end{bmatrix}$$

Fig.3.1: A matriz de dados

3.1.2.1.2 – Visualização dos Dados

A visualização dos dados é importante para se obter informações preliminares sobre os mesmos. Esse procedimento permite avaliar as faixas de valores em que as variáveis se distribuem e qual o grau de correlação entre elas, auxiliando na escolha do tipo de pretratamento e na identificação de objetos (amostras) anômalos (“*outliers*”).

3.1.2.2 – Pretratamento de Dados I

O pretratamento tem por finalidade a adequação dos dados registrados pelo conjunto dos instrumentos de medição aos métodos de análise de dados.

A adequação dos dados pode ser conduzida por meio de operações realizadas nas linhas e/ou colunas da matriz. Quando realizada nas linhas, a operação é denominada de

“Transformação nos Dados”, enquanto o procedimento nas colunas é denominado de “Preprocessamento de Dados” (Martens e Naes, 1989; Malinowski, 1991; Beebe *et alli*, 1998). Deve-se, todavia, chamar a atenção para o fato de que o preprocessamento também pode ser realizado nas linhas da matriz. Rigorosamente falando, o tipo de dados determinará o procedimento mais adequado. Uma generalização que pode ser aceita é a de que a transformação se processe nos objetos ou amostras (isto é, nas linhas), enquanto o preprocessamento seja realizado tanto nas variáveis quanto nos objetos (isto é, nas colunas e/ou nas linhas). Em quimiometria pode-se considerar que, em geral o preprocessamento é aplicado apenas às colunas da matriz. Isto porque, diferentemente das ciências sociais onde as variáveis são categóricas, em quimiometria são, via de regra, variáveis físico-químicas (contínuas).

3.1.2.2.1 – Transformação nos Dados

Envolve diferentes propósitos tais como a eliminação de ruídos ou a linearização de uma resposta instrumental. As operações mais usadas em quimiometria são:

- a) Derivada 1ª
- b) Derivada 2ª
- c) “Alisamento”
- d) Logaritmo
- e) Normalização
- f) Multiplicação
- g) Subtração

3.1.2.2.2 – Preprocessamento de Dados I

Os tipos mais comuns de preprocessamento são:

- a) Dados centrados na média (*cm*)– essa operação apenas desloca o centróide do conjunto de dados para a origem do sistema de eixos.

$$x_j = \frac{1}{I} \sum_{i=1}^I x_{ij}^* \quad (3.1)$$

$$x_{ij(cm)} = x_{ij} = x_{ij}^* - x_j \quad (3.2)$$

onde x_j representa o valor médio tomado sobre a variável j (também em b e c abaixo).

$x_{ij(cm)} = x_{ij}$ representa o objeto centrado na média e x_{ij}^* o valor original do objeto.

b)Dados escalados pela variância (sv) – nesse caso cada variável é ponderada pelo recíproco de seu desvio padrão.

$$s_j^2 = \frac{1}{I-1} \sum_{i=1}^I (x_{ij}^* - x_j)^2 \quad (3.3)$$

$$x_{ij(sv)} = \frac{x_{ij}}{s_j} \quad (3.4)$$

c)Dados autoescalados – nesse tipo de preprocessamento os dados são centrados na média e simultaneamente escalados pela variância.

$$x_{ij} = \frac{x_{ij}^* - x_j}{s_j} \quad (3.5)$$

d)Dados escalados pela variação (sr) ou (“range”) – as variáveis são escaladas segundo seus valores máximo e mínimo, isto é, pela faixa de variação dos mesmos.

$$x_{ij(sr)} = \frac{x_{ij}^* - x_{j(min)}}{x_{j(max)} - x_{j(min)}} \quad (3.6)$$

Todos os tipos de preprocessamento citados acima são sensíveis a amostras anômalas, porém, pelo fato de depender especificamente de dois pontos (os valores de máximo e de mínimo) e não de medidas de tendência geral como média e variância, o escalamento pela variação é, dentre eles, o mais sensível àquelas amostras.

3.1.3 – Análise de Componentes Principais (PCA)/Análise de Fatores (FA)

Esse tópico destina-se a dar uma visão geral sobre os dois métodos básicos e de longe mais usados na redução de dimensionalidade do espaço ocupado por um conjunto de dados, e a mostrar as principais diferenças entre eles. São eles: a Análise de Componentes Principais (PCA) e a Análise de Fatores (FA).

As análises de componentes principais e de fatores, remontam ao início do século. PCA em 1901 com Pearson (Malinowski, 1991, p.2; Jackson, 1991 p.6,420; Jolliffe, 1986, p.5-7), o qual utiliza uma abordagem geométrica aplicada no estudo de planos e retas que melhor se ajustam a um conjunto de pontos no hiperespaço, e a FA em 1904 com Spearman (Jackson, 1991, p.421; Jolliffe, 1986, p.5-7), posteriormente fundamentado (1933) por Hotteling (Malinowski, 1991, p.2; Jackson, 1991, p.421; Jolliffe, 1986, p.5-7), o qual utiliza uma abordagem algébrica envolvendo multiplicadores de Lagrange, e que termina com a solução de um problema de autovalores/autovetores. Essas abordagens apesar de similares guardam algumas diferenças entre si. A principal delas reside na forma algébrica por meio da qual podem ser expressas cada uma das análises, PCA ou FA, como pode ser visto na Figura 3.2, para a k-ésima componente principal em PCA, e para o k-ésimo fator em FA.

Análise de componentes principais:

$$\begin{array}{ccc}
 \text{Cominação linear} & & \\
 \text{PC}_k = p_{k1}x_1 + p_{k2}x_2 + \cdots + p_{kJ}x_J = (p_{k1}, p_{k2}, \cdots, p_{kJ}) \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_J \end{pmatrix} & & (3.7) \\
 \Downarrow & & \Downarrow \\
 \text{Componente principal} & & \text{Variáveis originais}
 \end{array}$$

Análise de fatores:

$$\begin{array}{ccc}
 x_k = a_{k1}f_1 + a_{k2}f_2 + \cdots + a_{kJ}f_J = (a_{k1}, a_{k2}, \cdots, a_{kJ}) \cdot \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_J \end{pmatrix} & & (3.8) \\
 \Downarrow & & \Downarrow \\
 \text{Variável original} & & \text{Fatores}
 \end{array}$$

Fig. 3.2: Comparação usual entre os métodos PCA e FA

Duas questões se colocam de imediato. A primeira refere-se às duas formas diferentes de se expressar as variáveis, e a segunda, refere-se ao significado da expressão "redução de dimensionalidade".

É importante, então, em primeiro lugar, fazer distinção entre variáveis originais (definidas pelos vetores originais de base), e novas variáveis (definidas pelos novos vetores de base). Aqui, a redução da dimensionalidade do sistema é necessariamente acompanhada de uma mudança de vetores de base, operação realizada tanto em PCA quanto em FA.

Em segundo lugar deve-se fazer distinção entre dados originais e variáveis originais. Em terceiro lugar, deve-se fazer distinção entre as diferentes formas de se expressar as variáveis, ou seja, apreender o significado subjacente às aplicações de PCA e de FA.

Dados originais são as entradas da matriz em consequência das medições realizadas sobre as variáveis originais. Em outras palavras, quando ocorre uma mudança de base, as

variáveis deixam de ser as originais, mas podem ser obtidas por meio dos dados originais. Entretanto, por meio de pré-tratamento, é possível realizar operações sobre as entradas originais de dados (portanto sobre os dados originais) sem que se faça qualquer mudança de base. Ou seja, os vetores que representam os objetos (e consequentemente suas projeções sobre as variáveis originais) passam a ter comprimentos diferentes daqueles dos dados originais, porém pertencendo ao sistema de base original.

A constatação acima conduz a duas situações diferentes, com suas respectivas implicações, no que diz respeito à operação de escalamento. Na primeira, o escalamento pode ser realizado sobre os autovetores, portanto na nova base, originando (Jackson, 1991, p.63): a) autovetores ortonormais produzindo componentes principais com variâncias iguais aos correspondentes autovalores; b) autovetores com coeficientes nas mesmas unidades dos dados, produzindo componentes principais com variâncias iguais ao quadrado dos autovalores; c) autovetores cujas componentes principais possuem variância unitária. Nesses casos, as PCs obtidas contém exatamente a mesma quantidade de informação.

Na segunda, quando o escalamento é realizado sobre os dados originais (como já visto na seção sobre pré-tratamento), os resultados dependem do método empregado e, inclusive da ausência de escalamento (Jackson, 1991, p.63).

Isto significa que os parâmetros obtidos dependem da matriz de dados, de sua transformação e do seu pré-processamento (Wold, 1987).

Resumindo: escalamento (ou ausência dele) refere-se a procedimentos realizados nas entradas de dados (escalamento de dados originais, ou de autovetores), enquanto mudança de base refere-se a procedimentos realizados no conjunto de base do sistema (variáveis originais para novas variáveis, as PCs). Esse assunto é explorado em detalhes na Seção 3.2 (ver Jackson, 1991, p.16-18, 63-75).

Após a aplicação de uma Análise de Componentes Principais (PCA), obtém-se novas variáveis em um espaço de dimensão reduzida em relação àquele ocupado pelos dados na matriz original (após pré-tratamento ou não). Essas novas variáveis são, na verdade, um conjunto de novos eixos (definidos por novos vetores de base) nos quais as entradas correspondentes às linhas da matriz (objetos ou amostras), são projetadas.

Na Análise de Componentes Principais (PCA), essas novas variáveis são expressas como uma combinação linear das variáveis originais enquanto na Análise de Fatores as

variáveis originais é que são expressas em função de novas variáveis (ou novos eixos), como mostra a Figura 3.2. Os aspectos algébricos sobre esse assunto são discutidos na Seção 3.2.

Um outro fator importante na análise é a definição do sistema de base da decomposição. Ou seja, se o espaço das linhas é definido como sistema de eixos, as entradas correspondentes às colunas são projetadas nele. Caso contrário, se o espaço das colunas é definido como sistema de eixos, as entradas correspondentes às linhas da matriz são projetadas nesses eixos. Também aqui a organização dos dados na matriz ajuda a orientar a decomposição.

A segunda questão (subseção 3.1.3, p.48) refere-se à redução de dimensionalidade.

Essa questão não é tão trivial quanto possa parecer. Em geral, na utilização dos métodos de Análise Multivariada confunde-se o significado dessa operação.

A dimensão de um sistema é dada pela quantidade de seus vetores de base. Se um sistema é composto por um conjunto de variáveis medidas sobre várias amostras, e for definido que a base do sistema é formada por aquelas variáveis, reduzir a dimensionalidade do sistema não significa necessariamente eliminar algumas das variáveis que o constituem. Embora isso possa ser realizado, poderá acarretar em perda de informação. Então, na redução, deve-se evitar a eliminação das variáveis definidas pela base do sistema (exceção feita a casos de identificação inequívoca de variáveis que não contém informação significativa, tais como análise de dados em QSAR – Quantitative Structure Activity Relationships – (Dunn, 1989), QSPR – Quantitative Structure Propriety Relationships –, ou linha de base em dados espectroscópicos).

O que se faz, na verdade, é aplicar uma rotação ao conjunto de base do sistema de modo que todas as variáveis que o compõem possam captar mais informação do que no sistema não “rotacionado”. Ao se realizar essa operação obtém-se um conjunto de novas variáveis que são combinações lineares das variáveis originais. Assim o sistema pode ser descrito por um número menor de novas variáveis (eixos) sem ocorrer qualquer perda de informação. Inclusive ruídos não identificados *a priori*, se houver, estarão contidos nesse novo sistema de base. A arte em Análise Multivariada é descobrir quantos novos eixos (novas variáveis) contém informação significativa. A Seção 3.2 aborda o assunto mais detalhadamente, quando são discutidos os significados dos termos base, dimensão e modo.

A discussão que se faz adiante toma por base vários autores que se dedicaram a estudar o assunto tais como, Wold *et alli* (1987), Geladi e Kowalski (1986), Beebe e Kowalski (1987), Thomas (1994) e Brown (1995).

Tendo definido superficialmente os conceitos de base e dimensão de um sistema, suponha-se um conjunto de dados obtidos por meio de medições de, por exemplo, $J=2$ variáveis sobre $I=5$ objetos.

Organize-se esses dados em uma matriz, $\mathbf{X}(I \times J)$ e proceda-se a uma análise de componentes principais (PCA). A Figura 3.3 mostra o sistema original (variáveis x_1 e x_2) com as coordenadas (projeções dos objetos) nesse sistema original, e a redução da dimensão para uma única variável (a nova variável, ou a componente principal), PC_1 , com as novas coordenadas dos pontos do espaço, usualmente denominadas escores (as projeções dos objetos nesse novo sistema). Aqui, por questões didáticas, a segunda componente principal, PC_2 , também é mostrada para evidenciar a mudança dos valores dos escores (projeções) dos objetos ao se passar do sistema de base original para o novo. O efeito causado sobre as variâncias dos dados em cada base é discutido na Seção 3.2.

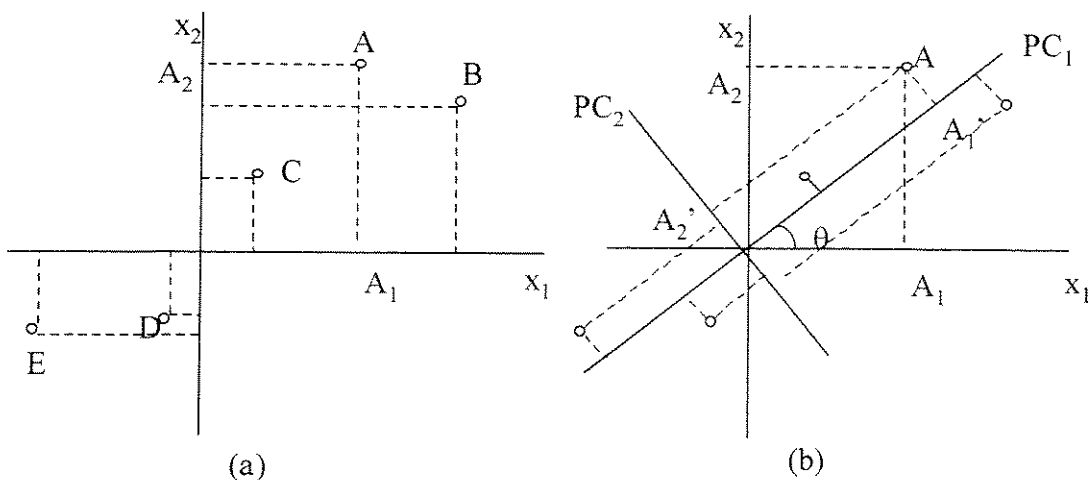


Fig.3.3: a) Projeções de cinco pontos, representativos de cinco objetos, na base que contém as variáveis originais (x_1, x_2). b) Projeções desses cinco pontos na nova base que contém as novas variáveis (PC_1, PC_2). São destacadas as projeções do ponto A nos dois sistemas, para que se observe a variabilidade dos dados em cada base.

Observando-se o gráfico (a) verifica-se que as projeções (A_1, A_2) do ponto A sobre os eixos (variáveis originais) x_1 e x_2 , possuem valores aproximadamente iguais. No gráfico (b), entretanto, a projeção (A_1') do mesmo ponto A sobre o primeiro novo eixo (nova variável) PC_1 , possui valor bem maior do que (A_2') no segundo novo eixo (segunda nova variável) PC_2 . A rotação do sistema levou a um novo espaço no qual a variabilidade dos dados está praticamente restrita a apenas um dos novos eixos: a primeira componente principal PC_1 . As projeções dos objetos no novo sistema de base mostram que os pontos estão mais espalhados na direção de PC_1 . Isto significa que o sistema pode ser descrito por apenas uma dimensão, isto é, aquela que contém PC_1 .

Aqui a redução deu-se de duas para uma dimensão, porém a utilidade de PCA (e também de FA) se faz sentir, de fato, quando o sistema possui muitas variáveis (centenas ou milhares) podendo ser reduzido para apenas duas ou três componentes (ou fatores). O procedimento descrito acima mostra que PCA, na verdade, vale-se da colinearidade presente nas variáveis originais para trazer à superfície um comportamento latente que não está nelas evidenciado.

Todavia, deve-se chamar a atenção para o fato de que o processo de obtenção das novas variáveis não muda a natureza dos dados presentes na matriz sob análise. Isto é, tal procedimento transfere para o novo sistema todas as características presentes nos dados, inclusive o ruído experimental. Como mencionado anteriormente, a arte está em se descobrir quantas componentes principais contém, de fato, informação significativa.

A redução da dimensionalidade do sistema “concentra” as variações sistemáticas presentes nos dados em um número menor de novos eixos. Isto significa que as últimas componentes contém principalmente variações não sistemáticas tais como o erro experimental, que devem ser eliminados do “modelo”. A Figura 3.4 mostra, de forma esquemática, como PCA pode ser usada para separar a estrutura inerente aos dados, do ruído experimental, a partir da uma matriz de dados, $\mathbf{X}$.

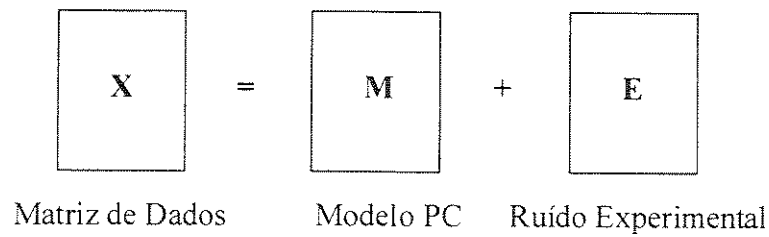


Fig.3.4: A matriz de dados, $\mathbf{X}$, pode ser considerada como sendo constituída de uma estrutura intrínseca aos dados (Modelo PC), $\mathbf{M}$, e outra parte contendo o ruído experimental, $\mathbf{E}$.

O significado da Figura 3.4 pode ser expresso de forma algébrica (ver equação 3.49 na subseção 3.2.2, p72; e equação 3.51, no tópico 3.2.5.2, p82). De forma simplificada, pode ser escrito como:

$$\mathbf{X}(I \times J) = [\text{Matriz de Escores}] [\text{Matriz de Pesos}] + [\text{Matriz de Resíduos}] \quad (3.9)$$

onde $\mathbf{X}(I \times J)$ = matriz de dados, [Matriz de Escores] = contém os escores para as componentes principais, [Matriz de Pesos] = contém os pesos para as componentes principais, e [Matriz de Resíduos] = matriz de resíduos deixados pelo modelo.

3.1.4 – Análise Confirmatória de Dados

Como já discutido em outras partes desse texto, os avanços tecnológicos que resultaram no interfaceamento de computadores à instrumentação analítica permitiram ao pesquisador a geração de uma quantidade de dados cada vez maior. Proporcionou inclusive a criação dos aparelhos hifenados, nos quais diferentes instrumentos podem extrair simultaneamente dados referentes a uma mesma amostra. Essa crescente quantidade de dados tem que ser transformada em informação para que seja de utilidade ao químico e à sociedade em geral. Então, os métodos capazes de analisar grande quantidade de dados tornaram-se uma necessidade para o analista e para o pesquisador de um modo geral.

Os principais métodos de calibração multivariada são: a) Quadrados Mínimos Clássico – CLS (Haaland *et alli*, 1985; Haaland, 1987); b) Quadrados Mínimos Inverso – ILS, também chamado de Regressão Linear Múltipla MLR – (Brown, 1982; Brown, 1988); c) Regressão por Componentes Principais – PCR (Fredericks *et alli*, 1985; Haaland e Thomas, 1988; Haaland, 1988; Cahn, 1988; Donahue *et alli*, 1988; Seasholtz *et alli*, 1989); e d) Regressão por Quadrados Mínimos Parciais – PLS (Lindberg *et alli*, 1983; Fuller *et alli*, 1988; Haaland e Thomas, 1988; Haaland, 1988; Cahn, 1988; Naes e Martens, 1985; Naes e Martens, 1988; Donahue *et alli*, 1988; Seasholtz *et alli*, 1989).

Essa tese não utiliza os métodos mencionados acima porque a abordagem se situa no plano da análise exploratória de dados, não sendo, portanto, discutidos aqui os métodos de calibração multivariada.

3.2 – BASE MATEMÁTICA EM ANÁLISE MULTIVARIADA

3.2.1 – Base, Dimensão e Modo

Com o objetivo de situar o estudo de análise multivariada no contexto das operações da disciplina álgebra linear, essa subseção preocupa-se em discutir os conceitos de base e dimensão de um sistema, e em discorrer sobre as diferenças e os significados dos termos dimensão e modo do mesmo. A abordagem concentra-se no “terreno” das operações algébricas, as quais dão sustentação às possibilidades dos resultados obtidos na análise e interpretação das aplicações quimiométricas tanto no uso das matrizes convencionais quanto no uso das estruturas multimodo.

Uma base para um espaço de vetores é formada por um conjunto de vetores linearmente independentes que geram tal espaço (Strang, 1988, p.83-85; Noble e Daniel, 1988, p.188,196-197; Watkins, 1991, p.170). Por sua vez, um espaço de vetores é aquele no qual são possíveis as operações de adição de quaisquer dois vetores e a multiplicação por escalares (Strang, 1988, p.63; Noble e Daniel, 1988, p.182,). Um espaço contendo J vetores coluna é representado por $\mathbb{R}^J$.

Os vetores de um conjunto são linearmente independentes se apenas a combinação trivial é igual a zero. Isto é, dado um conjunto de vetores $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_k$, se a combinação trivial, $c_1\mathbf{v}_1, c_2\mathbf{v}_2, c_3\mathbf{v}_3, \dots, c_k\mathbf{v}_k = 0$ ocorrer somente para $c_1 = c_2 = c_3 = \dots = c_k = 0$, os vetores $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_k$ são linearmente independentes (Strang, 1988, p.80; Noble e Daniel, 1988, p.189-193;). Uma base é, portanto, constituída por esses vetores que geram o espaço.

Um espaço tridimensional, por exemplo, é representado por $\mathbb{R}^3$. Dois vetores quaisquer de tal espaço descrevem, geometricamente, um plano passando através da origem. Tal plano é, portanto, ele próprio um espaço de vetores. Isto ilustra uma das mais fundamentais idéias na teoria da álgebra linear: *o subespaço* (Strang, 1988, p.64-67; Noble e Daniel, 1988, p.184-185; Watkins, 1991, p.169). O referido plano é um subespaço do espaço original $\mathbb{R}^3$ (Strang, 1988, p.64).

Não obstante um espaço de vetores não possua uma única base, quaisquer duas bases do mesmo contém o mesmo número de vetores. Esse número é denominado dimensão do espaço de vetores (Strang, 1988, p.85-87; Noble e Daniel, 1988, p.197-199; Watkins, 1991, p.170). Para dados arranjados em matrizes, interessa saber qual é a

dimensão do espaço de vetores: o número de vetores coluna, o de vetores linha ou outro número diferente.

Os dados a analisar são organizados em matrizes. Assim, deseja-se abordar o significado que, espaço de vetores, base e dimensão desse espaço, além de outros termos, adquirem em matrizes.

Dada uma matriz $\mathbf{A}(I \times J)$, os vetores coluna de $\mathbf{A}$ possuem I componentes. Eles são um subespaço de $\mathcal{R}^I$. Os vetores linha possuem J componentes e são um subespaço de $\mathcal{R}^J$.

Dependendo dos valores de I e J , três situações são possíveis:

$$I = J \quad (\text{i})$$

$$I > J \quad (\text{ii})$$

$$I < J \quad (\text{iii})$$

No caso de $I = J$, os subespaços são de mesmo “tamanho”. Quando $I > J$, o espaço das colunas é um subespaço de $\mathcal{R}^I$. Quando $I < J$, o espaço das linhas é um subespaço de $\mathcal{R}^J$. É necessário, então, um procedimento que oriente o caminho para se encontrar a dimensão e a base do sistema sob análise.

Mesmo que não se saiba ainda qual é a base do sistema pode-se saber qual é a dimensão do mesmo, a partir da definição do termo base. Como ela é formada por um conjunto de vetores linearmente independentes, o número de vetores do conjunto deve dar a dimensão do sistema. Adicionalmente, se esses vetores puderem ser identificados, fica também identificada a base do sistema.

Procura-se um número que dê informações sobre a dimensão do sistema. O caminho tem início em um processo chamado Eliminação Gaussiana (Strang, 1988, p.12-16, 71-77, 90-95; Noble e Daniel, 1988, p.83-91; Watkins, 1991, p.41-42). Por meio da eliminação chega-se a uma matriz triangular superior e encontra-se também o número procurado, o qual é denominado *posto* (“rank”, R) da matriz $\mathbf{A}$ (Strang, 1988, p.80, Noble e Daniel, 1988, p.143; Watkins, 1991, p.159-166).

O posto foi introduzido de forma puramente computacional e representa o número de pivôs no processo de eliminação (ou equivalentemente, é igual ao número de linhas não zero da matriz triangular superior resultante da eliminação). Entretanto, o posto não é apenas resultado da aplicação de um algoritmo para efeito de operações computacionais.

Ele tem um significado. Ele leva em conta o número de linhas genuinamente independentes na matriz $\mathbf{A}$ (Strang, 1988, p.80).

O pivôs aparecem em algumas das colunas da matriz triangular superior, isto é, eles identificam as colunas linearmente independentes na matriz. Essas colunas correspondem às mesmas colunas da matriz $\mathbf{A}$. Esse número, quantidade de pivôs, fornece a dimensão do espaço das colunas, o qual é igual ao posto, R , da matriz $\mathbf{A}$.

Deve ser observado que a referência a uma matriz $\mathbf{A}(I \times J)$ como tendo dimensões I por J não tem nenhuma relação com o posto, R , da matriz, não obstante o posto possa ser I ou J .

Quando $J = I$ (número de colunas igual ao de linhas), o posto pode ser, no máximo, igual $R = J = I$. Quando $J < I$ (número de colunas menor do que o de linhas), o posto pode ser, no máximo, igual a $R = J < I$. Quando $J > I$ (número de colunas maior do que o de linhas), o posto pode ser, no máximo igual $R = I < J$.

Resumindo: a dimensão do espaço das colunas de $\mathbf{A}$ é igual ao posto R , e é também igual à dimensão do espaço das linhas. O número de colunas independentes é igual ao número de linhas independentes. Uma base para o espaço das colunas de $\mathbf{A}$ é formado pelas R colunas de $\mathbf{A}$, as quais correspondem às colunas da matriz triangular superior (eliminação de Gauss) que contém pivôs (Strang, 1988, p.93).

O termo modo surge na análise multimodo em substituição ao termo dimensão. Em geral, ao se referir a uma matriz $\mathbf{A}(I \times J)$, diz-se que a mesma tem dimensões I por J . Por exemplo, a matriz $\mathbf{A}(5 \times 3)$ tem dimensões 5 por 3. Esses valores referem-se ao “tamanho físico” da matriz, ou seja, 5 linhas e 3 colunas. Como uma matriz pode ser quadrada ou retangular, essas figuras geométricas podem ser usadas para representá-las na forma de esquemas.

Quando os dados a analisar são arranjados em várias matrizes, as quais são empilhadas para formar um cubo ou paralelepípedo de dados, por exemplo, o termo modo é introduzido para fazer referência a cada uma das dimensões da figura geométrica cúbica (ver discussão na subseção sobre análise multimodo).

3.2.2 – Análise de Componentes Principais - PCA

Interpretações filosófica, algébrica e geométrica

A idéia central de Análise de Componentes Principais é a de reduzir a dimensionalidade de um conjunto de dados constituído de um grande número de variáveis interrelacionadas, aqui dada pelo número de colunas J , retendo o máximo possível da variação presente nos dados (Jolliffe, 1986, p.1). Ou seja, representar uma estrutura de dados J -dimensional em uma dimensão menor, usualmente duas ou três (Massart, 1988, p.339), dada pelo pseudo posto. Isto, como já discutido introdutoriamente no Capítulo III, é realizado através da transformação a um novo conjunto de variáveis, as componentes principais, PC's, as quais são não correlacionadas, e em ordem decrescente de retenção da variação presente em todas as variáveis originais.

Essa clara variabilidade dos dados é encontrada fornecendo-se variância máxima à primeira nova variável. A segunda nova variável é então selecionada de tal modo que seja não correlacionada com a primeira e possua a segunda maior variância, e assim sucessivamente. Do ponto de vista geométrico isso significa que as novas variáveis são eixos ortogonais.

Para uma melhor compreensão do problema é conveniente abordá-lo primeiro sob duas dimensões e depois considerá-lo em J dimensões.

Existem várias maneiras (técnicas, métodos, algoritmos etc) diferentes para a obtenção das componentes principais. Dentre elas pode-se destacar o cálculo do ângulo de rotação, a solução do problema de autovalores-autovetores, o algoritmo NIPALS (Wold, 1966, p.411-444; Wold *et alli*, 1983, p.147-188; Geladi e Kowalski, 1986; Wold *et alli*, 1987), a decomposição por valor singular – SVD – (Lawson e Hanson, 1974; Stoer e Bulirsch, 1980).

A abordagem por meio do cálculo do ângulo de rotação apresenta-se como uma das mais simples, do ponto de vista didático, para que seja compreendido o significado da análise PCA. Ela é discutida aqui mais detalhadamente nos seus aspectos algébricos e geométricos, enquanto as outras são apenas mencionadas.

3.2.2.1 – Componente Principal em Duas Dimensões

PCA por meio do cálculo do ângulo de rotação

O significado da redução de dimensionalidade (compressão de dados) referida acima pode ser compreendida considerando-se duas variáveis, x_1 e x_2 , sobre as quais se realizam quatro observações (Figura 3.5). O sistema em estudo é, portanto, constituído de quatro objetos (amostras) e duas variáveis.

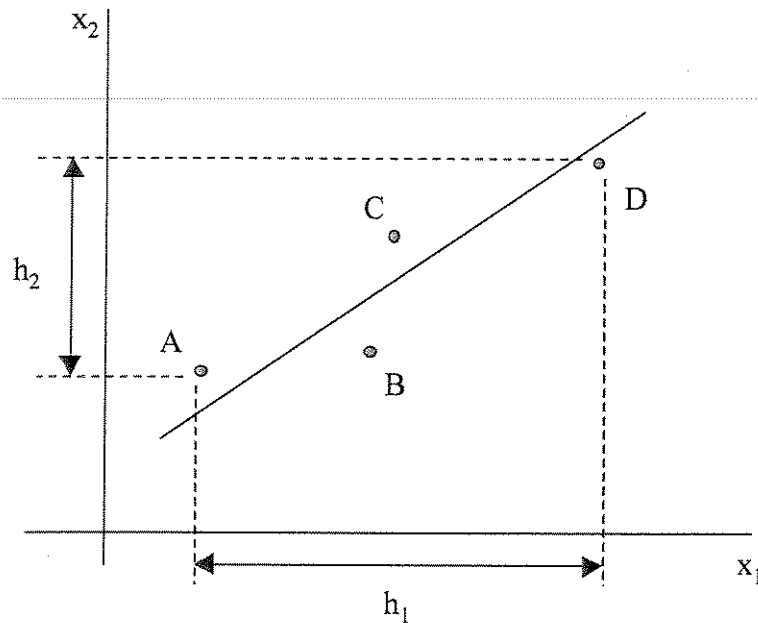


Fig.3.5: Variabilidade dos dados e colinearidade entre as variáveis x_1 e x_2

O gráfico da Fig.3.5 sugere que as variáveis x_1 e x_2 possuem uma alta correlação. Essas variáveis têm um alto grau de redundância (são colineares). Alta colinearidade entre variáveis, reveláveis estatisticamente por medidas de correlação ou covariância, é uma forte indicação de que é possível encontrar uma nova base que expresse melhor a informação presente nos dados. Alta colinearidade entre variáveis implica também que os dados residem em um subespaço do espaço total definido pelas variáveis originais (Brown, 1995).

De fato, na Figura 3.5, verifica-se a existência de uma relação linear entre as variáveis. Além disso, verifica-se que o espalhamento dos pontos na direção de x_1 é um pouco maior do que aquele na direção de x_2 , ($h_1 > h_2$). Por outro lado na Figura 3.6, o

espalhamento ao longo da linha ou eixo PC_1 é maior do que nos dois primeiros casos ($h_3 > h_1 > h_2$).

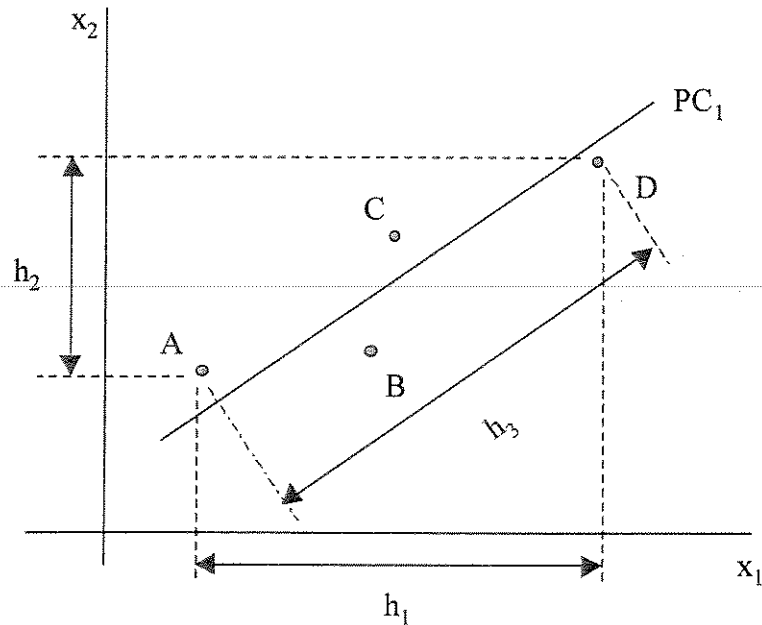


Fig.3.6: Variabilidade nos dados na direção do eixo PC_1 ,

A utilidade dessas constatações, como se vê, é justamente a possibilidade de se poder reduzir o sistema de duas, para apenas uma dimensão; aquela na qual os pontos apresentam a maior variação. Ou seja, PC_1 deve ser escolhido de forma a reter a maior variabilidade possível dos dados. Isto significa tornar máxima a variação ao longo desse novo eixo e torná-la mínima em torno dele.

Considere-se, agora, o ponto “O” como sendo o centróide da nuvem de pontos ao longo do eixo PC_1 (Figura 3.7).

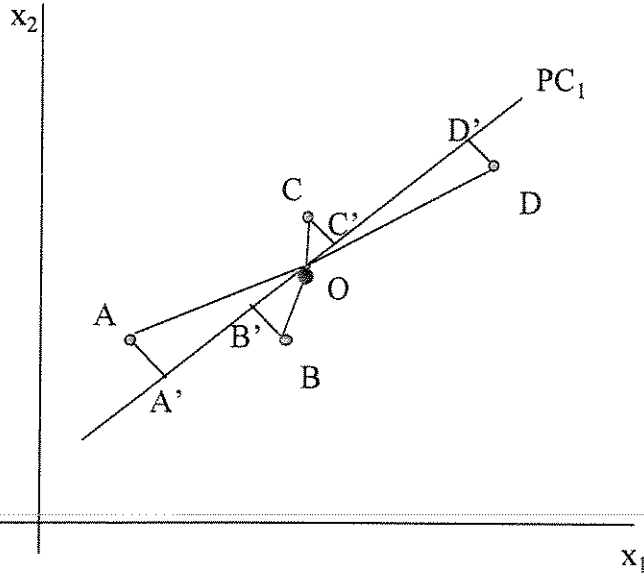


Fig.3.7: Construção do centróide e variabilidade dos pontos ao longo e em torno de PC_1

Tome-se, por exemplo, o ponto A da Figura 3.7. Para o triângulo OAA' , e para cada um dos outros triângulos formados pelos demais pontos e suas projeções, com o centróide, pode-se escrever:

$$(OA)^2 = (OA')^2 + (AA')^2 \quad (3.10)$$

$$(OB)^2 = (OB')^2 + (BB')^2 \quad (3.11)$$

$$(OC)^2 = (OC')^2 + (CC')^2 \quad (3.12)$$

$$(OD)^2 = (OD')^2 + (DD')^2 \quad (3.13)$$

O espalhamento dos pontos ao longo de PC_1 pode ser representado algebricamente como uma soma de quadrados de suas distâncias ao centróide (Massart, 1988, p.339-341).

$$S^2 = (OA)^2 + (OB)^2 + (OC)^2 + (OD)^2 \quad (3.14)$$

$$S^2 = (OA')^2 + (OB')^2 + (OC')^2 + (OD')^2 + (AA')^2 + (BB')^2 + (CC')^2 + (DD')^2 \quad (3.15)$$

Encontra-se o eixo PC_1 tornando máxima a soma dos termos $(OA')^2, \dots, (OD')^2$, a qual representa a variação ao longo de PC_1 , enquanto, simultaneamente, torna-se mínima a soma de $(AA')^2, \dots, (DD')^2$, a qual representa a variação em torno de PC_1 .

Uma operação usual em PCA constitui-se no deslocamento da origem do sistema formado pelas variáveis originais para o centróide da nuvem de pontos (Figura 3.7). Considerar-se-á aqui, as variáveis originais já *centradas na média* ou *corrigidas pela média*, x_1 e x_2 , conforme técnica de pré-processamento já abordada no item 3.12.2.2.

Sejam, então, duas variáveis originais centradas na média representadas pelos eixos x_1 e x_2 . Uma transformação linear geral transforma-as em duas novas variáveis representadas pelos eixos PC_1 e PC_2 .

$$PC_1 = a_{11}x_1 + a_{12}x_2 \quad (3.16)$$

$$PC_2 = a_{21}x_1 + a_{22}x_2 \quad (3.17)$$

Que relação existe entre a transformação linear geral algébrica das variáveis originais x_1 e x_2 , respectivamente, em PC_1 e PC_2 , e sua interpretação geométrica?

Como poderá ser visto, tal transformação linear significa, geometricamente, nada mais do que uma rotação dos eixos originais x_1 e x_2 de um ângulo θ , de modo a se obter o novo sistema formado pelos eixos PC_1 e PC_2 .

Para dar consistência à explanação, dentro do rigor matemático formal, é oportuno introduzir aqui algumas considerações:

1. O espaço ocupado por variáveis e respectivas observações feitas sobre elas é o Euclidiano.
2. O centro do sistema de coordenadas é o vetor $\mathbf{0}^T$ (zero transposto), de coordenadas (0,0) no caso bidimensional.
3. As direções dos eixos representativos das variáveis x_1 e x_2 no espaço bidimensional são caracterizadas por vetores de comprimento unitário, de coordenadas, respectivamente, (1,0) e (0,1).
4. Os vetores unitários acima constituem uma base padrão do sistema no qual são feitas as observações sobre as duas variáveis x_1 e x_2 .
5. Em qualquer base, original ou transformada, do espaço Euclidiano considerado, qualquer variável, quando corrigida pela média, pode assumir, nos eixos

definidores da base, valores no intervalo $(-\infty, +\infty)$ independentemente, e sem perda, de seu significado físico.

A transformação linear pode ser ortogonal (as novas variáveis são necessariamente não correlacionadas) ou oblíqua. Em PCA a transformação será sempre ortogonal (ou ortonormal se, além de ortogonal um ao outro, cada vetor tem comprimento unitário).

O propósito da transformação é o de encontrar uma nova direção na qual, ao longo dela, a variância se torne máxima. Essa nova direção, PC_1 , é mostrada na Figura 3.8, abaixo

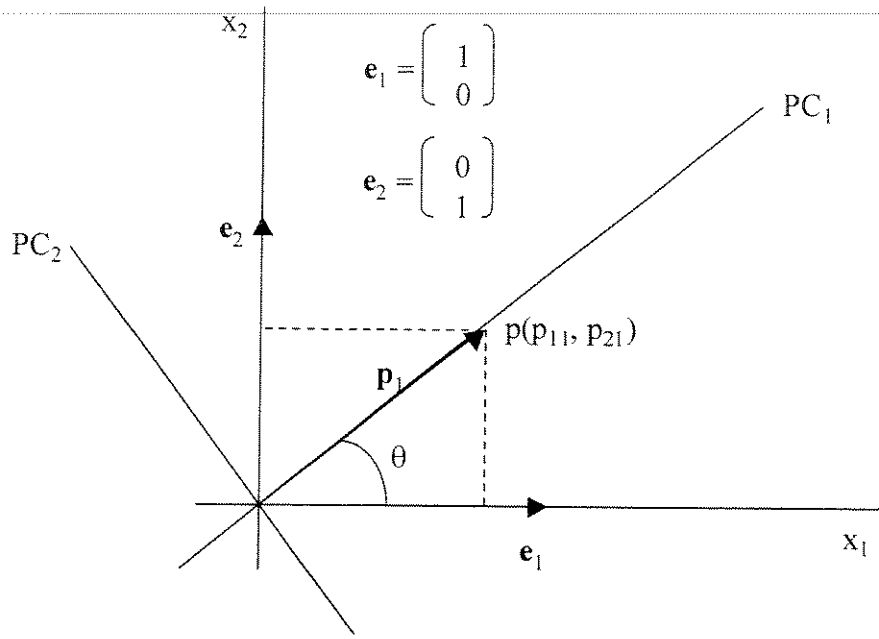


Fig.3.8: Obtenção das direções dos novos eixos

Tome-se sobre o eixo PC_1 , um vetor qualquer, $\mathbf{p}_1$, com extremidade no ponto $p(p_{11}, p_{21})$. As coordenadas de $\mathbf{p}_1$, expressas em termos da base formada pelos eixos x_1 e x_2 , são:

$$\begin{aligned} \mathbf{p}_1 &= (\cos \theta)(\mathbf{e}_1) + (\sin \theta)(\mathbf{e}_2) = (\cos \theta) \begin{bmatrix} 1 \\ 0 \end{bmatrix} + (\sin \theta) \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} \cos \theta \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \sin \theta \end{bmatrix} = \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix} \end{aligned} \quad (3.18)$$

Analogamente, para o vetor $\mathbf{p}_2$, sobre o eixo PC_2 tem-se:

$$\begin{aligned}\mathbf{p}_2 &= (-\sin \theta)(\mathbf{e}_1) + (\cos \theta)(\mathbf{e}_2) = (-\sin \theta) \begin{bmatrix} 1 \\ 0 \end{bmatrix} + (\cos \theta) \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} -\sin \theta \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \cos \theta \end{bmatrix} = \begin{bmatrix} -\sin \theta \\ \cos \theta \end{bmatrix}\end{aligned}\quad (3.19)$$

Nas expressões acima, o vetor $\mathbf{e}_1^T$ tem coordenadas (1, 0) e o vetor $\mathbf{e}_2^T$ tem coordenadas (0, 1). Esses vetores definem a base padrão para os eixos que representam as variáveis originais x_1 e x_2 .

Assim, as combinações lineares

$$\mathbf{p}_1 = (\cos \theta)(\mathbf{e}_1) + (\sin \theta)(\mathbf{e}_2) \quad (3.20)$$

$$\mathbf{p}_2 = (-\sin \theta)(\mathbf{e}_1) + (\cos \theta)(\mathbf{e}_2) \quad (3.21)$$

forneem as coordenadas para os vetores $\mathbf{p}_1$ e $\mathbf{p}_2$ e, portanto, as respectivas direções para os novos eixos, representativos das novas variáveis, PC_1 e PC_2 . Esses novos eixos, (as novas variáveis), são chamados de Componentes Principais, PCs, cujas direções são dadas pelos vetores:

$$\mathbf{p}_1 = \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix} \quad e \quad \mathbf{p}_2 = \begin{bmatrix} -\sin \theta \\ \cos \theta \end{bmatrix} \quad (3.22)$$

os quais podem ser organizados na matriz,

$$\mathbf{P} = (\mathbf{p}_1 \quad \mathbf{p}_2) = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \quad (3.23)$$

Tomando-se a transposta da matriz $\mathbf{P}$, obtém-se:

$$\mathbf{P}^T = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \quad (3.24)$$

Observe-se que a primeira linha da matriz $\mathbf{P}^T$ (o vetor linha $\mathbf{p}_1^T$) fornece os coeficientes para os vetores da base padrão, $\mathbf{e}_1$ e $\mathbf{e}_2$, na transformação linear que permite obter as coordenadas, $(\cos \theta, \sin \theta)$, para o vetor $\mathbf{p}_1$ e, portanto, a direção do primeiro novo

eixo, o qual representa a primeira nova variável, PC_1 . Analogamente, a segunda linha da matriz $\mathbf{P}^T$ (o vetor linha $\mathbf{p}_2^T$) fornece os coeficientes para os vetores da mesma base padrão, na transformação linear que permite obter as coordenadas, $(-\sin \theta, \cos \theta)$, para o vetor $\mathbf{p}_2$ e, portanto, a direção do segundo novo eixo, o qual representa a segunda nova variável, PC_2 .

Nas equações 3.20 e 3.21, denominando-se os coeficientes de $\mathbf{e}_1$ e $\mathbf{e}_2$, $(\cos \theta)$ e $(\sin \theta)$, em $\mathbf{p}_1$, respectivamente de a_{11} e a_{12} ; e $(-\sin \theta)$ e $(\cos \theta)$, em $\mathbf{p}_2$, de a_{21} e a_{22} , tem-se:

$$\mathbf{p}_1^T = a_{11} \mathbf{e}_1 + a_{12} \mathbf{e}_2 \quad (3.25)$$

$$\mathbf{p}_2^T = a_{21} \mathbf{e}_1 + a_{22} \mathbf{e}_2 \quad (3.26)$$

Nas relações acima,

$$a_{11} = a_{22} = \cos \theta \quad (3.27)$$

$$a_{12} = -a_{21} = \sin \theta \quad (3.28)$$

Como a transformação é ortogonal, $\mathbf{p}_1$ é perpendicular a $\mathbf{p}_2$. Então seu produto escalar é igual a zero. E, desde que os vetores $\mathbf{e}_i$ são necessariamente não nulos, a condição de ortogonalidade fica:

$$a_{11} \cdot a_{21} + a_{12} \cdot a_{22} = 0 \quad (i)$$

A equação i mostra que as variâncias ao longo dos eixos PC_1 e PC_2 podem se tornar arbitrariamente grandes se os coeficientes a_{11} , a_{12} e a_{21} , a_{22} forem multiplicados por uma constante qualquer. Então é necessário impor restrições aos coeficientes. A restrição mais conveniente é:

$$\left. \begin{aligned} (a_{11})^2 + (a_{12})^2 &= 1 \\ (a_{21})^2 + (a_{22})^2 &= 1 \end{aligned} \right\} \quad (ii)$$

A condição (i) implica que a nova base, caracterizada por $\mathbf{p}_1$ e $\mathbf{p}_2$ é ortogonal, e a condição (ii) implica que esses vetores têm comprimento unitário.

Então a transformação das variáveis originais, x_1 e x_2 , nas novas variáveis, PC_1 e PC_2 , em duas dimensões pode ser representada por:

$$\mathbf{p}_1^T \mathbf{x} = a_{11}x_1 + a_{12}x_2 = \sum_{j=1}^2 a_{1j}x_j \quad (3.29)$$

$$\mathbf{p}_2^T \mathbf{x} = a_{21}x_1 + a_{22}x_2 = \sum_{j=1}^2 a_{2j}x_j \quad (3.30)$$

onde $\mathbf{p}_2^T \mathbf{x}$ é uma função linear dos elementos de $\mathbf{x}$, a qual tem variância máxima e é não correlacionada com $\mathbf{p}_1^T \mathbf{x}$ (Jolliffe, 1986, p.2).

Resumindo, uma transformação linear operada por uma *rotação de eixos* no “sentido contrário ao dos ponteiros do relógio” (counterclockwise) pode ser efetuada pré multiplicando-se o vetor contendo as variáveis originais pela matriz transformação, $\mathbf{P}^T$.

Representando-se a transformação por meio de matrizes, tem-se:

$$\begin{bmatrix} PC_1 \\ PC_2 \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad (3.31)$$

No sistema constituído pelas variáveis originais, x_1 e x_2 , cada ponto representativo de uma determinada observação (amostra), é identificado pelas coordenadas do vetor que dá a sua posição geométrica na base original. Após a rotação de eixos, não obstante a distribuição de pontos no espaço não sofra qualquer alteração, modificam-se suas coordenadas em relação aos novos eixos, as componentes principais.

O ponto de partida para a transformação linear é a matriz de dados. É conveniente, portanto, expressar as coordenadas dos pontos nos novos eixos, em função das coordenadas (conhecidas) desses pontos nos eixos originais, em operação semelhante à realizada na transformação de variáveis.

Seja, então, $\mathbf{x}_i$ um vetor de pontos das I observações independentes, $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_I$, sobre um vetor de J elementos aleatórios (vetor contendo as $J = 2$ variáveis originais). As coordenadas de $\mathbf{x}_i$ nas duas primeiras componentes principais são dadas por:

$$t_{i1} = \mathbf{p}_1^T \mathbf{x}_i \quad (3.32)$$

$$t_{i2} = \mathbf{p}_2^T \mathbf{x}_i \quad (3.33)$$

onde $i = 1, 2, \dots, I$ e, t_{i1} e t_{i2} representam as coordenadas dos pontos da amostra i (medidas sobre as variáveis x_1 e x_2) nos novos eixos representados por PC_1 e PC_2 .

O vetor $\mathbf{p}_1^T$ torna máxima a variância da amostra, $[1 / (I-1)] \sum_{i=1}^I (t_{i1} - t_{1\text{médio}})^2$ sujeito a restrição de normalização, $\mathbf{p}_1^T \mathbf{p}_1 = 1$. Analogamente, $\mathbf{p}_2^T$ torna máxima a variância de t_{i2} sujeito a $\mathbf{p}_2^T \mathbf{p}_2 = 1$, e sujeito também a que os t_{i2} 's sejam não correlacionados com os t_{i1} 's. Esse processo é uma versão para a amostra da definição de componentes principais que se dá para a transformação de variáveis (Jolliffe, 1986, p.24), isto é, a obtenção das coordenadas das observações nos novos eixos é operada pela mesma matriz transformação.

Assim, uma observação qualquer X de coordenadas (X_1, X_2) na base original formada pelos eixos x_1 e x_2 , terá como coordenadas os valores (t_1, t_2) , na nova base de eixos PC_1 e PC_2 (Fig.3.9).

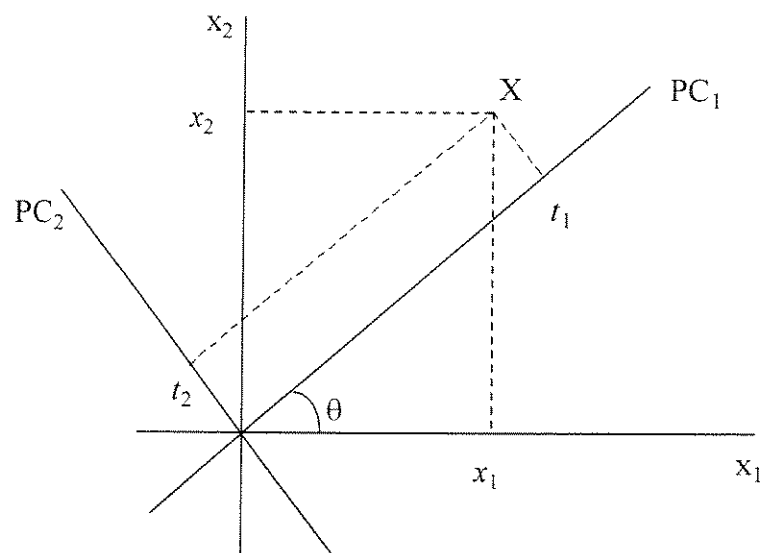


Fig.3.9: Coordenadas de um ponto X no sistema original, (x_1, x_2) , e no novo sistema, (PC_1, PC_2)

Representando-se a transformação por meio de matrizes tem-se:

$$\begin{bmatrix} t_1 \\ t_2 \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \quad (3.34)$$

onde t_1 e t_2 são chamados de scores das observações (amostras) para as componentes principais PC_1 e PC_2 , respectivamente.

A equação 3.34 pode ser escrita como $\mathbf{t} = \mathbf{P}^T \mathbf{x}$, onde $\mathbf{t} = [t_1 \ t_2]^T$ é o vetor contendo os scores das observações, $\mathbf{P}^T$ é a matriz transformação, e $\mathbf{x} = [X_1 \ X_2]^T$ é o vetor que contém as I observações, $\mathbf{x}_i$, sobre as J variáveis, $\mathbf{x}_j$.

Tomando-se a transposta da equação $\mathbf{t} = \mathbf{P}^T \mathbf{x}$ obtém-se

$$\mathbf{t}^T = (\mathbf{P}^T \mathbf{x})^T = \mathbf{x}^T \mathbf{P}, \quad (3.35)$$

a qual em representação matricial pode ser escrita

$$\begin{bmatrix} t_1 & t_2 \end{bmatrix} = \begin{bmatrix} X_1 & X_2 \end{bmatrix} \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \quad (3.36)$$

Deve ser observado que os vetores $\mathbf{t} = [t_1 \ t_2]^T$ e $\mathbf{x} = [X_1 \ X_2]^T$ são vetores coluna, portanto, os vetores $\mathbf{t}^T = [t_1 \ t_2]$ e $\mathbf{x}^T = [X_1 \ X_2]$ são vetores linha. Isto pode ser verificado tomando-se as transpostas dos vetores $\mathbf{t}$ e $\mathbf{x}$ como mostrado abaixo.

$$\mathbf{t}^T = ([t_1 \ t_2]^T)^T = [t_1 \ t_2] \quad (3.37)$$

$$\mathbf{x}^T = ([X_1 \ X_2]^T)^T = [X_1 \ X_2] \quad (3.38)$$

Estendendo-se o procedimento às I observações, $\mathbf{x}_i$, sobre as J variáveis, $\mathbf{x}_j$, obtém-se como resultado a matriz dos scores de todas as observações sobre todas as variáveis, em função da matriz contendo as entradas, dos dados originais, x_{ij} , o qual denota o (i, j) ésimo elemento dessa matriz, tal que x_{ij} é o valor da j -ésima variável medida para a i -ésima observação (Jolliffe, 1986, p.24-25).

Então é mais conveniente expressar a matriz de scores das PCs do seguinte modo:

$$\mathbf{T}_J = \mathbf{X} \mathbf{P}_J \quad (J \leq I, \text{subscrito } J \text{ indica o uso de todas as PCs}) \quad (3.39)$$

Na equação 3.39, $\mathbf{T}$ é uma matriz cujas linhas são os vetores $\mathbf{t}_i^T$ contendo os escores das I observações para as J variáveis no novo sistema constituído pelas componentes principais, as PCs, e $\mathbf{X}$ é uma matriz cujas linhas são os vetores $\mathbf{x}_i^T$ ($i=1, 2, \dots, I$) contendo as I observações para as J variáveis originais x_j ($j=1, 2, \dots, J$).

É interessante observar que a transformação ortogonal acima foi realizada por uma *rotação de eixos* no “sentido contrário ao dos ponteiros do relógio” (counterclockwise), a qual apresenta os mesmos resultados da transformação ortogonal por *rotação de pontos* no “sentido dos ponteiros do relógio” (clockwise) (Green e Carrol, 1978, p.112).

Deve ser apontado, ainda, que expressar a rotação em termos de um único ângulo (no caso o ângulo θ) é restrito a apenas duas dimensões. A notação envolvendo os quatro (para duas dimensões) ângulos, α_{11} , α_{12} , α_{21} , α_{22} , é mais geral porque o conceito de cosenos diretores se aplica para duas, três ou mais direções (Green e Carrol, 1978, p.115).

Nesse caso obter-se-á a seguinte igualdade:

$$\begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} = \begin{bmatrix} \cos \alpha_{11} & \cos \alpha_{21} \\ \cos \alpha_{12} & \cos \alpha_{22} \end{bmatrix} \quad (3.40)$$

onde, em α_{ij} na equação (3.40), i = eixos originais (variáveis originais), e j = novos eixos (novas variáveis, as PCs).

Como já visto, o ângulo de rotação, θ , deve ser calculado em função da direção de maior variabilidade dos dados, isto é, aquela na qual a variância ao longo de PC1 é máxima. Para tal deve-se considerar a matriz geral de variância-covariância das novas variáveis PC_j (Massart, 1988, p.345).

$$\mathbf{S} = \begin{bmatrix} s_{PC_1}^2 & s_{PC_1 PC_2} \\ s_{PC_1 PC_2} & s_{PC_2}^2 \end{bmatrix} \quad (3.41)$$

onde,

$$(s_{PC1})^2 = \text{variância de } PC_1$$

$$(s_{PC2})^2 = \text{variância de } PC_2$$

$$s_{PC1PC2} = \text{covariância entre } PC_1 \text{ e } PC_2$$

A variância da combinação linear das PCs é escrita em termos das variáveis originais,

$$s_{PC1}^2 = (\cos^2\theta) (s_{x1}^2) + (\sin^2\theta) (s_{x2}^2) + 2 \cos\theta \sin\theta s_{x1x2} \quad (3.42)$$

$$s_{PC2}^2 = (\sin^2\theta) (s_{x1}^2) + (\cos^2\theta) (s_{x2}^2) + 2 \sin\theta \cos\theta s_{x1x2} , \quad (3.43)$$

tal que a variância de PC_1 será máxima quando a sua derivada em relação a θ for igual a zero.

$$\frac{\partial}{\partial \theta} (s_{PC1}^2) = 0 \quad (3.44)$$

A partir daí calcula-se o ângulo θ para o qual a variância é máxima:

$$\theta = \frac{1}{2} \arctg[(2s_{x1x2})/(s_{x1}^2 - s_{x2}^2)], \quad \text{se } s_{x1}^2 > s_{x2}^2$$

$$\theta = 90 + \frac{1}{2} \arctg[(2s_{x1x2})/(s_{x1}^2 - s_{x2}^2)], \quad \text{se } s_{x1}^2 < s_{x2}^2$$

$$\theta = 45^\circ \text{ para variáveis positivamente correlacionadas} \quad \text{se } s_{x1}^2 = s_{x2}^2$$

$$\theta = -45^\circ \text{ para variáveis negativamente correlacionadas} \quad \text{se } s_{x1}^2 = s_{x2}^2$$

Conhecido o ângulo de rotação, θ , ficam estabelecidas as direções para os novos eixos, as PCs, podendo ser calculadas as coordenadas das observações (amostras) na nova base formada pelas componentes principais. No caso de mais de duas variáveis, o raciocínio pode ser estendido via abordagem algébrica porque, não obstante a existência de significado geométrico, é extremamente difícil a sua visualização e trabalhosa a sua realização via cálculo do ângulo de rotação.

Não obstante o procedimento para encontrar as PCs via cálculo do ângulo de rotação seja, sob o ponto de vista didático, mais simples para se compreender o significado da redução de dimensionalidade, a decomposição da matriz $\mathbf{X}$ é, via de regra, realizada por meio da decomposição por valor singular, SVD (ver tópico 3.2.5.2).

3.2.3 – Distinção Entre PCA e FA

Na seção 3.1.3 foram discutidos alguns dos principais aspectos filosóficos, geométricos e algébricos relacionados à decomposição de matrizes de dados por meio da aplicação de Análise de Componentes Principais e de Análise de Fatores. Não obstante a discussão introdutória naquele capítulo, ainda não parece visível, “palpável”, que PCA e FA são procedimentos, de fato, diferentes. Esse tópico destina-se a tentar explicitar, de forma algébrica, tais diferenças entre eles.

Para efeito da exposição a seguir, suponha-se que J observações sejam realizadas sobre I objetos constituindo um sistema formado por I objetos e J variáveis (originais). Para tornar a comparação o mais simples possível, sejam $J \leq I$ e $W < J$, tal que PCA e FA podem ser representados, respectivamente, como:

PCA

$$\begin{aligned} PC_1 &= p_{11}x_1 + p_{12}x_2 + \cdots + p_{1J}x_J \\ PC_2 &= p_{21}x_1 + p_{22}x_2 + \cdots + p_{2J}x_J \\ &\vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\ PC_J &= p_{J1}x_1 + p_{J2}x_2 + \cdots + p_{JJ}x_J \end{aligned} \tag{3.45a}$$

ou,

$$PC_k = p_{k1}x_1 + p_{k2}x_2 + \cdots + p_{kJ}x_J \tag{3.45b}$$

ou,

$$PC_k = \sum_{j=1}^J p_{kj}x_j \tag{3.45c}$$

FA

$$\begin{aligned} x_1 &= a_{11}f_1 + a_{12}f_2 + \cdots + a_{1W}f_W \\ x_2 &= a_{21}f_1 + a_{22}f_2 + \cdots + a_{2W}f_W \\ &\vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\ x_J &= a_{J1}f_1 + a_{J2}f_2 + \cdots + a_{JW}f_W \end{aligned} \tag{3.46a}$$

ou,

$$x_k = a_{k1}f_1 + a_{k2}f_2 + \cdots + a_{kW}f_W \tag{3.46b}$$

ou,

$$x_k = \sum_{w=1}^W a_{kw}f_w \tag{3.46c}$$

Em notação de matriz, as expressões acima podem ser expressas de forma concisa como:

$$\begin{array}{ll} \text{PCA} & \mathbf{T}_J = \mathbf{X} \mathbf{P}_J \end{array} \quad (3.47)$$

$$\begin{array}{ll} \text{FA} & \mathbf{X} = \mathbf{A} \mathbf{F} \end{array} \quad (3.48)$$

onde, na equação 3.47, $\mathbf{T}_J$ é a matriz contendo os escores para as componentes principais, $\mathbf{P}_J$ é a matriz de pesos, e o subscrito J indica que todas as componentes são usadas (J componentes), tal que o resíduo é igual a zero.

Tomando-se $R_{pp} < J$ componentes, a equação 3.47 pode ser modificada tornando-se:

$$\mathbf{T} = \mathbf{X} \mathbf{P} + \mathbf{E} \quad (3.49)$$

onde $\mathbf{E}$ representa uma matriz de resíduos.

Como já mencionado na Seção 3.1, PCA preocupa-se com a variabilidade dos dados. Essa variabilidade pode ser evidenciada por meio da medida das variâncias das variáveis. Nesse caso, ao reduzir a “dimensionalidade” do espaço de dados, PCA está tornando máximas as variâncias das poucas primeiras PCs. Esse processo tem como consequência tornar mínimas, se possível nulas, as variâncias das demais PCs. Considera-se que essas PCs com autovalores pequenos ou nulos descrevem apenas ruídos sendo, portanto, o erro atribuído à análise. Estatisticamente, isto significa que PCA procura tornar nulas, ou o mais próximo possível disto, as variâncias do erro (a variância específica) da análise.

Por outro lado, a análise de fatores preocupa-se, explicitamente, conforme postula o modelo, em explicar a estrutura subjacente aos dados, a qual pode ser evidenciada por meio das correlações entre variáveis. Isto significa que FA ocupa-se de tornar mínimos, se possível nulos, os valores das covariâncias.

Em FA devem ser destacados:

1) *Fatores Comuns* – referem-se aos fatores que levam em conta a variabilidade em pelo menos duas das variáveis originais, onde a comunalidade, $(s_i^2 - e_i)$, na qual s_i^2 é

a variância e e_i o erro), é a variabilidade associada com cada uma das variáveis originais, explicada pelos fatores comuns.

2) *Fatores Únicos* – são os elementos da diagonal da matriz de resíduos, $\mathbf{E}$, do modelo. Eles representam a variabilidade em cada variável original que não tem nada em comum com quaisquer outras variáveis.

Como FA, diferentemente de PCA, procura, por imposição do modelo, tornar mínimos, se possível nulos, os valores das covariâncias, isto significa que a matriz $\mathbf{E}$ deve ser diagonal. Ou seja, os elementos fora da diagonal de $\mathbf{E}$ são todos nulos. Então, após a obtenção dos fatores relevantes para o sistema, não existirá covariância residual implicando que a estrutura está totalmente explicada. Exatamente aí reside a diferença fundamental entre PCA e FA (Jackson, 1991, p.391).

3.2.4 – Métodos, Modelos, Algoritmos

Existem alguns métodos, modelos e algoritmos de uso mais freqüente em análise multivariada para a obtenção das componentes principais. São eles, o “power method” (muito lento nas operações computacionais), a solução do problema de autovalores-autovetores (eigenvalue analysis), Non Iterative Partial Least Squares – NIPALS, e a decomposição por valor singular – SVD. Exceto SVD, discutido no Tópico 3.2.5.2 (estruturas multimodo), os demais são abordados, introdutoriamente, no Apêndice A1.

3.2.5 – Análise Multimodo

Essa seção aborda os Dados de Ordem Superior ou Estruturas Multimodo. Aqui é apresentada uma visão geral desses métodos/modelos sob o ponto de vista histórico. Em seguida são abordados os aspectos geométricos e, finalmente, os aspectos algébricos dos modelos PCA Desdobrado (PCA-SUP), PARAFAC e TUCKER.

3.2.5.1 – Uma Visão Geral

Qualquer que seja a finalidade subjacente a uma coleta de dados, ela só faz sentido se esses dados puderem servir como meio de interpretação para o fim proposto. Em muitos casos deseja-se “trabalhar” tais dados de modo que possam fornecer previsões, ou apenas para se obter uma resposta qualitativa plausível sobre um fenômeno em estudo.

Não obstante esforços se coloquem no sentido de uma padronização de terminologia e de notação (Kiers, 2000), por possuírem “*status*” de “portal para a comunicação” no espaço da análise multivariada, esses aspectos ainda ensejam alguma revisão, que padronize termos de uso corrente em análise convencional e em análise multimodo, pelo menos no campo das aplicações químicas. Não devem, portanto, ser negligenciados. Assim, mesmo sem a intenção de adentrar a fundo na questão, é oportuno uma breve menção à discussão existente sobre o significado do termo “modelo” nesse campo de aplicações de conhecimentos da estatística e da álgebra linear.

Todavia, deve ser enfatizado que não se pretende apresentar, nessa tese, uma proposta de padronização na terminologia, mas, levantar a discussão, adotando, sob justificativa, termos que se entenda sejam mais adequados (não necessariamente de um mesmo autor) aos propósitos das análises em tela, e que não venham a afetar o significado de procedimentos e operações inerentes àquelas aplicações.

Em seu artigo sobre a máquina multilinear, a qual lida com problemas onde um arranjo de dados deve ser aproximado com o uso de uma expressão matemática, Paatero denomina modelo, a formulação matemática das expressões aproximadas para os dados (Paatero, 1999). Ou seja, um modelo, X_{est} , é uma aproximação do conjunto de dados, X (Bro, 1998). Partindo-se da premissa de que os dados levantados devem exibir um determinado comportamento passível de ser modelado por um método adequado, deve-se entender então a aproximação dos dados como sendo um modelo que expressa o

comportamento real dos mesmos. Em outras palavras, X est deverá revelar, dentro de um erro experimental aceitável, o comportamento subjacente à parte modelável¹ de X .

Como um modelo é, em geral, construído para fornecer previsões, fica uma dúvida sobre a conveniência de tal denominação quando aquele não for o objetivo da pesquisa. Assim, uma forma de contornar a dubiedade seria “qualificar” a finalidade do modelo, por exemplo, especificando-o se, para fins confirmatórios ou exploratórios.

Entretanto, particularmente para o termo modelo, seguir-se-á, nessa tese, a orientação de Bro (1998, p.14). Assim, o termo modelo não é usado aqui com o mesmo sentido que ele possui na estatística clássica. Distinguir-se-á um modelo pelas propriedades definidas pela especificação intrínseca a ele, as quais são: a *base estrutural*, as *restrições* e a *função objetiva* (Bro, 1998, p.14).

Outro aspecto que deve ser considerado em relação à terminologia é a distinção entre método e modelo. Em geral esses termos têm sido usados com o mesmo significado, embora Kroonenberg faça distinção entre eles classificando (Kroonenberg, 1989, p.8-10) como métodos: (a)estrutura de covariância; (b)quadrados mínimos alternantes (ALS); (c) quadrados mínimos parciais (PLS); (d)Tucker I; (e)Tucker II; e (f)Tucker III; e como modelos: (a)CANDECOMP; (b)IDIOSCAL; (c)INDSCAL; (d)PARAFAC; (e)Escalamento em três modos; (f)Tucker2; (g)Tucker3; e (h)Tucker com fator comum.

Na literatura química os termos método e técnica se confundem e se "intercambiam" inclusive em um mesmo contexto de um único artigo. Tal fato também ocorre, ainda que em menor escala, na literatura matemática. Não obstante caiba, não será incluída no escopo dessa tese uma definição categórica para o termo método, em que pese não se possa furtar a apontar distinções entre este e outros termos de uso generalizado na literatura corrente.

À parte de definições, as raízes históricas da formulação de terminologia parecem estar ligadas diretamente aos diferentes campos de aplicação dos métodos, técnicas e modelos. Assim, a especificidade de área, aliada à necessidade cada vez maior do desenvolvimento de procedimentos de análise, tornou-se o fator determinante na “cunhagem” de termos nas diferentes áreas de aplicação de análise multivariada. Especificamente, na área de quimiometria, uma remanescente falta de visibilidade na importância da convergência de terminologia, parece comum à maioria dos autores da área,

¹ Mesmo para X ausente de erros qualquer modelo poderá não se ajustar 100% aos dados.

mesmo àqueles que oferecem importantes contribuições no desenvolvimento de métodos/modelos.

Provavelmente isto ocorre porque a compreensão dos diferentes significados desses termos não aborta o desenvolvimento de métodos/técnicas/modelos, não influencia o desempenho dos mesmos, nem tampouco impede que resultados cada vez mais importantes surjam no espaço de publicações em análise multivariada.

Todavia, mesmo aceitando ser esta uma questão polêmica, poder-se-ia arriscar como definição para método num sentido mais amplo, como sendo um conjunto de procedimentos realizados segundo uma hierarquia ditada pelos objetivos que se deseja alcançar. Por seu lado, a técnica pode ser referida também como um conjunto de procedimentos, porém ancorada em determinadas propriedades intrínsecas ao fenômeno sob estudo. Na área da química analítica são exemplos de técnica, a espectroscopia, a espectrometria entre outras, que se valem de uma propriedade inerente aos materiais quando estimulados pela incidência de ftons de luz sobre a sua superfície, ao que respondem absorvendo e/ou refletindo parcialmente a energia incidente. A técnica seria pois, um procedimento dentro de um método. Este último, por sua vez, estaria inserido em uma metodologia.

Em análise multivariada e, especificamente em análise multimodo (ou dados de ordem superior), não obstante a classificação de Kroonenberg, a ausência de diferenças conceituais entre métodos e modelos não justificariam, a não ser por razões históricas, tais distinções na terminologia. Por exemplo, a solução inicial implementada pelo algoritmo TUCKALS3 no modelo Tucker3 (formulação de Kroonenberg e Jan de Leeuw – FKJL) é, de fato, a mesma solução do método I de Tucker (Kroonenberg, 1994; Tucker, 1966) (formulação do próprio Tucker – FT). E se, assegurado o mesmo número de colunas em cada modo, à superdiagonal do arranjo do núcleo for imposto o valor um (unidade) e zeros nas demais posições, obtém-se o modelo PARAFAC, caracterizando este como um caso particular do modelo Tucker3 (FKJL) ou do método I de Tucker (FT).

Uma incursão através do artigo original de Ledyard R. Tucker (Tucker, 1966) permite esclarecer de algum modo a motivação para a utilização de tal terminologia. Tucker, na verdade, formula um "Modelo Geral" para a análise em três modos, a partir da estrutura anteriormente denominada "Three Mode Factor Analysis" por Levin (1963), e que

ficou posteriormente conhecido como "Modelo Geral de Tucker". Quando o desenvolveu, Tucker pensou em uma estrutura em três modos capaz de ajustar-se aos dados coletados para I indivíduos e J testes em K diferentes ocasiões. O grande poder de tal estrutura para ajustar-se àqueles dados e, em última análise, a modelá-los tão bem, parece ter induzido Tucker a denominá-lo de "Fundamental Model for Three Mode Factor Analysis". Essa noção do significado do termo "modelo", também compartilhada por Kroonenberg (Kroonenberg, 1989, p.14-20; Geladi, 1989), está em total concordância com a adotada nessa tese, para fins de análise exploratória.

Problemas de ordem computacional estimularam Tucker à criação de rotas alternativas na aplicação do Modelo Geral. Para pequenos conjuntos de dados, o Modelo Geral foi chamado de Método I. Quando um dos modos² tornava-se muito grande, podendo vir a causar dificuldades computacionais, Tucker mantinha-o na forma não reduzida e propunha uma solução algébrica diferente para contornar aqueles problemas, denominando o caminho-solução de Método II. Quando além dessa dificuldade os arranjos multimodo expressavam características múltiplas (multitrait-multimode), Tucker usava uma terceira solução algébrica, porém denominando-a de Método III (Tucker, 1966).

A formulação de Kroonenberg e Jan de Leeuw (Kroonenberg e Jan de Leeuw, 1980) para o modelo em três modos corresponde àquela do modelo geral de que trata Tucker em seu artigo de 1966 (no qual ele fornece as vias para descrever a estimativa dos parâmetros de seu modelo). Assim como Tucker, Kroonenberg e de Leeuw também distinguiram três casos: o primeiro deles, o caso geral (três modos reduzidos), denominaram-no "Tucker3 model"; e os dois casos especiais, um deles (dois modos reduzidos), denotaram-no por "Tucker2 model", e o outro (apenas um modo reduzido), chamaram-no "Tucker1 model".

Deve ser observado que os modelos Tucker3 e Tucker1 propostos por Kroonenberg e de Leeuw correspondem, respectivamente, às soluções dos métodos I e II de Tucker. Finalmente, o modelo Tucker2 corresponderá ao método III de Tucker quando o arranjo daquele apresentar, adicionalmente, características múltiplas (por exemplo, dois modos forem representados por matrizes de covariâncias).

² O termo modo é introduzido em substituição ao termo dimensão ou direção. O primeiro causa confusão com o conceito de dimensão de um espaço de fatores em análise de fatores (ou de componentes em análise de componentes principais), e o segundo não é específico. Ver Apêndice A2 para distinção entre os termos modos (modes) e vias (ways).

Para resolver os problemas de ajuste de sua formulação para os modelos Tucker3 e Tucker2, Kroonenberg e de Leeuw desenvolveram um algoritmo por quadrados mínimos alternantes (ALS). Denominaram TUCKALS3 e TUCKALS2, respectivamente, àqueles usados para estimar os parâmetros dos modelos Tucker3 e Tucker2. Tecnicamente, a estimativa dos parâmetros no modelo Tucker2 não apresenta qualquer problema, bastando eliminar um dos passos no TUCKALS3. Do ponto de vista computacional, entretanto, é mais eficiente resolver o modelo diretamente pelo algoritmo análogo ao TUCKALS3, ou seja, pelo TUCKALS2 (Kroonenberg e Jan de Leeuw, 1980). No caso do modelo Tucker1, não é necessário um algoritmo específico para resolvê-lo porque, organizando-se apropriadamente a entrada de dados, a análise pode ser conduzida com qualquer programa usado para PCA (Kroonenberg e Jan de Leeuw, 1980).

Nessa tese adotar-se-á para os arranjos multimodo a seguinte terminologia: modelos Tucker1, Tucker2 e Tucker3 (todos FKJL), e modelos PARAFAC e PCA Desdobrado.

Um terceiro aspecto a ser considerado é a distinção entre método e algoritmo. De acordo com Malinowski (Malinowski, 1991, p.18), “em Análise de Fatores, o procedimento para se calcular a solução (abstrata) envolve um método matemático chamado ‘problema de autovalores-autovetores’ (eigenanalysis) gerando autovalores (eigenvalues) e autovetores (eigenvectors) associados”. O autor acrescenta ainda que “os quatro métodos mais usados para isto são o Power Method, o de Jacobi, a Decomposição por Valor Singular (SVD) e o NIPALS” (Malinowski, 1991, p.18). Ou seja, o método para a obtenção dos autovetores de uma matriz se utiliza de um dos quatro principais “métodos” acima para pôr em prática o procedimento.

Devido à ausência de rigor no uso da terminologia, por um lado, o mesmo termo é usado para procedimentos diferentes e, por outro, faz-se referência ao mesmo procedimento por meio de termos distintos. Assim, em que pese, método, técnica e algoritmo, possuam, dentro de suas semelhanças, diferenças intrínsecas, acabam se revestindo de uma aparente similitude. Então, quando se faz referência, por exemplo, a Análise de Componentes Principais (PCA), não se deve confundir com o algoritmo usado para extrair os componentes na análise (Bro, 1998, p.5). Assim, o NIPALS deve ser entendido como um algoritmo já que o processo para a obtenção das componentes, baseado na função objetiva, a qual leva em conta os resíduos do modelo, é iterativo. Isto é, é um processo de cálculo

baseado em um conjunto de operações bem definidas, dentro de um ordenamento particular, contendo um número finito de etapas.

O desenvolvimento de instrumentação analítica capaz de gerar outros espaços além daqueles ocupados pelas linhas e colunas da matriz convencional, ou a estrutura mesma de um estudo no qual uma matriz de dados deve ser construída para várias localidades (ocasiões) diferentes, exigiram um arranjo mais complexo para representar os dados. São exemplos, um experimento no qual a luminescência total de cada amostra analisada tem o espectro de emissão determinado em diferentes comprimentos de onda de excitação (Reis e Ferreira, 1999) (gerando um cubo de dados), e outro no qual, diferentes conjuntos de amostras são submetidos à medição de diversas variáveis, em épocas distintas e em diferentes estações de amostragem (gerando um hiper-cubo de dados).

Nessa tese preservar-se-á a noção de dimensão como aquela naturalmente percebida pela disciplina álgebra linear (Strang, 1988; p.63-69, 74-75, 80-87 e 90-99; Watkins, 1991; p.287-346). Assim, se um outro espaço for incorporado, ao invés da representação tradicional, quadrada ou retangular da matriz convencional, obter-se-á um cubo ou um paralelepípedo de dados. Nesse caso, cada matriz de dados será um plano desse cubo ou paralelepípedo e o arranjo possuirá três modos, um para cada dimensão da figura geométrica cúbica. Fica implícito a partir dos argumentos acima, que arranjos superiores a três modos (hiper-cubos) podem ser construídos.

Várias denominações têm surgido para estruturas dessa natureza. Nessa tese são usadas as denominações “*Análise Multimodo*” e, alternativamente “*Dados de Ordem Superior*”. Nesse sentido, enfatizando o parágrafo precedente e para melhor compreensão, buscar-se-á na análise tensorial, explicação e justificativa para a adoção do termo “*Ordem Superior*”. A teoria de tensores (Sanches e Kowalski, 1988a e 1988b; Zeng e Hopke, 1992; Burdick, 1995) define escalares, vetores e matrizes, respectivamente, como tensores de ordens zero, um e dois. Assim, um tensor de ordem dois (uma matriz) possui dois modos. Quando um ou mais modos são adicionados, o arranjo formado pelas entradas de dados nessa estrutura é denominado “*Dados de Ordem Superior*”. As operações algébricas envolvendo o produto de Kronecker (Apêndice IV) também sugerem essa terminologia. A denominação “Multimodo” se justifica também com base no parágrafo anterior, no qual o termo “modo” foi introduzido em substituição ao termo “dimensão”. A Figura 3.10 mostra

um esquema dessa estrutura com as diferentes possibilidades de se “fatiá-la” em suas correspondentes matrizes.

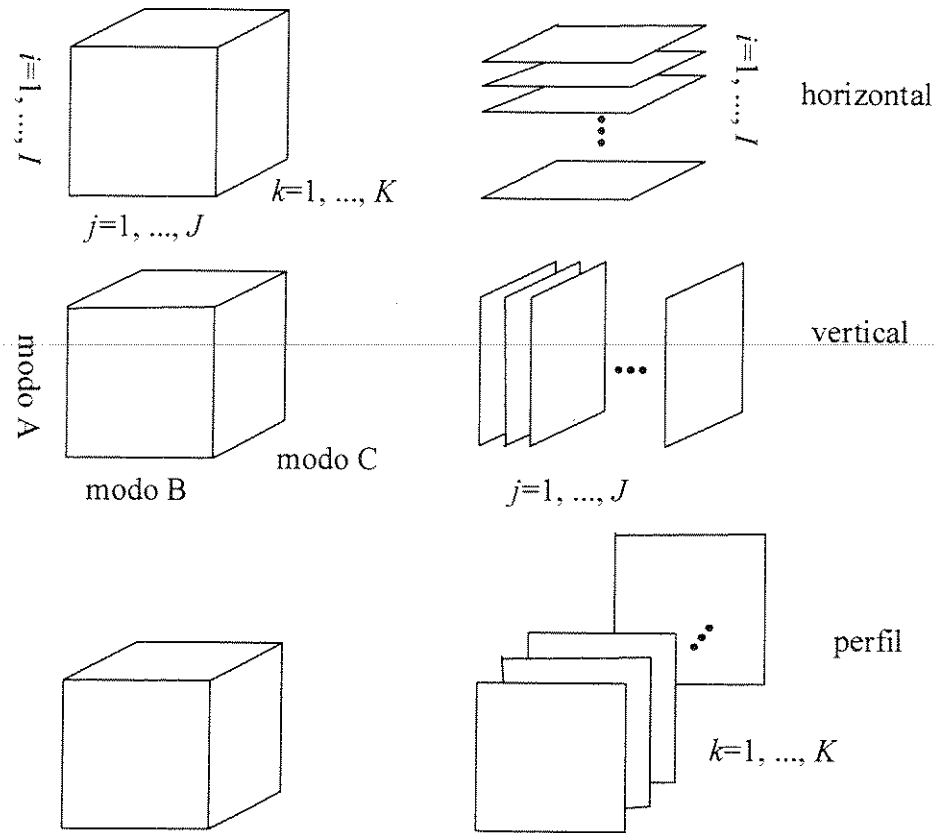


Fig.3.10: Estrutura de “Dados de Ordem Superior” ou “Multimodo” e suas “fatias” (a) horizontais, (b) verticais e (c) de perfil.

Como se observa na Figura 3.10, os índices I , J e K são reservados para indicar as três entidades que constituem as dimensões do arranjo de dados ao longo dos eixos, respectivamente, vertical, horizontal e do último, perpendicular aos dois primeiros (Kiers, 2000). Esses três conjuntos de índices definem os três modos do arranjo. No caso de arranjos com somente três modos, estes serão denotados por A, B e C, para os, respectivamente, primeiro, segundo e terceiro modos (Kiers, 2000).

Um arranjo em três modos (cubo de dados), pode ser considerado em termos de um conjunto de matrizes (Kiers, 2000), conforme visto acima. Como se observa na Figura 3.10, essas matrizes formam as “fatias” (ou planos) horizontais, laterais e frontais do arranjo (Kiers, 2000). A denominação para uma das “fatias”, a lateral, em Kiers (Kiers, 2000), recebe em Bro (Bro, 1998, p.9) o nome de vertical.

Esses métodos vêm experimentando ampla e crescente utilização por pesquisadores das diversas áreas de engenharia. Estes estão bastante familiarizados com a disciplina geometria descritiva (de Luigi Cremona e outros grandes mestres), a qual trata de projetar em duas dimensões os objetos construídos no espaço tridimensional (e vice-versa). O plano conhecido como frontal em análise multimodo, é denominado plano de perfil em geometria descritiva; um plano que é perpendicular aos dois planos de projeção, o horizontal, π (lê-se plano pi), e o vertical, π' (lê-se plano pi-linha). Como não faz diferença para pesquisadores de outras áreas, as denominações mais adequadas para as “fatias” seriam **horizontal**, **vertical** e **perfil** (para aquela conhecida como frontal). Discussões sobre denominações para os demais sub-arranjos (por exemplo, “fibras” ou “tubos”) da estrutura de ordem superior podem ser encontradas em Bro (Bro, 1998, p.7-9) e Kiers (Kiers, 2000).

Existe na literatura contribuição de vários autores às estruturas multimodo. Dentre as mais recentes podem ser destacados trabalhos que enfatizam a abordagem via teoria de tensores, como os de Sanches e Kowalski (Sanches e Kowalski, 1988a e 1988b), de Zeng e Hopke (Zeng e Hopke, 1992), de Burdick (Burdick, 1995); e de outros que apresentam os arranjos de ordem superior como uma extensão natural das estruturas da matriz convencional de dados, como Kroonenberg (Kroonenberg, 1989, p.14-20; Kiers (Kiers, 1991), Geladi (Geladi, 1989), Henrion (Henrion, 1994), Harsman e Lundy (Harsman e Lundy, 1994), Andersson e Bro (Andersson e Bro, 1998), Bro e Andersson (Bro e Andersson, 1998), Bro (Bro, 1998, p.2,5,14-32,43-49), Paatero (Paatero, 1999). Kroonenberg e Jan de Leeuw também usam a teoria de tensores para resolver o problema de ajuste dos modelos por quadrados mínimos alternantes.

3.2.5.2 – Estruturas Multimodo ou Dados de Ordem Superior

Comparado aos convencionais, os arranjos multimodo, além de suportarem, em geral, maior diversidade de dados, possuem estruturas mais complexas, o que causa reflexos diretos nas operações computacionais. Assim, o algoritmo, cujo coração tem sede na *função objetiva* (loss function) e nas *restrições* (constraints) definidas na *especificação do modelo*, adquire importância fundamental na aplicação dos mesmos. Isto sugere que, doravante, nas definições dos métodos/modelos haja preocupação em evidenciar tais aspectos.

Após as necessárias considerações históricas, antes de introduzir as estruturas multimodo, é conveniente resgatar um método convencional e um tipo de decomposição que mantém estrita relação com aquelas estruturas. Aqui, para efeito de especificação, PCA será também “enquadrado” nesses mesmos critérios definidores referidos acima.

- **P C A: a identificação do método**

No Tópico 3.2.2.1 foi abordada a Análise de Componentes Principais. Encontrou-se uma expressão matricial (equação 3.39) para se expressar os escores para as componentes principais quando são usadas todas as componentes. Tal expressão, $\mathbf{T}_J = \mathbf{X} \mathbf{P}_J$, pode ser modificada para representar a decomposição via PCA para uma determinada matriz de dados. Assim, pode-se obter a expressão

$$\mathbf{X} = \mathbf{T}_J \mathbf{P}_J^T \quad \text{ou} \quad \mathbf{X} = \mathbf{T}_R \mathbf{P}_R^T \quad (R = J) \quad (3.50)$$

onde $\mathbf{X}$ é a matriz de dados, $\mathbf{T}_R$ a matriz de escores, $\mathbf{P}_R$ é a matriz de loadings, J é o número de variáveis originais, e R é o posto da matriz $\mathbf{X}$.

Para o caso de uma matriz de dados, $\mathbf{X}(I \times J)$, mesmo não deficiente de posto, R pode ser menor do que J , por exclusão do ruído experimental tal que a equação 3.50 torna-se:

$$\mathbf{X} = \mathbf{T} \mathbf{P}^T + \mathbf{E} \quad \text{ou} \quad \mathbf{X}_{\text{est}} = \mathbf{T} \mathbf{P}^T \quad (3.51)$$

onde $\mathbf{X}(I \times J)$ é a matriz de dados, $\mathbf{T}(I \times R_{pp})$ é a matriz de escores, $\mathbf{P}(J \times R_{pp})$ é a matriz de loadings, $\mathbf{E}(I \times J)$ é a matriz de resíduos e R_{pp} é o pseudo posto da matriz $\mathbf{X}$.

A abordagem das estruturas nessa tese tenta enfatizar a semelhança estrutural entre o método PCA, a decomposição SVD, o modelo PARAFAC e o modelo de Tucker. Assim, para efeito da comparação entre a hierarquia e a estrutura dos modelos que é apresentada na seção 3.2.5.3, é interessante mostrar como os critérios definidores dos modelos se aplicam à análise PCA.

Um modelo, $\mathbf{X}_{est}$, obtido a partir de um conjunto de dados, $\mathbf{X}(I \times J)$, por meio de PCA, pode ser definido da seguinte maneira (Bro, 1998, p.15):

Dada uma matriz $\mathbf{X}(I \times J)$, existem matrizes $\mathbf{T}$ e $\mathbf{P}$ de dimensões coluna, R_{pp} , tal que

$$\mathbf{X}_{est} = \mathbf{T} \mathbf{P}^T \quad (3.52)$$

com a solução de

$$\min_{\mathbf{T}, \mathbf{P}} \|\mathbf{X} - \mathbf{T} \mathbf{P}^T\|_F^2 \mid \mathbf{T}^T \mathbf{T} = \Delta, \mathbf{P}^T \mathbf{P} = \mathbf{I} \quad (3.53)$$

onde Δ é uma matriz diagonal, $\mathbf{I}$ é a matriz identidade, R_{pp} é o pseudo-posto (pseudo-rank) de $\mathbf{X}$, a barra vertical “ $\mid$ ” significa “tal que”, e $\|\cdot\|_F^2$ é o quadrado da norma de Frobenius (uma norma de matrizes consistente com a norma Euclidiana de vetores, isto é, $\|\mathbf{x}\|_F^2 = \sum_i \sum_j x_{ij}^2 = \text{tr}(\mathbf{X}^T \mathbf{X})$).

As propriedades para identificação da análise de componentes principais podem ser resumidas da seguinte maneira:

Método	PCA
Dados	$\mathbf{X}$
Modelo dos Dados	$\mathbf{X}_{est}$
Base Estrutural	$\mathbf{X}_{est} = \mathbf{T} \mathbf{P}^T$
Restrições	$\mathbf{T}^T \mathbf{T} = \Delta, \mathbf{P}^T \mathbf{P} = \mathbf{I}$
Função Objetiva	$\ \mathbf{X} - \mathbf{T} \mathbf{P}^T\ _F^2$
Especificação do Método	$\min_{\mathbf{T}, \mathbf{P}} \ \mathbf{X} - \mathbf{T} \mathbf{P}^T\ _F^2 \mid \mathbf{T}^T \mathbf{T} = \Delta, \mathbf{P}^T \mathbf{P} = \mathbf{I}$

Uma discussão sobre base estrutural, restrições, função objetiva (loss function), e especificação do método, está bem fundamentada em Bro (Bro, 1998, p.15-19).

- **Decomposição por Valor Singular – S V D: origem e propriedades**

O procedimento SVD parece ter suas raízes no século XIX. Os nomes de pelo menos cinco renomados matemáticos clássicos, J. Sylvester (1814-1897), E. Beltrami (1835-1899), C. Jordan (1838-1921), E. Schmidt (1876-1959) e H. Weyl (1885-1955), parecem estar associados ao desenvolvimento da teoria da SVD (Datta, 1994, p.551).

Na literatura matemática, SVD não está associada a qualquer nome embora seja uma junção da fatoração LU (com origem na eliminação Gaussiana) com a fatoração QR do processo de ortogonalização de Gram–Schmidt. Assim, SVD é um tipo de fatoração de matrizes simplesmente conhecido como Decomposição por Valor Singular (Strang, 1988, p.1-2, 172-176, 195-196, 442-443).

A decomposição da matriz $\mathbf{X}$, na análise PCA, pode ser realizada, por exemplo, via SVD. De forma resumida:

Dada uma matriz $\mathbf{X}(I \times J)$, existem matrizes $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_I]$ e $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \dots, \mathbf{p}_J]$, tais que

$$\mathbf{U}^T \mathbf{X} \mathbf{P} = \Lambda \{(\lambda_1)^{1/2}, (\lambda_2)^{1/2}, (\lambda_3)^{1/2}, \dots, (\lambda_R)^{1/2}, (\lambda_{R+1})^{1/2}, (\lambda_{R+2})^{1/2}, \dots, (\lambda_W)^{1/2}\},$$

$$W = \min\{I, J\} \quad (3.54)$$

$$\text{onde tem-se: } \lambda_1 = \sigma_1^2 \geq \lambda_2 = \sigma_2^2 \geq \lambda_3 = \sigma_3^2 \geq \dots \geq \lambda_R > 0 \quad (\text{iii})$$

$$\sigma_{R+1}^2 = \sigma_{R+2}^2 = \sigma_{R+3}^2 = \dots = \sigma_W^2 = 0. \quad (\text{iv})$$

$$\Lambda = [\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_R, \lambda_{R+1}, \lambda_{R+2}, \lambda_{R+3}, \lambda_W], \quad R = \min\{I, J\} \quad (\text{v})$$

Em (iii) e (iv), os λ 's são autovalores de ambos $\mathbf{X} \mathbf{X}^T$ e $\mathbf{X}^T \mathbf{X}$, e os σ 's são valores singulares de $\mathbf{X}$.

Então pode-se expressar a matriz $\mathbf{X}$ (de posto completo, R) como:

$$\mathbf{X}_{(I,J)} = \mathbf{U}_{(I,I)} \Lambda_R \mathbf{P}_{(J,J)}^T, \quad R = \min\{I, J\} \quad (3.55)$$

onde $\mathbf{X}_{(I,J)}$ = matriz de dados originais, $\mathbf{U}_{(I,I)}$ = matriz de vetores singulares à esquerda de $\mathbf{X}$ (matriz de autovetores de $\mathbf{X}\mathbf{X}^T$) ; $\mathbf{P}_{(J,J)}$ = matriz de vetores singulares à direita de $\mathbf{X}$ (matriz de autovetores $\mathbf{X}^T\mathbf{X}$) e $\Lambda_R = \Lambda_{(I,R)} =$ matriz diagonal (para $J < I$) cujas entradas são os valores singulares de $\mathbf{X}$. Analogamente, se $I < J$, $\Lambda_R = \Lambda_{(R,J)}$.

Uma inspeção na equação 3.55 permite verificar que a matriz diagonal Λ_R determina a dimensão da base do sistema, isto é, determina o número de autovetores que gera aquele espaço. Assim, quando se tem $\mathbf{X}(I \times J)$ de posto completo (com $J < I$), o posto R na equação 3.55 assume o valor J . Desse modo a equação pode ser escrita:

$$\mathbf{X}_{(I,J)} = \mathbf{U}_{(I,I)} \Lambda_{(I,J)} \mathbf{P}_{(J,J)}^T = \mathbf{U}_{(I,I)} \Lambda_{(I,R)} \mathbf{P}_{(R,J)}^T \quad (3.56)$$

ou, equivalentemente, e para enfatizar o posto R :

$$\mathbf{X}_{(I,J)} = \mathbf{U}_{(I,I)} \Lambda_{(R)} \mathbf{P}_{(R)}^T, \text{ ou } \mathbf{X}_{(R)} = \mathbf{U}_{(I)} \Lambda_{(R)} \mathbf{P}_{(R)}^T \quad (R = J < I) \quad (3.57)$$

Analogamente, se $\mathbf{X}(I \times J)$ é de posto completo (com $I < J$), o posto R na equação 3.55 assume o valor I . Desse modo a equação pode ser escrita:

$$\mathbf{X}_{(I,J)} = \mathbf{U}_{(I,I)} \Lambda_{(I,J)} \mathbf{P}_{(J,J)}^T = \mathbf{U}_{(I,R)} \Lambda_{(R,J)} \mathbf{P}_{(R,J)}^T \quad (3.58)$$

ou, equivalentemente, e para enfatizar o posto R :

$$\mathbf{X}_{(I,J)} = \mathbf{U}_{(I,R)} \Lambda_{(R)} \mathbf{P}_{(J)}^T, \text{ ou } \mathbf{X}_{(R)} = \mathbf{U}_{(R)} \Lambda_{(R)} \mathbf{P}_{(J)}^T \quad (R = I < J) \quad (3.59)$$

onde a representação com o subscrito R enfatiza a dimensionalidade de $\mathbf{X}$.

Verifica-se também, para esse caso geral, em que $\mathbf{X}(I \times J)$ não é de posto completo, isto é, quando $R < \min\{I, J\}$, que a equação 3.55 deve ser representada como:

$$\mathbf{X}_{(I,J)} = \mathbf{U}_{(I,R)} \Lambda_{(R,R)} \mathbf{P}_{(R,J)}^T \quad (3.60)$$

Nesse caso, quando se tem $\mathbf{X}(I \times J)$ deficiente de posto, isto é, $R < \min\{I, J\}$, o posto R não assume o valor J nem o valor I , e por analogia às equações 3.57 e 3.59:

$$\mathbf{X}_{(R)} = \mathbf{U}_{(R)} \Lambda_{(R)} \mathbf{P}_{(R)}^T, \quad (\text{se } R < J < I, \text{ ou } R < I < J) \quad (3.61)$$

Para esse caso geral da decomposição por valor singular, SVD, vale acrescentar, em relação à equação 3.54:

a) os parâmetros λ e σ são denominados, respectivamente, autovetores de ambos, $\mathbf{X}^T \mathbf{X}$ e $\mathbf{X} \mathbf{X}^T$, e valores singulares de $\mathbf{X}$,

b) das J colunas (variáveis) da matriz $\mathbf{X}(I \times J)$, apenas R são linearmente independentes, portanto, $\sigma_k^2 = 0$ para $k=R+1, \dots, W$. Isto é, existem $W = \min(I, J)$ valores singulares de $\mathbf{X}$, e se R é o posto de $\mathbf{X}$, existem R valores singulares positivos, os quais são raízes quadradas dos autovalores não zero de $\mathbf{X}^T \mathbf{X}$ (e de $\mathbf{X} \mathbf{X}^T$). Os restantes, $W - R$ (se $R < W$) valores singulares são iguais a zero.

c) para o caso específico no qual $R = W = \min\{I, J\}$, a matriz $\mathbf{X}(I \times J)$ possui posto completo tal que 3.56 torna-se a equação 3.55:

Decorre de 3.55 e 3.60, respectivamente, que (representando $\mathbf{X}(I \times J)$ por $\mathbf{X}$):

$$\mathbf{X} \mathbf{P}_{(J, J)} = \mathbf{U}_{(I, I)} \Lambda_{(I, R)}, \quad R = \min\{I, J\} \quad (3.62a)$$

ou, simplesmente,

$$\mathbf{X} \mathbf{P}_{(J, J)} = \mathbf{U}_{(I, I)} \Lambda_R, \quad R = \min\{I, J\} \quad (3.62b)$$

e

$$\mathbf{X} \mathbf{P}_{(J, R)} = \mathbf{U}_{(I, R)} \Lambda_{(R, R)}, \quad R < \min\{I, J\} \quad (3.62c)$$

ou, simplesmente,

$$\mathbf{X} \mathbf{P}_{(J, R)} = \mathbf{U}_{(I, R)} \Lambda_R, \quad R < \min\{I, J\} \quad (3.62d)$$

resultados que correspondem às matrizes de escores das amostras para as PC's, e $\mathbf{P}_{(J,J)} = \mathbf{P}_J$ e $\mathbf{P}_{(J,R)} = \mathbf{P}_R$, correspondem às matrizes de pesos para as PC's, em cada caso. Encontra-se, desse modo, a mesma equação 3.39, obtida no tópico 3.2.2.1(p.69)

$$\mathbf{X} \mathbf{P}_J = \mathbf{T}_J \quad \text{ou} \quad \mathbf{X} \mathbf{P}_R = \mathbf{T}_R \quad (3.62e)$$

onde $\mathbf{U}_{(I,I)} \mathbf{\Lambda}_R = \mathbf{T}_J$ ou $\mathbf{U}_{(I,R)} \mathbf{\Lambda}_R = \mathbf{T}_R$, as quais podem ser transformadas em:

$$\mathbf{X} = \mathbf{T}_J \mathbf{P}_J^T \quad \text{ou} \quad \mathbf{X} = \mathbf{T}_R \mathbf{P}_R^T \quad (3.62f)$$

Deve ser enfatizado aqui o fato de que nas equações em que $R = W = \min\{I, J\}$, a matriz $\mathbf{X}(I \times J)$ é de posto completo, e quando $R < W = \min\{I, J\}$, ela é deficiente de posto.

Considerando-se apenas as componentes que possuem informação significativa, isto é, retirando-se o ruído experimental, o número de componentes é determinado pelo pseudo-posto, R_{pp} , tal que as equações 3.62, independentemente de deficiência ou não de posto, tornam-se

$$\mathbf{X} = \mathbf{T} \mathbf{P}^T + \mathbf{E} \quad (3.63)$$

onde $\mathbf{X}$ = matriz de dados ($I \times J$), $\mathbf{T}$ = matriz de escores ($I \times R_{pp}$), $\mathbf{P}$ = matriz de pesos ($J \times R_{pp}$), $\mathbf{E}$ = matriz de resíduos ($I \times J$), e R_{pp} é o *pseudo-posto* de $\mathbf{X}$.

Como se observa a equação 3.63 é idêntica às equações 3.49 (subseção 3.2.3, p72), 3.51 (tópico 3.2.5.2, p82) e representa a equação 3.9 na subseção 3.1.3 (p53).

Uma excelente discussão sobre os diferentes subespaços ocupados pelos autovetores de $\mathbf{X}\mathbf{X}^T$ e de $\mathbf{X}^T\mathbf{X}$ na aplicação de SVD, pode ser encontrada em Strang (Strang, 1988, p.338-350.), Golub (Golub, 1989, p.427), e em Watkins (Watkins, 1991, p.390-421). No Apêndice A3 pode-se encontrar mais detalhes sobre SVD.

A Figura 3.11 mostra, de forma esquemática, um resumo da discussão acima para uma decomposição por valor singular.

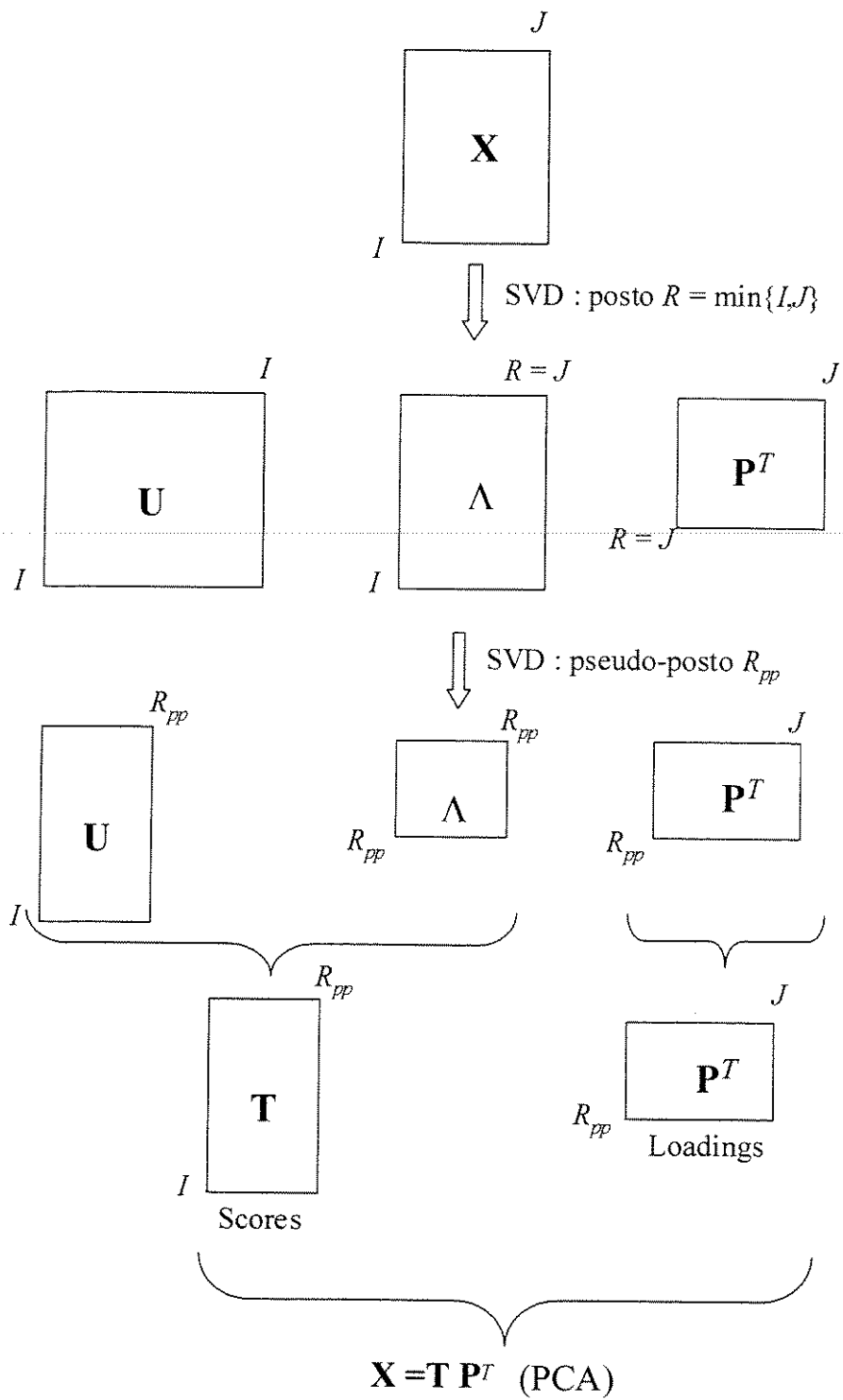


Fig.3.11: Representação esquemática de uma decomposição SVD. Considera-se uma matriz de posto completo (com $J < I$), e distinção entre o posto R , e o pseudo posto R_{pp} .

- **A Estrutura Multimodo**

O problema acima (para PCA) envolve as medições de variáveis distintas sobre as diferentes amostras. Porém, se tais dados forem coletados em várias localidades (ou condições), um terceiro modo tem que ser acrescentado ao arranjo obtendo-se uma estrutura multimodo. As amostras podem ser representadas no modo um (ou modo A), as variáveis no modo dois (ou modo B), e as localidades (ou condições) no modo três (ou modo C). Deve-se observar que a disposição dos modos dessa maneira é totalmente arbitrária, não existindo uma regra para tal ordenamento (Kiers, 1991). Para efeito de implementação dos algoritmos, o arranjo deve ser desdobrado de acordo com os planos que o representam. Isso é feito a partir da função algébrica que representa esse desdobramento e da função objetiva a ser tornada mínima (ver equações 3.66, 3.67, 3.69 e 3.70 adiante). A Figura 3.12 mostra um esquema de um cubo de dados e seu desdobramento nas matrizes que o compõem.

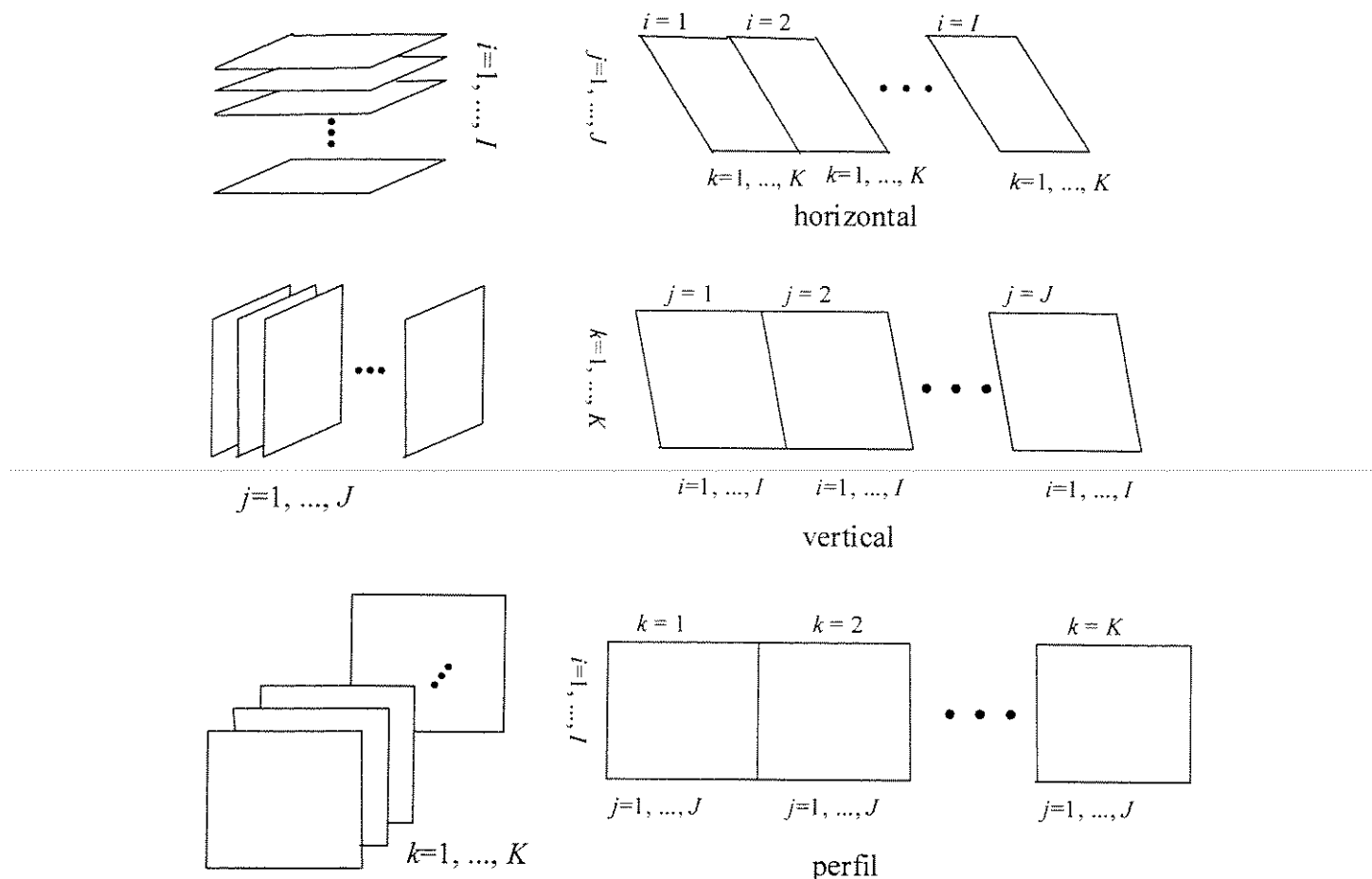


Fig.3.12: Desdobramento (“unfolding”) de um cubo de dados

O próximo tópico abordará os aspectos algébricos dos métodos/modelos multimodo. Não obstante exista um bom número de autores nesse campo da análise multivariada, na discussão dos modelos de ordem superior tratados nessa tese, serão seguidas, no que diz respeito ao tratamento algébrico, as abordagens, principalmente, de Kiers (Kiers, 1991), Bro (Bro, 1998), Kroonenberg (Kroonenberg, 1989; 1994), e Kroonenberg e Jan de Leeuw (Kroonenberg e Jan de Leeuw, 1980), Harshman e Lundy (Harshman e Lundy, 1994), Paatero (Paatero, 1999), Geladi (Geladi, 1989) e Henrion (Henrion, 1994).

3.2.5.3 – Modelos de Tucker, PARAFAC e PCA multimodo (PCA-SUP)

A análise multivariada em dados de ordem superior ou análise multimodo, tem recebido diversas denominações para o mesmo modelo, fato que, felizmente, do ponto de vista da análise em si mesma, acarreta apenas questões de semântica. Entretanto, como já anteriormente apontado, devido a confusões que podem surgir, é desejável um esforço no sentido de uma uniformização. Assim, numa tentativa de se dirigir a uma padronização de notação e de terminologia em análise multimodo, esse trabalho utilizará, com pequenos acréscimos e/ou modificações, as sugestões de Kiers (Kiers, 2000) envolvendo os modelos CANDECOMP/PARAFAC e TUCKER.

A gênese da análise de dados de ordem superior tem estado presente em vários artigos que tratam com diferentes graus de aprofundamento teórico, esse importante conjunto de métodos e de modelos de análise de dados.

Resumidamente, a sua história surgiu com a introdução da Análise de Componentes Principais em Três Modos por Tucker em 1963 e 1964 (Kroonenberg, 1989, p.3; Henrion, 1994), e de Diferenças de Escalamento Individual (Kroonenberg, 1989, p.3) por Bloxom (Bloxom, 1968b), Carroll e Chang (Carroll e Chang, 1970), e Horan (Horan, 1969). Desde então as publicações têm crescido em duas distintas direções. Uma voltada apenas para a aplicação dos modelos e outra voltada não só para a aplicação como também para o aperfeiçoamento dos mesmos.

Os modelos fundamentais usados em análise de dados de ordem superior são os propostos por Tucker (Kiers, 1991), e aqueles propostos em 1970 por Carroll e Chang, (Carroll e Chang, 1970, em Harshman e Lundy, 1994) e, independentemente, também em 1970, por Harshman (Harshman, 1970, em Harshman e Lundy, 1994; em Kiers, 1991).

Os problemas de ajuste dos modelos, como já visto, foram contornados quando Kroonenberg e de Leeuw (Kroonenberg e Jan de Leeuw, 1980) propuseram um algoritmo para o ajuste dos mesmos por quadrados mínimos. Para distinguir seu modelo daqueles que usam ajuste indireto, como análise de fatores, denominaram seu procedimento de “análise de componentes principais em três modos” (Kiers, 1991). Detalhes sobre as distinções entre os diversos métodos/modelos podem ser encontrados em Kroonenberg (Kroonenberg, 1989, p.48-62; Geladi, 1989), Kiers (Kiers, 1991; 2000) e Bro (Bro, 1998, p.22-49).

Vale acrescentar, todavia, por sua importância particularmente na resolução de espectros, um modelo (também trilinear) que tem recebido muita atenção e grande aplicação em quimiometria. Ele foi desenvolvido a partir de 1978 por Ho e colaboradores (Ho *et alli*, 1978, 1980), inicialmente como um algoritmo denominado Rank Annihilation Factor Analysis (RAFA). Sucessivas tentativas para melhorar o seu desempenho foram acompanhadas de correspondentes denominações: Generalized Rank Annihilation Factor Analysis (GRAFA), Generalized Rank Annihilation Method (GRAM), Trilinear Decomposition (TLD) e Direct Trilinear Decomposition (DTD), envolvendo a participação de vários pesquisadores tais como Appellof e Davidson, 1983; Lorber, 1984, 1985; Sanches e Kowalski, 1986; Wilson *et alli*, 1989; Sanches e Kowalski, 1990; Li *et alli*, 1992; Zeng e Hopke, 1992; Faber *et alli*, 1993; Booksh e Kowalski, 1994; Faber *et alli*, 1994a, 1994b; e Faber, 1997. O GRAFA é idêntico ao GRAM enquanto o TLD é idêntico ao DTD (Bro, 1997). Há que se acrescentar também Evolving Factor Analysis – EFA – (Maeder e Zilian, 1988).

Sem prejuízo do reconhecimento da importância de cada um dos autores envolvidos na evolução do método, podem ser destacados alguns deles: a) Sanches e Kowalski (1986), os quais, apresentam uma generalização do método de anulação de posto (Generalized Rank Annihilation Method – GRAM), do qual o método não iterativo de Lober (1985) é um caso particular (Sanches e Kowalski, 1986); b) Wilson, Sanches e Kowalski (1989), os quais apresentam um algoritmo mais estável para o GRAM (Wilson *et alli*, 1989); c) Sanches e Kowalski, (1990), os quais apresentam uma extensão do GRAM denominada Direct Trilinear Decomposition – DTD (Sanches e Kowalski, 1990); e d) Li, Hamilton e Gemperline (1992), os quais discutem o problema do aparecimento de valores complexos na estimativa do modelo, tanto para o GRAM como para a DTD, e oferecem alternativas para a solução do problema (Li *et alli*, 1992).

Antes de discorrer sobre os modelos PARAFAC e TUCKER, é necessário mencionar um tipo de análise cuja aplicação ainda não está bem compreendida (no sentido da preparação da estrutura de dados para a análise). Em dados de ordem superior essa análise é conhecida como PCA-SUP (Kiers, 1991) e em quimiometria é referida como “PCA desdobrado” (unfolding PCA) ou “PCA multimodo” (Multiway PCA) (Kiers, 2000, Wold *et alli*, 1987). Ela é, na realidade, o modelo Tucker1 (Kiers, 1991), no qual a estrutura

dos dados é rearranjada em uma nova matriz, sendo esta, submetida a uma análise PCA convencional (Bro, 1998, p.43).

A compreensão do modo de construção desse arranjo, como visto adiante, é bastante instrutiva porque se utiliza de uma operação extremamente simples e útil denominada “vetorização” (Kiers, 2000), por meio do operador algébrico, $\text{Vec}(\cdot)$, bastante usado em análise multimodo.

Supondo uma coleta de dados constituída de I objetos, J variáveis e K localidades, estes podem ser analisados via PCA convencional de duas maneiras:

- a) Organizando-se os dados verticalmente (Fig.3.13) tal que sejam empilhadas (justapostas verticalmente) K matrizes de dimensões $I \times J$. Nesse caso, a nova matriz obtida constitui-se de diferentes conjuntos de objetos (um conjunto para cada uma das K localidades) sob as mesmas variáveis.
- b) Organizando-se os dados horizontalmente (Fig.3.13) tal que sejam reunidas lado a lado (justapostas horizontalmente) K matrizes de diferentes conjuntos de variáveis (um conjunto para cada uma das K localidades).

Observe-se que essas formas de se empilhar as matrizes diferem totalmente daquela usada para análise multimodo. É útil comparar a Figura 3.13, a qual ilustra os casos a e b citados acima, com a Figura 4.8 para arranjos multimodo e com as Figuras 3.14 e 3.15 para PCA Multimodo (PCA-SUP) abordado adiante.

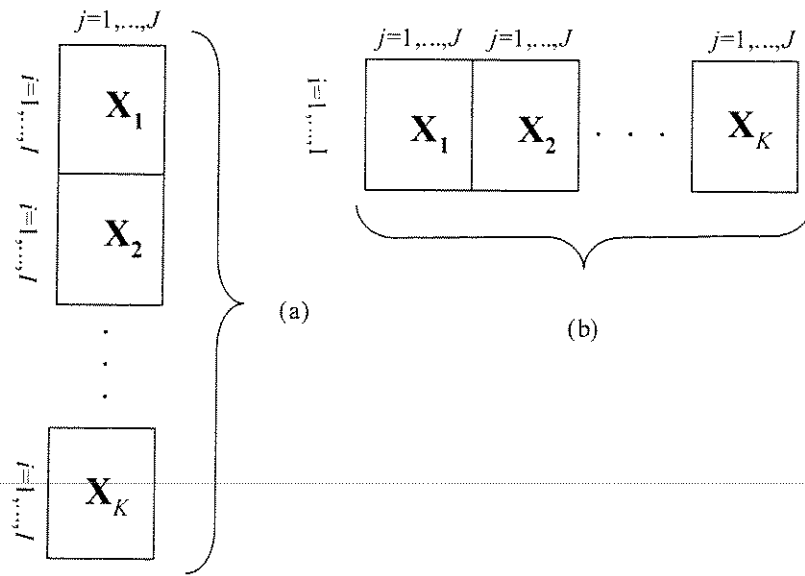


Fig.3.13: Matrizes $X_1, X_2, \dots, X_K$, justapostas (a) verticalmente, e (b) horizontalmente. Observe que a estrutura (b) é idêntica ao desdobramento de um paralelepípedo de dados em seus planos de perfil. Entretanto, a estrutura (a) não tem qualquer relação com os outros tipos de desdobramentos. Compare os índices de (a) e de (b) dessa Figura com aqueles da Figura 3.12.

Esses casos, a e b, são denominados conjuntos múltiplos de dados (multiple data sets) e não podem ser considerados como dados de ordem superior porque os três modos (localidades, objetos e variáveis) não são totalmente cruzados na análise (Kiers, 1991). No Apêndice A4 encontra-se uma discussão sobre a aplicação desses casos e dos resultados que ele pode oferecer quando são analisados nas formas descritas em a e b.

- **Modelo PCA multimodo (PCA-SUP)**

Os conjuntos múltiplos de dados podem, todavia, ser colocados como dados de ordem superior mediante modificação apropriada. Para o caso de vários conjuntos de observações sobre as mesmas variáveis, pode-se proceder da seguinte maneira (Figuras 3.14 e 3.15):

- 1) Toma-se a matriz, $\mathbf{X}_1$, de I observações por J variáveis, para a primeira localidade, k_1 . Em seguida, cada vetor coluna de $\mathbf{X}_1(I \times J)$ é organizado, um abaixo do outro, formando um vetor de dimensões $IJ \times 1$, denominado $\text{Vec}(\mathbf{X}_1)$.
- 2) Toma-se a matriz, $\mathbf{X}_2$, de I observações por J variáveis, para a segunda localidade, k_2 . Em seguida, cada vetor coluna de $\mathbf{X}_2(I \times J)$ é organizado, um abaixo do outro, formando um segundo vetor de dimensões $IJ \times 1$, denominado $\text{Vec}(\mathbf{X}_2)$.
- 3) Repete-se o procedimento até que todas as K localidades tenham sido usadas obtendo-se uma nova matriz, $\mathbf{X}$, cujas colunas são ocupadas pelos vetores $\text{Vec}(\mathbf{X}_1), \text{Vec}(\mathbf{X}_2), \text{Vec}(\mathbf{X}_3), \dots, \text{Vec}(\mathbf{X}_K)$. Essa matriz possuirá dimensões $IJ \times K$.

Para PCA, o modelo para os dados de ordem superior é dado pela projeção das colunas de $\mathbf{X}$ (as quais são as variáveis $\text{Vec}(\mathbf{X}_1), \text{Vec}(\mathbf{X}_2), \dots, \text{Vec}(\mathbf{X}_K)$), em um subespaço ocupado pelas componentes principais (PC) ou fatores ($\mathbf{A}$), $\text{Vec}(\mathbf{A}_1), \text{Vec}(\mathbf{A}_2), \text{Vec}(\mathbf{A}_3), \dots, \text{Vec}(\mathbf{A}_R)$, os quais são organizados em uma matriz $\mathbf{A}$ ($IJ \times K$). Então a solução pode ser obtida tornando-se mínima a função objetiva (loss function), a qual pode ser escrita (Kiers, 1991)

$$\text{PCA-SUP}(\mathbf{B}, \mathbf{A}) = \|\mathbf{X} - \mathbf{A} \mathbf{B}^T\|_F^2 \quad (3.64)$$

onde $\mathbf{B}(K \times R)$ = matriz de pesos ("loadings") para as localidades, e $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_R$, de dimensões $(I \times J)$ contém as coordenadas para as IJ combinações objetos-variáveis.

A operação para tornar mínima a equação acima sobre matrizes arbitrárias $\mathbf{A}$ e $\mathbf{B}$ de dimensões apropriadas, é simplesmente a aplicação de PCA à matriz $\mathbf{X}(IJ \times K)$. Em PCA multimodo, as colunas dessa matriz de dados referem-se às localidades (ou condições) e as linhas às combinações entre objetos (amostras) e variáveis (Kiers, 1991).

As Figuras 3.14 e 3.15 mostram, esquematicamente, a construção da estrutura descrita acima.

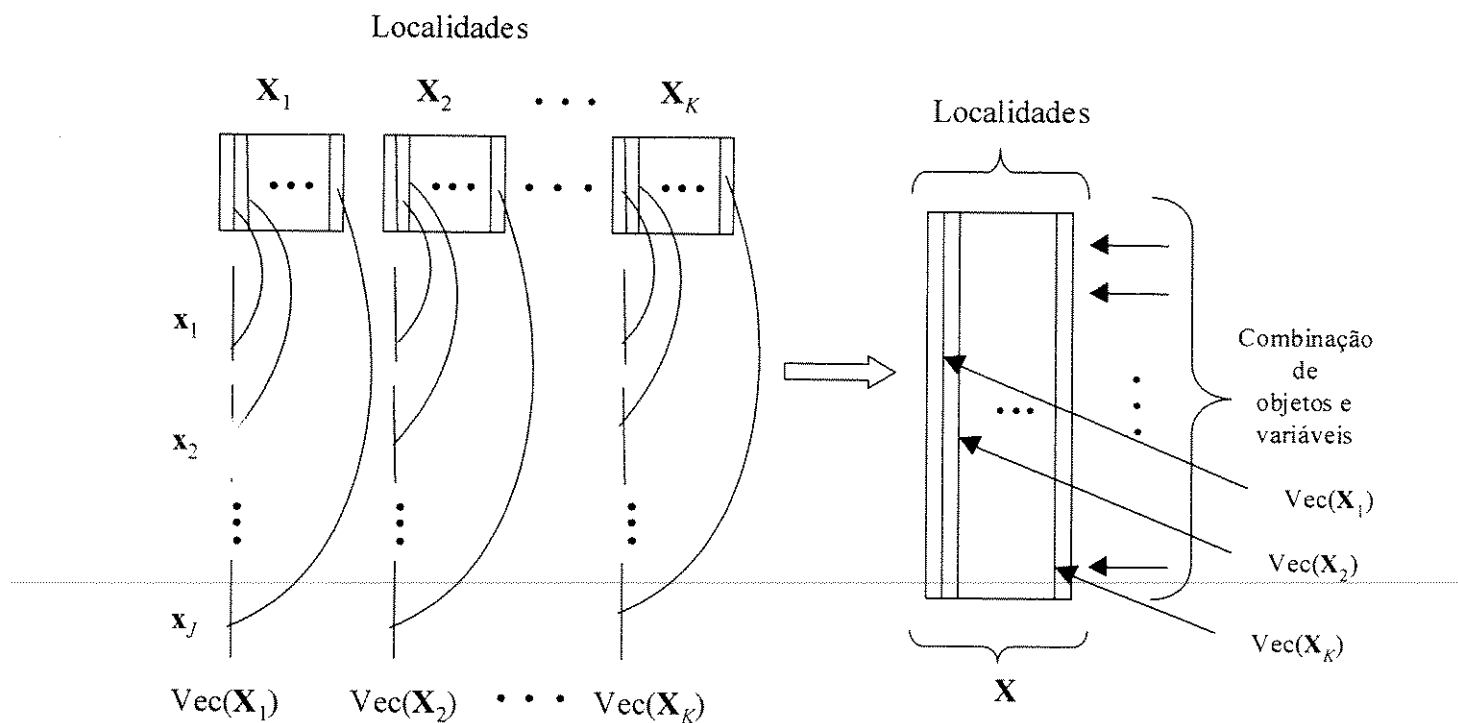


Fig.3.14: “Vetorização” de $X_1, X_2, \dots, X_K$

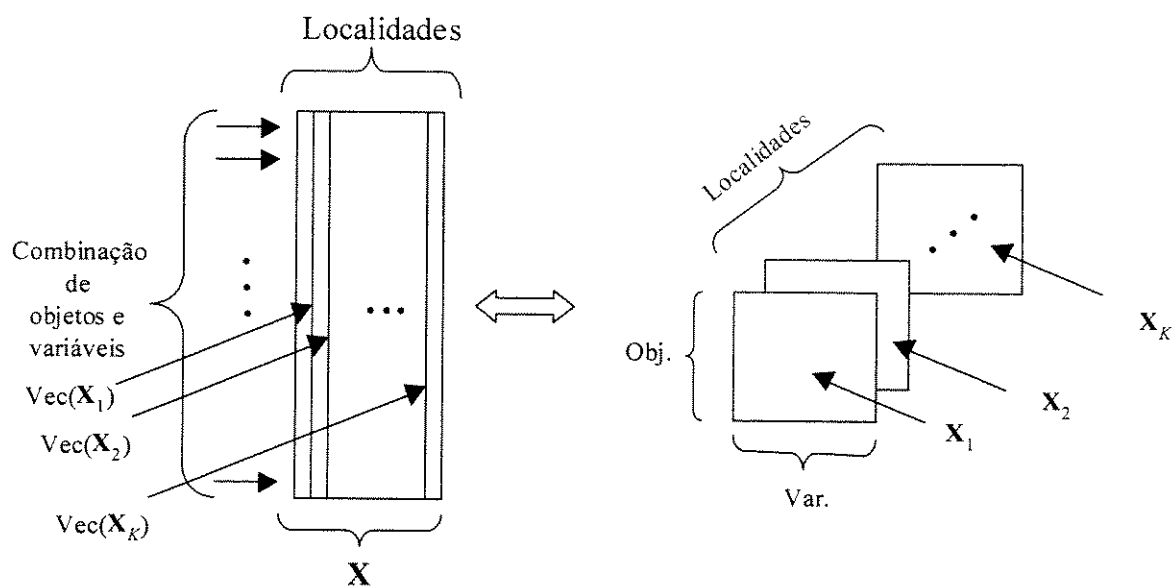


Fig.3.15: Os vetores $\text{Vec}(X_1), \text{Vec}(X_2), \dots, \text{Vec}(X_K)$ correspondem às localidades representadas pelas matrizes $X_1, X_2, \dots, X_K$

- **Modelo de TUCKER**

Um dos caminhos para se compreender os métodos/modelos de ordem superior pode ser por meio do estudo daquele considerado como o mais geral. E o modelo mais geral para a análise de dados de ordem superior é o de Tucker (Bro, 1998, p.43-49; Bro *et alli*, 1997; Kroonenberg, 1989, p.48-62; Tucker, 1966; Kiers, 1991; Geladi, 1989; Zeng e Hopke, 1992), o qual pode ser formulado na forma escalar da seguinte maneira:

$$x_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R a_{ip} b_{jq} c_{kr} g_{pqr} + e_{ijk} \quad (3.65)$$

onde a_{ip} , b_{jq} , c_{kr} , são os elementos das matrizes componentes **A** (no 1º modo ou modo A), **B** (no 2º modo ou modo B) e **C** (no 3º modo ou modo C) de dimensões $I \times P$, $J \times Q$ e $K \times R$, respectivamente (Kiers, 2000), e, x_{ijk} , e_{ijk} e g_{pqr} são elementos de arranjos de ordem superior, isto é, são elementos, respectivamente, dos arranjos multimodo, **X**, **E**, e **G**, sendo esse último denominado *núcleo* do modelo.

Suponha-se que seja dado um arranjo **X**($I \times J \times K$), e que se procura um modelo Tucker de pseudo posto P no primeiro, Q no segundo, e R no terceiro modos. Fazendo-se o desdobramento de maneira a se obter **X**($I \times JK$), o modelo pode ser escrito na forma matricial (Bro, 1998, p.44; Kiers, 2000; Kroonenberg e Jan de Leeuw, 1980)

$$\mathbf{X} = \mathbf{A} \mathbf{G} (\mathbf{C}^T \otimes \mathbf{B}^T) + \mathbf{E} \quad (3.66)$$

onde o símbolo $\otimes$ representa o produto de Kronecker (ver Apêndice A5). Observe que **X**($I \times JK$), **E**($I \times JK$) e **G**($P \times QR$) são representados na equação (3.66) na forma de matrizes sendo, portanto, estruturas em dois modos.

O algoritmo proposto por Kroonenberg e de Leeuw (Kroonenberg e Jan Leeuw, 1980), denominado TUCKALS-3, ajusta o modelo por quadrados mínimos alternantes, tornando mínima a função (Bro, 1998, p.70; Kroonenberg e Jan Leeuw, 1980; kiers, 1991; Henrion, 1994)

$$\text{TUCKALS3}(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{G}) = \| \mathbf{X} - \mathbf{A} \mathbf{G} (\mathbf{C}^T \otimes \mathbf{B}^T) \|^2_F \quad (3.67)$$

com $\mathbf{G}$ de dimensões $(P \times QR)$, e com $\mathbf{A}$, $\mathbf{B}$ e $\mathbf{C}$ ortogonais (Bro, 1998, p.70), sem perda de generalidade (Kiers, 1991). Isto é, tanto para Bro como para Kiers se $\mathbf{A}$, $\mathbf{B}$ e $\mathbf{C}$ não forem ortogonais chega-se ao mesmo mínimo, porém a restrição de ortogonalidade é visto pelo primeiro como uma necessidade e pelo segundo não como uma necessidade, mas sim como uma conveniência computacional e de interpretação (Kiers, 1991).

- **Modelo PARAFAC**

O modelo PARAFAC é dado, na forma escalar pela expressão

$$x_{ijk} = \sum_{r=1}^R a_{ir} b_{jr} c_{kr} + e_{ijk} \quad (3.68)$$

onde a_{ir} , b_{jr} , e c_{kr} são os elementos das matrizes componentes $\mathbf{A}$ (no 1º modo ou modo A), $\mathbf{B}$ (no 2º modo ou modo B) e $\mathbf{C}$ (no 3º modo ou modo C), porém de dimensões $I \times R$, $J \times R$ e $K \times R$, respectivamente (Bro, 1998, p.22-27; Kiers, 1991).

Comparando-se as equações 3.65 e 3.68 acima observa-se que o modelo PARAFAC é um caso particular de Tucker, ou seja, é uma versão do modelo Tucker com restrição já que todos os elementos do núcleo são iguais a zero, com exceção dos elementos g_{111} , g_{222} ... g_{rrr} , os quais são iguais a um (Kiers, 2000). Para se obter a versão do modelo PARAFAC em notação matricial, pode-se partir do modelo Tucker substituindo-se o núcleo deste por um arranjo de dimensões $R \times R \times R$ com o valor um na super-diagonal desse arranjo, o qual é chamado super-diagonal unidade, $\mathbf{I}_{\text{sdu}} (R \times RR)$ (Kiers, 2000). O modelo pode ser, então, escrito da seguinte maneira

$$\mathbf{X} = \mathbf{A} \mathbf{I}_{\text{sdu}} (\mathbf{C}^T \otimes \mathbf{B}^T) + \mathbf{E} \quad (3.69)$$

O modelo PARAFAC pode ser ajustado aos dados minimizando-se a função objetiva (Bro, 1998, p.59; Kiers, 1991)

$$P(\mathbf{A}, \mathbf{B}, \mathbf{C}) = \|\mathbf{X} - \mathbf{A} \mathbf{I}_{\text{sdu}} (\mathbf{C}^T \otimes \mathbf{B}^T)\|_F^2, \quad (3.70)$$

- **Estrutura e Hierarquia dos Modelos**

Nesse t3pico pretende-se mostrar as estreitas rela33es que os modelos de ordem superior mant3m com PCA e SVD. Os modelos de ordem superior tratados nessa tese s3o, aqueles em tr3s modos, chamados coletivamente de “three mode component analysis”, e podem ser considerados uma generaliza33o de PCA convencional, ou mais precisamente, de SVD (Kroonenberg, 1989, p.20; Bro, 1998, p.43; Geladi, 1989; Smilde, 1992; Henrion, 1994; Kiers, 1991). Kiers (1991), inclusive mostra que todos os modelos s3o uma vers3o, com restri33es, de PCA sobre as localidades (ou ocasi3es).

Para observar essas rela33es hier3rquicas deve-se observar as estruturas que os m3todos/modelos acomodam. A estrutura subjacente a PCA 3 a de um modelo bilinear (Bro, 1998, p.22), assim como SVD. J3 nos modelos de Tucker e PARAFAC a estrutura 3 trilinear. Isto pode ser verificado (para pseudo posto R , e omitidos os termos para o erro em cada caso) comparando-se as equa33es 3.71, 3.72, 3.73 e 3.74, respectivamente, para PCA, SVD, PARAFAC e TUCKER.

$$x_{ij} = \sum_{r=1}^R a_{ir} b_{jr} \quad \text{PCA} \quad (3.71)$$

$$x_{ij} = \sum_{r=1}^R u_{ir} b_{jr} g_{rr} \quad \text{SVD} \quad (3.72)$$

$$x_{ijk} = \sum_{r=1}^R a_{ir} b_{jr} c_{kr} g_{rrr} \quad \text{PARAFAC} \quad (3.73)$$

$$x_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R a_{ip} b_{jq} c_{kr} g_{pqr} \quad \text{TUCKER} \quad (3.74)$$

Vale notar que os significados de a_{ir} e u_{ir} , respectivamente, nas equa33es 3.71 (para PCA) e 3.72 (para SVD) s3o diferentes. Enquanto na equa33o 3.71 os a_{ir} s3o elementos dos autovetores de $\mathbf{X}$, na equa33o 3.72 os u_{ir} representam os elementos dos vetores singulares 3 esquerda de $\mathbf{X}$ (elementos dos autovetores de $\mathbf{X} \mathbf{X}^T$). No entanto, o interesse aqui est3 na estrutura dos modelos.

A formula33o SVD para PCA dada pela equa33o 3.72 pode ser posta alternativamente como (Bro, 1998, p.44)

$$\mathbf{X} = \mathbf{U} \mathbf{G} \mathbf{B}^T \quad (3.75)$$

Como pode ser observado, os índices dos valores singulares, g_{rr} , indicam que $\mathbf{G}$ é diagonal. Isto significa que as r -ésimas colunas de $\mathbf{U}$ interagem somente com as r -ésimas colunas de $\mathbf{B}$ (Bro, 1998, p.44). Então, o “elo” de ligação que promove essa relação de proximidade é a decomposição por valor singular, SVD. A matriz diagonal $\mathbf{G}$ aqui é a mesma matriz diagonal $\mathbf{\Lambda}$ usada na equação 3.54, assim como $\mathbf{B}$ corresponde à matriz $\mathbf{P}$ da equação 3.54 (quando nessa equação considera-se não o posto R , mas o pseudo posto, R_{pp} , de $\mathbf{X}$).

Então, PCA (para $\mathbf{X}$ de posto R) pode ser entendido como uma soma de R produtos externos

$$\mathbf{X} = \mathbf{u}_1 g_{11} \mathbf{b}_1^T + \mathbf{u}_2 g_{22} \mathbf{b}_2^T + \dots + \mathbf{u}_R g_{RR} \mathbf{b}_R^T + \mathbf{E} \quad (3.76)$$

ou,

$$\mathbf{X} = \mathbf{a}_1 \mathbf{b}_1^T + \mathbf{a}_2 \mathbf{b}_2^T + \dots + \mathbf{a}_R \mathbf{b}_R^T + \mathbf{E} \quad (3.77)$$

Por outro lado, para $\mathbf{X}$ de pseudo posto R_{pp} , PCA pode ser entendido como uma soma de R_{pp} produtos externos.

As equações 3.76 e 3.77 mostram que os autovetores $\mathbf{a}$ correspondem ao produto dos vetores singulares $\mathbf{u}$ pelos elementos diagonais g_{rr} . Todavia, para o caso em que os elementos da diagonal, g_{rr} , assumam somente o valor unitário, a versão SVD de PCA para um modelo com duas componentes pode ser formulada como (Bro, 1998, p.46)

$$\mathbf{X} = [\mathbf{a}_1 \ \mathbf{a}_2] \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} [\mathbf{b}_1 \ \mathbf{b}_2]^T \quad (3.78)$$

A equação 3.78 corresponde à uma generalização de PCA ao modelo PARAFAC. Isto pode ser constatado por verificação da matriz $\mathbf{G}$, usada em cada caso, PCA e PARAFAC. Então pode-se concluir que PARAFAC corresponde a um tipo de generalização de PCA.

Entretanto, a equação 3.78 pode ser formulada de um modo mais geral como

$$\mathbf{X} = [\mathbf{a}_1 \ \mathbf{a}_2] \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} [\mathbf{b}_1 \ \mathbf{b}_2]^T \quad (3.79)$$

Nesse caso, a equação 3.72e pode ser generalizada ao modelo Tucker3 (ver equação 3.74). Assim, o modelo Tucker3 pode ser visto como uma generalização de PCA à ordens superiores (Bro, 1998, p.46).

Quando, a um conjunto de dados, aplica-se PCA via decomposição SVD, a matriz diagonal Λ (ver equações 3.54 e 3.55) contém os valores singulares (raízes quadradas dos autovalores da matriz sendo decomposta). Os quadrados dos valores singulares fornecem os “pesos” para as projeções dos objetos (os escores) nos novos eixos (as PCs). Uma representação esquemática, como a da Figura 3.11, mostra esses aspectos, tal que as operações algébricas envolvidas tornam-se mais evidentes.

Quando a decomposição é efetuada sobre estruturas multimodo, os “pesos” são “distribuídos” por meio de um arranjo, o núcleo, $\underline{\mathbf{G}}$, para o modelo de Tucker, e a super diagonal, $\underline{\mathbf{I}}$, para o modelo PARAFAC. Tais arranjos contém um número de modos igual ao da estrutura original, $\underline{\mathbf{X}}$ (ver equações 3.65 e 3.66 para o modelo de Tucker, e as equações 3.68 e 3.69 para o modelo PARAFAC). O esquema da Figura 3.16 mostra a decomposição da matriz de dados, $\underline{\mathbf{X}}$, para o modelo de Tucker.

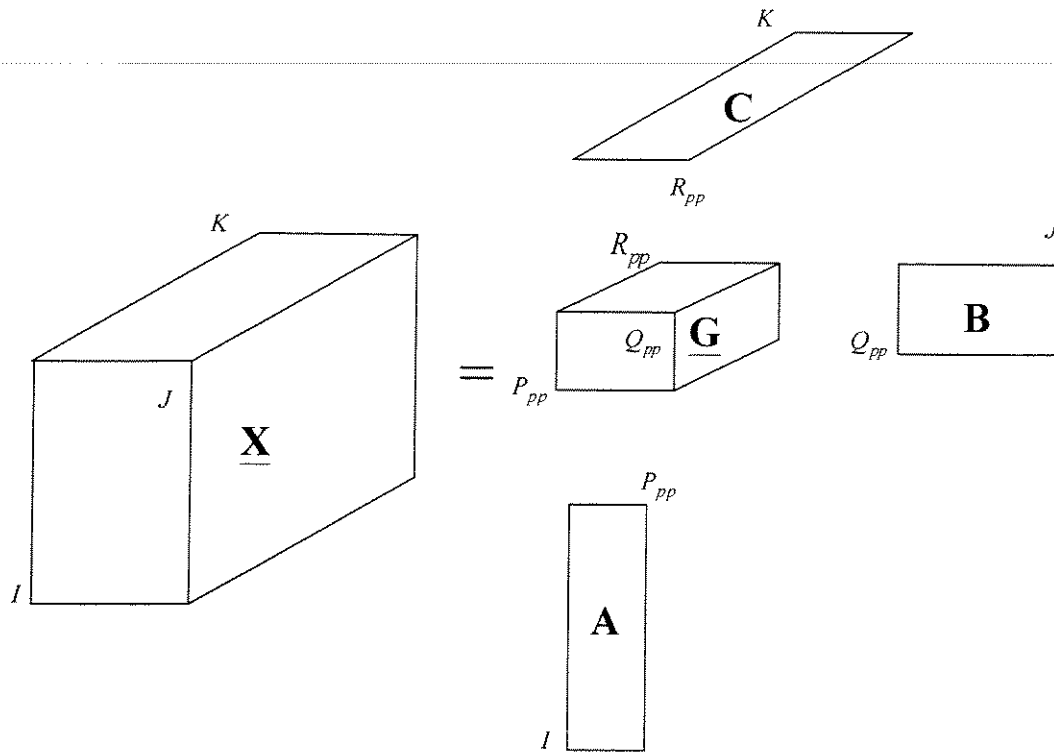


Fig.3.16: Esquema de uma decomposição para o modelo de Tucker. É importante observar as dimensões das matrizes de componentes, $\underline{\mathbf{A}}$, $\underline{\mathbf{B}}$ e $\underline{\mathbf{C}}$ e do núcleo, $\underline{\mathbf{G}}$. No núcleo, P_{pp} , Q_{pp} e R_{pp} correspondem aos pseudo postos, respectivamente, de $\underline{\mathbf{A}}$, $\underline{\mathbf{B}}$ e $\underline{\mathbf{C}}$.

O esquema da Figura 3.17 mostra a decomposição da matriz de dados, $\underline{\mathbf{X}}$, para o modelo PARAFAC.

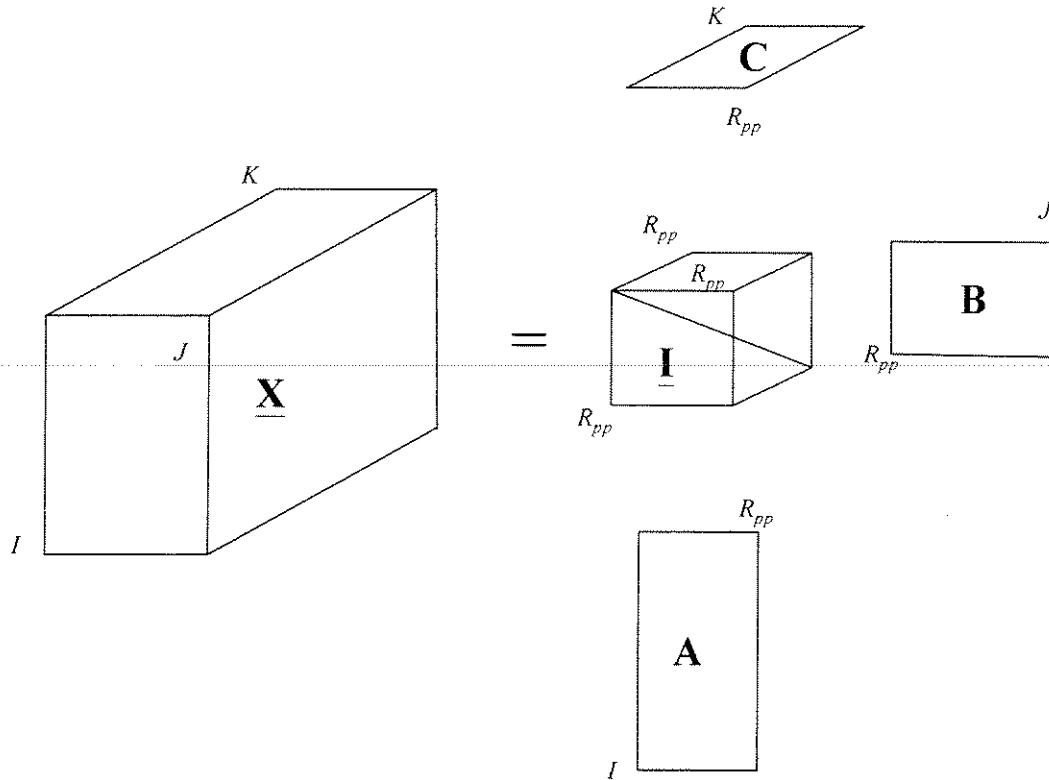


Fig.3.17: Esquema de uma decomposição para o modelo PARAFAC. É importante observar as dimensões das matrizes de componentes, $\mathbf{A}$, $\mathbf{B}$ e $\mathbf{C}$ e do arranjo contendo a superdiagonal, $\underline{\mathbf{I}}$. Como se vê, esse arranjo tem dimensões R_{pp} nos três modos e os elementos fora de sua superdiagonal são nulos, indicando que o PARAFAC é um caso particular do modelo de Tucker.

Uma comparação entre os modelos mostra que a decomposição para o modelo de Tucker ocorre com a participação de um núcleo, $\underline{\mathbf{G}}$, o qual contém os escores que dão a importância relativa às matrizes de componentes dos loadings, $\mathbf{A}$, $\mathbf{B}$ e $\mathbf{C}$. No caso do modelo PARAFAC, a decomposição é também mediada por um arranjo em três modos, a superdiagonal, $\underline{\mathbf{I}}$ (diagonal do cubo), porém com todos os elementos fora da superdiagonal, nulos. Comparando-se esses modelos com PCA e com SVD, observa-se que as funções de $\underline{\mathbf{G}}$ e de $\underline{\mathbf{I}}$ para as estruturas multimodo, são assumidas nos casos em dois modos pela matriz diagonal Λ . Porém, como discutido acima, do ponto vista estrutural, o arranjo $\underline{\mathbf{G}}$ pode ser

construído de forma a se encontrar dois tipos diferentes de generalizações de PCA. Um deles se generaliza ao modelo PARAFAC e o outro ao modelo Tucker3.

- **Unicidade dos Modelos**

Uma breve comparação entre os modelos no que diz respeito à solução única construir-se-á como um dos argumentos para justificar, como será visto na Seção 3.3, a qual discute as bases da metodologia, o uso do modelo PARAFAC como uma etapa intermediária antes da aplicação do modelo de TUCKER.

Uma vantagem do modelo PARAFAC sobre os outros é a unicidade da solução. Em modelos bilineares (Bro, 1998, p.24) como a solução obtida pelo método PCA, por exemplo, existe o bem conhecido problema de liberdade de rotação. Isto fez desencadear o surgimento de uma série de métodos diferentes para a obtenção de modelos melhor interpretáveis do que o PCA, ou a rotação da própria solução PCA às soluções mais apropriadas. Entretanto, a maioria dessas tentativas levou a métodos mais ou menos arbitrários ou possuindo propriedades mal definidas, diferentemente do que ocorre com o PARAFAC. Devido às propriedades de unicidade do PARAFAC, ele é, às vezes, referido como modelo de fatores únicos. O modelo PARAFAC não apenas determina de forma única o subespaço ocupado pelos dados, como também a posição dos eixos que o define. Daí o nome eixos ou fatores únicos (Bro, 1998, p.25).

O significado matemático de unicidade é o de que as matrizes de componentes estimadas pelo modelo PARAFAC não podem sofrer rotação sem perda de ajuste (Bro, 1998, p.27). Para verificar isto, suponha-se um modelo qualquer em dois modos (por ex. PCA) com R componentes,

$$\mathbf{X} = \mathbf{A} \mathbf{B}^T + \mathbf{E} \quad (3.80)$$

Sobre tal modelo pode-se aplicar uma rotação por meio de qualquer matriz não singular, $\mathbf{P}(R \times R)$ de modo que

$$\mathbf{A} \mathbf{P} (\mathbf{P}^{-1} \mathbf{B}^T) = [\mathbf{A} \mathbf{P}] [\mathbf{B} (\mathbf{P}^{-1})^T]^T \quad (3.81)$$

Fazendo

$$\mathbf{Y} = [\mathbf{A} \mathbf{P}] \quad (3.82)$$

e

$$\mathbf{Z} = [\mathbf{B} (\mathbf{P}^{-1})^T] \quad (3.83)$$

e substituindo em uma expressão para um mesmo modelo bilinear

$$\mathbf{V} = \mathbf{Y} \mathbf{Z}^T \quad (3.84)$$

obtem-se:

$$\mathbf{V} = \mathbf{Y} \mathbf{Z}^T = [\mathbf{A} \mathbf{P}] [\mathbf{B} (\mathbf{P}^{-1})^T]^T = \mathbf{A} \mathbf{P} \mathbf{P}^{-1} \mathbf{B}^T = \mathbf{A} \mathbf{B}^T = \mathbf{X} - \mathbf{E} \quad (3.85)$$

A rotação foi aplicada à $\mathbf{A} \mathbf{B}^T$, portanto, $\mathbf{E}$ permanece inalterado. Como se observa, comparando-se as equações 3.80 e 3.85, o modelo obtido tem exatamente o mesmo ajuste para os dados apesar da rotação. Isto é freqüentemente denominado “*ambigüidade de rotação*” de PCA (Bro, 1998, p.27).

O elemento x_{ij} de um modelo em dois modos pode ser representado pela expressão (Harshman and Lundy, 1994):

$$x_{ij} = a_{i1}b_{j1} + a_{i2}b_{j2} + \dots + a_{iR}b_{jR} + e_{ij} \quad (3.86)$$

o qual pode ser generalizado a um modelo em três modos mediante a equação

$$x_{ijk} = a_{i1}b_{j1}c_{k1} + a_{i2}b_{j2}c_{k2} + \dots + a_{iR}b_{jR}c_{kR} + e_{ijk} \quad (3.87)$$

o qual pode ser escrito de forma concisa como

$$\mathbf{X}_k = \mathbf{A} \mathbf{D}_k \mathbf{B}^T + \mathbf{E}_k \quad (3.88)$$

onde $\mathbf{X}_k$ ($I \times J$) é o k-ésimo plano de um arranjo $\underline{\mathbf{X}}$ ($I \times J \times K$), $\mathbf{A}$ ($I \times R$) é a matriz de componentes do modo A, $\mathbf{B}$ ($J \times R$) é a matriz de componentes do modo B, e $\mathbf{D}_k$ é uma matriz diagonal ($R \times R$) com os elementos da diagonal tomados da k-ésima linha da matriz de componentes do modo C, $\mathbf{C}$ ($K \times R$), e $\mathbf{E}_k$ é o k-ésimo plano do arranjo $\underline{\mathbf{E}}$.

O tipo de “indeterminação” citado acima para o modelo bilinear não existe no PARAFAC como se pode observar quando ao modelo

$$\mathbf{X}_k = \mathbf{A} \mathbf{D}_k \mathbf{B}^T \quad (3.89)$$

.....aplica-se uma rotação, podendo o mesmo vir igualmente a ser representado como.....

$$(\mathbf{A} \mathbf{T}) (\mathbf{T}^{-1} \mathbf{D}_k \mathbf{S}) (\mathbf{S}^{-1} \mathbf{B}^T) = (\mathbf{A} \mathbf{T}) (\mathbf{T}^{-1} \mathbf{D}_k \mathbf{S}) [\mathbf{B} (\mathbf{S}^{-1})^T]^T \quad (3.90)$$

Fazendo

$$\mathbf{W} = (\mathbf{A} \mathbf{T}) \quad (3.91)$$

$$\mathbf{Y} = (\mathbf{T}^{-1} \mathbf{D}_k \mathbf{S}) \quad (3.92)$$

e

$$\mathbf{Z} = [\mathbf{B} (\mathbf{S}^{-1})^T] \quad (3.93)$$

e substituindo-se em uma expressão para um modelo trilinear

$$\mathbf{U} = \mathbf{W} \mathbf{Y}_k \mathbf{Z}^T, \quad (3.94)$$

obtem-se:

$$\mathbf{U} = \mathbf{W} \mathbf{Y}_k \mathbf{Z}^T = (\mathbf{A} \mathbf{T}) (\mathbf{T}^{-1} \mathbf{D}_k \mathbf{S}) [\mathbf{B} (\mathbf{S}^{-1})^T]^T \quad (3.95)$$

ou,

$$\mathbf{U} = \mathbf{W} \mathbf{Y}_k \mathbf{Z}^T = (\mathbf{A} \mathbf{T}) (\mathbf{T}^{-1} \mathbf{D}_k \mathbf{S}) (\mathbf{S}^{-1}) \mathbf{B}^T \quad (3.96)$$

A equação 3.89 indica que em vez de $\mathbf{A}$, $\mathbf{B}$ e $\mathbf{D}_k$, poderiam ser igualmente usadas as matrizes de loadings $\mathbf{AT}$, $\mathbf{B}(\mathbf{S}^{-1})^T$ e $\mathbf{T}^{-1}\mathbf{D}_k\mathbf{S}$, respectivamente, causando, à primeira vista, a impressão de que a solução é idêntica àquela que não sofreu rotação. Entretanto, para que o

modelo PARAFAC se mantenha, a matriz $\mathbf{Y} = (\mathbf{T}^{-1} \mathbf{D}_k \mathbf{S})$, do 3º modo, deve ser diagonal. Isto significa que somente algumas matrizes $\mathbf{T}$ e $\mathbf{S}$ são válidas: aquelas que preservem $\mathbf{Y}$ como diagonal (Harshman, 1972, em Bro, 1998, p.27).

É oportuno avaliar a liberdade de rotação apresentada também pelo modelo de TUCKER, o qual não é estruturalmente único como o PARAFAC. Isto pode ser verificado quando ao modelo

$$\mathbf{X} = \mathbf{A} \mathbf{G} (\mathbf{C}^T \otimes \mathbf{B}^T) + \mathbf{E} \quad (3.97)$$

aplica-se uma rotação na matriz de loadings do primeiro modo (ou mais) por meio de uma matriz não singular quadrática, $\mathbf{S}$. Uma contra rotação em $\mathbf{G}$ por meio da inversa da matriz $\mathbf{S}$ conduz a um modelo que é equivalente ao original indicando a possibilidade de rotação. De fato, tem sido mostrado que o modelo estrutural é tão redundante que vários parâmetros (mais da metade) no núcleo, $\mathbf{G}$, podem ser forçados a assumir o valor zero sem que se mude o ajuste do modelo (Kiers *et alli*, 1997, em Bro, 1998, p.47). Essa característica do modelo é particularmente explorada nessa tese.

Para a realização das análises são usados os modelos PCA, PARAFAC e de Tucker. A utilização de PCA está fundamentada no fato de que, do ponto de vista de ajuste, PCA (o modelo mais flexível) sempre ajusta melhor os dados do que os modelos de Tucker e o PARAFAC, nessa ordem. Porém, o último usa um menor número de graus de liberdade sendo, então, o mais restritivo, o que significa que se o PARAFAC pode modelar um conjunto de dados, este também poderá sê-lo pelos modelos de Tucker e PCA (Bro, 1997).

Por outro lado, o PARAFAC apresenta solução única (Bro, 1997), o que possibilita encontrar o número correto de componentes, apesar de não permitir interações entre os pesos de índices diferentes nas matrizes de componentes (Kiers, 2000), fato que não ocorre com o modelo de Tucker, justificando, finalmente, a sua utilização.

Desse modo, foi possível considerar as qualidades dos modelos PCA (melhor ajuste) com as do PARAFAC (solução única) na aplicação do modelo de Tucker, como é mostrado na discussão sobre as bases da metodologia.

3.3 – AS BASES DA METODOLOGIA

3.3.1 – A Construção da Metodologia

A metodologia adotada nessa tese vem sendo discutida desde a Seção 3.1. Na presente seção pretende-se abordar os pressupostos que nortearam a construção da metodologia na qual se inserem os métodos e modelos aplicados.

O ponto de partida nesse trabalho possui como base o estudo do comportamento de vários compostos na atmosfera, mais precisamente na baixa troposfera, oriundos de diferentes fontes de emissão. O interesse ficou circunscrito à obtenção de informações sobre perfil dos poluentes com o objetivo de identificar a principal fonte (ou principais fontes) de emissão de poluentes para a atmosfera. O trabalho ficara inicialmente restrito ao estudo do *Ciclo do Monóxido de Carbono*, CO, sendo, posteriormente, estendido ao estudo do *Ciclo do Metano*, CH₄, e ao *Ciclo dos Hidrocarbonetos Não Metano*, NMHC.

Essas denominações referem-se aos ciclos criados para destacar a influência que cada um dos grupos de reações químicas ocorrendo na atmosfera, relacionadas, respectivamente, com o CO, o CH₄ e os NMHC, possui nos mecanismos e na acumulação de vários poluentes na baixa troposfera..

O primeiro passo dentro da metodologia foi dado em observância aos objetivos. Estes indicaram a Análise Exploratória de Dados como caminho rumo àquelas informações sobre os dados coletados. Após circunscrever o trabalho no âmbito da análise exploratória de dados, buscou-se construir o pilar central da “arquitetura” metodológica. Isto significa que a metodologia empregada tem fundamento em alguns pressupostos chave.

O primeiro deles baseia-se na constatação de que a aplicação de modelos bilineares, embora necessária, não é suficiente para a resolução do problema que se propõe investigar. Se o problema envolve a medição de variáveis sobre várias amostras, pode-se usar PCA, por exemplo, para obter informações sobre o comportamento dos dados. Porém, o problema envolve a medição de variáveis, ao longo de um dado período e em diferentes locais. Apenas isto já seria motivo suficiente para se aplicar estruturas multimodo aos dados. Entretanto não foi esse requisito que norteou a construção da metodologia.

Das fontes existentes, duas se apresentavam como potencialmente emissoras: *a de veículos automotores e a das indústrias*.

Aqui entra o segundo pressuposto. O problema é atacado a partir da emissão do poluente. Este, após a emissão, fica sujeito a interações com outras substâncias presentes na atmosfera (ou que são emitidas simultaneamente com ele pela mesma fonte ou por fontes diferentes), e ficam também à mercê do deslocamento de massas de ar, tal que o destino final é função de uma série de fatores. Pode-se considerar, entretanto, o destino como um resultado de emissão, reação e dispersão. Nessa tese os termos reação e dispersão são acomodados em uma componente que recebe o nome de *acumulação*. O conjunto dos fatores que influencia o destino final de um poluente na atmosfera pode, então, ser acomodado em dois fundamentais: *emissão* e *acumulação*. Assim, o destino pode ser considerado como um resultado de emissão e acumulação.

O primeiro pressuposto por si só (necessidade de estruturas multimodo), e apenas ele, entretanto, não é suficiente para alcançar os objetivos propostos: identificar fonte(s) de emissão de poluentes.

Um olhar inadvertido sobre os dados poderia apenas indicar: “compostos foram medidos na atmosfera por um determinado período, em várias localidades. Isto induziria à organização dos dados em (períodos $\times$ variáveis $\times$ localidades)”.

Ou seja, sob esse prisma, o problema envolve no mínimo três modos. O modo cuja dimensão é representado pelas amostras ou linhas (períodos), o modo para as variáveis (os poluentes) e o modo para as condições ou localidades (as estações). Isto sugere a aplicação de modelos de ordem superior. Mas, como organizar cada modo no arranjo de ordem superior? A resposta não é imediata. O objetivo não é simplesmente o de se encontrar algumas correlações e a redução de dimensionalidade do sistema seja de ordem superior ou não. Observe-se que realizou-se a medição de variáveis a cada hora ao longo de um ano inteiro, na verdade, ao longo de quatro anos. A análise deve ser cuidadosamente planejada.

Foi necessário, então, dividir o problema grande em problemas menores, analisá-los adequadamente e depois juntá-los aos poucos até à solução final. Ao invés de se concentrar na análise de escores e pesos (loadings) para obtenção de informação interpretável sob o ponto de vista da “separação” dos dados como pontos no subespaço representado por escores e pesos, procurou-se concentrar a atenção nas curvas representativas das variações sistemáticas que os dados viessem a apresentar. Estabeleceu-se como meta identificar tais variações sistemáticas por meio de métodos baseados em decomposições.

Projetar cada vetor correspondente a um ponto (um dado medido) no subespaço formado pelos pesos poderia ser o caminho, porque se a estrutura for organizada adequadamente, esse subespaço conterá vetores com seus respectivos pesos e, portanto, as contribuições que cada uma das variáveis possui neles, dando informações sobre o seu comportamento. Assim, a forma de entrada dos dados para a análise é fundamental para que se possa observar claramente essas variações sistemáticas.

Explora-se na construção da metodologia, exatamente a forma adequada para se encontrar tais variações. Isto é, como construir o arranjo de dados que permita obter as informações desejadas – *as variações sistemáticas presentes nos dados*. O resultado que se extrai dessas componentes (pesos) mostra claramente variações sistemáticas, as quais podem ser associadas à(s) fonte(s). Isto remete o problema para o formato da tabela contendo os dados como será visto adiante.

O modelo proposto visa, em primeiro lugar, encontrar variações sistemáticas em variável (poluente) medida, em uma dada estação (local de medição), e verificar se essas variações sistemáticas podem ser associadas à uma (ou mais) fonte(s) de emissão de poluentes. Em segundo lugar, estudar as interações intra e inter ciclos. Finalmente estender a análise às outras estações e aos outros anos (períodos).

3.3.2 – As Etapas da Construção da metodologia

As medições foram realizadas a cada hora, durante 24 horas, para todos os dias da semana, durante um ano inteiro, em cada estação de medição.

Para um dado poluente, por exemplo, o CO, medido em uma dada estação durante um ano, deseja-se obter informações sobre suas variações sistemáticas diária e semanal. Observe-se então que, a partir dos dados, pode-se construir um arranjo constituído por uma variável medida, o CO, durante o período de um ano. Isto é um vetor de dados. Se outras variáveis são adicionadas, pode-se ter uma matriz de dados (períodos $\times$ variável). Finalmente, se outras estações de medição forem incluídas no arranjo, pode-se chegar a um “cubo” ou “paralelepípedo” de dados (períodos $\times$ variável $\times$ estação).

Entretanto, deseja-se conhecer as variações sistemáticas apenas para o CO, em uma única estação, durante um ano, a partir de informações contidas em um vetor de dados.

Uma inspeção sobre esse vetor de dados (Figura 3.18) não revela qualquer variação sistemática, embora ela esteja, de fato, presente nele.

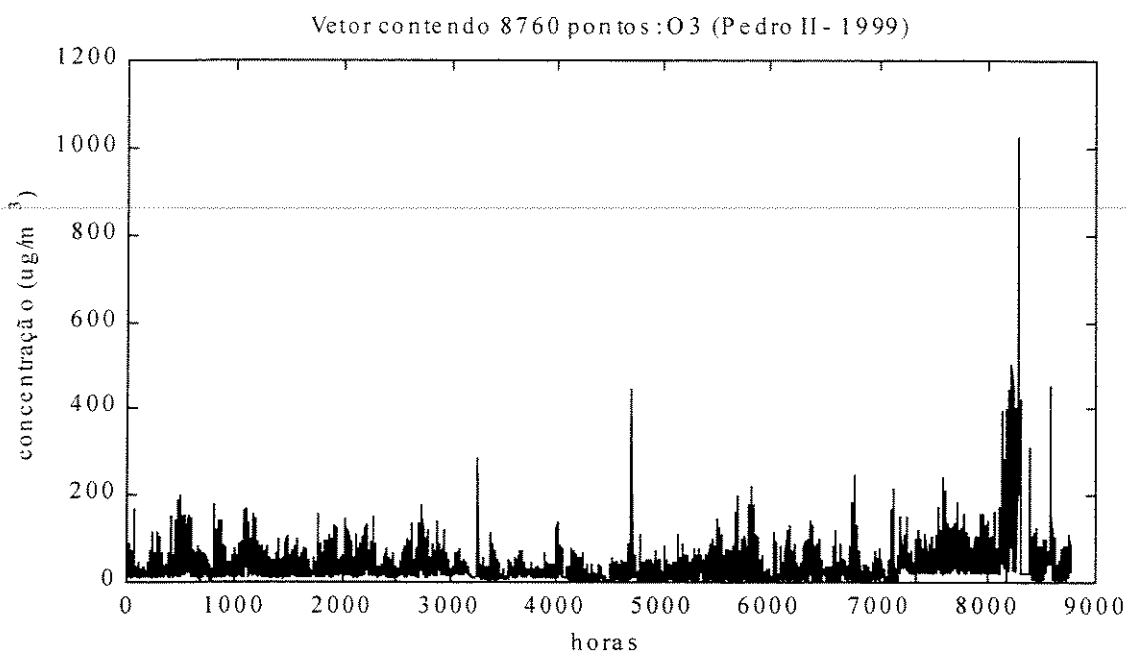


Fig.3.18 : Gráfico contendo 8760 dados.

Como se observa, não é possível perceber qualquer variação sistemática nos dados da Figura 3.18. São visíveis alguns pontos nos quais os valores são muito elevados. Isto pode apenas indicar que tais dados sejam anômalos, porém não se pode avaliar que tipo de perfis estão por trás desses dados na forma em que estão apresentados.

É necessário, portanto organizá-los de modo que possam fornecer não apenas qualquer informação, mas aquela, de fato, significativa, isto é, informação que possa ser considerada útil para o pesquisador e para a sociedade em geral.

Assim, dados tomados a cada hora, todos os dias da semana, durante todas as semanas do ano, podem ser armazenados em um único vetor. Isto é, pode ser construído um vetor de observações contendo várias variáveis. E esse vetor pode, como é o caso, não revelar qualquer informação.

Vale lembrar aqui a pergunta de Paul Geladi (1989) “... os métodos univariados realmente existem?” Ou formular de outra maneira: os problemas univariados realmente existem?

Assim, se os dados coletados durante todo o ano puderem ser vistos como dados gerados a cada hora, ao longo das 24 horas de cada dia, nos 7 dias da semana, durante as 52 semanas do ano, perceber-se-á um caminho sendo construído naturalmente em busca de variações sistemáticas presentes nos dados em cada um dos períodos (diário e semanal) citados. Configura-se assim, a necessidade de utilização de estruturas multimodo mesmo tratando-se de apenas um poluente e apenas uma estação. Isto significa transformar, inicialmente, o vetor de dados em uma matriz e, posteriormente, em um “paralelepípedo” de dados.

Independentemente da existência de vários compostos na atmosfera e de reações ocorrendo entre eles, se uma determinada substância possuir um perfil de comportamento (curável), este poderá ser revelado pela aplicação adequada de algum método de análise. Pelas características das principais possíveis fontes emissoras, é, de fato, razoável supor que deve existir um perfil diário e um semanal. Organizar corretamente os dados, é fundamental para que se possa extrair tais informações, inferindo sobre que tipo de decomposição poderia revelar as variações presentes nos mesmos.

O problema pode ser formulado matematicamente de várias maneiras. Uma delas (inspirado em Kroonenberg, 1989, p.16-20) é começar a analisar o problema pelo final, isto é, imaginar uma situação ideal, passar para uma aparente (fornecida pelo instrumento de medição, finalizando com a situação real (a que existe, de fato). Tenta-se entender a realidade a partir de situações ideais. Matematicamente significa assumir que a situação real é constituída de combinações lineares de situações ideais supostamente (ou não em muitos casos) conhecidas.

A situação ideal, seria aquela na qual não existe nenhum tipo de erro instrumental (portanto, no ato de medir), nem qualquer influência externa, como por exemplo, condições nas quais se realiza a medição. A situação ideal, então, não obstante imaginável, jamais existirá. A situação aparente é aquela na qual a medição considera o erro instrumental, porém não consegue controlar os fatores externos (condições). A medição é dependente das condições em que é realizada (por exemplo, condições climáticas, como em dados ambientais). Finalmente, a situação real é aquela na qual se leva em conta todos os fatores (instrumental e condições). Esta é a situação que os dados levantados representam. Infelizmente, estes são de difícil interpretação, não correspondem à realidade conhecida *a priori*, ou não possuem significado físico. Em suma, não se constituem como informação significativa.

Inversamente ao artifício representado pelo caminho sugerido acima para chegar, finalmente, à matriz de dados, na verdade, é ela, a própria matriz de dados, $\mathbf{X}$, o ponto de partida da análise. A decomposição deverá levar em conta alguns pressupostos, restrições etc, tal que será possível obter uma aproximação da matriz original, $\mathbf{X}$, que evidencie certas características, propriedades ou padrões, que não sejam perceptíveis no conjunto de dados originais.

Buscou-se na representação acima a aplicação do segundo pressuposto (o ataque do problema a partir das emissões, citado na p125). Buscar-se-á em seguida usar um raciocínio semelhante para inferir sobre que tipo decomposição poderia revelar a variação presente nos dados. A suposição aqui é a de que os dados medidos podem ser expressos como uma combinação linear de situações que devem representar de modo mais significativo o que eles, de fato, representam. Em suma, aplicar uma transformação (de eixos) na matriz de dados originais, de modo a evidenciar o padrão subjacente a eles.

Um dia com condições climáticas favoráveis à dispersão do poluente deve dificultar a sua medição, quando comparado a um dia com condições climáticas menos favoráveis à dispersão. Assim, um modelo multiplicativo como PCA (isto é, $\mathbf{X}=\mathbf{TP}^T$ ou $\mathbf{X}=\mathbf{AB}^T$, onde $\mathbf{A}=\mathbf{T}$ e $\mathbf{B}^T=\mathbf{P}^T$) foi primeiramente usado para avaliar a possibilidade (o grau de dificuldade) de se medir o poluente em um dado dia durante o ano.

Supôs-se que o conjunto de dados é constituído de duas partes:

1ª) emissão (contribuição da fonte)

2ª) acumulação (reação e dispersão – inclui reações químicas e dispersão devido à condições climáticas, o que provoca um grau de dificuldade na medição).

A primeira parte é um peso proporcional à quantidade de gás emitido pela fonte a uma dada hora. Tal peso é colocado nas colunas da matriz **B**, a qual fica constituída de linhas (horas) × colunas (fontes) como se vê na figura 3.19.

A segunda parte é um peso proporcional à contribuição de um determinado dia. Esse peso é colocado nas colunas da matriz **A**, a qual fica constituída de linhas (dias) × colunas (fontes) como se vê na figura 3.19.

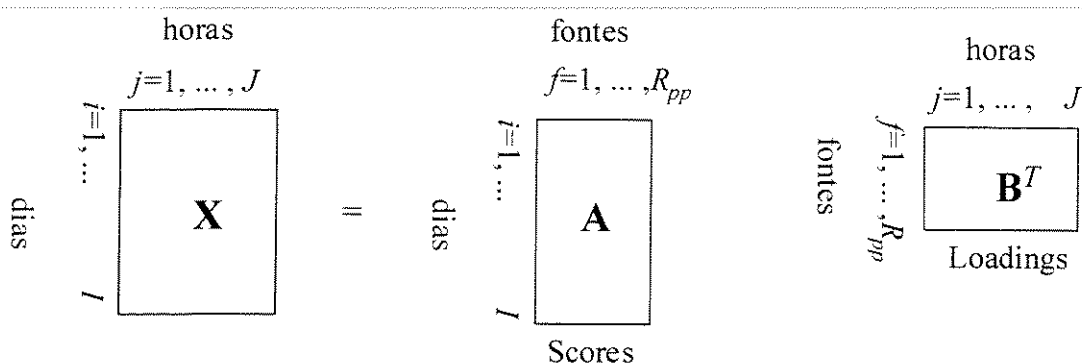


Fig.3.19: Decomposição da matriz de dados, **X**, nas matrizes **A** e **B^T**.

Para se extrair as variações sistemáticas diárias o modelo propõe que se arranje os dados em matrizes contendo um dos modos para as 24 horas do dia e o outro modo para os 365 dias do ano. A análise (PCA) dessas matrizes para cada um dos poluentes individualmente revela suas variações sistemáticas diárias ao longo de todo o ano. Este resultado é bastante representativo de tais variações devido a grande quantidade da dados, isto é, existem $24 \times 365 = 8760$ dados para cada poluente no ar. Obtém-se aqui um primeiro resultado: o perfil diário dos poluentes.

Retornando à tabela de dados, cada poluente é medido a cada hora durante 24 horas de cada dia, ao logo dos 365 dias do ano. Então tem-se um vetor de dados com 8760 pontos. Para construir a matriz de 365 dias × 24 horas, procedeu-se como na Figura 3.20. O vetor inteiro é composto de 365 partes (dias do ano) contendo cada uma 24 pontos (horas do dia) conforme mostra a Figura 3.20.

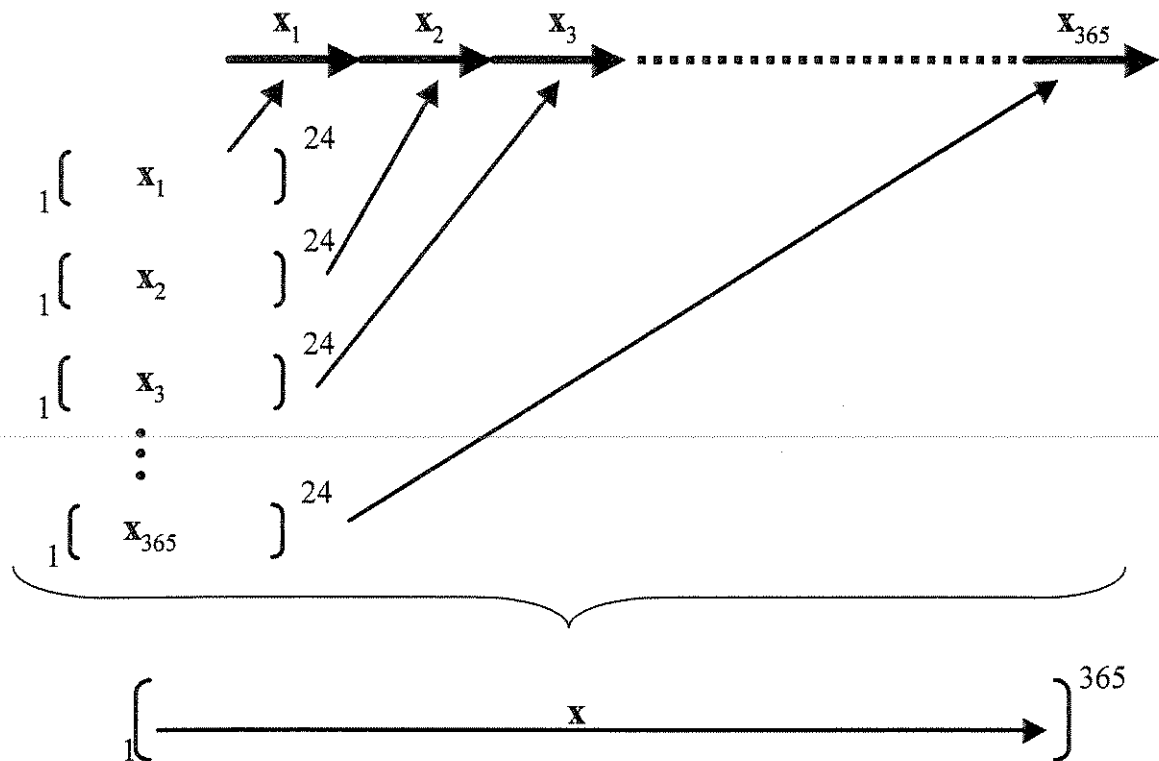


Fig.3.20: Esquema mostrando o vetor inteiro (1×8760) e suas 365 partes (1×24).

Cada vetor, $x_1, x_2, x_3, \dots, x_{365}$, de dimensões $x_i (1 \times 24)$, representa um dia do ano. Mantendo-se em mente as variações sistemáticas, pergunta-se: como se comporta o CO, por exemplo, a uma hora da manhã de cada dia, durante todo o ano? O esquema acima indica que a resposta está em um vetor que contém como entradas somente a 1ª hora de todos os dias do ano. O mesmo raciocínio vale para a 2ª, 3ª, ..., 24ª horas do dia. Isto nada mais é do que transformar o vetor com 8760 entradas em uma matriz de 365 linhas (dias) x 24 colunas (horas), conforme esquema abaixo (Figura 3.21).

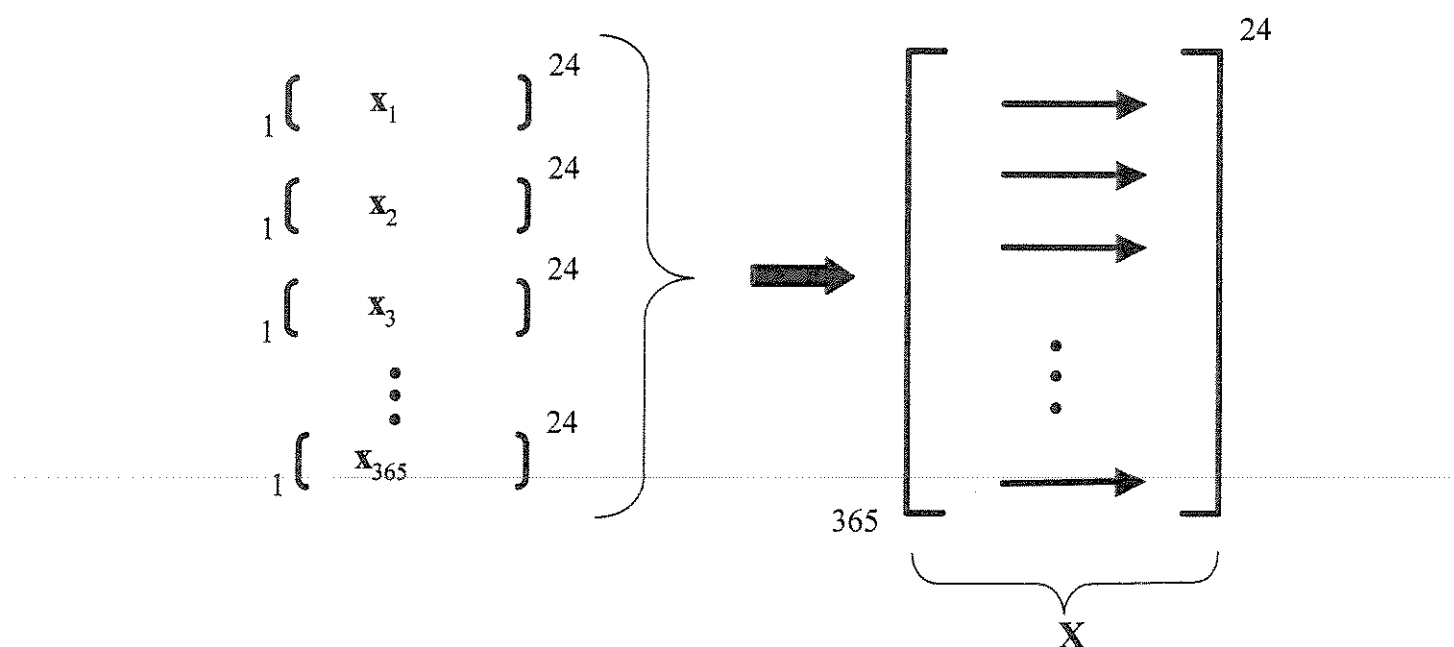


Fig.3.21: Construção da matriz de 365 (dias) $\times$ 24 (horas)

Aplicado aos dados reais, a versão acima para o monóxido de carbono, por exemplo, leva à Figura 3.22. Ela é o resultado prático da aplicação do procedimento descrito acima às concentrações medidas para o CO. A figura mostra também a decomposição da matriz X (365 dias $\times$ 24 horas) para o CO, em suas matrizes de escores (“scores”) e de pesos (“loadings”).

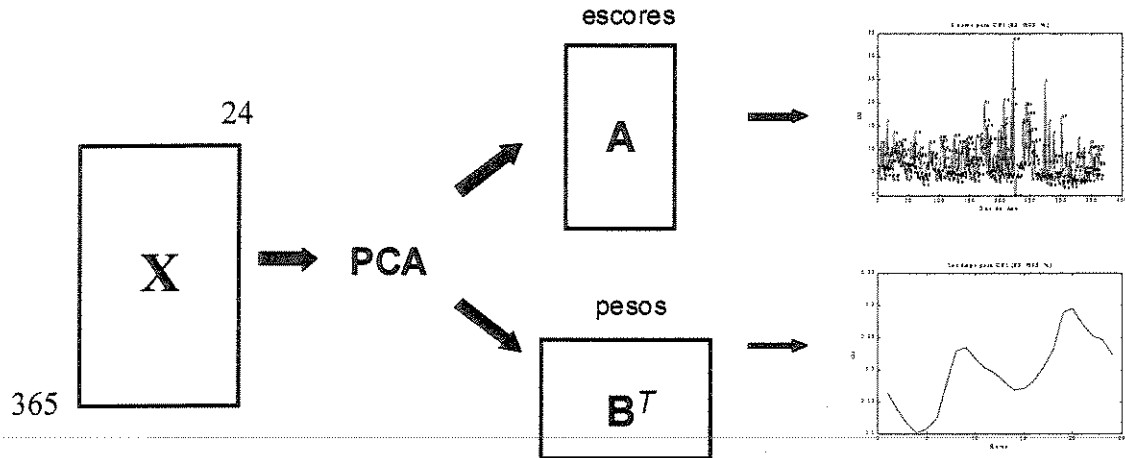


Fig.3.22: Construção da matriz 365 (dias) $\times$ 24 (horas) para o CO e sua decomposição em matrizes de escores e de pesos (loadings).

Um outro resultado útil é a avaliação dos perfis em diferentes épocas do ano. A tese nesse caso é a de que a relação entre os perfis pode se modificar como uma consequência de mudanças nas condições atmosféricas, porém os perfis em si mesmos não se modificarão, a menos que ocorra drástica modificação nas fontes emissoras (ou das condições meteorológicas). Isto está relacionado com a variação sazonal (períodos de seca e de chuvas) e com a variação semanal, sendo discutido em seguida.

Para a introdução dos modelos de ordem superior, é interessante continuar pensando na idéia das variações sistemáticas. Se um determinado composto exibe tais variações ao longo dos sete dias da semana, é razoável supor que isto se repetirá para as outras semanas ao longo do ano. Como discutido anteriormente, pelas características das principais possíveis fontes de poluição, é possível assumir a existência de um perfil diário e um semanal. Para extrair essa informação foi inicialmente usada a análise de componentes principais para se avaliar o grau de dificuldade de se medir a concentração diária do poluente através do ano. Aqui leva-se em consideração que as condições climáticas afetam as medições. Assumiu-se, como já mencionado, que os dados medidos recebem dois tipos de contribuição: a da fonte de emissão e a do grau de dificuldade de se realizar a medição.

A primeira é colocada nas colunas da matriz **B**(horas × fontes), a segunda é colocada nas colunas da matriz **A**(dias × fontes). Isto conduziu a um arranjo com um dos modos para os 365 dias do ano e outro para as 24 horas do dia, cuja análise (PCA) revelou as variações sistemáticas diárias durante o ano. Essa análise PCA corresponde a $K = 52$ matrizes de dimensões 24 horas × 7 dias como pode ser visto nas “fatias” de perfil da Figura 3.23. Dessa mesma figura, pode-se observar que as outras “fatias” têm “dimensões” 52 semanas × 24 horas (“fatias verticais”), e 7 dias × 52 semanas (“fatias horizontais”). Isto significa que a matriz **A** em PCA pode descrever dois tipos diferentes de contribuição: 1)a do efeito sazonal durante o ano (condições climáticas favoráveis ou não à dispersão dos poluentes, como nas “fatias” de perfil), isto é, (dias × fontes), e 2)a contribuição dos dias da semana (semanas × fontes). Entretanto, devido ao assumido inicialmente, isto é, **A**(horas × fontes) e **B**(dias × fontes), PCA não pode extrair essa informação, a não ser que os dados sejam adequadamente manuseados (ver discussão sobre essa afirmação no Apêndice A6).

Das considerações acima, para que se possa extrair o perfil semanal, sugere-se uma estrutura em três modos para cada poluente: 24 (horas) × 7 (dias) × 52 (semanas). Se essa estrutura de ordem superior para um dado poluente for estendida a um segundo poluente (ou mais), o raciocínio é o mesmo, porém a nova estrutura obtida possui quatro modos ao invés de três, isto é, $\mathbf{X} = \mathbf{AG} (\mathbf{D}^T \otimes \mathbf{C}^T \otimes \mathbf{B}^T) + \mathbf{E}$, onde a matriz **D** representa o modo que contém as contribuições dos poluentes.

A suposição de que as variações sistemáticas semanais apresentam o mesmo perfil a despeito das variações sazonais, devem ser confirmadas por meio de algum modelo.

Ainda mantendo-se em mente as variações sistemáticas pergunta-se: como o CO se comporta durante os sete dias da semana, ao longo de todo o ano? Tomando-se, as sete 1^{as} linhas da matriz da Figura 3.21, obtém-se uma matriz de 24 (horas) x 7 (dias). Essa matriz contém o comportamento na primeira semana. Tomando-se as sete próximas linhas (linhas 8 à 14) obtém-se uma segunda matriz de 24 (horas) x 7 (dias). Essa segunda matriz contém o comportamento para a segunda semana. Se o processo é realizado para as 52 semanas do ano, obtém-se a resposta. Tal processo significa organizar, paralelamente, 52 matrizes de 24 x 7 em um arranjo, em forma de um “cubo” ou paralelepípedo de dados como mostra a Figura 3.23.

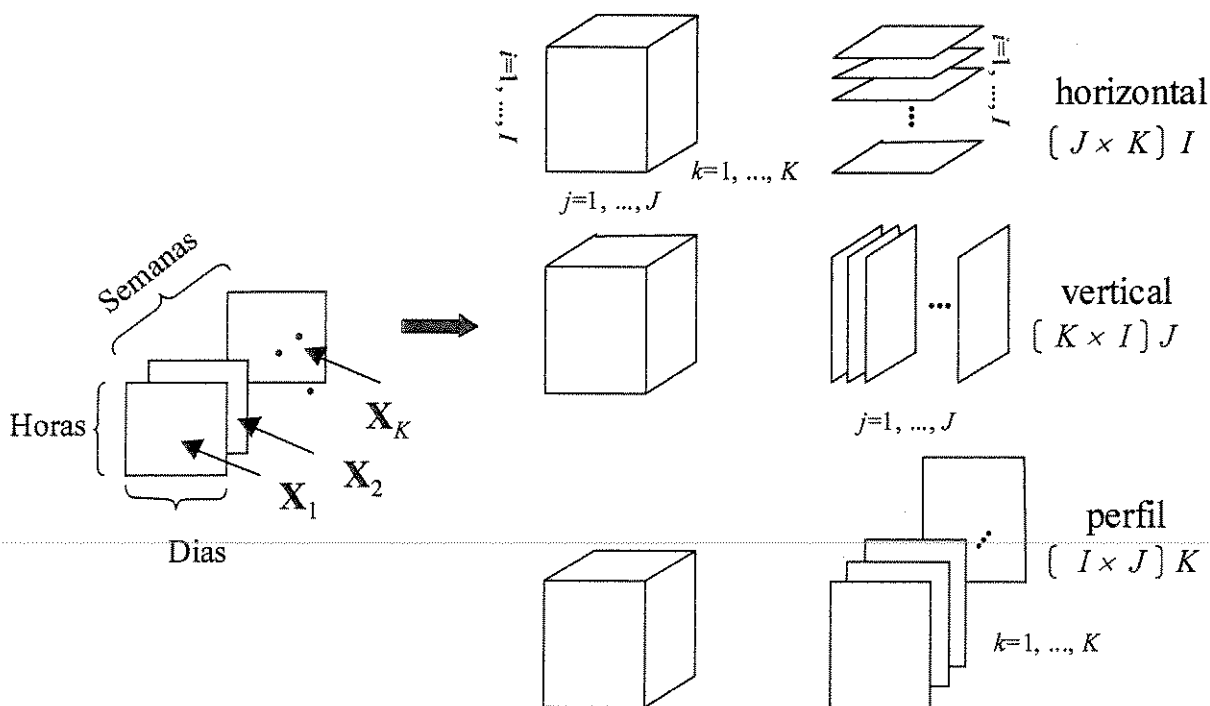


Fig.3.23: Construção da estrutura multimodo mostrando a disposição dos dias, horas e semanas nas matrizes.

- Analogamente, tem-se a versão de ordem superior (Figura 3.24) para o CO quando às suas concentrações aplica-se o procedimento acima. A figura mostra também a decomposição de $\underline{\mathbf{X}}$ nas matrizes de componentes $\mathbf{A}$ (perfil diário), $\mathbf{B}$ (perfil semanal) e $\mathbf{C}$ (variação anual).

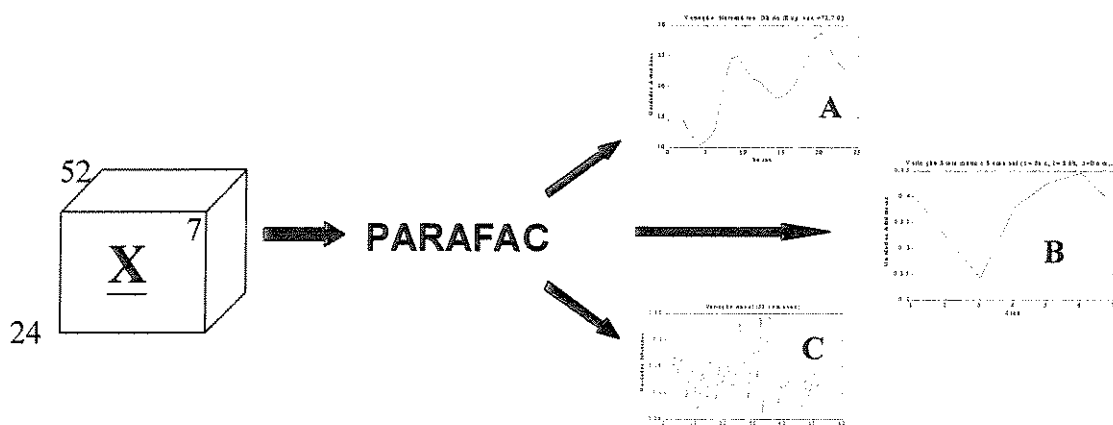


Fig.3.24: Arranjo multimodo para o CO (matriz $\underline{\mathbf{X}}$) e sua decomposição via modelo PARAFAC nas matrizes de componentes $\mathbf{A}$, $\mathbf{B}$ e $\mathbf{C}$.

As Figuras 3.23 e 3.24 mostram como o arranjo é construído, e o resultado de sua decomposição. A confirmação do perfil deve possuir também um fundamento matemático, o qual é discutido a seguir.

Quando se extrai do arranjo de 24 (horas) x 7 (dias) x 52 (semanas), apenas os dados da primeira semana, obtém-se uma matriz de 24 (horas) por 7 (dias), isto é, a matriz $\mathbf{X}_1$ na Figura 3.23. Se esta é decomposta (SVD) obtém-se como base do espaço, no qual são projetados os scores, um conjunto de autovetores. O autovetor que corresponde ao maior autovalor é a componente principal que captura a maior parte da variância dos dados.

Se uma outra matriz também de 24 (horas) por 7 (dias), referente à segunda semana for analisada de forma semelhante, obter-se-á também uma componente principal que captura a maior parte da variância dos dados para aquela semana. As soluções obtidas em cada caso sugerem que a orientação de tais eixos deve ser a mesma nos dois casos, ficando a diferença entre as duas soluções restrita apenas às proporções de quantidade de variância capturada de um conjunto para o outro. Denominando-se a matriz para a primeira semana de $\mathbf{X}_1$ e supondo-se que ela possa ser adequadamente modelada por $\mathbf{U}_1 \mathbf{\Lambda}_1 \mathbf{P}_1^T = \mathbf{A} \mathbf{B}^T$, de posto dois, tal modelo pode ser expresso por (Bro, 1998)

$$\mathbf{X}_1 = {}^1\mathbf{a}_1 {}^1\mathbf{b}_1^T c_{11} + {}^1\mathbf{a}_2 {}^1\mathbf{b}_2^T c_{12} \quad (3.98)$$

Se a matriz para a segunda semana, $\mathbf{X}_2$, puder ser descrita pelos mesmos conjuntos de vetores de escores e de pesos, diferindo apenas em suas proporções, tal decomposição pode ser expressa por

$$\mathbf{X}_2 = {}^2\mathbf{a}_1 {}^2\mathbf{b}_1^T c_{21} + {}^2\mathbf{a}_2 {}^2\mathbf{b}_2^T c_{22} \quad (3.99)$$

onde ${}^1\mathbf{a}_1$, ${}^1\mathbf{a}_2$, ${}^1\mathbf{b}_1^T$ e ${}^1\mathbf{b}_2^T$ são vetores singulares de $\mathbf{X}_1$, ${}^2\mathbf{a}_1$, ${}^2\mathbf{a}_2$, ${}^2\mathbf{b}_1^T$ e ${}^2\mathbf{b}_2^T$ são vetores singulares de $\mathbf{X}_2$, e c_{21} e c_{22} não são, em geral, iguais a c_{11} e c_{12} . Os dois modelos, os quais consistem dos mesmos perfis (paralelos) diferindo apenas em suas proporções, podem ser expressos como um modelo combinado fazendo (Bro, 1998)

$$\mathbf{X}_k = \mathbf{A} \mathbf{D}_k \mathbf{B}^T \quad (3.100)$$

onde $\mathbf{D}_k$ é uma matriz diagonal com os elementos da k-ésima linha, c_{kr} , de $\mathbf{C}$, em sua diagonal (ver equações 3.86 a 3.88).

Então, esses eixos, ${}^1\mathbf{b}_1^T$ com ${}^2\mathbf{b}_1^T$, e ${}^1\mathbf{b}_2^T$ com ${}^2\mathbf{b}_2^T$, devem ser paralelos nos seus respectivos espaços-solução, existindo, na verdade, um conjunto único de direções de eixos em ambos os espaços (raciocínio análogo vale para os vetores singulares " $\mathbf{a}$ "). Descobrimos essa tal posição única, pode-se determinar empiricamente a orientação aproximada que tais componentes possuíam quando os dados foram gerados (Harshman e Lundy, 1994). Isto constitui a base para o modelo PARAFAC (PARAlell FACtor analysis) tendo sido Cattell (1944) o primeiro a mostrar que a presença de perfis paralelos proporcionais pode levar a uma decomposição isenta de ambigüidades (Bro, 1998).

É importante destacar o fato de que a estrutura acima, construída em três modos, possui variações proporcionais nas contribuições de um dado fator do nível um para o próximo, em todos os três modos do arranjo, como se deve esperar, por exemplo, de emissões de fontes automotivas. Em outras palavras, espera-se existir um perfil diário de emissão diferindo apenas no "peso" de um dia para outro (comparando-se, por exemplo, as concentrações do poluente em uma quarta-feira com as concentrações do mesmo em um domingo, ou essas concentrações em semanas sucessivas) ao longo de cada semana do ano.

Para o caso de ajuste direto (Harshman e Lundy, 1994), isto está em total concordância com o princípio de "Variação do Sistema" ("System Variation"), cujas variações proporcionais devem se manter em pelo menos três dos modos de um arranjo multimodo, exigência a ser cumprida para que ao problema possa ser aplicado o modelo PARAFAC (Harshman e Lundy, 1994). De qualquer maneira, esse é um obstáculo facilmente transponível, bastando que se faça o ajuste indireto.

O modelo PARAFAC (assim como o Tucker) não apenas identifica a fonte, como também separa os poluentes primários dos secundários. Entretanto, o modelo não permite a interação entre os pesos (loadings) de índices diferentes nas matrizes de componentes (por exemplo, não é possível calcular em um único peso o efeito sazonal para os poluentes primários ou secundários, quando isto é necessário). Então, perde-se a possibilidade do cruzamento dessas informações. Em outras palavras, é desejável que se possa verificar a possibilidade de desmembramento das contribuições de diferentes fontes às concentrações

de um dado poluente, representadas, por exemplo, por perfis diferentes. Ou ainda, a contribuição de um poluente primário na formação de um secundário.

A aplicação do modelo de Tucker, como será visto adiante, oferece respostas interessantes ainda não obtidas até aqui. A análise por meio do modelo de Tucker será enfatizada devido à elegância da sua formulação e ao escopo de sua solução.

A construção metodológica pode ser resumida nas seguintes etapas abaixo:

- 1) Validação dos dados da tabela original fornecida pela CETESB,
- 2) Identificação dos vetores de dados para posterior construção das matrizes e preenchimento dos dados faltantes com conjuntos vazios,
- 3) Construção das matrizes de $365 \text{ dias} \times 24 \text{ horas}$,
- 4) Correções de valores negativos de concentração (considerados como dados faltantes e estimados pelo modelo),
- 5) Pretratamento dos dados (identificação de outliers, substituição destes por dados faltantes, aplicação de PCA para cálculo da média anual e, finalmente, substituição dos dados anômalos e das variações bruscas por média anual),
- 6) Correções de “lacunaridade” (ver definição adiante), e avaliação de “esparsidade”,
- 7) Aplicação de PCA às matrizes $\mathbf{X}(365 \times 24)$ para cada poluente individualmente, com a finalidade de extração do perfil diário dos mesmos.
- 8) Aplicação do modelo PARAFAC às estruturas multimodo $\underline{\mathbf{X}}(24 \times 7 \times 52)$ para cada poluente individualmente, visando a extração dos perfis diário e semanal, e da variação anual dos mesmos.
- 9) Aplicação do modelo PARAFAC aos arranjos multimodo $\underline{\mathbf{X}}(24 \times 7 \times 52 \times \text{NP})$ para os poluentes do ciclo do monóxido de carbono, CO, para os compostos do ciclo do metano, CH₄, para os compostos do ciclo dos hidrocarbonetos não metano, NMHC, e para os demais poluentes (NP = N° de Poluentes). São obtidos os perfis diários e semanais, e a variação anual dos poluentes.
- 10) Aplicação do modelo de Tucker aos arranjos multimodo $\underline{\mathbf{X}}(24 \times 7 \times 52 \times \text{NP})$ aos três ciclos para obtenção dos perfis diários e semanais, a variação anual, e os perfis de acumulação dos poluentes.

3.3.3 – Pretratamento de Dados II

Devido à grande quantidade de dados levantados e em função dos pressupostos firmados na metodologia, várias rotinas (“scripts”) foram construídas para colocar os dados em condições de análise. Isto é, um vetor com a totalidade dos dados possuiria 8760 dados. Porém, devido à “existência de dados faltantes”, os vetores para cada variável possuíam um número inferior de pontos. Os “scripts” cuidaram de “preencher” as lacunas com conjuntos vazios, os quais foram estimados como dados faltantes. Desse modo foi possível trabalhar sempre com 8760 dados e, por conseguinte, com matrizes de 365 dias $\times$ 24 horas. O mesmo vale para as estruturas multimodo de 24 horas $\times$ 7 dias $\times$ 52 semanas. Tal providência permitiu, sobretudo, identificar a qualquer momento, a uma dada hora de um determinado dia, uma concentração específica para um poluente em uma dada estação.

Para efeito da discussão que se fará ao longo deste e dos demais capítulos, do ponto de vista da terminologia, são usadas as seguintes denominações:

1)*Dados originais* – organização dos dados originais contidos nas tabelas (CETESB) em um arranjo denominado matriz de dados originais.

2)*Dados corrigidos* – dados que sofreram correções para dados faltantes e dados anômalos.

3)*Dados suspeitos* – correções específicas em dados corrigidos, de modo a tornar possível comparações (ver análise sobre o SO₂).

4)*Concentração* – são os valores para os dados nas matrizes de dados originais, dados corrigidos, dados anômalos, variações súbitas, dados suspeitos e dados estimados, excetuando-se a temperatura (°C), e a umidade (%).

5)*Escores (Scores)* – o termo *escores (scores)* é usado para os dados obtidos após aplicação dos modelos. Como discutido, os valores dos *escores* após aplicação dos modelos são apenas proporcionais aos valores das concentrações dos dados originais. Esses últimos podem ser obtidos a partir dos primeiros por meio de adequada transformação de retorno. Para PCA, por exemplo, basta examinar a equação 3.51 (p82), na qual os pesos na matriz **P** (ou **B**, Fig.3.19, p114) , obtidos na decomposição, permitem a transformação dos dados originais (corrigidos) da matriz **X**, nos *escores* da matriz **T** (ou **A**, Fig.3.19, p114), e vice-versa.

3.3.3.1 – Os Dados

Os dados foram fornecidos em tabelas contendo informações sobre data, hora, concentração, estação e critérios de validação. Os critérios de validação são classificados como: 1) válido (valid), 2) validado (user valid), 3) não validado (user invalid), 4) incerto (uncertain), 5) não válido (not valid) e 6) não disponível (not available). Uma tabela contendo saída típica de dados fornecidos pela CETESB encontra-se no Apêndice A7.

3.3.3.1.1 – Validação dos Dados

Em geral, os valores de concentração segundo tais critérios não eram discrepantes. Para verificar os critérios, foram realizadas algumas análises PCA (CO, NO, NO₂ e O₃, para o ano de 1999) com cada poluente individualmente, incluindo todas as entradas disponíveis para o referido poluente. Em seguida, a cada vez (uma espécie de leave-one-out), o conjunto de dados dentro de uma mesma classe segundo os critérios de validação acima era retirado e substituído por dados faltantes, sendo novamente analisado por PCA. Esse resultado mostrou que os perfis encontrados eram muito semelhantes tal que, para efeito de análise exploratória, a classificação acima não mostrou interferência nos resultados.

O procedimento adotado nesse caso foi manual, isto é, um conjunto de dados dentro de uma mesma classe, segundo o critério de validação da CETESB era “deletado”, sendo o novo vetor, contendo maior quantidade de dados faltantes, submetido à rotina geradora da matriz. Essa nova matriz era submetida a PCA e o resultado comparado com a análise PCA da matriz contendo todas as classes. Em seguida, os dados da classe inicialmente retirada eram repostos e a classe seguinte era retirada para nova análise PCA e posterior comparação com o conjunto completo. O procedimento foi realizado do sexto critério para o primeiro de modo que ficava preservada a maior quantidade de dados válidos segundo a hierarquia dos critérios. Isto foi realizado até que todas as classes fossem testadas. As classes 5 e 6 não foram usadas para validação das outras quatro, sendo apenas substituídas como dados faltantes cuja estimativa foi usada para validação das 4 primeiras.

O procedimento descrito acima foi aplicado individualmente para os poluentes CO, NO, NO₂ e O₃ para o ano de 1999. Como os resultados (perfis, dados estimados e resíduos) sempre foram muito semelhantes, considerou-se que os dados fornecidos pela CETESB estavam validados para as etapas posteriores da análise.

3.3.3.1.2 – Dados Anômalos (“Outliers”)

Dados anômalos foram identificados por inspeção visual do gráfico da matriz de dados originais, para cada poluente individualmente. Foram considerados dados faltantes quando seus valores eram superiores a dez vezes o valor da média anual para a variável.

Tomou-se a média aritmética porque é cientificamente mais significativa (pelo menos para dados de concentrações), e mais adequada do ponto de vista de saúde pública (Parkurst, 1998). Em segundo lugar, pelo critério de tornar mínima a tendenciosidade (bias), a média aritmética é sempre preferível à geométrica, a qual, em geral leva a resultados tendenciosamente baixos (Parkurst, 1998). Em terceiro lugar, devido à validade do teorema do limite central (Mardia, 1979, p51-52; Morrison, 1990, p8-9, 85) aos dados. Ademais, o perfil de emissão é independente da forma da distribuição, o que é suficiente para fins exploratórios. Finalmente, a CETESB também usa médias aritméticas, exceto para *partículas totais em suspensão* onde é usada a *média geométrica anual*.

Com relação ao valor de corte para dados anômalos deve ser acrescentado que:

1) as concentrações médias anuais das médias máximas sempre foram, historicamente (1981-1999), inferiores aos padrões nacionais de qualidade do ar (ver relatório da CETESB, 1999),

2) no período 1997-1999 (CETESB, 1999), tomando-se as concentrações médias das máximas (média das 8 horas para o CO), o padrão não foi ultrapassado nos locais Pedro II, Cerqueira César e Centro. Tomando-se as concentrações médias das máximas (média de 1 hora para o O₃), o padrão não foi ultrapassado nos locais Pedro II, São Caetano, Ibirapuera e Lapa. Por fim, tomando-se as concentrações médias (média anual para o NO₂), o padrão não foi ultrapassado nos locais Pedro II, São Caetano, Ibirapuera, Cerqueira César, Centro, Lapa Osasco. Isto, mesmo levando-se em conta, para o caso do ozônio, o aumento de suas concentrações médias anuais das máximas em certas localidades após 1990,

3) o valor igual a dez vezes a média anual, para um dado poluente, encontrado aqui sempre foi superior ao padrão (primário ou secundário) adotado pela CETESB,

Assim, considerando-se os itens acima e os valores para as médias anuais, tomou-se aquele valor de corte como sendo razoável para o que se propõe investigar. Por outro lado, quando um pico era constituído por dados que apresentavam uma certa tendência, não se aplicou estritamente esse valor de corte para não causar distorções nos resultados.

A média anual (ver expressão adiante no tópico 3.3.3.1.4, p142), foi calculada tomando-se os valores de todas as horas registradas durante o ano (média sobre 8760 dados). Tomou-se a matriz contendo os dados originais, aplicou-se PCA (rotina mdpca) para se obter uma estimativa para os dados faltantes, os quais foram substituídos por sua estimativa. A média anual foi calculada sobre essa nova matriz. Aplicou-se o ponto de corte para dados anômalos, os quais foram substituídos pela média anual. O gráfico dessa matriz foi inspecionado visualmente quanto à variações bruscas. O ponto de corte para essas variações foi ajustado pela obtenção do melhor compromisso entre a distribuição da quantidade de variância capturada pelas três primeiras componentes principais, comparação visual da matriz contendo os dados estimados com a matriz de dados originais, e o erro envolvido na análise (considerando-se três PCs).

As Figuras 3.25 a 3.30 mostram o resultados obtidos com o procedimento adotado. Os dados são para o monóxido de carbono, CO, na Estação de Medição Ibirapuera (Ibir) para o ano de 1999.

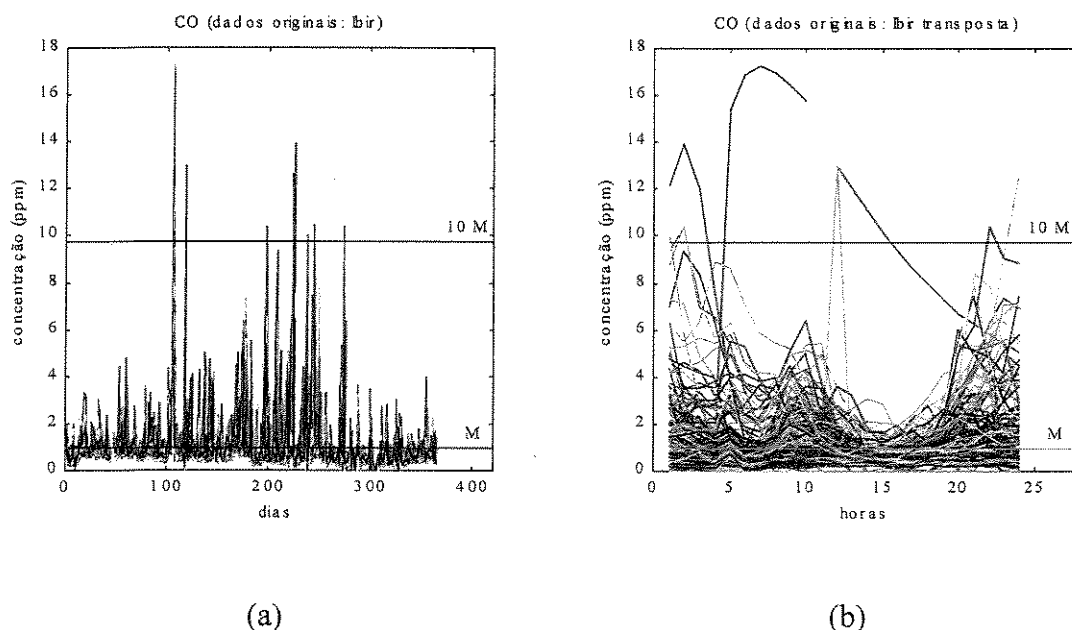


Fig.3.25: Dados anômalos ("outliers"). M = média e 10 M = valor de corte para "outliers", para o poluente monóxido de carbono, CO, na estação Ibirapuera no ano de 1999. a)Gráfico da concentração $\times$ dia, b)Gráfico da concentração $\times$ hora.

A Figura 3.25a mostra os dados originais para o monóxido de carbono na localidade Ibirapuera para o ano de 1999, enquanto a Figura 3.25b mostra o gráfico para a matriz transposta desses mesmos dados. Notam-se as linhas assinalando o valor médio, M , e o valor para o ponto de corte, $10 M$.

Após o corte e a substituição dos dados anômalos obtém-se os gráficos apresentados na Figura 3.26.

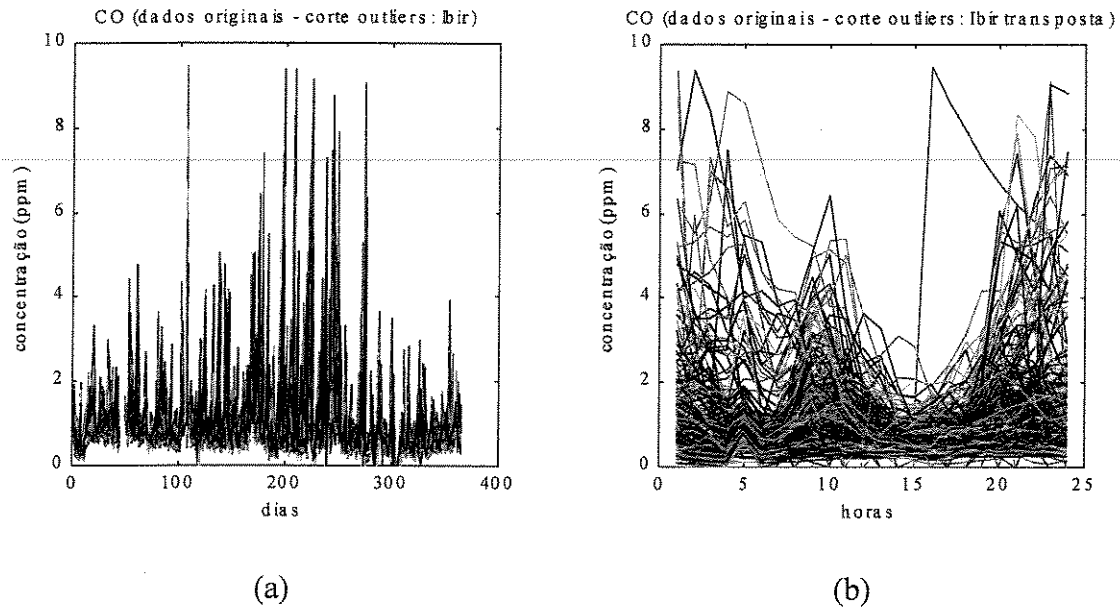


Fig.3.26: a) Gráfico concentração $\times$ dia sem os dados anômalos (“outliers”) para o CO no Ibirapuera (1999), e b) concentração $\times$ hora. (M = média).

A nova matriz (acima) é examinada quanto à existência de variações súbitas. De fato, uma inspeção na Figura 3.26b mostra que valores acima de aproximadamente 4.0 escapam do conjunto que representa as variações sistemáticas dos dados. Dados acima desse valor podem ser considerados como variações súbitas. Tais dados são substituídos pela média anual obtendo-se, finalmente, a matriz que é submetida à análise PCA para obtenção das variações sistemáticas diárias. Na seqüência, a Figura 3.27 mostra essa matriz para o monóxido de carbono no Ibirapuera em 1999. O estabelecimento desse ponto de corte leva em consideração a distribuição da quantidade de variância capturada pelas três primeiras componentes principais (Figura 3.28), a matriz contendo os dados estimados (posto completo, Figura 3.29) e o erro envolvido na análise (Figura 3.30).

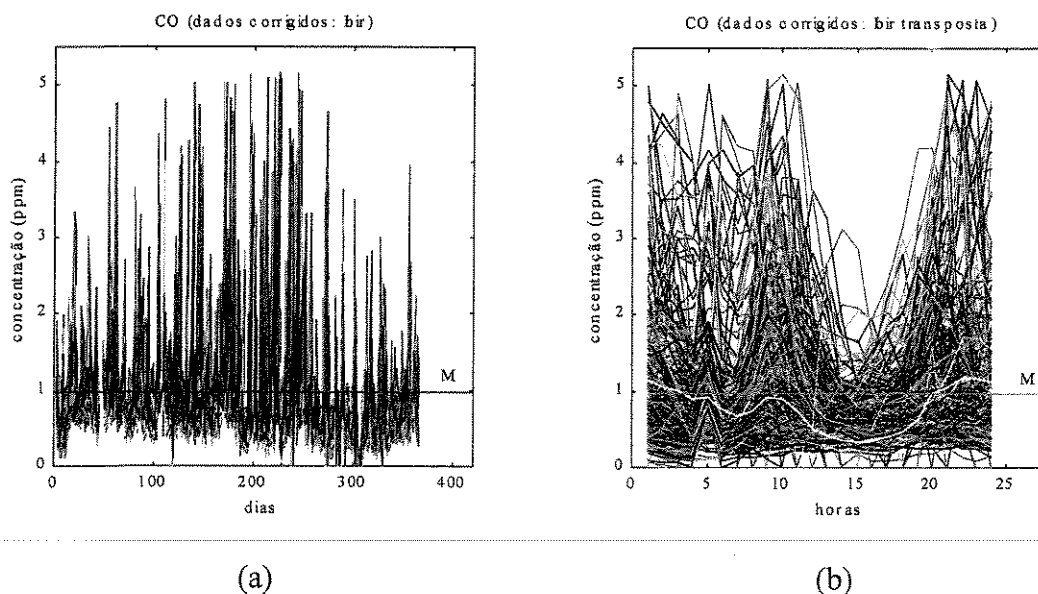


Fig.3.27: a) Gráfico concentração $\times$ dia sem os dados anômalos (“outliers”) e sem variações súbitas para o CO no Ibirapuera (1999), e b) concentração $\times$ hora. (M = média). A curva na cor branca representa a variação do desvio padrão ao longo das 24 horas.

Uma comparação entre as Figuras 3.25, 3.26 e 3.27 mostra que a informação significativa (variação sistemática diária) está muito bem representada na Figura 3.27.

Now working on iteration number 35		
Now forming final PCA model		
Percent Variance Captured		
by PCA Model		
Based on Known Data Only		
PC#	%Var	%TotVar
1.0000	79.8743	79.8743
2.0000	10.0318	89.9061
3.0000	3.2821	93.1882
»		

Fig.3.28: Saída de dados para a análise PCA com a variância capturada pelas três primeiras PCs (correção dados anômalos; em seguida, são corrigidos para variações súbitas).

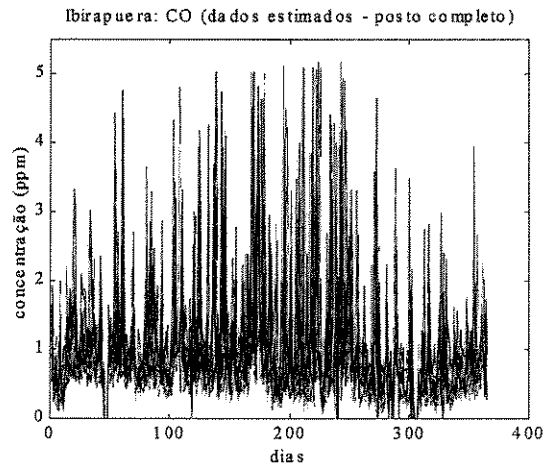


Fig.3.29: Dados estimados por meio da análise PCA (comparar com a Fig.3.27a).

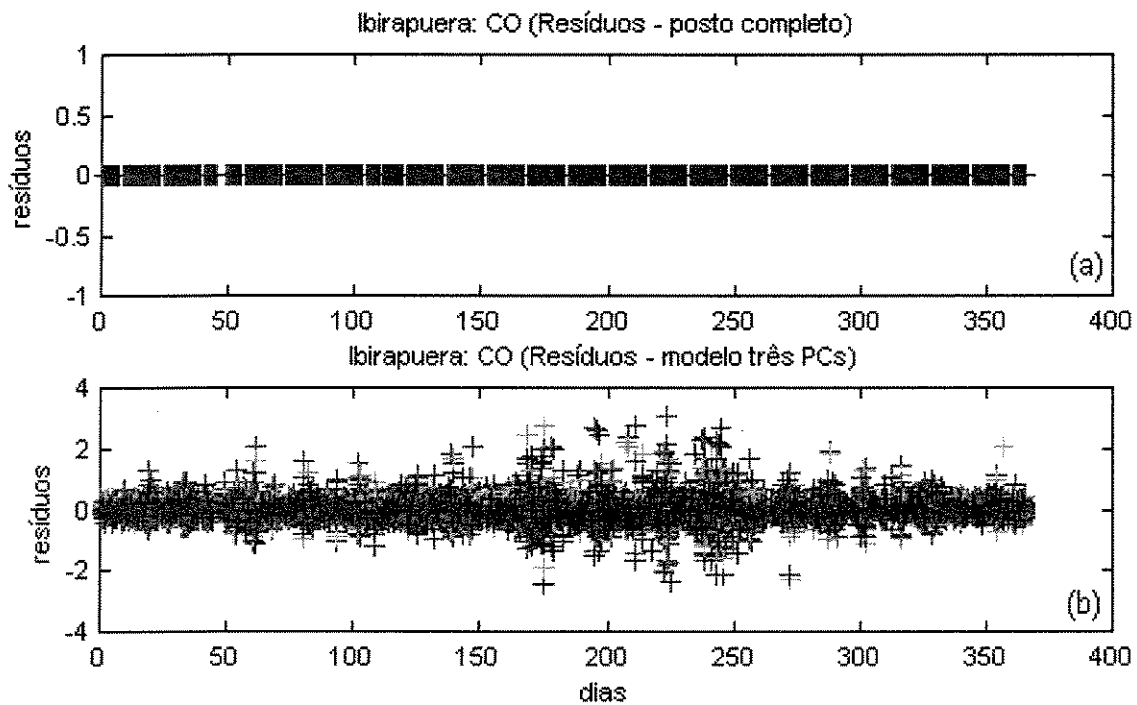


Fig.3.30: Resíduos deixados pelo modelo considerando-se posto completo (a), e as três primeiras componentes principais (b).

A Figura 3.30a mostra, como esperado, a ausência de erro quando se leva em conta todas as PCs e a Figura 3.30b uma distribuição aleatória para o erro tomando-se três PCs.

3.3.3.1.3 – Dados Faltantes

Na maioria dos casos, os dados faltantes não se constituíram como um problema em função de sua quantidade, mas sim em função de sua distribuição.

Em certos casos matrizes esparsas poderiam causar alguma influência nos resultados. Isto ocorria quando valores iguais a zero surgiam de forma súbita, isoladamente ou em um dado período, seguido ou precedido de dados faltantes e/ou dados anômalos. A substituição dos dados faltantes e anômalos pela média anual resolveu o problema.

No caso de dados faltantes, o problema passava a existir quando a seqüência era relativamente extensa, e o período contendo esses dados era precedido e/ou seguido de alguma tendência. Especificamente para esse caso de dados faltantes, tal ocorrência foi denominada “lacunaridade” (Figura 3.31).

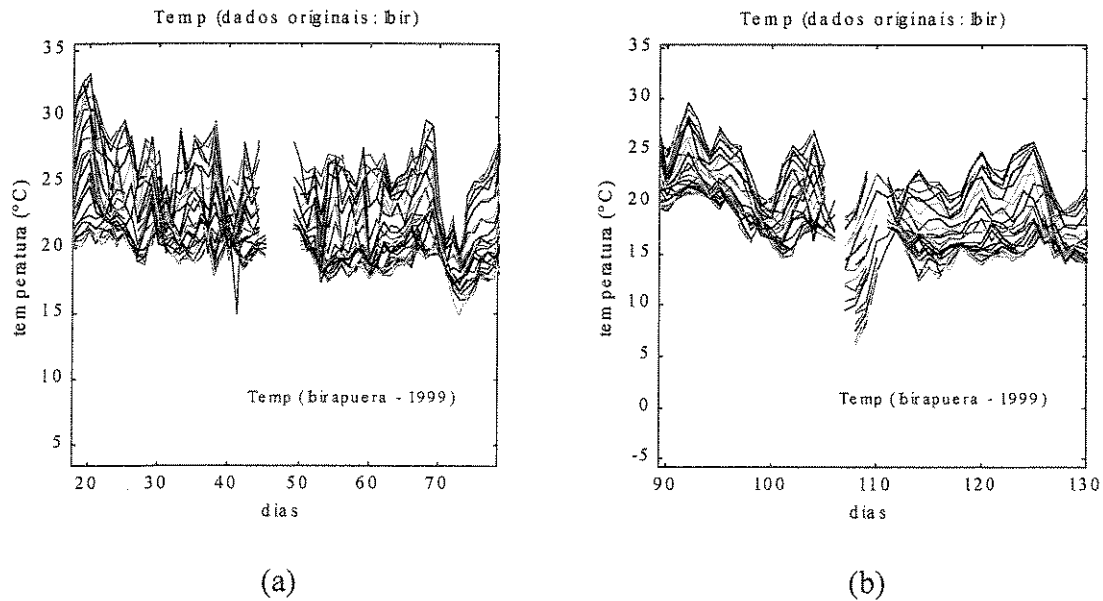


Fig.3.31: Ocorrência de “lacunaridade” (Temp., Ibirapuera – 1999)

Quando ocorria o problema acima, a rotina “mdpca” usada realizava a estimativa conservando a tendência encontrada nos dados existentes, os quais sofriam abrupta interrupção. Assim a rotina substituía os dados faltantes pela estimativa conservando uma tendência que conduzia a valores muito altos ou muito baixos (negativos) no período estimado.

A Figura 3.31a mostra um intervalo com ausência de dados. Nesse caso, como a extremidade esquerda do intervalo não apresenta uma tendência específica e a extremidade direita apresenta leve tendência para cima, o modelo, para compensar estima para baixo (ver Figuras 3.32a, 3.33a e 3.36a) os dados no intervalo. Na Figura 3.31b o problema é mais grave. Os dois extremos do intervalo apresentam a mesma tendência, porém o lado à direita do intervalo apresenta maior queda nos dados originais (variação de mais ou menos 18 °C a aproximadamente 7 °C). A rotina mdpca interpreta esses dados como uma tendência constante de queda e estima valores muito baixos no intervalo.

No primeiro caso, a influência não é forte o suficiente para causar resíduos elevados, porém o é para causar grandes distorções nos valores estimados, os quais escapam do padrão exibido pelo conjunto dos dados.

No segundo caso, mesmo para um intervalo menor de dados faltantes, a tendência seguida pelos dados, pré e pós lacuna, causa influência tão grande no algoritmo que estima os valores no intervalo com ausência de dados, que os mesmos atingem valores negativos inaceitáveis para as condições climáticas na região estudada, ou mesmo em qualquer região do Brasil. Sendo esse comportamento considerado irreal, correções devem ser realizadas nos dados de modo a torná-los mais consistentes com a realidade da região sob investigação.

Não obstante a análise dos resíduos com todas as PCs não indique a anomalia (Figura 3.32) eis que a mesma é absorvida pelo algoritmo da rotina, quando se utiliza um número reduzido de PCs, fica patente a influência causada pelo problema como mostra a Figura 3.33 para uma componente principal (Temp., Ibirapuera – 1999).

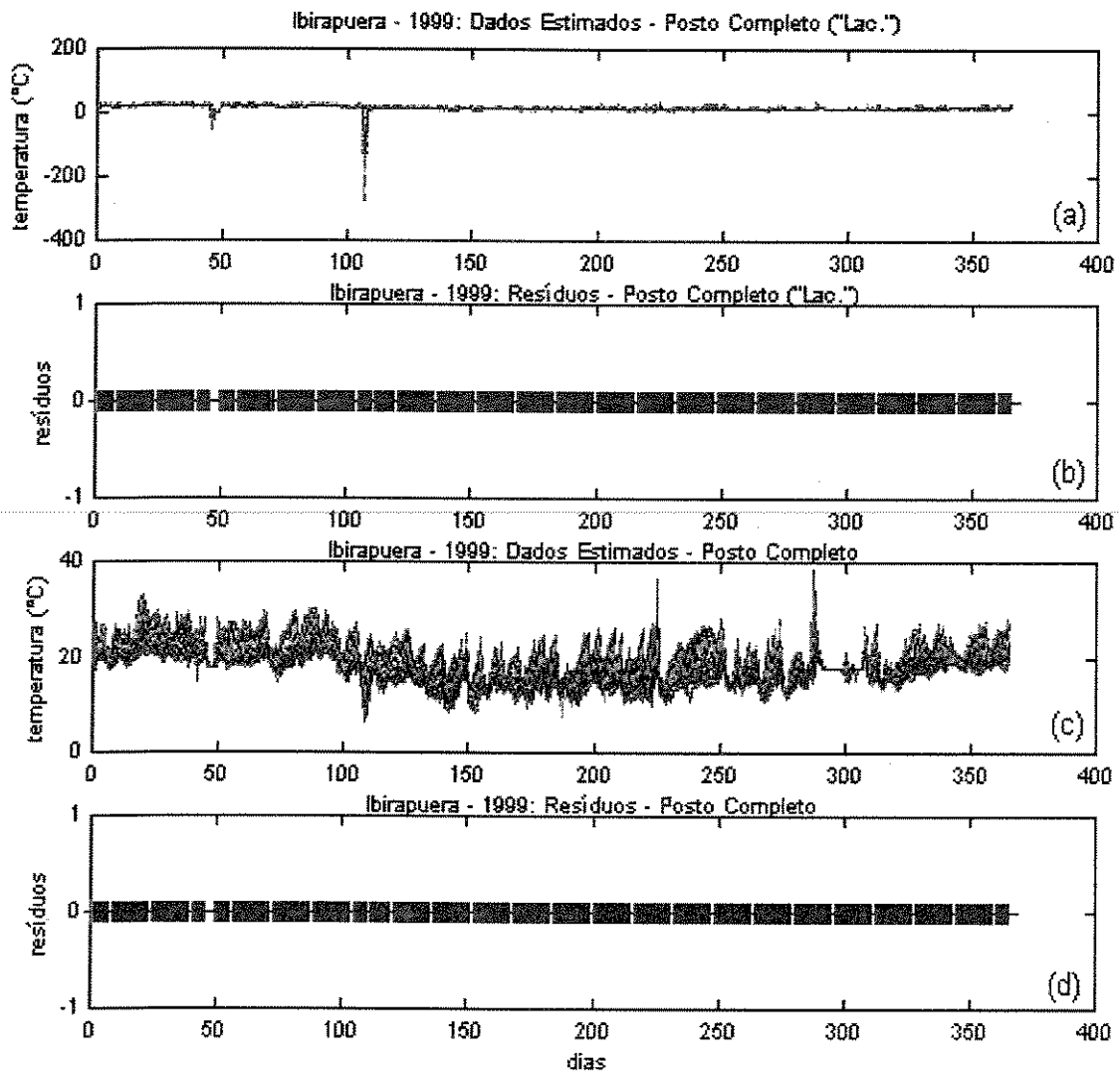


Fig.3.32: Dados estimados (a) e resíduos (b) considerando-se todas as PCs e a influência do problema lacunar. Dados estimados (c) e resíduos (d) para todas as PCs, corrigido o problema (Temp., Ibirapuera – 1999).

A Figura 3.32a mostra que os valores para os dados estimados considerando-se posto completo podem chegar a cerca de -30°C na primeira lacuna e -300°C na outra. A análise pura e simples dos resíduos (Figura 3.32b) não permitiria detectar a anomalia. Quando se examina a Figura 3.32c, pode-se observar que a correção aplicada leva a um resultado aceitável sob o ponto de vista da realidade local. Do mesmo modo, a Figura 3.32d mostra que os resíduos são iguais ao caso sem a correção.

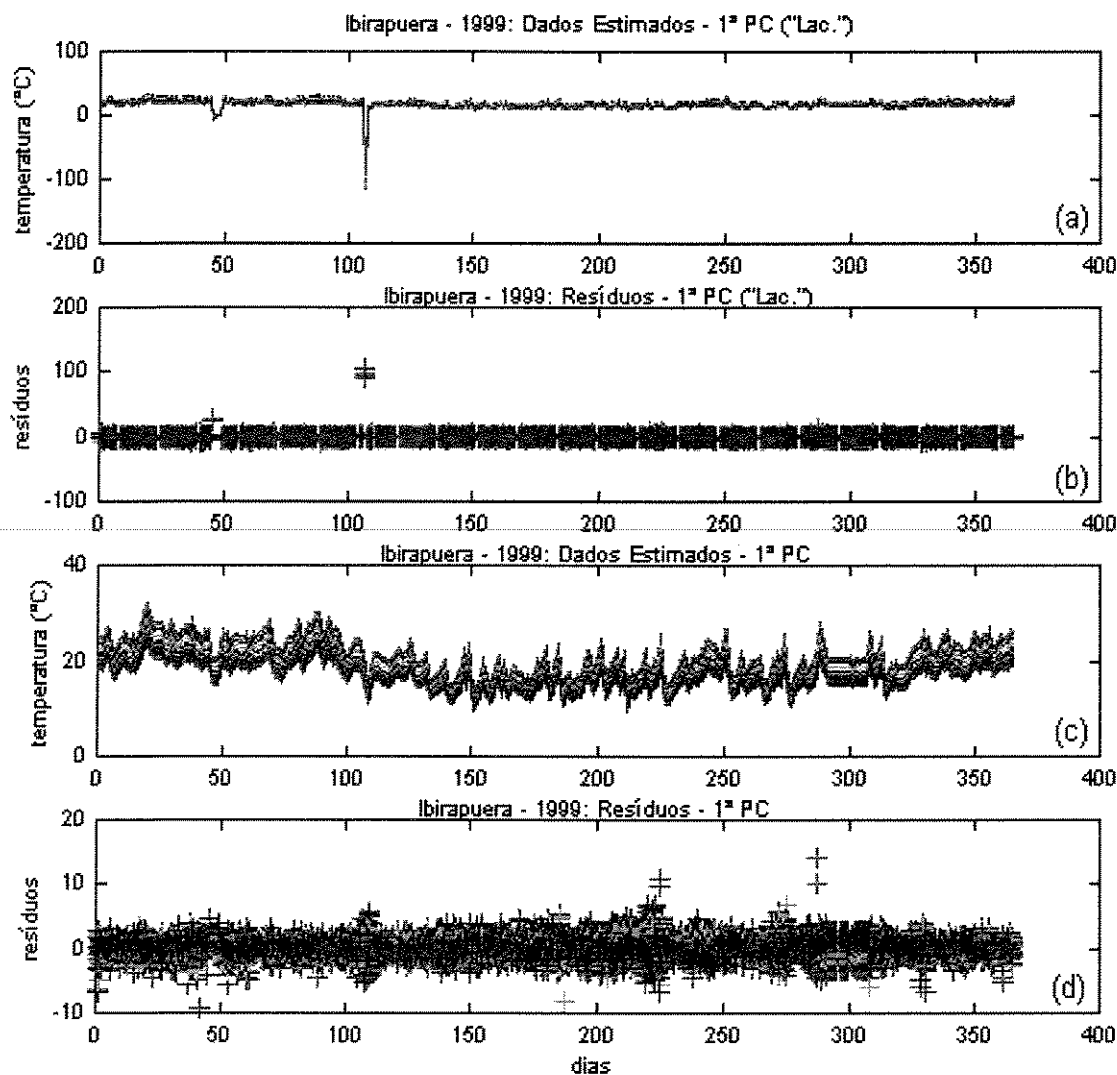


Fig.3.33: Dados estimados (a) e resíduos (b) considerando-se a primeira PC e a influência do problema lacunar. Dados estimados (c) e resíduos (d) para a primeira PC, corrigido o problema (Temp., Ibirapuera – 1999).

A Figura 3.33a, considerando-se apenas a primeira PC, mostra resultado semelhante ao da Figura 3.32a com todas as PCs. A Figura 3.33c (Com a 1ª PC) mostra um padrão semelhante ao da Figura 3.32c. Entretanto, quando se examinam os resíduos para uma PC (Figuras 3.33b e d) nota-se a influência causada pela problema de “lacunaridade”. No primeiro caso (Figura 3.33b) os resíduos, especialmente no segundo período lacunar

atingem valores bastante elevados, enquanto no segundo caso (Figura 3.33d) os valores para os resíduos encontram-se em intervalos aceitáveis. Observa-se ainda, por meio dessa última figura, que a correção pode ser mais refinada ainda por volta dos 225° e 290° dias, os quais ainda apresentam resíduos um pouco altos.

Se as correções sobre a ‘lacunaridade’ não forem realizadas os perfis sofrerão distorções. Por exemplo, os valores de pico para a temperatura em cada caso, sob e sem os efeitos do problema de “lacunaridade”, encontram-se em horários diferentes. Além disso, quando são calculados os escores para as componentes principais, ocorrem valores discrepantes em relação àqueles que se obtém quando a anomalia é corrigida. Uma segunda constatação é aquela encontrada nos resíduos como já discutida anteriormente. Um outro ponto a ser verificado está relacionado com a influência dos dados faltantes (ver explanação na p139) como explicitado pelo gráfico correspondente a essa avaliação (Figura 3.37). Outro aspecto a ser considerado refere-se à análise com o modelo PARAFAC para a variável em questão. Quando não são postas em prática as correções, o problema passa a se manifestar nas variações sistemáticas semanais e, claro, na variação anual.

Um conjunto de figuras adiante destaca as diferenças para cada caso (com e sem “lacunaridade”). A primeira delas (Figura 3.34) mostra as distinções para as três primeiras componentes principais. Esse gráfico mostra que a influência pode se manifestar não apenas na primeira componente principal. Em seguida, são destacados os perfis para a primeira componente principal, em cada caso (Figura 3.35). Passo seguinte, são apresentados os escores para a primeira componente principal para os dois casos (Figura 3.36).

Para verificar o efeito provocado por conjunto de dados contendo dados faltantes agravados com problemas de “lacunaridade”, são mostradas figuras contendo essa influência em cada caso (Figura 3.37). Finalmente, são apresentadas figuras com a análise PARAFAC (Figuras 3.38 e 3.39, para a Temp – 1999 no Ibirapuera).

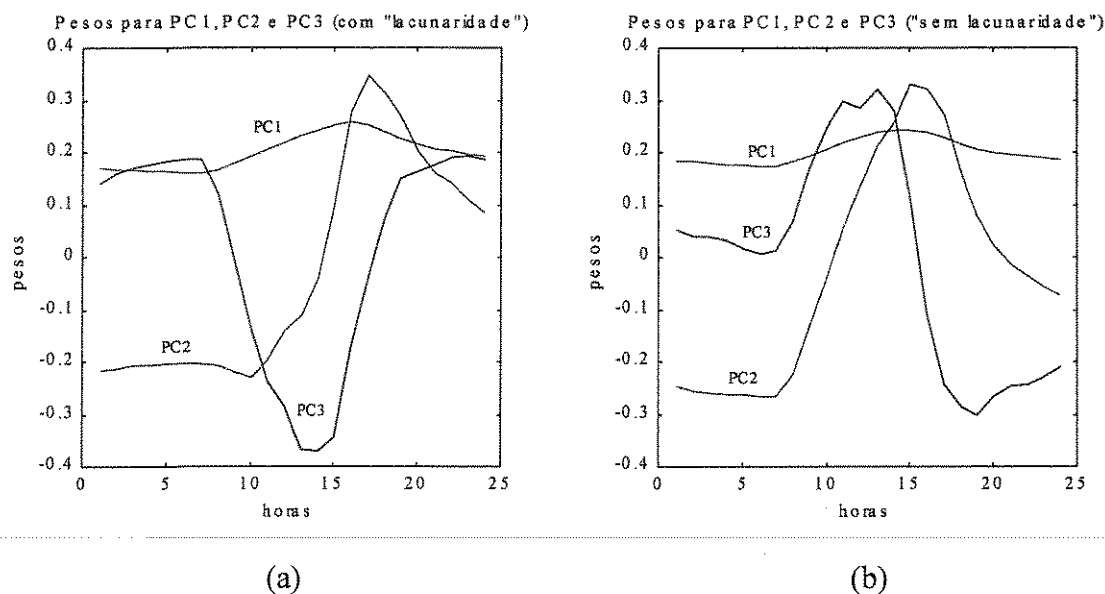


Fig.3.34: Pesos para PC1, PC2 e PC3 com (a) e sem (b) "lacunaridade".

Observa-se por meio da Figura 3.34 que as curvas para os pesos das três PCs apresentam diferenças. PC1, em (a), está mais deslocada para a direita em relação a (b), indicando que o pico para a temperatura ocorre em torno das 16:00 h. PC2 em (a) também está deslocada para a direita em relação a (b). Por fim, a curva para PC3 em (a) tem aspecto totalmente diferente daquele apresentado por PC3 em (b). O efeito para PC1 é destacado em figura à parte (Figura 3.35).

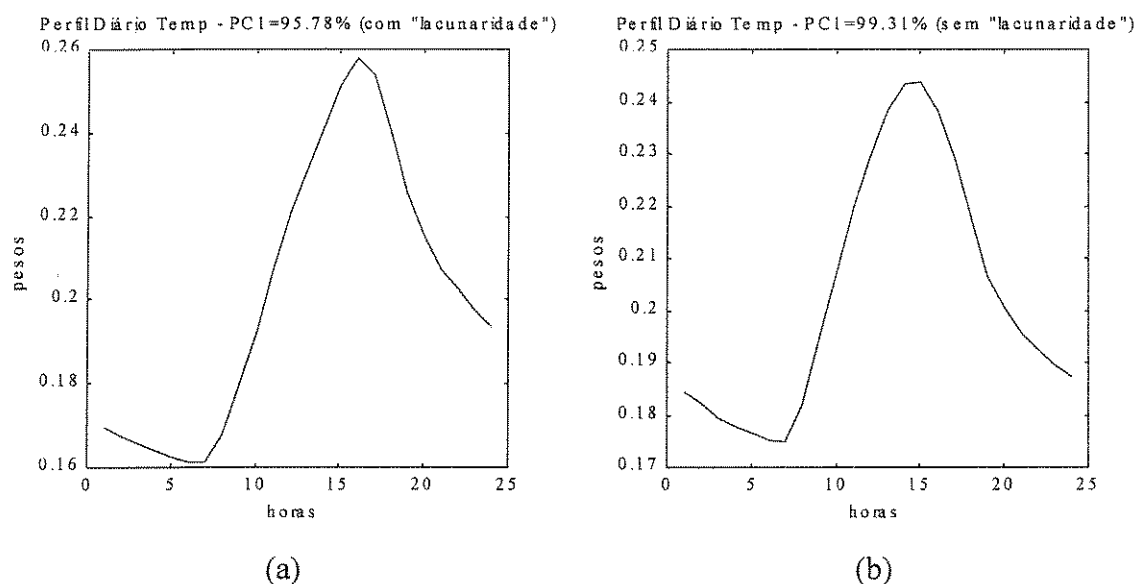
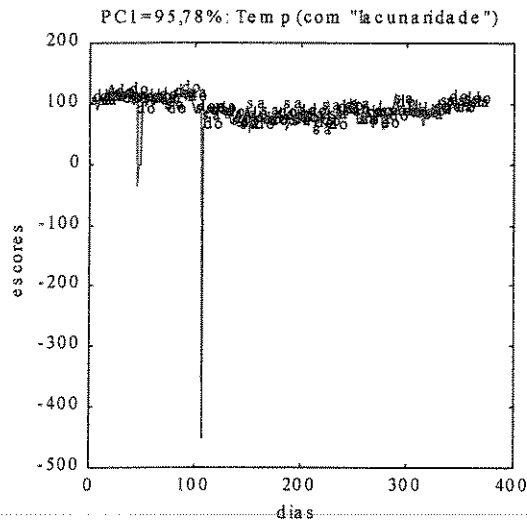
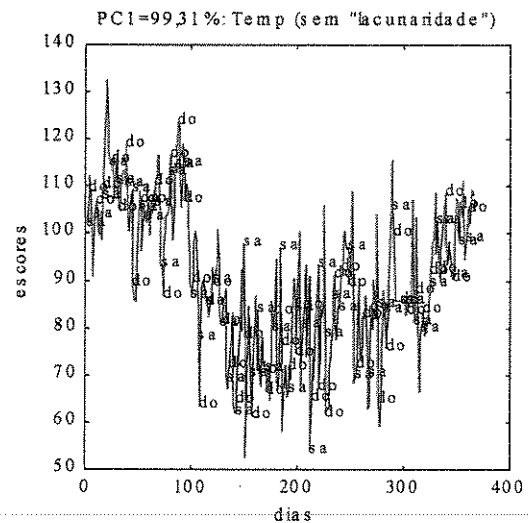


Fig.3.35: Pesos para PC1



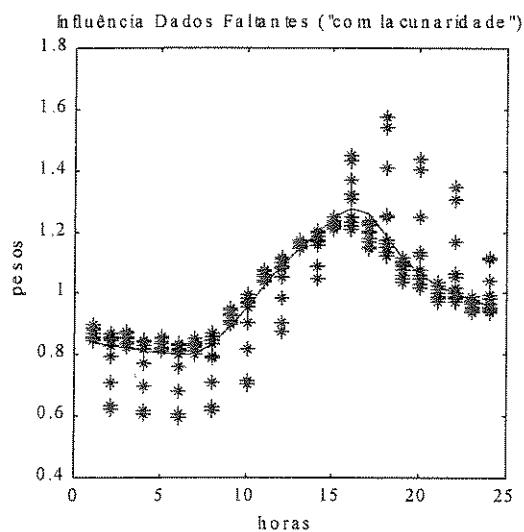
(a)



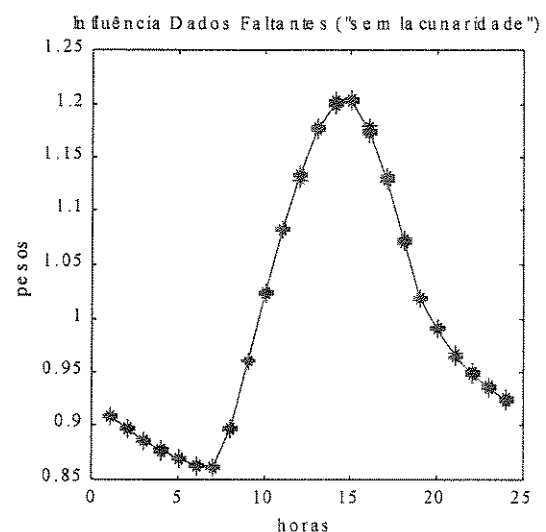
(b)

Fig.3.36: Escores para PC1 com (a) e sem (b) “lacunaridade”.

Comparando-se os gráficos (a) e (b) da Figura 3.36 nota-se claramente o efeito causado quando não se faz a correção de “lacunaridade”. Sem a correção os valores para os escores nos períodos considerados alcançam valores totalmente discrepantes do restante da distribuição ao longo do ano.



(a)



(b)

Fig.3.37: Influência dos dados faltantes em PC1. (a) com e (b) sem “lacunaridade”.

A Figura 3.37 mostra o efeito da “lacunaridade” sobre os dados faltantes, quando se comparam PC1 em cada caso.

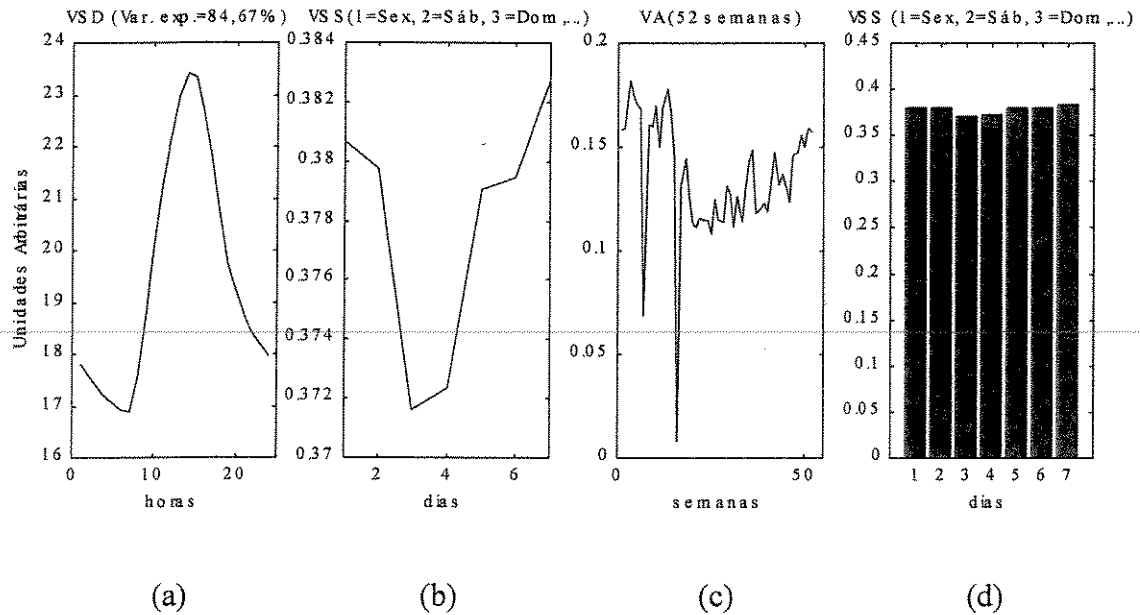


Fig.3.38: (a)Variação sistemática diária (VSD), (b)Variação sistemática semanal (VSS), (c)Variação anual e (d)VSS em gráfico de barras.

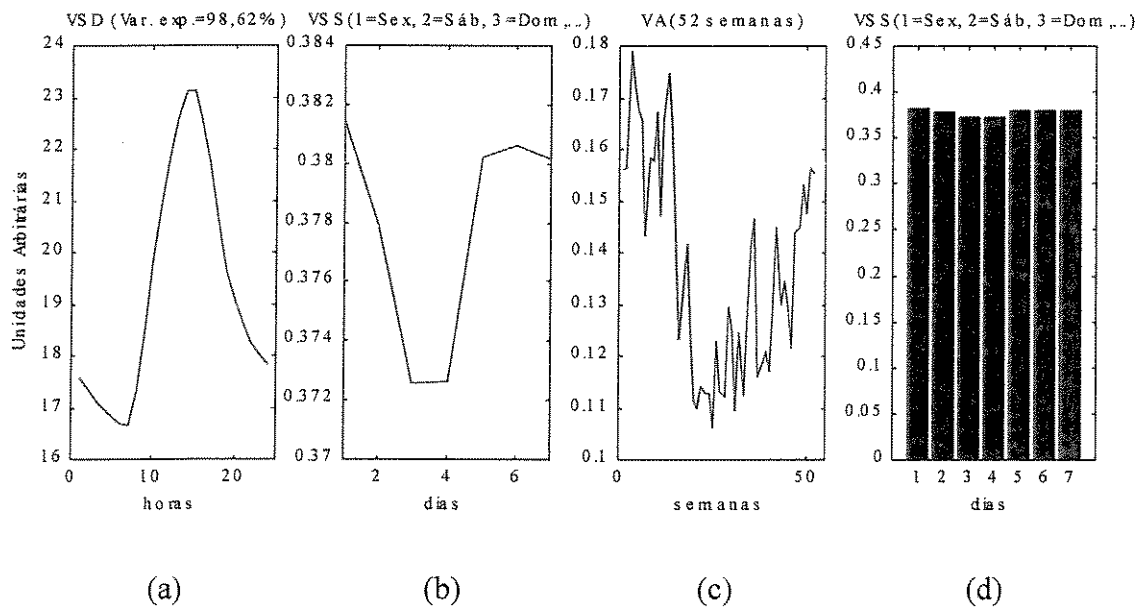


Fig.3.39: (a)Variação sistemática diária (VSD), (b)Variação sistemática semanal (VSS), (c)Variação anual e (d)VSS em gráfico de barras.

As Figuras 3.38 e 3.39 mostram o efeito causado pelo problema nas curvas obtidas por meio do modelo PARAFAC. Verifica-se por meio das Figuras 3.38b e 3.39b que as variações sistemáticas semanais conservam suas características principais (queda das emissões nos finais de semana), mas apresentam pequenas diferenças no perfil.

A Figura 3.38c mostra a ocorrência de valores muito baixos para a variação anual nos períodos afetados pela “lacunaridade”. Tal efeito não ocorre quando o problema é corrigido como mostra a Figura 3.39c.

Para solucionar os problemas de “lacunaridade”, dados faltantes em períodos afetados pelo problema foram substituídos pela média anual. O procedimento envolveu duas etapas. Na primeira os dados foram estimados diretamente pelo modelo. Em seguida esses dados foram substituídos pela média anual aí obtida, sendo a análise PCA novamente aplicada à matriz.

Em função da existência de variação sazonal, o critério acima foi comparado com um segundo procedimento, no qual foi usada média dos dados nos períodos adjacentes ao lacunar. Isto é, a substituição dos dados lacunares pela média nos períodos imediatamente antes e após o período “lacunar”. Tal providência teve como objetivo verificar se a média anual era adequada ao problema.

O procedimento aqui também envolveu duas etapas. Na primeira, foram identificados tais dados e o período no qual se encontravam. Na segunda, esses dados foram substituídos pela média dos dados não lacunares dentro do período, sendo PCA novamente aplicada à matriz de dados. O critério para estabelecer a extensão de dados para efeito de cálculo da média naqueles períodos pré e pós lacunares foi a avaliação da distribuição do erro na matriz analisada, isto é, a avaliação sob o ponto de vista da distribuição aleatória do erro. Segundo o critério acima, os períodos pré e pós lacunares variaram de 15 a 30 dias (360 a 720 dados).

Diferenças desprezíveis foram encontradas nos valores das componentes obtidas segundo os dois procedimentos. Tais distinções não foram significativas o suficiente para gerar perfis diferentes para os poluentes em cada caso.

Em suma, como os resultados segundo os dois procedimentos não mostraram qualquer diferença nos perfis obtidos, optou-se por adotar o primeiro deles (substituição

dos dados lacunares pela média anual) porque, além de maior rapidez e facilidade de aplicação, é mais geral.

A quantidade de dados faltantes em uma dada Estação de Medição é sempre a mesma para todas as variáveis. Por exemplo, na localidade Parque D. Pedro II, o número de dados ausentes para CO, NO, NO₂, O₃, SO₂, CH₄, NMHC, Temp e Umid, sempre totalizava 101 dados. Conforme discutido acima, a quantidade não se configurou como um problema, mas sim a sua distribuição. Tal problema foi resolvido da maneira descrita anteriormente. Entretanto, considerou-se de utilidade expressar o resultado acima em uma única saída para cada poluente analisado.

Assim, para verificar tal influência, isto é, se os dados faltantes, sob o ponto de vista de quantidade e lacunaridade afetariam a análise PCA, tomou-se, aleatoriamente, um dos sete dias de cada uma das 52 semanas. Com esse procedimento, foram construídas dez matrizes de $(364 - 52 = 312)$ dias. As matrizes de 312×24 , nas quais as linhas são os dias e as colunas são as horas, foram divididas em duas: uma para as horas ímpares e outra para as horas pares. Essas matrizes foram submetidas a PCA. Desse modo foram obtidas, para cada variável sob análise, vinte curvas de loadings, as quais representam as variações sistemáticas nas 24 horas.

Esses resultados estão lançados nas Figuras 3.40 a 3.48 a seguir para as variáveis CO, NO, NO₂, O₃, SO₂, CH₄, NMHC, Temp e Umid (Pedro II – 1999).

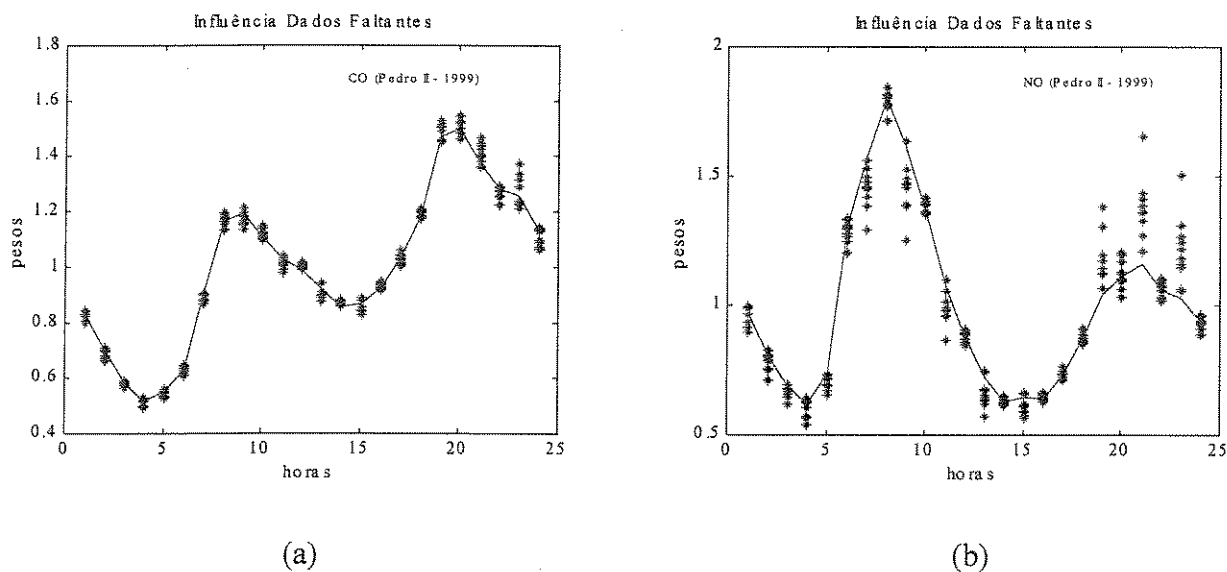
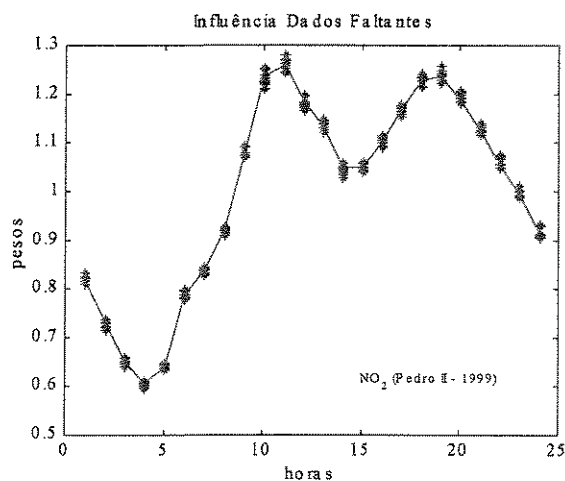
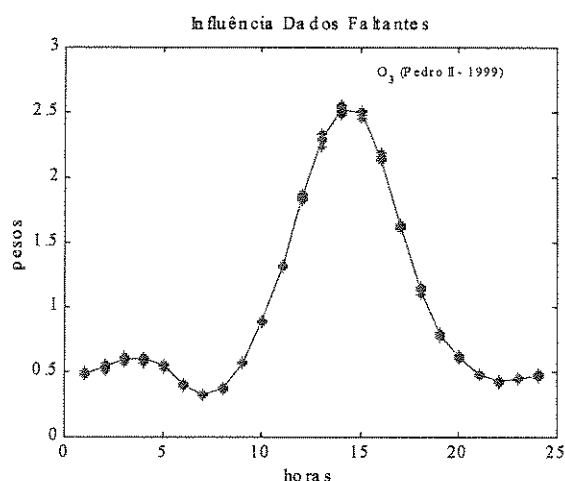


Fig.3.40: Teste para verificar a influência dos dados faltantes. (a) CO, (b) NO.

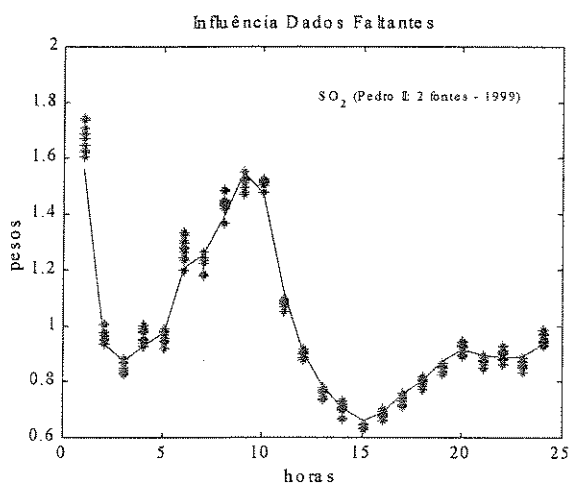


(a)

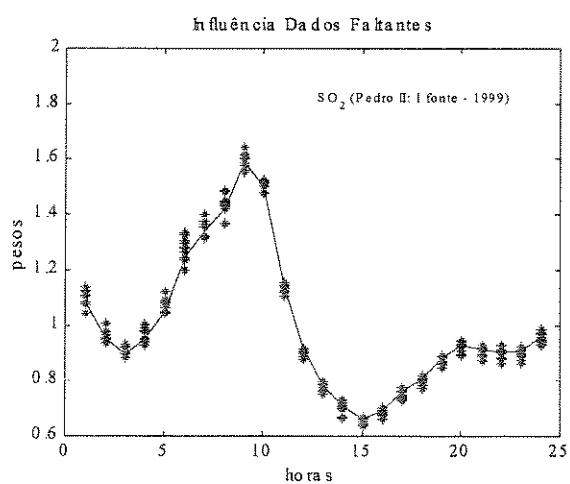


(b)

Fig.3.41: Teste para verificar a influência dos dados faltantes. (a) NO_2 , (b) O_3 .



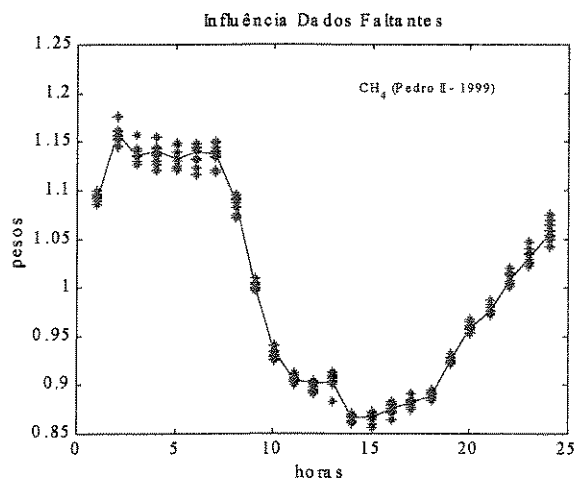
(a)



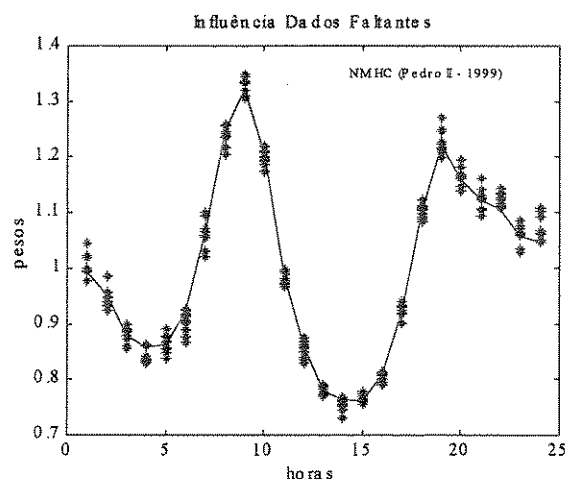
(b)

Fig.3.42: Teste para verificar a influência dos dados faltantes. (a) SO_2 com duas fontes de emissão, (b) SO_2 com apenas 1 fonte de emissão.

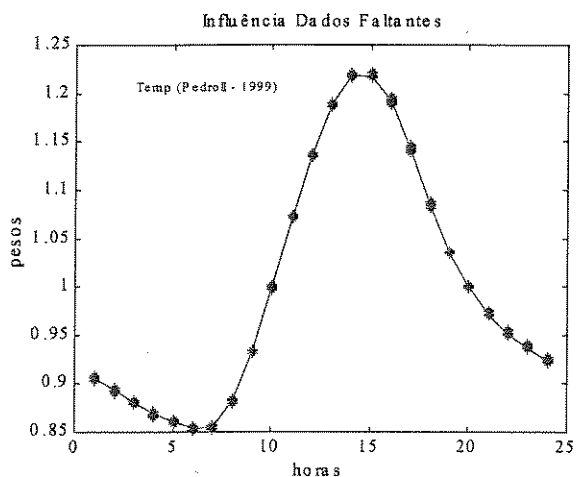
Os aspectos referentes à análise do SO_2 são discutidos de forma mais completa no Capítulo VI (Resultados e Discussão Preliminar).



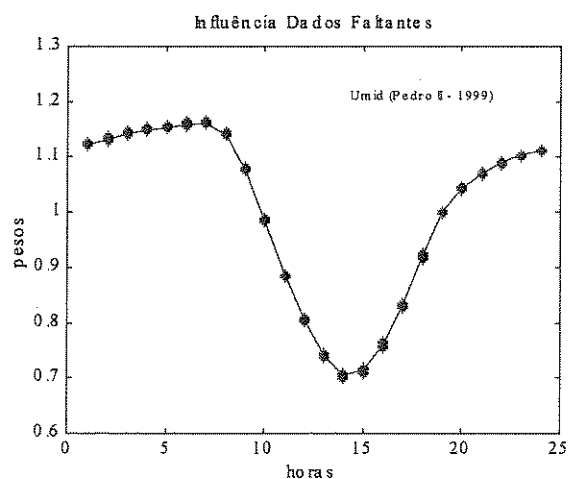
(a)



(b)



(c)



(d)

Fig.3.43: Teste para verificar a influência dos dados faltantes. a) CH_4 , b) NMHC, c) Temp., e d) Umid.

Uma análise nos resultados do teste para verificação da influência que a quantidade dos dados faltantes tem sobre os resultados mostra que, do ponto de vista da quantidade, a influência não é significativa.

3.3.3.1.4 – Preprocessamento de Dados II

Nenhum tipo de transformação foi aplicado aos dados. Para análise dos poluentes individualmente por PCA, nenhum preprocessamento foi aplicado aos dados. Para amenizar o efeito de variações súbitas locais (horárias), quando a análise foi realizada com o modelo PARAFAC para uma variável, e para os modelos PARAFAC e Tucker, para duas ou mais simultaneamente, os dados foram padronizados pelos valores médios do poluente em questão, isto é, para um dado poluente, cada entrada, foi dividida pelo valor médio tomado sobre todos os dados relativos àquele composto (ver equação 3.1, p45), tal que a influência entre eles causada por eventuais variações súbitas fosse atenuada.

3.3.3.1.5 – Restrições

Nessa tese, os dados correspondem à concentrações de poluentes levando-se em conta todos os dias da semana, diferentemente da abordagem de Reis e Ferreira (2002), no qual apenas os dados referentes aos chamados dias úteis da semana (Segunda a Sexta-feira) foram usados. Assim, não foi possível evitar trabalhar com quedas bruscas em valores de concentração, oriundas de quedas nas emissões da frota de veículos durante os finais de semana, o que poderia levar a valores negativos para a concentração em alguns pontos para os perfis obtidos em cada modo nas análises de ordem superior. Assim, como não tem sentido valores negativos para concentração, impôs-se, para a análise multimodo, restrição do tipo não-negatividade aos dados.

Deve ser observado que não obstante ocorra queda nas emissões nos finais de semana, esse fato não invalida o que foi assumido em relação à variações bruscas na metodologia. Essas variações repentinas só escapam à formulação do modelo quando aleatórias (sendo consideradas como dados faltantes, conforme explicado acima). No caso das quedas durante os finais de semana o comportamento caracteriza-se como sistemático eis que ocorrendo periodicamente ao longo do ano.

3.3.3.1.6 – Rotinas desenvolvidas

O primeiro passo dado em relação aos dados foi providenciar a sua validação. Essa etapa inicial, como já mencionado, foi realizada manualmente. Todas as etapas seguintes foram acompanhadas do desenvolvimento ou da utilização de rotinas já existentes. Abaixo segue uma relação daquelas que foram criadas, com explicações sobre o seu funcionamento.

1 – Para colocar os dados da tabela original fornecida pela CETESB em condições de serem submetidos ao pretratamento de dados, construiu-se uma rotina que armazena os dados em matrizes contendo o nome da Estação de Medição, a data, a hora, e a concentração do poluente (e os valores para as variáveis temperatura e umidade), tomando-se por base as informações contidas na tabela original.

Algumas tabelas originais continham informações em padrão diferente. Isto exigiu modificação da rotina para atender a esses casos específicos.

Rotina 1: “Organização da tabela de Dados” (Apêndice A8)

2 – Para construir as matrizes de 365 dias $\times$ 24 horas, foi desenvolvida uma rotina denominada Estação, a partir de um máximo de 8760 vetores de dados. A rotina realiza uma “varredura” nas matrizes contendo informações sobre concentração, data, hora e nome da Estação, estabelecendo vínculo entre essas informações e a matriz de valores de concentração (365 $\times$ 24). Para poder realizar essa operação, a rotina foi construída considerando-se: dia de 24 horas, semana de 7 dias, meses contendo 28 (Fevereiro, 29 dias para ano bissexto – 1996), 30 e 31 dias, e ano de 12 meses. Desse modo, quando o conjunto de dados apresenta menos de 8760 entradas (quase totalidade dos casos), as informações sobre a data (dia, mês, ano) e hora ficam preservadas, tal que a ausência de uma entrada de dado fica, ato contínuo, com uma lacuna. A existência desses conjuntos vazios, denominados dados faltantes garante que todas as matrizes possuam o mesmo tamanho físico (365 $\times$ 24).

Rotina 2: “Construção das matrizes de Dados” (Apêndice A8)

As Figuras que levam como título “Dados Originais” e “Dados Corrigidos”, são exemplos de resultados da aplicação dessas duas rotinas aos dados das tabelas originais fornecidas pela CETESB.

3 – Foi desenvolvida uma rotina especialmente para avaliar a influência da quantidade de dados faltantes nas análises. Infelizmente, quando da sua construção não se previra problemas de “lacunaridade” tal que, quando eles se manifestaram, tiveram que sofrer correção, sem o que a rotina desenvolvida não funcionava a contento.

Rotina 3: “Influência de Dados Faltantes” (Apêndice A8)

A Figura 3.44 (Temp99 – Ibirapuera) é um exemplo de casos nos quais os problemas citados acima ocorreram (Ver também Figuras 3.31 a 3.39).

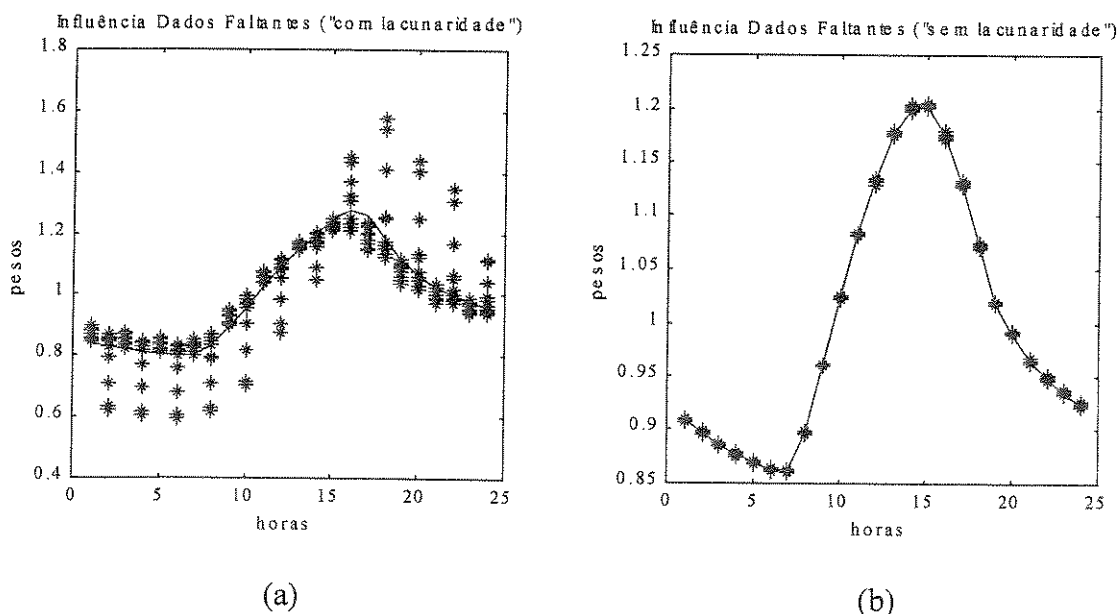


Fig.3.44: a)Aplicação da rotina para verificar a influência de dados faltantes à matrizes com problemas de “lacunaridade”, e b) problema corrigido.

A repercussão nos resultados das análises pode ser avaliada, por exemplo, quando se comparam as matrizes para os dados estimados em cada caso (Figura 3.45), e/ou quando se comparam escores para PCs. Essa última comparação é mostrada na Figura 3.46 para a primeira componente principal.

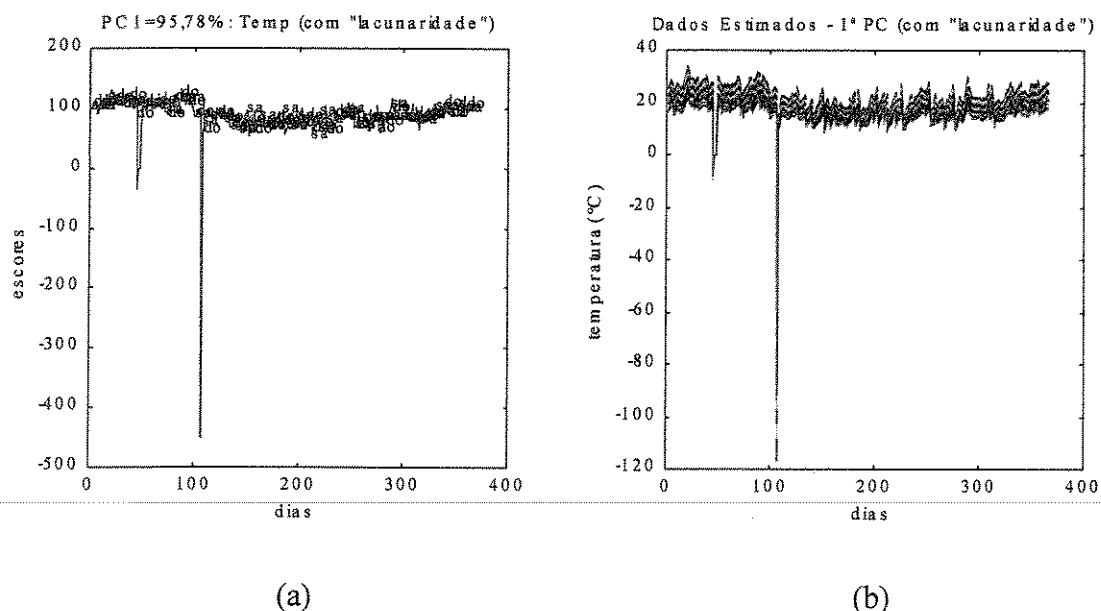


Fig.3.45: a)Escore para PC_1 , e b)dados estimados. Dados exibindo “lacunaridade”.

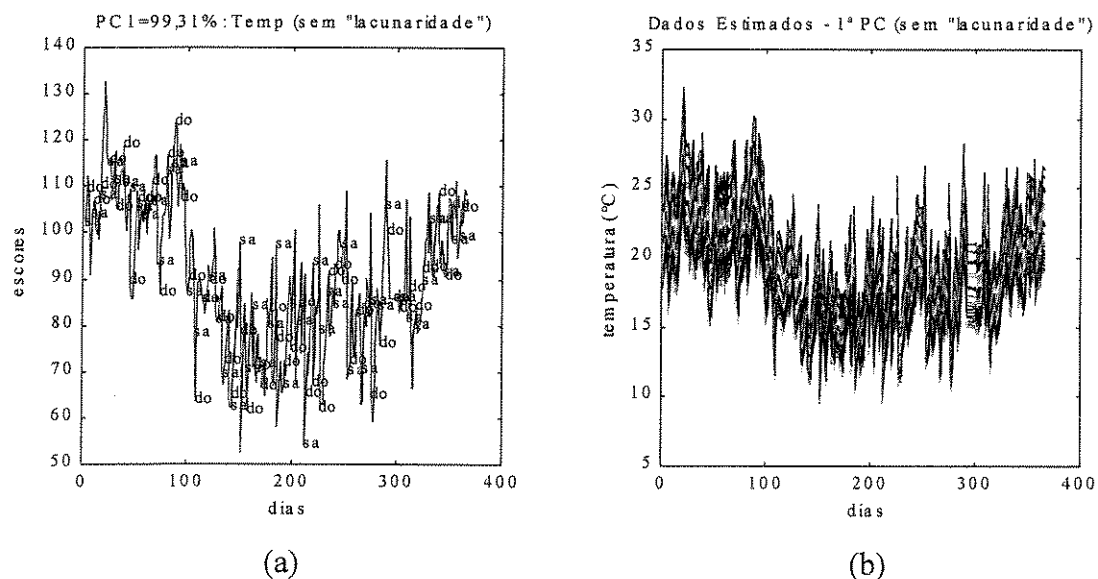


Fig.3.46: a)Escore para PC_1 , e b)dados estimados. Correção de “lacunaridade”.

As Figuras 3.45 e 3.46 mostram resultados com e sem os efeitos causados por problemas de “lacunaridade” (Temp. 1999 – Ibirapuera). É importante observar (Figura 3.45b) que sem as correções de “lacunaridade” a temperatura atinge valores inferiores a cem graus Celsius negativos (-100°C), inaceitáveis para qualquer região do Brasil.

4 – Vários outros pequenos “scripts” (ver Apêndice A8) foram criados. Estão embutidos nas seqüências utilizadas para as análises PCA, PARAFAC e Tucker.

5 – Uma seqüência ampla de saídas para análise PCA com várias informações, incluindo influência de dados faltantes, e modelo PARAFAC (um poluente, CO, para as sete Estações de Medição - 1998), com vários gráficos, é apresentada no Capítulo intitulado Análise de Dados (“Scripts” desenvolvidos: Apêndice A8).

6 – Seqüências típicas de saídas para os modelo PARAFAC e Tucker para o ciclo do monóxido de Carbono (4 poluentes, CO, NO, NO₂, O₃), ciclo do metano (5 poluentes, CO, NO, NO₂, O₃ e CH₄) e ciclo dos hidrocarbonetos não metano, NMHC, (5 poluentes, CO, NO, NO₂, O₃ e NHMC) são lançadas no Capítulo VI (Resultados e Discussão). Os “Scripts” desenvolvidos encontram-se no Apêndice A8.

3.3.3.1.7 – Estrutura Computacional e Programas (“Softwares”)

As análises foram realizadas em um processador AMD K6-2 com velocidade de processamento de 333 MHz e 64 Mb de memória RAM. O software utilizado foi o MATLAB versão 5.1 instalado em plataforma Windows 95.

• Análises em Dois Modos (PCA)

Para as análises em dois modos (PCA) foi usada a rotina MDPCA interna ao conjunto de funções denominado PLS-toolbox.

A rotina MDPCA (Principal Components for matrices with missing data) é uma função que trabalha calculando um modelo PCA e em seguida troca os dados faltantes por valores que são mais consistentes com o modelo. Um novo modelo PCA é então calculado e o processo é repetido até que a estimativa para os dados faltantes convirja.

As entradas são a matriz de dados, o número de componentes principais (PCs) que se julga sejam significativas, e o teste (“flag”) usado para os dados faltantes. Nessa tese foi usado o valor 1,0000EE12 (ou 999999999999). Os dados faltantes são substituídos por esse valor tal que ao encontrá-lo o algoritmo “interpreta-o” como um valor a ser estimado. A variável que controla a tolerância para a convergência (tol) foi usada com o valor 1,0000EE-10 (o valor de “default” é igual a 1,0000EE-05).

As saídas são os scores e loadings estimados para as componentes principais (PCs), e uma estimativa da matriz de dados com os valores estimados pelo algoritmo para os dados faltantes incluídos nessa matriz. O formato de entrada e saída (I/O format) é especificado como:

```
[scores,loads,estdata] = mdpca(data,pcs,flag,tol);
```

• Análises Multimodo (PARAFAC e Tucker)

As análises dos arranjos multimodo (PARAFAC e Tucker) foram realizadas com o programa “The N-way Toolbox version 1.02 for MATLAB” desenvolvido por R. Bro & C. A. Andersson (Chemometrics Group, Food Technology, Department of Food and Dairy Science – The Royal Veterinary and Agricultural University/ Denmark) encontrado no endereço <http://newton.mli.kvl.dk/foodtech.html>.

♦ PARAFAC

Para um arranjo em três modos, a solução para o modelo PARAFAC pode ser encontrada por meio da aplicação de quadrados mínimos alternantes assumindo-se, sucessivamente, que os pesos (“loadings”) em dois modos (1° e 2°) são conhecidos e, estimando-se o conjunto de parâmetros desconhecidos do último modo (Bro, 1997). Esse procedimento é realizado para cada um dos modos.

Tomando como ponto de partida a equação (3.69), e fazendo-se $\mathbf{Z} = (\mathbf{C}^T \otimes \mathbf{B}^T)$, o modelo definindo $\mathbf{A}$ torna-se:

$$\mathbf{X} = \mathbf{A} \mathbf{Z} \quad (3.101)$$

onde $\mathbf{X}$ é o arranjo $\underline{\mathbf{X}}$ desdobrado em dois modos ($I \times J K$) e $\mathbf{A}$ é ($I \times R_{pp}$).

A estimativa condicional por meio de quadrados mínimos para $\mathbf{A}$ é:

$$\mathbf{A} = \mathbf{X} \mathbf{Z}^T (\mathbf{Z} \mathbf{Z}^T)^{-1} \quad (3.102)$$

Para estimar $\mathbf{B}$, o arranjo multimodo $\underline{\mathbf{X}}$ é desdobrado em uma matriz ($J \times I K$). Desse modo, $\mathbf{Z}$ torna-se uma matriz ($R_{pp} \times J K$) calculada a partir de $\mathbf{A}$ e $\mathbf{C}$. Assim, a matriz $\mathbf{B}$ torna-se $\mathbf{X} \mathbf{Z}^T (\mathbf{Z} \mathbf{Z}^T)^{-1}$. O mesmo procedimento é usado para estimar $\mathbf{C}$.

O algoritmo ALS geral para o modelo PARAFAC pode ser escrito da seguinte maneira (Bro, 1997):

- (0) Decidir o número de componentes, R_{pp} (o pseudo posto de $\mathbf{A}$, $\mathbf{B}$ e $\mathbf{C}$),
- (1) “Inicializar” $\mathbf{B}$ e $\mathbf{C}$,
- (2) Estimar $\mathbf{A}$ a partir de $\underline{\mathbf{X}}$, $\mathbf{B}$ e $\mathbf{C}$, por meio de regressão por quadrados mínimos,
- (3) Estimar $\mathbf{B}$ da mesma maneira,
- (4) Estimar $\mathbf{C}$ da mesma maneira,
- (5) Continuar o processo até ocorrer convergência (pequena mudança no ajuste ou nos “loadings”).

♦ **Tucker**

Considerando-se também aqui um arranjo em três modos $\underline{\mathbf{X}} (I \times J \times K)$, procura-se um modelo Tucker3 de pseudo postos P_{pp} , Q_{pp} e R_{pp} , respectivamente, para os modos $\mathbf{A}$, $\mathbf{B}$ e $\mathbf{C}$. Desdobrando-se $\underline{\mathbf{X}}$ em $\mathbf{X}_A(I \times J \times K)$ obtém-se (Andersson e Bro, 1998):

$$\mathbf{X}_A = \mathbf{A} \mathbf{G}_A (\mathbf{C}^T \otimes \mathbf{B}^T) + \mathbf{E}_A \quad (3.103)$$

onde $\mathbf{A} \mathbf{G}_A (\mathbf{C}^T \otimes \mathbf{B}^T)$ é o modelo para $\mathbf{X}_A$ e $\mathbf{E}_A$ é a parte não modelada (resíduo). Na matriz $\mathbf{X}_A(I \times J \times K)$, o subscrito A indica o modo que está sendo preservado. A ordem $J \times K$ indica que para cada $k=1$ até $k=K$, J está variando de $j=1$ até $j=J$. O núcleo $\underline{\mathbf{G}}$ é desdobrado em uma matriz $\mathbf{G}_A(P_{pp} \times Q_{pp} \times R_{pp})$.

Analogamente, tem-se para $\mathbf{X}_B$ e $\mathbf{X}_C$

$$\mathbf{X}_B = \mathbf{B} \mathbf{G}_B (\mathbf{C}^T \otimes \mathbf{A}^T) + \mathbf{E}_B \quad (3.104)$$

$$\mathbf{X}_C = \mathbf{C} \mathbf{G}_C (\mathbf{B}^T \otimes \mathbf{A}^T) + \mathbf{E}_C \quad (3.105)$$

O núcleo $\underline{\mathbf{G}}$ para $\mathbf{A}$, $\mathbf{B}$ e $\mathbf{C}$ pode ser encontrado por meio da projeção de $\underline{\mathbf{X}}$ sobre $\mathbf{A}$, $\mathbf{B}$ e $\mathbf{C}$, obtendo-se para os modos A, B e C:

$$\mathbf{G}_A = \mathbf{A}^T \mathbf{X}_A (\mathbf{C} \otimes \mathbf{B}) \quad (3.106)$$

$$\mathbf{G}_B = \mathbf{B}^T \mathbf{X}_B (\mathbf{C} \otimes \mathbf{A}) \quad (3.107)$$

$$\mathbf{G}_C = \mathbf{C}^T \mathbf{X}_C (\mathbf{B} \otimes \mathbf{A}) \quad (3.108)$$

Para a explanação que segue considerar-se-á que $\mathbf{A}$, $\mathbf{B}$ e $\mathbf{C}$ são ortogonais com dimensões coluna, respectivamente, iguais a seus pseudo postos P_{pp} , Q_{pp} e R_{pp} . Se o modelo é perfeito, $\underline{\mathbf{G}}$ contém exatamente a mesma informação que $\underline{\mathbf{X}}$, embora expresso por meio de coordenadas diferentes (Bro, 1998, p.70).

Substituindo-se a equação 3.106 na equação 3.103 e omitindo-se os resíduos, obtém-se:

$$\mathbf{X}_A = \mathbf{A} \mathbf{G}_A (\mathbf{C}^T \otimes \mathbf{B}^T) = \mathbf{A} \mathbf{A}^T \mathbf{X}_A (\mathbf{C} \otimes \mathbf{B}) (\mathbf{C}^T \otimes \mathbf{B}^T) \quad (3.109)$$

Fixando-se $\mathbf{B}$ e $\mathbf{C}$, e considerando-se a função objetiva para o modelo Tucker3 (ver equação 3.67, isto é, $\min_{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{G}} \|\mathbf{X} - \mathbf{A} \mathbf{G} (\mathbf{C}^T \otimes \mathbf{B}^T)\|_F^2$), encontrar o melhor valor para $\mathbf{A}$ significa tornar mínima a norma de $(\mathbf{X} - \mathbf{A} \mathbf{A}^T \mathbf{X}_A (\mathbf{C} \otimes \mathbf{B}) (\mathbf{C}^T \otimes \mathbf{B}^T))$.

Considerando-se o projetor ortogonal, $(\mathbf{A} \mathbf{A}^+)$, sobre o espaço das colunas de $\mathbf{A}$, e o projetor, $(\mathbf{I} - \mathbf{A} \mathbf{A}^+)$, sobre o espaço nulo de $\mathbf{A}$, e considerando-se que se $\mathbf{A}$ tem colunas ortonormais $\mathbf{A}^+ = \mathbf{A}^T$, então $\mathbf{A} \mathbf{A}^T = \mathbf{A} \mathbf{A}^+$ é também um projetor. O mesmo se aplica à $\mathbf{B} \mathbf{B}^T$ e $\mathbf{C} \mathbf{C}^T$, e à $(\mathbf{C} \otimes \mathbf{B}) (\mathbf{C}^T \otimes \mathbf{B}^T)$, tal que o problema pode ser expresso (Bro, 1998, p.71):

$$\min_{\mathbf{A}} \|\mathbf{X} - \mathbf{A} \mathbf{A}^T \mathbf{X} (\mathbf{C} \otimes \mathbf{B}) (\mathbf{C}^T \otimes \mathbf{B}^T)\|_F^2 \quad (3.110)$$

ou

$$\min_{\mathbf{A}} \|\mathbf{X} \mathbf{I} - (\mathbf{C} \otimes \mathbf{B}) (\mathbf{C}^T \otimes \mathbf{B}^T) + \mathbf{X} (\mathbf{C} \otimes \mathbf{B}) (\mathbf{C}^T \otimes \mathbf{B}^T) - \mathbf{A} \mathbf{A}^T \mathbf{X} (\mathbf{C} \otimes \mathbf{B}) (\mathbf{C}^T \otimes \mathbf{B}^T)\| \quad (3.111)$$

Quando $\mathbf{I} - (\mathbf{C} \otimes \mathbf{B}) (\mathbf{C}^T \otimes \mathbf{B}^T)$ e $(\mathbf{C} \otimes \mathbf{B}) (\mathbf{C}^T \otimes \mathbf{B}^T)$ geram espaços ortogonais, a equação 3.111 pode ser expressa:

$$\min_{\mathbf{A}} \|\mathbf{X} \mathbf{I} - (\mathbf{C} \otimes \mathbf{B}) (\mathbf{C}^T \otimes \mathbf{B}^T)\|_F^2 + \|\mathbf{X} (\mathbf{C} \otimes \mathbf{B}) (\mathbf{C}^T \otimes \mathbf{B}^T) - \mathbf{A} \mathbf{A}^T \mathbf{X} (\mathbf{C} \otimes \mathbf{B}) (\mathbf{C}^T \otimes \mathbf{B}^T)\| \quad (3.112)$$

Quando $\|\mathbf{X} \mathbf{I} - (\mathbf{C} \otimes \mathbf{B}) (\mathbf{C}^T \otimes \mathbf{B}^T)\|_F^2$ é constante, a equação 3.112 é equivalente a:

$$\min_{\mathbf{A}} \|\mathbf{X} (\mathbf{C} \otimes \mathbf{B}) (\mathbf{C}^T \otimes \mathbf{B}^T) - \mathbf{A} \mathbf{A}^T \mathbf{X} (\mathbf{C} \otimes \mathbf{B}) (\mathbf{C}^T \otimes \mathbf{B}^T)\| \quad (3.113)$$

ou

$$\min_{\mathbf{A}} \|\mathbf{I} - \mathbf{A} \mathbf{A}^T \mathbf{X} (\mathbf{C} \otimes \mathbf{B}) (\mathbf{C} \otimes \mathbf{B}^T)\|_F^2 \quad (3.114)$$

ou ainda,

$$\max_A \left\| \mathbf{A} \mathbf{A}^T \mathbf{X} (\mathbf{C} \otimes \mathbf{B}) (\mathbf{C}^T \otimes \mathbf{B}^T) \right\|_F^2 \quad (3.115)$$

O máximo revelado pela expressão 3.115 é obtido quando $\mathbf{X} (\mathbf{C} \otimes \mathbf{B}) (\mathbf{C}^T \otimes \mathbf{B}^T)$ é projetado sobre seu próprio ‘sub-espço principal’, isto é, quando o espaço ocupado pelas colunas de $\mathbf{A}$ é, de fato, o mesmo sub-espço gerado pelos primeiros P_{pp} vetores singulares à esquerda de $\mathbf{X} (\mathbf{C} \otimes \mathbf{B}) (\mathbf{C}^T \otimes \mathbf{B}^T)$, ou seja, ao espaço gerado pelos primeiros P_{pp} autovetores da matriz $[\mathbf{X} (\mathbf{C} \otimes \mathbf{B}) (\mathbf{C}^T \otimes \mathbf{B}^T)] [\mathbf{X} (\mathbf{C} \otimes \mathbf{B}) (\mathbf{C}^T \otimes \mathbf{B}^T)]^T$.

Considerando-se que:

$$\begin{aligned} & [\mathbf{X} (\mathbf{C} \otimes \mathbf{B}) (\mathbf{C}^T \otimes \mathbf{B}^T)] [\mathbf{X} (\mathbf{C} \otimes \mathbf{B}) (\mathbf{C}^T \otimes \mathbf{B}^T)]^T = \\ & = [\mathbf{X} (\mathbf{C} \otimes \mathbf{B}) (\mathbf{C}^T \otimes \mathbf{B}^T)] [(\mathbf{C} \otimes \mathbf{B}) (\mathbf{C}^T \otimes \mathbf{B}^T)]^T \mathbf{X}^T = \\ & = [\mathbf{X} (\mathbf{C} \otimes \mathbf{B}) (\mathbf{C}^T \otimes \mathbf{B}^T)] \mathbf{X}^T = \mathbf{X} (\mathbf{C} \otimes \mathbf{B}) (\mathbf{C}^T \otimes \mathbf{B}^T) \end{aligned} \quad (3.116)$$

e usando as propriedades de simetria e “idempotenticidade” de operadores de projeção ortogonal, a matriz procurada pode ser obtida a partir da matriz $\mathbf{X} (\mathbf{C} \otimes \mathbf{B})$ por extração dos vetores singulares à esquerda (Bro, 1998, p.72). Desse modo, pode-se obter um algoritmo genérico para o modelo Tucker3, como indicado em seguida.

Algoritmo Tucker3:

“Inicialize” $\mathbf{B}$ e $\mathbf{C}$,

$$1. [\mathbf{A}, \mathbf{\Lambda}, \mathbf{P}] = \text{svd} (\mathbf{X}_A (\mathbf{C} \otimes \mathbf{B}), P_{pp}),$$

$$2. [\mathbf{B}, \mathbf{\Lambda}, \mathbf{P}] = \text{svd} (\mathbf{X}_B (\mathbf{C} \otimes \mathbf{A}), Q_{pp}),$$

$$3. [\mathbf{C}, \mathbf{\Lambda}, \mathbf{P}] = \text{svd} (\mathbf{X}_C (\mathbf{B} \otimes \mathbf{A}), R_{pp}),$$

4. Vá para a etapa 1 até que a mudança no ajuste seja relativamente pequena,

$$6. \mathbf{G} = \mathbf{A} \mathbf{X} (\mathbf{C} \otimes \mathbf{B})$$

No algoritmo acima, $[\mathbf{A}, \mathbf{\Lambda}, \mathbf{P}] = \text{svd} (\mathbf{X}_A (\mathbf{C} \otimes \mathbf{B}), P_{pp})$, significa que a decomposição por valor singular de $[\mathbf{X}_A (\mathbf{C} \otimes \mathbf{B})]$ é truncada de modo que sejam retidas as primeiras P_{pp} componentes. A matriz $\mathbf{A}$ contém os primeiros P_{pp} vetores singulares à

esquerda, $\mathbf{P}$ contém os primeiros P_{pp} vetores singulares à direita, e Λ é uma matriz diagonal com os primeiros P_{pp} valores singulares em sua diagonal (Bro, 1998, p.72). O significado para SVD das etapas 2 e 3 pode ser apreendido por analogia à etapa 1.

CAPÍTULO IV

ORIGEM DOS DADOS

4.1 – A Origem dos Dados

Os dados analisados foram cedidos pela Companhia de Tecnologia e Saneamento Básico (CETESB), do Estado de São Paulo. Tais dados referem-se a medições realizadas pela rede telemétrica da companhia envolvendo as substâncias monóxido de carbono, CO, os óxidos de nitrogênio, NO e NO₂ (e a soma deles, NO_x), ozônio, O₃, dióxido de enxofre, SO₂, metano, CH₄, e os hidrocarbonetos não-metano (NMHC). Foram coletados também dados sobre a partícula inalável (PI), além dos valores de temperatura e umidade. O período total coberto vai de 1996 a 1999. Os registros são efetuados em intervalos de uma hora para cada entidade (variável medida) em sete estações de medição (Cerqueira Cesar, Centro, Lapa, Osasco, Parque D. Pedro II, São Caetano do Sul e Ibirapuera). O critério de escolha dessas estações foi principalmente baseado na existência da maior quantidade possível de dados comuns às mesmas.

4.2 – As Estações de Medição

A CETESB opera com uma rede automática de monitoramento do ar desde 1981. Possui duas redes manuais que medem os teores de dióxido de enxofre /fumaça desde 1973 e uma outra que mede as partículas totais em suspensão desde 1983 na RMSP e em Cubatão. Possui também, em diversas cidades do interior de São Paulo, uma rede de amostragem manual medindo dióxido de enxofre e fumaça desde 1986 (CETESB, 1998).

A rede manual da RMSP e de Cubatão é composta de oito estações de amostragem, as quais medem SO_2 e fumaça, e onze estações medindo partículas totais em suspensão. Todas as medições são efetuadas a cada seis dias.

A rede automática é composta de 24 (vinte e quatro) estações fixas e duas móveis. Vinte e dois locais de medição, além da sede (Pinheiros) situados na RMSP (Figura 4.1) e dois na área de Cubatão. A rede mede os seguintes parâmetros: CO , NO , NO_2 , O_3 , SO_2 , hidrocarbonetos, partículas inaláveis, temperatura, umidade, velocidade do vento, pressão e radiação solar (global e ultravioleta) como mostra a Tabela (V.1) no Apêndice (B1).

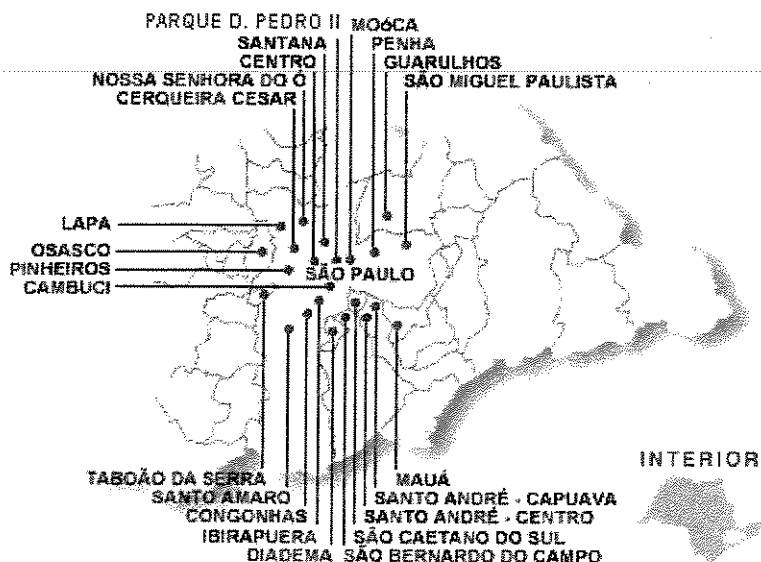


Fig.4.1: Estações de medição da rede telemétrica da RMSP

A Figura 4.1 mostra a localização das Estações de Medição da rede automática. O Parque D. Pedro II está em destaque por ser a localidade, que contém a maior quantidade de informações dentre as sete fornecidas e entre os quatro anos estudados.

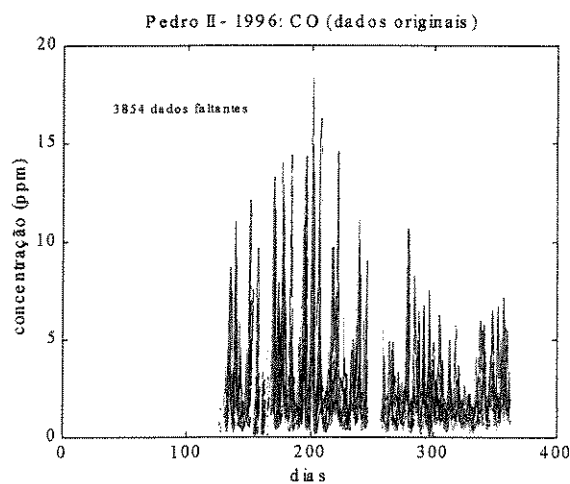
No Apêndice B1 é apresentada uma outra Figura da RMSP, na qual é possível uma melhor visualização da distribuição das Estações de Medição da CETESB na área.

4.3 – Os Dados

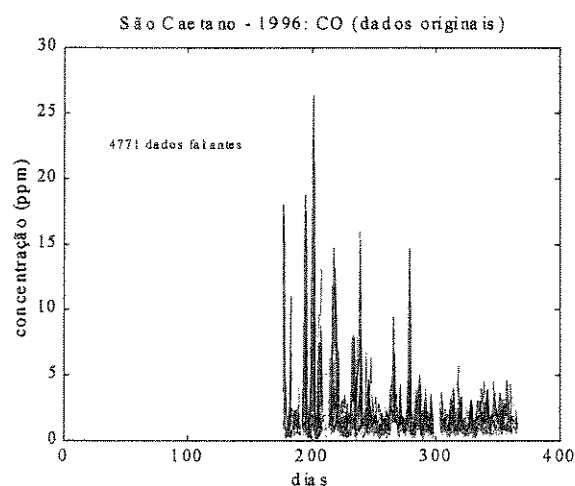
Apenas uma Estação de Medição não possuía dados faltantes (Centro – 1998). Todas as demais, para os anos de 1997 a 1999, possuíam quantidades de dados faltantes que não ultrapassavam cerca de 500 dados, sendo boa parte deles distribuída ao longo do ano, isto é, as lacunas não eram extensas o suficiente para causar sérios problemas nas análises (ou quando o foram, puderam ser resolvidos). Entretanto, enquanto para os anos de 1997 a 1999 o problema da quantidade de dados faltantes ficava limitada a algumas das Estações de Medição, para o ano de 1996 quantidade de dados faltantes era muito elevada em todas as localidades, inviabilizando as análises. Por exemplo, o CO, na Estação de Medição Pedro II, continha 3854 dados faltantes, dos quais, 3120 em uma única lacuna. O primeiro dado registrado nesse caso ocorreu no 130º dia. Em 1996 esta foi a variável que registrou um dos menores valores para a quantidade de dados faltantes. Os dados para os óxidos de nitrogênio, NO e NO₂, possuíam lacunas com valores ainda maiores (NO com 4709 e NO₂ com 5101 dados faltantes). E para os hidrocarbonetos não metano, NMHC, o problema é ainda mais grave, pois a lacuna alcançava a marca de 6121 dados faltantes, ou seja, 69,8% do total.

Não obstante, mesmo em um processo lento, o modelo usado possa extrair alguns perfis, o algoritmo estima todos os dados faltantes como zeros, muito provavelmente valores irreais. Além disso, o algoritmo não converge após o número máximo de iterações (50 iterações para o programa MDPCA), ficando a soma dos quadrados dos resíduos na estimativa dos dados faltantes (ordem de grandeza 10²), muito distante do valor estipulado pelo “flag” ($1,00 \times 10^{-10}$). Esses casos diferem daqueles nos quais os valores já são originalmente nulos. Ademais, diferem também pela quantidade muito elevada de dados faltantes em uma única lacuna.

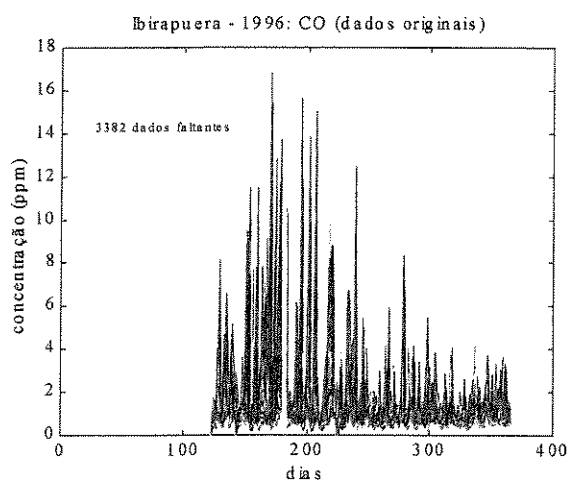
As figuras abaixo mostram os gráficos correspondentes aos dados originais para alguns dos casos citados acima.



(a)



(b)



(c)

Fig.4.2: Dados originais para o CO durante o ano de 1996. a)P. D. Pedro II contendo 3854 dados faltantes, b)São Caetano do Sul contendo 4771 dados faltantes, e c)Ibirapuera contendo 3382 dados faltantes.

A Figura 4.2 mostra a distribuição de dados faltantes para o CO nas três localidades (Pedro II, São Caetano e Ibirapuera: 1996). Deve ser observado que é possível analisar os dados a partir do ponto em que tem início a distribuição. Entretanto, perde-se a informação sobre a variação sazonal, o que provoca reflexos nas análises com os modelos multimodo, especialmente quando os dados são submetidos ao modelo de Tucker.

Para o NO e NO₂ (figuras ausentes) o problema é semelhante ao do CO em 1996.

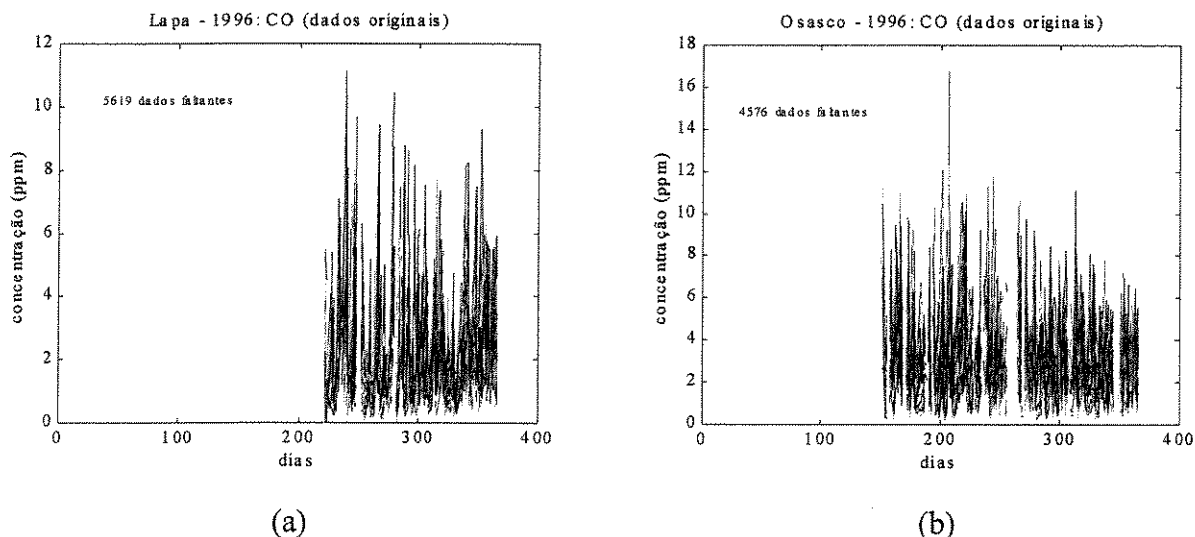


Fig.4.3: Dados originais para o CO durante o ano de 1996. a) Estação de Medição Lapa, 3619 dados faltantes, e b) Estação de Medição Osasco, 4576 dados faltantes.

As Figuras 4.3a e 4.3b mostram as distribuições para os dados faltantes para o CO durante o ano de 1999, respectivamente, nas localidades Lapa e Osasco. Existem casos nos quais a distribuição é diferente da mostrada acima. Isto é, não contém uma lacuna tão extensa, mas pequenos trechos de dados intercalados com grandes lacunas, chegando a um total acima de 6000 dados faltantes. A análise desses dados também apresenta sérias dificuldades pois mais da metade dos dados deve ser estimada pelo modelo.

Assim sendo, optou-se por excluir da análise, os dados referentes ao ano de 1996.

4.4 – Instrumentos de Medição

A CETESB vem realizando suas análises sobre indicadores de qualidade do ar por meio de uma rede mista, uma delas constituída de estações de coleta de material para análise em laboratório central (métodos “manuais”) e outra constituída de rede telemétrica.

A rede telemétrica, de onde se originam os dados usados nesse trabalho, é equipada com instrumentação capaz de medir on-line vários poluentes na atmosfera da Região Metropolitana de São Paulo (RMSP), e em várias localidades do interior do estado.

Na rede não automática são utilizados os seguintes métodos de análise:

Partículas Totais em Suspensão – Amostrador de Grandes Volumes

Partículas Inaláveis (PM10) – Separação Inercial/Filtração

Fumaça – Reflectância

Dióxido de Enxofre – Acidimétrico (peróxido de hidrogênio) e Pararosanilina

Monóxido de Carbono – Infravermelho Não-Dispersivo (contínuo)

Ozônio – Quimiluminescência

Dióxido de Nitrogênio – Quimiluminescência

Na rede automática:

Material Particulado em Suspensão (PM10) – Radiometria/atenuação β

Partícula Inalável – Radiometria/atenuação β

Dióxido de Enxofre – Fluorescência de Pulso (ultravioleta)

Monóxido de Carbono – Espectrometria de Correlação com Filtro de Gás e Infravermelho não Dispersivo

Óxidos de Nitrogênio (NO, NO₂) – Quimiluminescência

Ozônio – Fotometria UV

Hidrocarbonetos – Cromatografia Gasosa/Ionização de Chama

Detalhes adicionais sobre o funcionamento, aplicações, regimes de trabalho e limites de detecção (LOD) de alguns instrumentos encontram-se no Apêndice B.2.

A Tabela V.1 mostra as variáveis analisadas (coluna 1) com suas respectivas unidades (coluna 2).

Tabela IV.1: Variável e suas respectivas unidades.

Variável	Unidade
Monóxido de carbono – CO	ppm
Óxido nítrico – NO	$\mu\text{g}/\text{m}^3$
Dióxido de Nitrogênio – NO ₂	$\mu\text{g}/\text{m}^3$
Ozônio – O ₃	$\mu\text{g}/\text{m}^3$
Dióxido de enxofre – SO ₂	$\mu\text{g}/\text{m}^3$
Metano – CH ₄	$\mu\text{g}/\text{m}^3$
Hidrocarbonetos não metano – NMHC	$\mu\text{g}/\text{m}^3$
Temperatura – Temp	°C
Umidade – Umid	%
Partícula inalável – PI	$\mu\text{g}/\text{m}^3$

Fonte: Adaptado do Relatório de Qualidade do Ar no Estado de São Paulo (CETESB – 1999)

4.5 – Estação de Trabalho

Informações sobre a estrutura computacional, rotinas desenvolvidas e programas utilizados encontram-se no Capítulo III, seção 3.3, subseção 3.3.3, tópico 3.3.3.1.7, na página 159.

Os dados fornecidos encontravam-se agrupados por variáveis (poluentes) para cada ano. Com essa configuração, um vetor típico de dados deveria possuir 52560 dados (no caso das 7 estações esse valor passa para 61320) para cada ano.

Como os dados foram organizados incluindo data, hora, estação e concentração do poluente, a matriz construída para acomodá-los possui um tamanho considerável, exigindo, da máquina mencionada acima, um tempo computacional de cerca de duas horas de processamento para cada poluente no processo de construção de cada matriz (uma para cada variável).

Para a análise PCA, após a organização dos dados, o tempo computacional situou-se na faixa de poucos segundos à cerca de 20 segundos / poluente / ano / estação (não computado o tempo para construir os gráficos, cerca de dez ou mais, dependendo da análise). Só foram incluídos na tese os gráficos considerados relevantes para a interpretação dos resultados.

No caso do modelo PARAFAC, a convergência (default 5000 iterações) era sempre muito rápida quando o número de componentes estava correto e a “lacunaridade” dos dados não era um problema (alguns segundos, com menos de 100 iterações). Em outros casos atingia-se cerca de alguns minutos para atingir a convergência. A saída fornecia as matrizes de “loadings”, cujos gráficos, em geral 4 por poluente, eram função do número de modos e dos objetivos a explorar. Detalhes adicionais na capítulo sobre análise de dados.

No caso do modelo de Tucker, a convergência (gradual, desde 200 até um máximo de iterações igual a 2000) sempre levou mais tempo que no caso do PARAFAC. Isto era esperado porque a estrutura para o Tucker pode conter modos com dimensões diferentes, além das operações realizadas pelo núcleo para o cruzamento das informações contidas em pesos (loadings) de índices diferentes nas matrizes de componentes. Em alguns casos o sistema levou algumas horas para processar os dados. Vários gráficos foram obtidos por meio da aplicação desse modelo, o qual permitiu avaliar tanto as variações sistemáticas diárias e semanais e a variação anual, quanto a acumulação de poluentes na atmosfera.

Detalhes sobre essas análises encontram-se no Capítulo V (Análise de Dados), e no Capítulo VI (Resultados e Discussão).

CAPÍTULO V

ANÁLISE DE DADOS

5.1 – “ARQUITETURA” DA ANÁLISE DE DADOS

A quantidade de dados analisada é particularmente grande. Os métodos e modelos utilizados permitem que se obtenha um número muito grande de saídas, sejam elas gráficas ou não. Assim, optou-se por dividir o capítulo sobre a análise de dados, resultados e discussões em duas partes: a primeira dedicada a fornecer uma visão “panorâmica” do conjunto de dados, apresentada no presente capítulo denominado Análise de Dados. A segunda centrada na apresentação apenas dos resultados que fornecem suporte para discussão em torno dos perfis de emissão das fontes obtidos por meio das análises. Essa segunda parte, e a discussão ela própria, são apresentadas no Capítulo VI.

No tópico 5.1.1 é apresentado um conjunto de figuras destacando aspectos dos dados originais que sirvam como referência para comparações entre as variáveis estudadas. São apresentados também alguns resultados de análises tais como, os gráficos de escores e de pesos para as componentes principais, as matrizes contendo dados estimados e resíduos. Esses resultados são referentes às variáveis CO, NO, NO₂, O₃, SO₂, CH₄, NMHC, Temperatura e Umidade, na Estação de Medição Parque D. Pedro II, para o ano de 1999.

No tópico 5.1.2 são apresentados os resultados das análises para o monóxido de carbono, CO, em todas as Estações de Medição (Parque D. Pedro II, São Caetano, Ibirapuera, Cerqueira César, Centro, Lapa e Osasco), para o ano de 1998.

Finalmente, no tópico 5.1.3, tecem-se comentários sobre os resultados para algumas das variáveis, dentre as nove (CO, NO, NO₂, O₃, SO₂, CH₄, NMHC, Temperatura e Umidade), em algumas das sete Estações de Medição no período de 1997 a 1999.

Deve ser observado que esse capítulo se limita a apresentar resultados que sirvam como fonte introdutória de informações para subsidiar discussão que ocorre de forma incipiente no presente capítulo (Análise de Dados), e de forma mais completa, ainda que preliminar, no Capítulo VI (Resultados e Discussão).

5.1.1 – Análise PCA (P. D. Pedro II – Nove Variáveis: 1999)

Nos tópicos 5.1.1.1 a 5.1.1.9 pretende-se dar uma visão geral dos dados. Para isto, considerou-se uma forma padrão de apresentação dos resultados capaz de oferecer uma comparação entre os níveis de emissão, o perfil (ou perfis) de emissão das fontes, os dados estimados e os resíduos envolvidos nas análises. Desse modo a seqüência de resultados engloba figuras contendo gráficos para a matriz de dados originais, matriz de dados corrigidos, matriz de escores e de pesos para as componentes principais, a matriz de dados estimados pelo modelo, e a matriz de resíduos deixados pelo modelo.

5.1.1.1 – Análise PCA (P. D. Pedro II – CO: 1999)

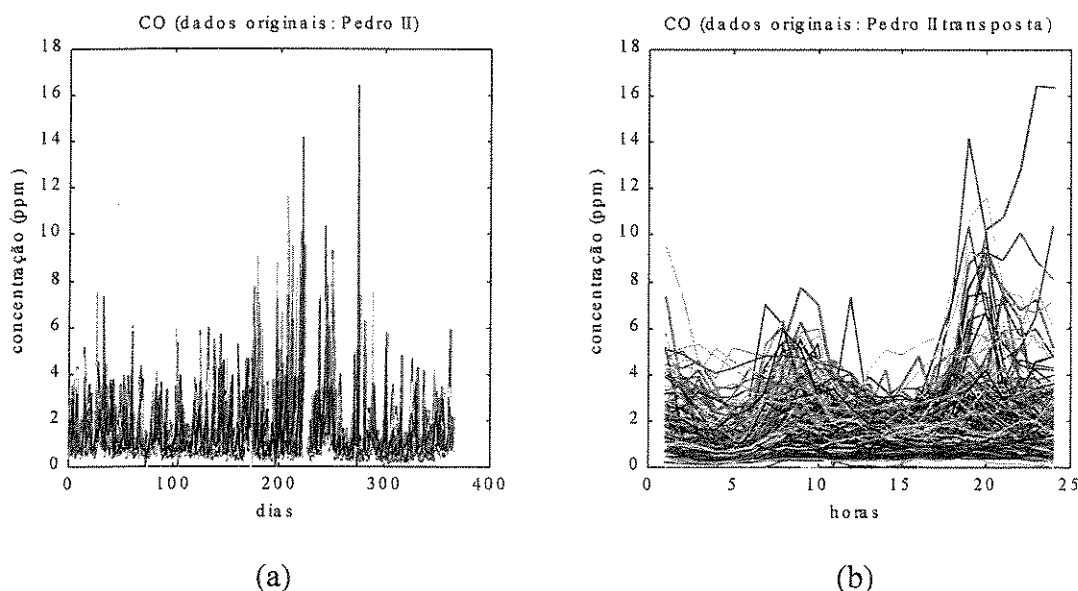
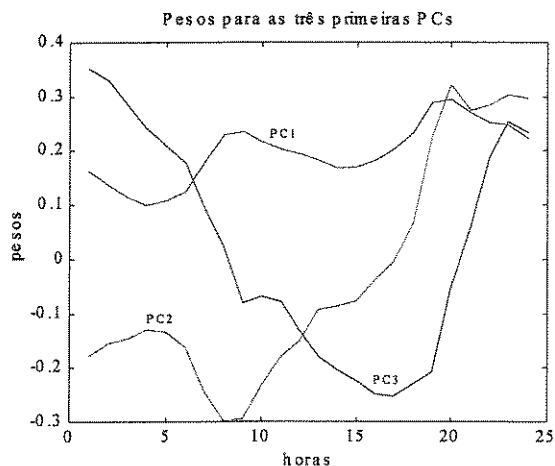
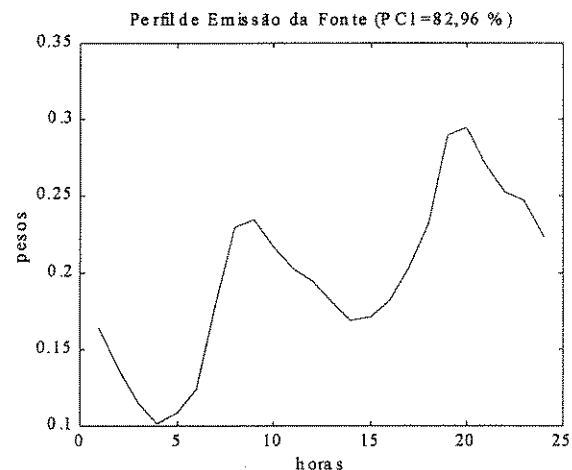


Fig.5.1: a)Gráfico de concentração × dia. b)Gráfico de concentração × hora.

A Figura 5.1 mostra a ocorrência de dados anômalos e de variações súbitas.



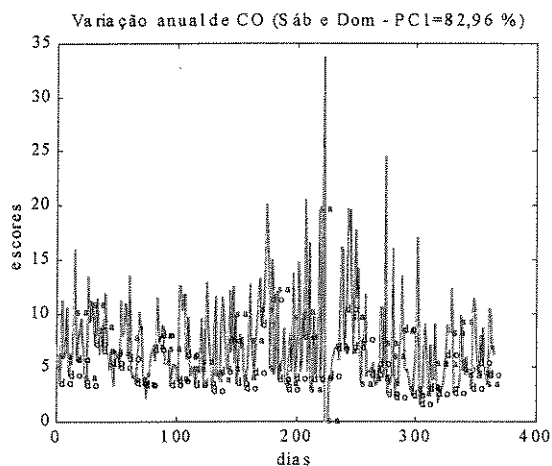
(a)



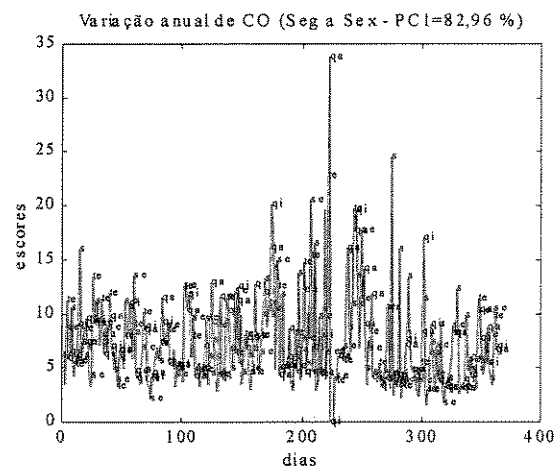
(b)

Fig.5.2: a)Pesos para PC1, PC2 e PC3. b)Pesos para PC1

A Figura 5.2a mostra os pesos para as três primeiras PCs por ordem decrescente de variância capturada (PC1=82,96%, PC2=7,72% e PC3=3,06% - sem correções). A Figura 5.2.b destaca PC1. O perfil encontrado para o monóxido de carbono, CO, é semelhante aos encontrados para os óxidos de nitrogênio, NO e NO₂, como será visto adiante.



(a)



(b)

Fig.5.3: a)Escore para PC1 (finais de semana). b)Escore para PC1 (dias úteis).

A Figura 5.3 mostra que os níveis de emissão de CO caem nos finais de semana.

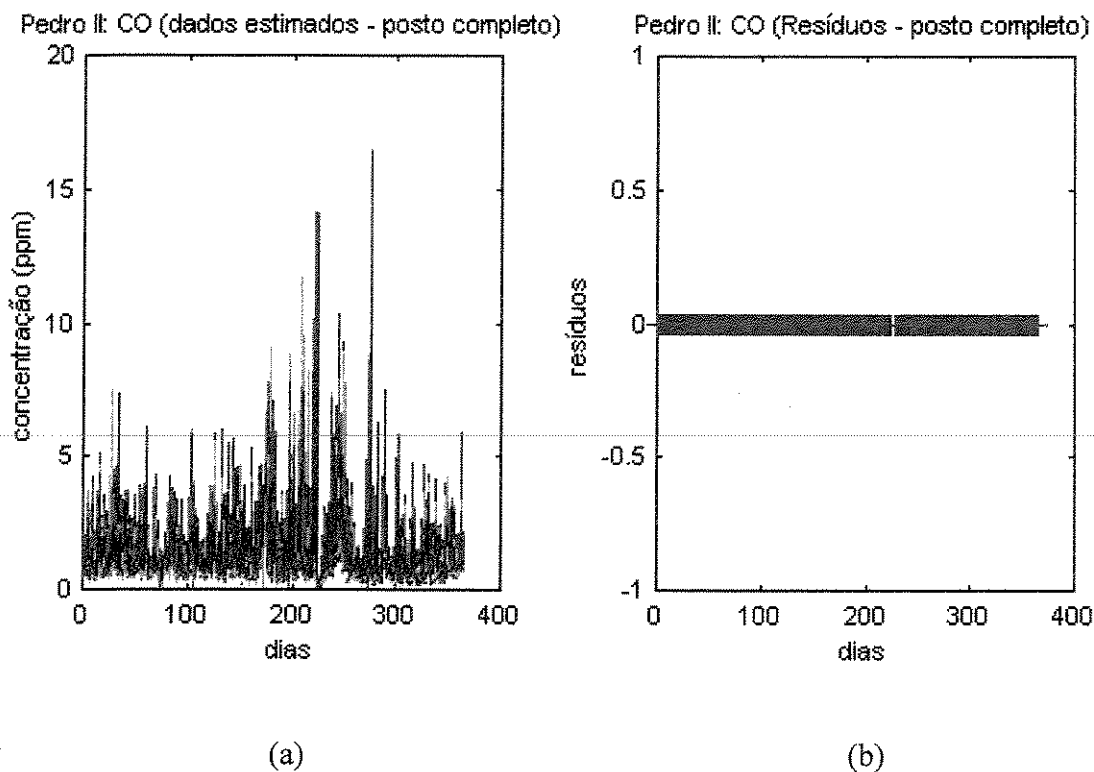


Fig.5.4: a)Dados estimados e b)Resíduos, considerando-se posto completo.

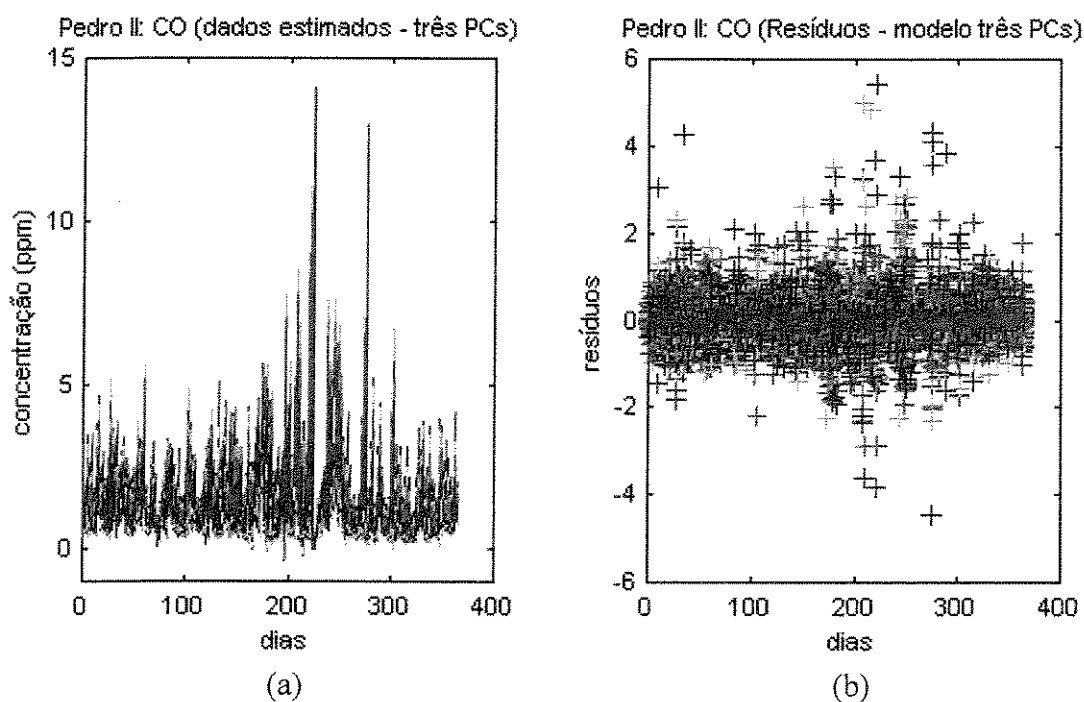


Fig.5.5: a) Dados estimados e b) Resíduos, considerando-se três PCs.

As Figuras 5.4 e 5.5 mostram os dados estimados e os resíduos considerando-se, respectivamente, posto completo e as três primeiras componentes principais.

5.1.1.2 – Análise PCA (P. D. Pedro II – NO: 1999)

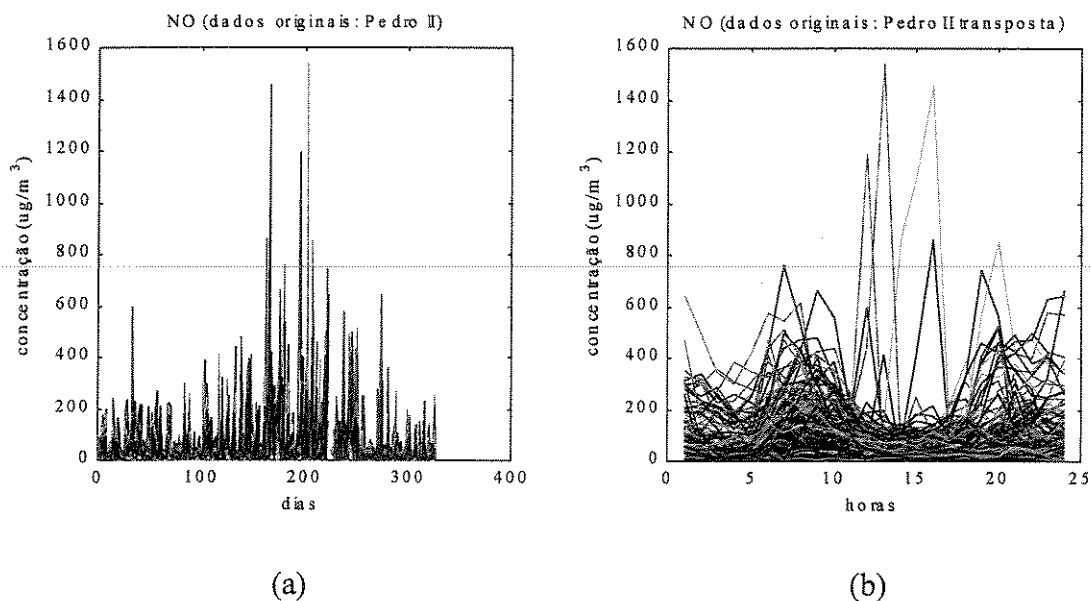


Fig.5.6: a)Gráfico da concentração $\times$ dia. b)Gráfico da concentração $\times$ hora.

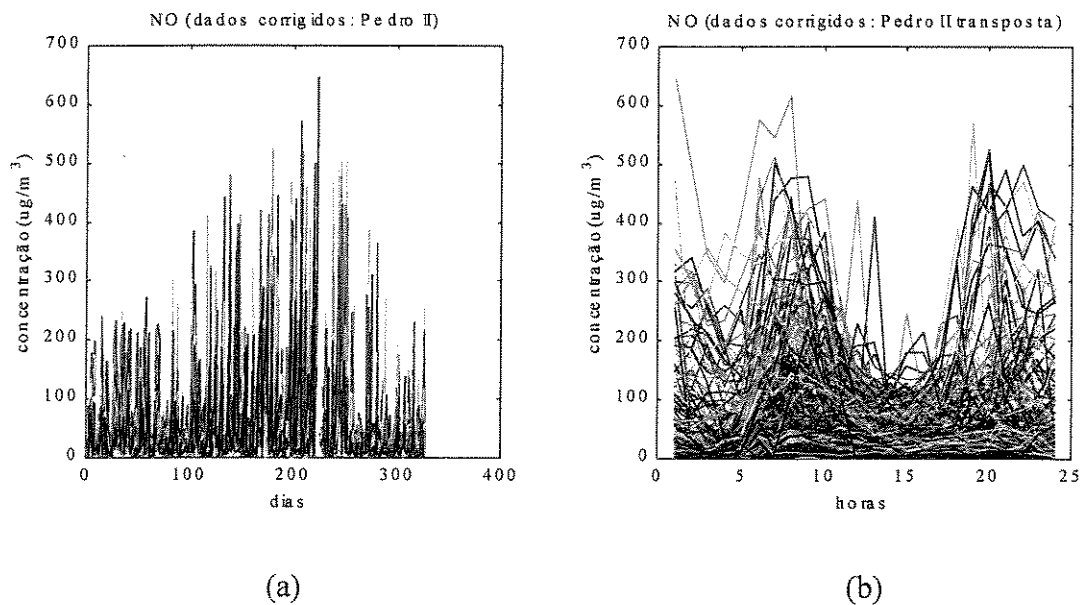
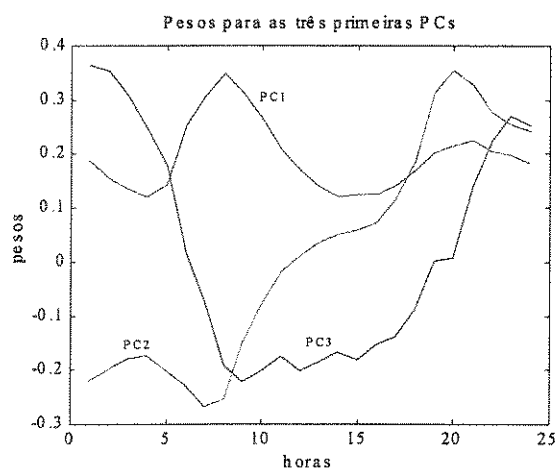
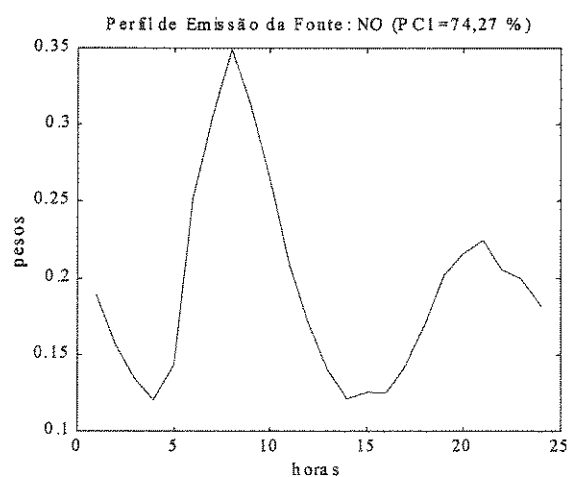


Fig.5.7: a)Gráfico da concentração $\times$ dia. b)Gráfico da concentração $\times$ hora.

As Figuras 5.6 e 5.7 mostram os dados originais sem e com correções.

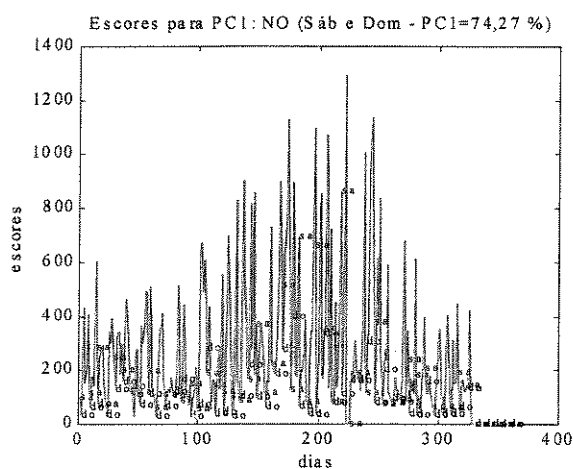


(a)

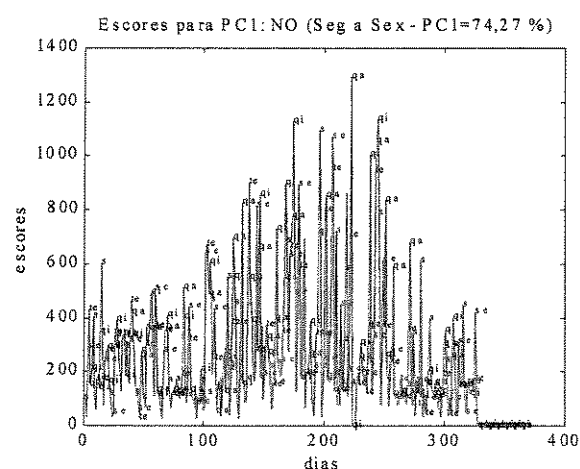


(b)

Fig.5.8: a)Pesos para PC1, PC2 e PC3. b)Pesos para PC1.



(a)



(b)

Fig.5.9: a)Escores para PC1 (finais de semana). b)Escores para PC1 (dias úteis).

A Figura 5.8 mostra os pesos para as três primeiras componentes principais, e a Figura 5.9 mostra os escores para a primeira componente principal considerando-se os finais de semana e os chamados dias úteis. Assim como acontece com o CO, as emissões para o NO também mostram queda nos finais de semana. Essa queda nos finais de semana é melhor observada na análise multimodo (modo B: variação sistemática semanal).

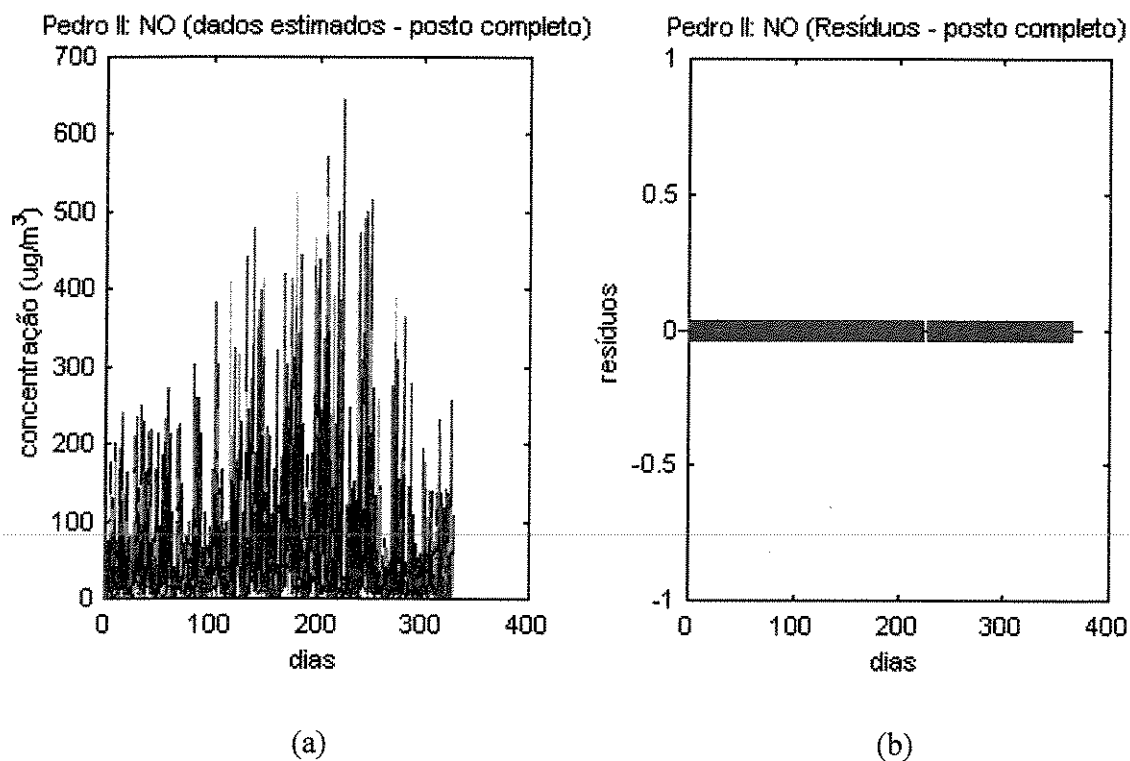


Fig.5.10: a)Dados estimados e b)Resíduos, considerando-se posto completo.

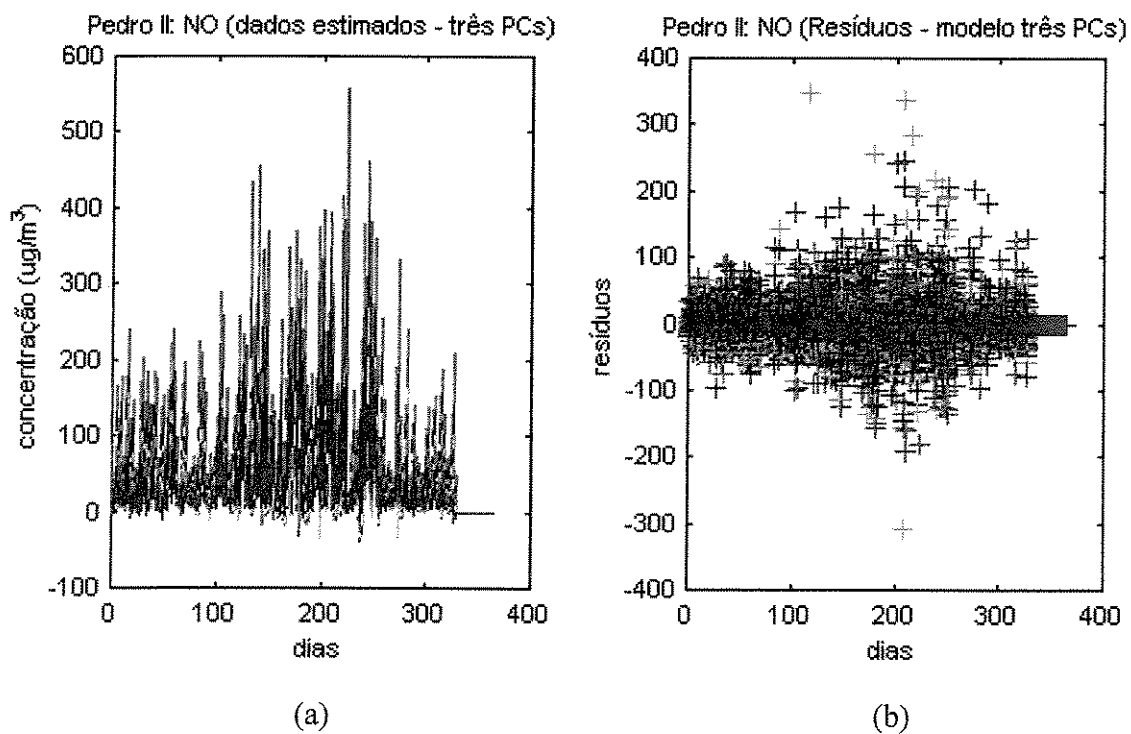


Fig.5.11: a) Dados estimados e b) Resíduos, considerando-se três PCs.

As Figuras 5.10 e 5.11 mostram os dados estimados e os resíduos para o NO.

5.1.1.3 – Análise PCA (P. D. Pedro II – NO₂: 1999)

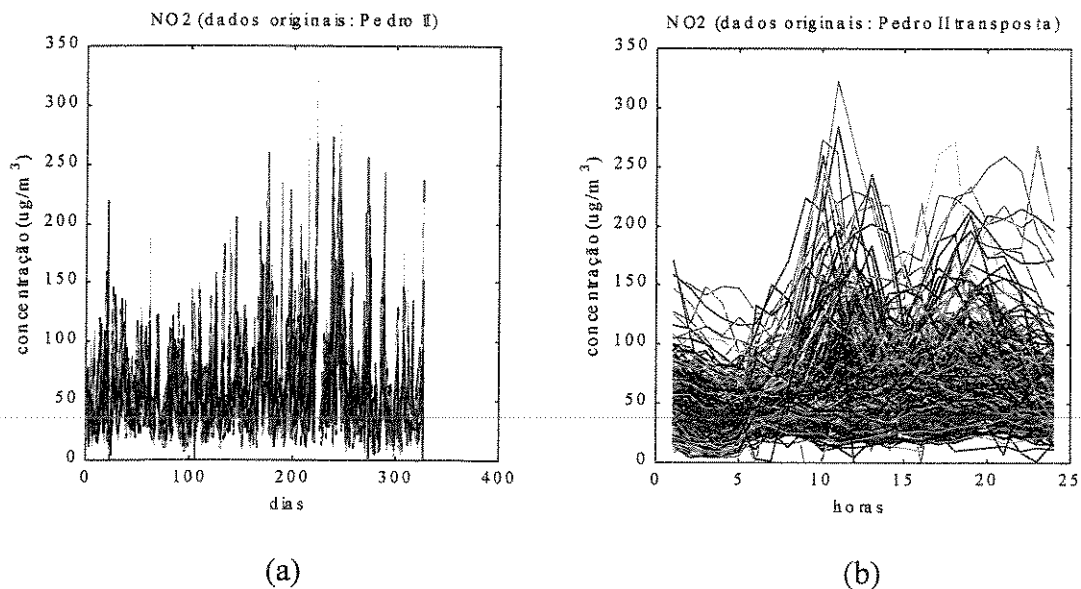


Fig.5.12: a)Gráfico da concentração × dia. b)Gráfico da concentração × hora.

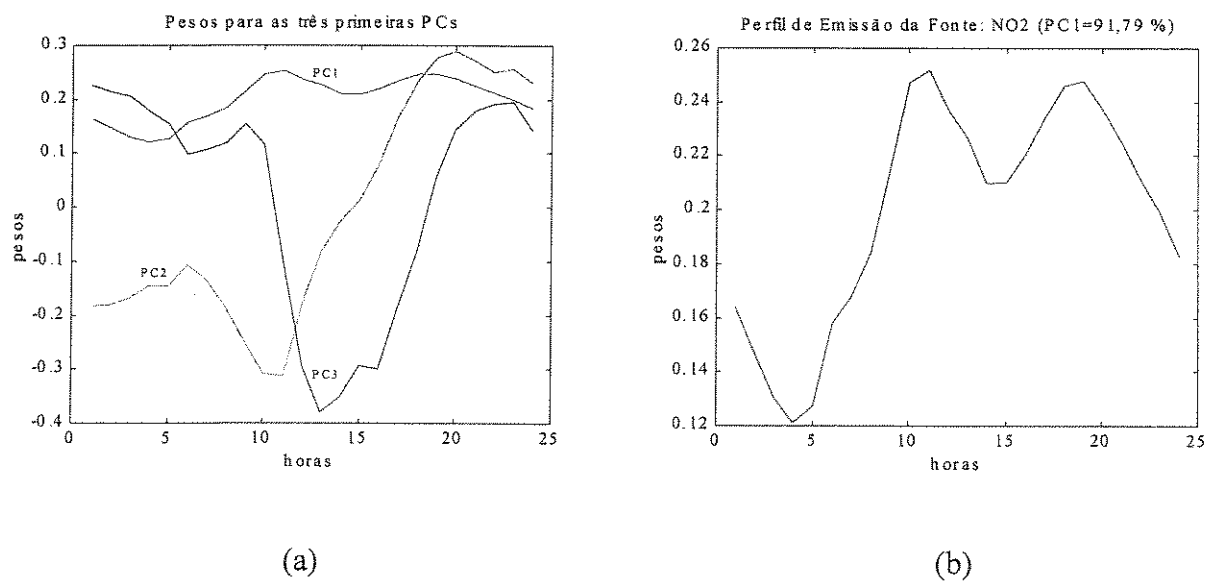
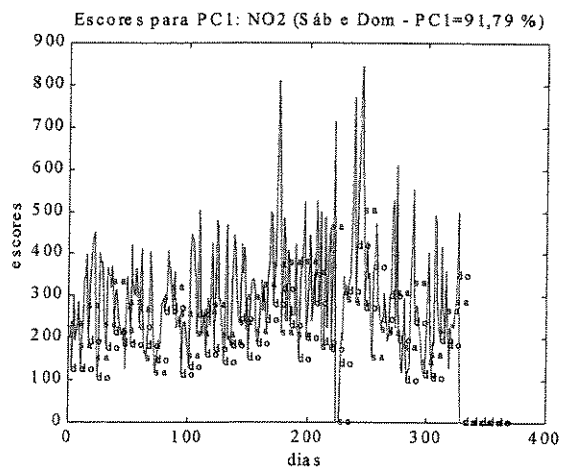
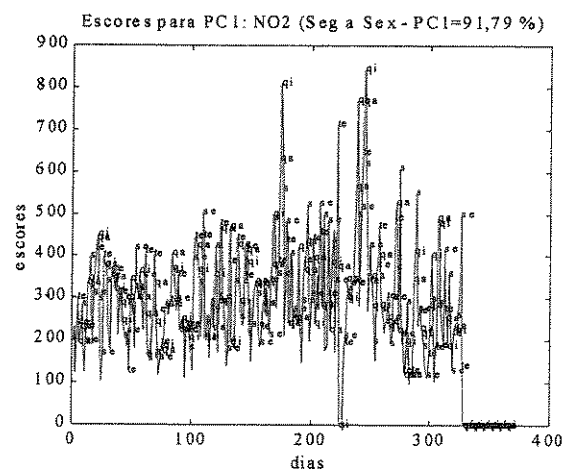


Fig.5.13: a)Pesos para PC1, PC2 e PC3. b)Pesos para PC1

A Figura 5.12 mostra os dados originais para o NO₂, e as Figuras 5.13a e 5.13b mostram, respectivamente, os pesos para as três primeiras PCs e para a primeira PC.



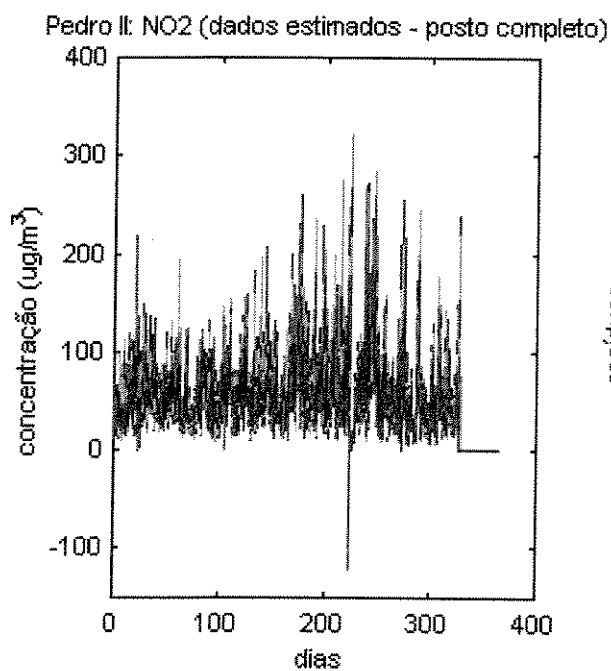
(a)



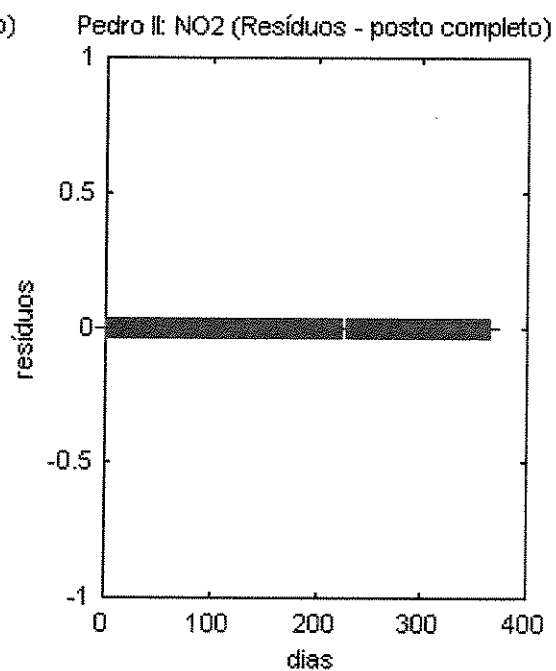
(b)

Fig.5.14: a)Escores para PC1 (finais de semana). b)Escores para PC1 (dias úteis).

As Figuras 5.14a e 5.14b mostram os escores para a primeira componente principal, respectivamente, para os finais de semana e para os dias úteis. Também para o NO_2 ocorre queda das emissões nos fins de semana.



(a)



(b)

Fig.5.15: a)Dados estimados e b)Resíduos, considerando-se posto completo.

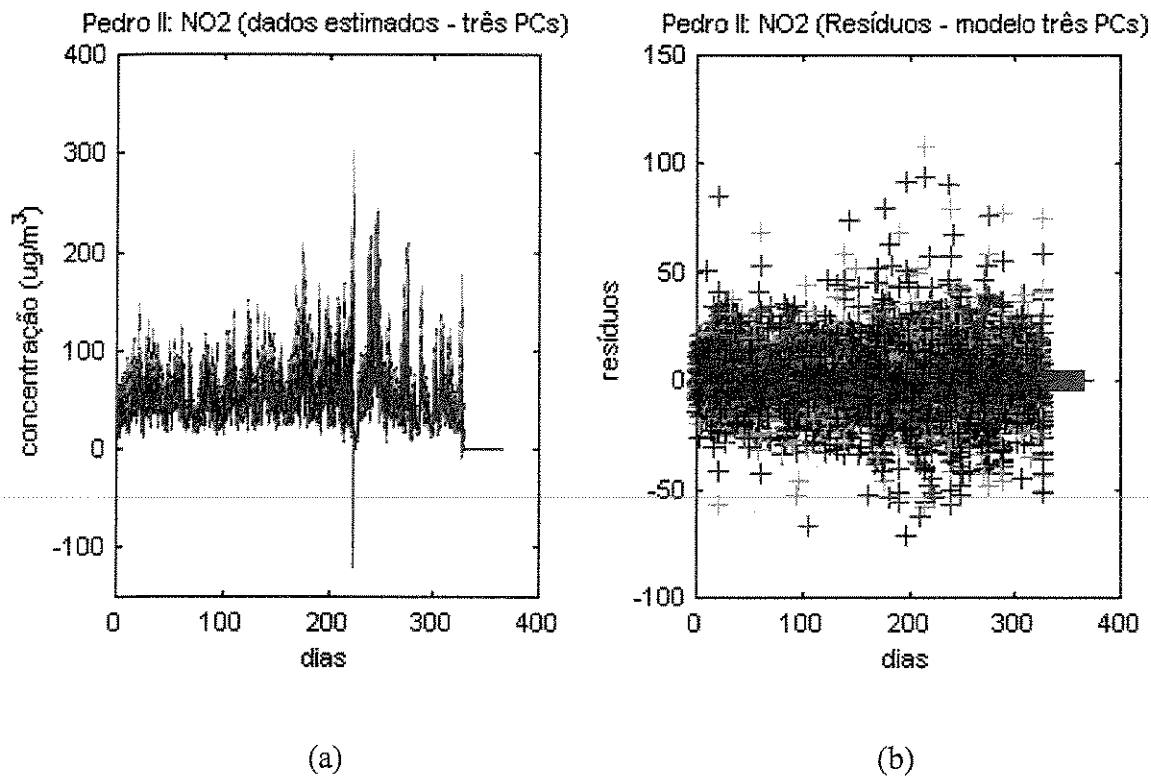


Fig.5.16: a) Dados estimados e b) Resíduos, considerando-se três PCs.

A Figura 5.15 mostra os dados estimados e os resíduos considerando-se posto completo, e a Figura 5.16 reproduz os casos acima porém levando em conta apenas as três primeiras componentes principais.

5.1.1.4 – Análise PCA (P. D. Pedro II – O₃: 1999)

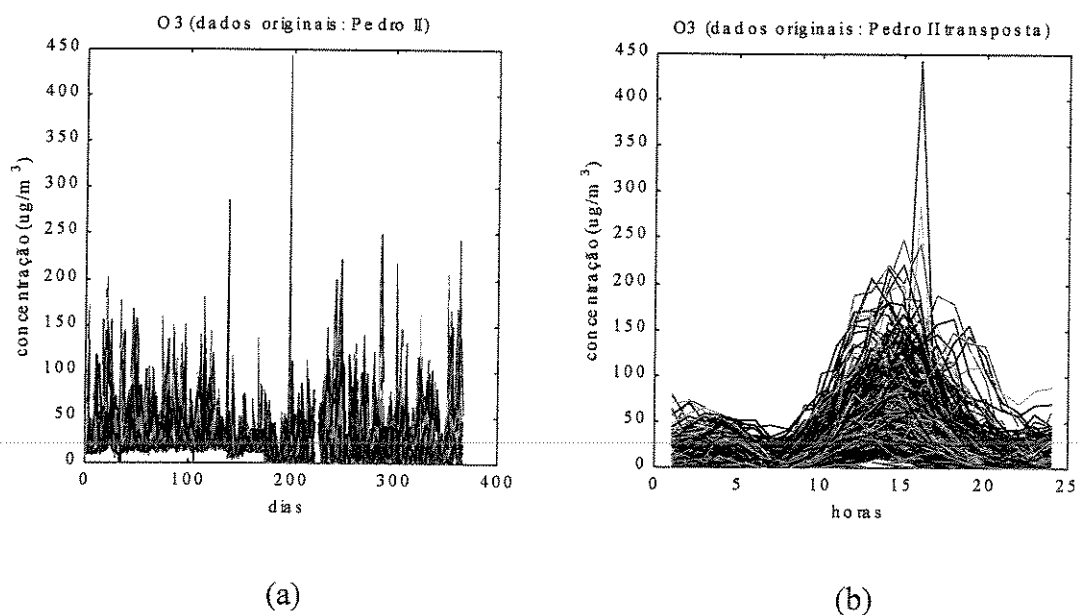


Fig.5.17: a)Gráfico da concentração × dia. b)Gráfico da concentração × hora.

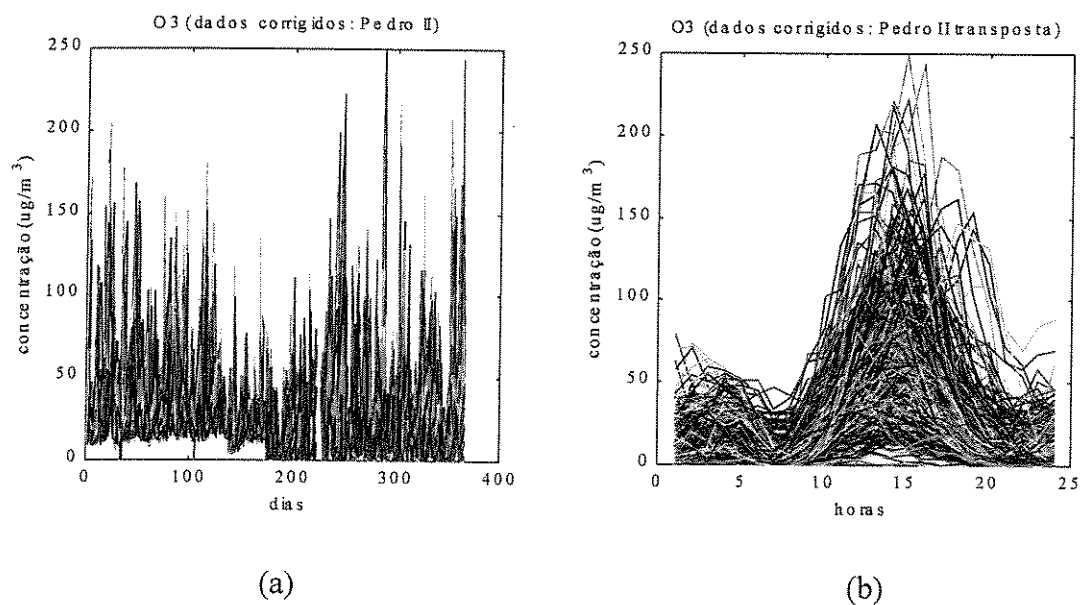
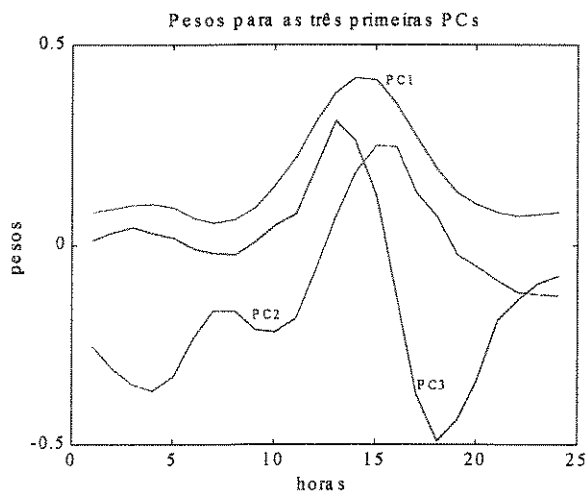
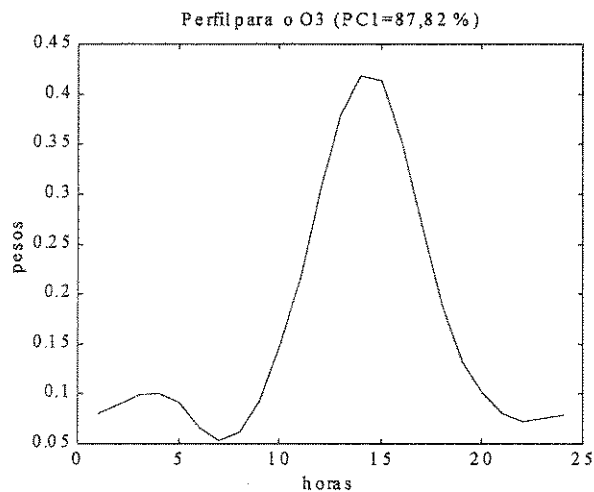


Fig.5.18: a)Gráfico da concentração × dia. b)Gráfico da concentração × hora.

As Figuras 5.17 e 5.18 mostram os dados originais, respectivamente, sem e com correções, para o ozônio. Nota-se claramente a ausência de dados anômalos (“outliers”), os quais são substituídos pela média anual.



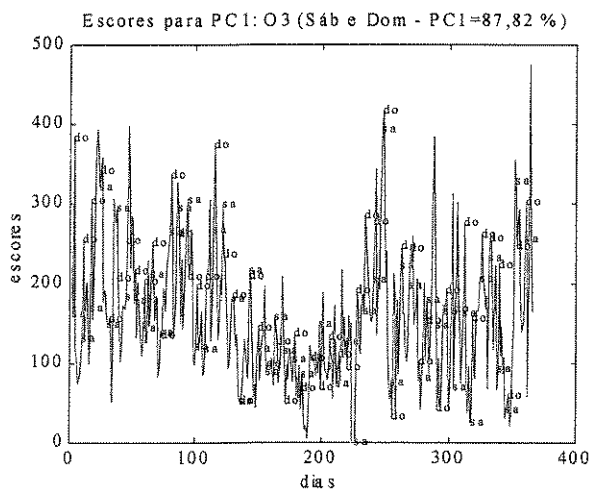
(a)



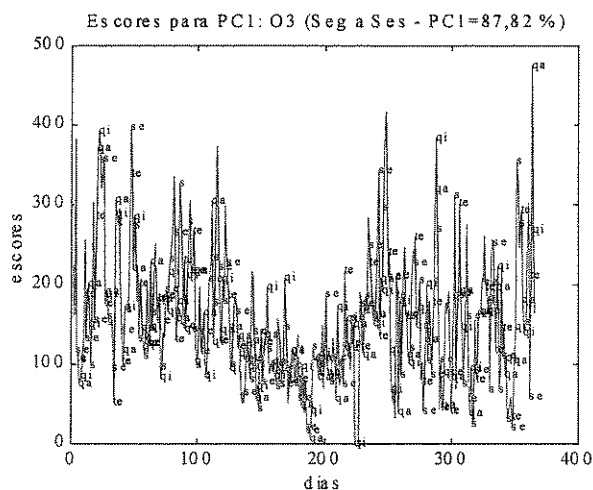
(b)

Fig.5.19: a)Pesos para PC1, PC2 e PC3. b)Pesos para PC1.

As Figuras 5.19a e 5.19b mostram, respectivamente, os pesos para as três primeiras PCs e para a primeira PC.



(a)



(b)

Fig.5.20: a)Escores para PC1 (finais de semana). b)Escores para PC1 (dias úteis).

As Figuras 5.20a e 5.20b mostram os escores para a primeira PC, respectivamente, para os finais de semana e dias úteis. É importante observar que, ao contrário do que ocorre com o CO, NO e NO₂, os valores mais elevados para o O₃ ocorrem nos finais de semana.

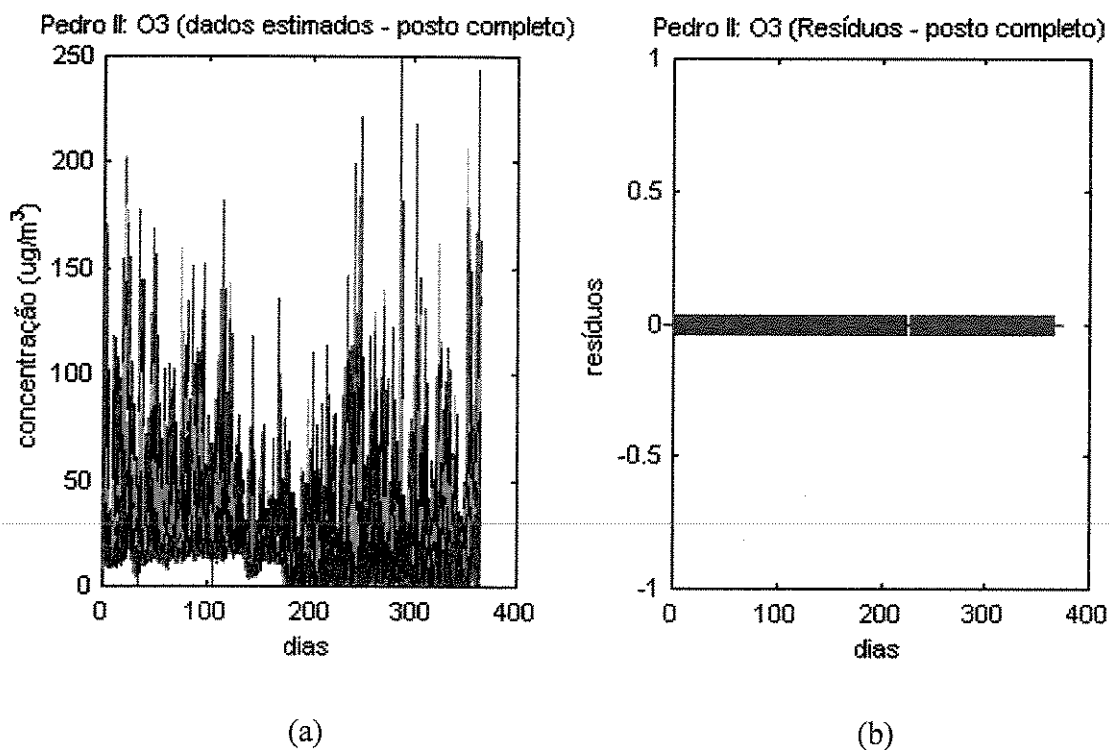


Fig.5.21: a)Dados estimados e b)Resíduos, considerando-se posto completo.

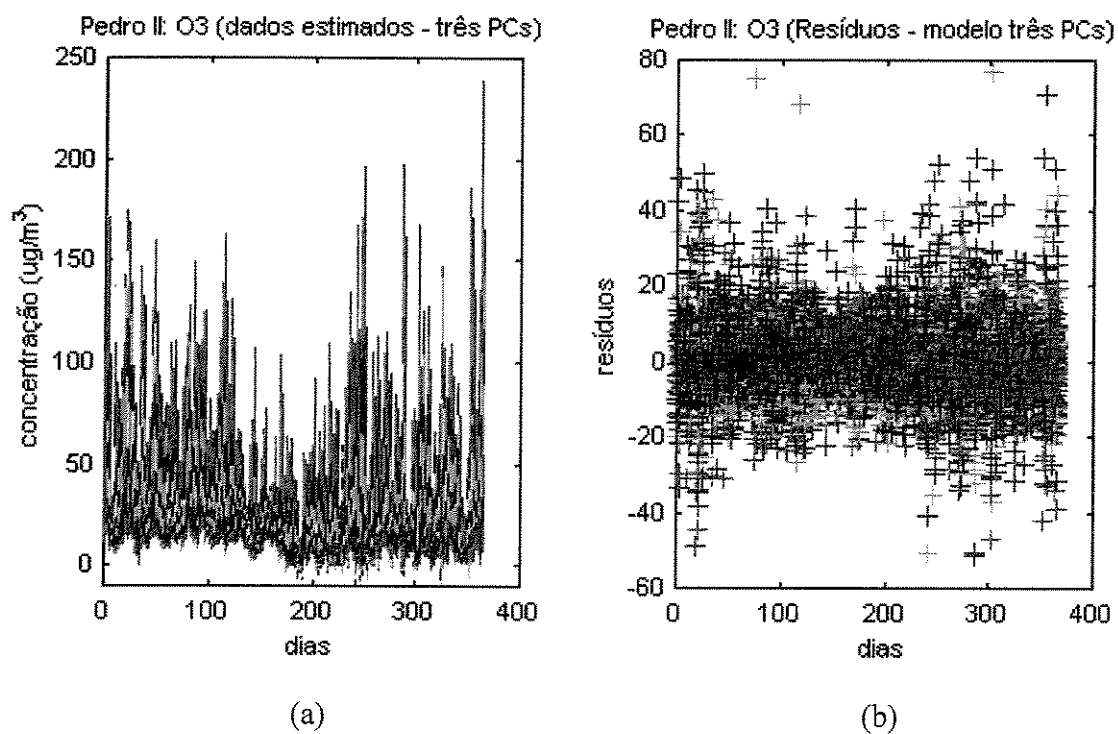


Fig.5.22: a) Dados estimados e b) Resíduos, considerando-se três PCs.

As Figuras 5.21 e 5.22 mostram os dados estimados e os resíduos.

5.1.1.5 – Análise PCA (P. D. Pedro II – SO₂: 1999)

(Duas Fontes)

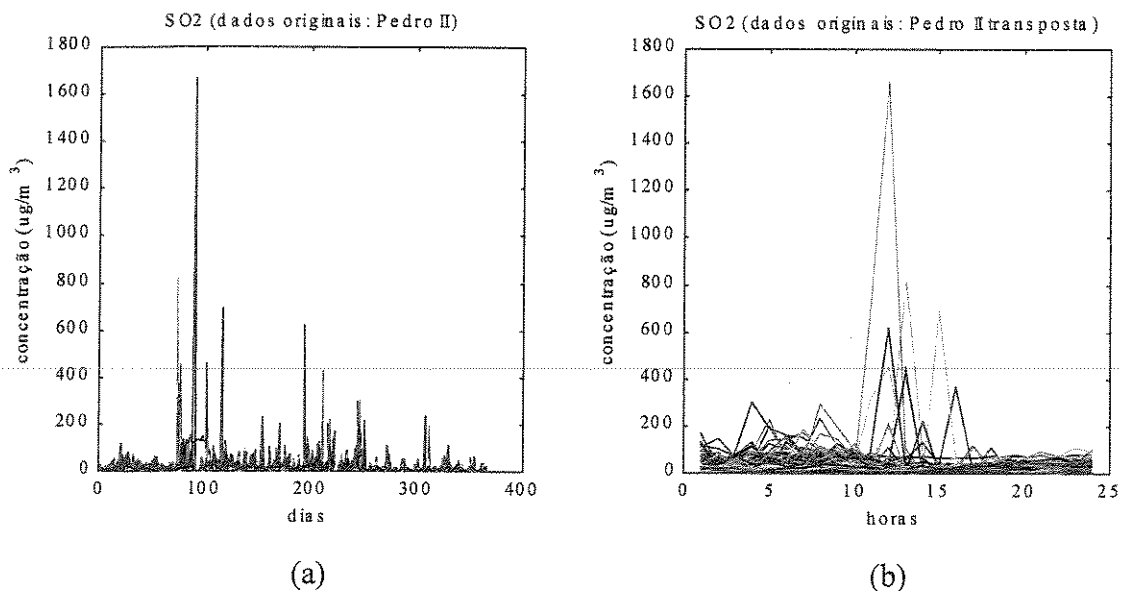


Fig.5.23: a) Gráfico concentração × dia. b) Gráfico concentração × hora.

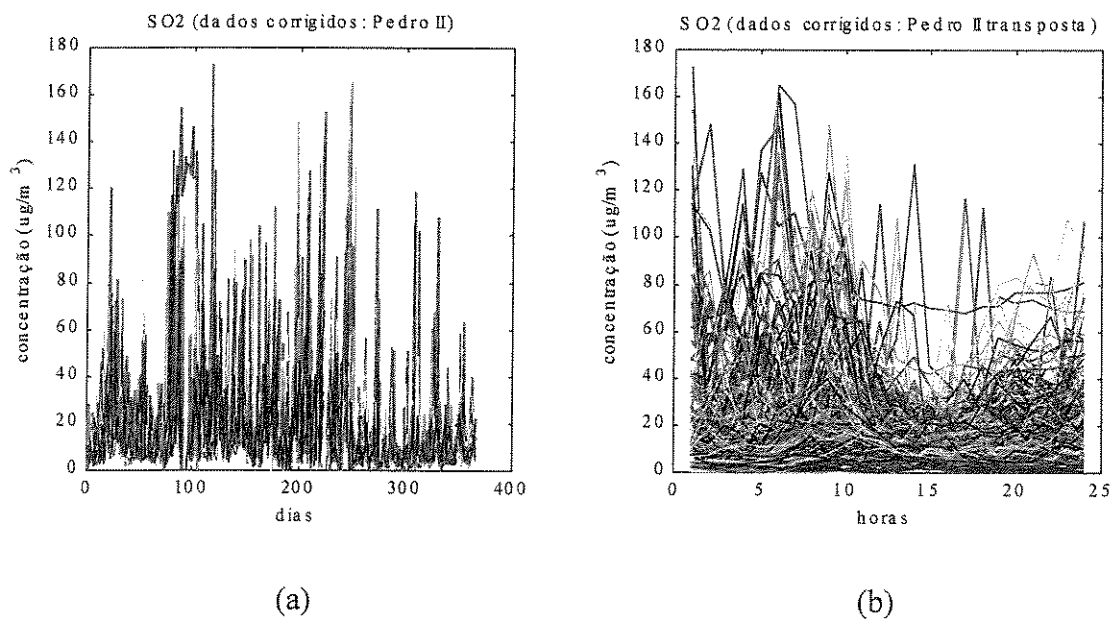
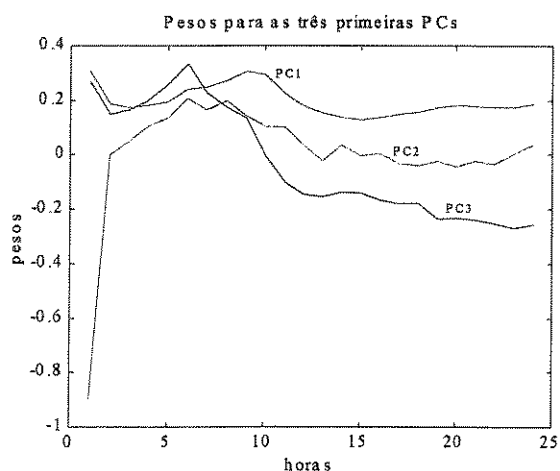
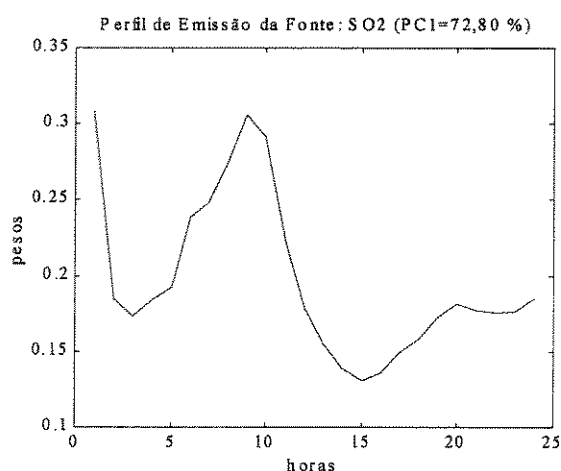


Fig.5.24: a) Gráfico concentração × dia. b) Gráfico concentração × hora.

As Figuras 5.23 e 5.24 mostram os dados originais, respectivamente, sem e com correções para o caso de duas fontes de emissão de SO₂ (discussão no Capítulo VI).

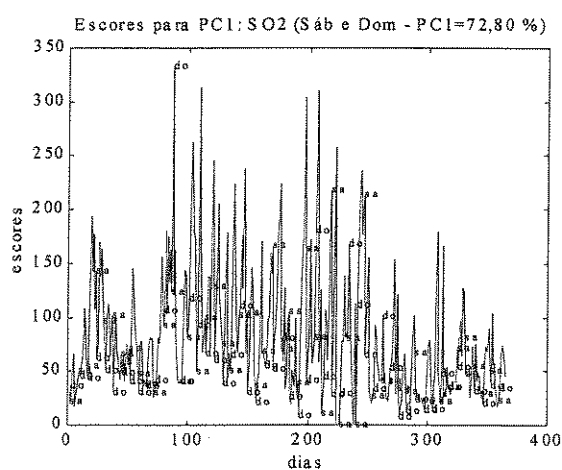


(a)

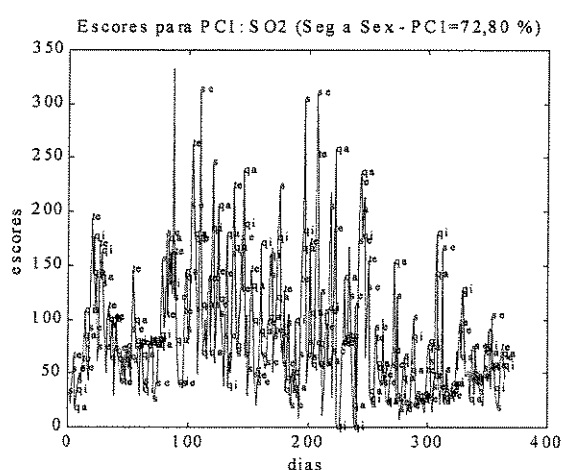


(b)

Fig.5.25: a)Pesos para PC1, PC2 e PC3. b)Pesos para PC1.



(a)



(b)

Fig.5.26: a)Escores para PC1 (finais de semana). b)Escores para PC1 (dias úteis).

As figuras 5.25a e 5.25b mostram, respectivamente, os pesos para as três primeiras PCs e para a primeira PC. Nota-se, em 5.25b, valores elevados à 1:00 h. (discussão sobre essas emissões no Capítulo VI). As Figuras 5.26a e 5.26b mostram, respectivamente, os escores para a primeira PC. Também para o poluente SO_2 , as emissões mostram valores mais baixos nos finais de semana, quando comparados com aqueles registrados nos chamados dias úteis.

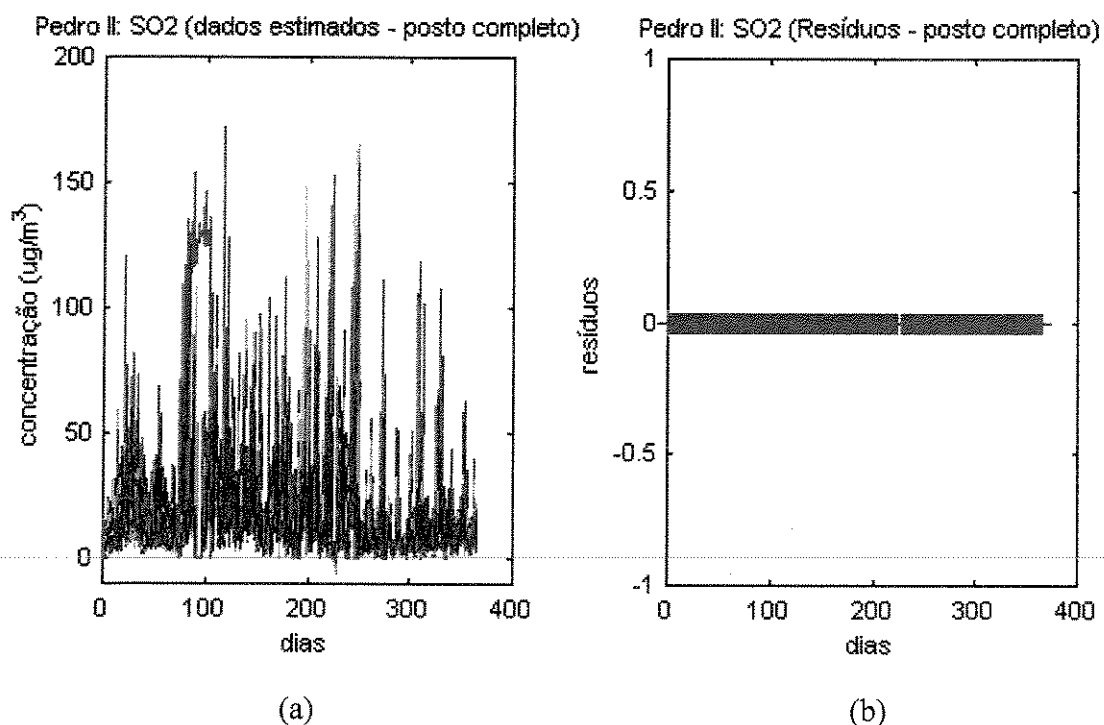


Fig.5.27: a)Dados estimados e b)Resíduos, considerando-se posto completo.

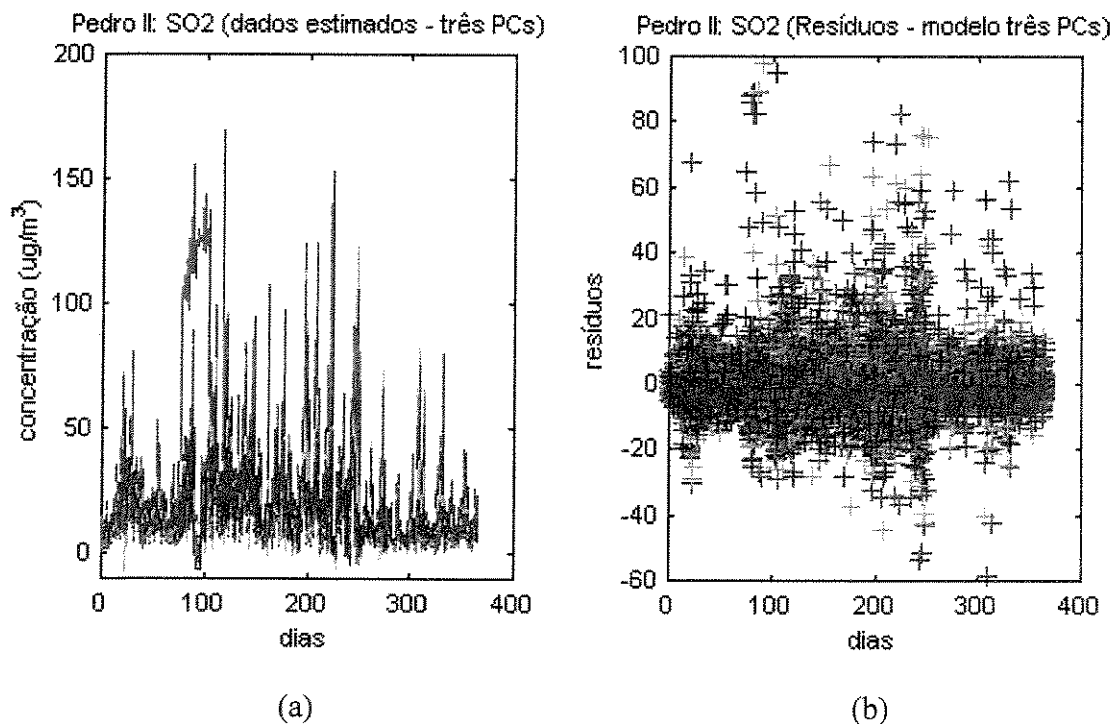


Fig.5.28: a) Dados estimados e b) Resíduos, considerando-se três PCs.

As Figuras 5.27 e 5.28 mostram os dados estimados e os resíduos (duas fontes).

(Uma Fonte)

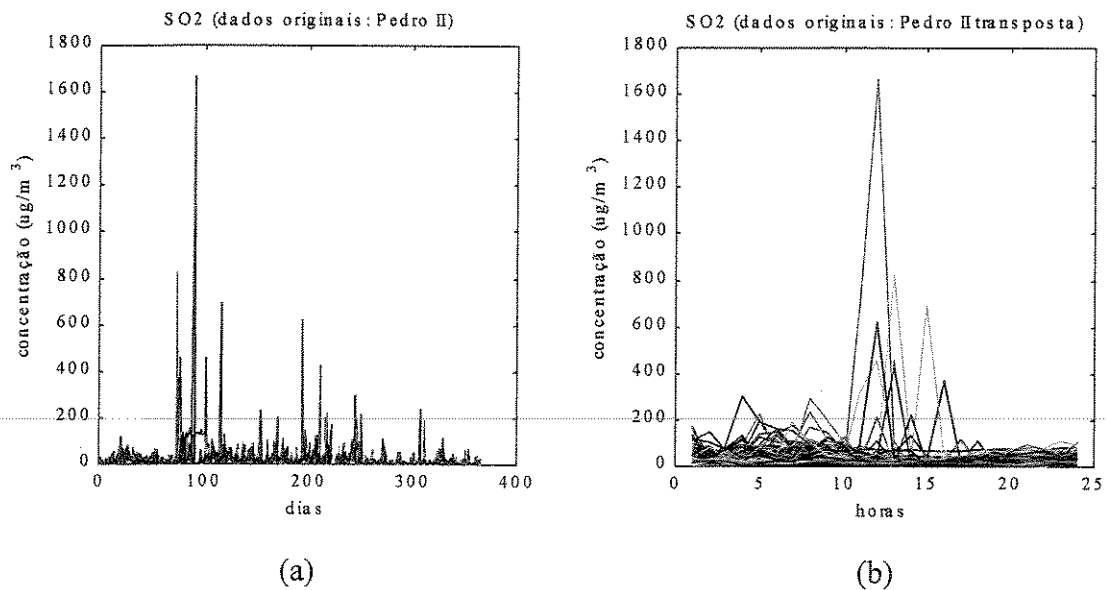


Fig.5.29: a)Gráfico concentração $\times$ dia. b) Gráfico concentração $\times$ hora.

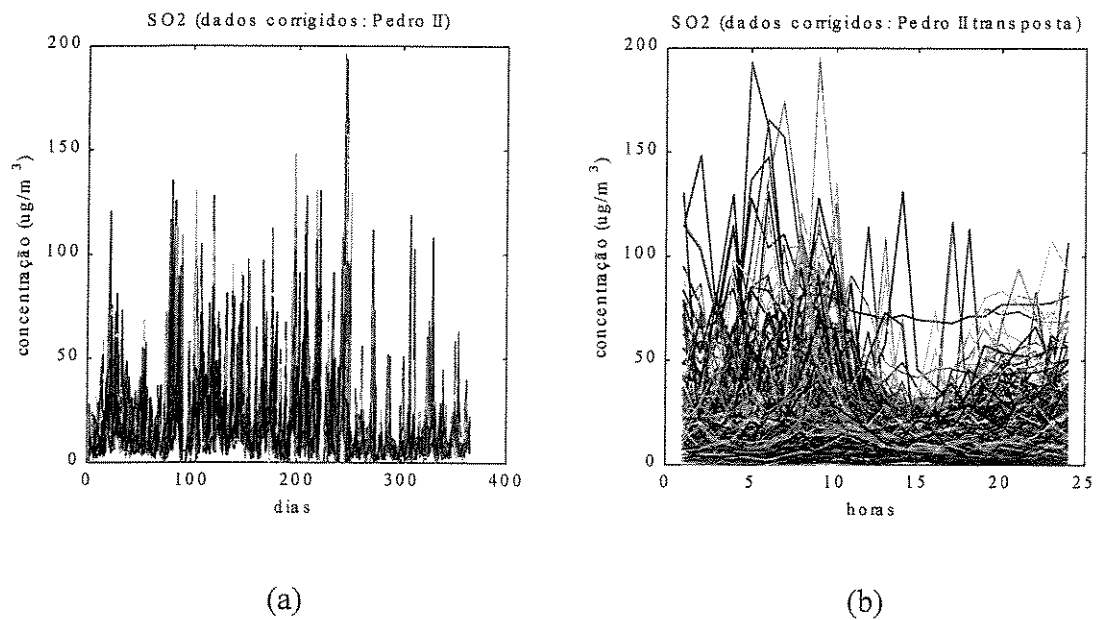
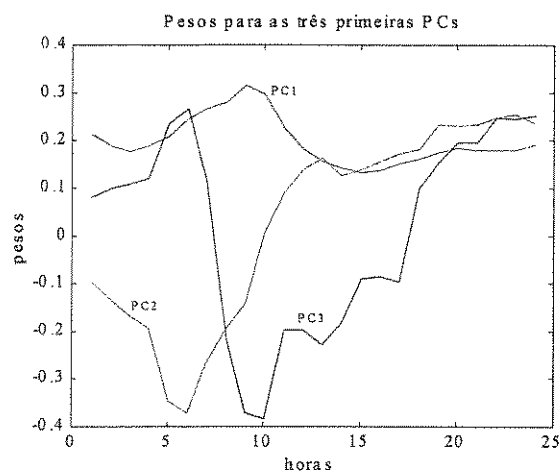
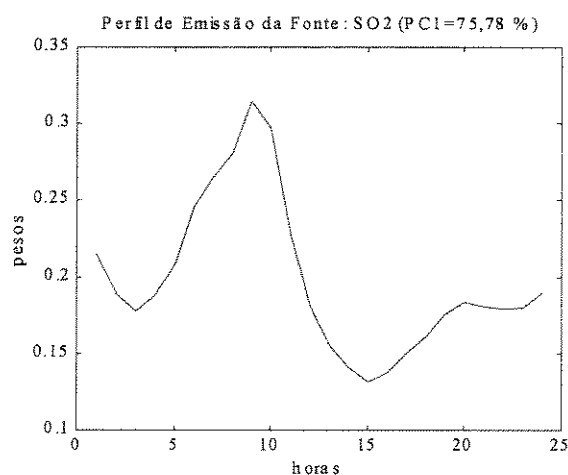


Fig.5.30: a)Gráfico concentração $\times$ dia. b) Gráfico concentração $\times$ hora.

As Figuras 5.29 e 5.30 mostram os dados originais, respectivamente, sem e com correções, especialmente para os dados coletados à 1:00 h (comparar Figuras 5.24 e 5.30), os quais exercem influência nos resultados (discussão no Capítulo VI).

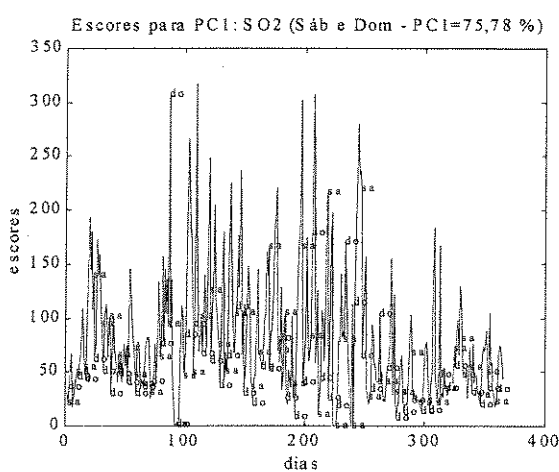


(a)

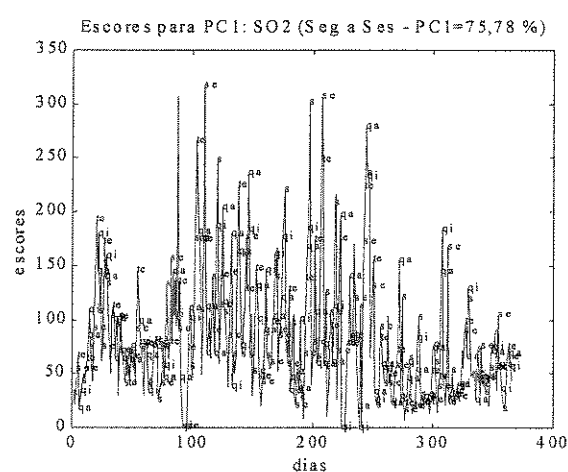


(b)

Fig.5.31: a)Pesos para PC1, PC2 e PC3. b)Pesos para PC1.



(a)



(b)

Fig.5.32: a)Escores para PC1 (finais de semana). b)Escores para PC1 (dias úteis).

As figuras 5.31a e 5.31b mostram os pesos, respectivamente, para as três primeiras PCs e para a primeira PC. É importante comparar as Figuras 5.25 e 5.31. Porém, quando são comparados os escores das Figuras 5.26 e 5.32, as pequenas diferenças não são tão evidentes. De fato, de acordo com as bases da metodologia e sua construção nessa tese, o que confere importância aos dados são os pesos como discutido no Capítulo III (ver discussão sobre os resultados para o SO_2 no Capítulo VI).

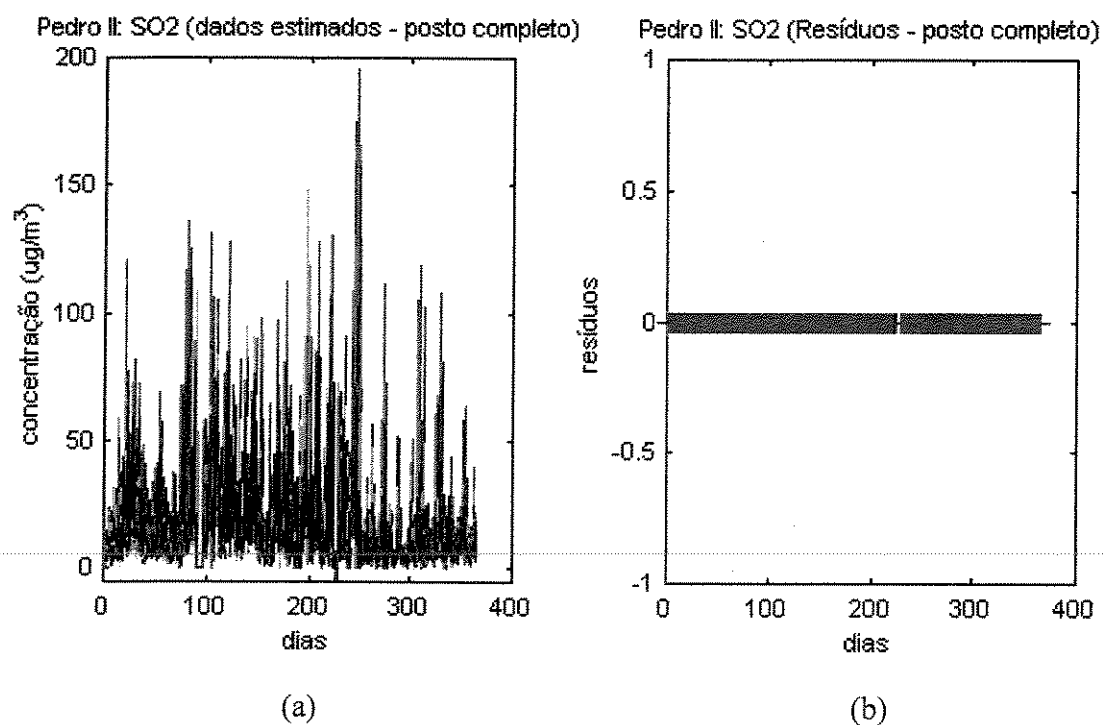


Fig.5.33: a)Dados estimados e b)Resíduos, considerando-se posto completo.

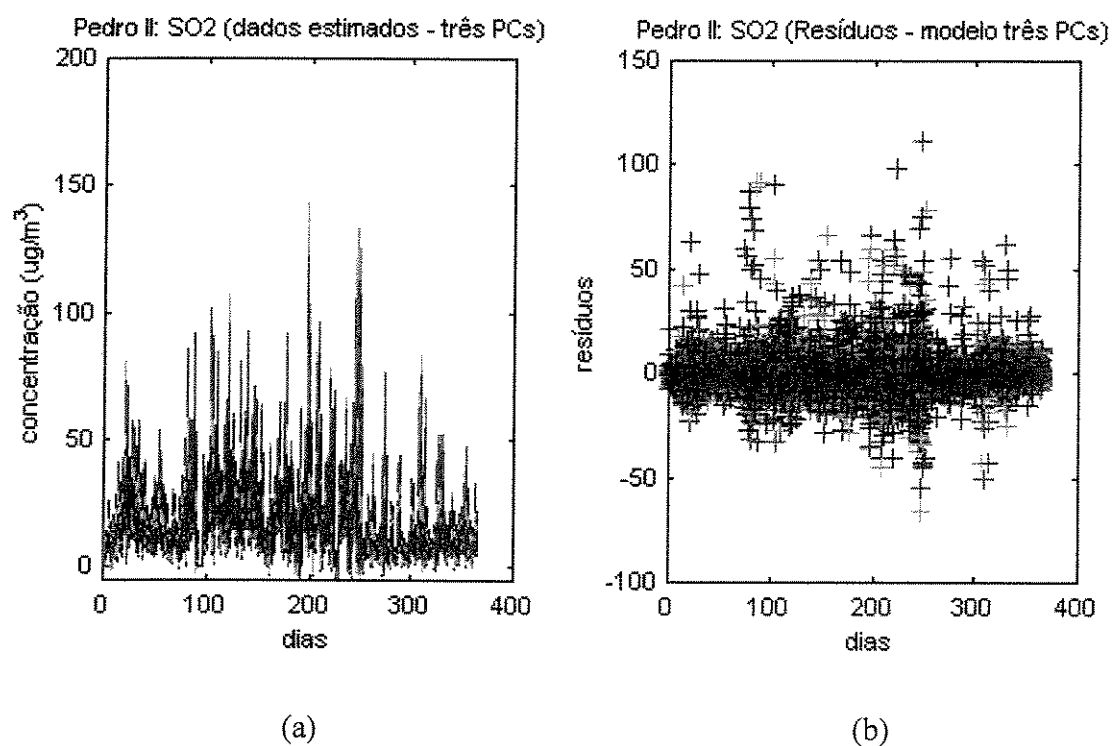


Fig.5.34: a) Dados estimados e b) Resíduos, considerando-se três PCs.

As Figuras 5.33 e 5.34 mostram os dados estimados e os resíduos (uma fonte).

5.1.1.6 – Análise PCA (P. D. Pedro II – CH₄: 1999)

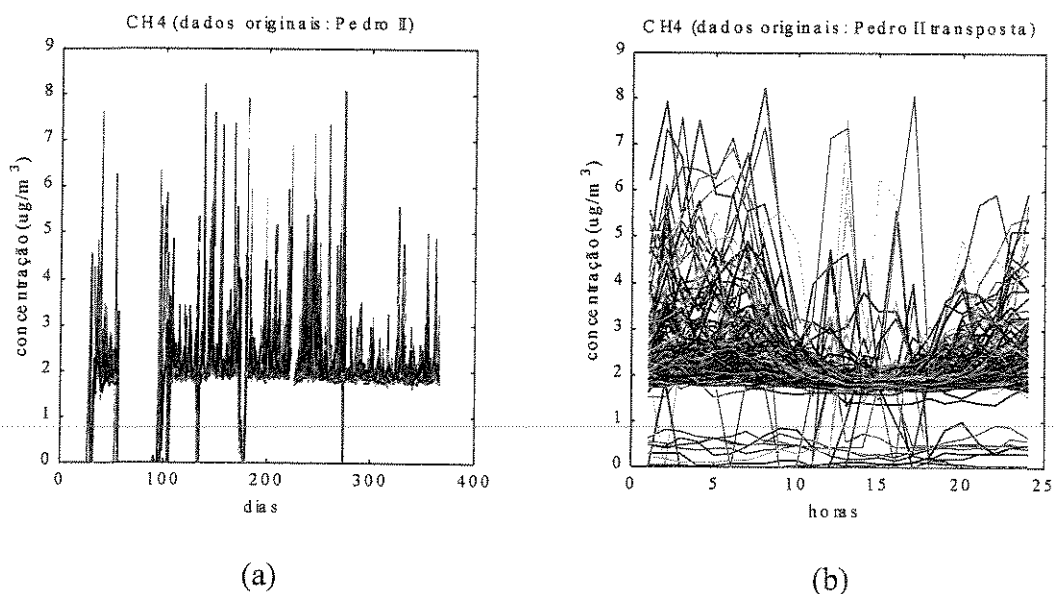


Fig.5.35: a) Gráfico concentração × dia. b) Gráfico concentração × hora.

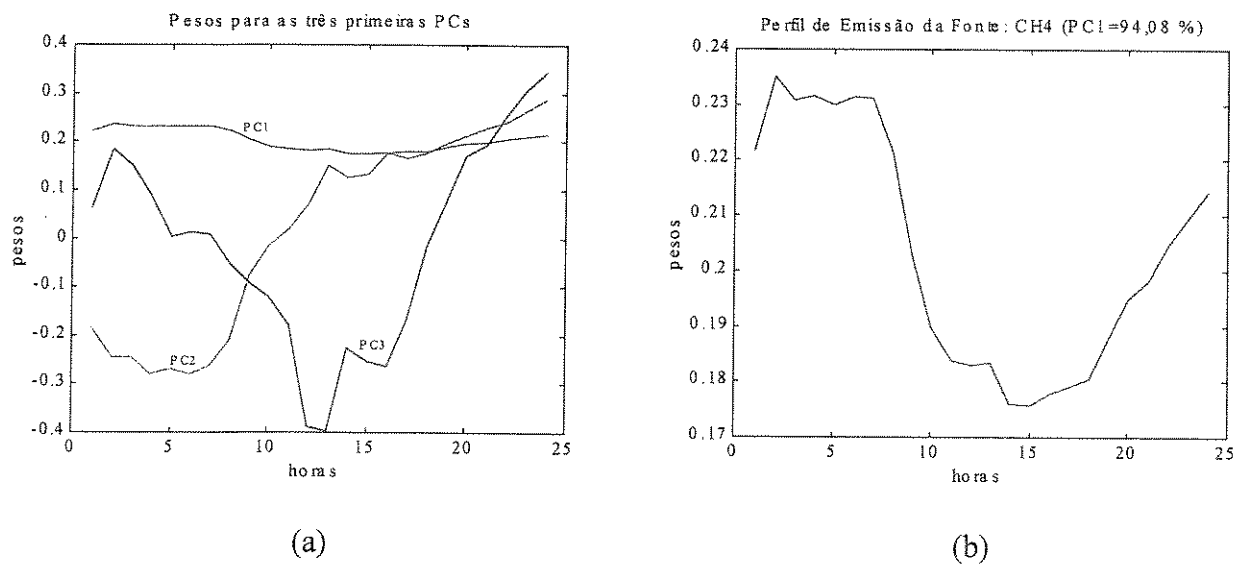
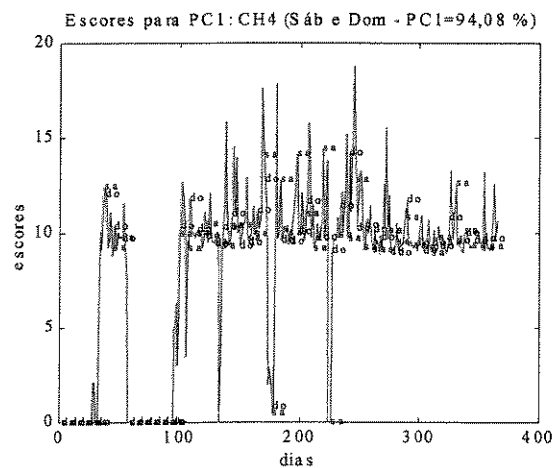
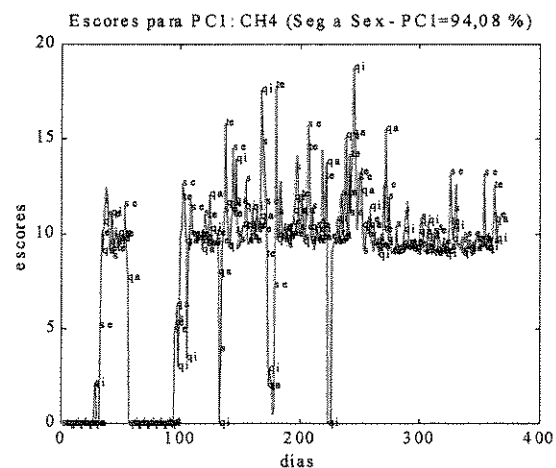


Fig.5.36: a) Pesos para PC1, PC2 e PC3. b) Pesos para PC1.

A Figura 5.35 mostra os dados originais. A Figura 5.36 mostra os pesos para o CH₄. Pode ser observado que mesmo sem correções PCA consegue capturar as variações sistemáticas diárias para o poluente. A Figura 5.36b mostra um perfil diferente daquele obtido para as emissões de CO, NO, NO₂ e SO₂, e também distinto do perfil para o O₃.

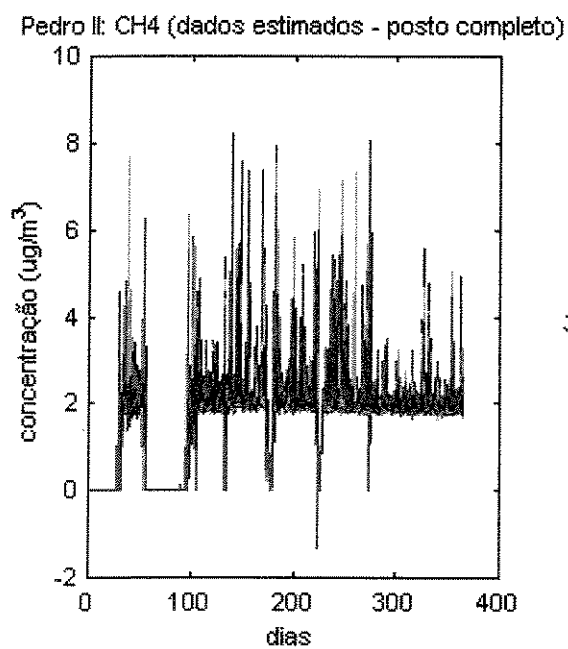


(a)

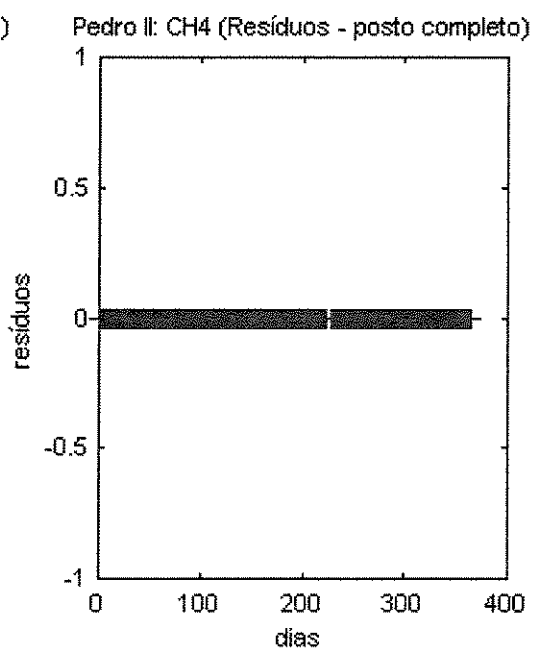


(b)

Fig.5.37: a)Escore para PC1 (finais de semana). b)Escore para PC1 (dias úteis).



(a)



(b)

Fig.5.38: a)Dados estimados e b)Resíduos, considerando-se posto completo.

A Figura 5.37 mostra os escores para PC1 e a Figura 5.38 mostra os dados estimados e os resíduos.

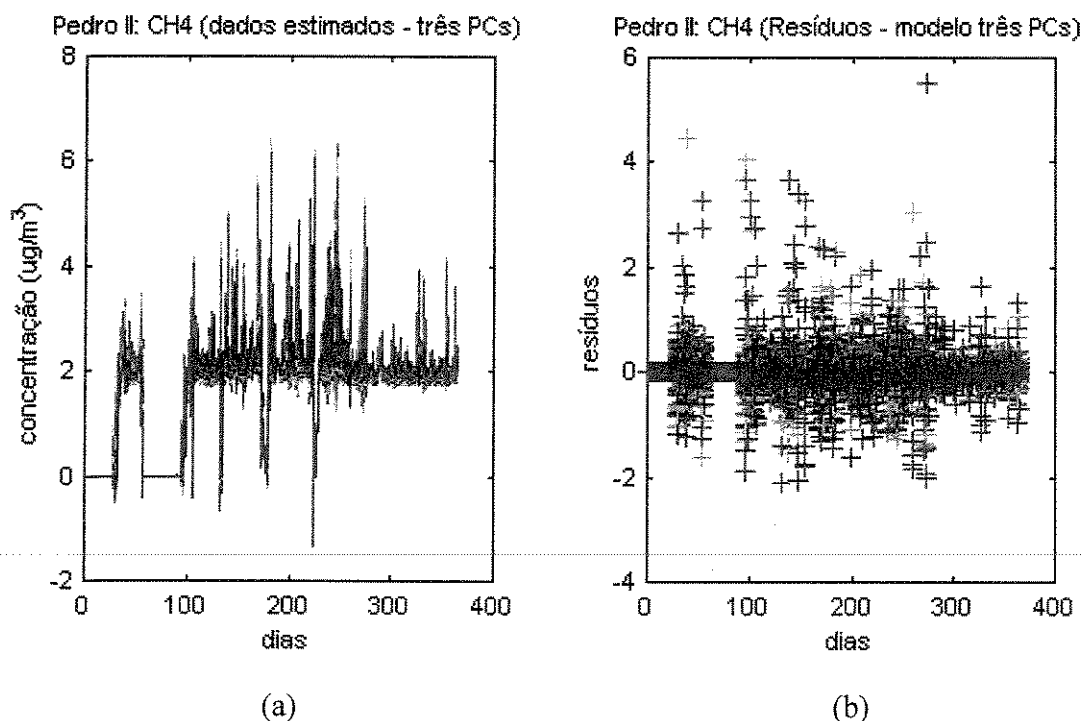


Fig.5.39: a) Dados estimados e b) Resíduos, considerando-se três PCs.

A Figura 5.39 mostra os dados estimados e os resíduos para o metano.

Deve ser observado que os valores iguais a zero para os escores na Figura 5.37 são originalmente nulos como mostra a Figura 5.35. Essas situações surgem para outras variáveis. Assim, é interessante ter em mente que nem sempre valores nulos encontrados para os escores após a aplicação dos modelos aos dados, são oriundos da análise ela própria. Os valores originalmente nulos, assim permanecem mesmo após as análises. Isto pode ser confirmado quando se observa o gráfico para os dados estimados, considerando-se posto completo (ver Figura 5.38)

5.1.1.7 – Análise PCA (P. D. Pedro II – NMHC: 1999)

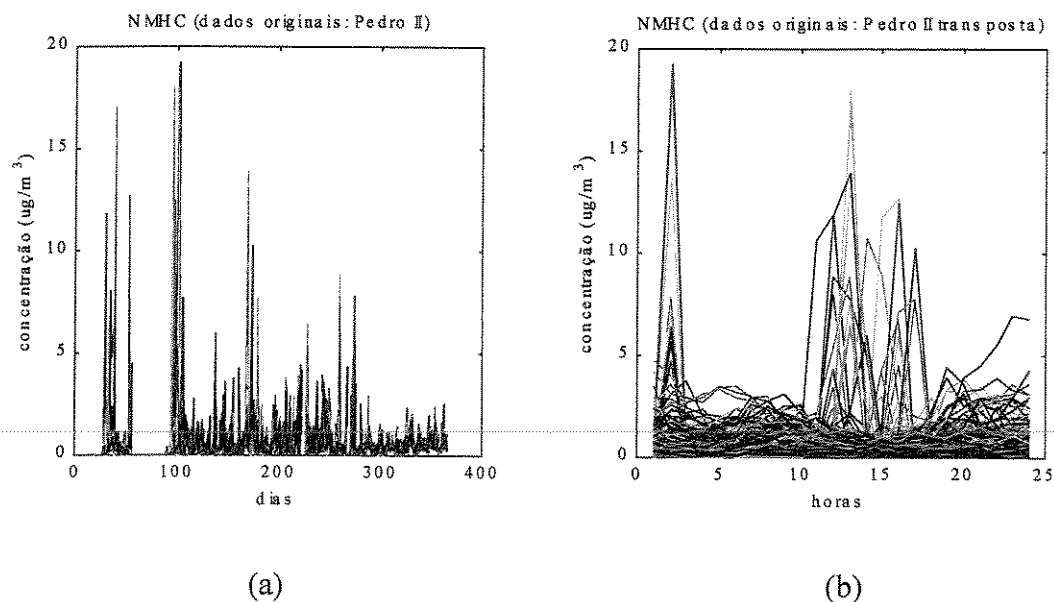


Fig.5.40: a) Gráfico concentração × dia. b) Gráfico concentração × hora.

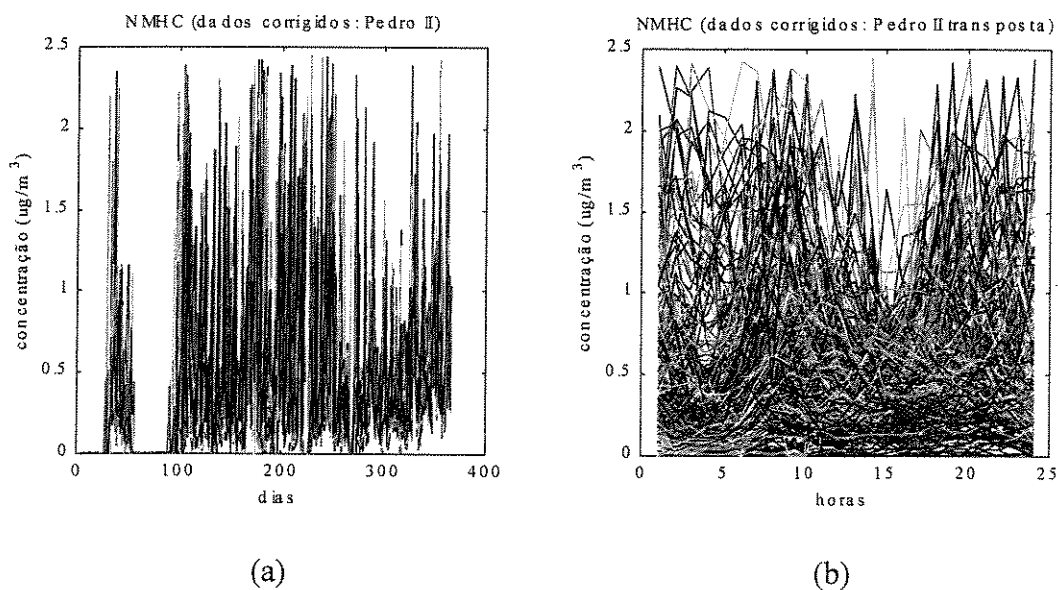
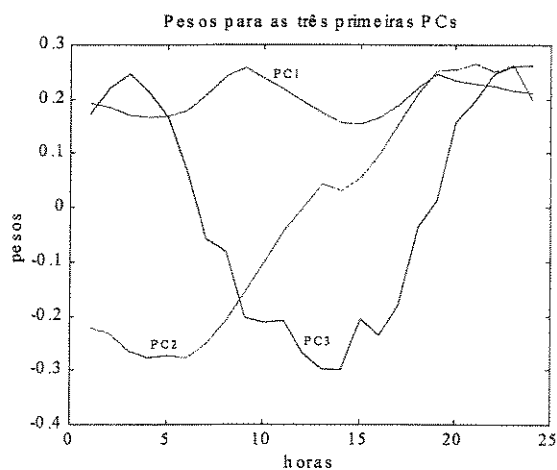
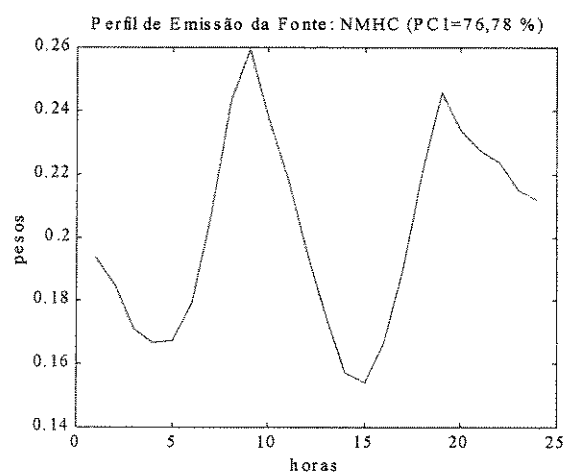


Fig.5.41: a) Gráfico concentração × dia. b) Gráfico concentração × hora.

As Figuras 5.40 e 5.41 mostram os dados originais, respectivamente, sem e com correções. Não obstante a quantidade de correções sobre os dados anômalos e variações súbitas, PC1 captura cerca de 77% da variância (Figura 5.42).

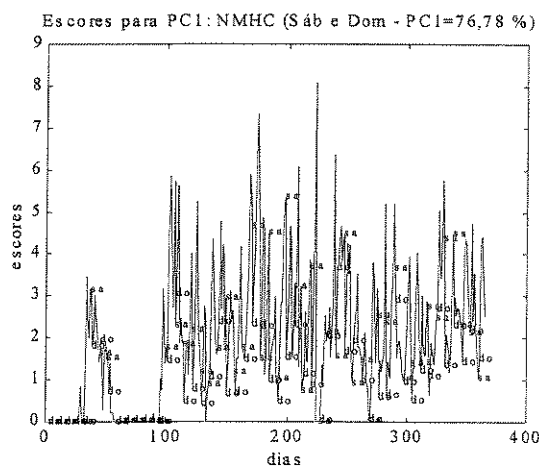


(a)

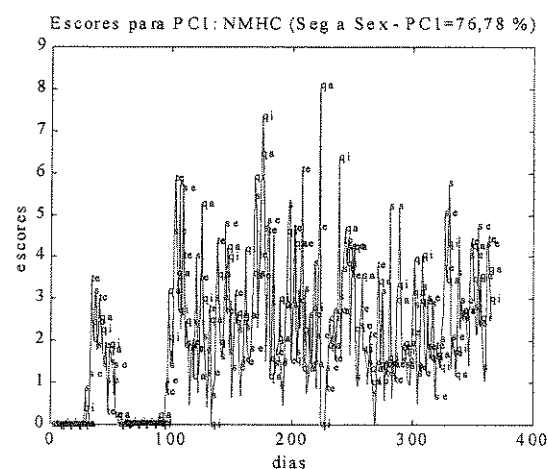


(b)

Fig.5.42: a)Pesos para PC1, PC2 e PC3. b)Pesos para PC1.



(a)



(b)

Fig.5.43: a)Escore para PC1 (finais de semana). b)Escore para PC1 (dias úteis).

A Figura 5.42 mostra os pesos para as PCs. PC1 captura 76,78% da variância (PC2=8,97% e PC3=3,40%). Nota-se a semelhança do perfil para os NMHC com aqueles para o CO, NO, NO₂ e SO₂. Por outro lado, há distinção em relação ao CH₄ (e ao O₃, o qual não é emitido pelas fontes em potencial, mas sim formado por meio de reações que ocorrem na atmosfera).

A Figura 5.43 mostra os escores para a primeira componente principal.

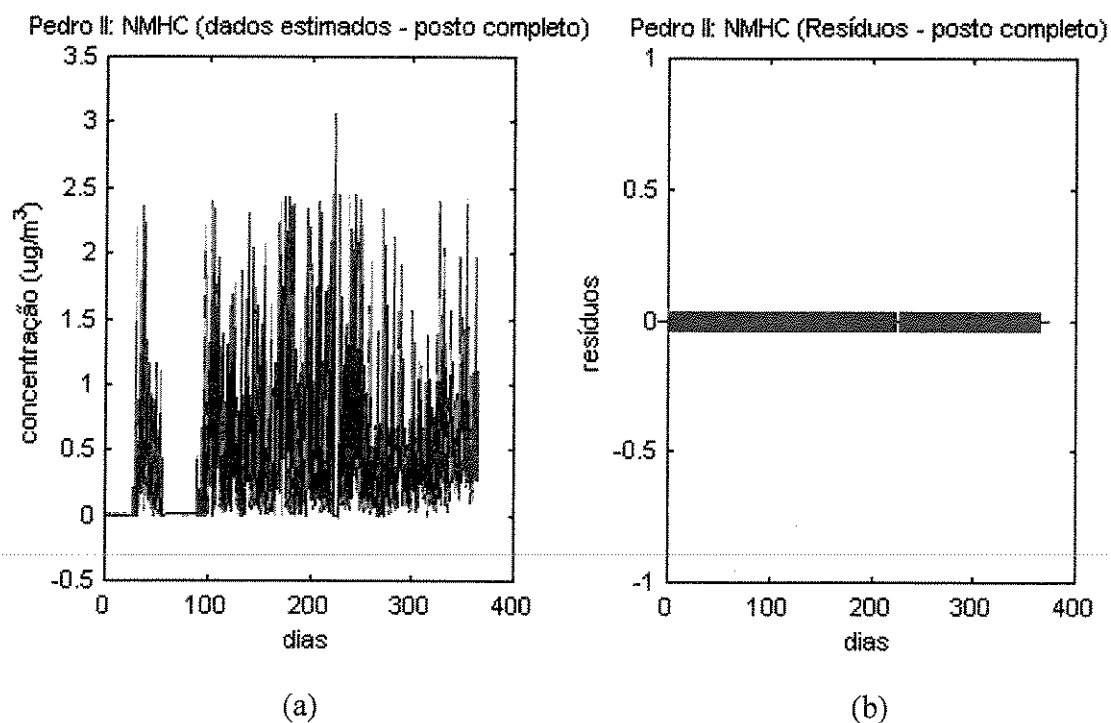


Fig.5.44: a)Dados estimados e b)Resíduos, considerando-se posto completo.

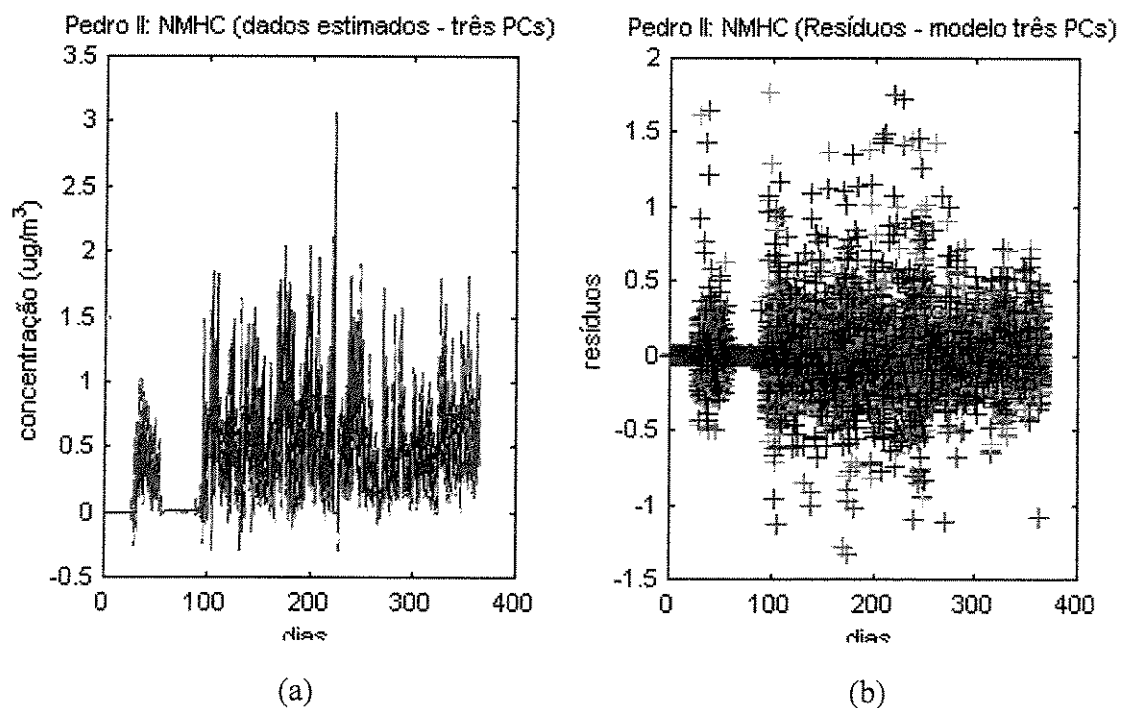


Fig.5.45: a) Dados estimados e b) Resíduos, considerando-se três PCs.

As Figuras 5.44 e 5.45 mostram os dados estimados e os resíduos.

5.1.1.8 – Análise PCA (P. D. Pedro II – Temp: 1999)

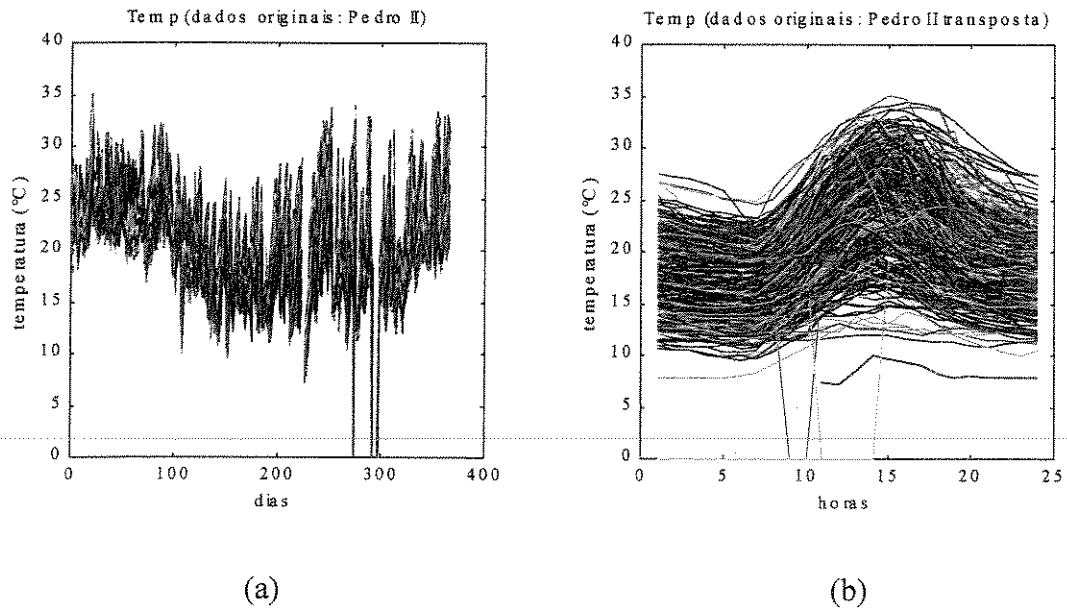


Fig.5.46: a)Gráfico concentração × dia. b) Gráfico concentração × hora.

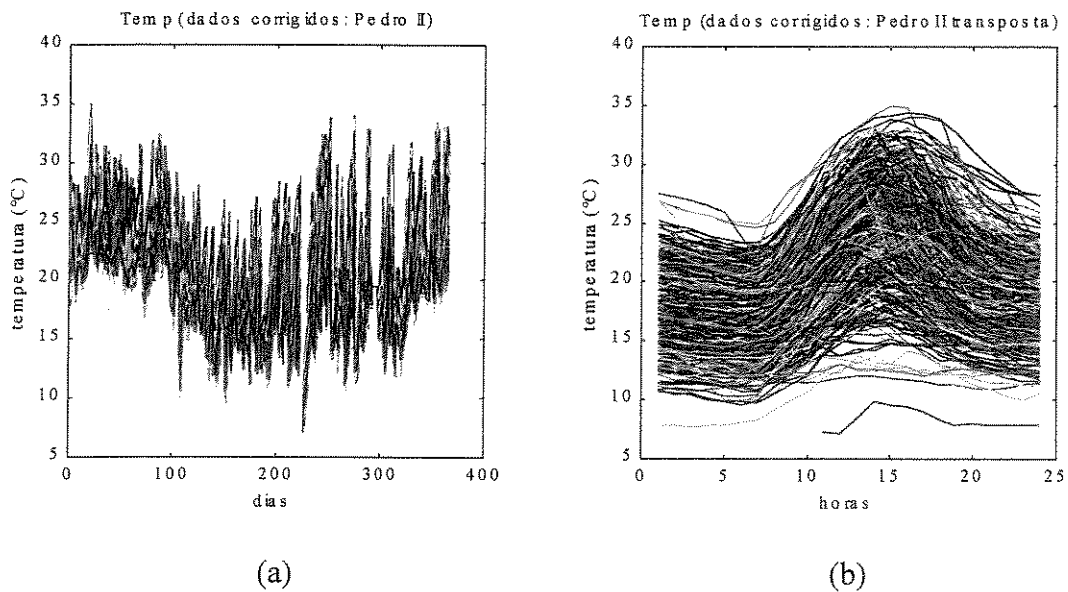
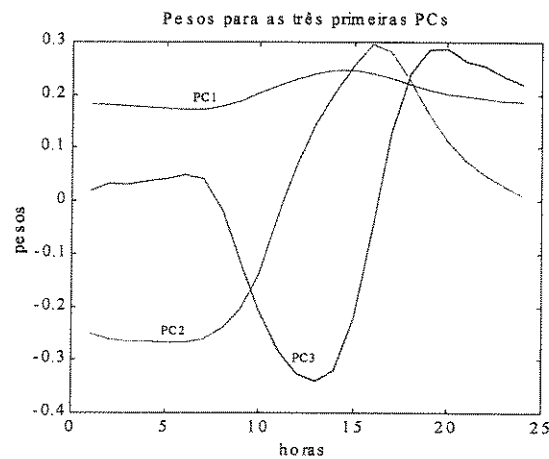
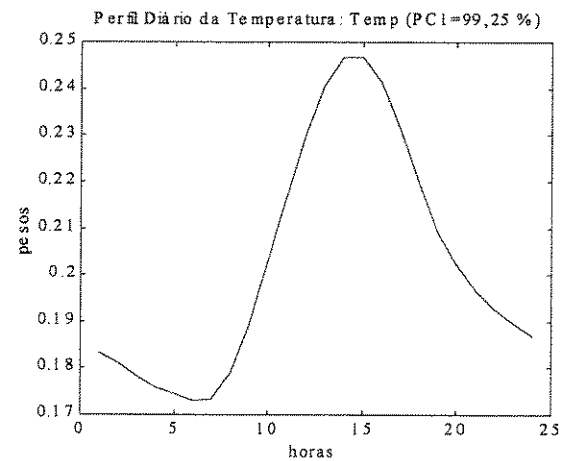


Fig.5.47: a)Gráfico concentração × dia. b) Gráfico concentração × hora.

As Figuras 5.46 e 5.47 mostram os dados originais, respectivamente, sem e com correções.

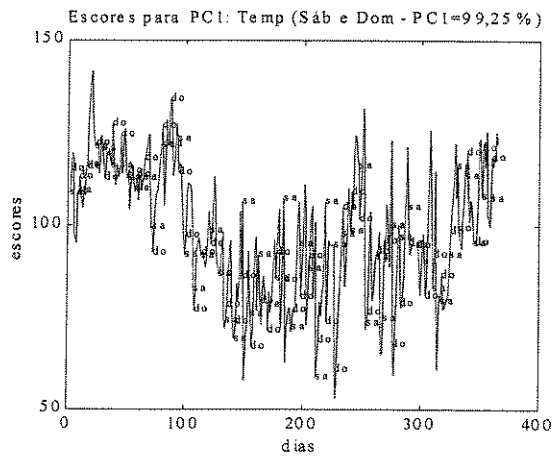


(a)

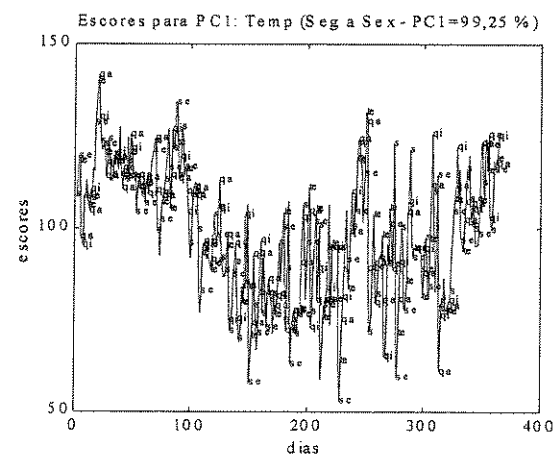


(b)

Fig.5.48: a)Pesos para PC1, PC2 e PC3. b)Pesos para PC1.



(a)



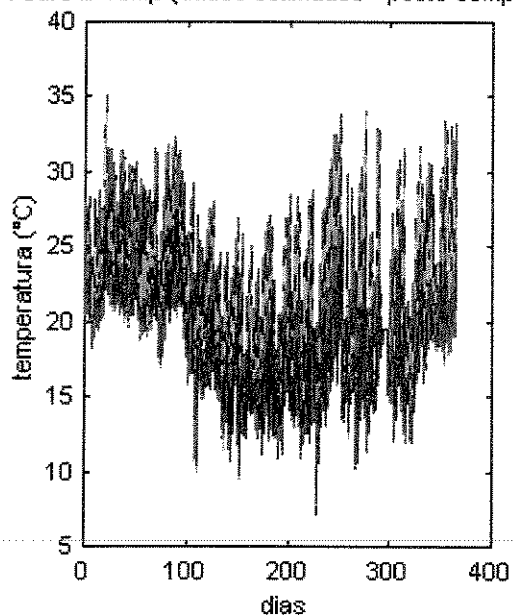
(b)

Fig.5.49: a)Escores para PC1 (finais de semana). b)Escores para PC1 (dias úteis).

A Figura 5.48 mostra os pesos para as PCs. É importante comparar esta com a Figura 5.54, para que se observe a correlação negativa existente entre as variáveis temperatura e umidade.

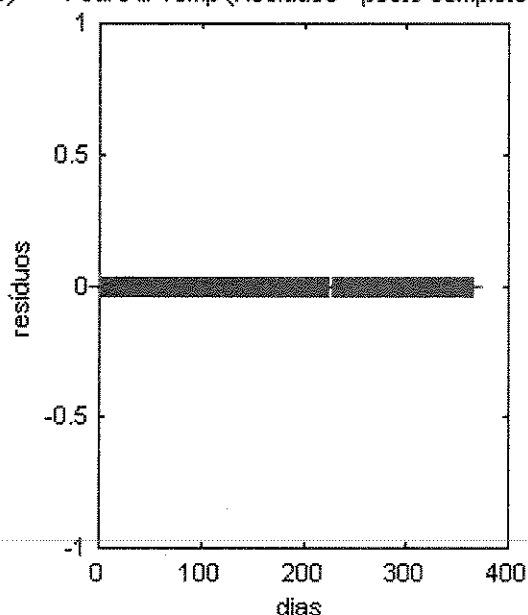
A Figura 5.49 mostra os escores para a primeira componente principal. Aqui também é útil a comparação com a Figura 5.55.

Pedro II: Temp (dados estimados - posto completo)



(a)

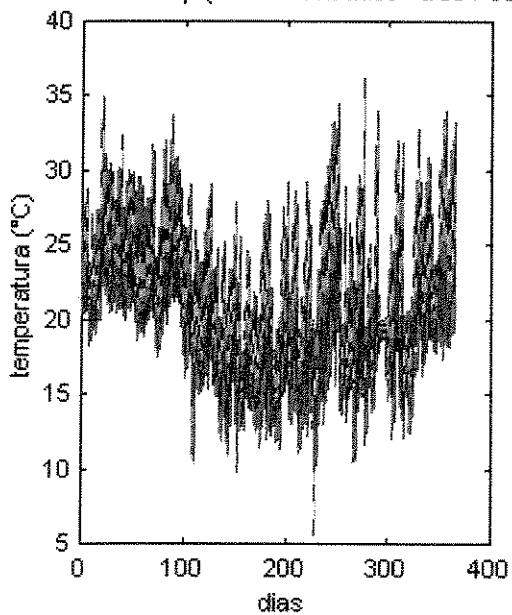
Pedro II: Temp (Resíduos - posto completo)



(b)

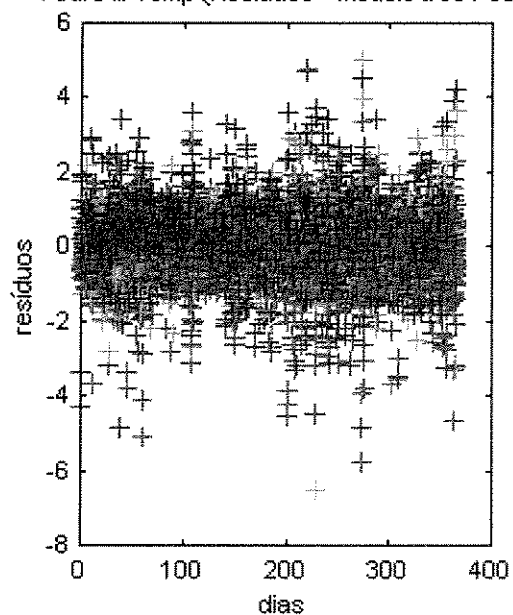
Fig.5.50: a)Dados estimados e b)Resíduos, considerando-se posto completo.

Pedro II: Temp (dados estimados - três PCs)



(a)

Pedro II: Temp (Resíduos - modelo três PCs)



(b)

Fig.5.51: a) Dados estimados e b) Resíduos, considerando-se três PCs.

As Figuras 5.50 e 5.51 mostra os dados estimados e os resíduos.

5.1.1.9 – Análise PCA (P. D. Pedro II – Umid: 1999)

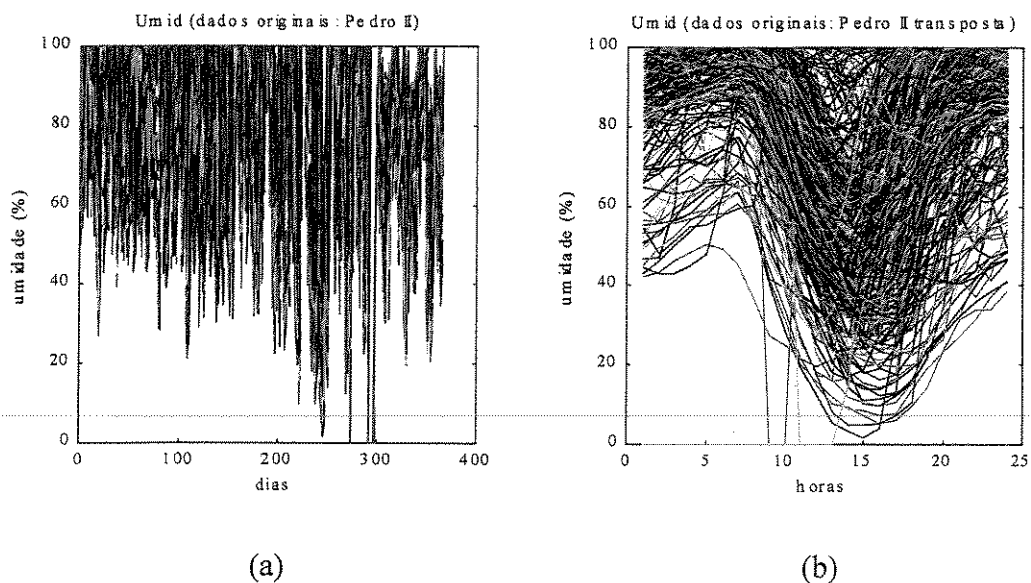


Fig.5.52: a)Gráfico concentração × dia. b) Gráfico concentração × hora.

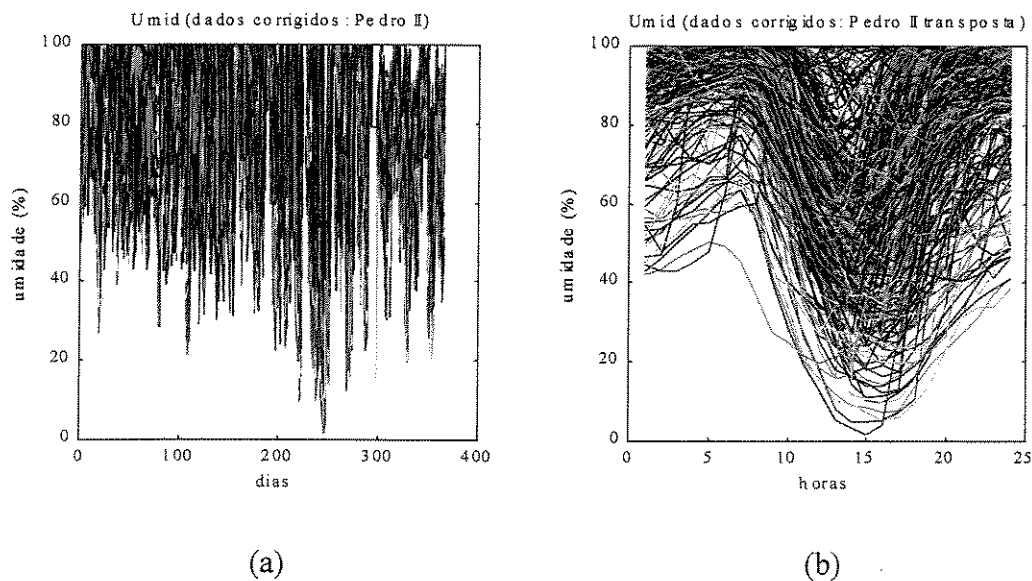
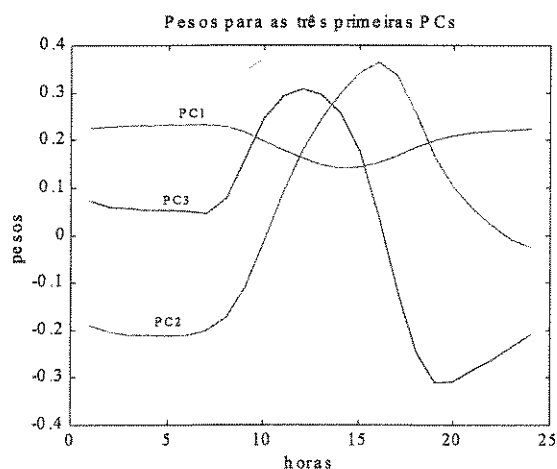
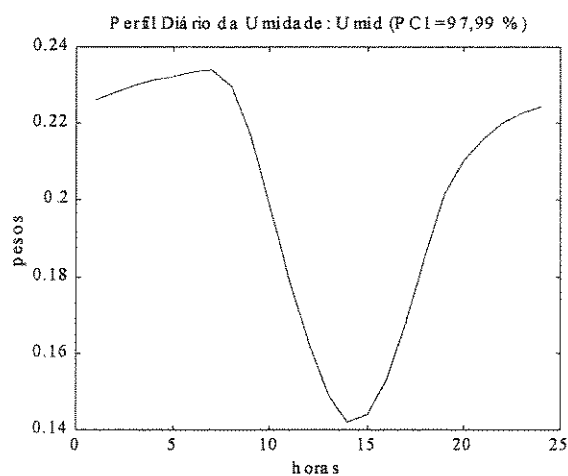


Fig.5.53: a)Gráfico concentração × dia. b) Gráfico concentração × hora.

As Figuras 5.52 e 5.53 mostram os dados originais, respectivamente, sem e com correções.

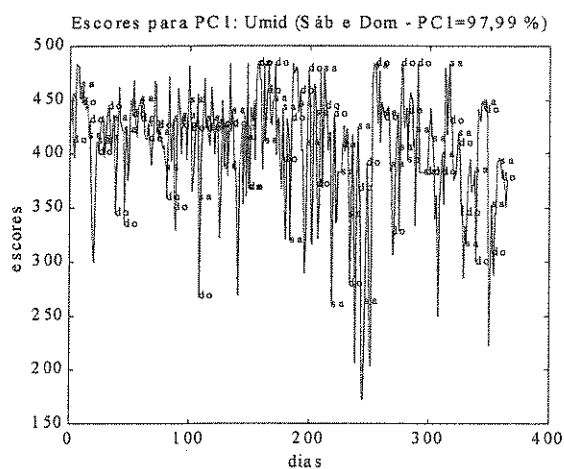


(a)

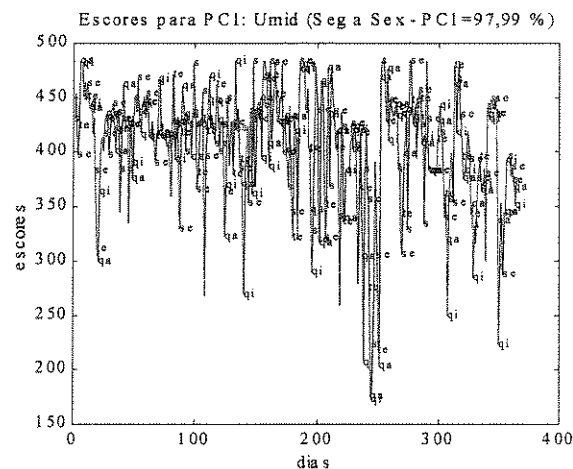


(b)

Fig.5.54: a)Pesos para PC1, PC2 e PC3. b)Pesos para PC1.



(a)



(b)

Fig.5.55: a)Escores para PC1 (finais de semana). b)Escores para PC1 (dias úteis).

A Figura 5.54 mostra os pesos para as componentes principais (as PCs). Nota-se a existência de correlação negativa entre os níveis obtidos para a temperatura e para a umidade.

A Figura 5.55 mostra os escores para a primeira componente principal.

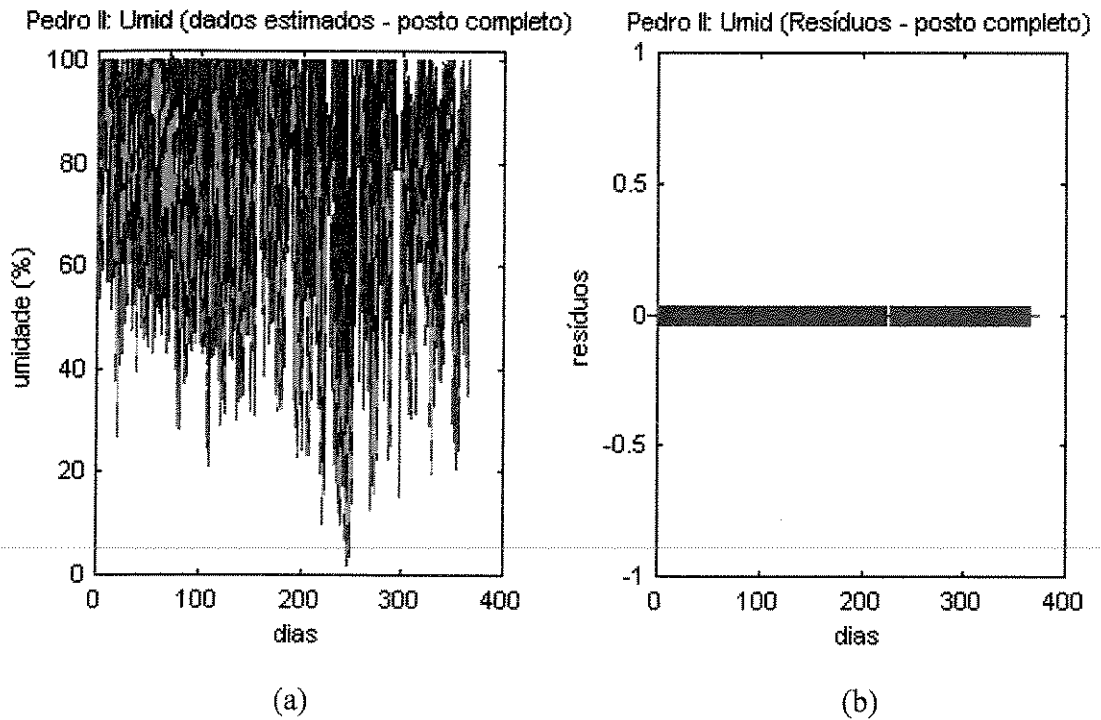


Fig.5.56: a)Dados estimados e b)Resíduos, considerando-se posto completo.

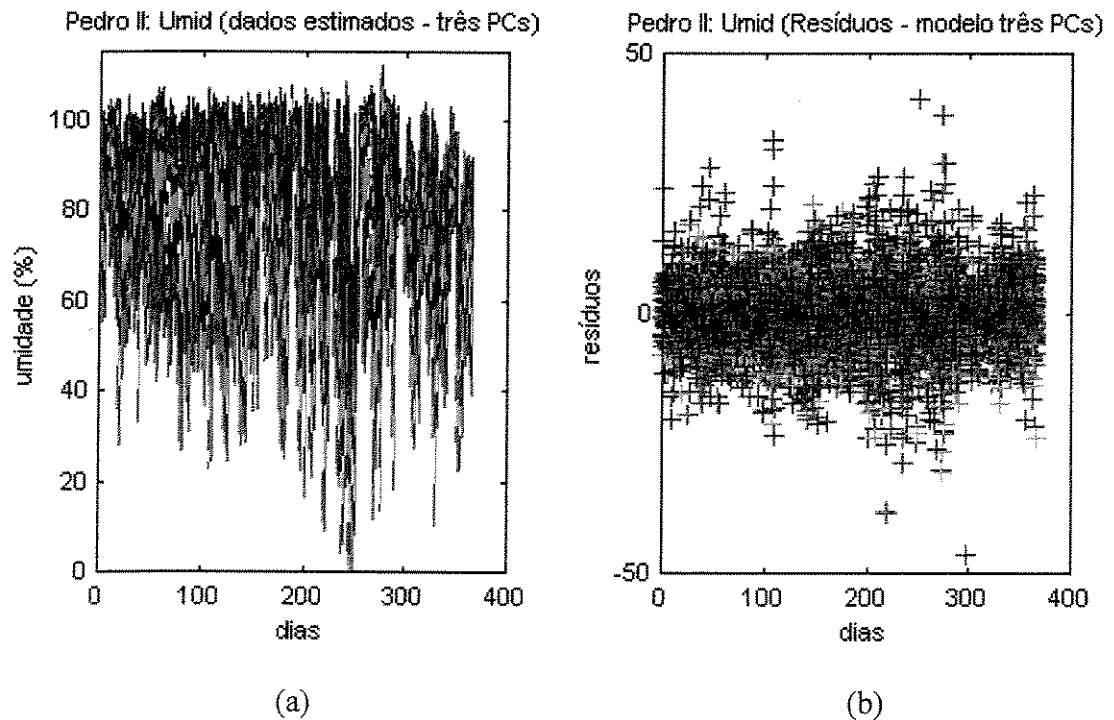


Fig.5.57: a) Dados estimados e b) Resíduos, considerando-se três PCs.

As Figuras 5.56 e 5.57 mostram os dados estimados e os resíduos.

Uma comparação entre as Figuras 5.48 e 5.54 mostra que no primeiro caso, para a temperatura, a terceira componente principal, PC_3 , não possui significado físico. Por outro lado, no segundo caso, para a umidade, ambas PC_2 e PC_3 , não possuem significado físico. O valor encontrado para PC_1 (98,0%) indica pois, uma única fonte para a umidade na atmosfera. Embora essa seja uma conclusão óbvia, ela serve para mostrar consistência nas análises, isto é, na adequação dos modelos aplicados.

No caso da temperatura, a variância capturada pela primeira componente principal, PC_1 , alcança 99,25%. Embora não se possa afirmar que PC_2 não possua significado físico, a variância residual a ser distribuída entre as duas outras PCs indica que mesmo que à PC_2 possa ser atribuída uma fonte de aquecimento para a atmosfera, ela pode ser considerada desprezível.

No Capítulo VI, quando são apresentadas as análises com os modelos de ordem superior, poderá ser visto que a variação da temperatura só adquire importância quando se consideram as variações sistemáticas diárias e a variação anual (a variação sazonal). Não obstante durante o dia e ao longo do ano os valores registrados para a temperatura apresentem variações significativas, o perfil semanal para essa variável é aproximadamente constante.

Discute-se no Capítulo VII, o problema das inversões. Aborda-se o problema da altura de mistura e suas correlações com a temperatura, a umidade e com a concentração de ozônio na troposfera.

5.1.2 – Análise PCA (Monóxido de Carbono – Sete Estações: 1998)

Nos tópicos 5.1.2.1 a 5.1.2.7 pretende-se dar também uma visão geral dos dados, porém, para o ano de 1998. Analisa-se, para efeito de comparação, o comportamento do poluente CO nas sete Estações de Medição. Como em 5.1.1, a seqüência de resultados também engloba figuras contendo a matriz de dados originais, matriz de dados originais corrigidos, matriz de escores e de pesos para as componentes principais, a matriz de dados estimados pelo modelo, e a matriz de resíduos deixados pelo modelo.

5.1.2.1 – Análise PCA (CO – P. D. Pedro II: 1998)

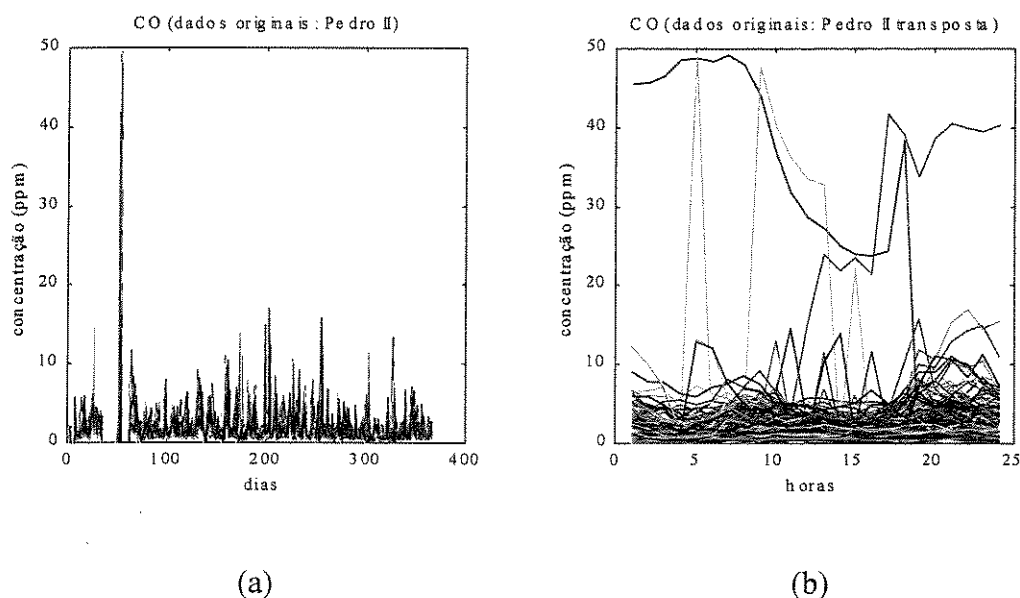


Fig.5.58: a)Gráfico concentração × dia. b) Gráfico concentração × hora.

A Figura 5.58a mostra a matriz de dados originais para o CO na Estação de Medição Parque D. Pedro II, durante o ano de 1998. A Figura 5.58b mostra a transposta da matriz de dados originais. Como se observa, a matriz transposta põe em evidência tanto os dados anômalos (“outliers”) quanto a maioria daqueles provenientes de variações súbitas. Por exemplo, o pico que aparece em 5.58a no 60º dia, é a mesma curva em vermelho (com valores muito elevados) no período de 1:00 h até 19:00 h, que se observa em 5.58b. Todos os pontos dessa curva, nesse intervalo, são considerados dados anômalos.

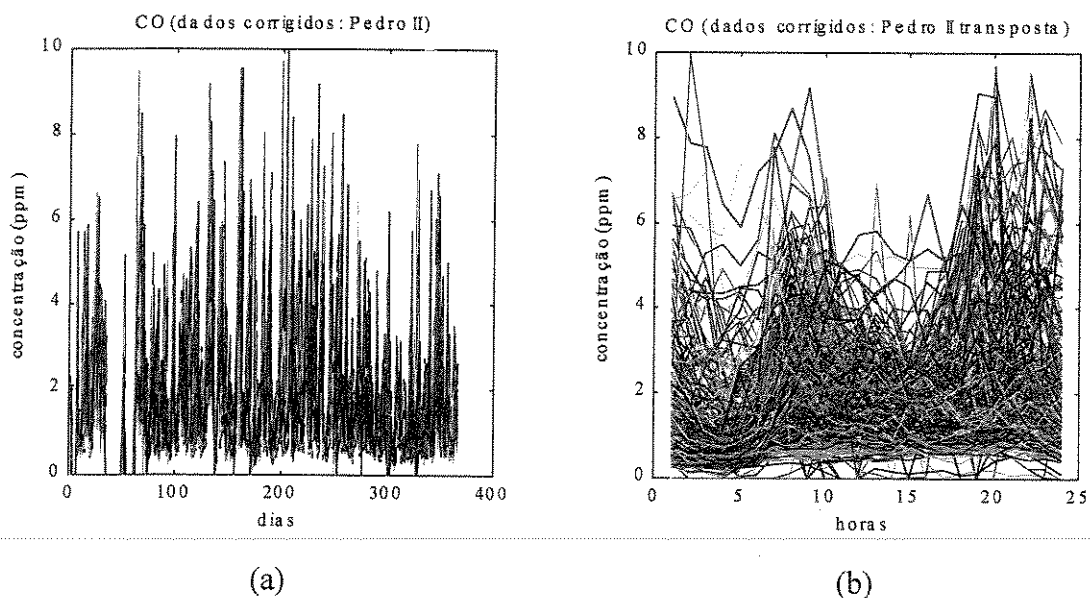


Fig.5.59: a)Gráfico concentração × dia. b) Gráfico concentração × hora.

A Figura 5.59 mostra os dados originais corrigidos. Notam-se ainda algumas variações súbitas, mas que não causaram influência na análise.

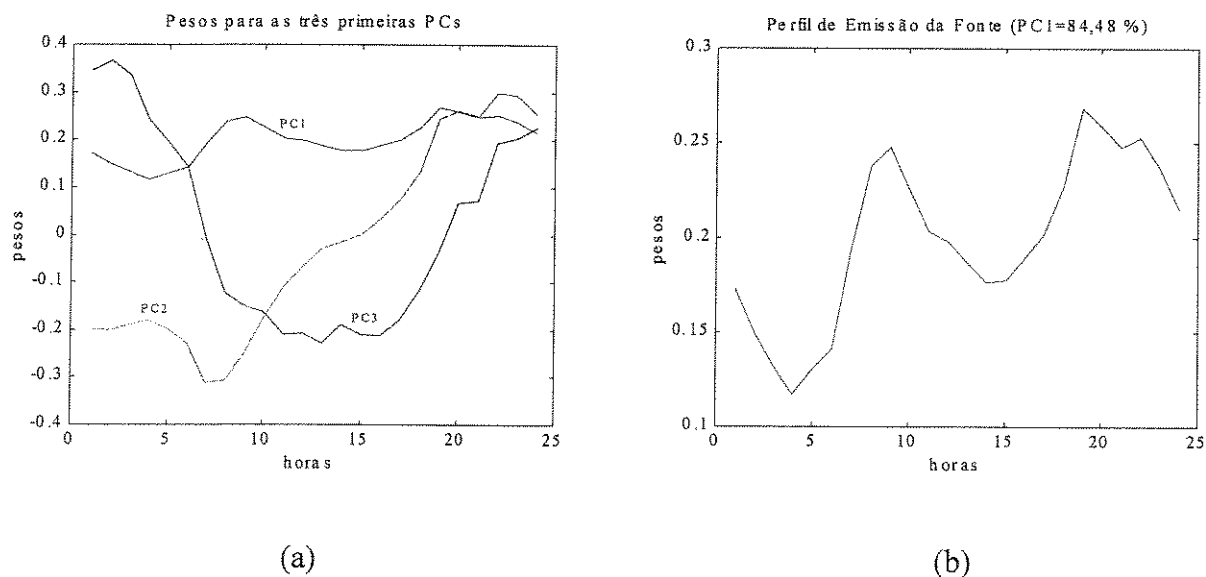
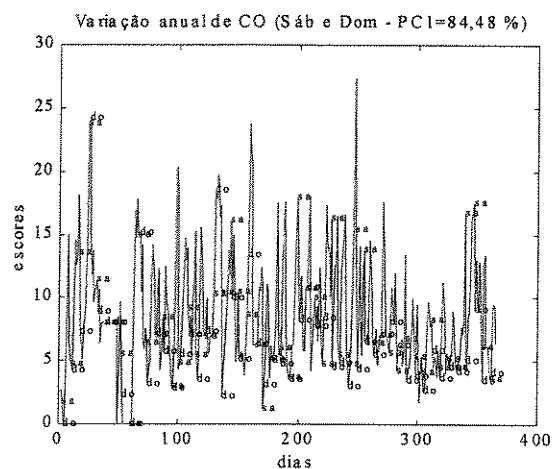
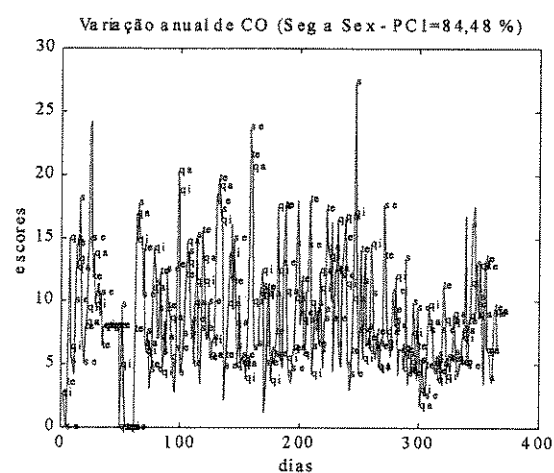


Fig.5.60: a)Pesos para PC1, PC2 e PC3. b)Pesos para PC1.

A Figura 5.60 mostra os pesos, respectivamente, para as três primeiras PCs e para a primeira PC. Nota-se a semelhança com os dados para o CO em 1999 nesse local.



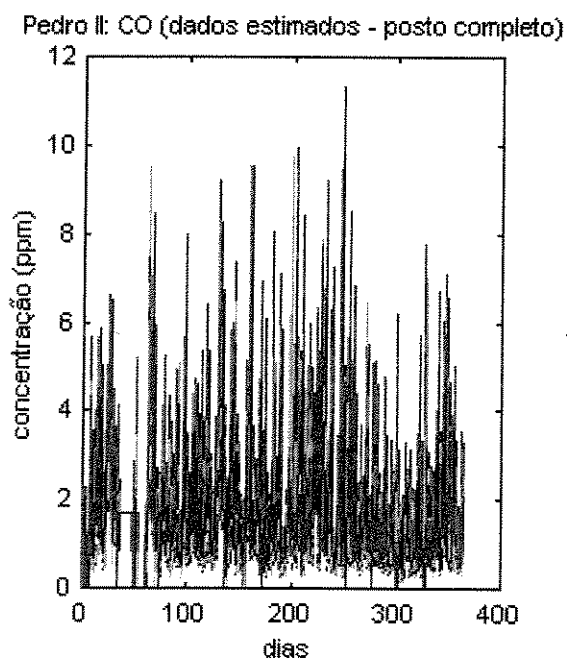
(a)



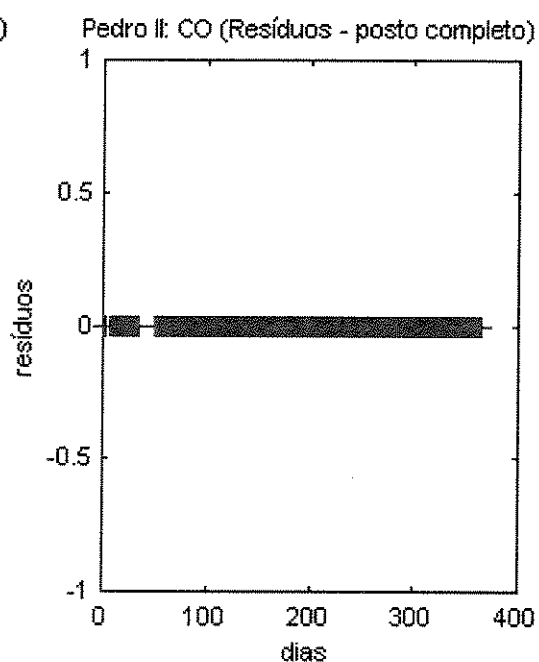
(b)

Fig.5.61: a)Escore para PC1 (finais de semana). b)Escore para PC1 (dias úteis).

A Figura 5.61 mostra os escores para a primeira componente principal.



(a)



(b)

Fig.5.62: a)Dados estimados e b)Resíduos, considerando-se posto completo.

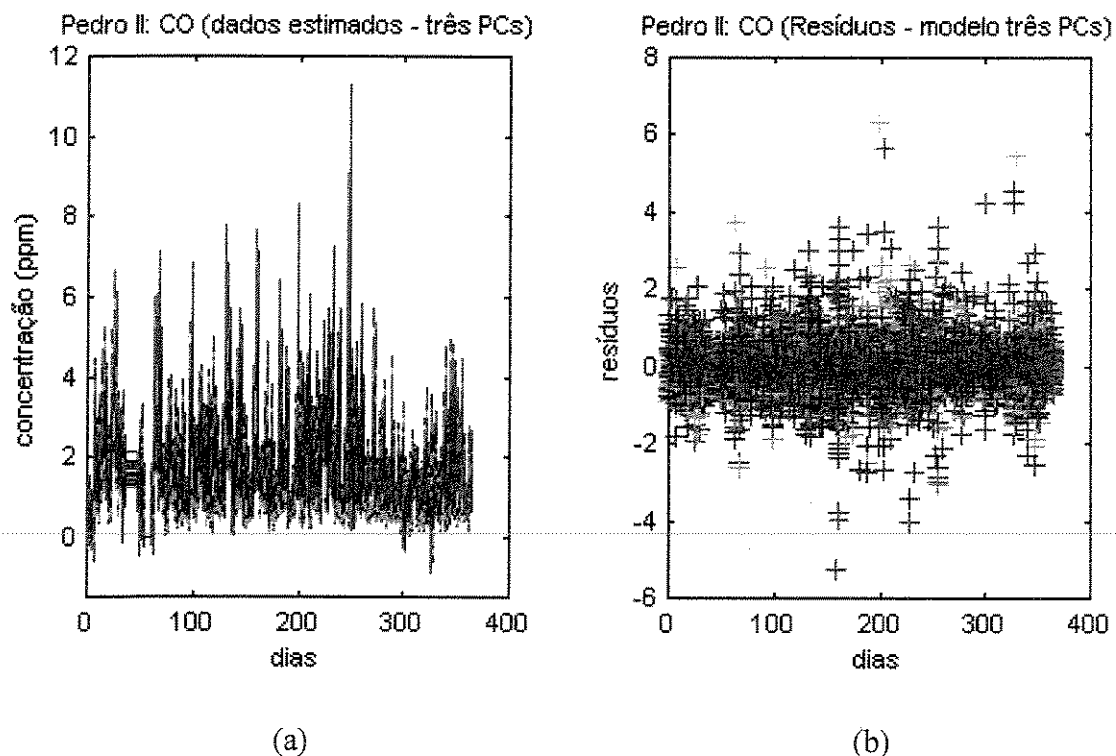


Fig.5.63: a) Dados estimados e b) Resíduos, considerando-se três PCs.

As Figuras 5.62 e 5.63 mostram os dados estimados e os resíduos considerando-se, respectivamente, posto completo e as três primeiras componentes principais.

Na Figura 5.61 pode ser observado que os valores dos escores para os finais de semana são baixos quando comparados àqueles para os dias úteis, semelhante ao que ocorreu em outros locais.

5.1.2.2 – Análise PCA (CO – São Caetano do Sul: 1998)

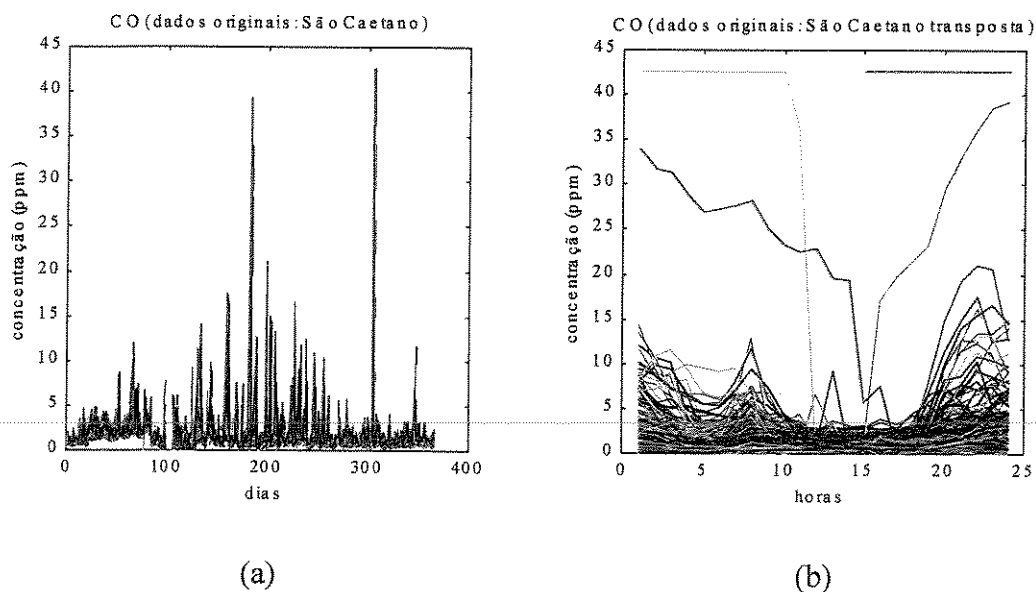


Fig.5.64: a)Gráfico concentração × dia. b) Gráfico concentração × hora.

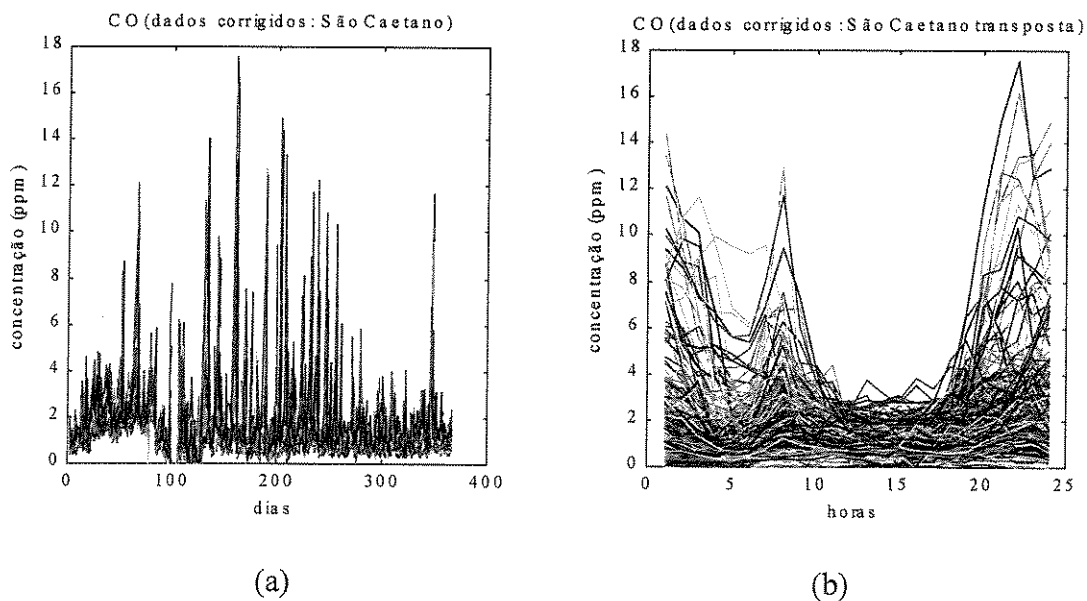
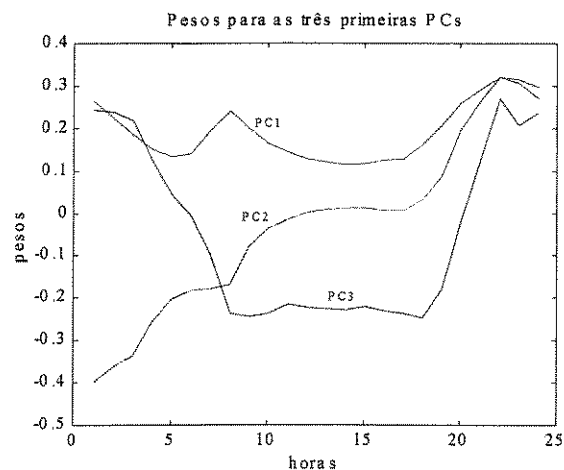
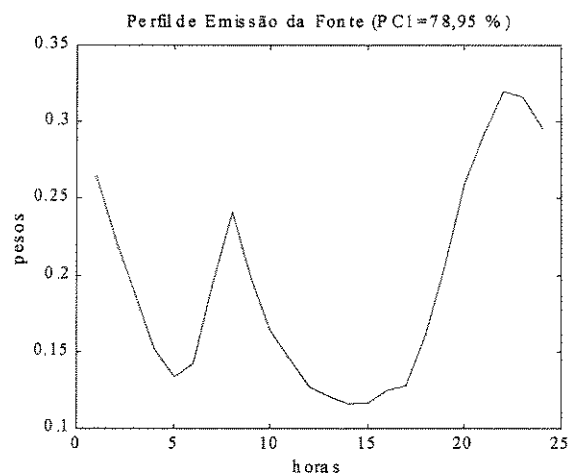


Fig.5.65: a)Gráfico concentração × dia. b) Gráfico concentração × hora.

As Figuras 6.64 e 5.65 mostram os dados originais, respectivamente, sem e com correções. Novamente em 5.64b é possível observar a ocorrência de dados anômalos com bastante facilidade.

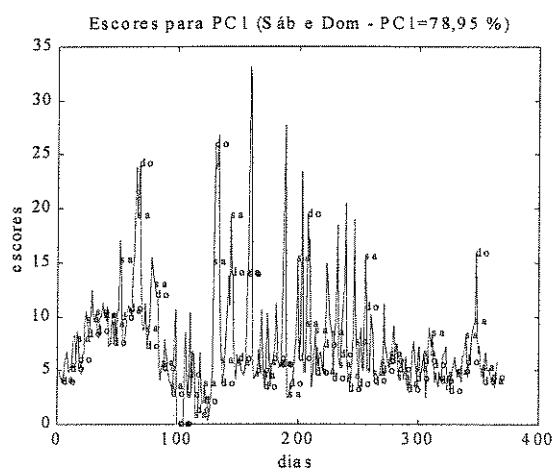


(a)

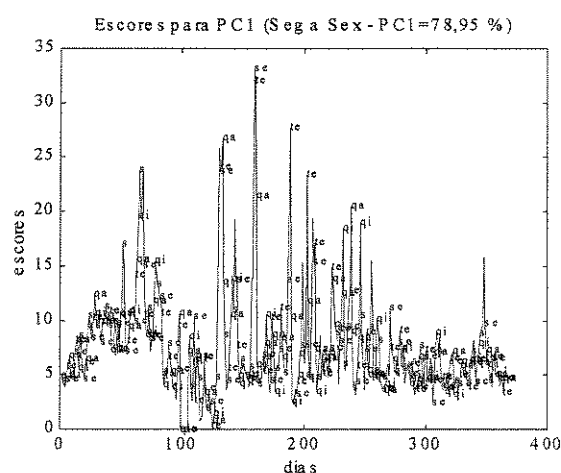


(b)

Fig.5.66: a)Pesos para PC1, PC2 e PC3. b)Pesos para PC1.



(a)



(b)

Fig.5.67: a)Escores para PC1 (finais de semana). b)Escores para PC1 (dias úteis).

A Figura 5.66 mostra os pesos para as PCs. Em 5.66b pode-se perceber a semelhança com o perfil para o poluente CO nessa mesma Estação de Medição durante o ano de 1999.

A Figura 5.67 mostra os escores para a primeira componente principal.

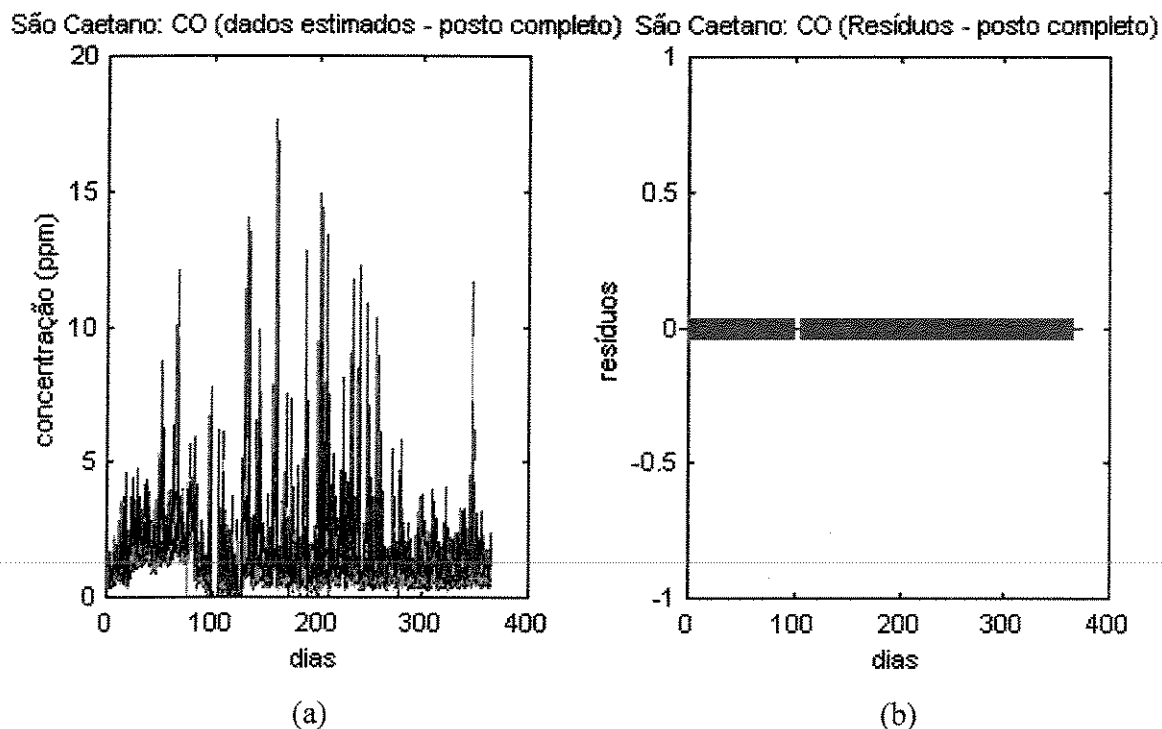


Fig.5.68: a)Dados estimados e b)Resíduos, considerando-se posto completo.

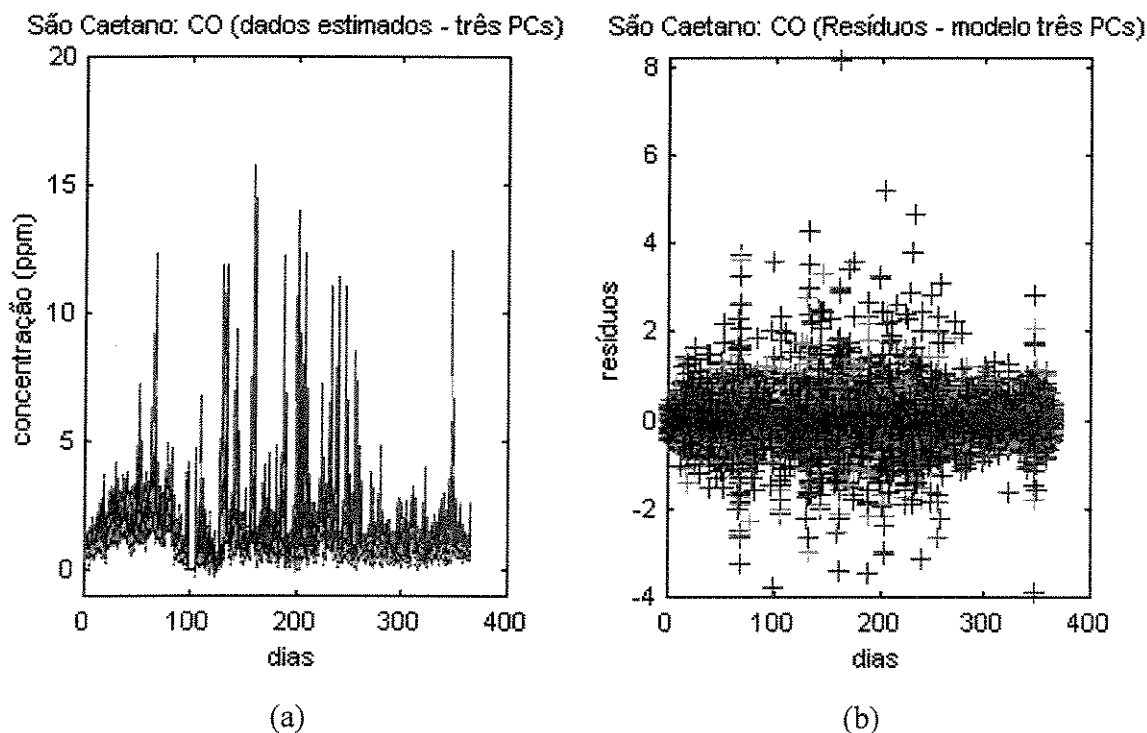


Fig.5.69: a) Dados estimados e b) Resíduos, considerando-se três PCs.

As Figuras 6.68 e 5.69 mostram os dados estimados e os resíduos.

5.1.2.3 – Análise PCA (CO – Ibirapuera: 1998)

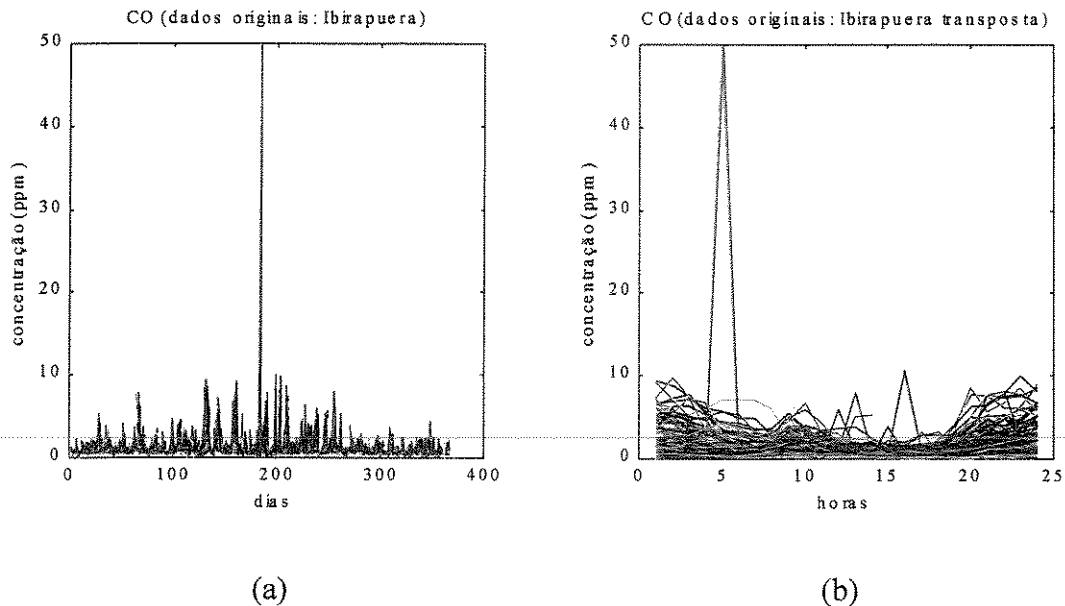


Fig.5.70: a)Gráfico concentração × dia. b) Gráfico concentração × hora.

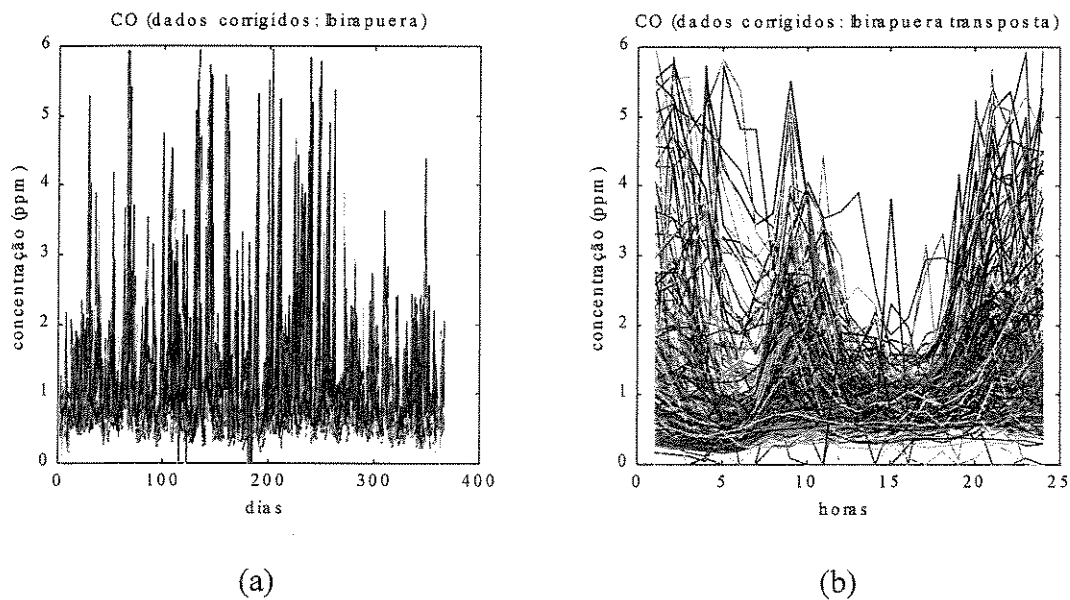
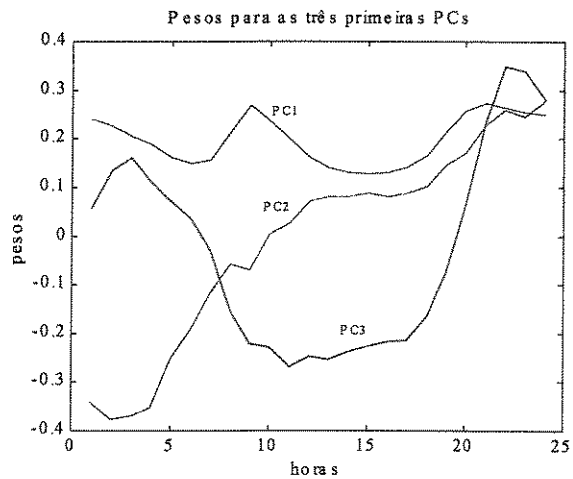
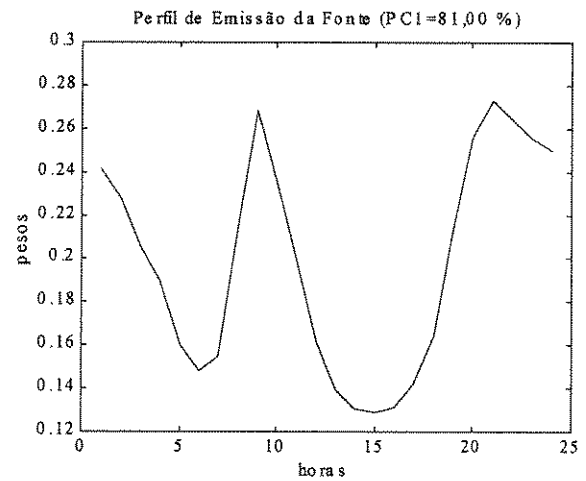


Fig.5.71: a)Gráfico concentração × dia. b) Gráfico concentração × hora.

As Figuras 5.70 e 5.71 mostram os dados originais, respectivamente, sem e com correções sobre os dados anômalos e variações súbitas.

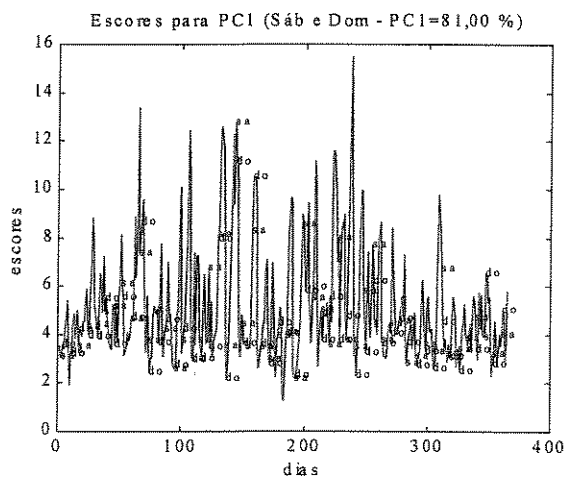


(a)

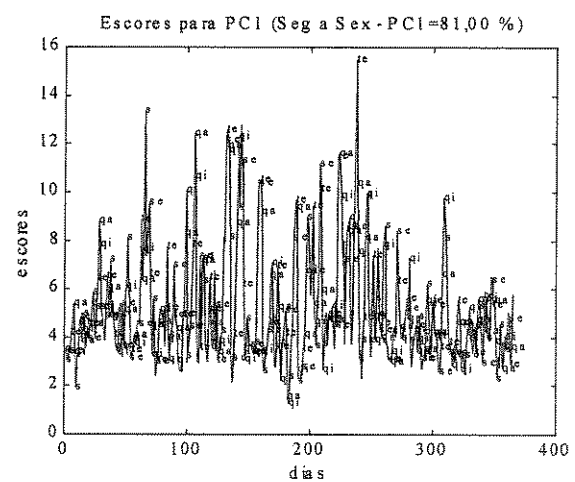


(b)

Fig.5.72: a)Pesos para PC1, PC2 e PC3. b)Pesos para PC1.



(a)

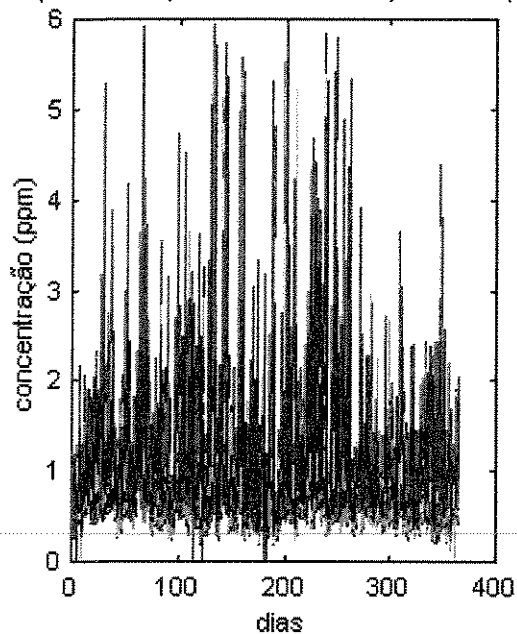


(b)

Fig.5.73: a)Escores para PC1 (finais de semana). b)Escores para PC1 (dias úteis).

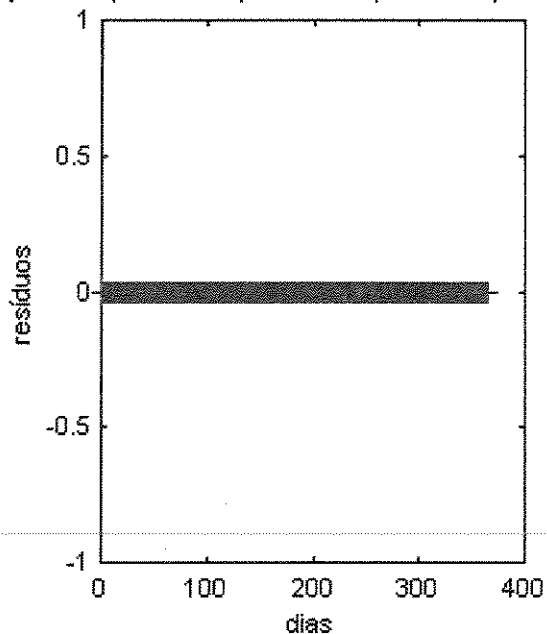
A Figura 5.73 mostra os pesos, respectivamente para as três primeiras componentes principais e para a primeira componente principal. A Figura 5.74 mostra os escores para a primeira componente principal.

Ibirapuera: CO (dados estimados - posto completo)



(a)

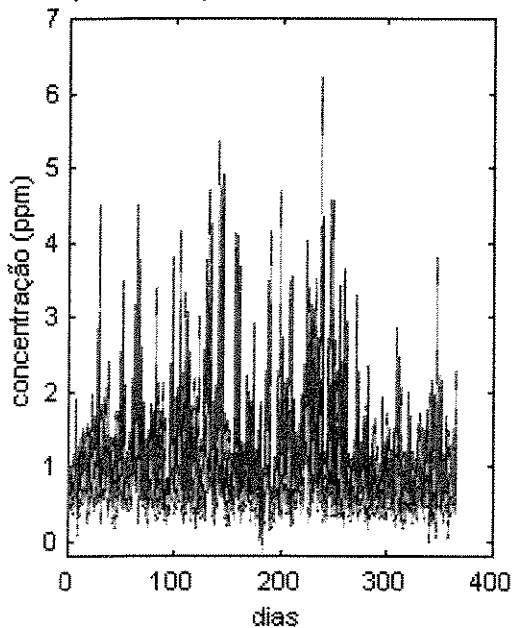
Ibirapuera: CO (Resíduos - posto completo)



(b)

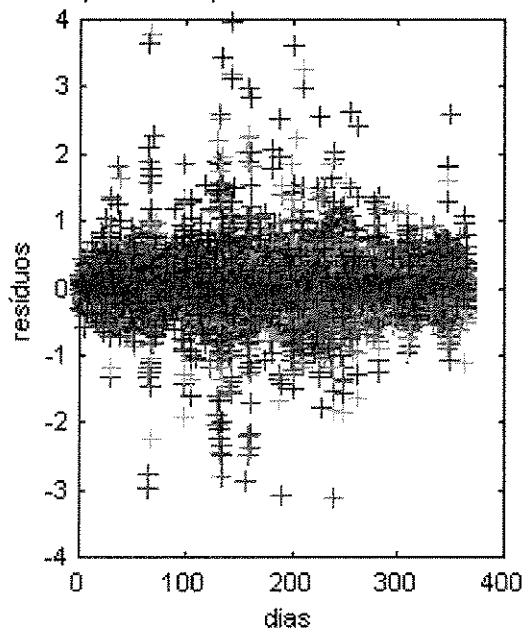
Fig.5.74: a)Dados estimados e b)Resíduos, considerando-se posto completo.

Ibirapuera: CO (dados estimados - três PCs)



(a)

Ibirapuera: CO (Resíduos - modelo três PCs)



(b)

Fig.5.75: a) Dados estimados e b) Resíduos, considerando-se três PCs.

As Figuras 5.74 e 5.75 mostram os dados estimados e os resíduos.

5.1.2.4 – Análise PCA (CO – Cerqueira César: 1998)

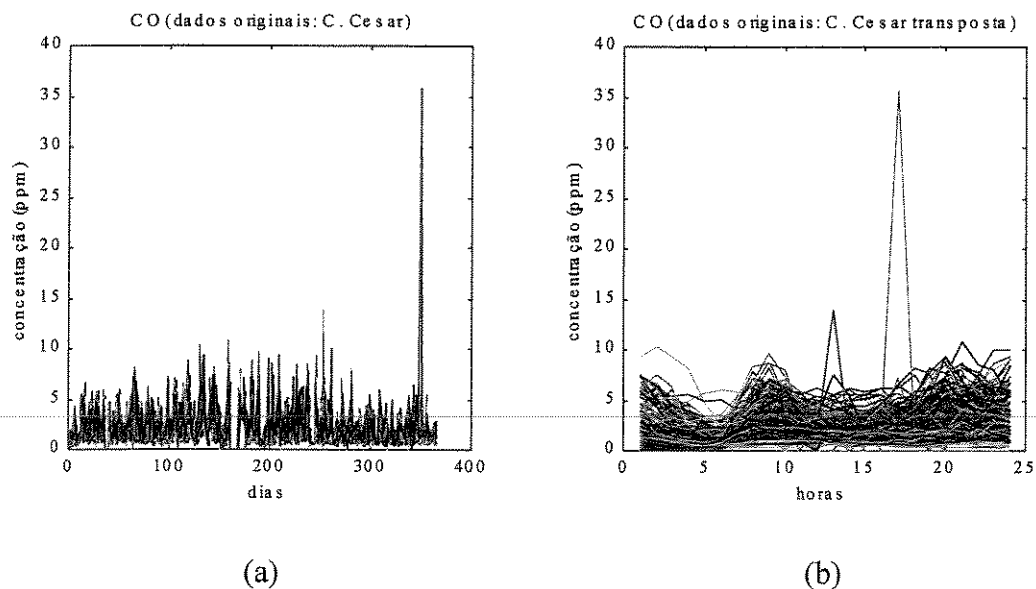


Fig.5.76: a)Gráfico concentração × dia. b) Gráfico concentração × hora.

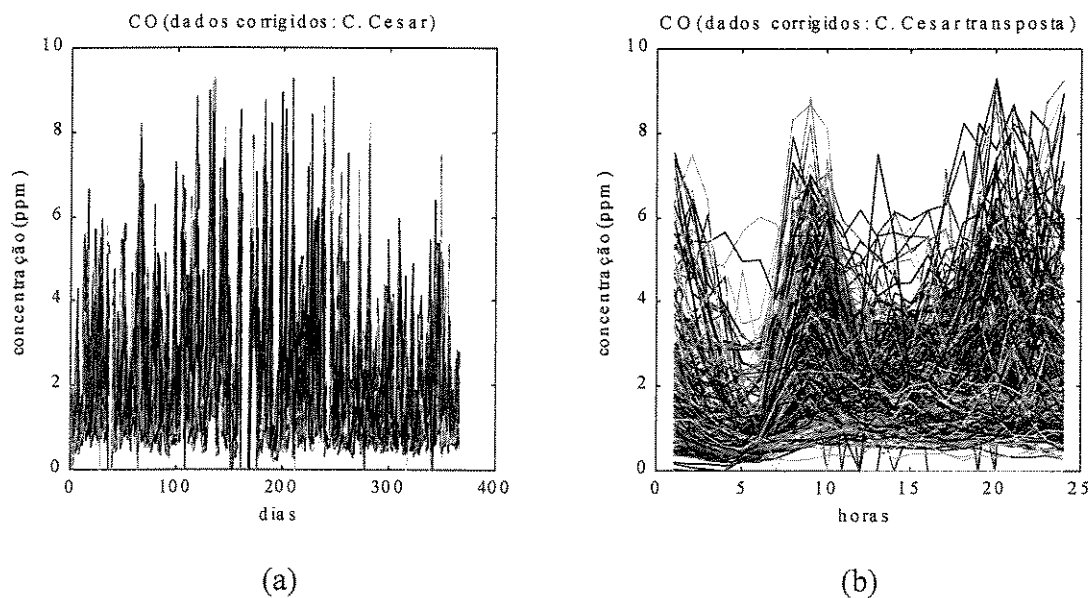
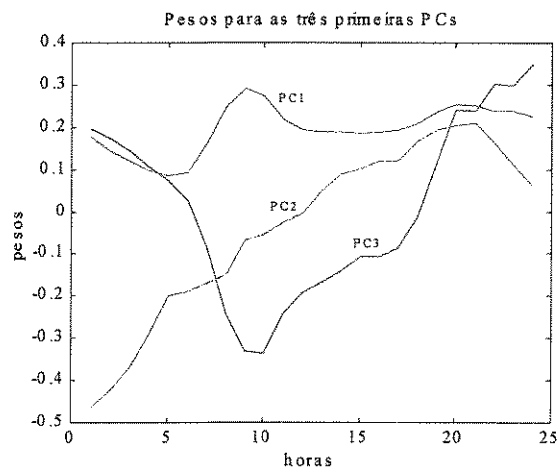
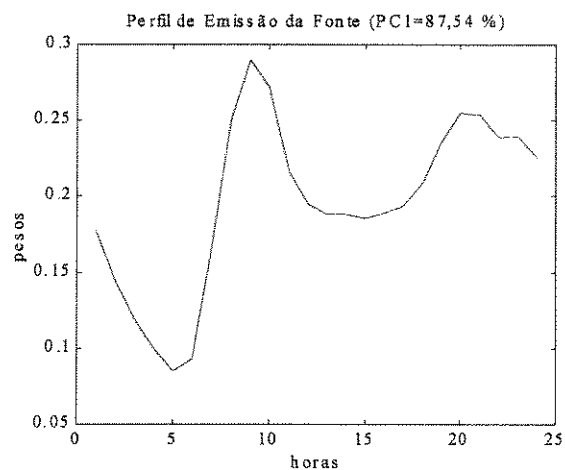


Fig.5.77: a)Gráfico concentração × dia. b) Gráfico concentração × hora.

As Figuras 5.76 e 5.77 mostram os dados originais, respectivamente, sem e com correções. Em 5.76 nota-se a presença de um dado anômalo (pico nas cores vermelha em a e amarela em b). O pico na cor azul (13:00 h em 5.76b) foi considerado variação súbita.

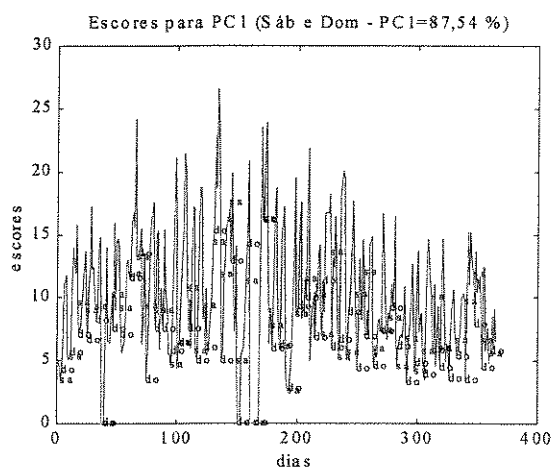


(a)

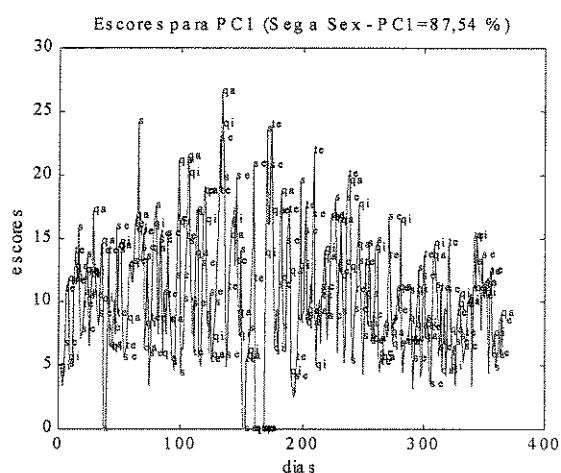


(b)

Fig.5.78: a)Pesos para PC1, PC2 e PC3. b)Pesos para PC1.



(a)



(b)

Fig.5.79: a)Escores para PC1 (finais de semana). b)Escores para PC1 (dias úteis).

A Figura 5.78 mostra os pesos para as PCs. Em 5.78b verifica-se que a queda em torno de 14:00 h não é tão pronunciada. Isto deve estar relacionado com menor queda na emissão local, mas principalmente com menor dispersão do poluente. Entretanto, o padrão (dois picos: um em torno das 9:00 h e outro em torno das 20:00h) se mantém com nos demais casos, indicando origem de mesma fonte.

A Figura 5.79 mostra os escores para a primeira componente principal.

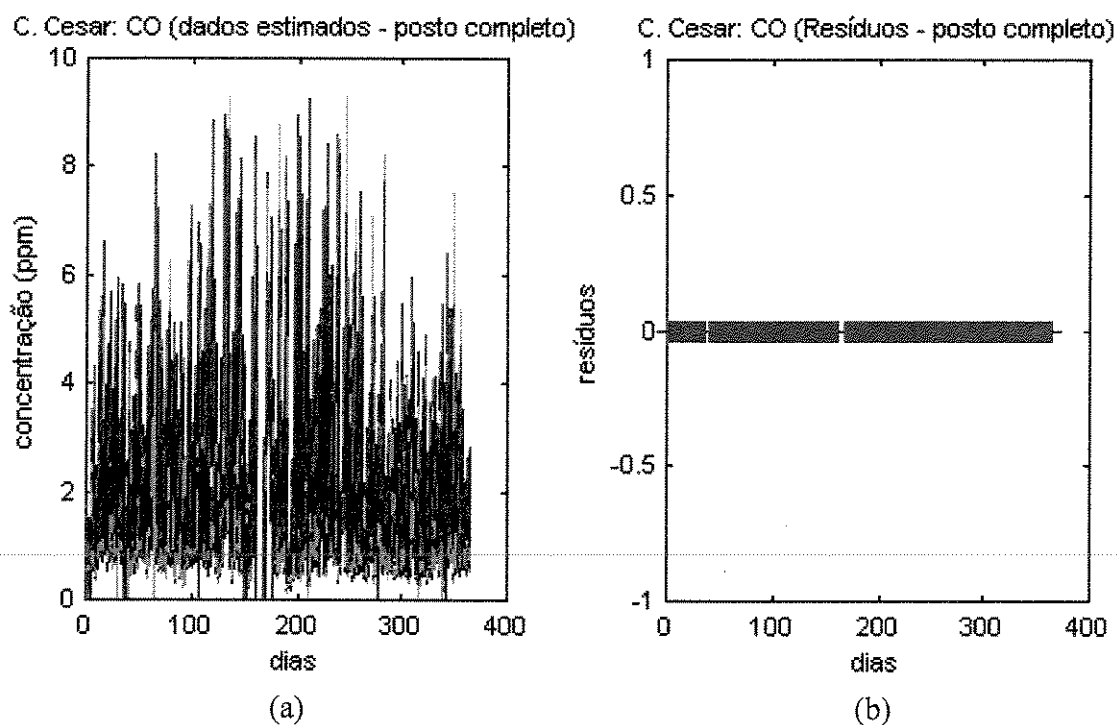


Fig.5.80: a)Dados estimados e b)Resíduos, considerando-se posto completo.

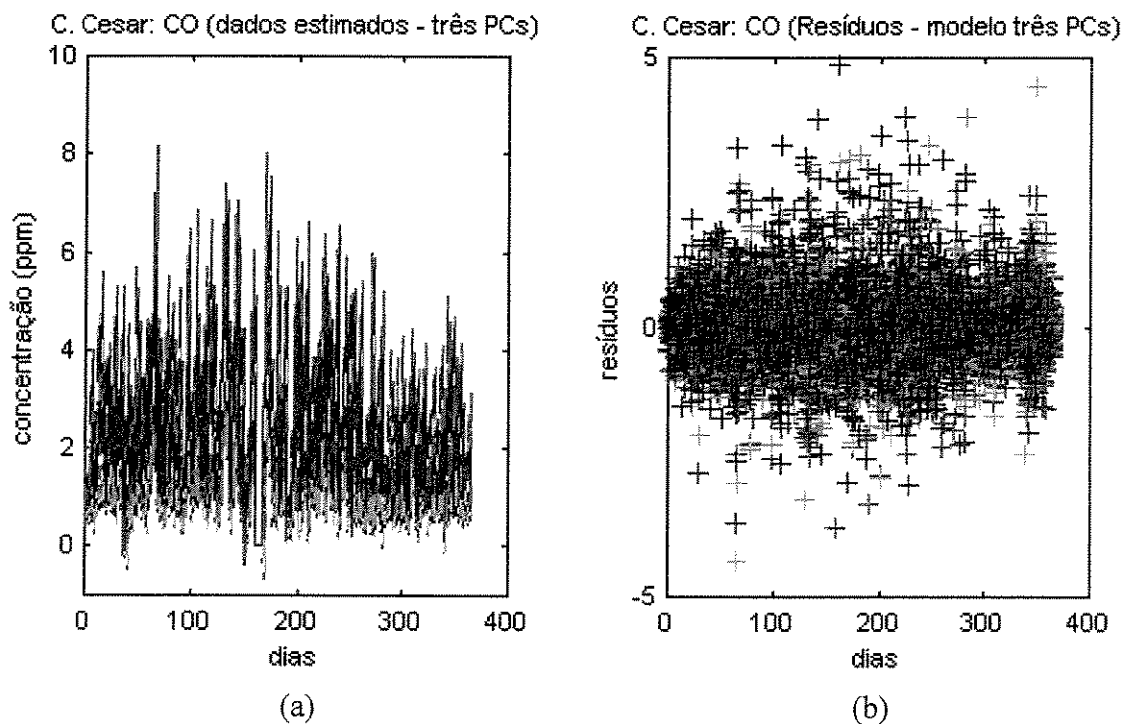


Fig.5.81: a) Dados estimados e b) Resíduos, considerando-se três PCs.

As Figuras 5.80 e 5.81 mostram os dados estimados e os resíduos.

5.1.2.5 – Análise PCA (CO – Centro: 1998)

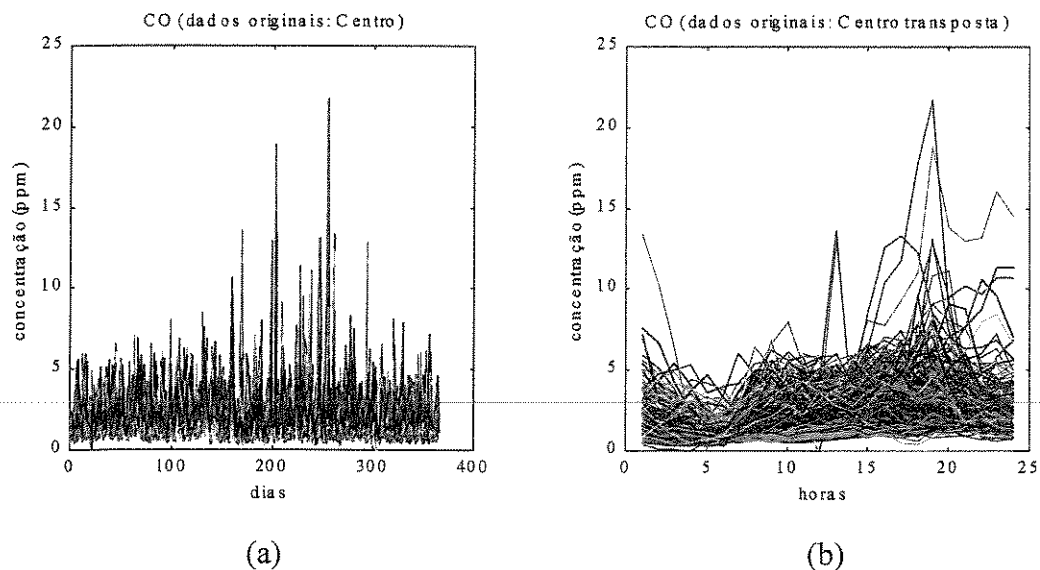


Fig.5.82: a)Gráfico concentração × dia. b) Gráfico concentração × hora.

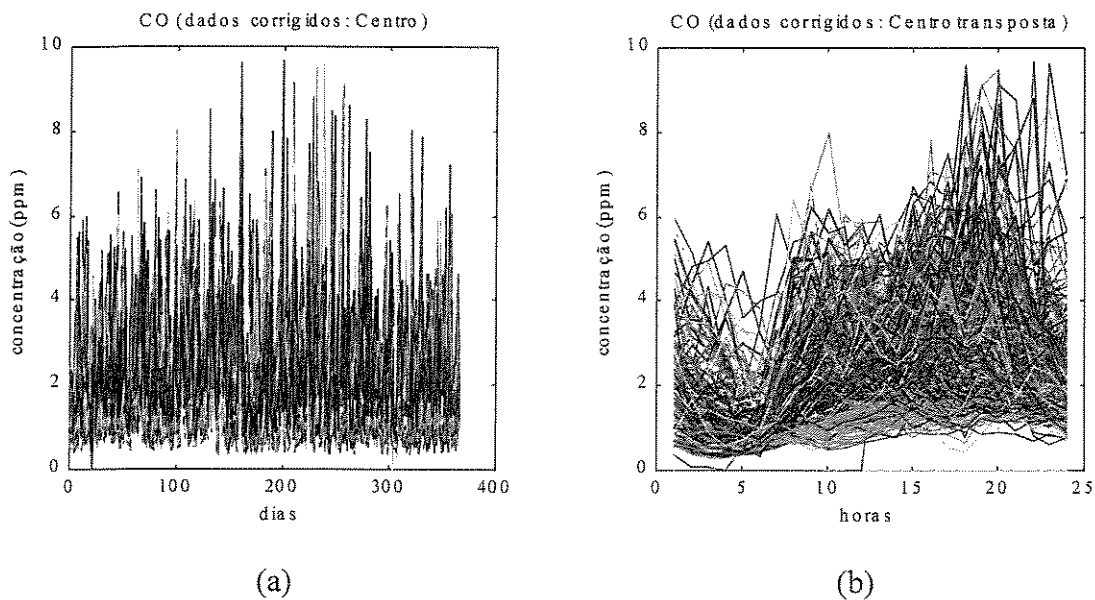
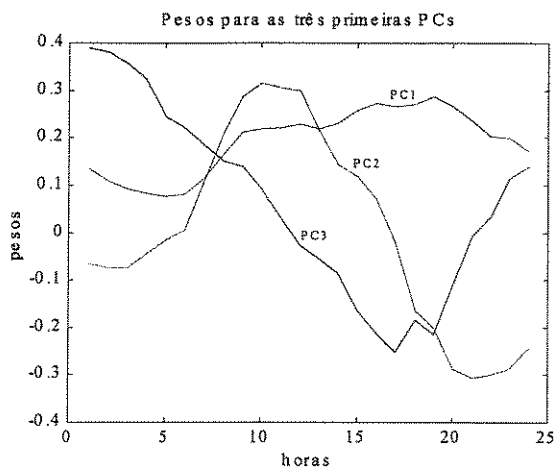
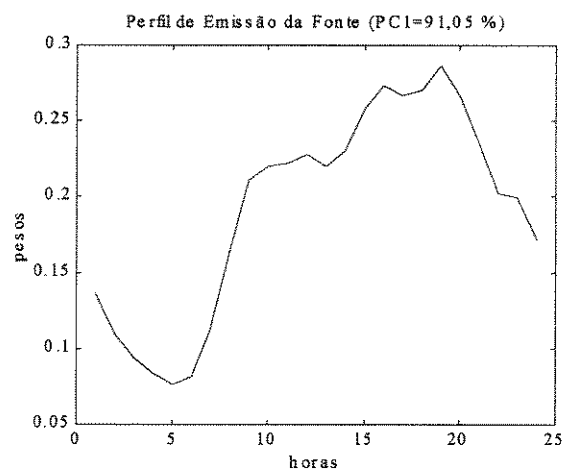


Fig.5.83: a)Gráfico concentração × dia. b) Gráfico concentração × hora.

As Figuras 5.82 e 5.83 mostram os dados originais, respectivamente, sem e com correções. Nota-se em 5.82b a ocorrência de dados anômalos e de variações súbitas. Nota-se também já nessas matrizes maior quantidade de emissões em torno das 14:00 h.

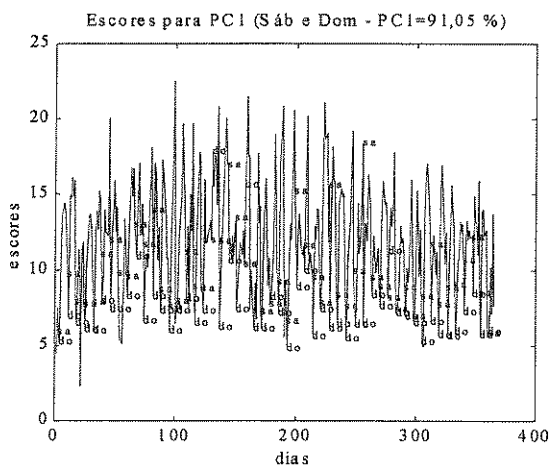


(a)

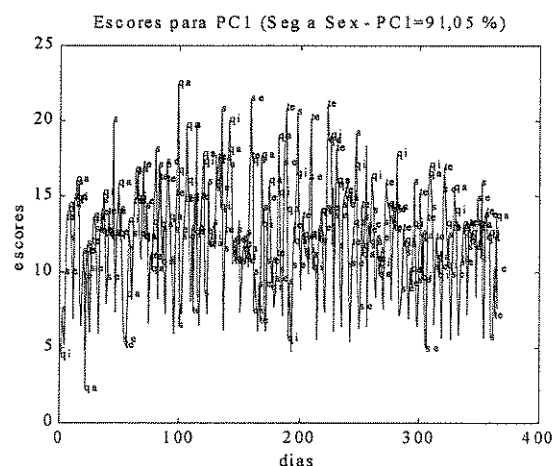


(b)

Fig.5.84: a)Pesos para PC1, PC2 e PC3. b)Pesos para PC1.



(a)



(b)

Fig.5.85: a)Escore para PC1 (finais de semana). b)Escore para PC1 (dias úteis).

A Figura 5.84 mostra os pesos para as PCs. Em 5.84b verifica-se que a queda em torno de 14:00 h é menos pronunciada ainda do que no caso da Figura 5.78b. Isto deve estar relacionado com menor queda na emissão no Centro em relação à Cerqueira Cesar, mas também com menor dispersão do poluente no Centro. Entretanto, o padrão (dois picos: um em torno das 9:00 h – 10:00 h e outro em torno das 20:00h) se mantém como nos demais casos, indicando mesma fonte de emissão.

A Figura 5.85 mostra os escores para as componentes principais.

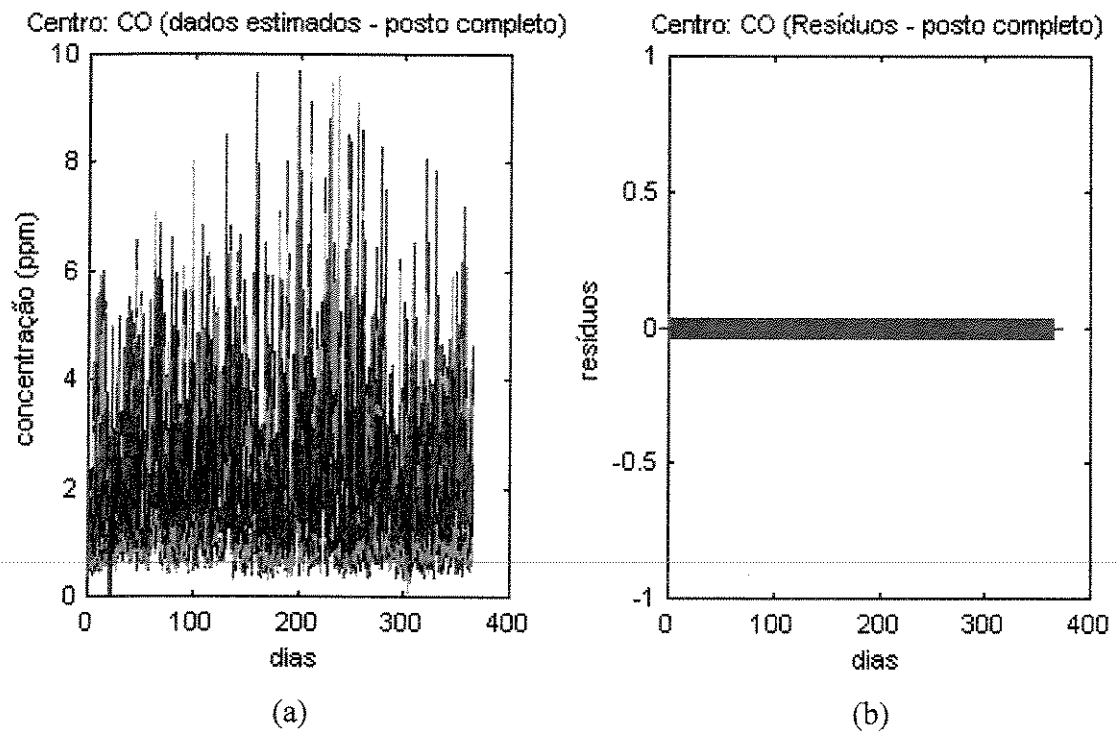


Fig.5.86: a)Dados estimados e b)Resíduos, considerando-se posto completo.

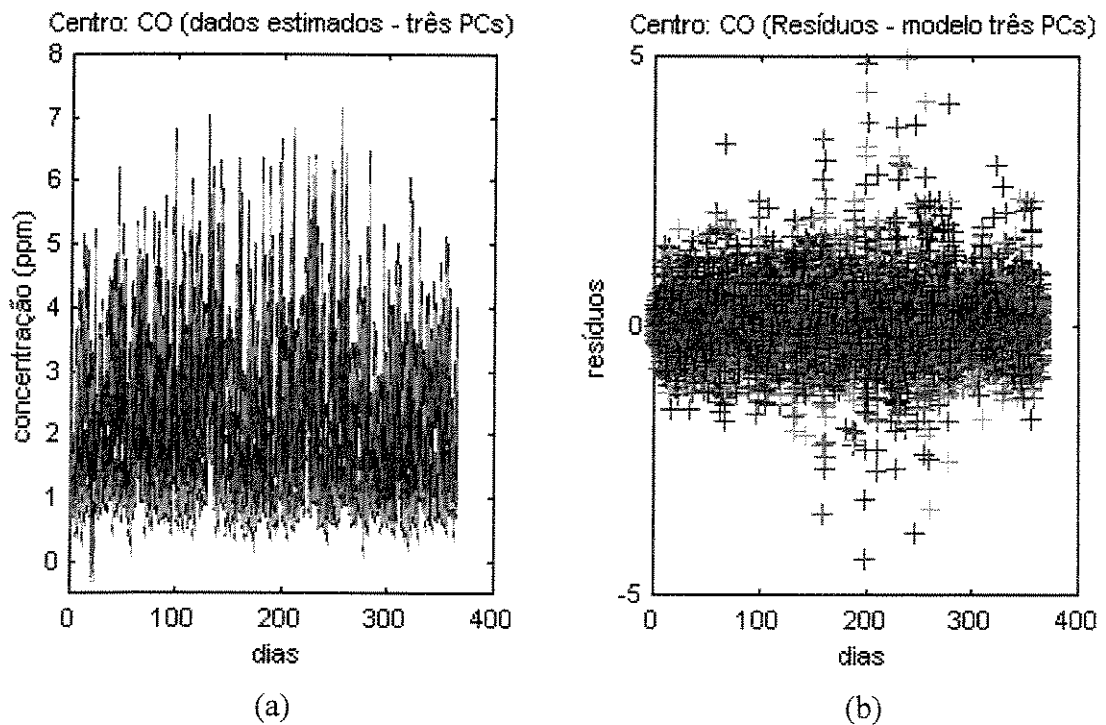


Fig.5.87: a) Dados estimados e b) Resíduos, considerando-se três PCs.

As Figuras 5.86 e 5.87 mostram os dados estimados e os resíduos.

5.1.2.6 – Análise PCA (CO – Lapa: 1998)

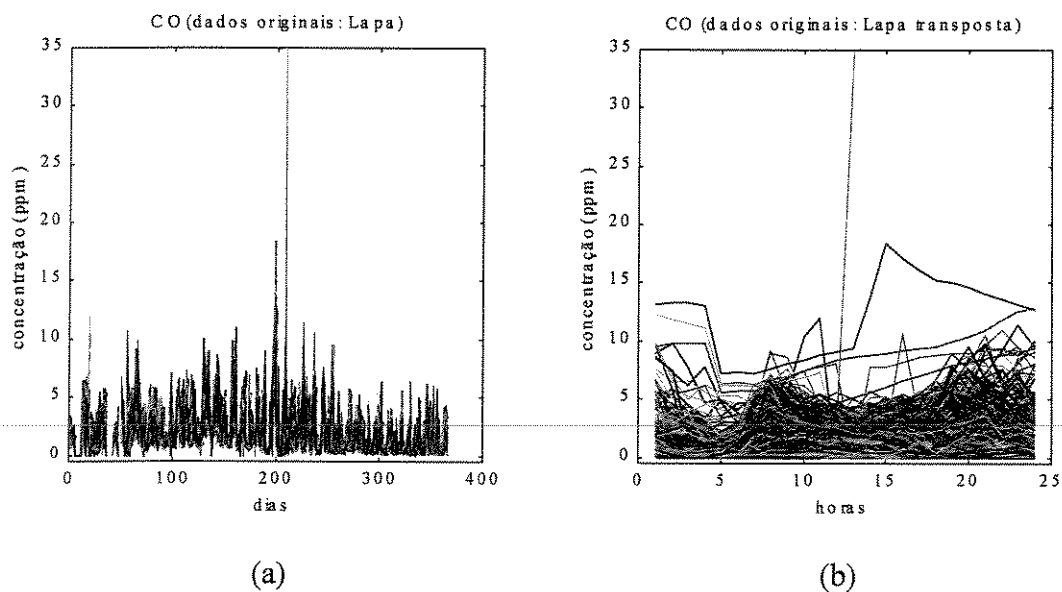


Fig.5.88: a)Gráfico concentração × dia. b) Gráfico concentração × hora.

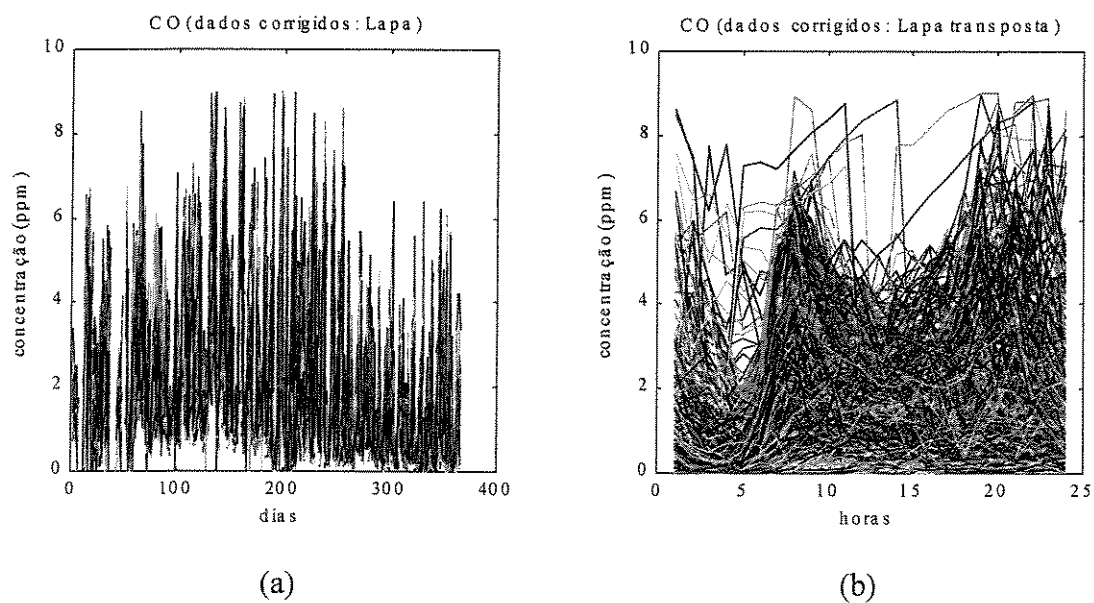


Fig.5.89: a)Gráfico concentração × dia. b) Gráfico concentração × hora.

As Figuras 5.88 e 5.89 mostram os dados estimados, respectivamente, sem e com correções. Em 5.89b nota-se que a correção para variações súbitas poderia ser melhorada, porém tais variações residuais não causaram problemas na análise.

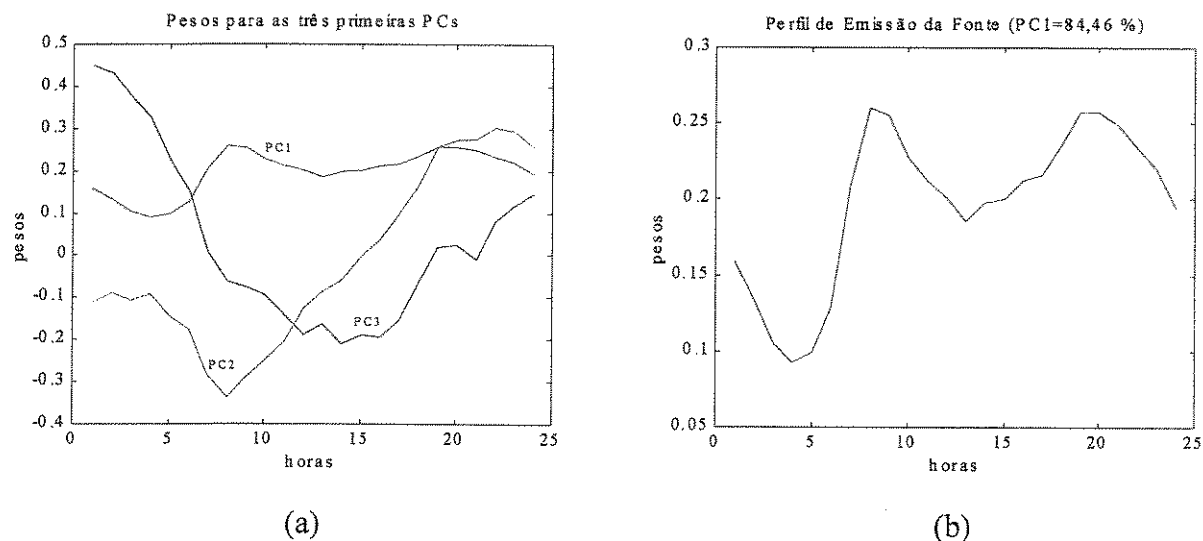


Fig.5.90: a)Pesos para PC1, PC2 e PC3. b)Pesos para PC1.

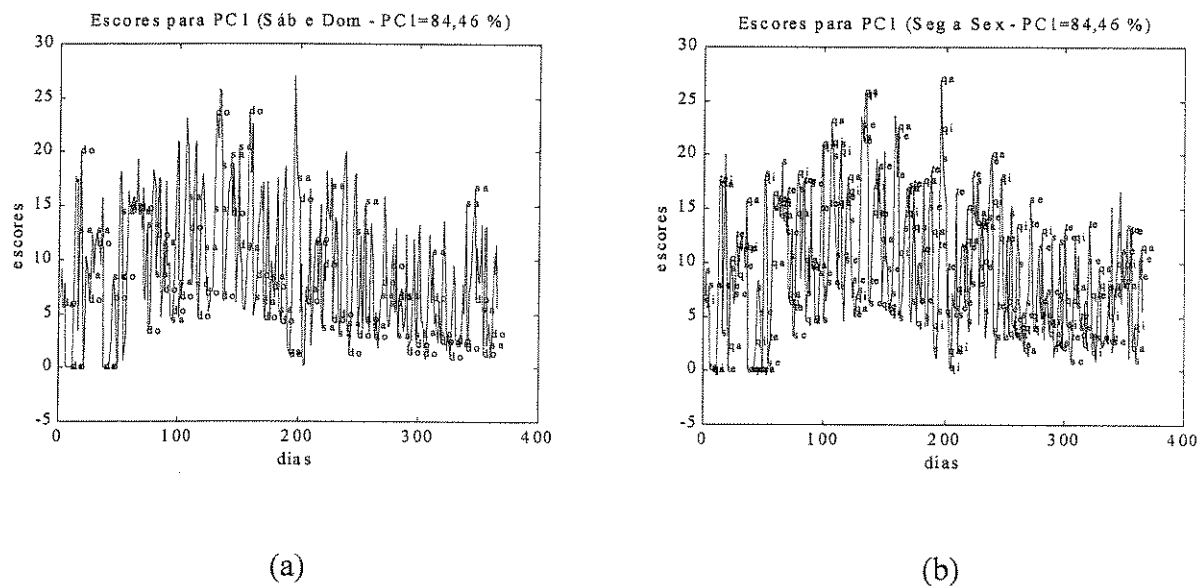


Fig.5.91: a)Escores para PC1 (finais de semana). b)Escores para PC1 (dias úteis).

A Figura 5.90 mostra os pesos para as componentes principais. Do mesmo modo que nas demais Estações de Medição ocorrem os dois picos característicos daquelas emissões (~9:00 h e ~20:00 h), como mostra a Figura 5.90b.

A Figura 5.91 mostra os escores para as componentes principais.

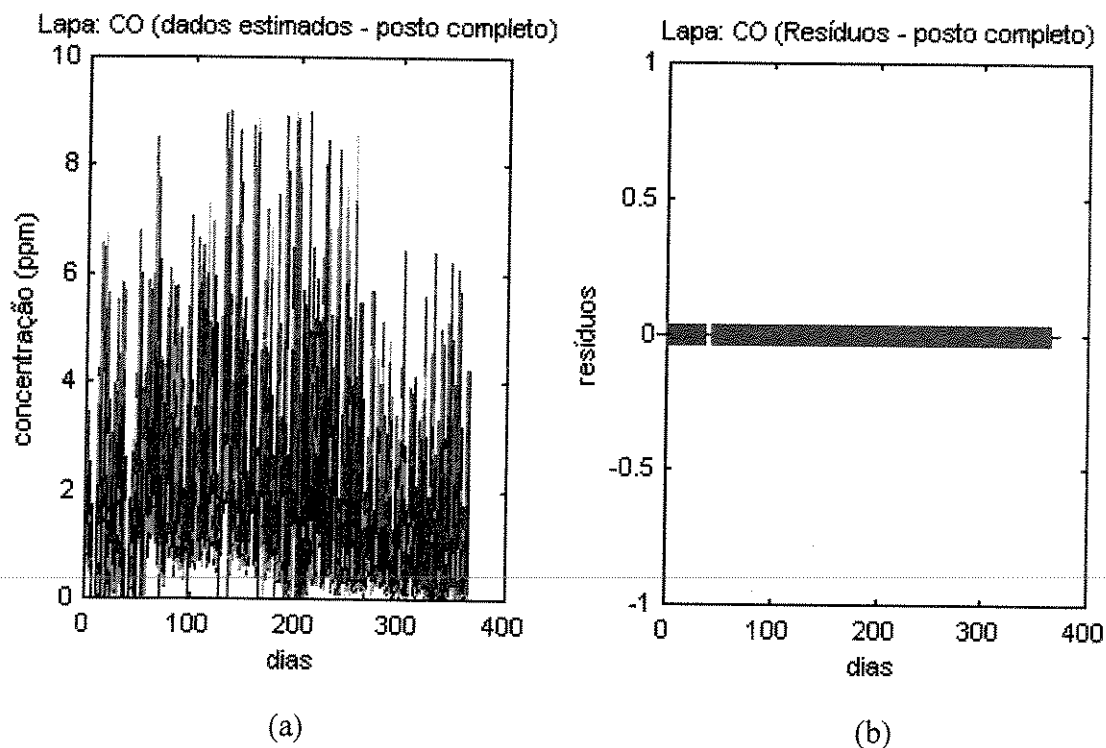


Fig.5.92: a)Dados estimados e b)Resíduos, considerando-se posto completo.

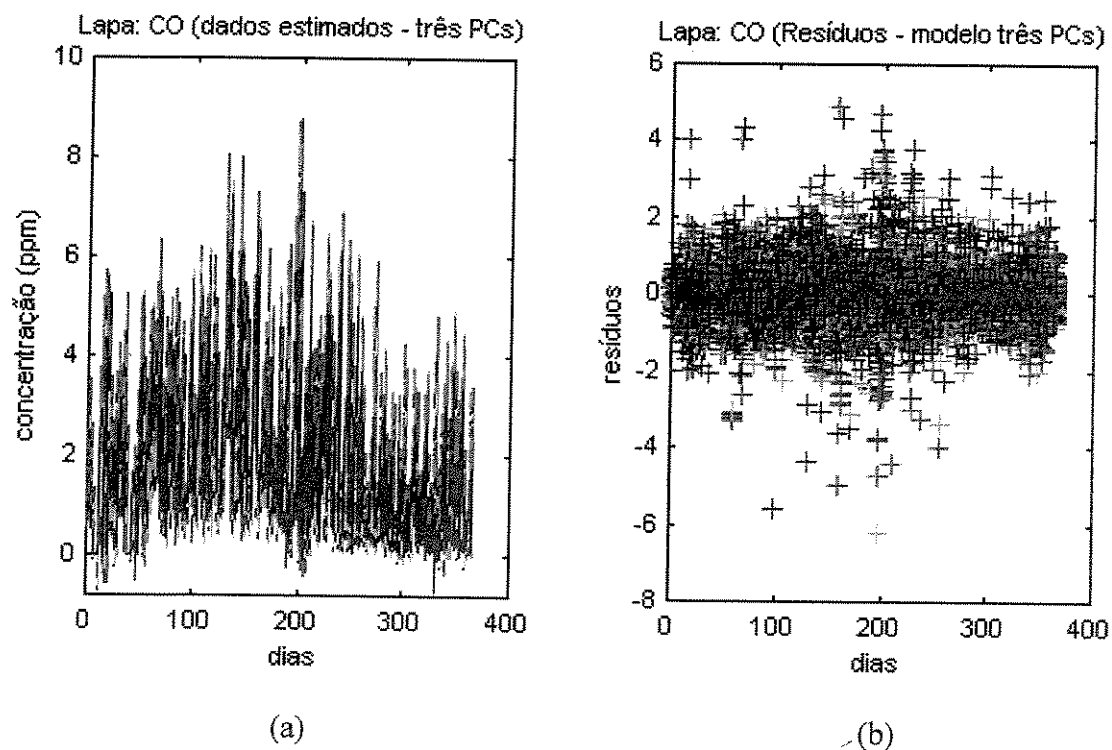


Fig.5.93: a) Dados estimados e b) Resíduos, considerando-se três PCs.

As Figuras 5.92 e 5.93 mostram os dados estimados e os resíduos.

5.1.2.7 – Análise PCA (CO – Osasco: 1998)

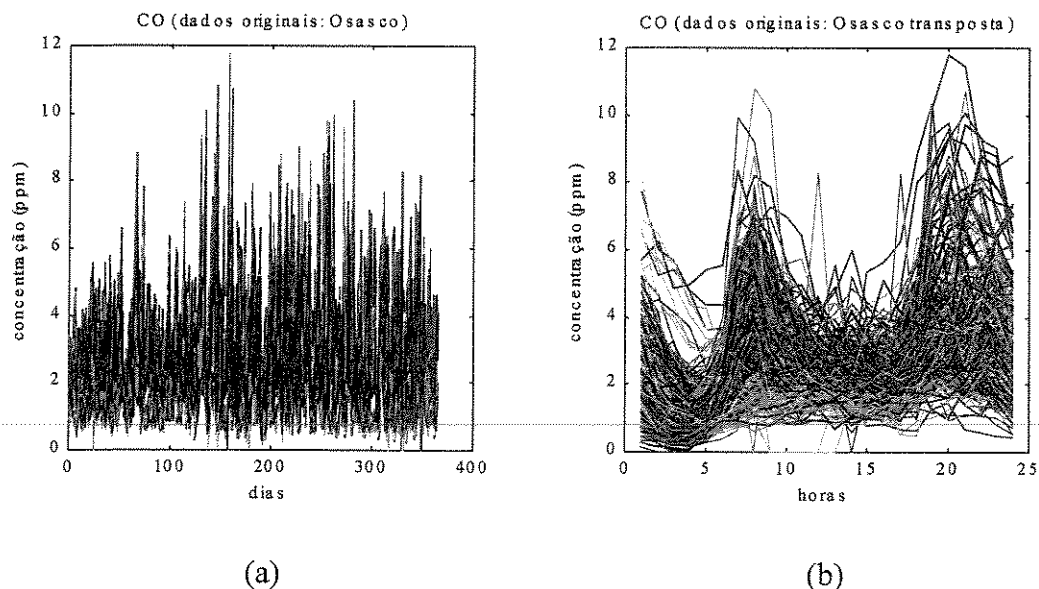


Fig.5.94: a) Gráfico concentração $\times$ dia. b) Gráfico concentração $\times$ hora.

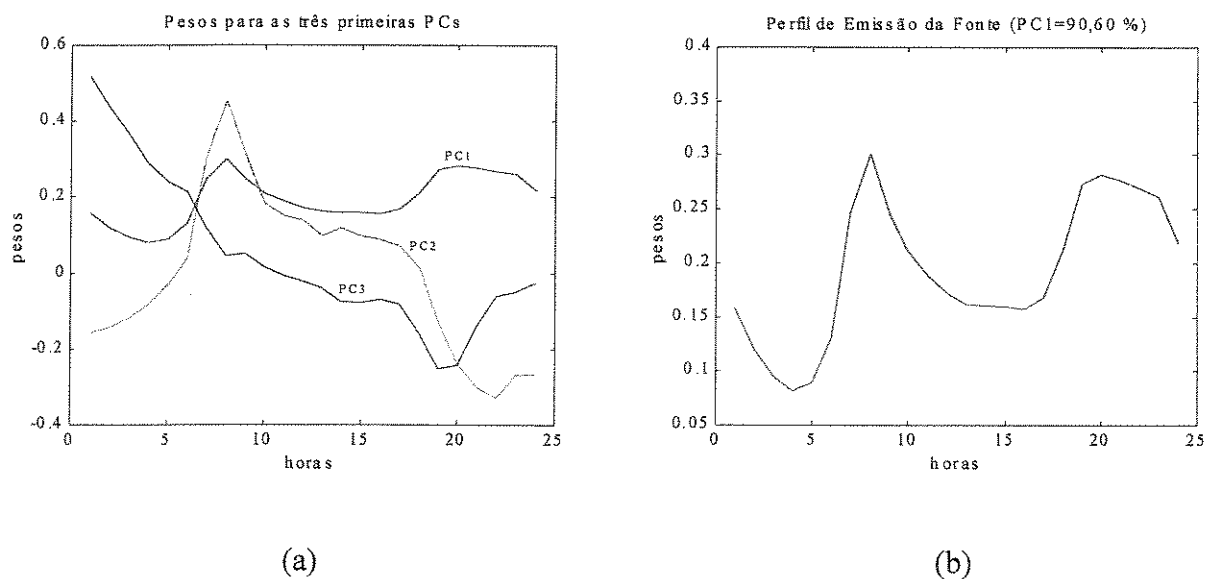
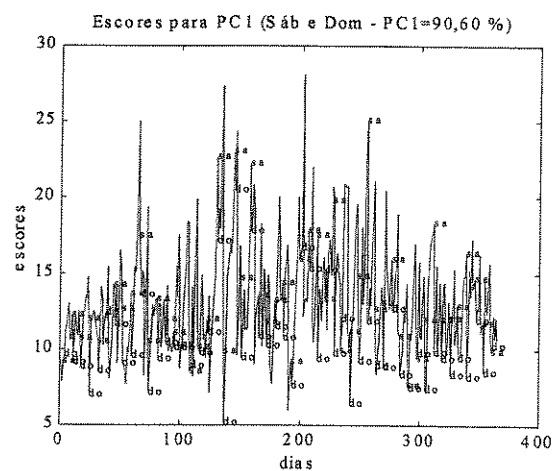
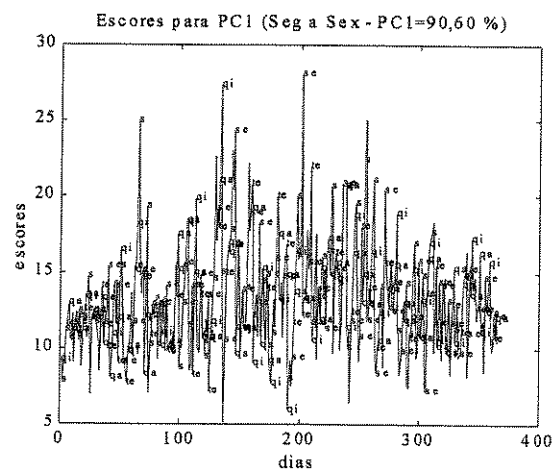


Fig.5.95: a) Pesos para PC1, PC2 e PC3. b) Pesos para PC1.

A Figura 5.94 mostra os dados originais (não houve necessidade de correções). A Figura 5.95 mostra os pesos para as componentes principais. Em 5.95b notam-se os dois picos característicos das emissões presentes nos outros casos.

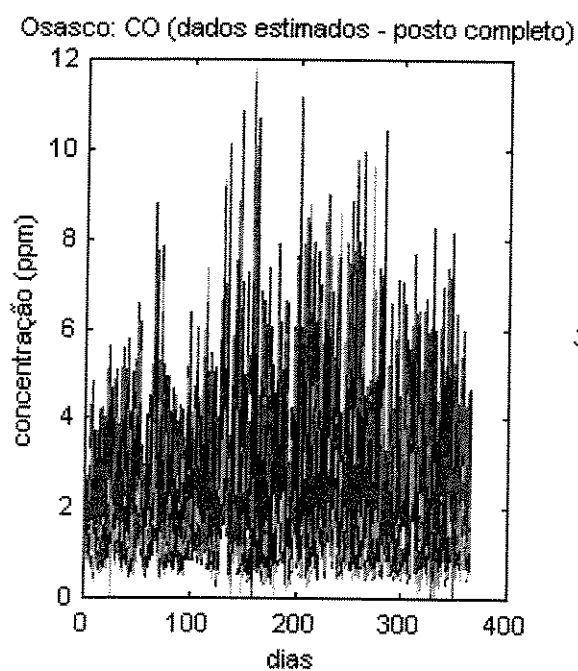


(a)

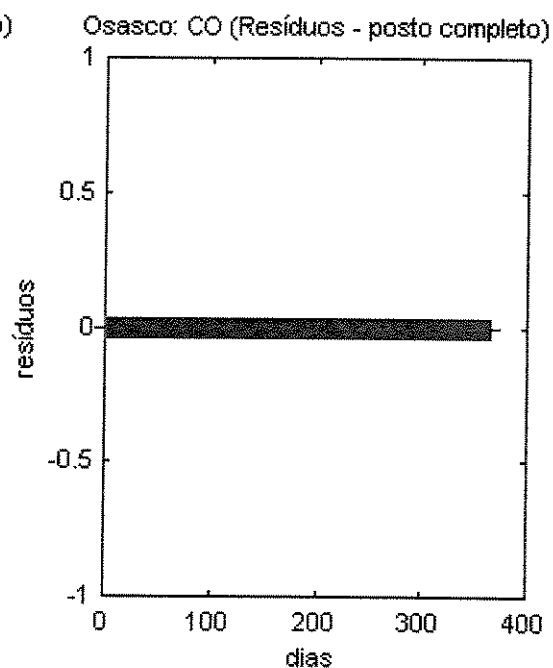


(b)

Fig.5.96: a)Escore para PC1 (finais de semana). b)Escore para PC1 (dias úteis).



(a)



(b)

Fig.5.97: a)Dados estimados e b)Resíduos, considerando-se posto completo.

A Figura 5.96 mostra os escores para as componentes principais. A Figura 5.97 mostra os dados originais e os resíduos considerando-se todas as PCs.

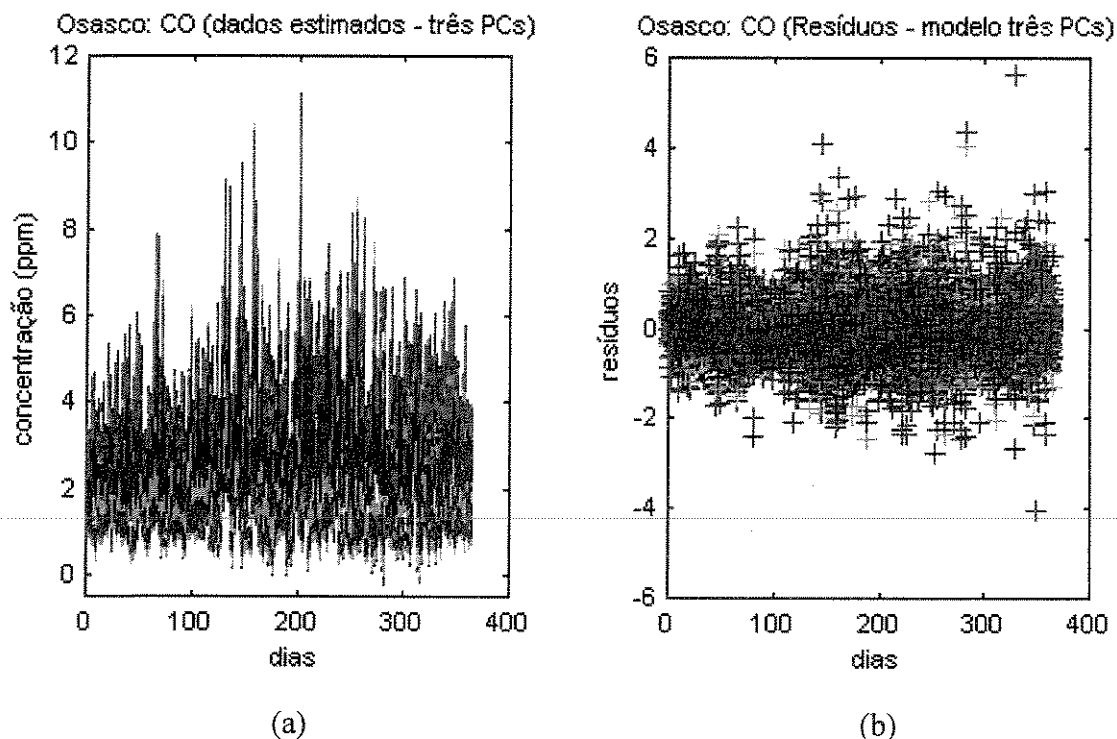


Fig.5.98: a) Dados estimados e b) Resíduos, considerando-se três PCs.

A Figura 5.98 mostra os dados estimados e os resíduos considerando-se apenas as três primeiras componentes principais.

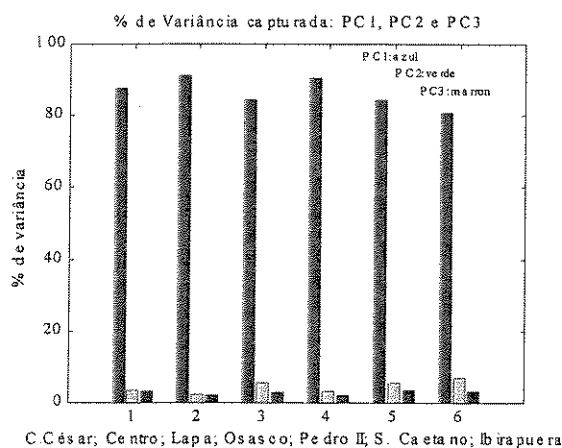


Fig.5.99: % de Variância capturada por PC₁, PC₂ e PC₃ nas sete Estações de Medição durante o ano de 1998.

Os resultados encontrados para o ano de 1998 nas 7 Estações de Medição são semelhantes aos obtidos para o ano de 1999. Do mesmo modo, como indica PC₁ na Figura 5.99, a principal fonte de poluição é a mesma em todos os casos.

5.1.3 – Análise PCA (1997 - 1999)

Nessa subseção são apresentados alguns dos resultados para algumas das variáveis, dentre as nove (CO, NO, NO₂, O₃, SO₂, CH₄, NMHC, Temperatura e Umidade) em algumas das sete Estações de Medição, para o período de 1997 a 1999.

São consideradas as localidades Cerqueira César, Centro, Lapa e Osasco, e as variáveis NO₂, NO, O₃ e SO₂, no período referido acima. Apenas os resultados das análises PARAFAC são apresentados.

Estação de Medição C. César: NO₂99, NO₂98 e NO₂97

NO99 e NO98

Estação de Medição Centro: NO₂99 e NO₂97

NO99

SO₂98

Estação de Medição Lapa: NO₂99

NO99

O₃97

Estação de Medição Osasco: NO₂99

NO99 e NO98

SO₂98

O₃97

A distribuição acima está posta em termos da Estação de Medição, porém para que se possa comparar o comportamento de cada poluente em um dado ano em uma dada localidade, os resultados são apresentados segundo a hierarquia poluente ano localidade. Assim o primeiro bloco de resultados contém a variável NO₂, durante os três anos e nas localidades consideradas, conforme abaixo:

C. César (NO₂99, NO₂98 e NO₂97),

Centro (NO₂99 e NO₂97),

Lapa (NO₂99),

Osasco (NO₂99).

Estação de Medição C. César (NO₂99, NO₂98 e NO₂97)
(NO₂99 – C. César)

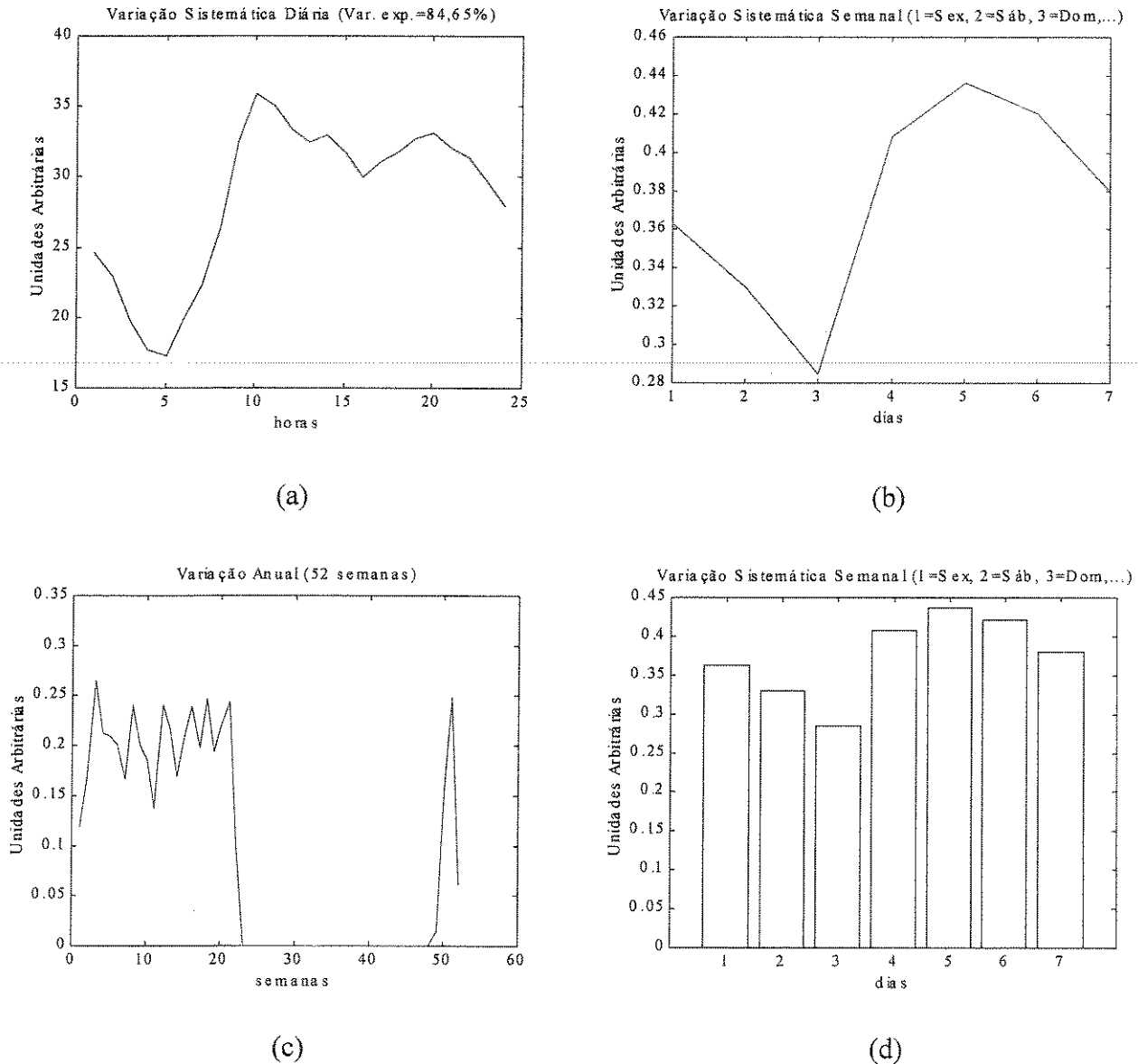


Fig.5.100: a)Variações sistemáticas com periodicidade diária. b) e d)Variações sistemáticas com periodicidade semanal. c)Variação anual. Dados para C. César – 1999.

Em 5.100a nota-se a existência dos dois picos característicos de emissões obtidos nos casos anteriores, porém com um vale menos pronunciado às 14:00 h. Em 5.100b (e d) nota-se a queda de emissões nos finais de semana. Em 5.100c, a avaliação para a variação anual fica prejudicada devido à grande quantidade de valores originalmente nulos (intervalo entre ~20^a e ~48^a semanas).

(NO₂98 – C. César)

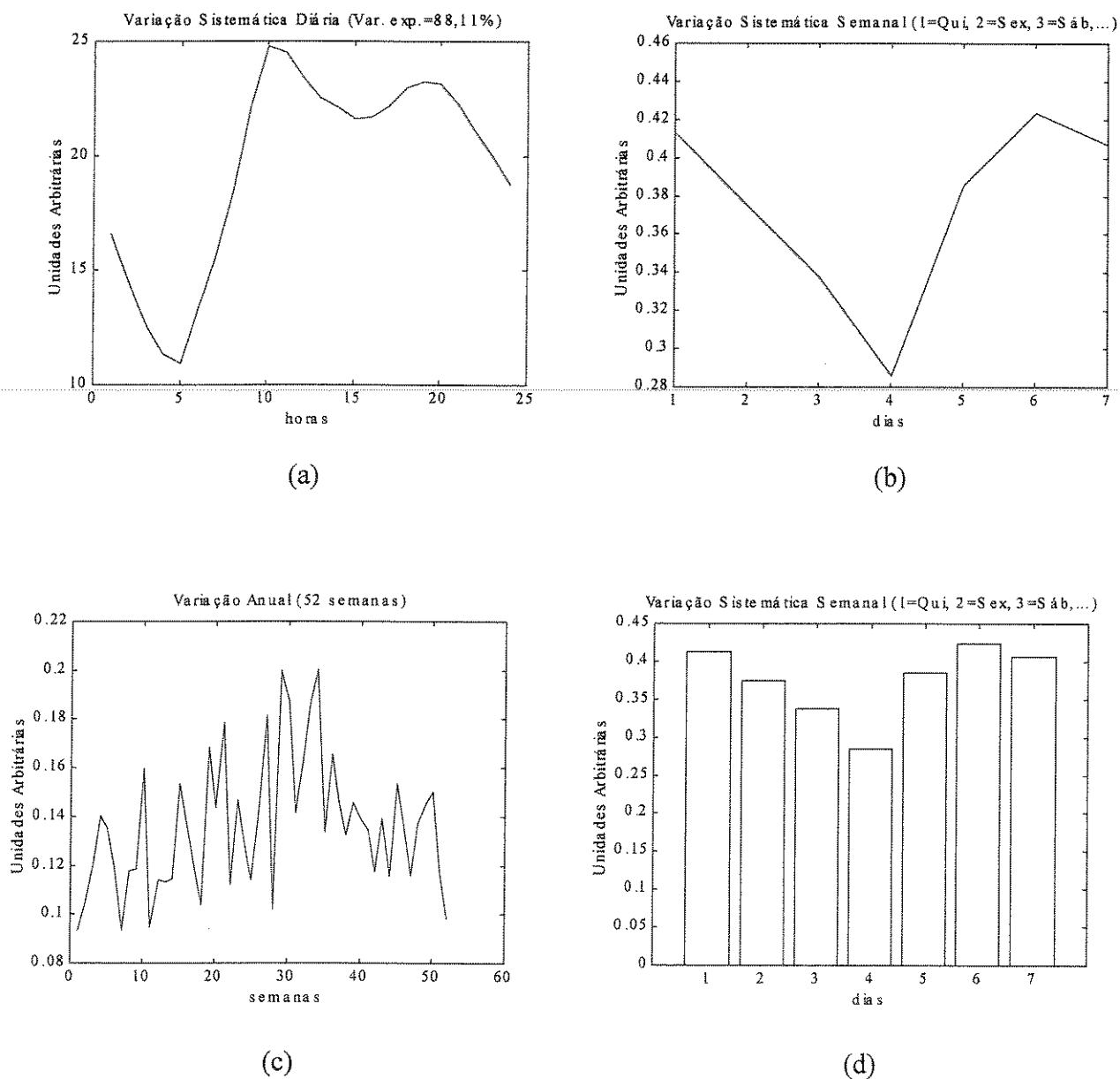


Fig.5.101: a)Variações sistemáticas com periodicidade diária. b)e d)Variações sistemáticas com periodicidade semanal. c)Variação anual. Dados para C. César – 1998.

Em 5.101a pode-se observar os mesmos picos característicos dos outros casos. Em 5.101b (e d) nota-se a queda nas emissões nos fins de semana. Aqui, deve ser observado que o 1º ponto é uma Quinta-feira, logo o dia assinalado com o número 4 no gráfico, é um domingo. Em 5.101c observa-se a maior acumulação no período de inverno.

(NO₂97 – C. César)

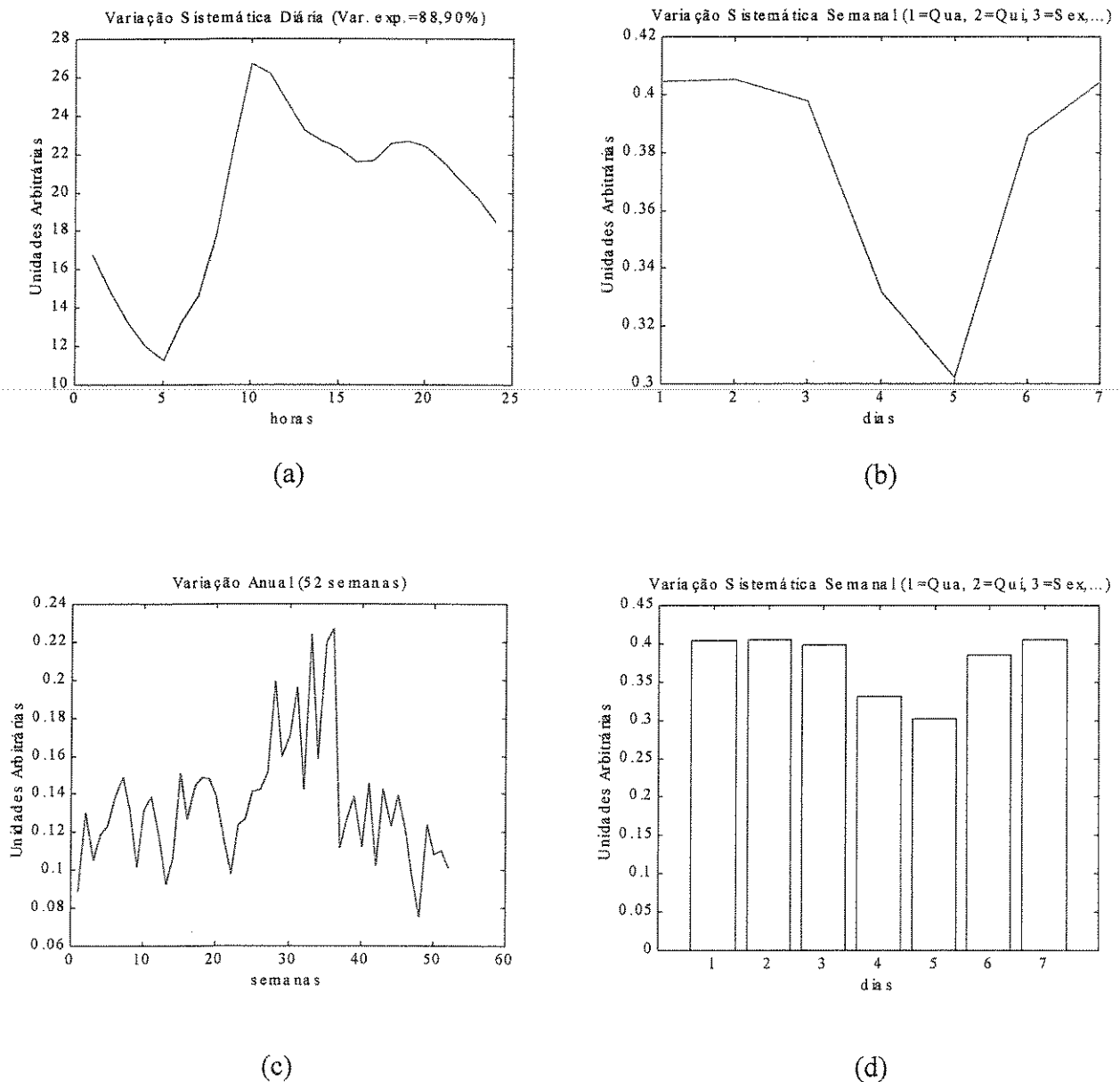


Fig.5.102: a)Variações sistemáticas com periodicidade diária. b) e d)Variações sistemáticas com periodicidade semanal. c)Variação anual. Dados para C. César – 1997.

Em 5.102a observam-se os mesmos picos anteriores. Aqui tem-se um 2º pico de menor intensidade. Em 5.102b (e d) nota-se a queda nos finais de semana. Também aqui deve ser notado que o Domingo é assinalado no gráfico como o 5º dia porque os dados referem-se ao ano de 1997. Em 5.102c nota-se a maior acumulação no período de inverno.

**Estação de Medição Centro (NO₂99 e NO₂97)
(NO₂99 – Centro)**

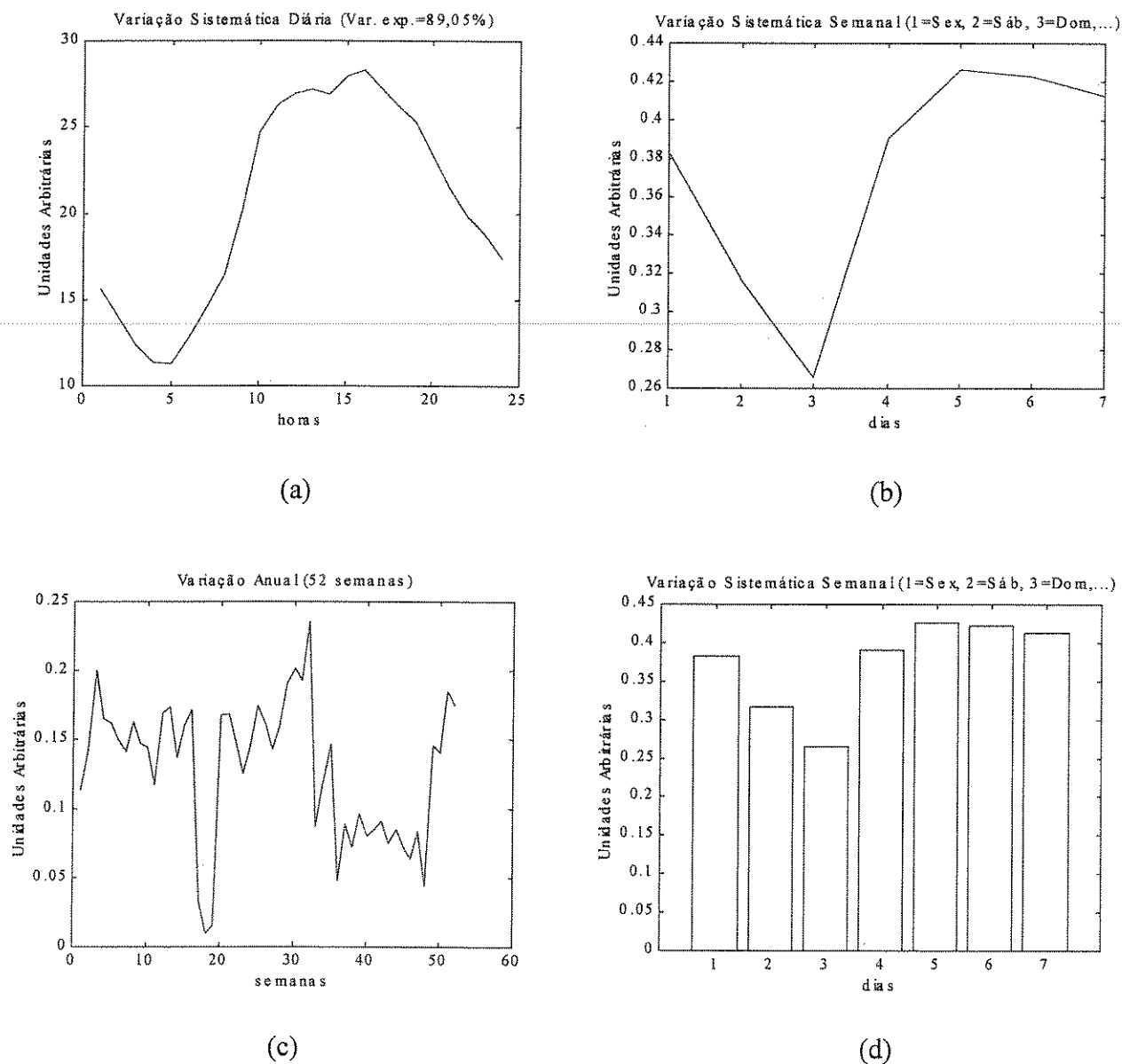
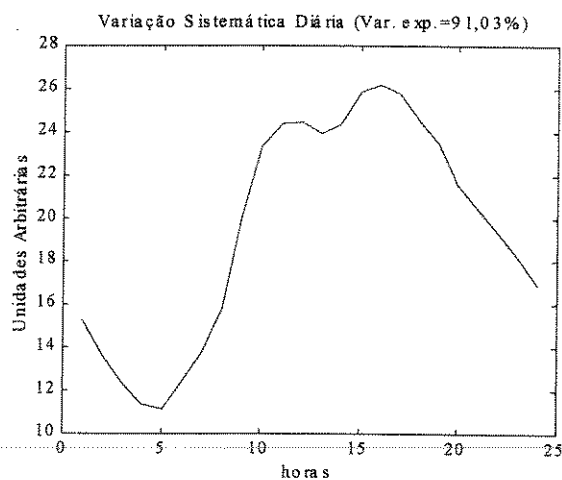


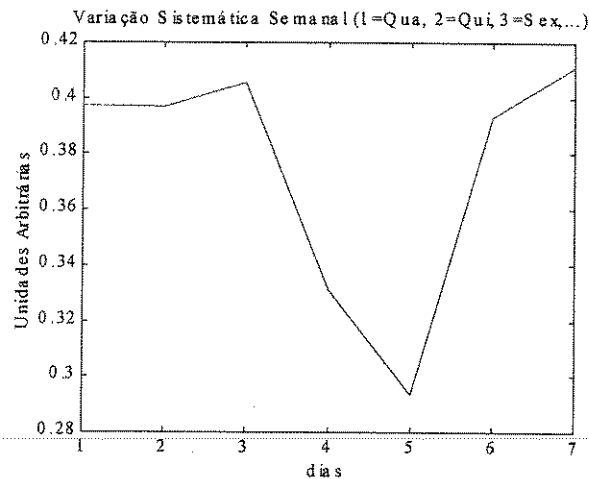
Fig.5.103: a)Variações sistemáticas com periodicidade diária. b) e d)Variações sistemáticas com periodicidade semanal. c)Variação anual. Dados para Centro – 1999.

Em 5.103a quase não se percebem os picos característicos das emissões anteriores tal a contínua emissão de poluentes no local. Em 5.103b (e d) nota-se a queda das emissões nos fins de semana. Em 5.103c, problemas com os dados originais causam quedas bruscas em dois períodos, prejudicando uma melhor avaliação da variação anual.

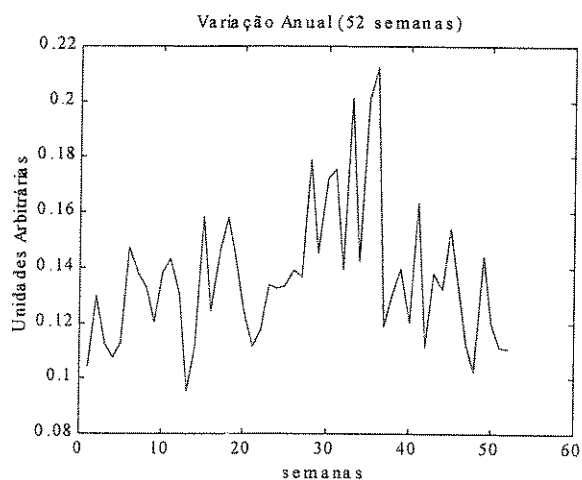
(NO₂97 – Centro)



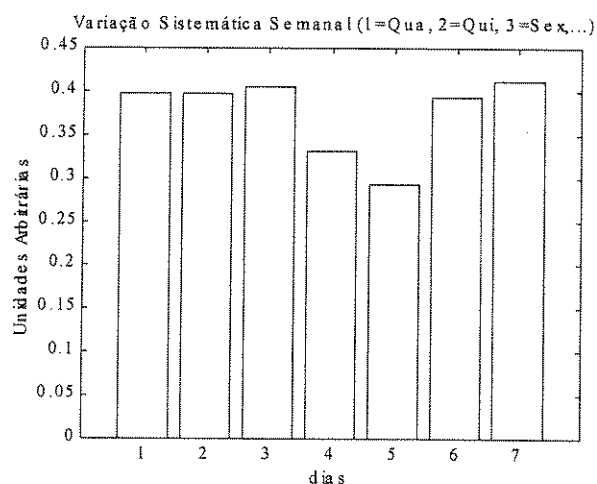
(a)



(b)



(c)



(d)

Fig.5.104: a)Variações sistemáticas com periodicidade diária. b) e d)Variações sistemáticas com periodicidade semanal. c)Variação anual. Dados para Centro – 1997.

Como no caso anterior, em 5.104a, quase não se percebem os picos característicos das emissões predominantes. Em 5.104b (e d) nota-se a queda nos fins de semana. Em 5.104c nota-se a maior acumulação no inverno.

Estação de Medição Lapa (NO₂99) (NO₂99 – Lapa)

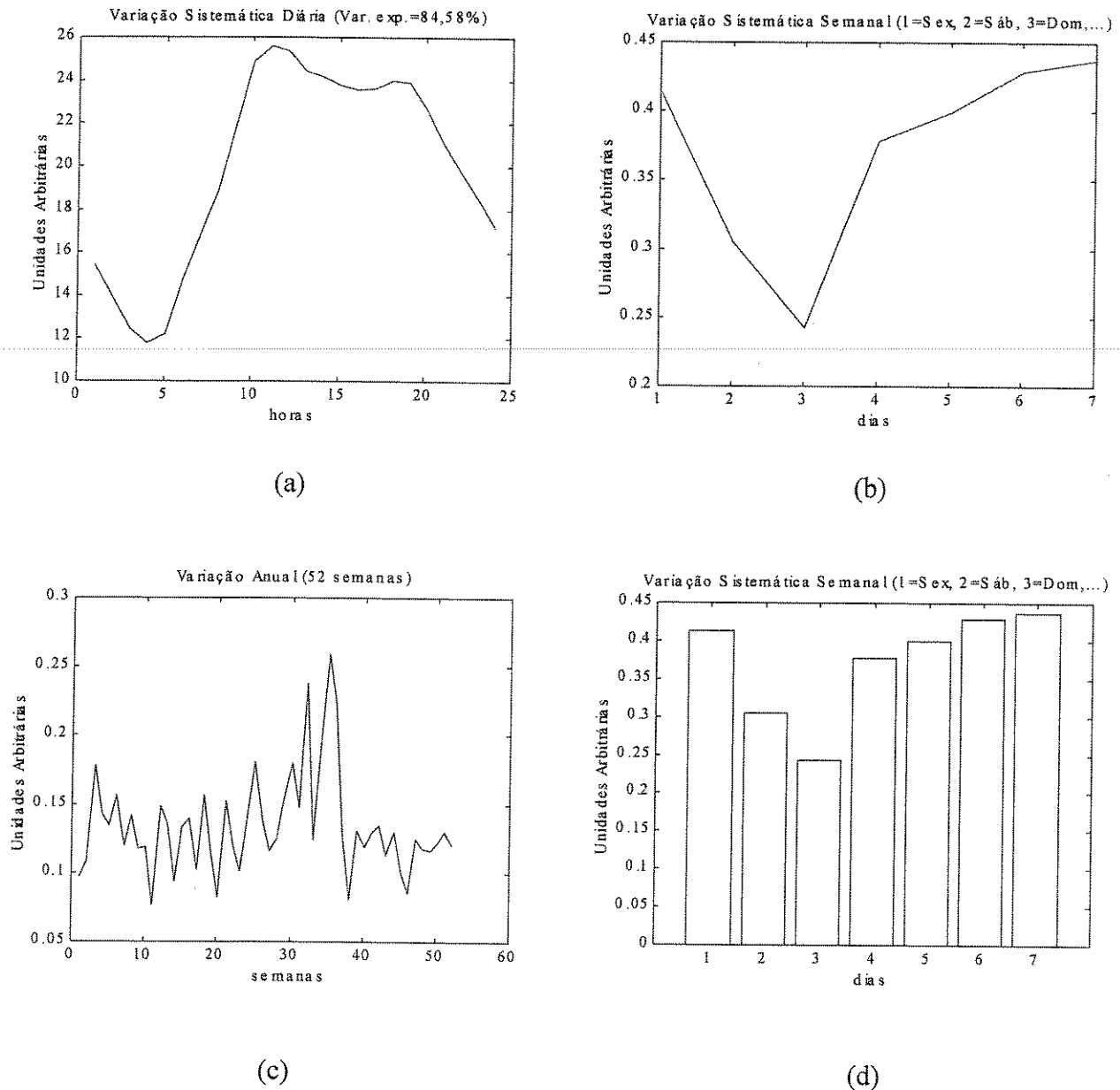


Fig.5.105: a)Variações sistemáticas com periodicidade diária. b) e d)Variações sistemáticas com periodicidade semanal. c)Variação anual. Dados para Lapa – 1999.

Em 5.105a o 2º pico característico de emissões da fonte é menos intenso e o vale é menos pronunciado. Em 5.105b (e d) nota-se a queda de emissão nos fins de semana. Em 5.105c observa-se maior acumulação no período de inverno.

Estação de Medição Osasco (NO₂99) (NO₂99 – Osasco)

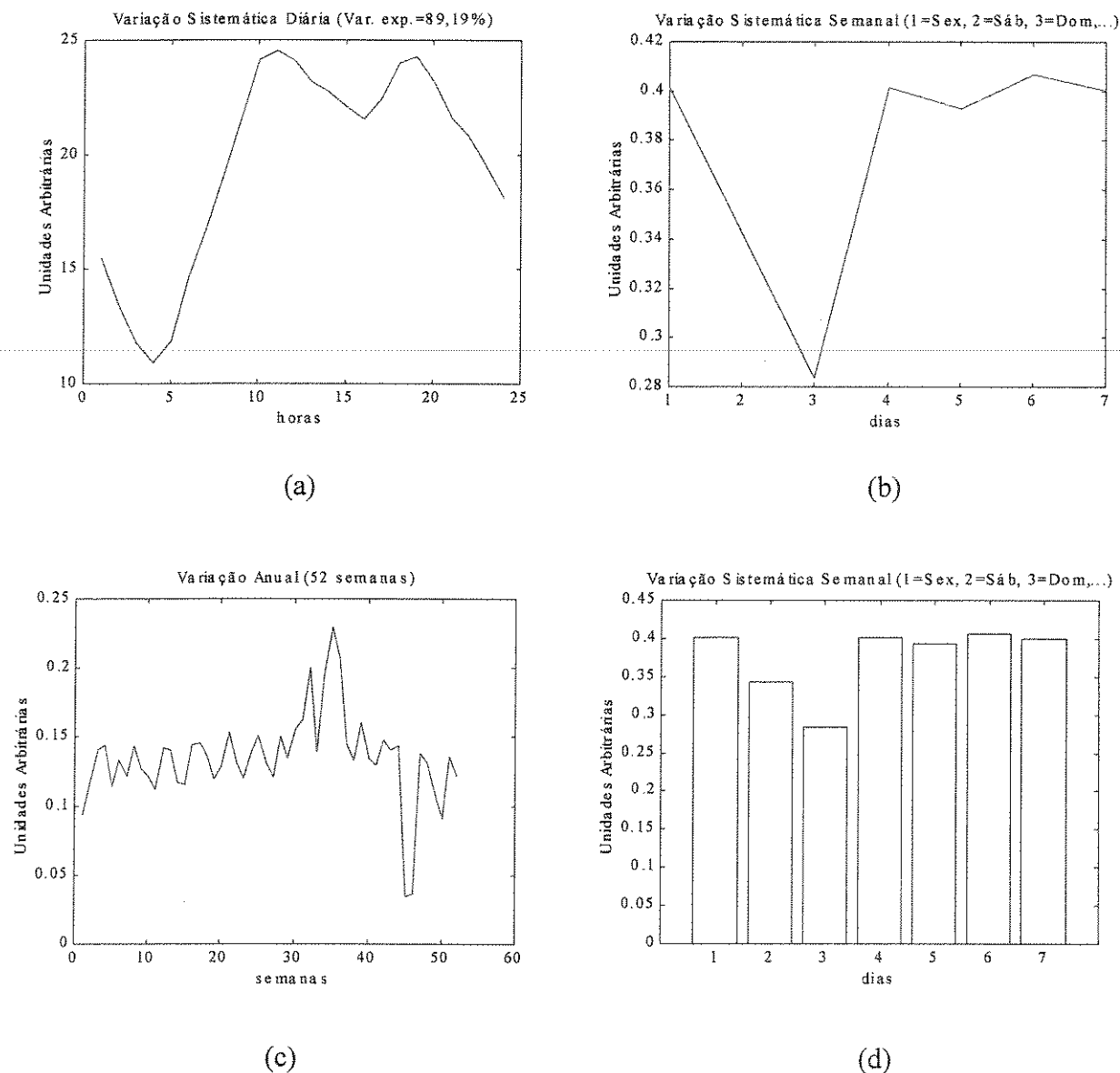


Fig.5.106: a)Variações sistemáticas com periodicidade diária. b) e d)Variações sistemáticas com periodicidade semanal. c)variação anual. Dados para Osasco – 1999.

Em 5.106a, notam-se os dois picos característicos das emissões predominantes. Em 5.106b e (d), nota-se a queda das emissões nos finais de semana. Em 5.106c nota-se maior acumulação no inverno.

**Estação de Medição C. César (NO99 e NO98)
(NO99 – C. César)**

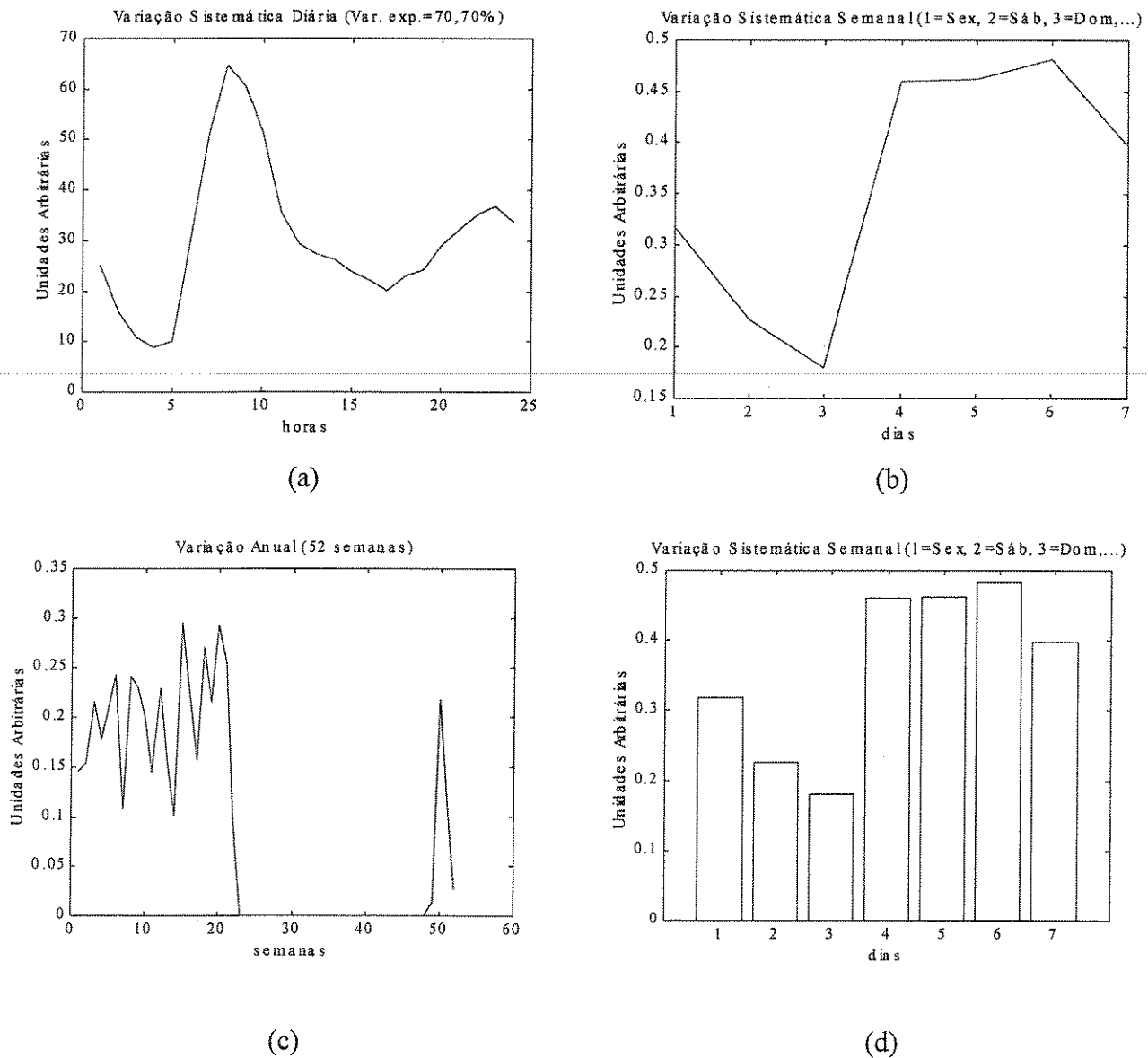
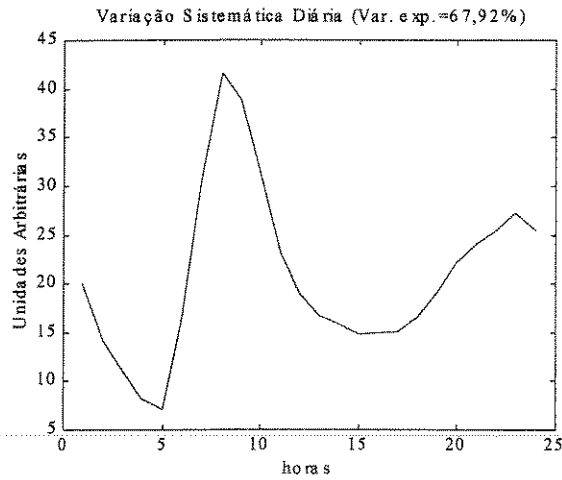
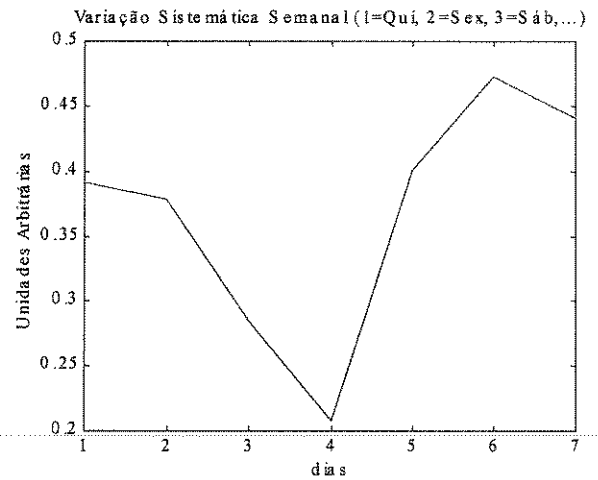


Fig.5.107: a)Variações sistemáticas com periodicidade diária. b) e d)Variações sistemáticas com periodicidade semanal. c)Variação anual. Dados para C. César – 1999)

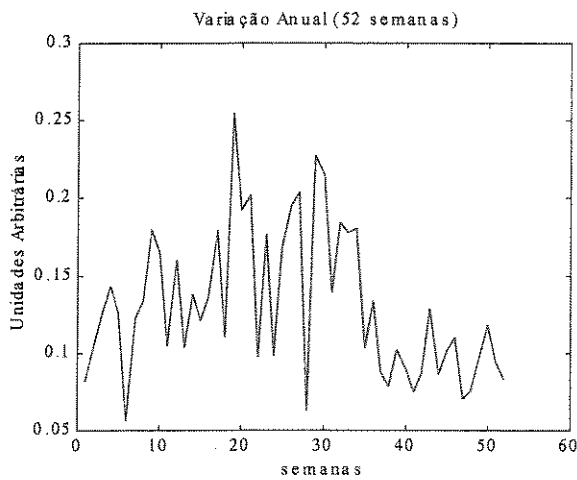
Em 5.107a nota-se o mesmo padrão de emissão. Em 5.107b (e d) nota-se a queda de emissões nos finais de semana. Em 5.107c não pode ser observada maior acumulação no período de inverno. Como aconteceu com o “NO₂99 – C. César”, também aparecem os valores originalmente nulos entre 23^a e 48^a semanas (deve ser lembrado que o NO₂ e NO são medidos e registrados pelo mesmo instrumento).



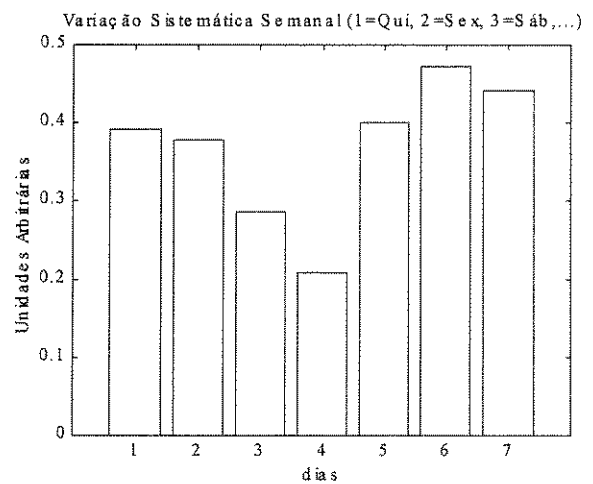
(a)



(b)



(c)



(d)

Fig.5.108: a)Variações sistemáticas com periodicidade diária. b) e d)Variações sistemáticas com periodicidade semanal. c)Variação anual. Dados para C. César – 1998.

Em 5.108a notam-se os picos característicos das emissões anteriores. Em 5.108b (e d) nota-se a queda das emissões nos finais de semana. Em 5.108c nota-se a maior acumulação no período de inverno.

Estação de Medição Centro (NO99) (NO99 – Centro)

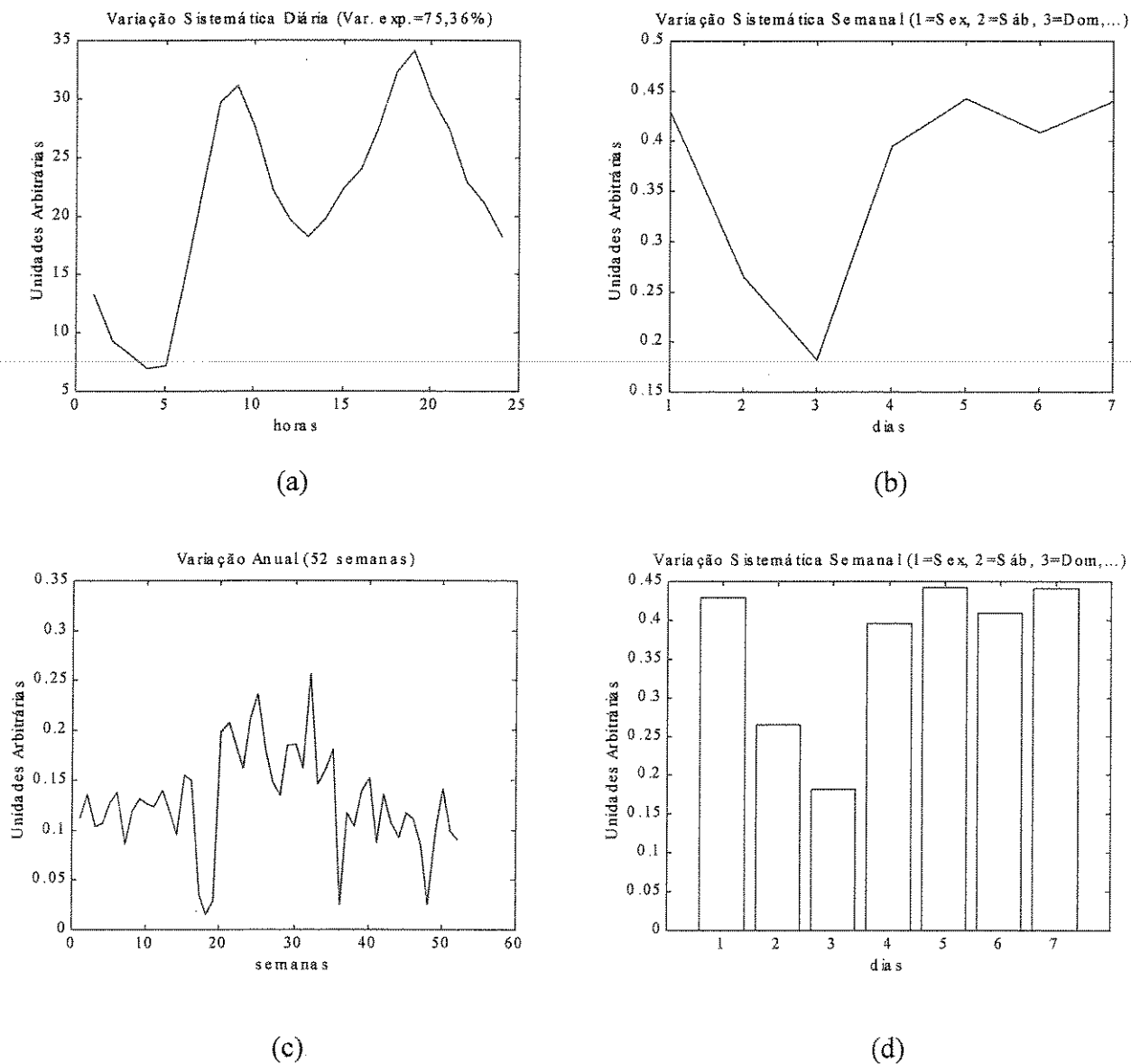


Fig.5.109: a)Variações sistemáticas com periodicidade diária. b) e d)Variações sistemáticas com periodicidade semanal. c)Variação anual. Dados para Centro – 1999.

Em 5.109a notam-se os picos característicos de emissões anteriores. Em 5.109b (e d) nota-se a queda de emissões nos finais de semana. Em 5.109c nota-se maior acumulação nos finais de semana.

Estação de Medição Lapa (NO99) (NO99 – Lapa)

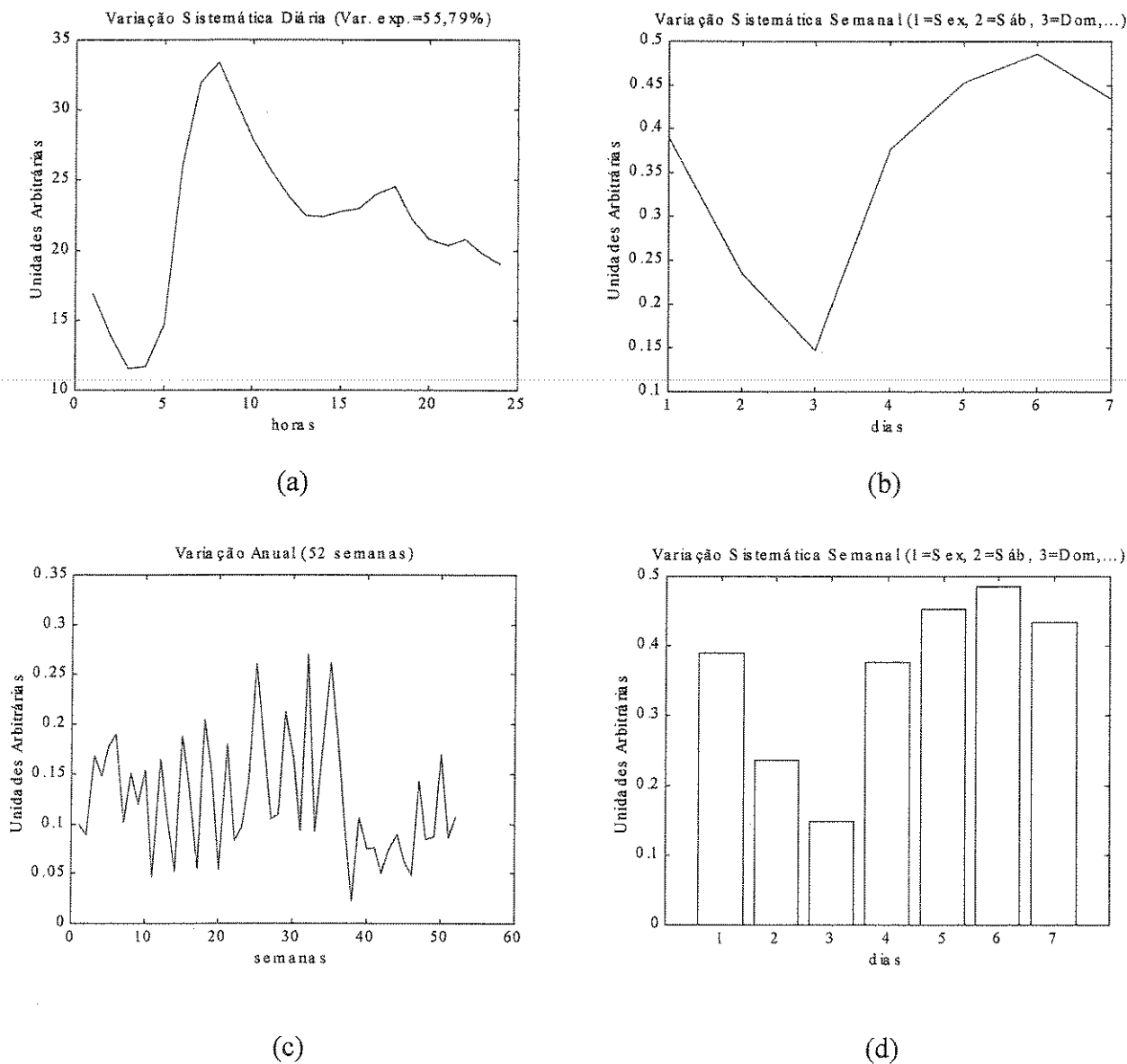


Fig.5.110: a)Variações sistemáticas com periodicidade diária. b) e d)Variações sistemáticas com periodicidade semanal. c)Variação anual. Dados para Lapa – 1999.

Em 5.110a, nota-se o mesmo padrão de emissão com os picos característicos. Em 5.110b (e d) nota-se a queda de emissões nos finais de semana. Em 5.110c nota-se maior acumulação do poluente.

Estação de Medição Osasco (NO99 e NO98)
(NO99 – Osasco)

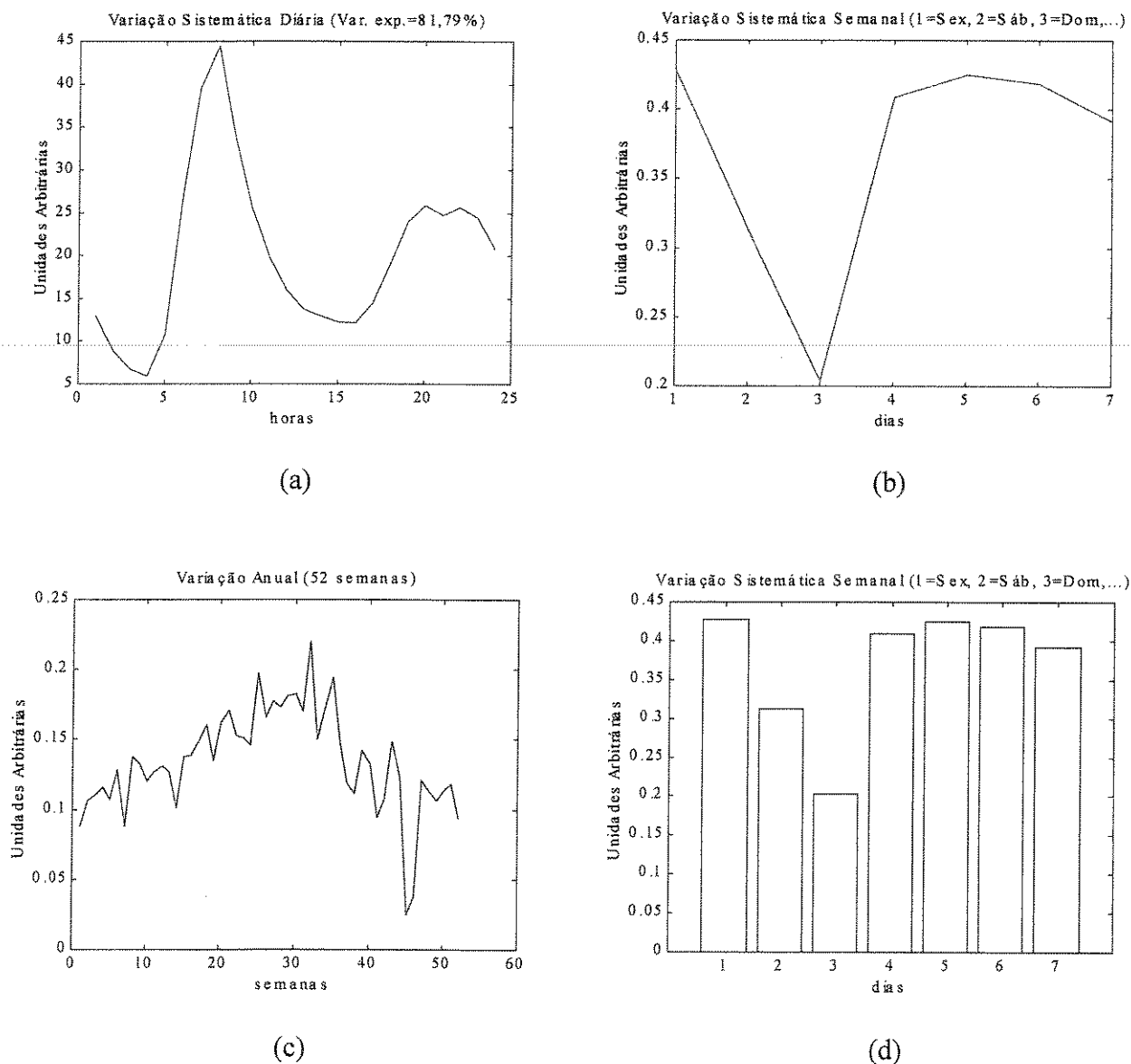


Fig.5.111: a)Variações sistemáticas com periodicidade diária. b) e d)Variações sistemáticas com periodicidade semanal. c)Variação anual. Dados para Osasco – 1999.

Em 5.111a nota-se o mesmo padrão de emissão com os picos característicos dos outros casos. Em 5.111b nota-se a queda da emissões nos fins de semana. Em 5.111c pode ser observada a maior acumulação no período de inverno.

(NO98 – Osasco)

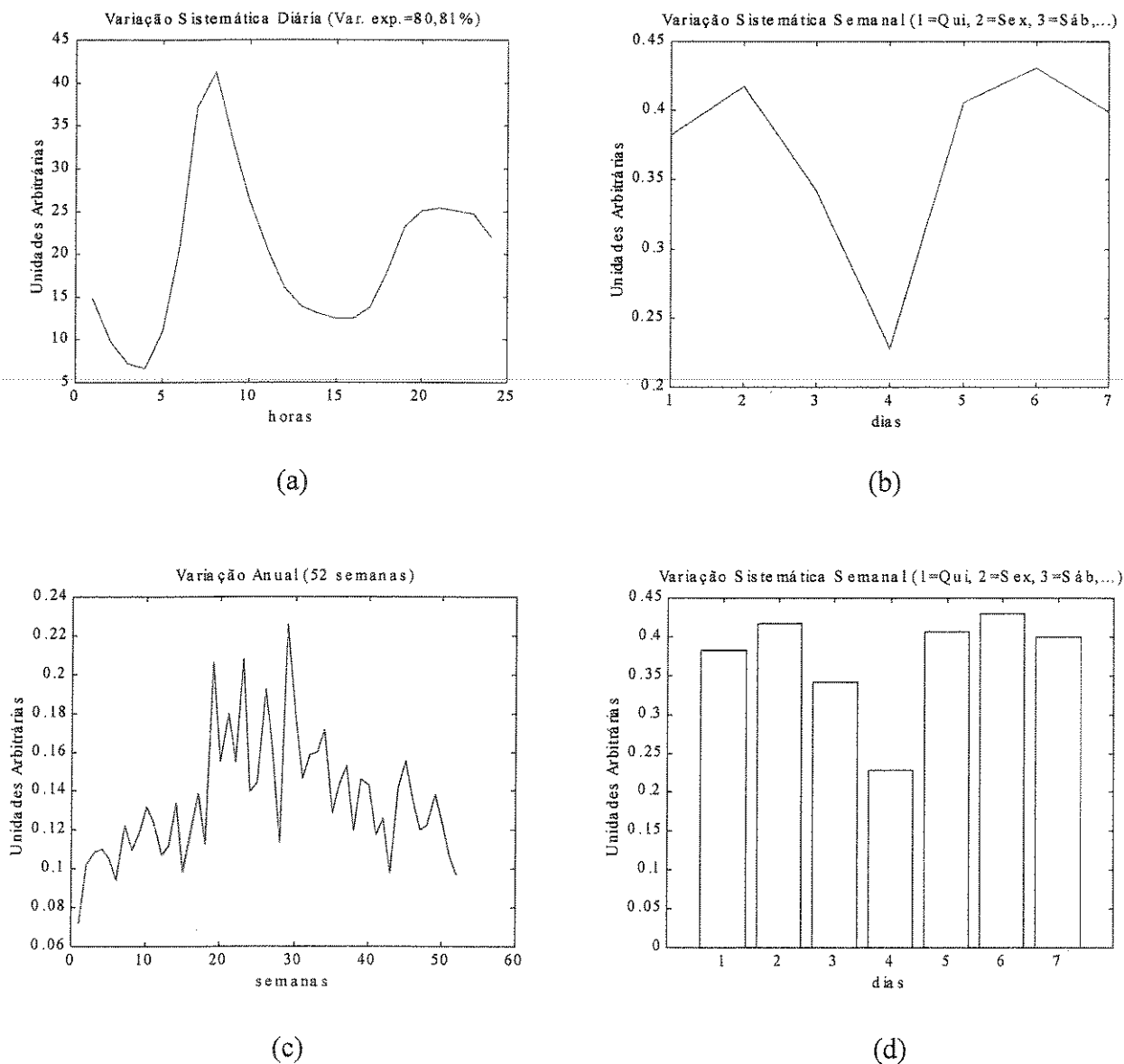


Fig.5.112: a)Variações sistemáticas com periodicidade diária. b) e d)Variações sistemáticas com periodicidade semanal. c)Variação anual. Dados para Osasco – 1998.

Em 5.112a notam-se os picos característicos encontrados nas emissões em outros casos. Em 5.112b (e d) pode ser observada a queda de emissões nos finais de semana. Em 5.112c nota-se maior acumulação no período de inverno.

Estação de Medição Lapa (O₃97) (O₃97 – Lapa)

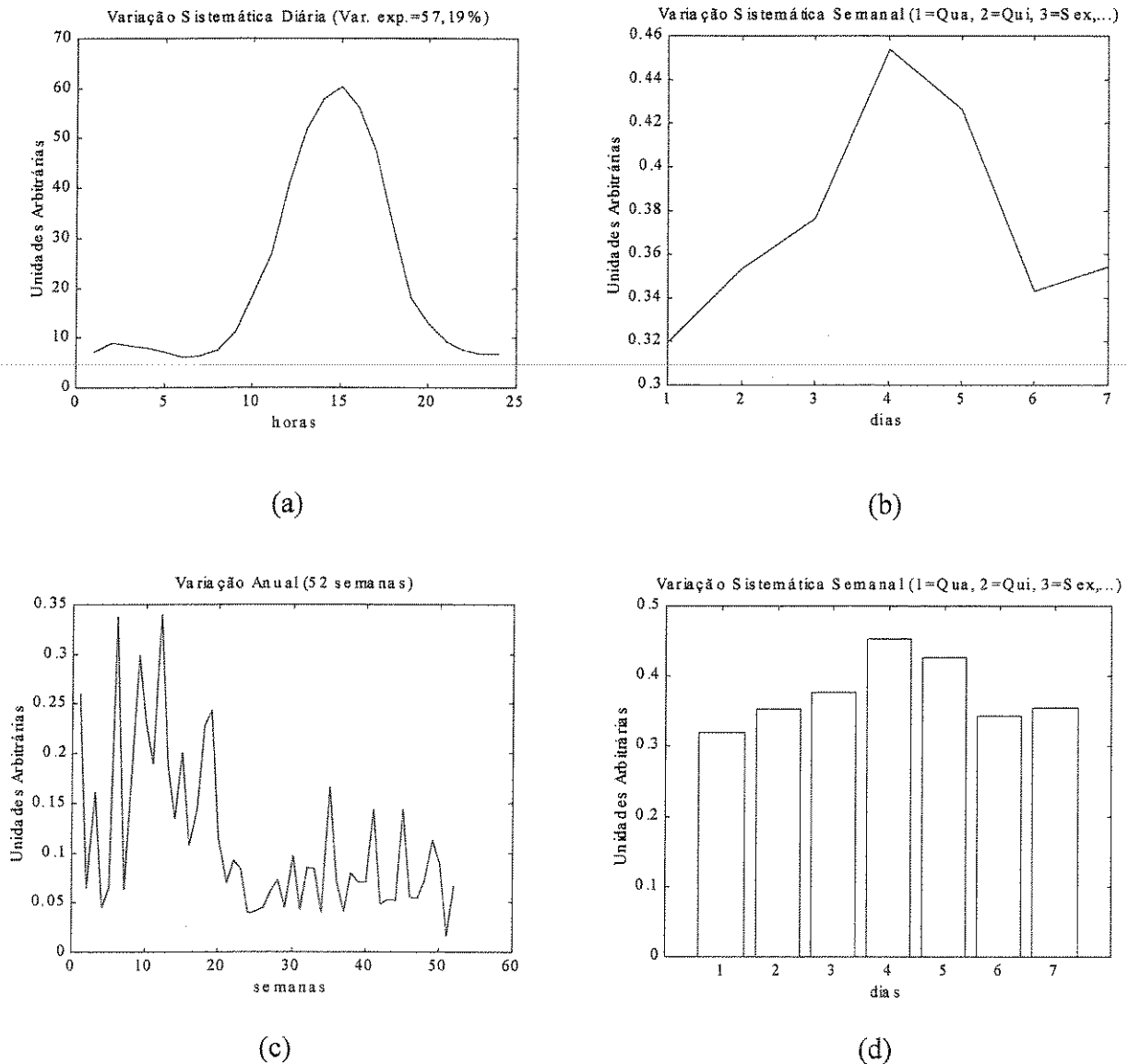


Fig.5.113: a)Variações sistemáticas com periodicidade diária. b) e d)Variações sistemáticas com periodicidade semanal. Variação anual. Dados para Lapa – 1997.

Em 5.113a pode ser observado o perfil característico do ozônio apresentado em todas as localidades. Em 5.113b (e d) nota-se o aumento dos níveis nos finais de semana. Em 5.113c observa-se a variação anual com queda pronunciada no inverno resultando em menor acumulação nesse período.

**Estação de Medição Osasco (O₃97 e SO₂98)
(O₃97 – Osasco)**

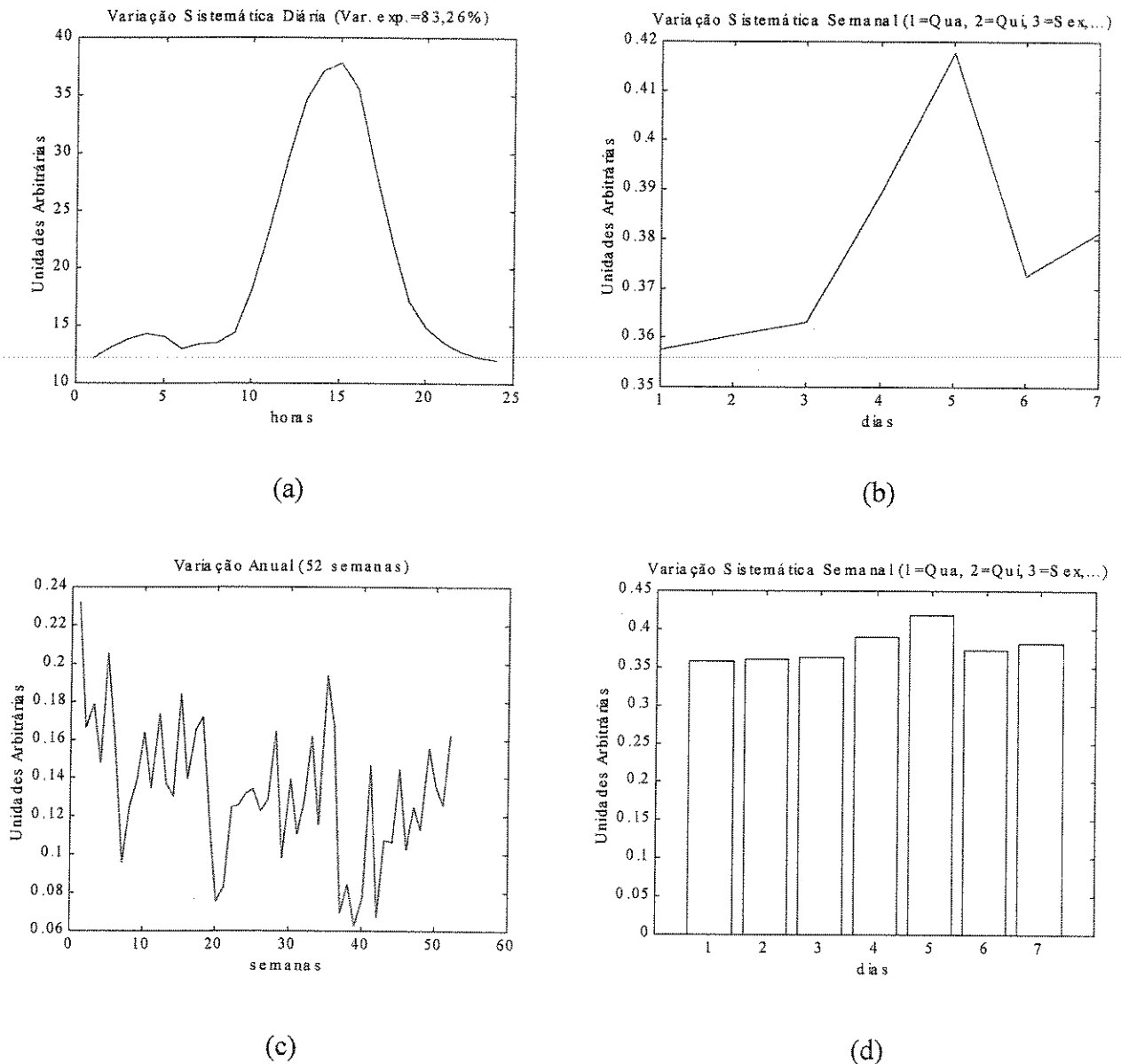


Fig.5.114: a)Variações sistemáticas com periodicidade diária. b) e d)Variações sistemáticas com periodicidade semanal. c)Variação anual. Dados para Osasco – 1997.

Em 5.114a nota-se o perfil característico do ozônio com o seu pico às 14:00 h. Em b (e d) notam-se os níveis mais elevado nos finais de semana. Em 5.114c é mostrada a variação anual. Ao final do período de inverno ocorre elevação dos níveis de O₃, seguida de queda brusca e nova elevação.

(SO₂98 – Osasco)

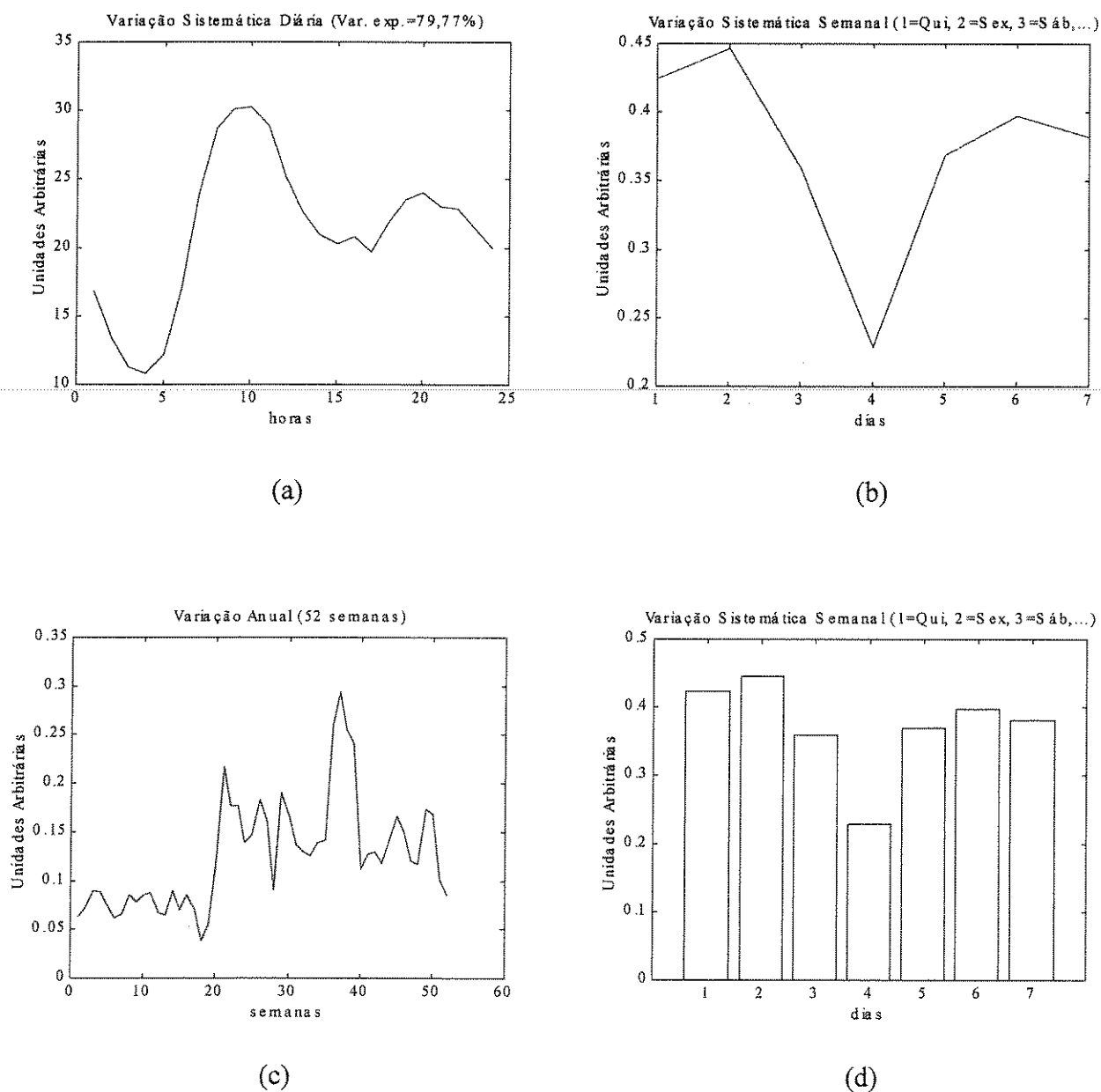


Fig.5.115: a)Variações sistemáticas com periodicidade diária. b) e d)Variações sistemáticas com periodicidade semanal. Variação anual. Dados para Osasco – 1998.

Em 5.115 a notam-se dois picos característicos de outros poluentes já analisados. Em 5.115b (e d) nota-se queda nas emissões nos finais de semana. Em 5.115c nota-se a maior acumulação no período de inverno.

Estação de Medição Centro (SO₂98) (SO₂98 – Centro)

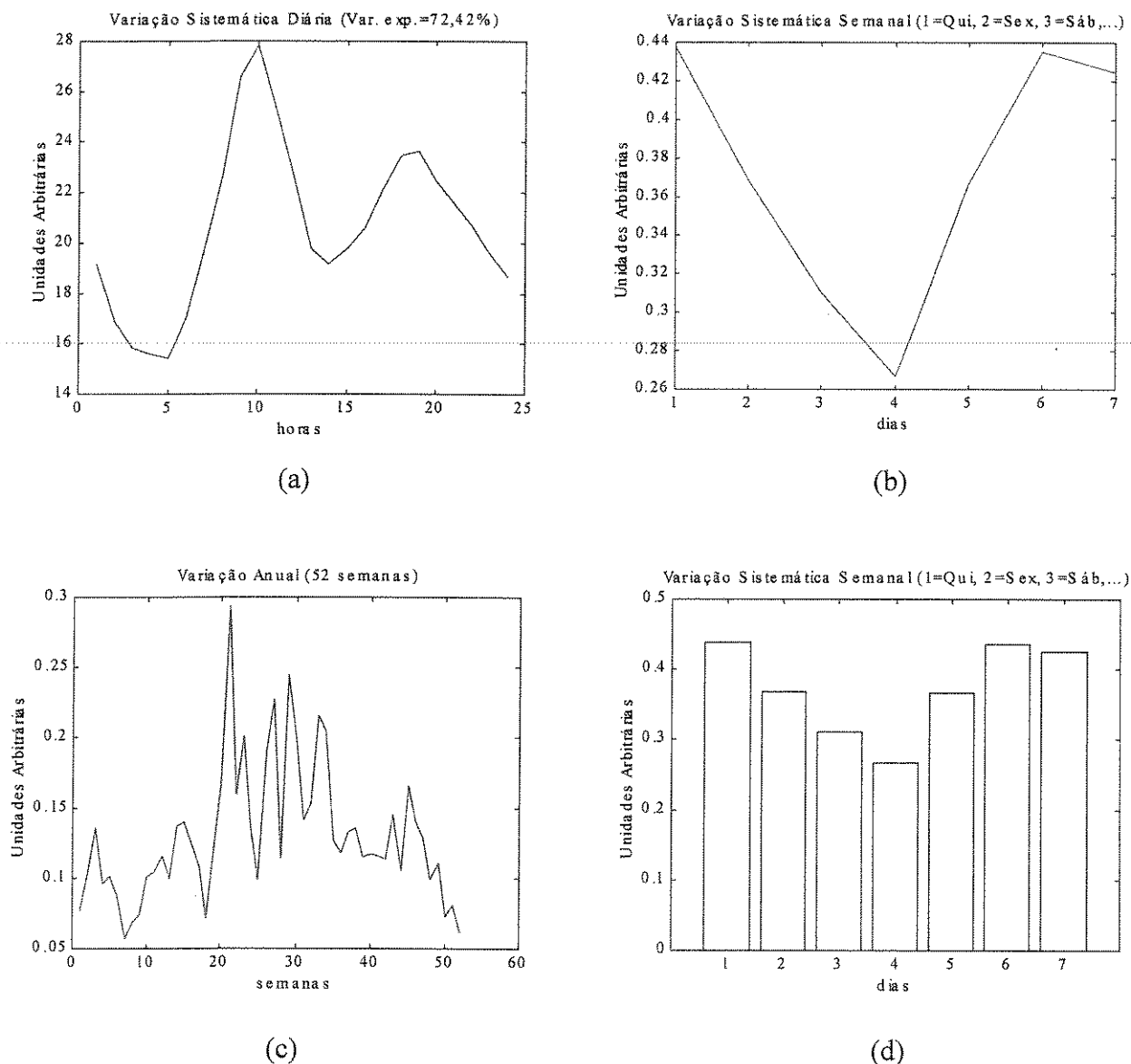


Fig.5.116: a)Variações sistemáticas com periodicidade diária. b) e d)Variações sistemáticas com periodicidade semanal. c)Variação anual. Dados para Centro – 1998.

Em 5.116 a, notam-se os picos característicos de emissões de outros poluentes e em outras localidades. Em 5.116 b (e d) nota-se queda nas emissões durante os finais de semana. Em 5.116c nota-se a maior acumulação no período de inverno.

Os resultados apresentados na seqüência de figuras com início na de número 5.100 até a de número 5.116 mostram que, de um modo geral, os padrões são os mesmos apresentados pelas variáveis sob investigação na Estação de medição P. D. Pedro II durante o ano de 1999.

O padrão marcado pela presença de dois picos, um deles pela manhã, em torno de 9:00 h, e o outro em torno de 20:00 h é um indício muito forte de que a fonte dessas emissões é automotiva. De fato, isto será confirmado mais adiante. Assim, a discussão que segue tomará como sendo proveniente de veículos, a principal fonte de emissão nos locais analisados até aqui.

As causas de algumas diferenças podem surgir em consequência de variações nas condições climáticas e/ou meteorológicas. Por exemplo, a maior acumulação de NO₂ em 1999 na localidade Centro (Fig.5.103a, p.219) pode ser atribuída à contínua emissão de poluentes no local, ou a problemas de dificuldade para a dispersão do mesmo. Não havendo queda apreciável da movimentação de veículos, não ocorre uma queda significativa nos níveis alcançados pelas concentrações de poluentes. Por outro lado, especificamente para o NO₂, pode ocorrer uma formação adicional do mesmo por meio de reação do NO com radicais livres, fato que, aliado à condições desfavoráveis à dispersão, causa maior acumulação desse poluente.

A Figura 5.104a (p.220) também mostra um perfil para o NO₂ em 1997 na Estação de medição Centro, semelhante àquele para 1999 na mesma localidade. Por outro lado, uma inspeção na Fig.5.109a (p.225), a qual mostra as variações sistemáticas para o NO em 1997 na localidade Centro, revela que a variável apresenta os dois picos característicos de emissões automotivas, e um vale bem pronunciado ~14:00 h, diferentemente do que ocorre para o NO₂ na mesma localidade e no mesmo ano. Desse modo, a explicação mais plausível para o fato é aquela cujos argumentos estão baseados em dificuldade de dispersão dos poluentes (ver discussão nos capítulos VI e VII).

Outro aspecto a destacar nos resultados desse bloco de análises refere-se à variação anual (não sistemática). A Figura 5.107c (p.223) mostra a variação anual para o NO na localidade Cerqueira César em 1999. Observa-se que entre ~23^a e 48^a semanas os níveis são iguais a zero. Esses valores não significam que tenha ocorrido ausência de emissões no período. É mais provável que o medidor não tenha conseguido registrar as emissões por

falha no seu funcionamento. Isto dificulta a avaliação da variação anual para esse caso. Por outro lado, às vezes, nota-se uma ou outra queda brusca na variação anual, mas isto não causa dificuldade na avaliação dessa variação. Esse é o caso mostrado na Figura 5.109c (p.225) para o NO na Estação de Medição Centro em 1999. Aqui a ocorrência de duas quedas bruscas (por curto período) durante o ano, não inviabilizou a avaliação.

Para os casos acima (variações sistemáticas ou variação anual), nos quais o perfil ou a variação escapava um pouco do que se previra, o recurso sempre foi recorrer às variações sistemáticas para confirmar a fonte de emissões. Claro, é conhecido *a priori* o perfil diário de movimentação de veículos, isto é, os horários de pico (“hush”) durante os dias úteis. Sabe-se também que nos finais de semana o tráfego de veículos cai drasticamente. Assim, se existe uma fonte automotiva importante, o perfil semanal pode confirmá-la sem qualquer dificuldade. Tal certeza da confirmação deriva do fato de que ocorre queda brusca nos finais de semana por um período muito mais prolongado do que durante um único dia. Nesse último caso, em primeiro lugar, variações nas condições de dispersão devem ficar restritas a algumas horas do dia. Em segundo lugar, elas devem ocorrer aleatoriamente, tal que se houver prolongamento do período, o episódio deve ser considerado como um fato isolado. Por outro lado, quedas bruscas nas emissões em todos os finais de semana não devem ser atribuídas a eventos aleatórios. Devem ser consideradas como variações sistemáticas e, nesse caso específico, tais variações devem ser atribuídas às emissões automotivas.

CAPÍTULO VI

RESULTADOS E DISCUSSÃO

6.1 – Principais Resultados e Discussão

A metodologia aplicada baseia-se na possibilidade de se mostrar a existência de variações sistemáticas nos dados. Esse procedimento se justifica porque as variações sistemáticas diárias e semanais, se existirem (e de fato existem), quando reveladas pelas variáveis latentes (componentes principais) podem ser associadas aos perfis de densidade de tráfego urbano e/ou das emissões industriais. Além disso, é possível verificar se o procedimento para encontrar tais variações não violam os mecanismos de reações já estabelecidos para a química da atmosfera.

Primeiramente mostrou-se a existência dessas variações para cada poluente isoladamente. Construiu-se um conjunto de matrizes de dimensões 365 (dias do ano) x 24 (horas do dia) para cada poluente individualmente. PCA desses dados revelou variações sistemáticas traçando, em última análise, o perfil característico das emissões diárias de cada um conforme mostrado nas Figuras 6.1 a 6.44 da seção 6.3.

Para a análise como dados de ordem superior as estruturas foram arranjadas em forma de cubo ou paralelepípedo de dados. Foram aplicados os modelos PARAFAC e TUCKER a arranjos, inicialmente, em três e quatro modos. Para a análise de variações sistemáticas diárias e semanais, e da variação anual, as dimensões foram de 24 (horas do dia) x 7 (dias da semana) x 52 (semanas do ano) para um dado poluente, possuindo, portanto, três modos. A adição de pelo menos mais um poluente ao modelo já exige a inclusão de mais um modo, sendo, portanto, uma estrutura em quatro modos. Outras

análises com os modelos PARAFAC e TUCKER envolveram arranjos de 24 (horas) x 7 (dias) x 52 (semanas) x poluentes x Estações de Medição, consistindo de estruturas em até 5 modos. Incluídos outros anos, as estruturas podem chegar a seis modos. Outras estruturas em cinco modos envolveram arranjos de 24 (horas) x 7 (dias) x 52 (semanas) x poluentes x anos. Incluídas outras estações de medição, as estruturas podem atingir, também seis modos. A escolha entre os dois tipos de arranjos em cinco e/ou seis modos acima dependerá dos objetivos a alcançar. Por exemplo, se é desejo estudar, para cada ano, as variações sistemáticas à luz de mudanças radicais nas fontes de emissão e/ou de mudanças climáticas consideradas suficientemente significativas para causar variações sensíveis nos perfis, cada ano comportar-se-á diferentemente dos outros, exigindo modos diferentes para cada um deles.

Não obstante tenha ocorrido (~junho/1997 a maio/1998) importantes variações climáticas (sazonais) causadas por deslocamento de massas de ar associadas a variações horizontais de temperatura, decorrentes do fenômeno conhecido como “El Niño”, com alterações significativas das condições de dispersão de poluentes na atmosfera da RMSP, nos dados analisados não se detectou diferenças significativas nos perfis diário e semanal para as emissões analisadas, e nas variações anuais para as variáveis em questão.

Deve ser enfatizado, então, que tais variações sazonais, não obstante tenham influído na dispersão dos poluentes nos diferentes anos, não se deram em extensão suficiente para causar modificações nos perfis de emissão dos mesmos.

A discussão dos resultados é realizada tendo em mente os principais resultados obtidos por meio de PCA, do modelo PRAFAC e do modelo de Tucker, tomando por base três ciclos para os referidos poluentes: o ciclo básico do monóxido de carbono, o qual considera-se constituído pelas variáveis CO, NO, NO₂ e O₃; o ciclo do metano (CH₄, NO, NO₂ e O₃) e do recém criado ciclo dos hidrocarbonetos não metano, no qual NMHC substitui o CO ou o CH₄. Depois estudou-se a interação entre os ciclos por meio dos modelos PARAFAC e de Tucker.

6.2 – Resultados

Devido a grande quantidade de resultados obtidos optou-se por expor a apresentação dos mesmos em uma hierarquia relacionada tanto com o número de variáveis estudadas em cada análise, quanto com os respectivos modelos aplicados nas mesmas para cada Estação

de Medição. Assim, são apresentados primeiramente os resultados para uma única variável por localidade. Para esse primeiro caso são apresentados os resultados obtidos por meio dos modelos PCA e PARAFAC, acompanhados de discussão preliminar. Em seguida, são apresentados os resultados para um conjunto de variáveis. Nesse caso, são apresentados os resultados obtidos com o modelo PARAFAC com discussão preliminar e, em seguida, os resultados oriundos da aplicação do modelo de Tucker.

Optou-se por essa forma de apresentação de resultados e discussões para evitar (o que não foi sempre possível) repetições. Entretanto, dadas a grande quantidade de dados analisados e a complexidade do comportamento das variáveis que constituem o sistema atmosférico, tornou-se necessário, em não raros casos, remeter o leitor à figuras distantes do ponto em discussão.

Espera-se que essa forma de apresentação de resultados cumpra o objetivo de identificação de fonte(s) de emissão de poluentes e seus desdobramentos. Eventuais dúvidas que parem em torno dos resultados poderão ser dirimidas por meio das Figuras e demais comentários adicionais contidos nos Apêndices.

6.3 – Análise em Dois Modos (PCA) e em Três Modos (PARAFAC)

Essa seção se preocupa em analisar por meio de PCA as variações com periodicidade diária para cada um dos poluentes, em cada estação, e em cada ano. Realiza-se também aqui a análise por meio do modelo PARAFAC para cada poluente/estação/ano, para se obter informações sobre os perfis diário e semanal das variáveis. O fato de se usar o modelo PARAFAC quando existe o de Tucker, mais geral, se justifica porque o PARAFAC é mais restrito e tem solução única indicando assim, sem ambigüidades, os perfis existentes. Os pesos (loadings) obtidos nessa análise são posteriormente “cruzados” no modelo de Tucker, conforme já discutido (Capítulo III, Seção 3.3).

A decomposição obtida por meio da análise de componentes principais fornece as matrizes de escores, e de pesos. Pode-se, conhecido o posto da matriz de dados decomposta, calcular a matriz de dados estimados. A diferença entre essas duas matrizes, isto é, o erro, deve ser o mínimo possível.

Nas análises realizadas nessa tese, a variância capturada pela primeira componente principal sempre foi muito maior (na maioria dos casos, maior do que 80%) do que a soma

das variâncias das outras PCs. Isto permitiria estabelecer como padrão para todas as análises a extração de três PCs (soma das variâncias para as três PCs foi, em geral, maior do que 90%). Entretanto, a existência de duas fontes de emissão em potencial já limitaria o número de PCs extraídas. Uma componente adicional às duas primeiras poderia levar em conta emissões de outras fontes menos importantes ou com variações não sistemáticas.

Como é óbvio, o erro calculado com todas as PCs deve ser zero. A distribuição do erro, porém, é importante para se ter uma idéia da qualidade da análise, isto é, se a variância explicada pelas PCs consideradas importantes contém, de fato, apenas informação significativa. A extensão do erro (em geral pequeno, como visto acima) não foi motivo de preocupação. Todavia, cuidou-se de avaliar a forma de sua distribuição. Assim, a partir dos parâmetros da análise, calculou-se o erro com as três primeiras PCs, com as duas primeiras, e também com a primeira PC, cujos resultados mostraram uma distribuição aleatória.

Pode-se considerar que a terceira PC esteja descrevendo variâncias sistemáticas de outras fontes menos importantes, variâncias sistemáticas desconhecidas, variações não sistemáticas, ou não possuir significado físico. Isto não se constitui em problema porque a identificação de fontes distintas deve ser melhor executada pelo modelo PARAFAC.

A disponibilidade de informações que permite organizar a maior quantidade de variáveis comuns a um determinado número de estações conduziu ao Quadro VI.1, o qual inclui três estações e sete variáveis. As estações são: Parque D. Pedro II, São Caetano do Sul e Ibirapuera. As variáveis: CO, NO, NO₂, O₃, SO₂, Temperatura e Umidade. Para que fosse possível estudar as influências de CH₄ e de NMHC, construiu-se um segundo quadro (Quadro VI.2) constituído de duas estações (Parque D. Pedro II e São Caetano do Sul) e nove variáveis (CO, NO, NO₂, O₃, CH₄, NMHC, SO₂, Temperatura e Umidade).

Apenas o ano de 1999 é apresentado no corpo do texto. Considera-se que isto seja suficiente para uma idéia completa tanto do conjunto de dados quanto do alcance da análise. Isto reduz drasticamente a parte gráfica, sem, contudo, causar prejuízos na análise e interpretação dos resultados.

6.3.1 – PCA e PARAFAC: 1999

Nessa seção são apresentados os resultados referentes às análises de três estações e sete poluentes. Os resultados estão lançados individualmente por poluente e por estação,

para o ano de 1999 (subseções 6.3.1.1 a 6.3.1.6), de acordo com hierarquia apresentada nos Quadros VI.1 e VI.2, mostradas a seguir.

Quadro VI.1: Hierarquia de apresentação de resultados PCA e PARAFAC para três Estações de Medição (1999)

ANO – 1999								
Variáveis								
/	CO	NO	NO ₂	O ₃	SO ₂	Temp	Umid	Modelo
Estações								
loadings	Figs.	Figs.	Figs.	Figs.	Figs.	Figs.	Figs.	
Estação 1	6.1	6.2a	6.2b	6.3	6.4	6.5a	6.5b	PCA
Pedro II	6.6	6.7	6.8	6.9	6.10.1 e 6.10.2	6.11	6.12	PARAFAC
Loadings	Figs.	Figs.	Figs.	Figs.	Figs.	Figs.	Figs.	
Estação 2	6.13	6.14a	6.14b	6.15	6.16	6.17a	6.17b	PCA
S. Caet.	6.18	6.19	6.20	6.21	6.22	6.23	6.24	PARAFAC
loadings	Figs.	Figs.	Figs.	Figs.	Figs.	Figs.	Figs.	
Estação 3	6.25	6.26a	6.26b	6.27	6.28	6.29a	6.29b	PCA
Ibirap.	6.30	6.31	6.32	6.33	6.34	6.35	6.36	PARAFAC
Resultados Adicionais								
Estação 1 Pedro II	Capítulo V e Apêndice C							
Estação 2 São Caetano	Capítulo V e Apêndice C							
Estação 3 Ibirapuera	Capítulo V e Apêndice C							

Quadro VI.2: Hierarquia de apresentação de resultados PCA e PARAFAC para duas Estações de Medição (1999)

ANO – 1999					
Estações / Variáveis	Estação 1 Parque D. Pedro II	Estação 2 São Caetano	Modelo	R e s u l t a d o s	Estações P. D. Pedro II e São Caetano
CO	Figs. 6.1 6.6	Figs. 6.13 6.18	PCA PARAFAC		
NO	Figs. 6.2a 6.7	Figs. 6.14a 6.19	PCA PARAFAC	A d i c i o n a i s	Capítulo V e Apêndice C
NO ₂	Figs. 6.2b 6.8	Figs. 6.14b 6.20	PCA PARAFAC		
O ₃	Figs. 6.3 6.9	Figs. 6.15 6.21	PCA PARAFAC		
CH ₄	Figs. 6.37 6.38	Figs. 6.41 6.42	PCA PARAFAC		
NMHC	Figs. 6.39 6.40	Figs. 6.43 6.44	PCA PARAFAC		
SO ₂	Figs. 6.4 6.10	Figs. 6.16 6.22	PCA PARAFAC		
Temp	Figs. 6.5a 6.11	Figs. 6.17a 6.23	PCA PARAFAC		
Umid	Figs. 6.5b 6.12	Figs. 6.17b 6.24	PCA PARAFAC		

No Quadro VI.2, CH₄ e NMHC estão destacados em vermelho para melhor identificação pois a opção de apresentação de resultados (inclusão de CH₄ e NMHC, e exclusão de Ibirapuera) implicou em truncamento da numeração.

Est.1: Estação P. D. Pedro II – 1999

6.3.1.1 – Análise de Monóxido de Carbono – CO

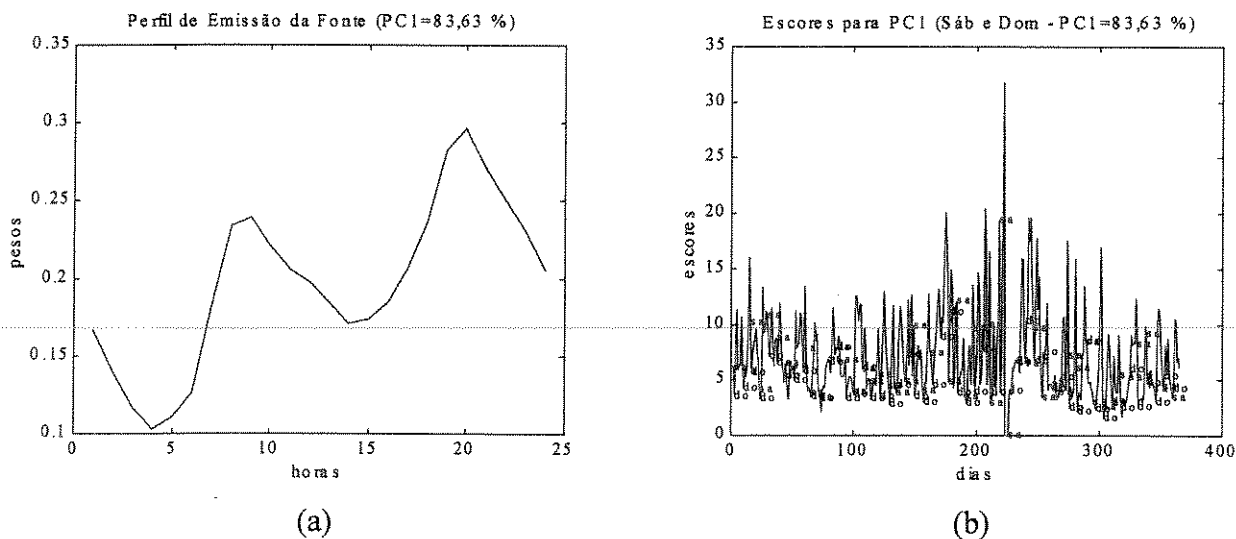


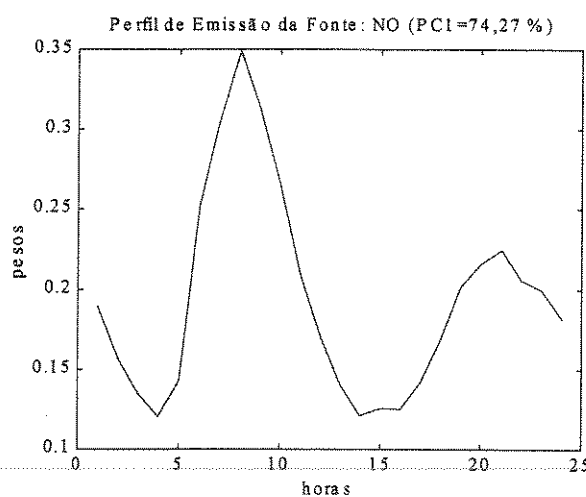
Fig.6.1: a)Pesos para PC₁. b)Escores para PC₁.

A análise PCA na matriz $X(365 \times 24)$ para o monóxido de carbono, CO, mostrou no seu perfil diário, um crescimento nas primeiras horas do dia, com um pico em torno das 9:00 h seguido de uma queda com um vale em torno das 14:00 h. A partir daí, um novo crescimento com um segundo pico aproximadamente às 20:00 h, voltando a cair. A figura 6.1a mostra que esse padrão se repete todos os dias do ano, variando apenas em amplitude. Essa variação de amplitude não é visível nos pesos extraídos nessa análise PCA, a não ser que a matriz de dados seja adequadamente organizada para tal.

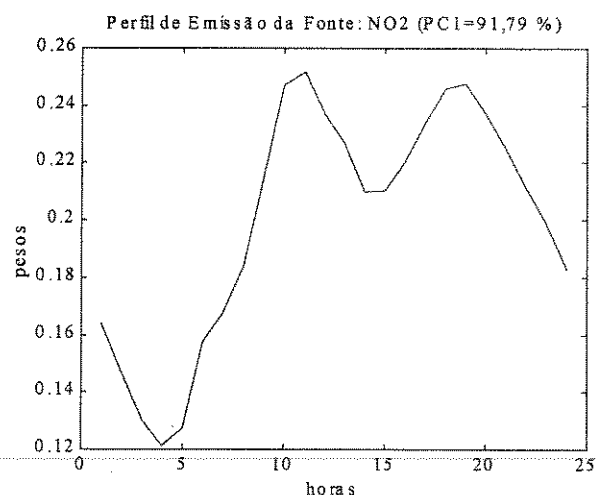
Entretanto, na Figura 6.1b é possível observar a influência do efeito sazonal na amplitude da variação anual, a qual atinge valores mais elevados durante o inverno (período de seca). Nessa figura estão “rotulados” apenas os finais de semana. Percebe-se que nos sábados (“sa”) e domingos (“do”) os escores são baixos indicando queda nas emissões no período.

Como se verá mais adiante, os pesos extraídos na decomposição por meio do modelo PARAFAC mostram, no modo C, a variação anual referida acima para as 52 semanas do ano, e no modo B, a variação semanal, de forma muito mais visível.

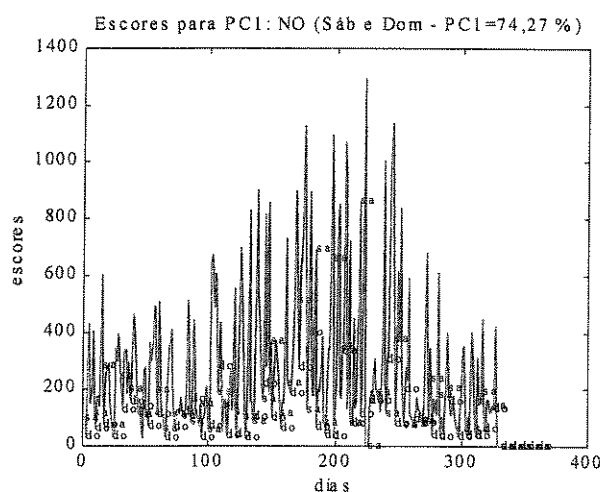
6.3.1.2 – Análise de Óxido Nítrico e Dióxido de Nitrogênio – NO e NO₂



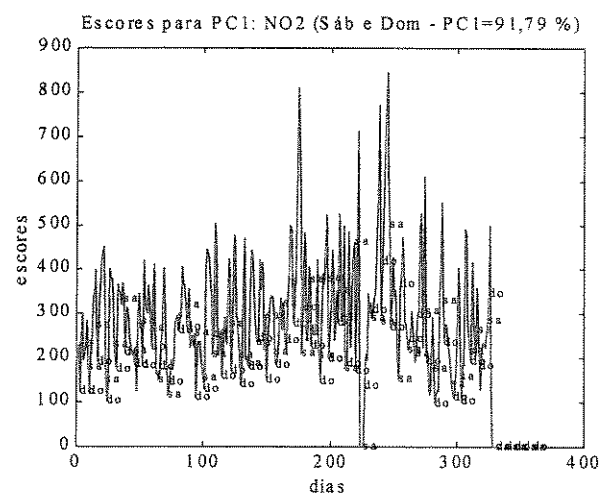
(a)



(b)



(c)



(d)

Fig.6.2: a)Pesos para a primeira componente principal, PC₁, para o NO. b)Pesos para a primeira componente principal, PC₁, para o NO₂. c)Escores para PC₁ (NO). d)Escores para PC₁ (NO₂).

Uma inspeção na Figura 6.2 mostra a semelhança entre os perfis de emissão da fonte para o NO e o NO₂ e CO mostrado na Figura anterior. Analisados por PCA os óxidos de nitrogênio mostraram, para o perfil diário, uma variação similar à do monóxido de

carbono. O NO apresenta dois picos, sendo um deles em torno das 8:00 h e o outro, menos intenso, aproximadamente às 20:00 h. Por outro lado, o NO₂ possui seu primeiro pico em torno das 10:00 h, e o segundo por volta das 19:00 h, ambos, ao contrário do NO, com valores de intensidade muito próximos. Isto pode ser observado nas Figuras 6.2a e 6.2b. Do mesmo modo que nos outros casos, esse padrão se repete ao longo de todo o ano.

A defasagem (“time lag”) entre os horários de pico para o NO e o NO₂, se devem, no primeiro caso (1º pico) à formação de NO₂ às custas de NO em presença de luz solar e, no segundo (2º pico), à combinação de vários fatores, os quais vão desde o efeito da densidade de tráfego, à reações ocorrendo com os poluentes. O fenômeno está, portanto, relacionado com o efeito causado pelas reações fotoquímicas sobre os compostos envolvidos no ciclo do monóxido de carbono, já abordado parcialmente no Capítulo II, e sendo detalhadamente discutido mais adiante nesse capítulo.

Nas Figuras 6.2c e 6.2d, respectivamente, para o NO e o NO₂, é possível observar a influência do efeito sazonal na amplitude da variação anual, a qual atinge valores mais elevados durante o inverno (período de seca). Nessa figura também estão “rotulados” apenas os finais de semana. Percebe-se que nos sábados e domingos os escores são baixos indicando queda nas emissões no período. A variação sazonal acima é melhor percebida quando é aplicado o modelo PARAFAC aos dados.

Ainda nas Figuras 6.2c e 6.2d, os valores nulos que aparecem na extremidade direita dos mesmos (a partir do 325º dia) correspondem à valores originalmente iguais a zero. Isto é, os valores de concentração contidos nas tabelas fornecidas pela CETESB, para as duas variáveis, são iguais a zero e sob a classificação “*não disponível*”. Esses valores aparecem nas tabelas no período de 23/11/1999 a 31/12/1999. Esses intervalos originalmente nulos são observados nos gráficos correspondentes às variáveis originais e, claro, surgem após as análises nos valores dos escores para as componentes principais. Entretanto, nem sempre é isto que acontece. Os valores iguais a zero para os escores para a primeira PC entre o 224º e 226º dias em 6.2c e 6.2d não são originalmente nulos, mas correspondem a dados faltantes no referido período. Esses dados apresentam pequenos problemas de “lacunaridade” (não corrigida), tal que o modelo aplicado leva a valores de escores e de dados estimados mais baixos do que o padrão das curvas, especialmente para o NO₂ (Figura 6.2d).

6.3.1.3 – Análise de Ozônio – O₃

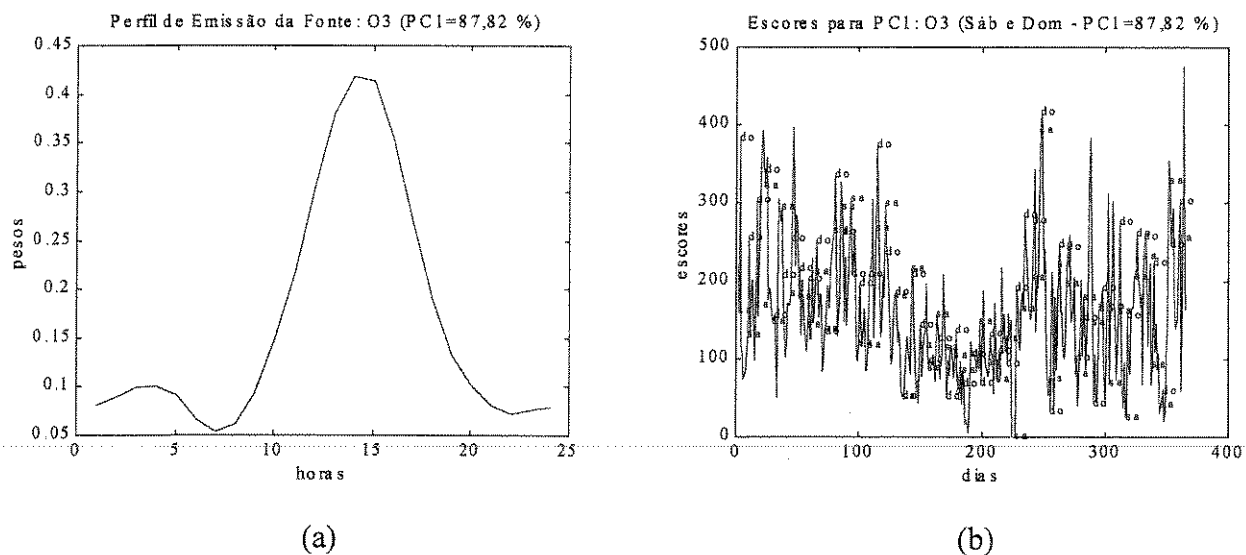


Fig.6.3: a)Pesos para PC₁. b)Escores para PC₁.

A análise PCA para o ozônio mostrou, no seu perfil diário, um pequeno crescimento nas primeiras horas da madrugada, com um pequeno pico em torno das 3:00 h, um vale aproximadamente às 7:00 h, um contínuo crescimento com um segundo (e mais intenso) pico em torno das 14:00 h e, então, uma queda contínua com um novo mínimo aproximadamente às 22:00 h, como mostra a Figura 6.3a. Esse padrão se repete ao longo de todo o ano. Comparando-se as Figuras 6.1a, 6.2a, 6.2b e 6.3a, observa-se uma correlação negativa entre os níveis para o poluente secundário, ozônio, e os poluentes primários, CO, NO e NO₂.

Comparando-se as Figuras 6.1b, 6.2c e 6.2d com 6.3b observa-se que a variação sazonal para o ozônio difere daquela apresentada pelos poluentes primários. Isto é melhor percebido quando se aplica o modelo PARAFAC como se verá adiante.

Do mesmo modo que nos casos anteriores, foram assinalados (rotulados) para o O₃ apenas os dias correspondentes aos finais de semana (sábados e domingos). Observa-se por meio da Figura 6.3b que os escores são mais elevados para o ozônio nos finais de semana quando comparados àqueles para os poluentes primários. Esse contraste também é muito mais perceptível na análise por meio do modelo PARAFAC.

6.3.1.4 – Análise de Dióxido de Enxofre – SO₂

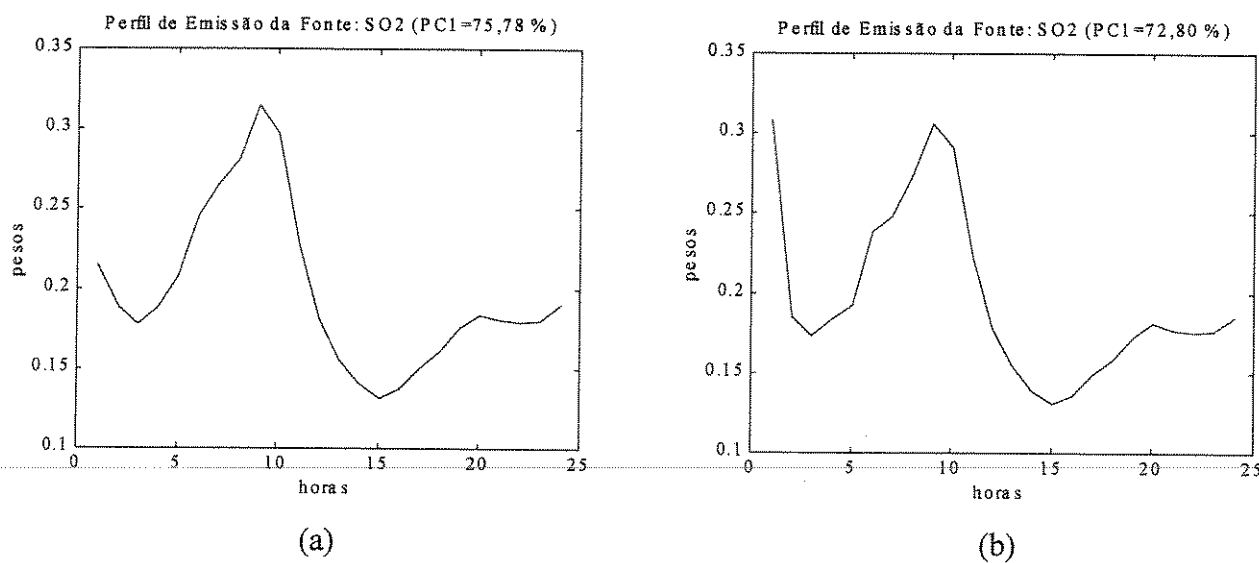


Fig.6.4: Pesos para PC₁. (a) Uma fonte, (b) Influência de uma segunda fonte.

O comportamento do SO₂ durante o ano de 1999 é semelhante ao dos outros poluentes primários no que diz respeito às fontes automotivas. Entretanto, ele é peculiar e guarda diferenças em relação ao ocorrido nos anos de 1997 e 1998. Em 1999, para cada uma das Estações de medição (P. D. Pedro II menor influência do que em S. Caetano, e esta menor do que no Ibirapuera) ocorreram, à 1:00 h em determinados períodos contínuos, valores “descolados” (ou deslocados) do padrão das emissões nos outros horários. Esses casos foram estudados sem e com a influência dos valores coletados à 1:00 h. Isto, por um lado, afastaria uma eventual hipótese de interferência no perfil das emissões provenientes dos veículos e, por outro, poderia confirmar a existência de outra(s) fonte(s) de emissão, por exemplo, indústrias.

A Figura 6.4a, sem a influência daqueles valores (ver explanação adiante), exibe um padrão parecido com o das Figuras 6.1a, 6.2a e 6.2b (respectivamente, para o CO, o NO e o NO₂). Porém o padrão assemelha-se mais ao da Figura 6.2a (para o NO). Nos dois casos (NO e SO₂), o primeiro pico (9:00 – 10:00 h) é mais intenso do que o segundo (20:00 – 21:00 h). Entretanto, enquanto no caso do NO ocorre uma queda após o segundo pico, para o SO₂ a queda é muito suave e a curva tende a subir novamente quando se aproxima das 24:00 h. O segundo pico, menos intenso, e a queda do NO estão relacionados por um lado, com a queda das emissões e, por outro, com reações de consumo de NO (por exemplo,

reação com O_3 e formação indireta de HONO – ver discussão na Seção 7.3). O modelo PARAFAC (Figura 6.10.1, p257) capta tais variações sistemáticas diárias e semanais, e a variação anual. Esses casos são discutidos no final desse capítulo.

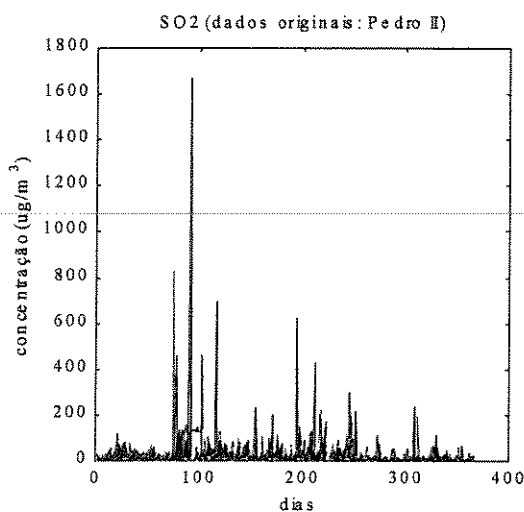
A Figura 6.4b, com a influência daqueles valores (explicação adiante), mostra o mesmo padrão apresentado na Figura 6.4a para as emissões automotivas, porém é visível a distinção à 1:00 h. O modelo PARAFAC (Figura 6.10.2, p258) capta as variações sistemáticas diárias e semanais, e a variação anual para as duas diferentes fontes emissoras.

No caso da Estação de Medição P. D. Pedro II, o período afetado (27 dias) vai do 76º ao 102º dias (quatro semanas, 17/03 a 12/04, num total de 648 dados). A operação para retirar a influência de tais valores foi realizada de duas maneiras. Na primeira delas, todos os valores para 1:00 h foram retirados e a análise foi feita apenas com os dados referentes às 23 horas do dia. Na segunda, calculou-se a média para as vinte e quatro horas de cada dia (27 médias), e os valores para 1:00 h em cada um dos dias afetados foram substituídos pelos correspondentes valores médios. Adicionalmente calculou-se a média para as vinte e três horas de cada dia (isto é, de 2:00h – 24:00h), sendo estas usadas em substituição àqueles valores para 1:00 h nos dias afetados. Em seguida os dados foram analisados por PCA e PARAFAC. Em que pese seja a média a opção recomendável, era necessário verificar se ocorreria discrepância entre os dois procedimentos. Ademais, na medida do possível, é desejável preservar os modos com $24 \text{ (horas)} \times 7 \text{ (dias)} \times 52 \text{ (semanas)}$ ao longo das análises.

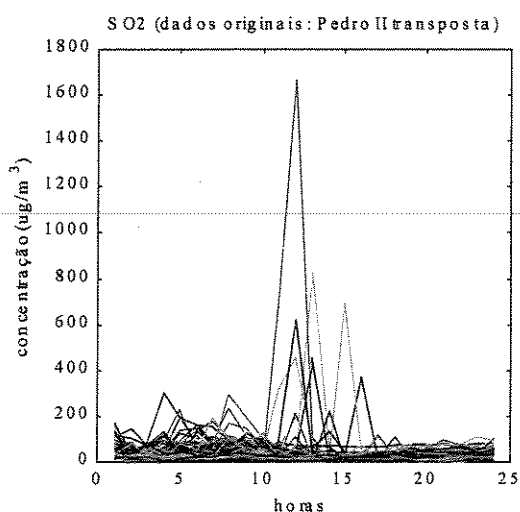
Assim, no caso de PCA (matrizes de $365 \text{ dias} \times 24 \text{ horas}$), as matrizes de pesos conservam um dos seus modos com as 24 horas. Nesse caso, a frase acima “*sem a influência daqueles valores*”, deve ser entendida como dados para 1:00 h substituídos pelas médias, respectivamente, para as 24 horas do dia (1:00 – 24:00 h) ou para as 23 horas do dia (2:00 – 24:00 h). Obviamente, a frase “*com a influência daqueles valores*”, significa dados para as 24 horas do dia sem correção para os valores à 1:00 h. Esses valores foram denominados de “*dados suspeitos*” e não devem ser confundidos com dados anômalos. Em resumo, as correções para dados suspeitos referidas aqui são realizadas sobre os dados já corrigidos para dados anômalos e variações súbitas.

Não houve diferenças para os perfis obtidos segundo cada média (1:24 h e 2:24 h). Isto, provavelmente porque o período afetado é curto e as concentrações alcançadas pelo SO₂ não são muito elevadas.

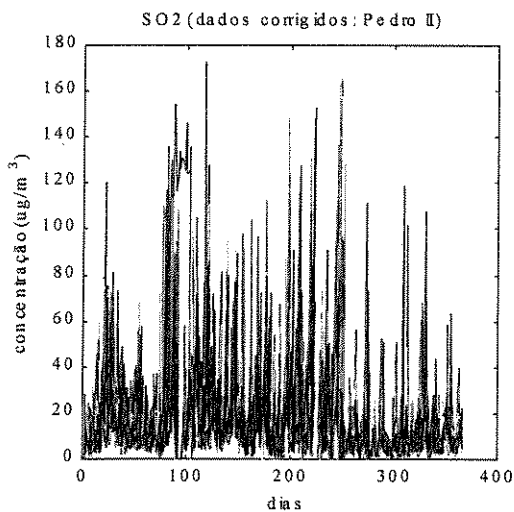
Em seguida são mostradas as matrizes contendo os dados originais e os dados corrigidos, para que se possa visualizar a curva para 1:00 h no período afetado.



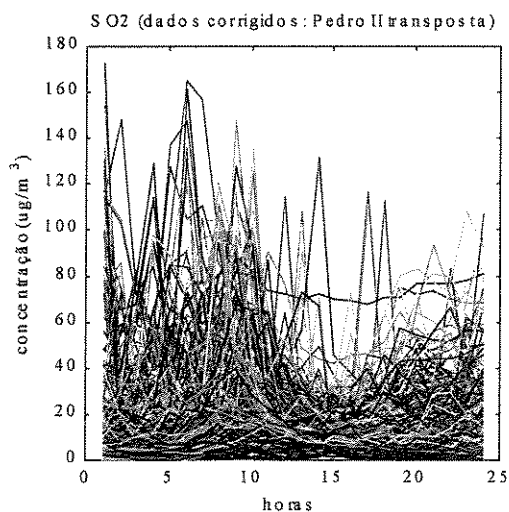
(a)



(b)



(c)



(d)

Fig.6.4.1: Dados originais: a) concentração $\times$ dia, b) concentração $\times$ hora. Dados corrigidos: c) concentração $\times$ dia, d) concentração $\times$ hora.

As Figuras 6.4.1a e 6.4.1b mostram, respectivamente, os dados originais e a matriz transposta desses dados. É difícil visualizar os valores para 1:00 h em a e/ou b. Após as correções para dados anômalos e variações súbitas, obtêm-se as Figuras 6.4.1c e 6.4.1d. Estas representam, respectivamente, a matriz de dados originais e sua transposta. Aqui já são mais visíveis os valores para 1:00 h.

Entretanto, é na próxima figura (6.4.2) que se pode observar mais claramente esse fato.

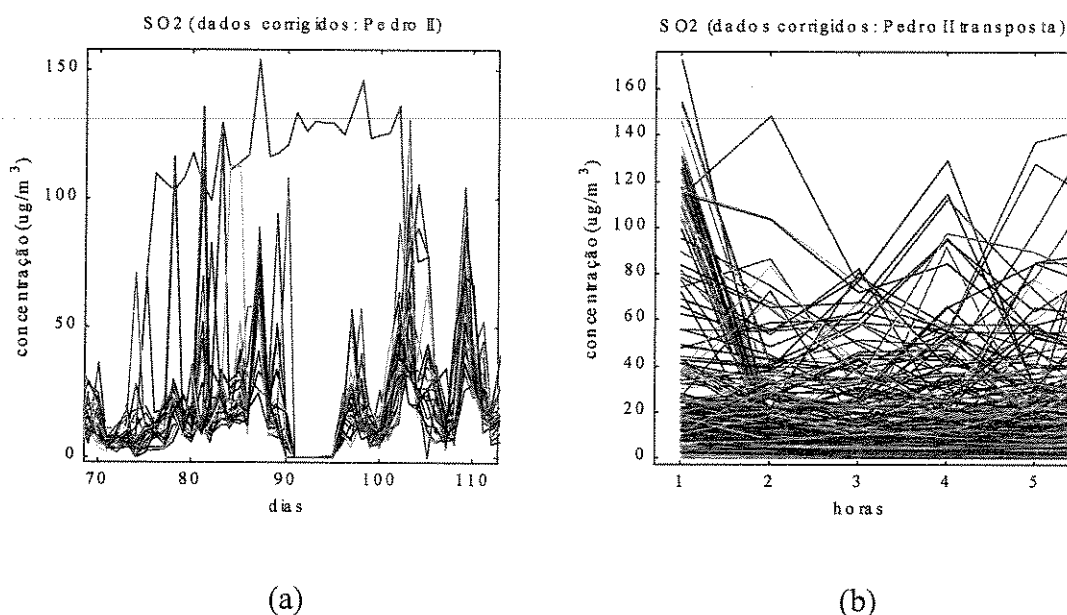


Fig.6.4.2: Ampliação ("zoom") sobre dados corrigidos no período afetado. a) concentração $\times$ dia. b) concentração $\times$ hora.

A Figura 6.4.2a mostra claramente por meio da curva na cor azul (na parte superior da figura), os valores para 1:00 h. Como se vê, esses valores escapam aos demais ao longo do espectro. Uma inspeção na Figura 6.4.2b mostra que um conjunto de valores, acima de cerca de 80 para os escores, forma um feixe com queda súbita para valores bem mais baixos (menores do que ~ 20) às 2:00 h.

Os resultados acima, para o SO₂ com a influência dos dados à 1:00 h, devem ser comparados com aqueles nos quais a influência daqueles dados foi retirada. A próxima figura (6.4.3) mostra a espectro obtido quando tal influência é removida.

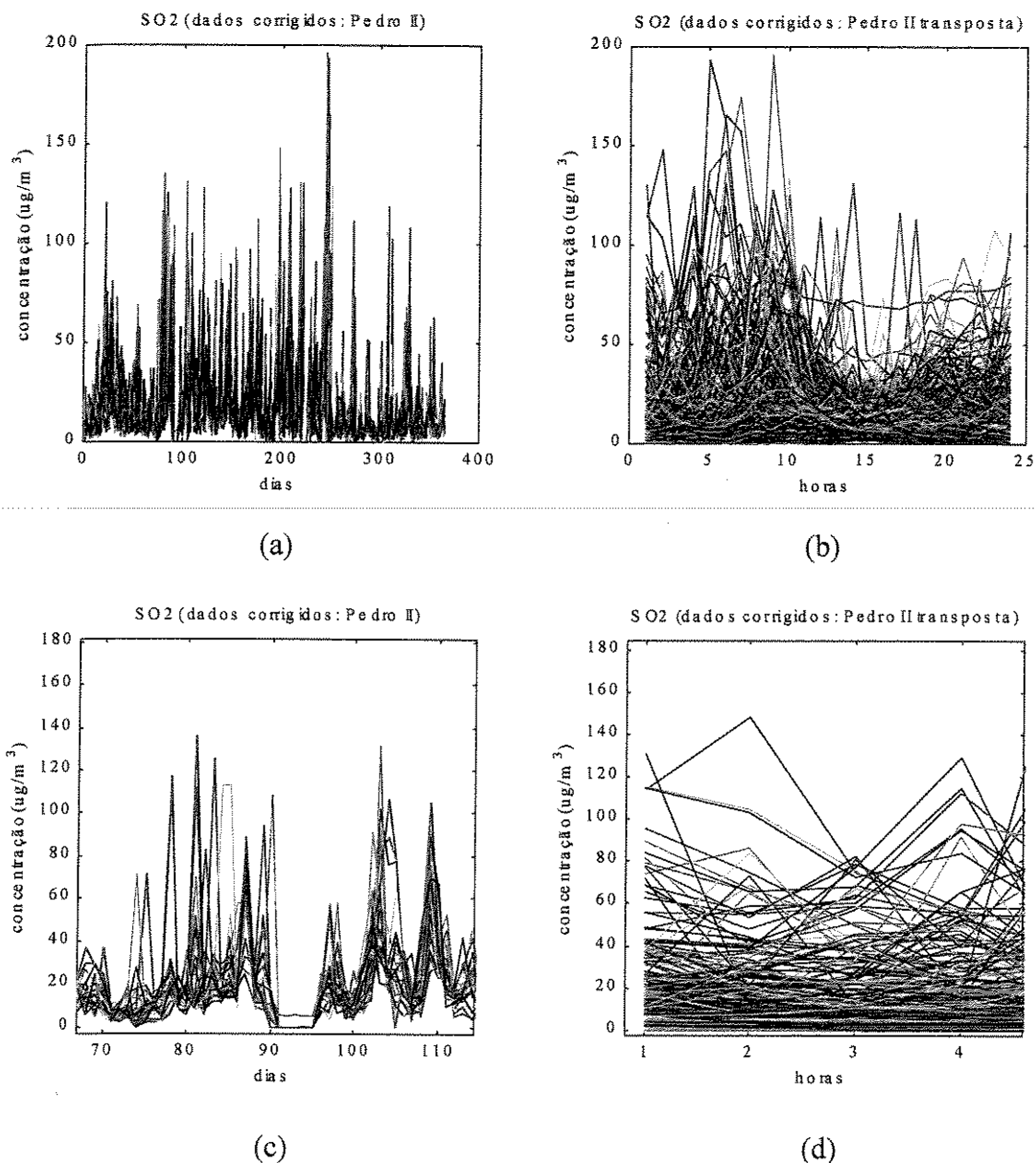


Fig.6.4.3: a) Dados corrigidos. a) concentração $\times$ dia. b) concentração $\times$ hora. Ampliação ("zoom") sobre dados corrigidos no período afetado sem os dados à 1:00 h. c) concentração $\times$ dia. b) concentração $\times$ hora.

Em 6.4.3a observa-se o espectro sem os dados para 1:00 h no período afetado. Nota-se, em 6.4.3b e d, a ausência do "feixe" de quedas súbitas ao passar de 1:00 h para 2:00 h.

O problema manifestado pelo SO₂ na Estação de Medição Parque D. Pedro II (e também em São Caetano e no Ibirapuera), durante o ano de 1999, exigiram a introdução de informações adicionais para que se possa tentar compreender o comportamento desse poluente na atmosfera da RMSP (Região Metropolitana de São Paulo).

Após a análise para o SO_2 no Ibirapuera faz-se uma discussão sobre essas emissões. Lá, figuras mostrando dados originais e dados estimados com números diferentes de PCs pela análise PCA, destacam as curvas para 1:00 h e para as 23 horas restantes, durante o ano inteiro. Esses fatos também são explorados por meio do modelo PARAFAC. Tais resultados tornam evidente a importância dessas emissões. Com base nelas e em dados sobre precipitações pluviométricas e de frequência de inversões térmicas, é possível fazer inferências sobre as causas de distinções entre as emissões para o ano de 1999 e para o período 1997-1998 no qual o “padrão à 1:00 h” não se verifica, nesses locais.

Como esse padrão à 1:00 h ocorre nas três localidades, optou-se por apresentar cada resultado separadamente, por Estação de Medição, seguido de breves comentários para, em seguida, apresentar uma discussão levando-se em consideração as três localidades. Se isto, por um lado vier a revelar-se tarefa cansativa para o leitor, por outro, revelar-se-á mais fácil sob o ponto de vista de interpretação, pois torna mais palpáveis inevitáveis comparações.

Cabe resgatar, nesse ponto, que a Figura 6.4a foi construída com os valores para 1:00 h substituídos pela média das 23 horas, isto é, média de 2:00 h a 24:00 h, a despeito de não se verificarem diferenças significativas entre os resultados com qualquer das médias. Isto foi adotado porque no caso da Estação de Medição do Ibirapuera, ocorrem diferenças significativas entre as duas médias. Assim, para que todos os casos fiquem submetidos ao mesmo critério, adotou-se em todos eles a média da 23 horas do dia (2:00 h – 23:00 h) para substituir os dados à 1:00 h, quando se desejava retirar a influência daqueles dados sobre o conjunto dos demais na respectiva localidade para o poluente em questão (SO_2).

6.3.1.5 – Análises da Temperatura e da Umidade – Temp e Umid

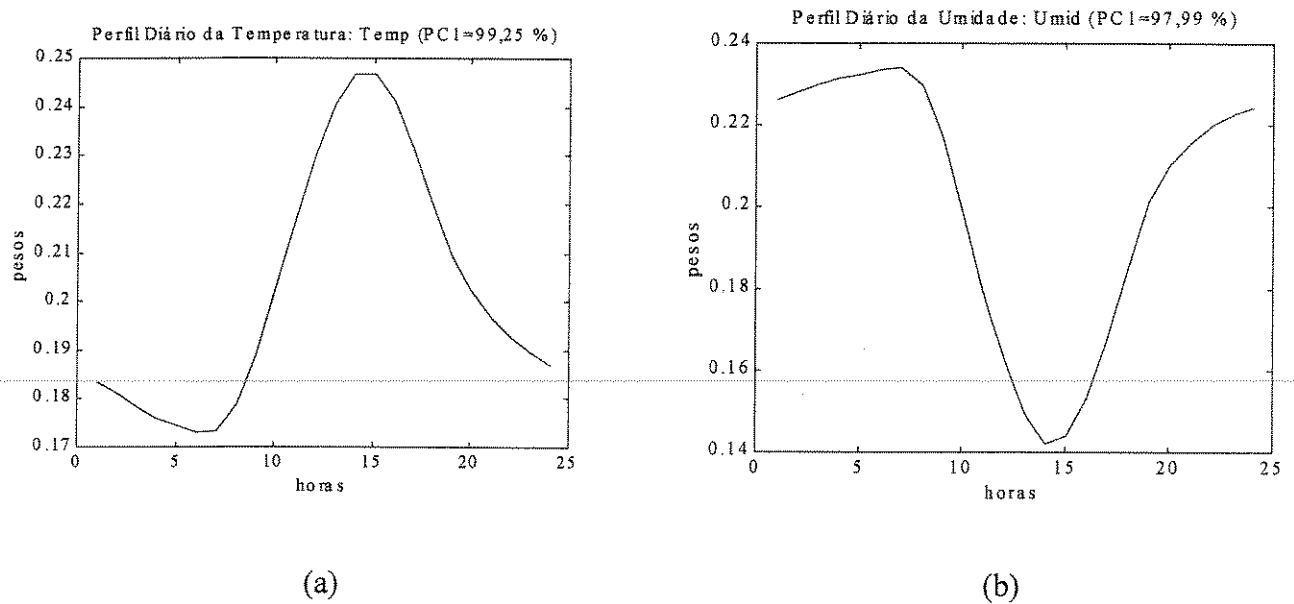
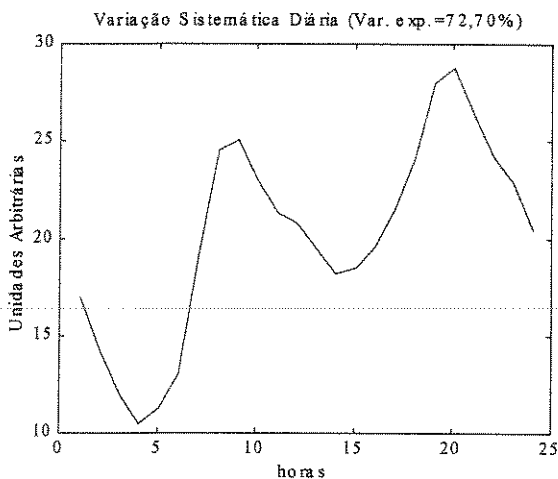


Fig.6.5: a) Pesos para a primeira componente principal, PC_1 , para a temperatura. b) Pesos para a primeira componente principal, PC_1 , para a umidade.

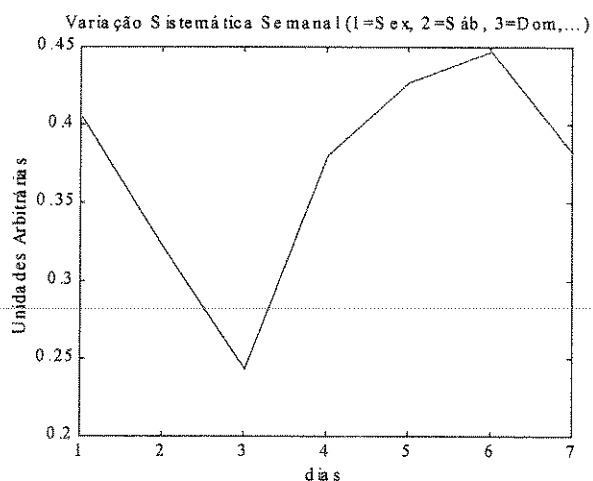
É interessante notar, na Figura 6.5 a simetria existente entre os perfis para a temperatura e para a umidade. Enquanto a temperatura atinge seu valor de pico próximo das 14:00 h, é nesse mesmo horário que a umidade possui seu valor mínimo. Os perfis mostram que existe uma correlação negativa entre os níveis da temperatura e da umidade, podendo-se inferir daí que, para um grupo importante de reações na baixa troposfera, as influências dessas duas variáveis são contribuições aditivas.

6.3.1.6 – Análises PARAFAC (Parque D. Pedro II)

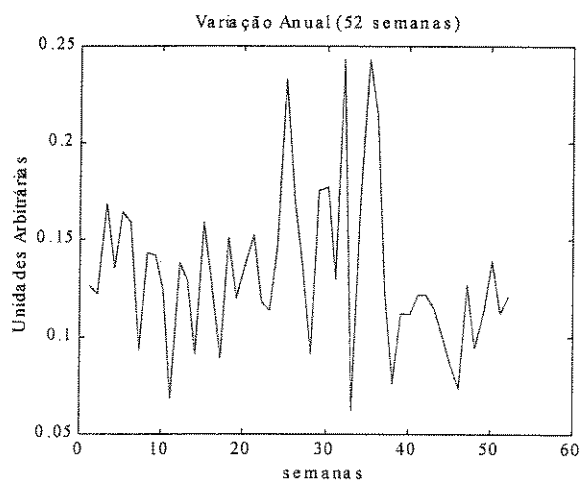
Modelo PARAFAC para o CO



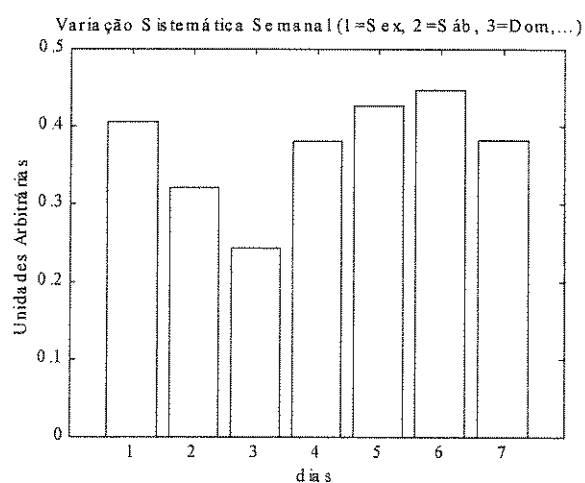
(a)



(b)



(c)



(d)

Fig.6.6: a) Variações sistemáticas com periodicidade diária. b) Variações sistemáticas com periodicidade semanal. c) Variação anual (não sistemática). Os gráficos b e d mostram que as emissões caem nos finais de semana, sendo mais elevadas nos chamados dias úteis. Nota-se, em c, o aumento de amplitude nas semanas correspondentes ao período seco.

Os resultados obtidos por meio de PCA e do modelo PARAFAC são consistentes entre si (comparar perfis diários encontrados em cada análise, respectivamente, Figuras 6.1a e 6.6a). Essas figuras mostram um pico em torno das 9:00-10:00 h e outro em torno das 19:00-20:00 h. Estes são horários de pico para o tráfego de veículos. Um outro aspecto interessante mostrado nos gráficos b e d da Figura 6.6, é a confirmação daquilo que já se adiantara no resultado obtido por meio de PCA para o CO: os escores para PC1 são baixos (ver Figura 6.1b). Esse fato é muito mais visível quando se observa o perfil semanal obtido por meio do modelo PARAFAC (Figuras 6.6b e 6.6d). Essa última análise mostra queda nas emissões nos finais de semana, fato que também se verifica com o tráfego de veículos.

As análises por meio de PCA e do modelo PARAFAC fornecem vários parâmetros de saída com informações adicionais sobre os procedimentos usados. Seria interessante mostrar uma dessas saídas típicas para uma melhor orientação sobre as análises. O Quadro VI.3 fornece essas informações para o CO na Estação de Medição P. D. Pedro II em 1999.

Quadro VI.3: Parâmetros de saída do software usado em PCA e PARAFAC

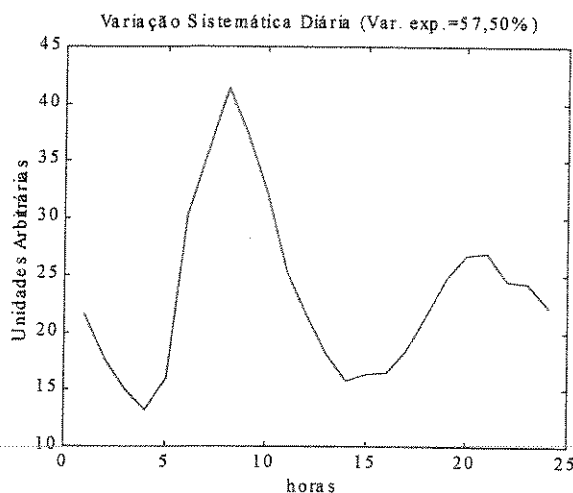
```

PCA: CO (Pedro II - 1999)
% Now working on iteration number 19
% Sum of squared differences in missing data estimates = 5.9779 e-11
% Now forming final PCA model
% Percent Variance Captured by PCA model
% Based on Known Data Only
% PC#    %Var    %TotVar
% 1.0000 83.6351 83.6351
% 2.0000 7.1046 90.7397
% 3.0000 2.9502 93.6899

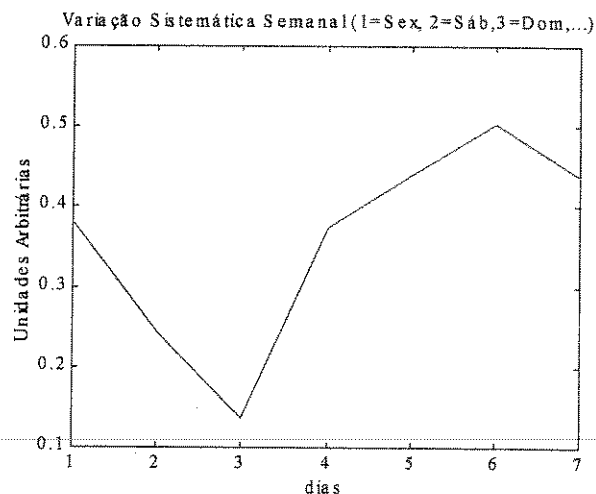
PARAFAC: CO (Pedro II - 1999)
%PRELIMINARY
% The array is a 3-way array with dimensions: 24 x 7 x 52
% A 1-component model will be fitted
% Nonnegativity on mode 1
% Nonnegativity on mode 2
% Nonnegativity on mode 3
% The convergence criterion is 1e-010
% Sum-of-Squares    Iterations    Explained    Tuckers congruence
% of residuals              variation      coefficient
% 4074.6563002128      6      71.6005      1.0000
% The algorithm converged

```

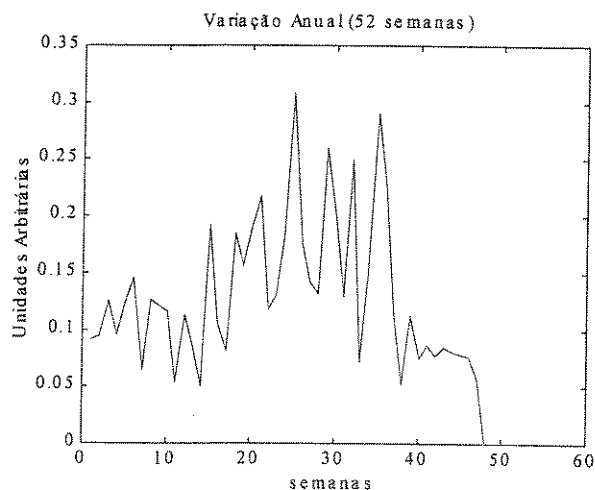
Modelo PARAFAC para o NO



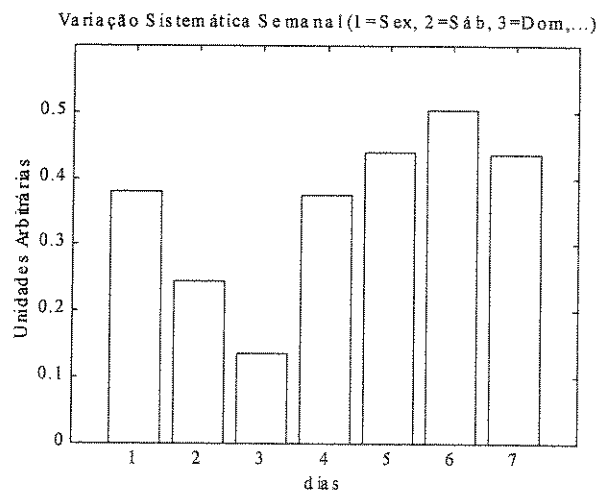
(a)



(b)



(c)

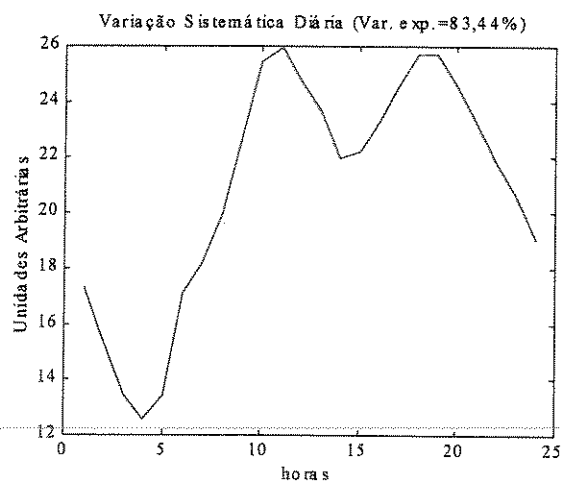


(d)

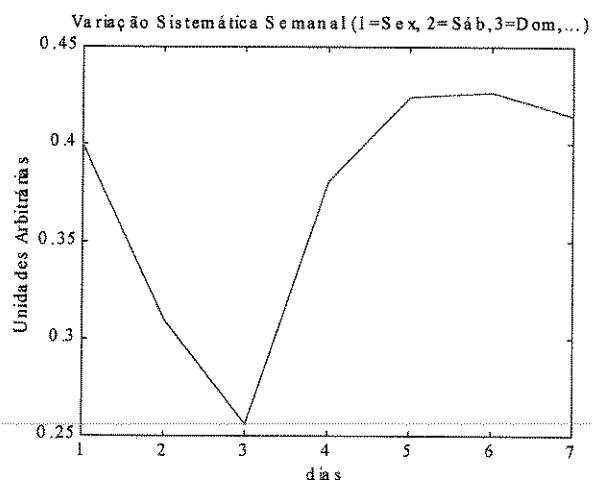
Fig.6.7: a) Variações sistemáticas com periodicidade diária. b) Variações sistemáticas com periodicidade semanal. c) Variação anual (não sistemática). Os gráficos b e d mostram que as emissões caem nos finais de semana, sendo mais elevadas nos chamados dias úteis. Nota-se também, no gráfico c, o aumento de amplitude nas semanas correspondentes ao período seco (inverno).

Observe-se a concordância com os resultados obtidos por meio da análise PCA para o perfil diário. Comparação entre os gráficos b e d das Figuras 6.6 e 6.7 mostra claramente a semelhança existente no perfil semanal para os gases CO e NO.

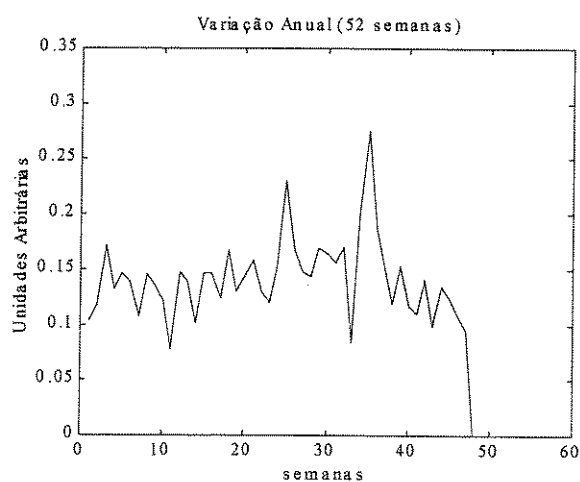
Modelo PARAFAC para o NO₂



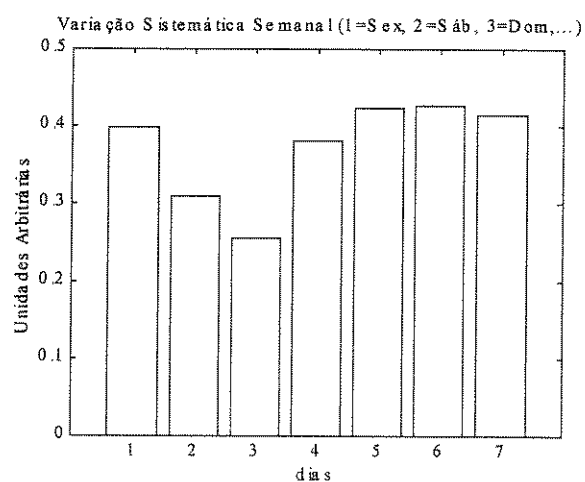
(a)



(b)



(b)



(d)

Fig.6.8: a) Variações sistemáticas com periodicidade diária. b) Variações sistemáticas com periodicidade semanal. c) Variação anual (não sistemática). Os gráficos b e d mostram queda das emissões durante os finais de semana com valores mais elevados para os outros dias. O gráfico c registra o aumento de amplitude durante o inverno.

Também para o NO₂ observa-se uma concordância com os resultados obtidos por meio de PCA para o perfil diário. Como no caso do NO, a queda brusca registrada na extremidade do gráfico c deve-se a valores inicialmente iguais a zero. Comparações entre os gráficos b e d nas Figuras 6.6, 6.7 e 6.8, mostra semelhança no perfil semanal para o CO, NO e NO₂.

Modelo PARAFAC para o O_3

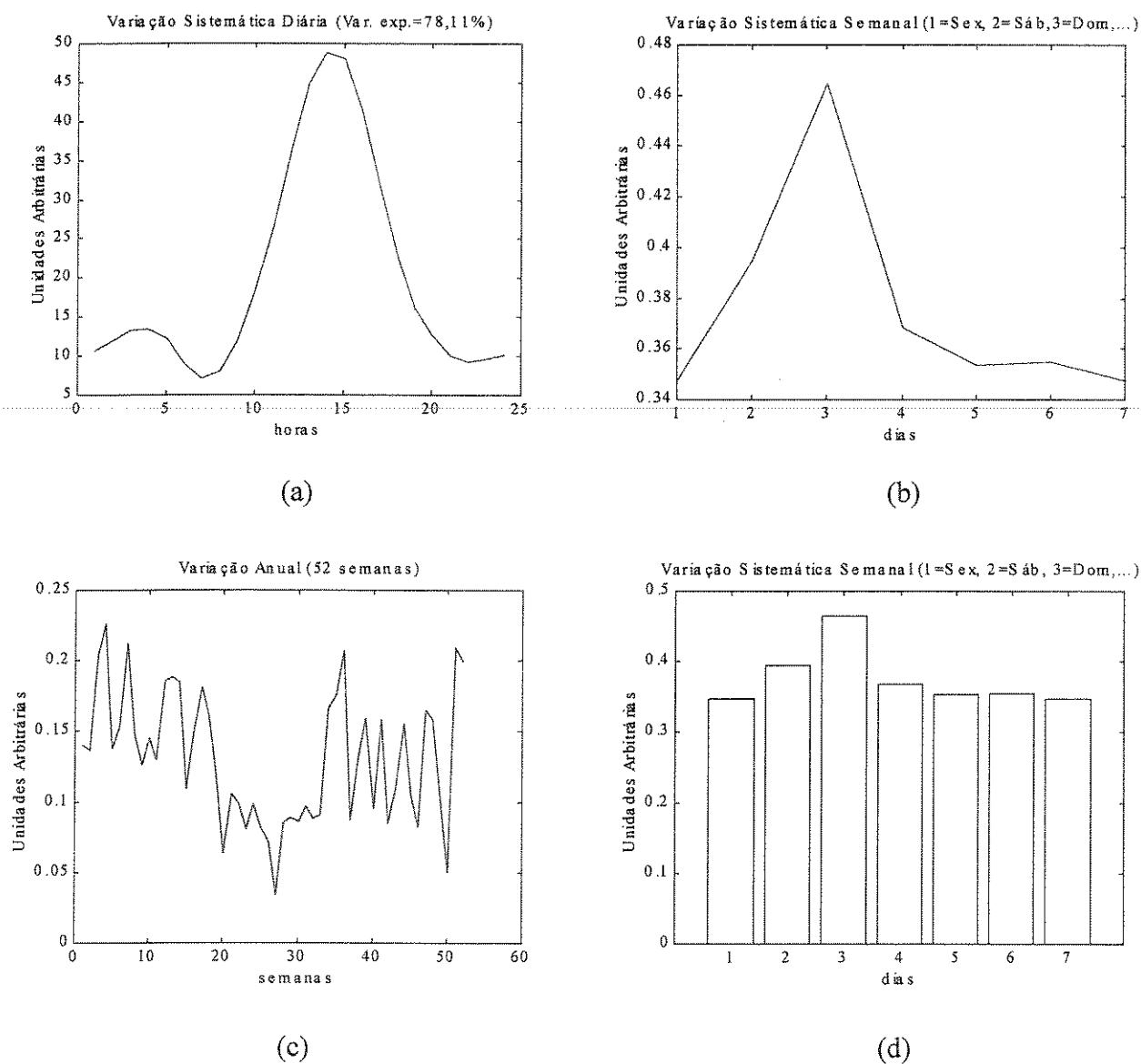


Fig.6.9: a) Variações sistemáticas com periodicidade diária. b) Variações sistemáticas com periodicidade semanal (idem gráfico d). c) Variação anual (não sistemática).

Como nos casos anteriores há concordância com os resultados obtidos com a análise PCA para o perfil diário. Os gráficos b e d mostram que, ao contrário do que ocorre com os poluentes primários (CO , NO e NO_2), para o O_3 há um acréscimo de emissões nos finais de semana. Esse resultado aparentemente contraditório é discutido no final desse capítulo. O gráfico c mostra que, diferentemente do que acontece com os poluentes primários, o O_3 acumula menos durante o inverno (resultado também amplamente discutido nesse capítulo).

Modelo PARAFAC para o SO₂ (sem influência de outra fonte)

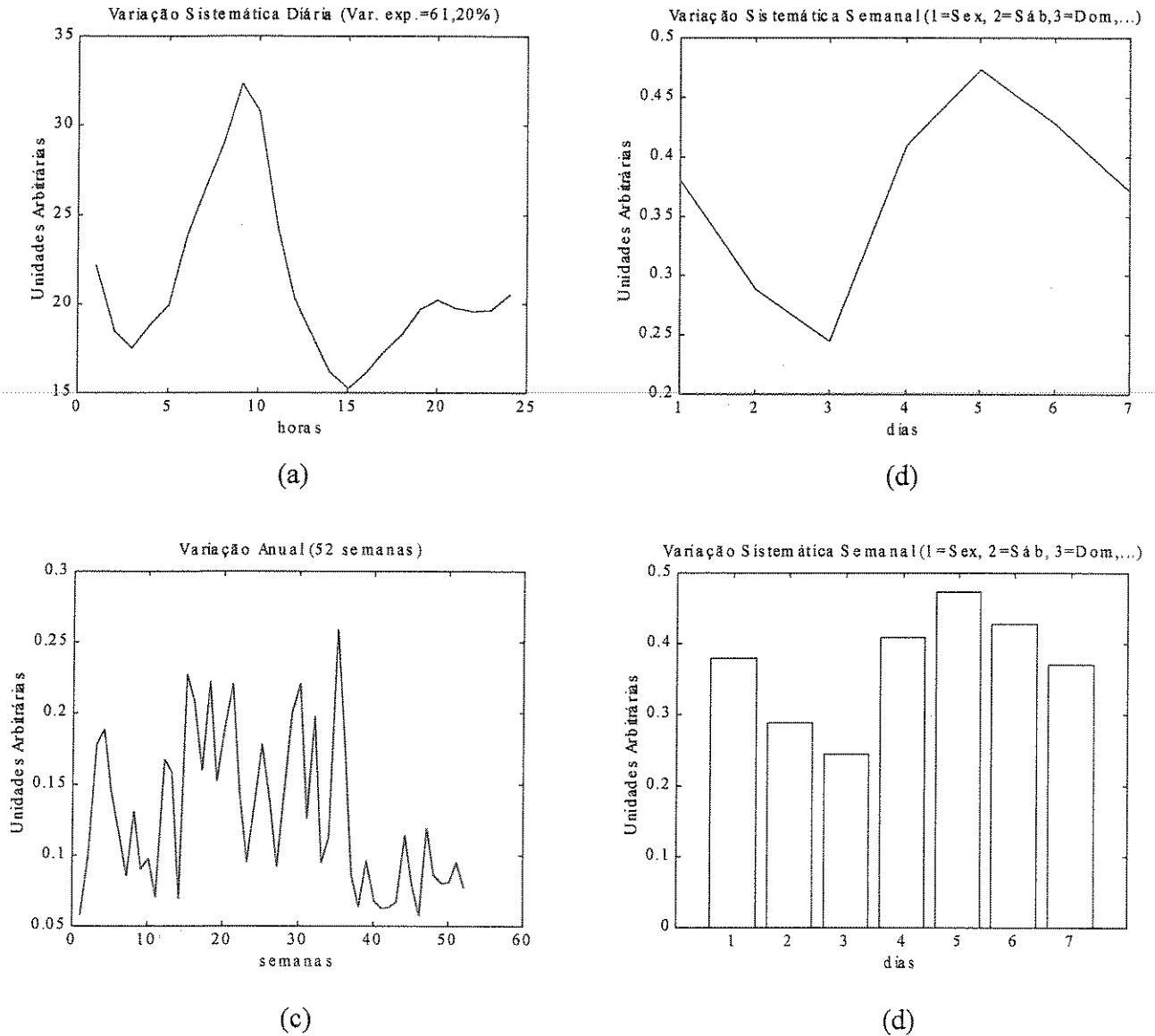


Fig.6.10.1: a) Variações sistemáticas com periodicidade diária. b) Variações sistemáticas com periodicidade semanal. c) Variação anual (não sistemática). Os gráficos b e d também mostram, como nos casos do CO, NO e NO₂, uma queda nas emissões durante os finais de semana. O gráfico c mostra aumento de amplitude durante o inverno, como nos casos dos poluentes primários.

Os resultados acima (sem o “padrão à 1:00 hora” do SO₂) para os cinco poluentes (CO, NO, NO₂, O₃ e SO₂) são consistentes em dois aspectos fundamentais: o primeiro é a consistência entre PCA e PARAFAC, e o segundo é a consistência em relação às emissões que se assemelham às das fontes automotivas (CO, NO, NO₂ e SO₂).

Modelo PARAFAC para o SO_2 (com a influência de outra fonte)

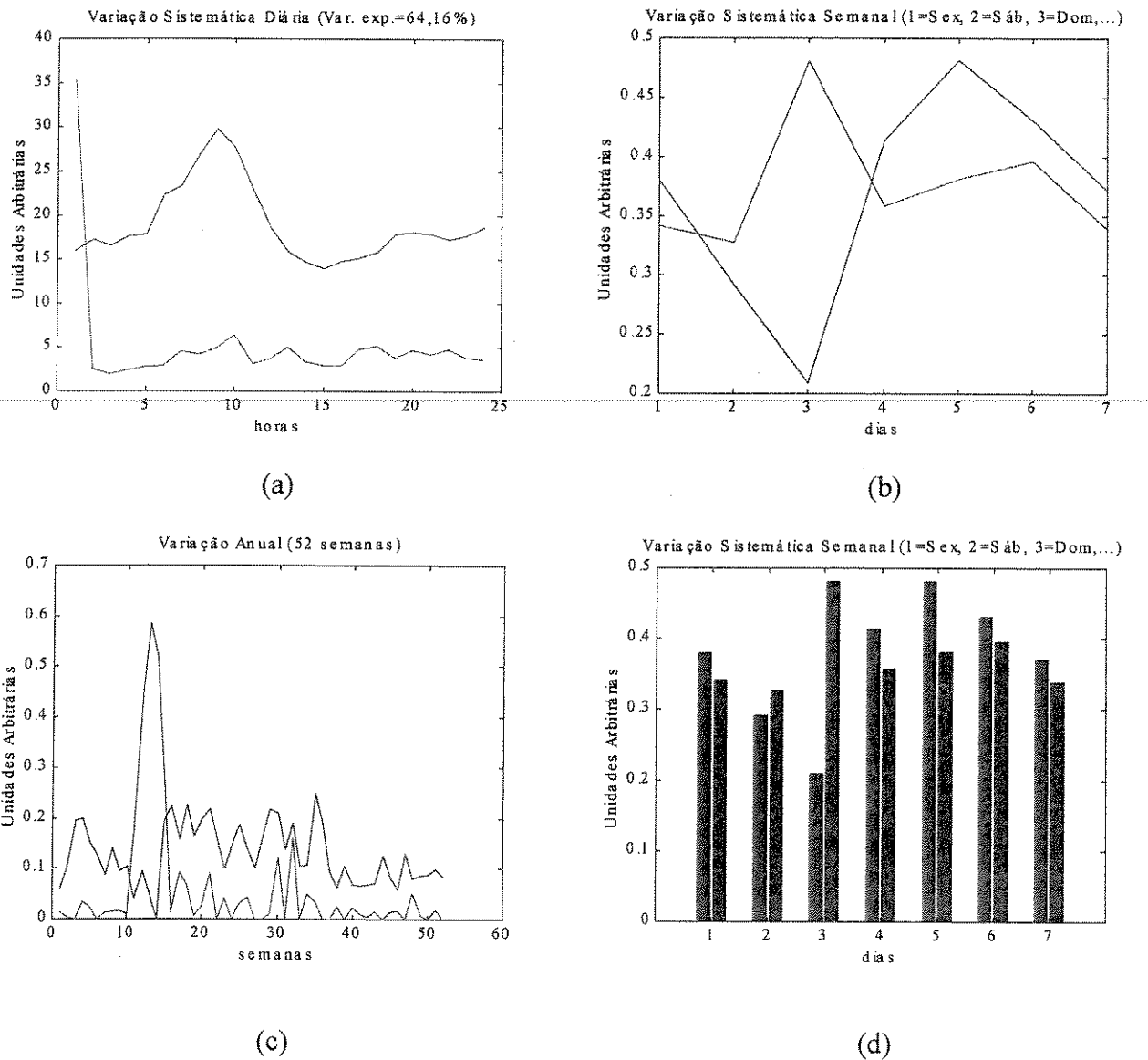


Fig.6.10.2: a) Variações sistemáticas com periodicidade diária. b) Variações sistemáticas com periodicidade semanal. c) Variação anual (não sistemática). d) Gráfico de barras para as variações sistemáticas semanais.

A Figura 6.10.2a mostra a influência das emissões à 1:00 h (valor mais elevado para a curva verde). Existe a possibilidade de que esse perfil seja causado por emissões pontuais (sempre naquele horário) de indústrias. O perfil na cor azul deve ser atribuído à fontes automotivas, o que pode ser confirmado por meio da Figura 6.10.2b e d, as quais mostram, em azul, o perfil de fontes automotivas (queda características nas emissões nos finais de

semana) e, em verde (6.10.1b e em marrom, 6.10.1d), as variações sistemáticas semanais para a outra fonte de emissão (muito provavelmente, industrial). A Figura 6.10.2c mostra o perfil anual. Note-se, em verde, a influência das emissões à 1:00 h.

O resultado apontado na Figura 6.4 (p245), especialmente, na Figura 6.4.2 (p248), conjugados com os das Figuras 6.4.1 (p247) e 6.10, respectivamente, para as análises PCA e PARAFAC indicam forte suspeita de que essas emissões sejam provenientes de indústrias. A curva na cor azul, na Figura 6.4.2a corresponde aos dados registrados pelo medidor à 1:00 h. Essa curva está totalmente deslocada do espectro correspondente aos dados registrados nas outras horas do dia, isto é, no período de 2:00 a 24:00 h. As figuras mostram que os valores em torno de 1:00 h no período afetado estão muito abaixo daqueles à 1:00 h. A Figura 6.4.2b mostra esse fato quando se passa de 1:00 h para 2:00 h.

A Tabela abaixo mostra as variâncias capturadas em cada caso, sem e com a influência do período afetado.

Tabela VI.1: Variâncias capturadas pelas três primeiras PCs (P. D. Pedro II – 1999)
 Percentagem de Variância capturada pelo modelo PCA

Sem a influência do período afetado			Com a influência do período afetado		
PC#	%Var.	%Var. Tot.	PC#	%Var.	%Var. Tot.
1	75.7814	75.7814	1	72.8013	72.8013
2	7.4802	83.2616	2	6.8472	79.6485
3	3.1768	86.4384	3	6.5211	86.1696

A variância capturada por PC₁, nos dois casos (sem e com a influência dos dados à 1:00 h), é muito maior do que a capturada pelas outras duas componentes, indicando que os frota de veículos é a principal fonte de emissões de SO₂ na localidade P. D. Pedro II. Isto deve estar associado principalmente com emissões de veículos movidos à óleo diesel tais como ônibus, caminhões e alguns veículos leves (Costa e Baldasano, 1996).

As variâncias para PC₂ e PC₃, considerando-se a influência dos dados à 1:00 h, são aproximadamente iguais. Isto pode significar a existência de dois tipos distintos de fonte industrial emitindo simultaneamente naquele horário (1:00 h) altas quantidades de SO₂. Por

outro lado podem ser emissões de fontes diferentes, uma delas, certamente, industrial. De qualquer modo, a fonte constituída pelas indústrias caracteriza-se como a segunda maior fonte de emissão de dióxido de enxofre naquela localidade.

A retirada da influência do período afetado considera a média das 23 horas do dia (no caso do P. D. Pedro II a média das 24 horas fornece o mesmo resultado) em substituição aos dados tomados à 1:00 h. Nesse caso PC_1 passa a capturar 75,78% da variância, enquanto PC_2 aumenta para 7,48% e PC_3 cai para 3,17%. Essa variação é esperada porque ao se substituir os dados para 1:00 h pela média das 23 horas (ou 24 horas) do dia transfere-se para a componente PC_1 maior parte da variação dos dados do que ocorre para PC_2 , enquanto a variância acumulada permanece aproximadamente a mesma (~86%).

Esse resultado mostra que PC_1 está, de fato, capturando variância de uma fonte distinta daquela que o faz PC_2 quando se considera o período afetado.

Modelo PARAFAC para a Temp

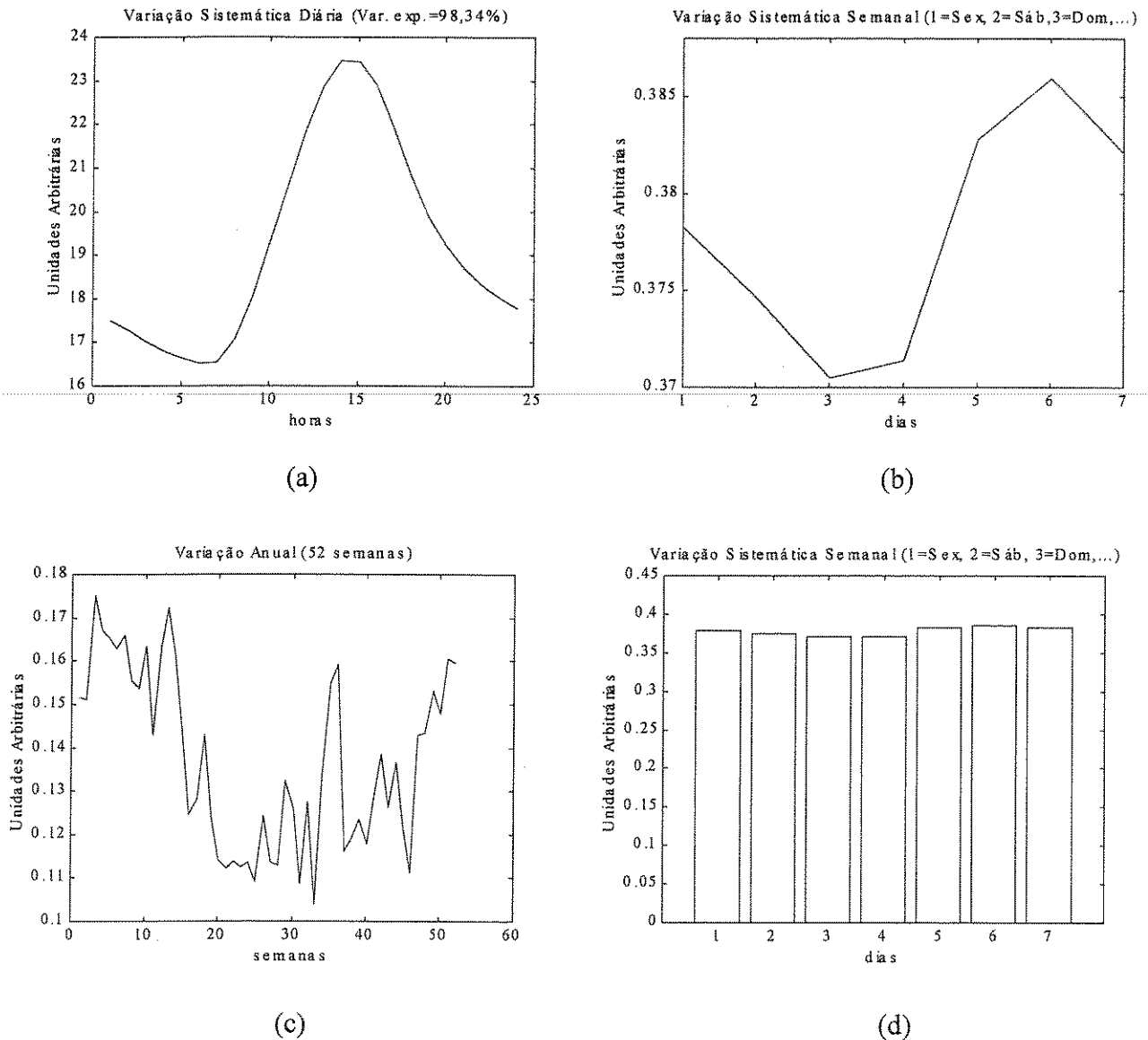


Fig.6.11: a) Variações sistemáticas com periodicidade diária. b) Variações sistemáticas com periodicidade semanal. c) Variação anual. d) Gráfico em barras para a variação anual. O perfil da temperatura registrado no gráfico c mostra queda da mesma durante o inverno.

Aqui também há concordância com PCA para o perfil diário. Os resultados para a temperatura também estão em concordância com o que se devia esperar para o perfil anual (temperaturas mais baixas no inverno) e para o perfil diário (temperaturas mais altas em torno das 14:00 horas). Uma inspeção nos valores para os pesos em b, e no gráfico d, mostram que durante a semana a temperatura mantém-se aproximadamente constante

Modelo PARAFAC para a Umid

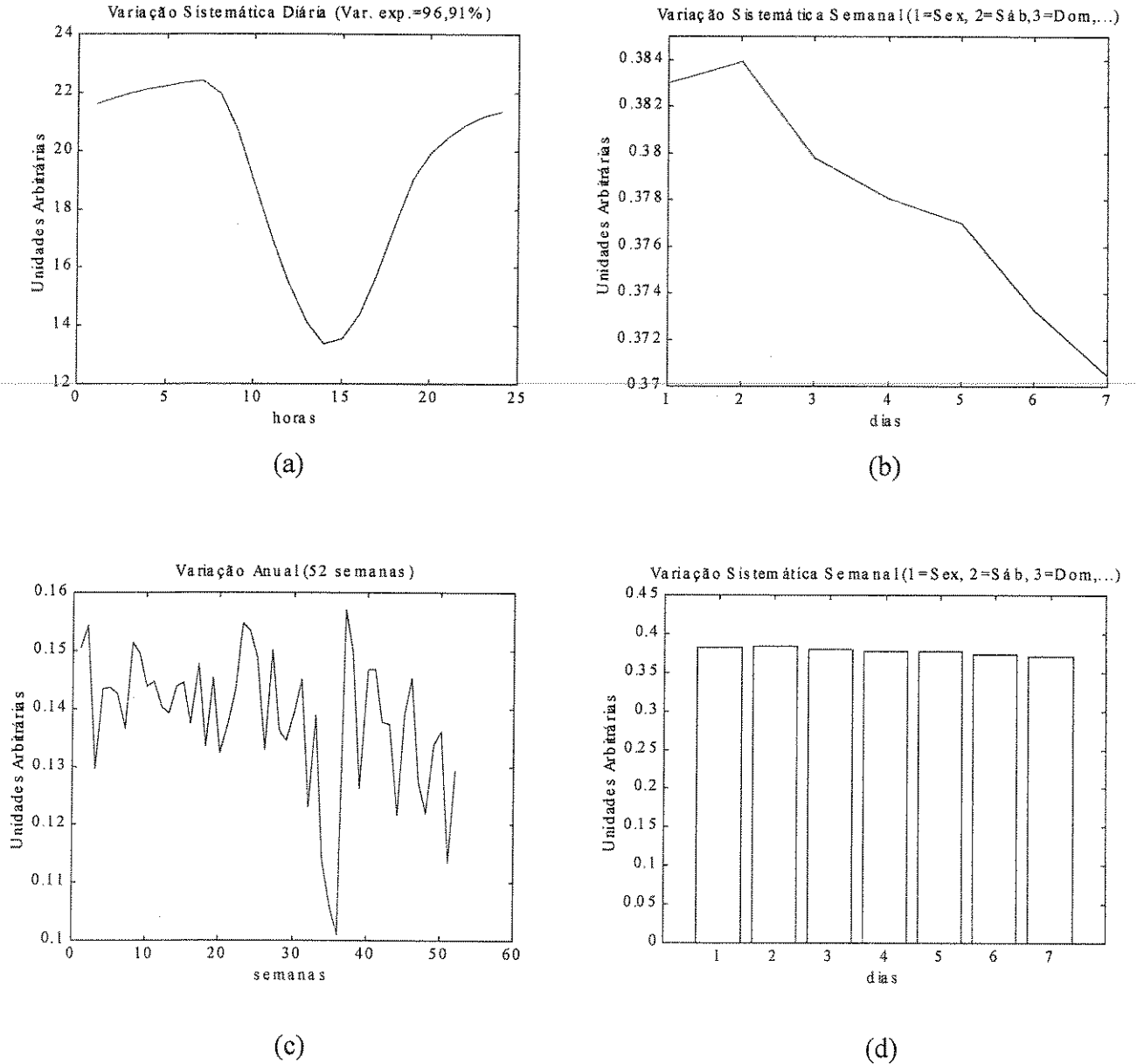


Fig.6.12: a) Variações sistemáticas com periodicidade diária. b) Variações sistemáticas com periodicidade semanal. c) Variação anual. d) Gráfico em barras para a variação semanal. O perfil do gráfico c mostra que a umidade é mais alta nos períodos de chuva.

Como nos demais casos há concordância com os resultados obtidos por meio de PCA para o perfil diário. Os resultados para a temperatura e umidade mostram concordância entre si. As Figuras 6.12b e 6.12d mostram que não ocorre variação significativa da umidade durante a semana.

Est.2: Estação São Caetano – 1999

6.3.1.7 – Análise de Monóxido de Carbono – CO

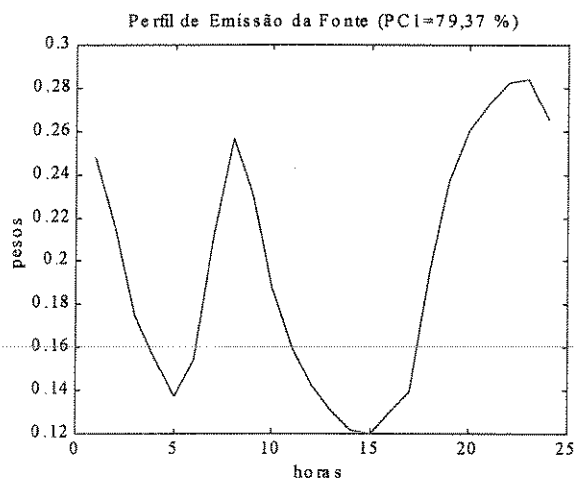


Fig.6.13: Pesos para a primeira componente principal, PC₁, para o CO.

A Figura 6.13 mostra o perfil de emissão da fonte (automotiva) para o monóxido de carbono na Estação de Medição São Caetano do Sul. O perfil tem as mesmas características dessas emissões para a Estação de Medição Parque D. Pedro II, com um pico na parte da manhã, em torno das 9:00 h e outro, o da noite, por de volta 22:00 h. As diferenças nos perfis (horário do segundo pico e profundidade dos vales) devem ser atribuídas às diferentes localizações das duas Estações de Medição, as quais estão sujeitas a diferentes condições atmosféricas e a diferentes quantidades de emissão. Entretanto, o que os dois locais possuem em comum, ou a característica que as tornam semelhantes, é o regime de tráfego a que estão sujeitas. Ou seja, as condições locais (velocidade e direção do vento, quantidade de emissões, tipo de emissões, etc) podem interferir causando variações na intensidade de cada pico, ou no horário de pico, porém o padrão de emissão não se altera. Em outras palavras, em função das condições locais, pode ocorrer variação nas quantidades que o medidor registra, isto é, entre a quantidade emitida pela fonte e aquela que o medidor, de fato, consegue captar. É essa variação que pode causar diferenças nos horários e alturas dos picos. Descartadas variações bruscas o suficiente para quebrar o padrão de emissão (por exemplo, tráfego de veículos vedado no local), se a fonte for automotora, o perfil de emissão sempre será o mesmo.

6.3.1.8 – Análise de Óxido de Nitrogênio e Dióxido de Nitrogênio – NO e NO₂

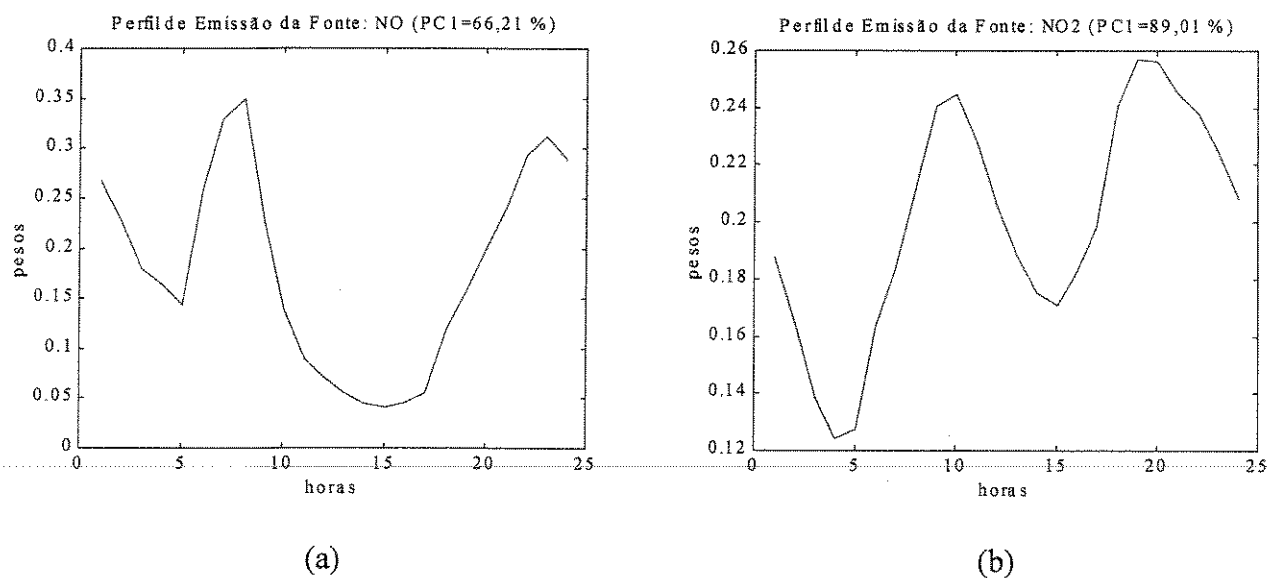


Fig.6.14: a) Pesos para a primeira componente principal, PC₁, para o NO. b) Pesos para a primeira componente principal, PC₁, para o NO₂.

As Figuras 6.14a e 6.14b mostram, respectivamente, os perfis de emissão do NO e do NO₂ na Estação de Medição São Caetano do Sul. Aqui também ocorrem diferenças na intensidade e no horário para o segundo pico do NO. Para o NO₂ essas diferenças já não se fazem notar em relação ao encontrado na Estação de Medição Parque D. Pedro II.

6.3.1.9 – Análise de Ozônio – O₃

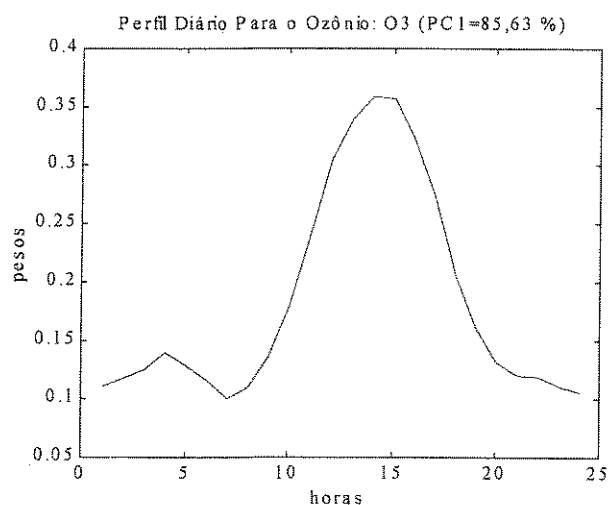


Fig.6.15: Pesos para a primeira componente principal, PC₁, para o O₃.

A Figura 6.15 mostra que os perfis para o O₃ em São Caetano e Pedro II são iguais.

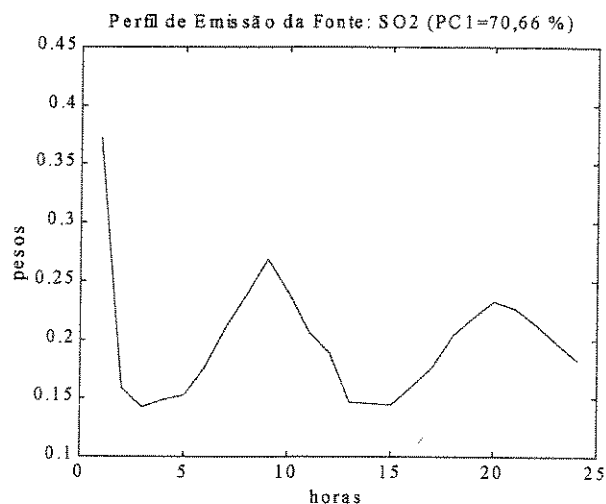
6.3.1.10 – Análise de Dióxido de Enxofre – SO₂

Como já citado anteriormente, o comportamento do SO₂ durante o ano de 1999 nas três localidades (Pedro II, S. Caetano e Ibirapuera) é distinto do ocorrido no período 1997-1998. Para a Estação de Medição São Caetano do Sul, o fenômeno que ocorre à 1:00 h é mais pronunciado do que na localidade Pedro II, isto é, sua influência é maior nesse segundo caso. Da mesma forma que no caso anterior, a análise foi conduzida sem e com a inclusão de tais dados (1:00 h) para se avaliar se eles afetavam o perfil devido às emissões provenientes do tráfego de veículos.

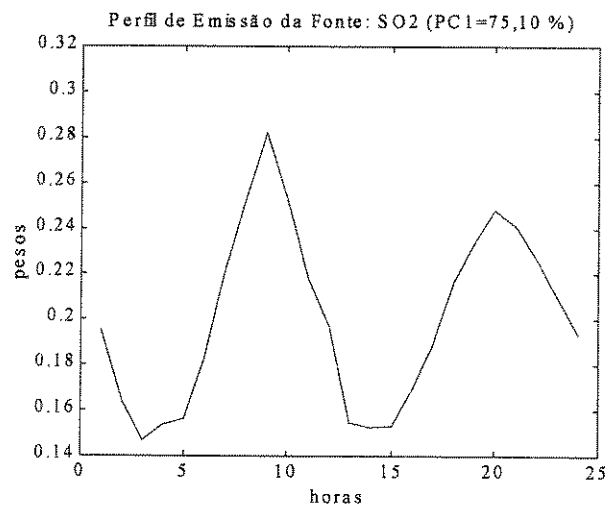
Para essa Estação, o período total afetado (42 dias, contendo 1008 pontos de dados) é constituído de três partes. A primeira (13 dias) vai do 128° ao 140° dias (~ duas semanas, 08/05 a 20/05, num total de 312 pontos de dados). A segunda (26 dias), no intervalo entre o 177° e 202° dias (três e meia semanas, 26/06 a 21/07, num total de 624 pontos de dados). A terceira (3 dias) no período do 215° ao 217° dias (meia semana, 02/09 a 04/09, num total de 72 pontos de dados). A operação envolvendo a influência dos dados aparentemente anômalos foi realizada como anteriormente. Na primeira, foram retirados todos os valores para 1:00 h no período afetado, sendo a análise conduzida com os 23 pontos de dados diários durante o período em questão. Na segunda, calculou-se a média para as vinte e quatro horas de cada dia (42 médias), sendo essas médias usadas em substituição aos valores correspondentes iniciais para 1:00 h naquele período. Os dados foram analisados por PCA e PARAFAC não sendo detectada qualquer discrepância entre as análises segundo as duas operações (sem e com os dados suspeitos).

Adicionalmente, foram calculadas as médias para as 23 horas do dia (de 2:00 a 24:00 h) para verificar uma possível influência de valores suspeitos na média, resultando em médias inadequadas. Para o caso da Estação de Medição Parque D. Pedro II, não houve diferença significativa nos resultados. Para a Estação São Caetano do Sul, ocorre uma ligeira diferença. Como será visto adiante, isto também ocorre no Ibirapuera. Assim, foi usada a média das 23 horas do dia em substituição àqueles para uma hora, quando se tratou de analisar os perfis sem a influência dos dados suspeitos.

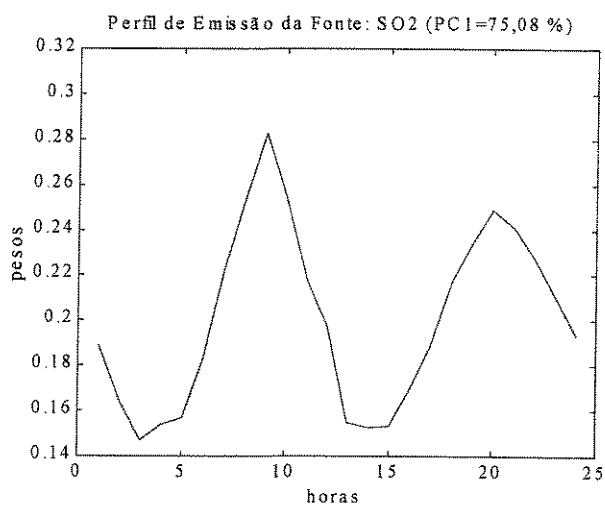
Esses resultados são mostrados nas Figuras 6.16a, b e c para a análise PCA. Para o modelo PARAFAC, são mostrados nas Figuras 6.22.1 a 6.22.3.



(a)



(b)



(c)

Fig.6.16: Pesos para PC₁, para o SO₂. a) 2 fontes. b) 1 fonte (média de 24 horas para os dados da 1:00 h). c) 1 fonte (média de 23 horas, isto é, de 2 a 24 horas, para os dados à 1:00 h).

A Figura 6.16a mostra os pesos para PC₁ para a variável SO₂ quando se leva em conta na análise PCA os dados referentes à 1:00 h. O efeito é bastante evidente no gráfico a. A figura 6.16b mostra o mesmo resultado quando os dados para 1:00 h são substituídos pela média das 24 horas de cada dia. A figura 6.16c também se refere a mesma análise, porém com os dados para a 1:00 h substituídos pela média das 23 horas (2:00 às 23:00 h) restantes de cada dia. É interessante notar que os dados referentes à 1:00 h não interferem no perfil diário da fonte automotiva. As Figuras 6.16b e c mostram que, na verdade, uma segunda fonte predomina sobre as emissões automotivas à 1:00 h (Fig. 6.16a).

Semelhantemente ao realizado para a localidade P. D. Pedro II, são mostradas, para São Caetano, as figuras correspondentes às matrizes de dados originais e de dados originais corrigidos, para melhor visualização da curva para 1:00 h no período afetado.

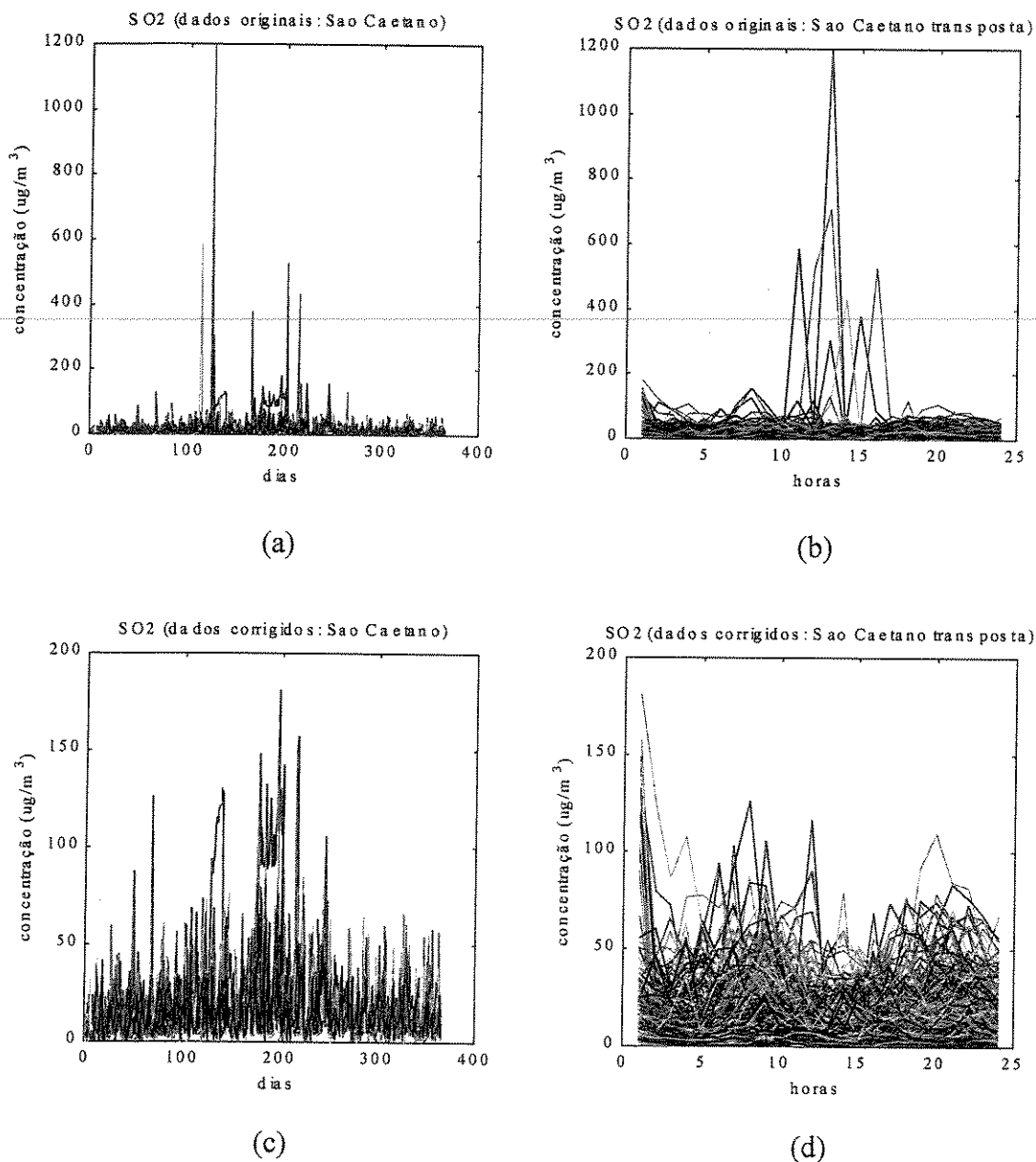


Fig.6.16.1: Dados originais: a) concentração $\times$ dia, b) concentração $\times$ hora. Dados corrigidos: c) concentração $\times$ dia, d) concentração $\times$ hora.

As Figuras 6.16.1a e 6.16.1b mostram, respectivamente, os gráficos correspondentes à matriz para os dados originais e à matriz transposta desses dados. Embora seja mais fácil

visualizar a curva para 1:00 h em São Caetano, quando comparado à Pedro II, ainda não está tão clara aquela curva. Após as correções para dados anômalos e variações súbitas, obtém-se as Figuras 6.16.1c e 6.16.1d, as quais representam, respectivamente, a matriz de dados originais e sua transposta.

Todavia, a ampliação do período afetado permite constatar de forma cabal a ocorrência do fato, como mostra a próxima figura (6.16.2).

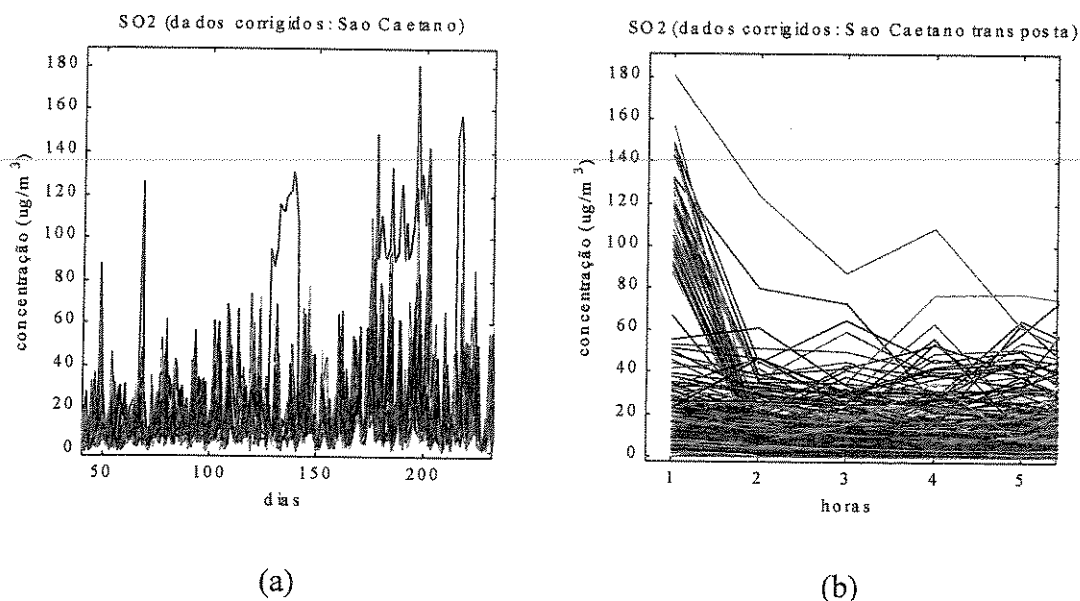
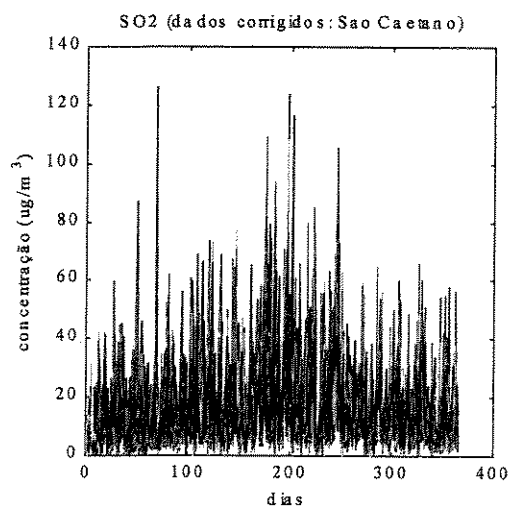


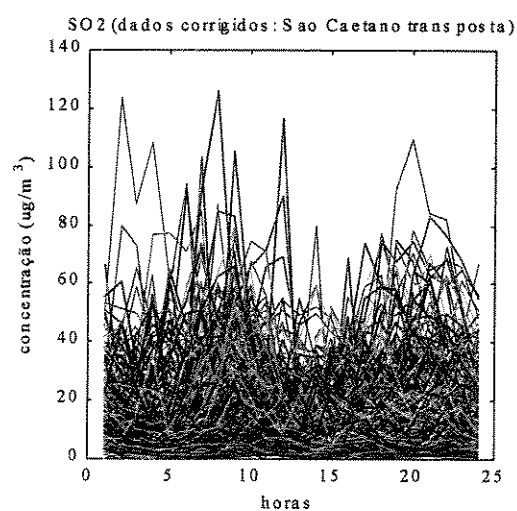
Fig.6.16.2: Ampliação ("zoom") sobre dados corrigidos no período afetado. a) concentração $\times$ dia, b) concentração $\times$ hora.

Os trechos contínuos de curva na cor azul (128° ao 140° dia, 177° ao 202° dia e do 215° a 217° dia), na parte superior da figura, mostram os valores para 1:00 h. Esse trecho de curva escapa do espectro mais denso situado mais abaixo na figura. Tal efeito pode ser visto na Figura 6.16.2b, na qual um conjunto de valores acima de aproximadamente 80 para os eixos, forma um feixe com queda súbita para valores muito mais baixos (abaixo de ~20) às 2:00 h (lado esquerdo do gráfico b).

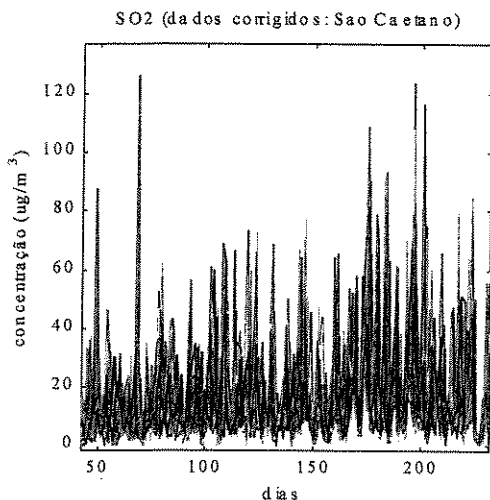
Os resultados para o SO₂ com a influência dos dados à 1:00 h obtidos para a Estação de medição São Caetano devem ser comparados com aqueles nos quais a influência daqueles dados foi retirada para essa mesma localidade. A Figura 6.16.3, à frente, mostra o espectro obtido quando tal influência é removida.



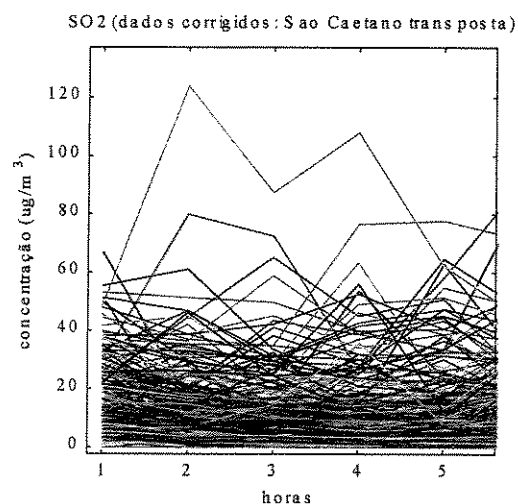
(a)



(b)



(c)



(d)

Fig.6.16.3: Dados corrigidos: a) concentração $\times$ dia, b) concentração $\times$ hora. Ampliação (“zoom”) sobre dados corrigidos no período afetado sem os dados à 1:00 h. c) concentração $\times$ dia, d) concentração $\times$ hora.

A Figura 6.16.3a mostra o espectro sem os dados para 1:00 h no período afetado e em 6.16b é mostrada a transposta desses dados (dados corrigidos). Não se observa em 6.16.3a, a curva em azul que aparece nas Figuras 6.16.1c e o feixe que aparece em 6.16.1d. Quando o período afetado é ampliado torna-se mais fácil constatar a ausência da curva azul e do feixe de quedas bruscas. A comparação deve ser feita com as Figuras 6.16.2a e

6.16.2b. Em 6.16.3b, nota-se a ausência do “feixe” de quedas súbitas na mudança de 1:00 h para 2:00 h.

Cabe assinalar que os dois picos na cor azul que aparecem na extremidade esquerda, tanto da Figura 6.16.3a quanto 6.16.3b, não são valores registrados à 1:00 h, mas sim às 8:00 h. Esses valores aparecem no 49º e 68º dias. Isto significa que o programa utilizado nas análises usa como “default” sete cores, com início na azul. Significa também que não se pode identificar somente pela cor da curva qual a hora que está sendo registrada. Para o caso da cor azul pode-se ter o dado para 1:00 h, para 8:00 h, para 15:00 h ou para 22:00 h. Há que se recorrer à matriz de dados para identificar exatamente o horário procurado.

Deve ser mais uma vez enfatizado que os dados para 1:00 h acima referidos não foram simplesmente retirados, mas corrigidos pela média das 23 horas (2:00 h a 23:00 h). Isto significa que fora do período afetado os dados estão corrigidos apenas para dados anômalos e variações súbitas. Obviamente, internamente ao período afetado, além dessas correções foram aplicadas aquelas sobre os valores à 1:00 h.

6.3.1.11 – Análise da Temperatura e da Umidade – Temp e Umid

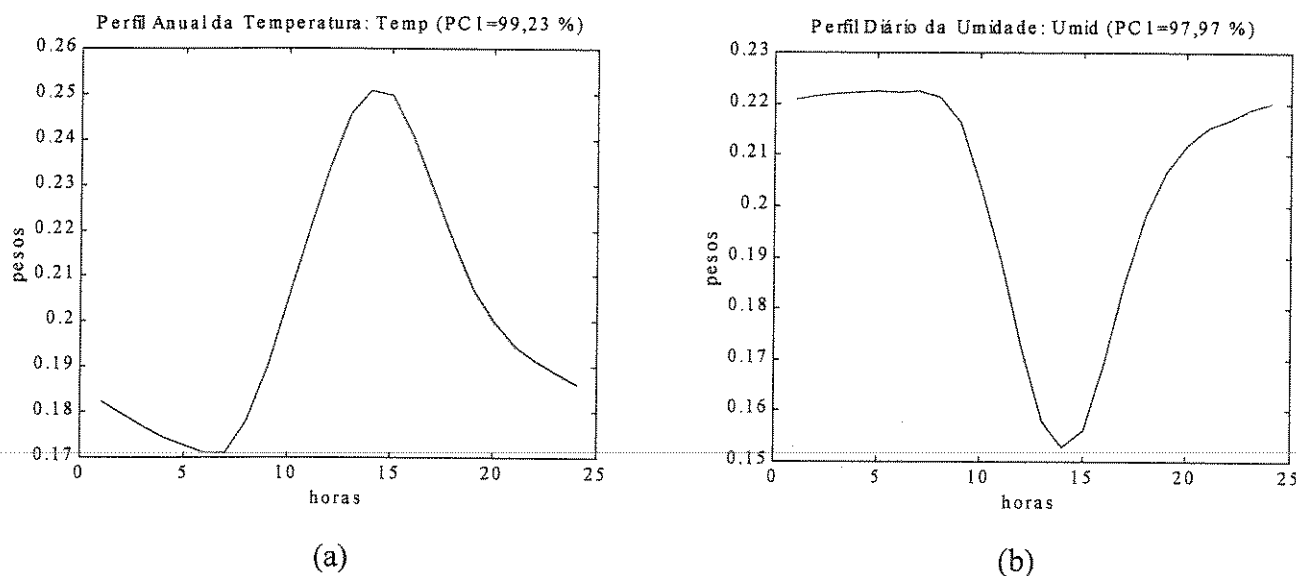


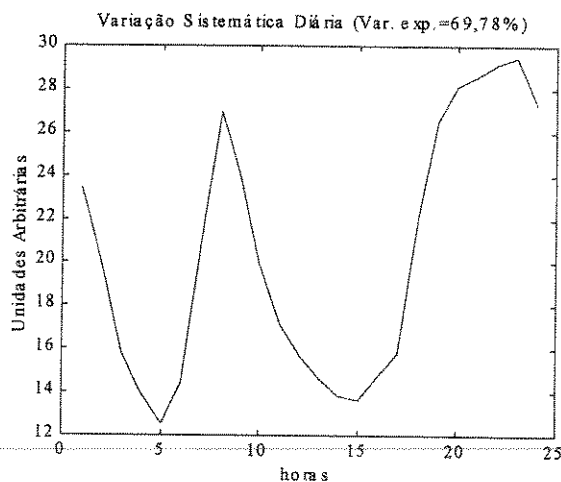
Fig.6.17: Pesos para a primeira componente principal, PC_1 . a) Temperatura e b) Umidade.

As Figuras 6.17a e 6.17b mostram, respectivamente, os perfis diários para a temperatura e para a umidade. Como se vê, tanto o pico para a temperatura, quanto o mínimo para a umidade, ocorrem às 14:00 h. Nota-se, portanto, a característica de simetria apresentada pelos perfis.

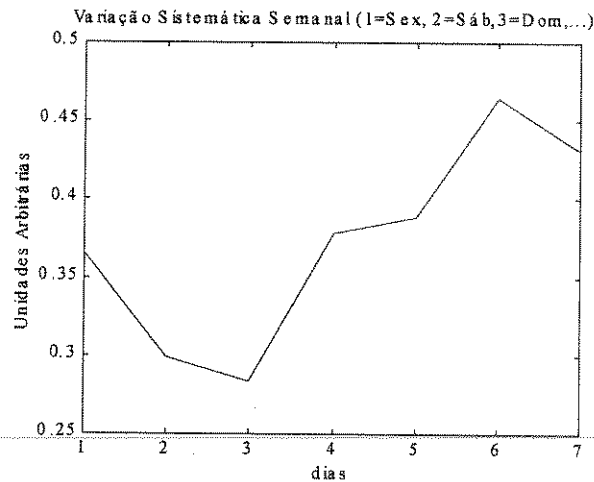
As próximas Figuras (6.18 a 6.24) mostram os resultados obtidos com o modelo PARAFAC para as variáveis CO, NO, NO₂, O₃, SO₂, Temp e Umid, para a Estação de Medição São Caetano do Sul durante o ano de 1999. Será útil comparar os pesos obtidos com a análise PCA e com o PARAFAC para o perfil diário nessa localidade. O modelo PARAFAC, como já antecipado, fornece também as variações sistemáticas semanais e a variação anual. Esses resultados deverão ser comparados com aqueles obtidos para o Parque D. Pedro II.

Como no caso do P. D. Pedro II, em São Caetano do Sul ocorre o mesmo problema, e em maior extensão, com os dados coletados à 1:00 h para o SO₂. Assim os resultados PARAFAC para essa variável também serão apresentados tendo em vista aquele fenômeno (isto é, sem e com a influência daqueles dados).

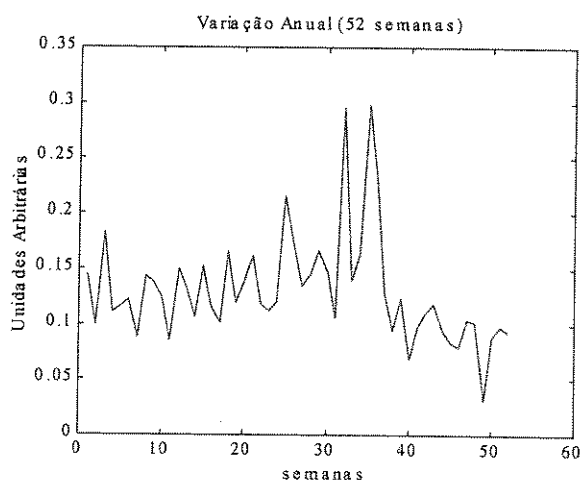
6.3.1.12 – Análises PARAFAC (São Caetano): Modelo PARAFAC para o CO



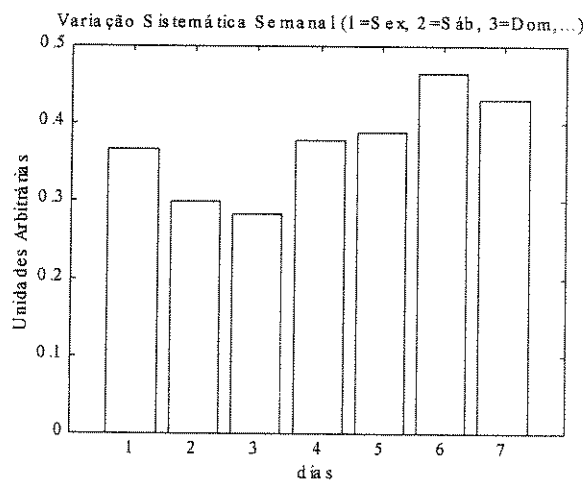
(a)



(b)



(c)

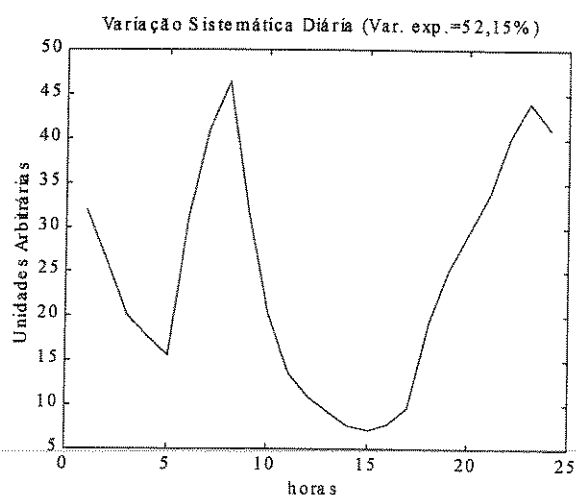


(d)

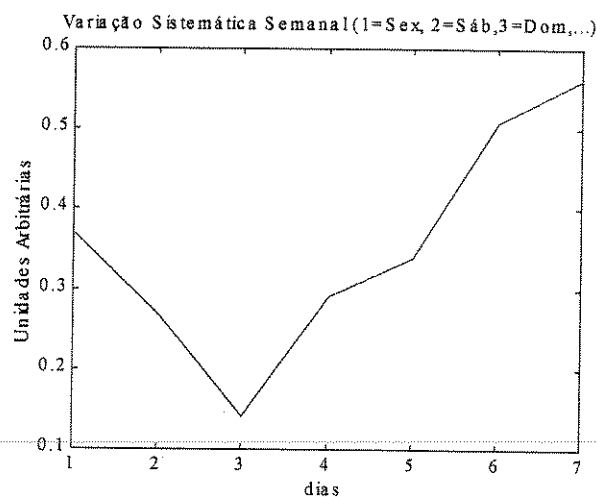
Fig.6.18: a) Variações sistemáticas com periodicidade diária. b) e d) Variações sistemáticas com periodicidade semanal. c) Variação anual (não sistemática). Os gráficos b e d mostram que as emissões caem nos finais de semana, sendo mais elevadas nos chamados dias úteis.

A Figura 6.18a mostra concordância com os resultados obtidos com a análise PCA para o perfil diário. As Figuras 6.18b e 6.18d mostram a queda das emissões nos finais de semana. Pode-se notar também, na Figura 6.18c, o aumento de amplitude nas semanas correspondentes ao período de inverno

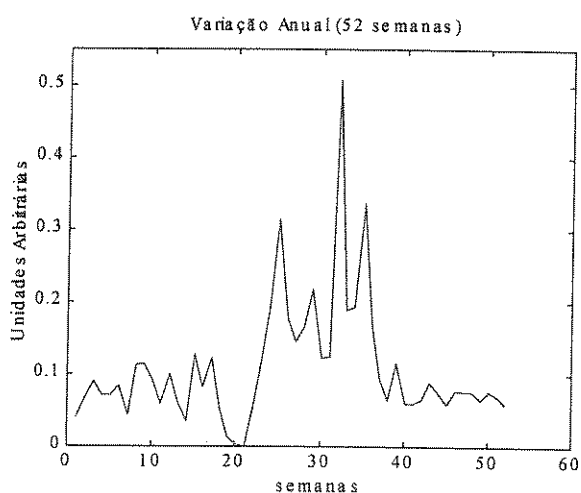
Modelo PARAFAC para o NO



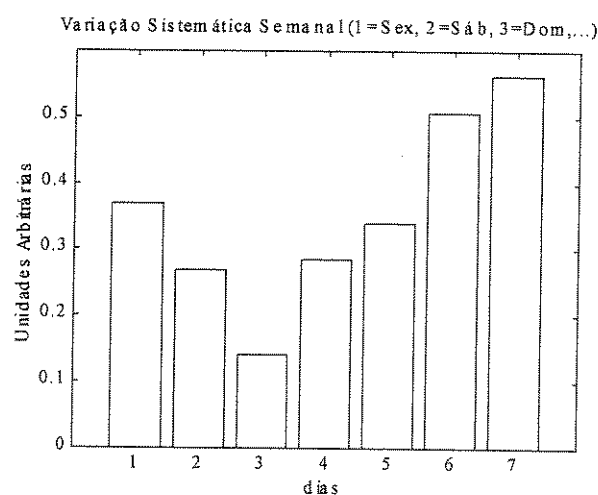
(a)



(b)



(c)

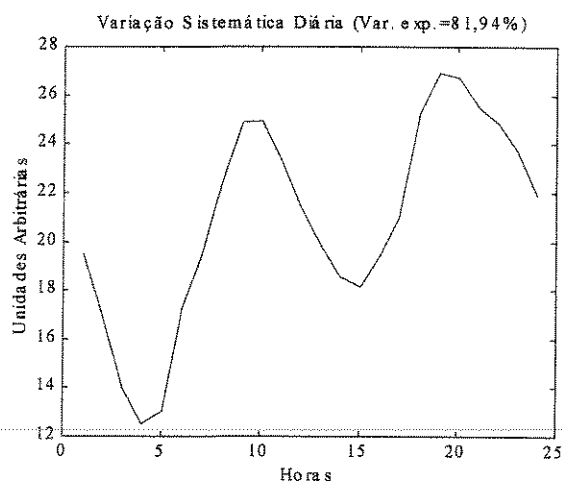


(d)

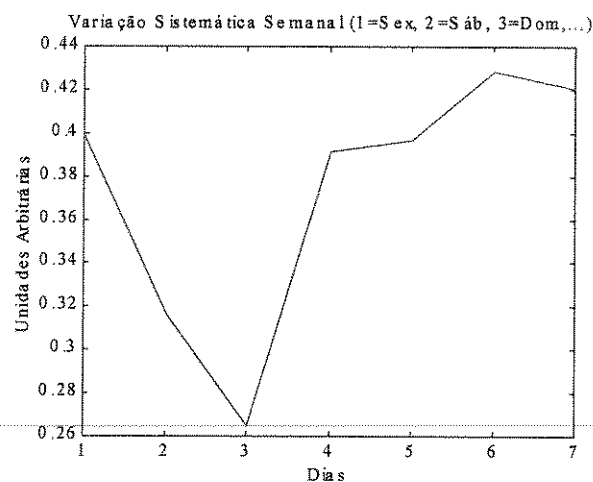
Fig.6.19: a) Variações sistemáticas com periodicidade diária. b) Variações sistemáticas com periodicidade semanal. c) Variação anual (não sistemática). d) Variações sistemáticas com periodicidade semanal.

Observe-se em 6.19a, a concordância com os resultados obtidos por meio da análise PCA para o perfil diário. Em 6.19b e 6.18d verifica-se a queda nas emissões durante os finais de semana, ao contrário do que ocorre nos chamados dias úteis. Nota-se também, em 6.19c, o aumento de amplitude nas semanas correspondentes ao período seco (inverno).

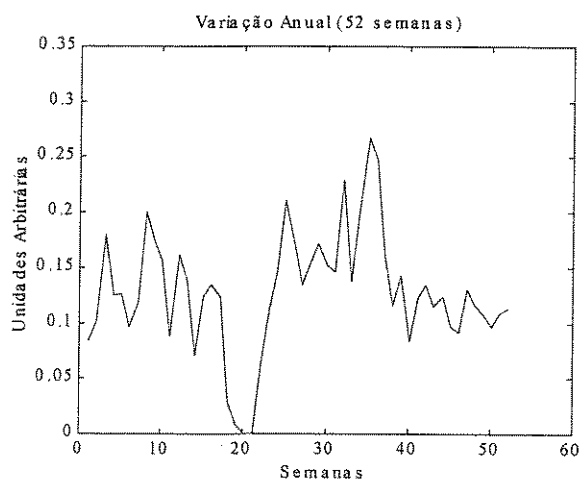
Modelo PARAFAC para o NO₂



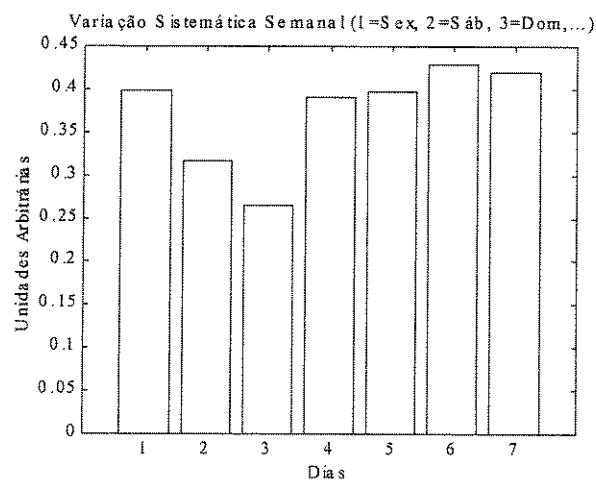
(a)



(b)



(c)



(d)

Fig.6.20: a) Variações sistemáticas com periodicidade diária. b) e d) Variações sistemáticas com periodicidade semanal. c) Variação anual (não sistemática).

A Figura 6.20 a mostra, para o NO₂, concordância com os resultados PCA para o perfil diário. As Figuras 6.20b e 6.20d mostram queda das emissões nos finais de semana. A Figura 6.20c registra o aumento de amplitude durante o inverno. Como no caso do NO, a queda brusca próximo da 20ª semana em 6.20c, deve-se aos valores originalmente nulos.

Modelo PARAFAC para o O_3

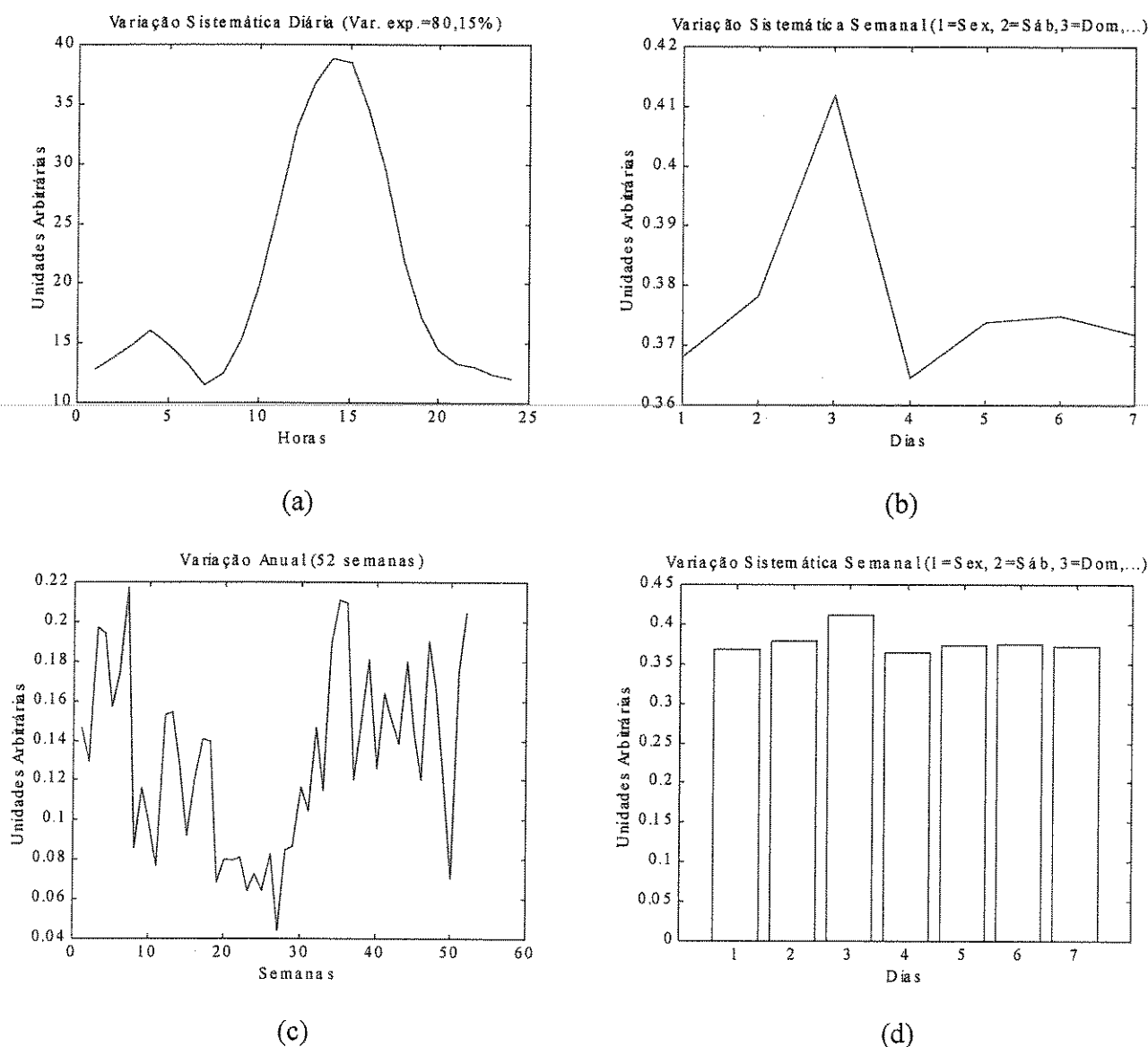


Fig.6.21: a) Variações sistemáticas com periodicidade diária. b) Variações sistemáticas com periodicidade semanal. c) Variação anual (não sistemática). d) Variações sistemáticas com periodicidade semanal em gráfico de barras.

Do mesmo modo que nos casos anteriores há concordância com os resultados obtidos com a análise PCA para o perfil diário (consistência nos picos e na correlação negativa com os poluentes primários), como mostra a Figura 6.21a. As Figuras 6.21b e 6.21d mostram também para São Caetano que, ao contrário do que ocorre com os poluentes

primários (CO, NO e NO₂), para o O₃ há um acréscimo de emissões nos finais de semana. Esse resultado é discutido mais adiante nesse capítulo. O gráfico 6.21c também mostra que, diferentemente do que acontece com os poluentes primários, o O₃ acumula menos durante o inverno (resultado também discutido mais adiante nesse capítulo). O perfil diário para o ozônio (Figura 6.21a) apresenta um pequeno pico em torno de 3:00 h – 4:00 h em todos os casos analisados. A explicação para isto não é trivial já que o ozônio é formado por meio da reação de oxigênio molecular com oxigênio atômico. Esse último existe (naturalmente) na troposfera em concentrações muito baixas. Então é a fotodissociação de NO₂ que leva à produção dos átomos de oxigênio necessários à formação de O₃ durante o dia. A ausência de luz solar no período noturno inviabiliza a produção de ozônio por essa via. Assim, procurou-se também no decorrer das discussões nesse capítulo, abordar o problema de modo a se oferecer contribuições aos estudos que consideram a existência do Ozônio noturno.

Modelo PARAFAC para o SO₂ (sem a influência de outra fonte)

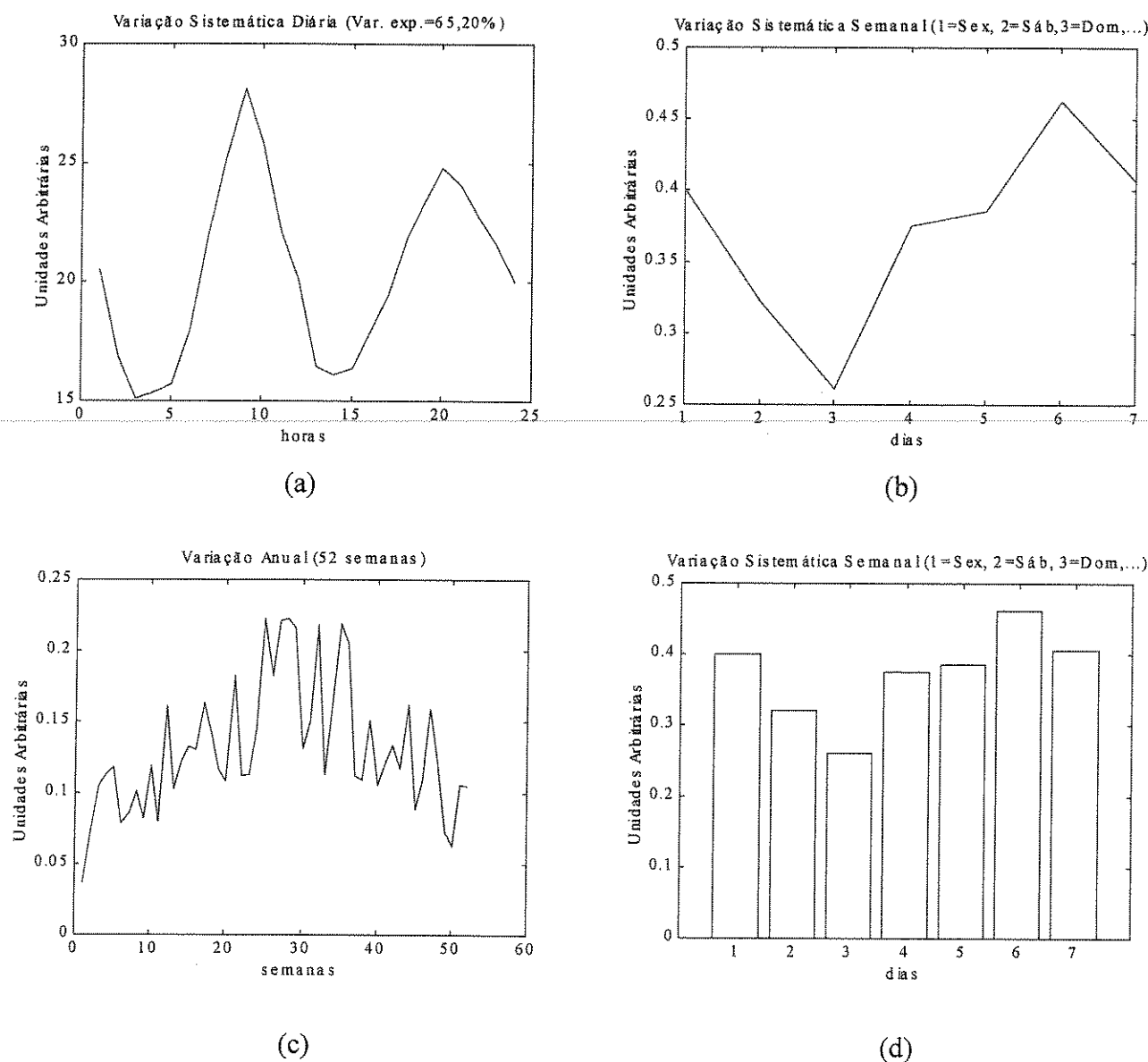


Fig.6.22.1: Modelo PARAFAC para o SO₂ na Estação de Medição São Caetano do Sul em 1999 (apenas uma fonte, a automotiva, e dados para 1:00 h substituídos pela média das 24 horas de cada dia no período afetado). a) Variações sistemáticas com periodicidade diária. b) e d) Variações sistemáticas com periodicidade semanal. c) Variação anual.

Os resultados registrados em 6.22.1a, 6.22.1b e 6.22.1d mostram que o perfil para a fonte de emissão automotiva são consistentes com todos os resultados anteriores. Adiante são apresentadas as Figuras 6.22.2 (apenas uma fonte, a automotiva, e dados para 1:00 h substituídos pela média das 23 horas, isto é, de 2:00 às 23:00 h, de cada dia no período afetado) e 6.23 (duas fontes). Em seguida a discussão em torno desses resultados.

Modelo PARAFAC para o SO_2 (sem a influência da 2ª fonte, com 2ª média)

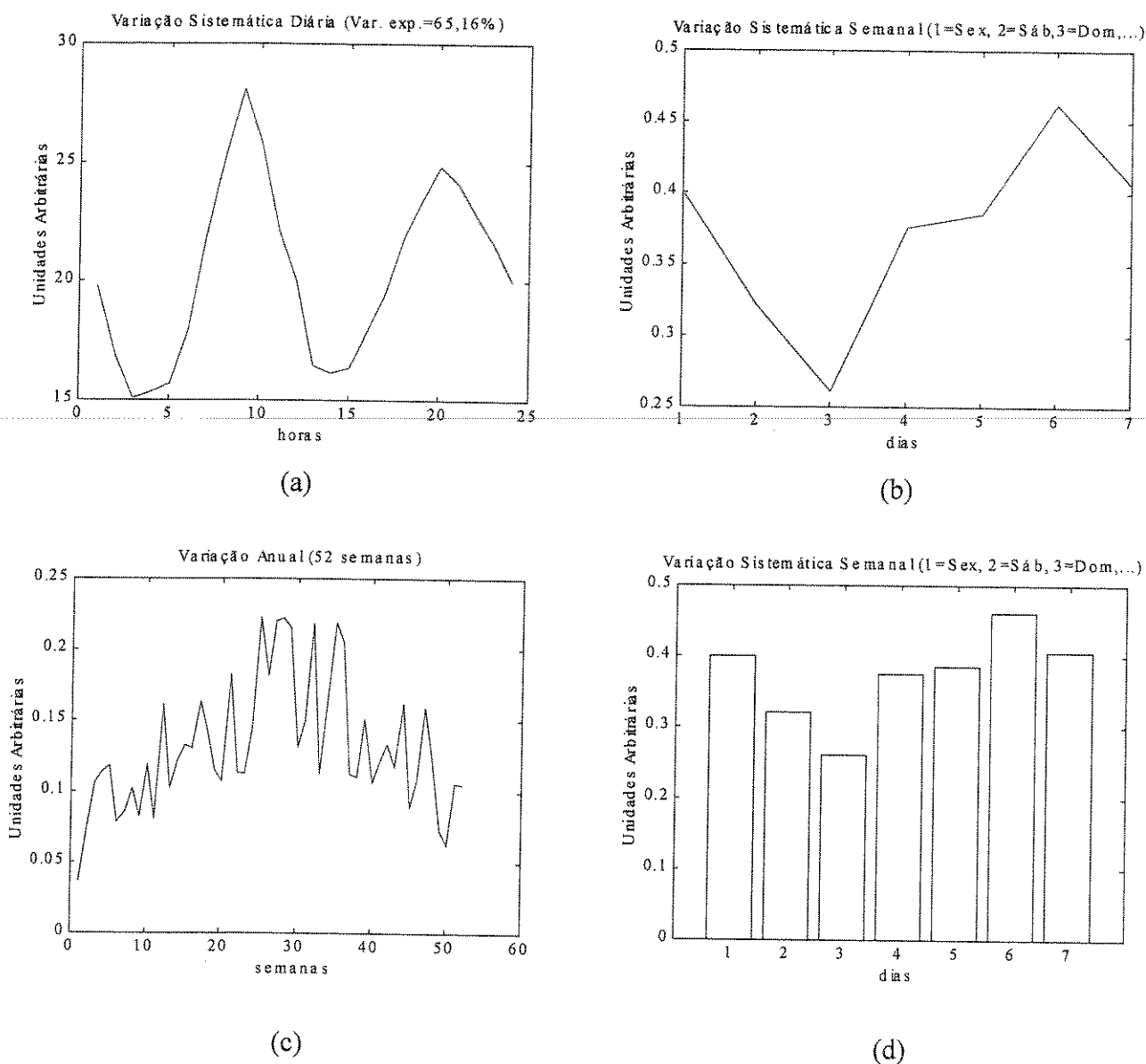


Fig.6.22.2: Modelo PARAFAC para o SO_2 na Estação de Medição São Caetano do Sul em 1999 (apenas uma fonte, a automotiva, e dados para 1:00 h substituídos pela média das 23 horas, isto é, de 2:00 às 23:00 h, de cada dia no período afetado). a) Variações sistemáticas com periodicidade diária. b) e d) Variações sistemáticas com periodicidade semanal. (c) Variação anual.

As Figuras 6.22.2a, 6.22.2b e 6.22.2d mostram que os perfis para a fonte de emissão automotiva são consistentes com todos os resultados anteriores. Do mesmo modo que no caso anterior (Figura 6.22.1c), ocorre maior acumulação do poluente no período de inverno (Figura 6.22.2c).

Modelo PARAFAC para o SO₂ (2 fontes)

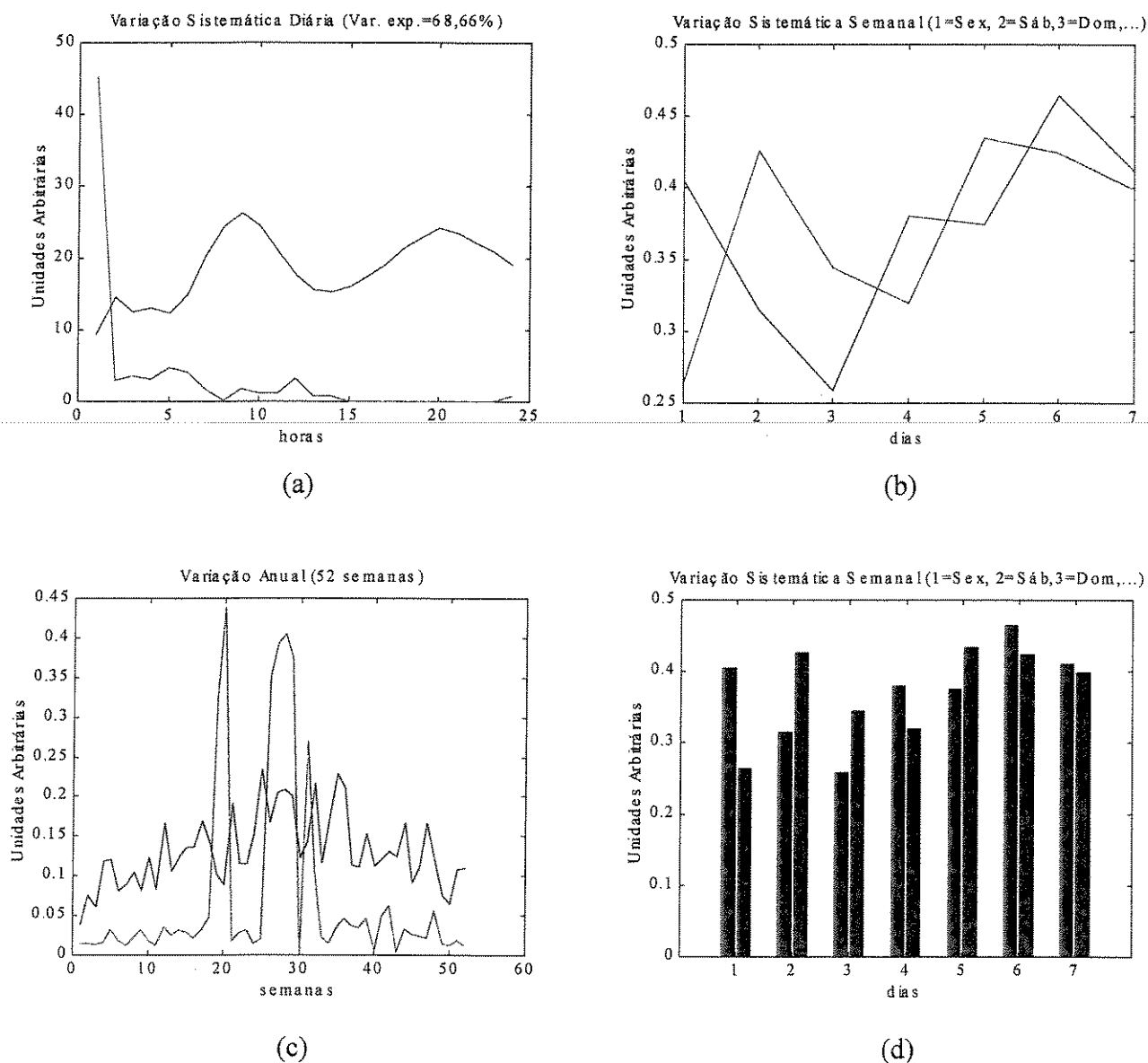


Fig.6.22.3: Modelo PARAFAC para o SO₂ na Estação de Medição São Caetano do Sul em 1999 (duas fontes). a) Variações sistemáticas com periodicidade diária. b) Variações sistemáticas com periodicidade semanal. c) Variação anual. d) Variações sistemáticas com periodicidade semanal. O perfis em azul, em a, b e c representam as emissões automotoras, enquanto o verde representa a 2^a fonte. A cor azul em d representa a fonte automotiva enquanto a marrom representa a 2^a fonte.

A Figura 6.22.3 mostra os resultados da aplicação do modelo PARAFAC aos dados da variável SO_2 na Estação São Caetano do Sul durante o ano de 1999. O gráfico a da referida figura mostra os perfis (os dois primeiros pesos da matriz de componentes) correspondentes às duas fontes de emissão captadas pelo modelo. O primeiro deles, na cor azul representa o perfil de emissão (diário) da fonte automotiva, e o segundo, em verde, representa uma segunda fonte. Nota-se que, nesse caso, os dados referentes à 1:00 h exercem uma grande influência no perfil da segunda fonte. Inclusive, capta (nesse horário) parte da variação que deveria pertencer à primeira fonte. Isto pode ser verificado observando-se os pontos para 1:00 h em cada curva. A curva azul, da fonte automotora, tem início um pouco abaixo do que deveria. Quando se compara essa figura com as Figuras 6.22.1a e 6.22.2a, pode-se perceber isto mais facilmente. Naqueles dois casos, os dados para 1:00 h foram substituídos, respectivamente, pela média das vinte e quatro horas, e pela média das vinte e três horas (de 2:00 h até 24:00 h) nos períodos correspondentes afetados.

O gráfico b mostra os perfis referentes às variações sistemáticas semanais. A cor azul representa a fonte automotiva e a verde a segunda fonte. A Figura 6.22.3d mostra essas mesmas variações. A cor azul refere-se à primeira fonte e a marrom à segunda.

A Figura 6.22.3c mostra as variações anuais para cada fonte. Em azul, a fonte automotora e em verde a outra fonte. Uma inspeção nessa figura mostra para a curva em verde, grandes aumentos na amplitude em três períodos. O primeiro, entre as décima nona e vigésima semanas, o segundo, entre as vigésima quinta e trigésima semanas, e o último, na trigésima segunda semana. Estes são os períodos em que ocorrem valores para 1:00 h característicos da segunda fonte. Ao final da apresentação de resultados para a Estação Ibirapuera serão mostradas algumas figuras contendo os espectros para os dados originais com os quais se poderá confirmar as afirmações acima.

As emissões de SO_2 que ocorrem à 1:00 h nas três Estações de Medição possuem a mesma característica comum: em certos períodos contínuos, os valores para 1:00 h são sempre muito mais altos do que aqueles registrados uma hora antes e uma hora depois (isto é, à meia noite e às 2:00 h). Este é um tipo de variação súbita e localizada que aparentemente não possui significado físico. Entretanto, pelo fato da mesma ocorrer sempre no mesmo horário e por um período de tempo prolongado, o modelo consegue captá-la. Por outro lado, para identificá-la é prudente a introdução de informação adicional.

Semelhantemente ao realizado para a Estação de medição P. D. Pedro II, a análise das variâncias capturadas pelo modelo PCA ajuda a entender as emissões pelas diferentes fontes. A Tabela abaixo mostra a variâncias capturadas em cada caso, sem e com a influência do período afetado.

Tabela VI.2: Variâncias capturadas pelas três primeiras PCs (São Caetano – 1999)

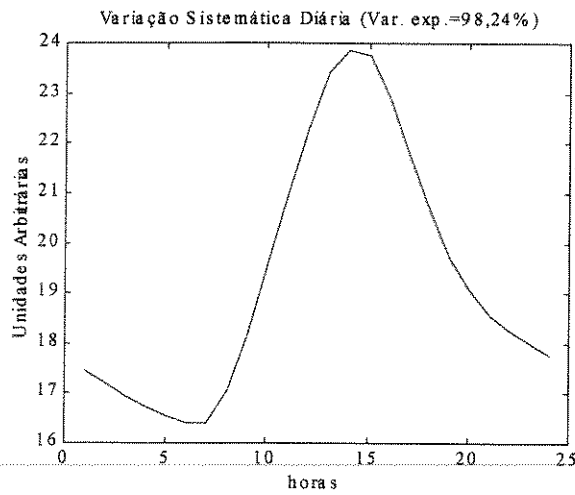
Percentagem de Variância capturada pelo modelo PCA					
Sem a influência do período afetado			Com a influência do período afetado		
PC#	%Var.	%Var. Tot.	PC#	%Var.	%Var. Tot.
1	75.0883	75.0883	1	70.6665	70.6665
2	7.1187	82.2070	2	9.6315	80.2980
3	3.6495	85.8565	3	5.3491	85.6471

O resultado é muito parecido com o que acontece no P. D. Pedro II. Observa-se que a variância capturada por PC₁, nos dois casos (sem e com a influência do período afetado), é muito maior do que a capturada pelas outras duas componentes. Isto indica as emissões automotivas (ônibus e caminhões) como a principal fonte de poluição por SO₂ no local.

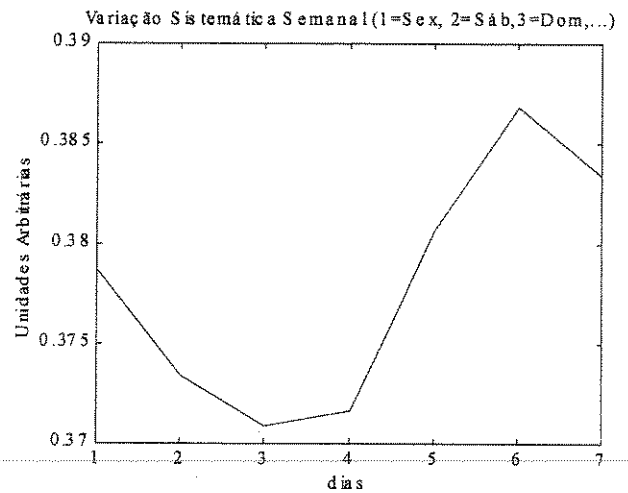
Deve ser observado que também para essa localidade, as percentagens de variâncias acumuladas, em cada caso, são aproximadamente iguais (85,55% e 85,64%).

A análise para as variâncias capturadas por PC₂ e PC₃ é análoga à realizada para a Estação de Medição P. D. Pedro II. Isto indica que PC₁ está capturando variância de uma fonte de emissão (automotiva) e PC₂ (e/ou PC₃) de outra fonte (indústrias).

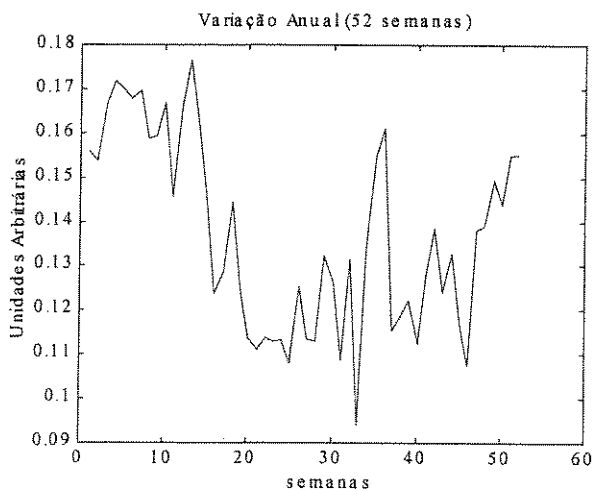
Modelo PARAFAC para a Temperatura



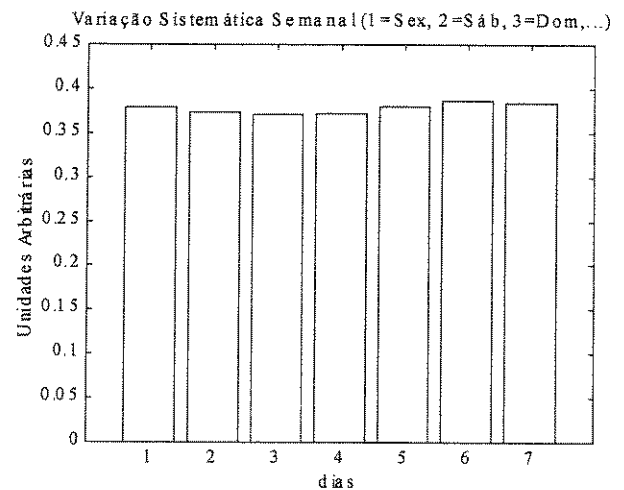
(a)



(b)



(c)



(d)

Fig.6.23: a) Variações sistemáticas com periodicidade diária. b) Variações sistemáticas com periodicidade semanal. c) Variação anual. d) Variações sistemáticas com periodicidade semanal.

Os perfis diário (Figura 6.23a), semanal (Figuras 6.23b e 6.23d), e a variação anual (Figura 6.23d) concordam com os resultados obtidos para as outras Estações de Medição.

Percebe-se, ao contrário variação sistemática semanal, a importância da variação sistemática diária e da variação anual da temperatura no comportamento da atmosfera.

Modelo PARAFAC para a Umidade

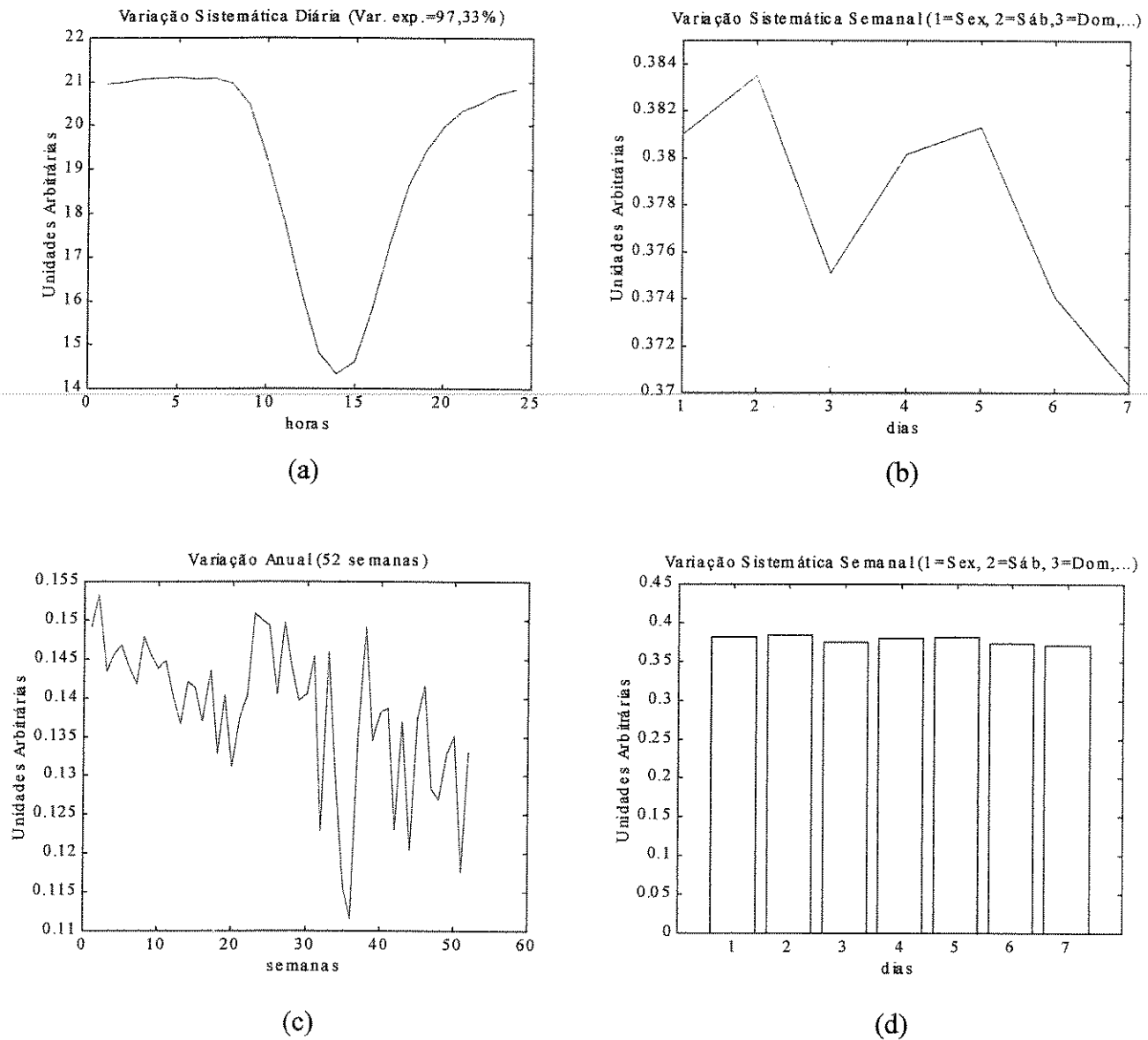


Fig.6.24: a) Variações sistemáticas com periodicidade diária. b) Variações sistemáticas com periodicidade semanal. c) Variação anual. d) Variações sistemáticas com periodicidade semanal.

Os perfis diário (Figura 6.24a), semanal (Figuras 6.24b e 6.24d), e a variação anual (Figura 6.24c) concordam com os resultados obtidos para as outras Estações de Medição.

Assim como para a temperatura, as variações sistemáticas semanais para a umidade não oferecem contribuição significativa ao estudo do comportamento da atmosfera. Por outro lado o perfil diário e a variação anual são importantes.

Est.3: Estação Ibirapuera – 1999

6.3.1.13 – Análise de Monóxido de Carbono – CO

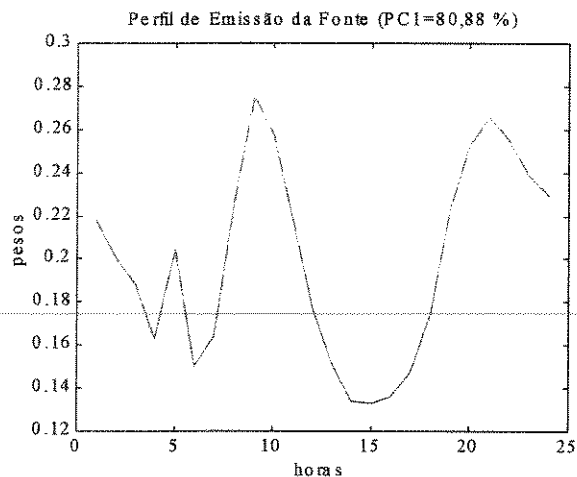
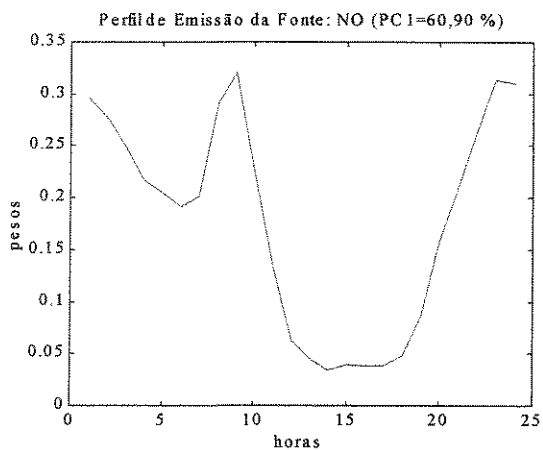


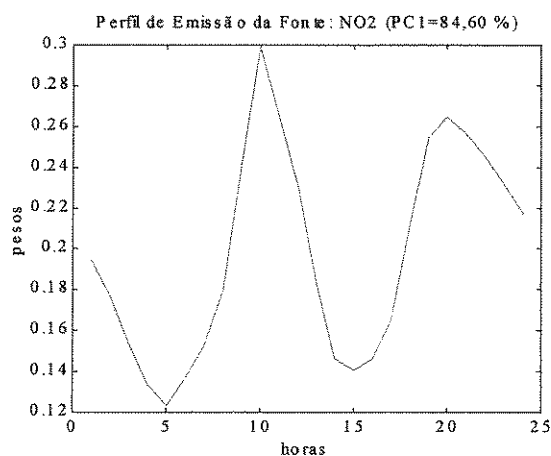
Fig.6.25: Pesos para a 1ª componente principal, PC₁, para o CO.

A Figura 6.25 mostra que no Ibirapuera a variável CO possui o comportamento típico que esse poluente primário apresenta nas outras localidades estudadas.

6.3.1.14 – Análise de Óxidos de Nitrogênio – NO e NO₂



(a)



(b)

Fig.6.26: Pesos para PC₁. a) Para o NO, e b) para o NO₂.

As Figuras 6.26a e 6.26b mostram que à despeito de variações na amplitude (e no horário do 2º pico para o NO), o padrão anterior para as variáveis, se repete no Ibirapuera.

6.3.1.15 – Análise de Ozônio – O₃

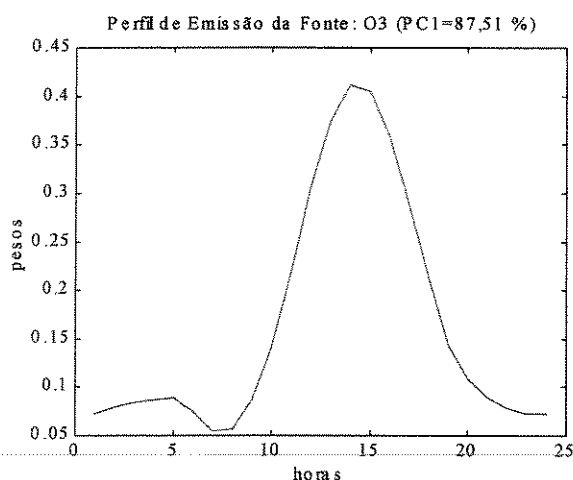


Fig.6.27: Pesos para PC₁, para o O₃.

A Figura 6.27 mostra que o perfil para o O₃ no Ibirapuera também segue o mesmo padrão encontrado nas outras Estações estudadas.

6.3.1.16 – Análise de Dióxido de Enxofre – SO₂

A característica peculiar ao SO₂ nas Estações de medição Parque D. Pedro II e São Caetano no tocante aos dados coletados 1:00 h, também se manifesta, e com maior intensidade, no Ibirapuera. A extensão dessa manifestação é tão grande que passa a ser captada pela primeira componente principal na análise PCA quando os dados levam em conta aquela contribuição. Nesse caso, PC₂ capta uma quantidade de variância bem maior do que aquelas encontradas para os casos das outras duas estações (P. D. Pedro II e São Caetano). Por outro lado, quando a influência dessa segunda fonte é retirada, PC₁ capta o perfil de emissão da fonte automotiva. Então, tal fenômeno para o SO₂ no Ibirapuera recebeu o mesmo tratamento dado às outras localidades com comportamento semelhante (isto é, sem e com a influência da 2^a fonte).

O procedimento por meio do qual os dados para 1:00 h foram substituídos pela média das 24:00 h no período afetado, não eliminou o problema, como esperado. No segundo procedimento, quando os dados foram substituídos pela média das 23:00 h (isto é, de 2:00 h a 24:00 h) no período afetado, o problema foi resolvido satisfatoriamente. Ou seja, foi possível retirar a influência daqueles dados, como mostra a Figura 6.28.1a. As Figuras 6.28.1b a 6.28.1d mostram os resultados, incluído o período afetado.

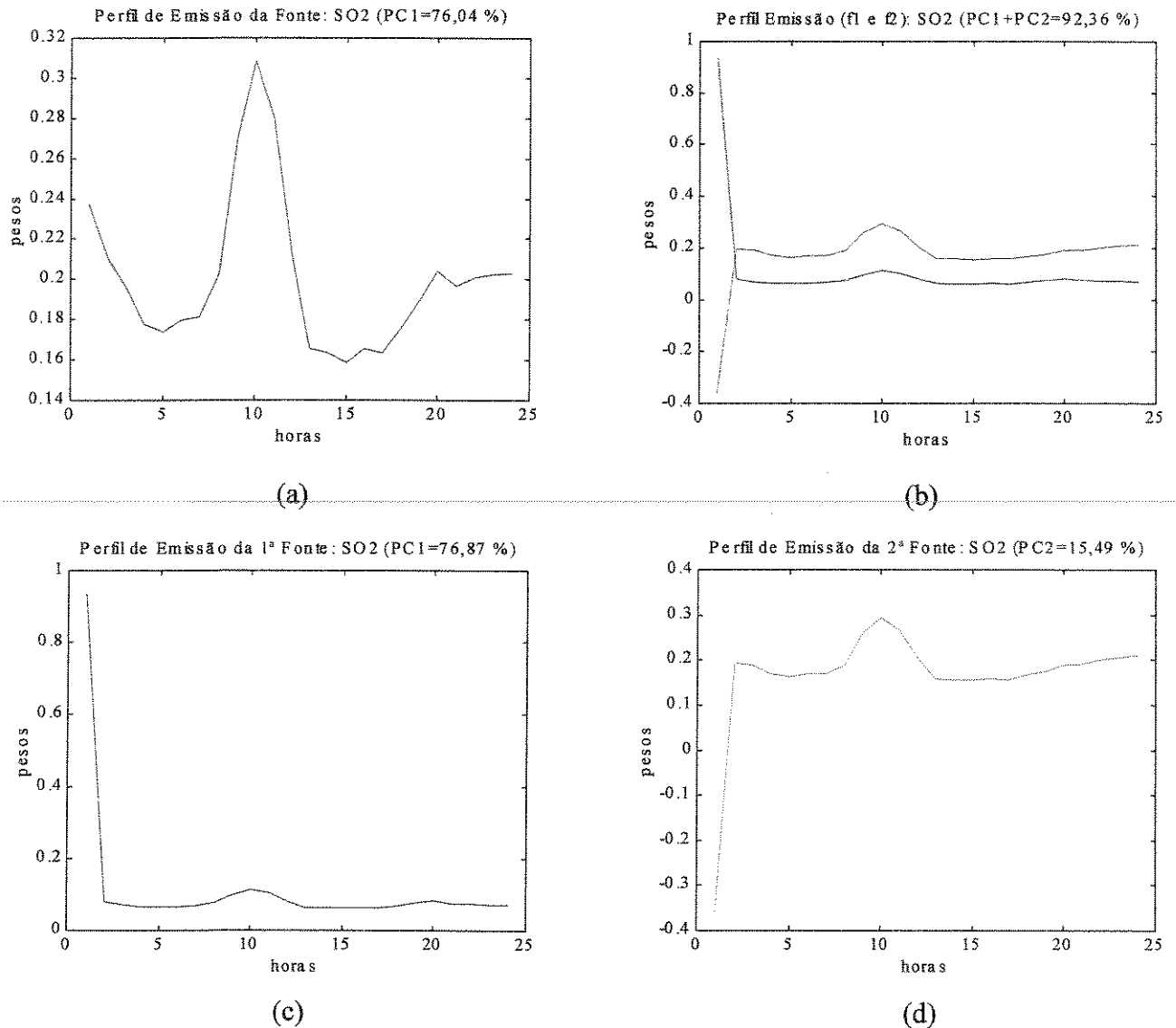
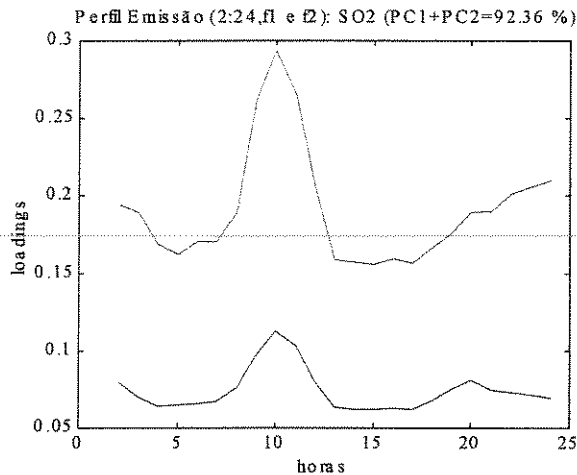


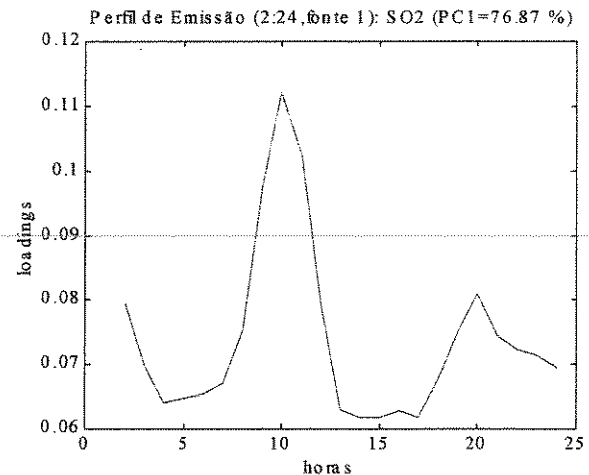
Fig.6.28.1: a) Pesos para a primeira componente principal, PC_1 (uma fonte), e b) Pesos para as duas primeiras componentes principais, PC_1 e PC_2 . c) Pesos para PC_1 e d) Pesos para PC_2 .

A Figura 6.28.1a mostra a semelhança com os perfis para a variação sistemática diária das emissões automotivas (retirada a influência do período afetado), enquanto a Figura 6.28.1b mostra PC_1 e PC_2 (incluído o período afetado). As Figuras 6.28.1c e 6.28.1d mostram, respectivamente, PC_1 e PC_2 . Observe-se que na Figura 6.28.1a não foram simplesmente retirados os valores para 1:00 h, mas sim recalculados segundo a média dos valores para o período afetado como discutido acima. Os gráficos b, c e d não exibem um padrão semelhante ao das emissões automotivas. Entretanto, quando são “plotados” sem os

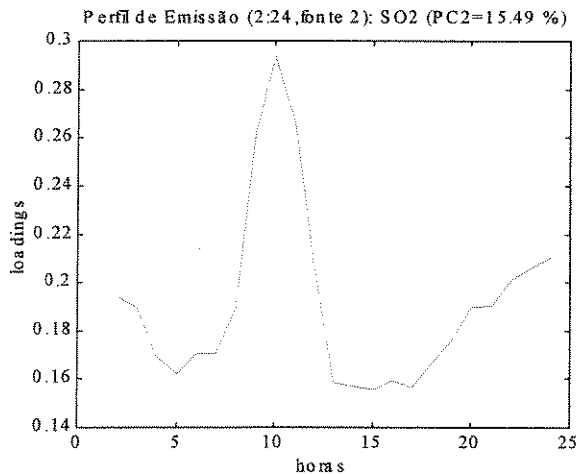
valores para 1:00 h (Figuras 6.28.2a, 6.28.2b e 6.28.2c) pode-se perceber o mesmo perfil de emissão exibido pela fonte de emissão dos veículos. À primeira vista isto poderia induzir à conclusão de que os dois perfis, PC_1 e PC_2 (Figuras 6.28.1b, 6.28.1c e 6.28.1d) estariam associados à fontes automotivas as quais, de algum modo, emitiriam valores muito altos naquele horário.



(a)



(b)



(c)

Fig.6.28.2: Pesos para PC_1 e PC_2 , para a variável SO_2 no Ibirapuera (2:00-24:00 h). a) PC_1 e PC_2 , b) PC_1 , e c) PC_2 .

As Figuras 6.28.2a, b e c, parecem mostrar, por meio de PC_1 e PC_2 , perfis de emissão de fontes automotivas. Mas não é o caso. Deve ser observado que a Figura 6.28.2 é gerada com os mesmos dados que deram origem a Figura 6.28.1. Porém, ao se construir os gráficos, os dados à 1ª hora foram retirados. Isto provoca um novo ordenamento na escala

para os pesos. Na Figura 6.28.1b a escala varia de $-0,4$ a $+1,0$. Ao se retirar a primeira hora a escala passa a variar no intervalo de $+0,05$ a $+3,0$ como mostra a Figura 6.28.2a. Analogamente, a escala que na Figura 6.28.1c varia no intervalo de $0,0$ a $+1,0$ passa a variar entre os limites de $0,06$ a $+0,12$ como mostra a Figura 6.28.2b. Do mesmo modo, para a Figura 6.28.1d cuja variação está situada entre $-0,4$ e $+4,0$, a escala passa a variar no intervalo de $+0,14$ a $+0,3$ (Figura 6.28.2c).

Deve ser observado que a operação acima, isto é, a retirada dos dados para uma hora, não deve afetar os outros horários, tal que os valores lançados correspondam às 23 horas do dia, isto é, entre 2:00 h e 24:00 h.

A interpretação dos resultados obtidos nas análises PCA para a Estação de Medição do Ibirapuera, no ano de 1999, torna-se mais difícil quando comparada à das outras duas localidades analisadas até aqui, principalmente porque o período afetado é de tal ordem que sua manifestação passa a ser capturada pela primeira componente principal, o que não aconteceu nos outros casos.

Cabe lembrar que o fenômeno que ocorreu para o SO_2 durante o ano de 1999 nas três localidades mencionadas, também foi registrado pelas Estações de Medição Cerqueira César, Centro e Osasco, no mesmo ano. Tal fato não foi registrado por nenhuma das sete Estações no período 1997 – 1998.

Vale acrescentar ainda que dentre as localidades analisadas, a que apresentou maiores problemas de interpretação foi a do Ibirapuera. Isto sugere que aquele local adquiriu, em 1999, características que o tornaram diferente dos outros em termos de comportamento em relação à algumas das emissões a que esteve sujeito.

Para tentar dirimir dúvidas que por ventura venham a pairar em relação as variações sistemáticas diárias para o SO_2 no Ibirapuera (e também no P. D. Pedro II, São Caetano, Cerqueira César, Centro e Osasco, eis que apresentaram o mesmo tipo de fenômeno em 1999), são apresentadas figuras contendo as seguintes informações:

- a) Figura 6.28.3a (dados originais) – matriz de dados originais,
- b) Figura 6.28.3b (dados corrigidos) – matriz de dados originais corrigidos apenas para dados anômalos e variações súbitas,
- c) Figura 6.28.3c (transposta dos dados corrigidos) – transposta da matriz de dados originais corrigidos, como em b, (para se observar os valores em torno de 1:00 h),

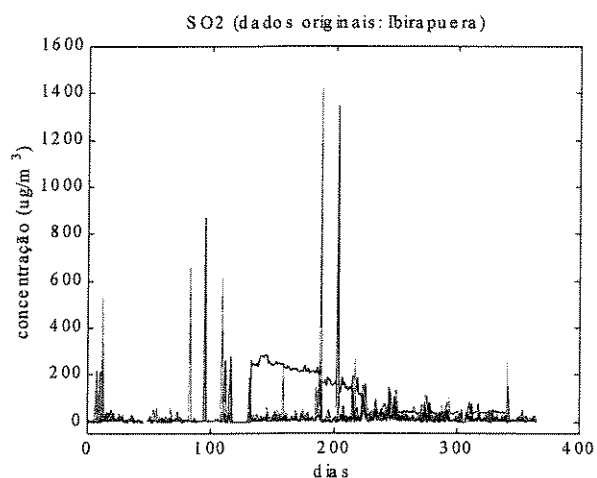
d) Figura 6.28.4a (transposta dos dados corrigidos) – ampliação da extremidade esquerda da Figura 6.28.3c (para melhor acuidade na observação dos escores à 1:00 h e às 2:00 h),

e) Figura 6.28.4b (transposta dos dados corrigidos) – ampliação da extremidade direita da Figura 6.28.3c (melhor observação dos escores às 24:00 h),

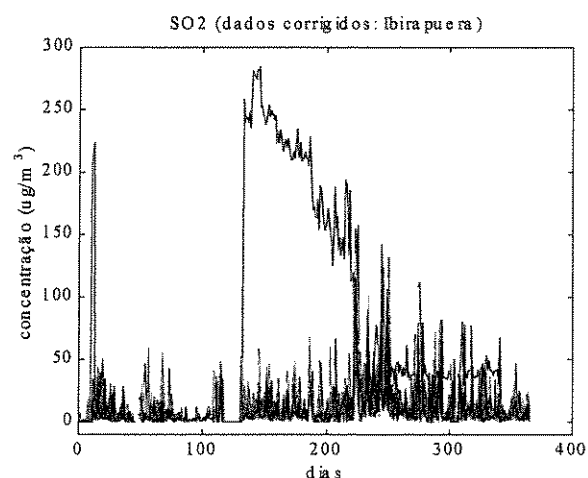
f) Figura 6.28.5a (período afetado) – referente aos dias correspondentes ao período afetado (dias 10 a 12, 133 a 222, 250 a 272, 274 a 283, 287 a 301 e 310 a 341),

g) Figura 6.28.5a (período afetado) – referente aos dias correspondentes ao período não afetado (dias 1 a 9, 13 a 132, 223 a 249, o dia 273, 284 a 286, 302 a 309 e 342 a 365).

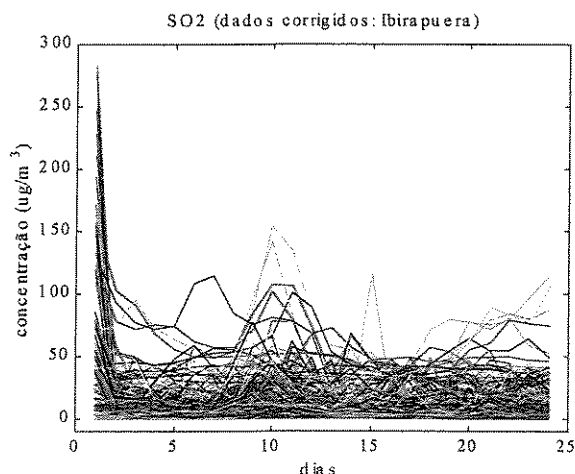
h) Figura 6.28.5c (período não afetado) – escores apenas para dados à 1:00 h.



(a)



(b)



(c)

Fig.6.28.3: a) Dados originais: concentração $\times$ dia. Dados corrigidos: b) concentração $\times$ dia, c) concentração $\times$ dia.

As Figuras 6.28.3a e b mostram, respectivamente os escores para os dados originais e os escores para os dados originais corrigidos. Nessas Figuras pode-se notar que os valores para 1:00 h estão, em determinados períodos (citados acima), “descolados” dos valores para as outras 23 horas do dia. Tomando-se a transposta da matriz de dados originais corrigidos (Figura 6.28.3b) obtém-se a Figura 6.28.3c. Nessa figura constata-se que os escores para os horários de 24:00 h e de 2:00 h são muito baixos em relação àqueles para 1:00 h. Isto é confirmado ampliando-se as extremidades do gráfico da Figura 6.28.3c, como mostram as Figuras 6.28.4a e 6.28.4b.

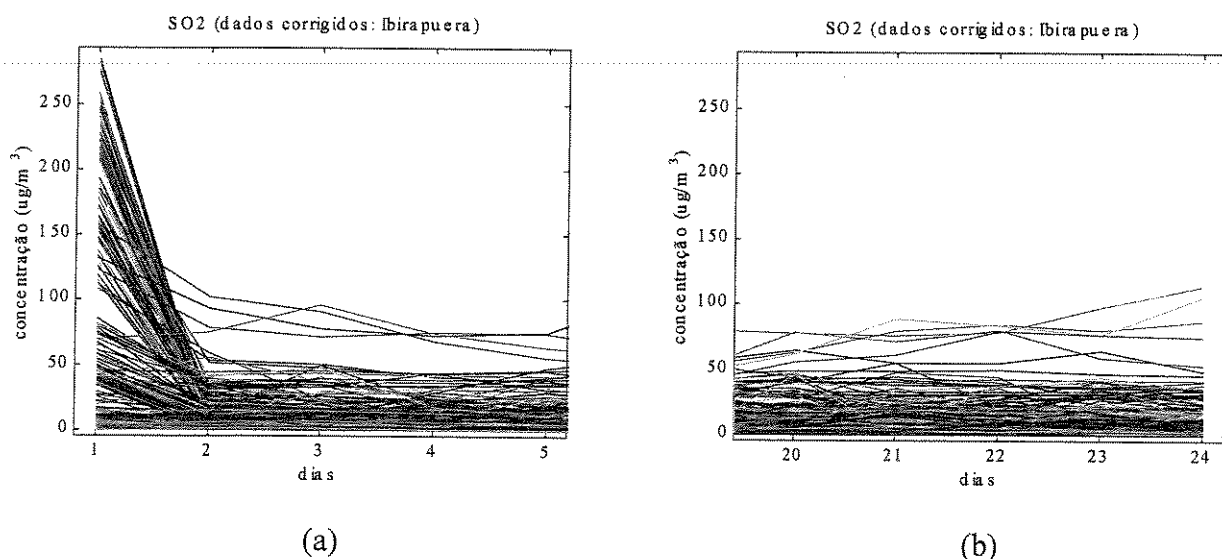
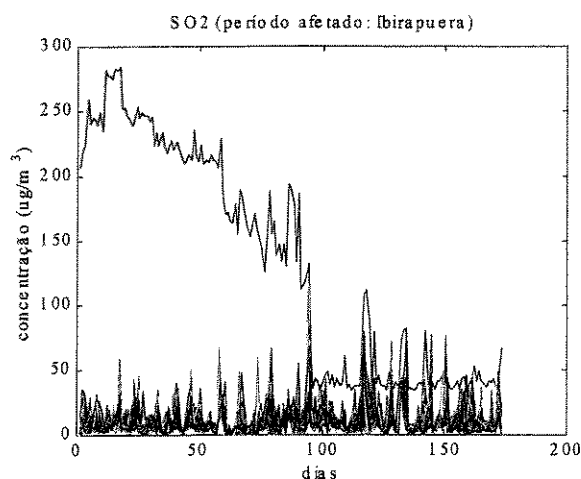
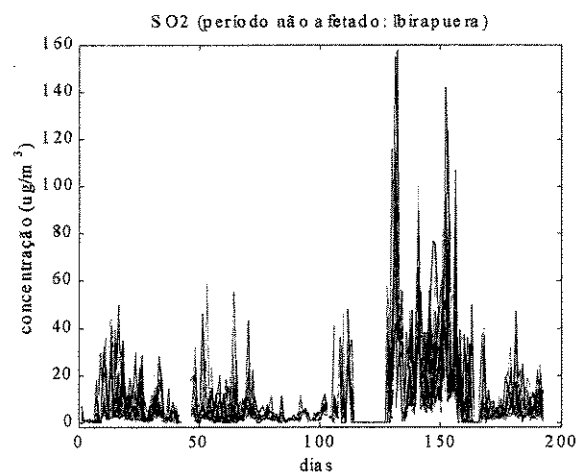


Fig.6.28.4: Ampliações das extremidades, respectivamente, a) esquerda e b) direita da Figura 6.28.3c. Na figura 6.28.4a percebe-se claramente a distribuição dos valores para 1:00 h em três grupos. O primeiro, entre ~ 100 e ~ 300 , o segundo entre ~ 35 e ~ 100 , e o terceiro abaixo de ~ 35 . Na Figura 6.28.4b nota-se que a grande parte dos dados situa-se abaixo de aproximadamente 50.

Não obstante os resultados acima sejam suficientes para se afirmar que aqueles valores medidos no horário de 1:00 h (no período afetado) não pertençam à emissões automotivas, são mostradas as figuras com as matrizes apenas para os dias correspondentes ao período afetado (Figura 6.28.5a), para os dias correspondentes ao período não afetado (Figura 6.28.5b), e para os dados registrados apenas à 1:00 h nos dias correspondentes ao período não afetado (Figura 6.28.5c).

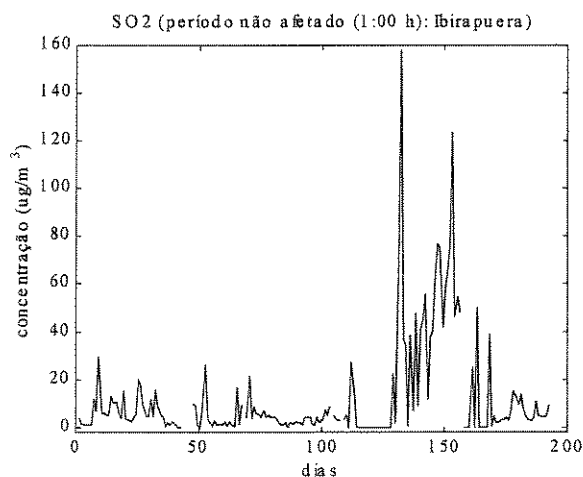


(a)



(b)

Fig.6.28.5: Escores para os dados corrigidos. a) período afetado (1:00-24:00 h). b) período não afetado (entre 1:00 e 24:00 h). c) período não afetado (com escores apenas para os dados tomados à 1:00 h no período não afetado).



(c)

A Figura 6.28.5a mostra os escores para os dados originais corrigidos (período afetado). Essa figura mostra claramente que os pontos coletados no horário de 1:00 h no período segue uma distribuição própria que não tem qualquer relação com o restante dos dados coletados nos outros horários (2:00-24:00 h). Por outro lado, quando se observa a Figura 6.28.5b, nota-se que os dados tomados nas 24 horas do dia possuem uma estrita relação entre eles. Isto é, os dados para 1:00 h não se destacam do espectro como uma curva à parte, como no primeiro caso. A figura 6.28.5c mostra a curva para 1:00 h no período não afetado. É visível a concordância com a Figura 6.28.5b, isto é, a curva para 1:00 h não se destaca do espectro geral no período como ocorre na Figura 6.28.5a.

Em seguida é apresentada a Figura 6.28.6 (dados originais corrigidos sem o período afetado) para fins de comparação com os resultados obtidos com os dados originais corrigidos, considerando-se o período afetado.

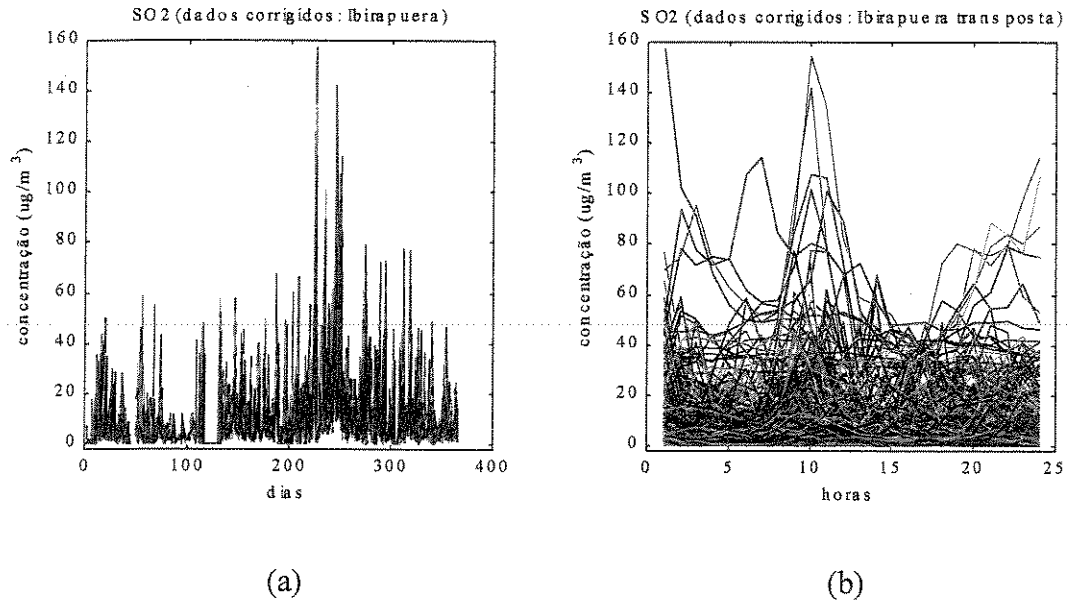


Fig.6.28.6: Dados corrigidos sem o período afetado (dados para 1:00 h substituídos pela média das 23 horas do dia). a) concentração $\times$ dia, b) concentração $\times$ hora.

A Figura 6.28.6a mostra o espectro sem os dados para 1:00 h no período afetado. Nota-se, em 6.28.6b, a ausência do “feixe” de quedas bruscas nos valores de concentração na passagem de 1:00 h para 2:00 h (comparar com Figuras 6.28.3a e 6.28.4a). Quando são comparadas as extremidades esquerdas, respectivamente, das Figuras 6.28.3a (e/ou 6.28.4a) e 6.28.6b, pode-se constatar o efeito provocado pela substituição dos valores à 1:00 h pela média das 23 horas de cada dia no qual o dado para 1:00 h foi substituído.

É útil, nesse ponto, como já realizado anteriormente, comparar as variâncias capturadas em cada caso, sem e com a influência do período afetado (Tabela VI.3 abaixo) para se apreender melhor a distinção entre os comportamentos apresentados em relação ao SO_2 , pelas diferentes localidades estudadas até aqui.

Tabela VI.3: Variâncias capturadas pelas três primeiras PCs nas três localidades

Porcentagem de Variância capturada pelo modelo PCA					
Sem a influência do período afetado			Com a influência do período afetado		
PC#	%Var.	%Var. Tot.	PC#	%Var.	%Var. Tot.
Estação de Medição Parque D. Pedro II (1999)					
1	75.7814	75.7814	1	72.8013	72.8013
2	7.4802	83.2616	2	6.8472	79.6485
3	3.1768	86.4384	3	6.5211	86.1696
Estação de Medição São Caetano do Sul (1999)					
1	75.0883	75.0883	1	70.6665	70.6665
2	7.1187	82.2070	2	9.6315	80.2980
3	3.6495	85.8565	3	5.3491	85.6471
Estação de Medição Ibirapuera (1999)					
1	76.0475	76.0475	1	76.8742	76.8742
2	8.1016	84.1491	2	15.4956	92.3698
3	3.3497	87.4987	3	2.6064	94.9762

Semelhantemente ao ocorrido nas outras duas Estações, a variância capturada por PC₁ nos dois casos (sem e com a influência do período afetado) é muito maior do que a capturada pelas outras duas componentes. Entretanto, surgem algumas diferenças quando se comparam as duas primeiras, com a presente localidade. A primeira delas refere-se à variância capturada pela primeira PC (sem e com a influência do período afetado) para cada uma das três localidades. A segunda refere-se à soma das variâncias para PC₂ e PC₃ (sem e com a influência do período afetado) em cada localidade. A terceira refere-se à variância acumulada pelas três PCs em cada caso para cada uma das três localidades.

Os dados da Tabela VI.3 mostram (sem a influência do período afetado) que as percentagens de variância capturadas por PC₁, PC₂ e PC₃ são aproximadamente iguais nos três locais. Isto implica que as percentagens de variância acumulada também possuem valores muito próximos.

Por outro lado, para Pedro II, as variâncias acumuladas sem e com o período afetado são, respectivamente, 86,43% e 86,16%, portanto, aproximadamente iguais. Analogamente,

para São Caetano, variâncias acumuladas sem e com o período afetado são, respectivamente, iguais a 85,85% e 85,64%, sendo, portanto, também, aproximadamente iguais (e apenas ligeiramente menores do que no P. D. Pedro II).

Para o Ibirapuera entretanto, a distribuição da variância entre as três PCs não se dá do mesmo modo que ocorre para os dois primeiros locais como mostra a tabela VI.3 acima. A primeira diferença está nas variâncias capturadas por PC_1 sem e com o período afetado, respectivamente, 76,04% e 76,87%, portanto iguais nos dois casos. Aqui, porém, a PC_2 (15,49% no período afetado) captura muito mais variância do que PC_2 (8,10%) no período não afetado, considerando-se que não há queda em PC_1 quando se comparam os dois períodos. Este resultado difere do ocorrido nas outras localidades nas quais há queda em PC_1 quando se comparam os dois períodos.

Um outro ponto a considerar é que a soma das variâncias de PC_2 e PC_3 sem e com o período afetado são, respectivamente, 11,45% e 18,10%, mostrando que no período afetado essa soma é bem mais expressiva no Ibirapuera.

Finalmente, no Ibirapuera, as variâncias acumuladas, sem e com o período afetado, são, respectivamente, iguais a 87,49% e 94,97%. Diferentemente do que ocorre nas duas primeiras Estações, a variância acumulada no Ibirapuera quando se considera o período afetado é significativamente maior.

Para Pedro II e São Caetano, as respectivas diferenças entre as percentagens de variância acumulada para os dois casos (sem e com o período afetado), ocorrem às custas de uma diminuição parcial nos valores de PC_1 . Para o Ibirapuera, no entanto, o aumento na percentagem de variância total ao se passar da primeira situação (sem o período afetado) para a segunda (período afetado incluso), ocorre sem que haja queda no valor para PC_1 .

A Figura 6.4b (p245) para Pedro II mostra que as emissões à 1:00 h no período afetado não interferem no perfil automotivo. Para São Caetano, os valores para 1:00 h (Figura 6.16a, p266) são um pouco mais elevados, porém não o suficiente para interferir no perfil automotivo. Observe-se que nos dois casos, as figuras são construídas incluindo-se o período afetado, porém o perfil emerge em apenas uma componente principal.

Por sua vez, a Figura 6.28.1b (p286) para o Ibirapuera (também incluído o período afetado) mostra que as emissões à 1:00 h são tão elevadas e por período tão prolongado, que interferem no perfil automotivo, não podendo ser capturadas em apenas uma PC.

A inclusão de mais duas figuras ajuda sobre maneira a interpretação que se pode dar aos perfis obtidos na análise PCA para o Ibirapuera.

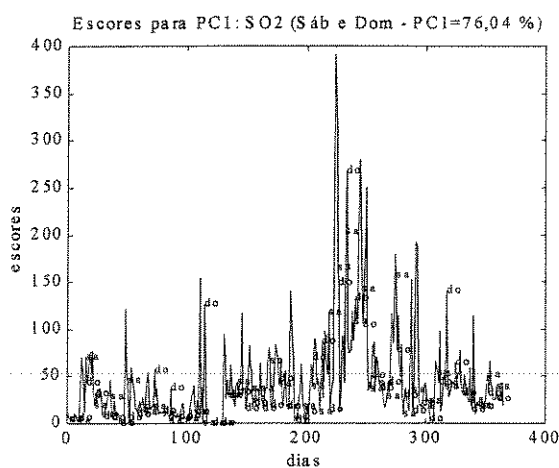
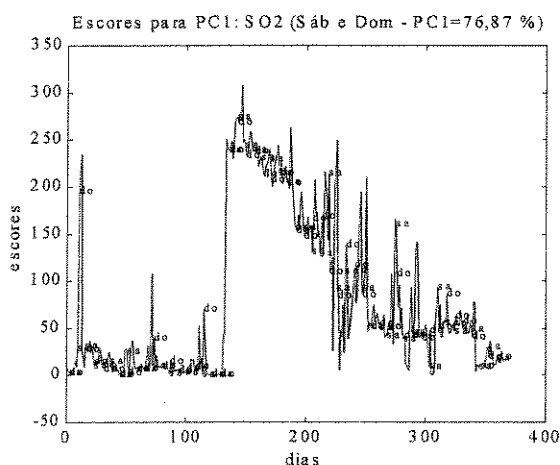
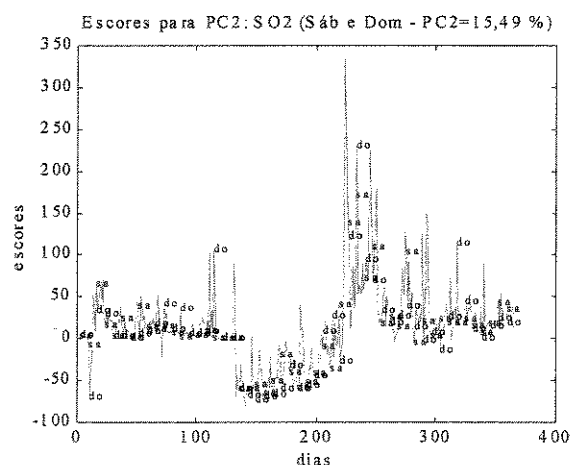


Fig.6.28.7: Escores para PC₁ (sem o período afetado).



(a)



(b)

Fig.6.28.8: a) Escores para PC₁, b) Escores para PC₂. Em ambos os casos, incluído o período afetado. Os “rótulos” para os finais de semana (sábados – sa, domingos – do).

A Figura 6.28.8a mostra os escores para PC₁ na análise PCA. Essa figura é semelhante à 6.28.3b, a qual mostra o gráfico correspondente à matriz de dados originais corrigidos (incluído o período afetado). Por outro lado, uma comparação entre as Figuras 6.28.7 (fonte automotiva) e 6.28.8b revela uma semelhança tal que não se pode furtar a deduzir que esta última também representa emissões de fontes automotivas.

Quando se realiza a análise com o modelo PARAFAC (mais adiante), não se encontra esse tipo de dificuldade para identificar as fontes. O modelo consegue extrair de uma única vez as variações sistemáticas diárias e semanais, e a variação anual.

7.3.1.16 – Análise da Temperatura e da Umidade – Temp e Umid

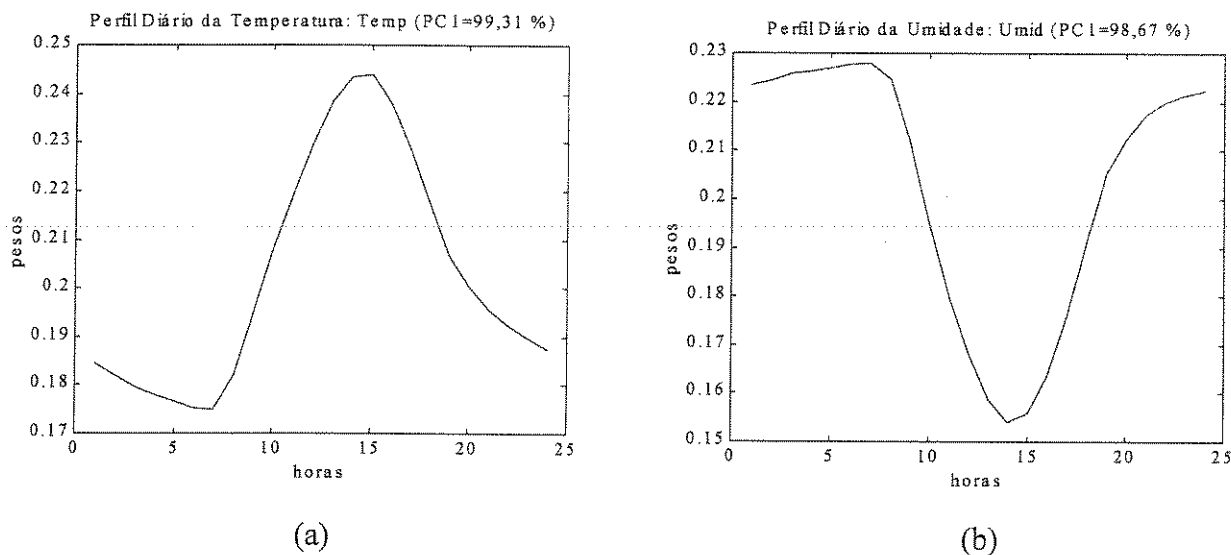


Fig.6.29: Pesos para PC₁. (a) Temperatura (Temp) e (b) Umidade (Umid).

As Figuras 6.29a e b mostram as variações sistemáticas diárias, respectivamente, para a temperatura e para a umidade. Também para o Ibirapuera essas variáveis apresentam o mesmo padrão de variação já apontado nas outras Estações de Medição, ou seja, a temperatura apresenta valor de pico em torno das 14:00 h enquanto a umidade apresenta seu mínimo em torno desse mesmo horário. A característica de simetria também se repete para as referidas curvas nessa localidade.

Para concluir esse conjunto de resultados obtidos nas três Estações de Medição (Pedro II, São Caetano e Ibirapuera) durante o ano de 1999, são apresentados a seguir (Figuras 6.30 a 6.36) os resultados das análises PARAFAC para essa última localidade.

6.3.1.18 – Análises PARAFAC (Ibirapuera): Modelo PARAFAC para o CO

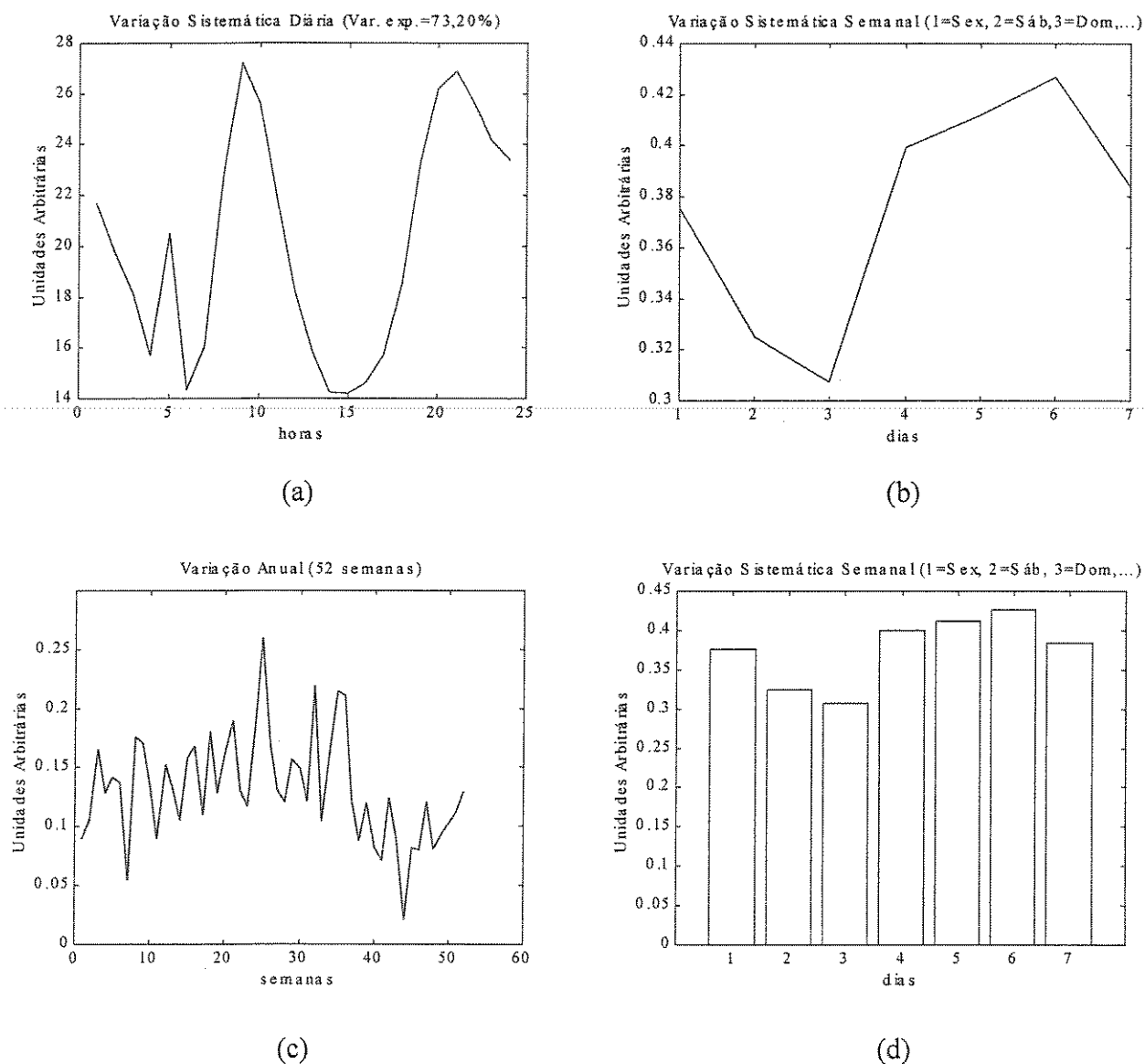


Fig.6.30: a) Variações sistemáticas com periodicidade diária. b) e d) Variações sistemáticas com periodicidade semanal. c) Variação anual. Em a verificam-se os picos da manhã e da noite. Em b e d observa-se a queda das emissões nos finais de semana. Em c a variação anual mostra valores mais elevados durante o inverno.

A Figura 6.30 mostra que os padrões de variação para o CO se repetem nessa Estação de Medição. Do mesmo modo que nos outros locais, os perfis diário (Figura 6.30a) e semanal (Figuras 6.30b e 6.30d) de emissão da fonte automotiva se reproduzem aqui. A Figura 6.30c mostra maior acumulação no inverno.

Modelo PARAFAC para o NO

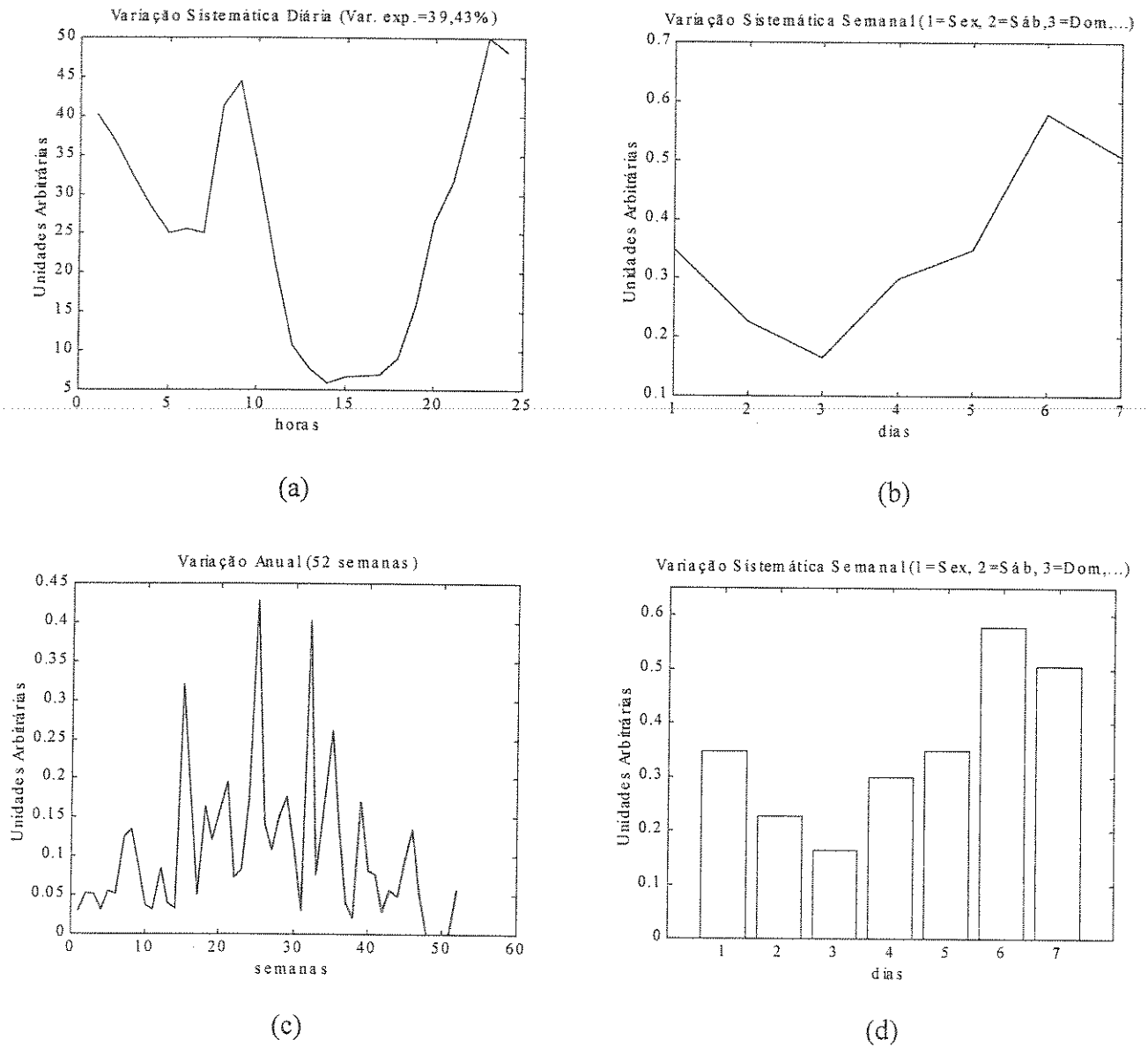


Fig.6.31: a) Variações sistemáticas com periodicidade diária. b) e d) Variações sistemáticas com periodicidade semanal. c) Variação anual. O perfil semanal apresenta a mesma queda nos finais de semana (b e d). A variação anual também mostra valores mais elevados no período de inverno.

A Figura 6.31 mostra que os mesmos padrões de variação para o NO se verificam nessa localidade. Nessa estação também ocorre, com algumas diferenças nos horários de pico para o perfil diário (Figura 6.31a), o mesmo padrão de variação encontrado nos outros locais, como queda nos fins de semana (6.32b e d), e maior acumulação no inverno (6.31c).

Modelo PARAFAC para o NO₂

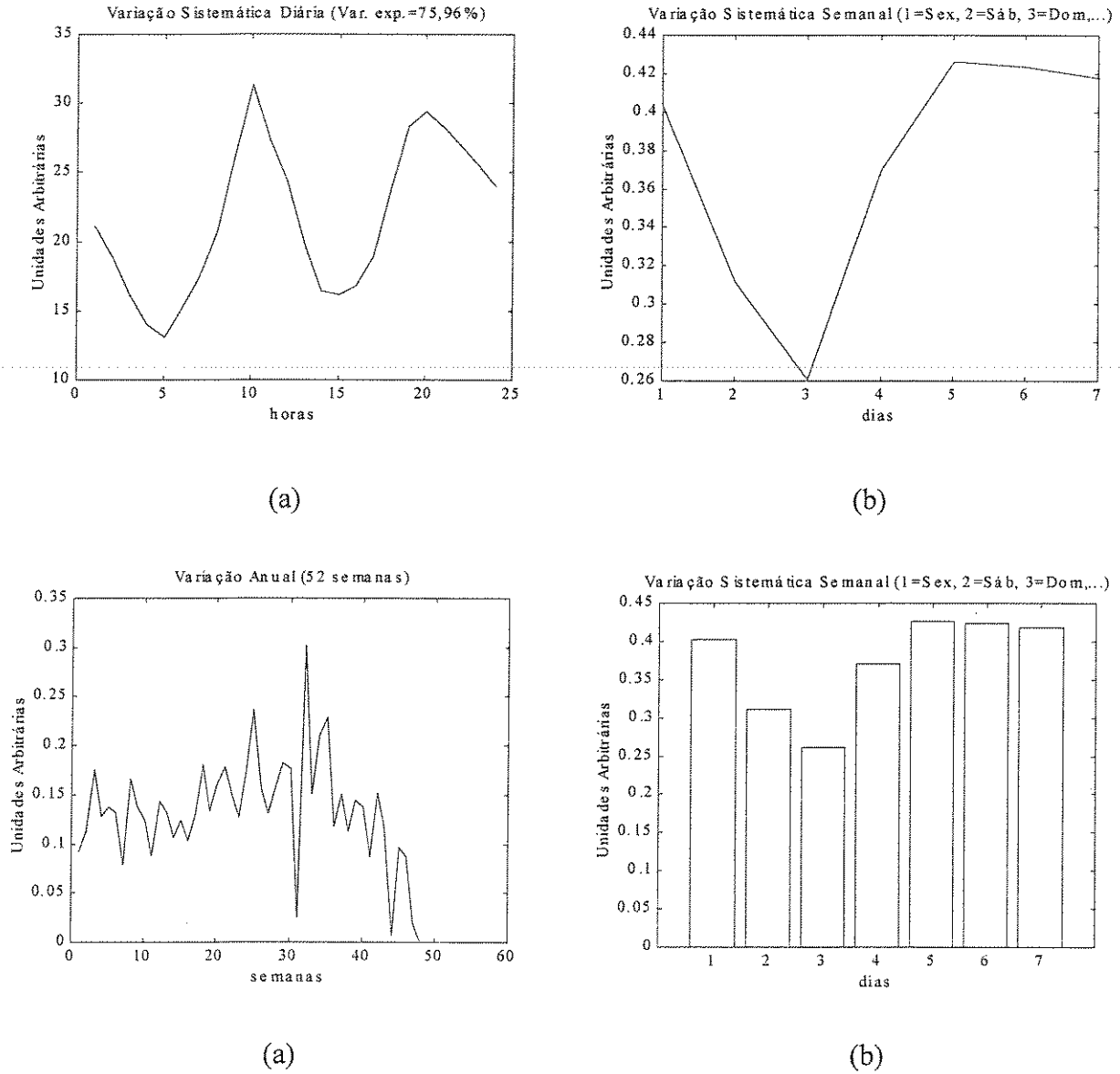


Fig.6.32: a) Variações sistemáticas com periodicidade diária. b) e d) Variações sistemáticas com periodicidade semanal. c) Variação anual. Em a verifica-se o padrão característico para o perfil de emissão da fonte. Em b e d também ocorre a queda de emissões nos finais de semana. Em c nota-se elevação da curva no período de inverno.

A Figura 6.32 mostra que para o NO₂, no Ibirapuera, verificam-se as mesmas características de emissão encontradas nas outras localidades.

Modelo PARAFAC para O₃

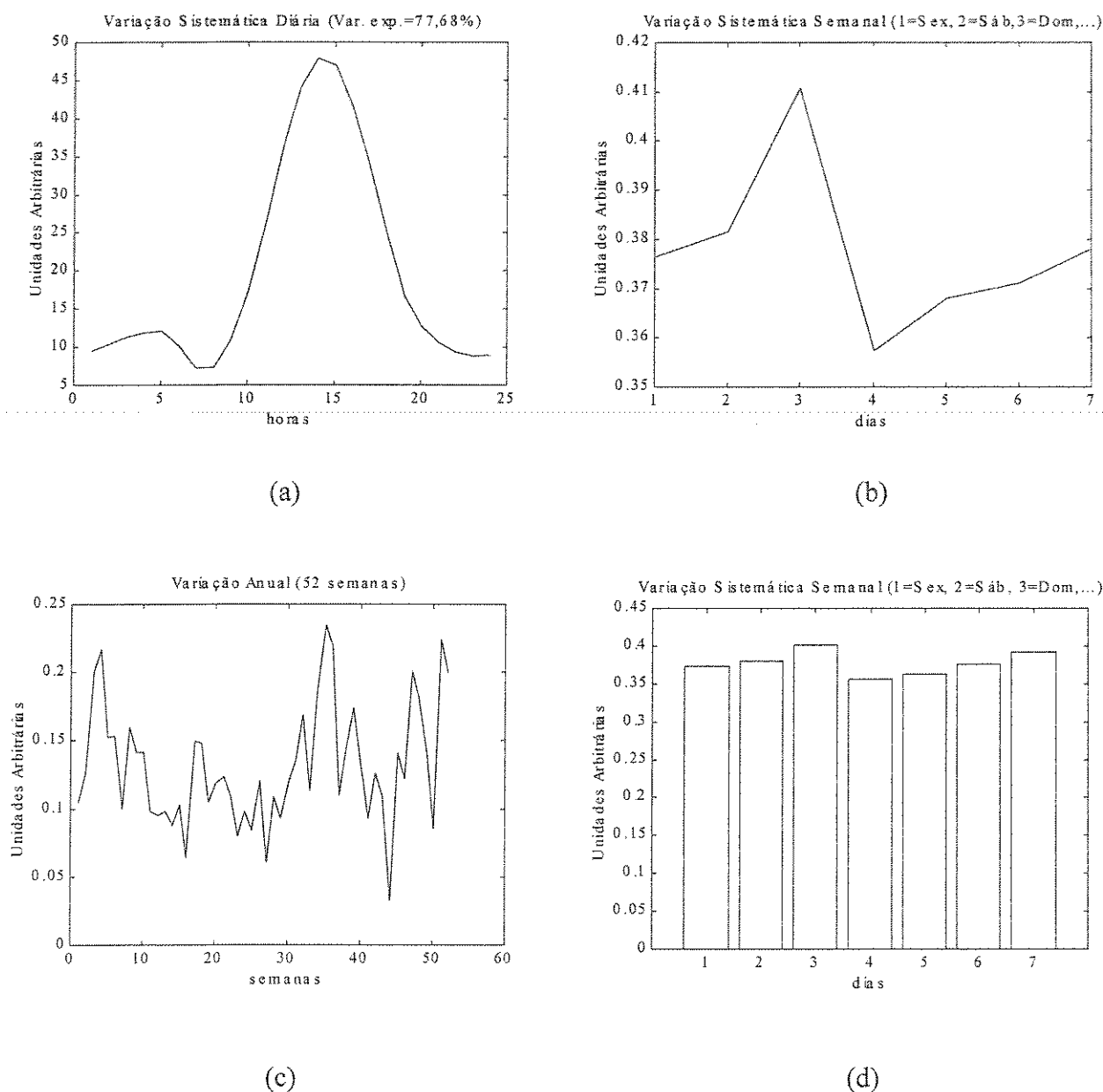


Fig.6.33: a) Variações sistemáticas com periodicidade diária. b) e d) Variações sistemáticas com periodicidade semanal. c) Variação anual para o O₃. Em b e d, pode-se observar o aumento nos finais de semana. Em c verifica-se a queda durante o inverno.

Contrariamente ao que ocorre com os poluentes primários, para o perfil diário do ozônio ocorrem os picos (~3:00 h e 14:00 h, também característico das outras localidades) como mostra a Figura 6.33a. Analogamente, repetem-se os padrões semanais (Figuras 6.33b e 6.33d) e a variação anual (Figura 6.33c).

Modelo PARAFAC para o SO₂ (sem a influência da 2ª fonte)

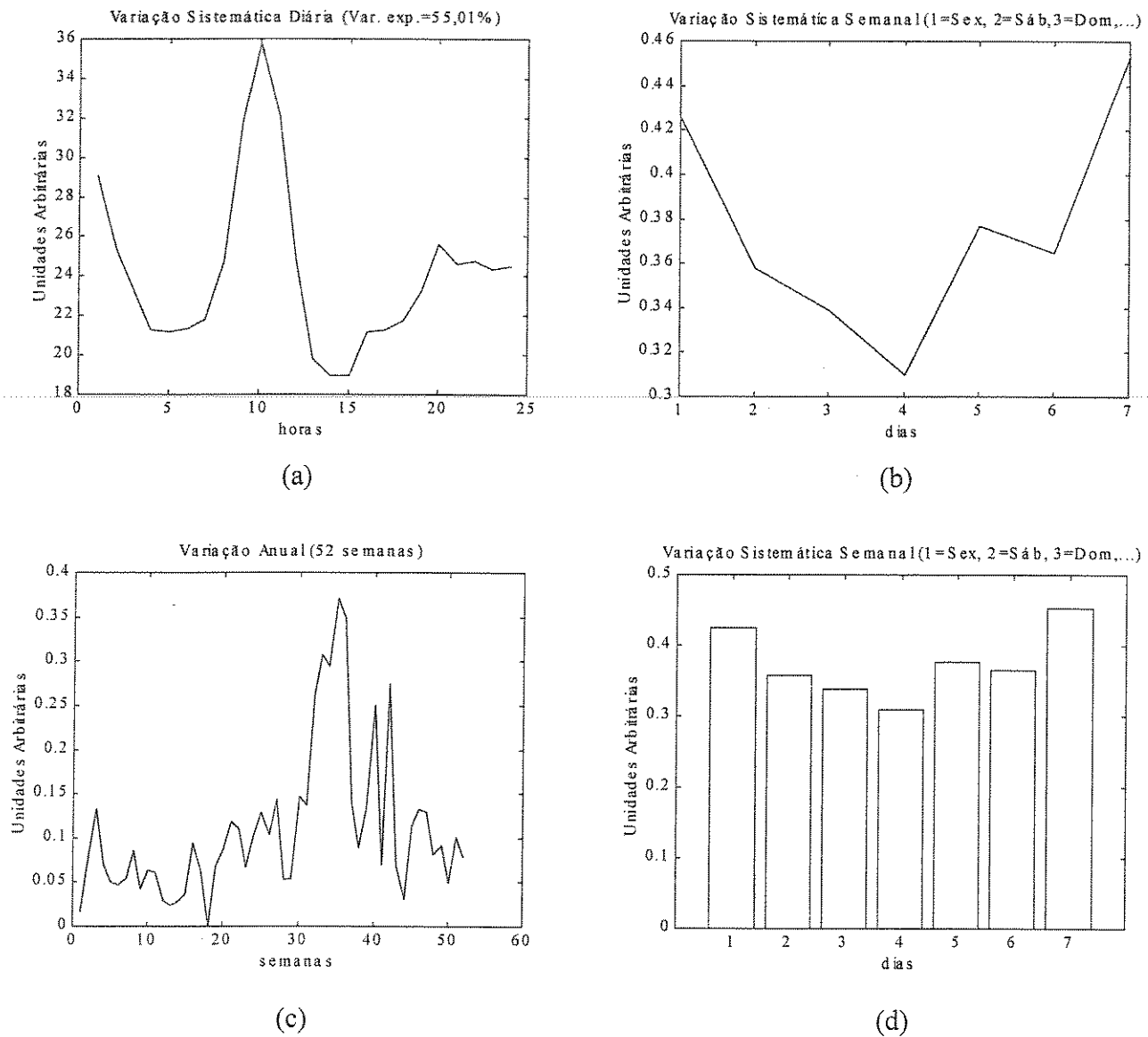


Fig.6.34.1: Modelo PARAFAC para o SO₂ no Ibirapuera em 1999 (sem o período afetado, isto é, dados à 1:00 h substituídos pela média das 23 horas – 2:00 h às 23:00 h – de cada dia no período afetado). a) Variações sistemáticas com periodicidade diária. b) e d) Variações sistemáticas com periodicidade semanal. c) Variação anual.

O perfil diário de emissão da fonte mostrado na Figura 6.34.1a confirma a existência dos dois picos para os poluentes primários. As Figuras 6.34.1b e d mostram a queda nos finais de semana, porém com valor mais baixo na segunda-feira.

Modelo PARAFAC para o SO₂ (duas fontes)

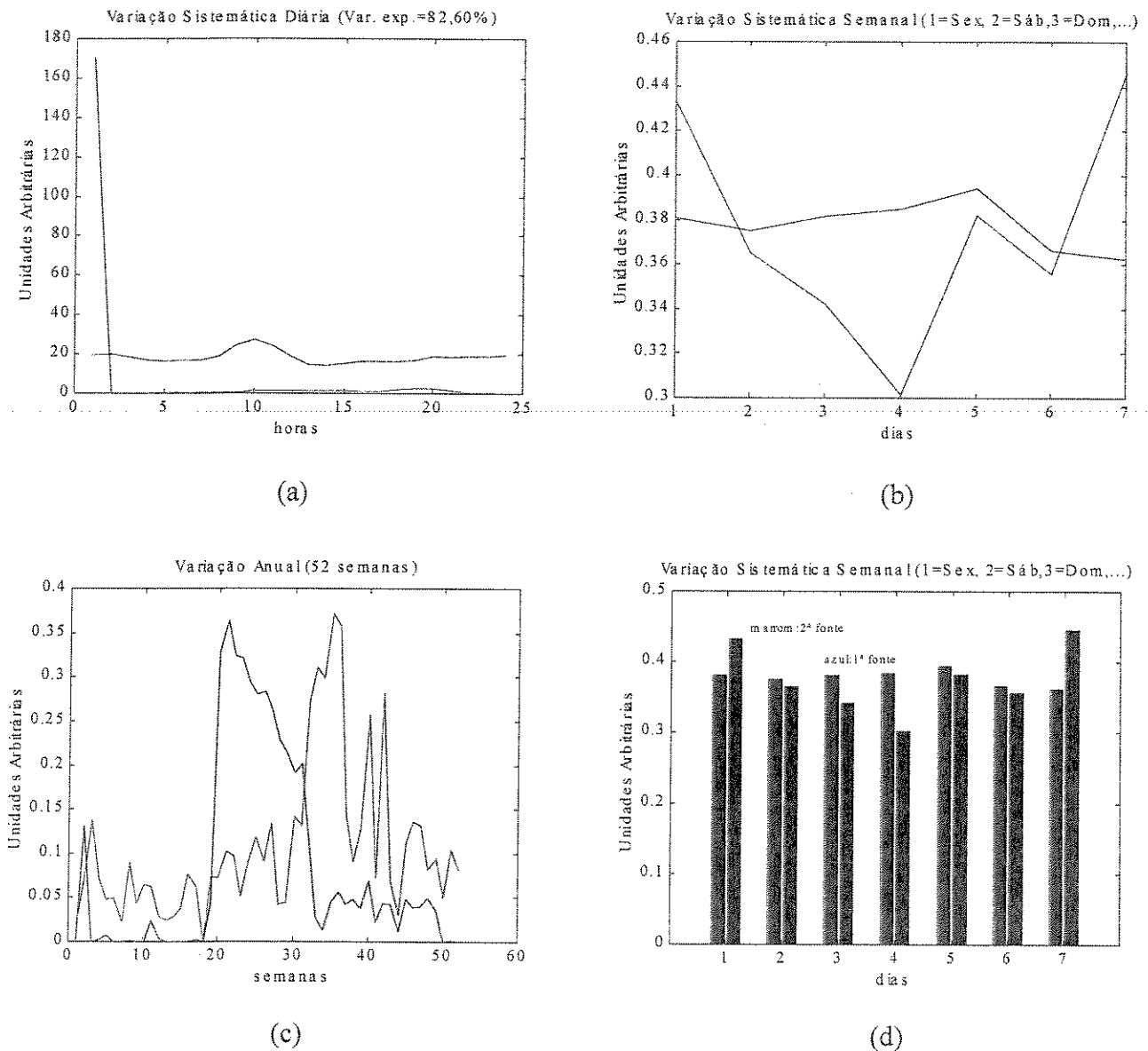


Fig.6.34.2: Modelo PARAFAC para o SO₂ na Estação de Medição Ibirapuera em 1999 (incluídos os dados à 1:00 h no período afetado). a) Variações sistemáticas com periodicidade diária. b) e d) Variações sistemáticas com periodicidade semanal. c) Variação anual.

A Figura 6.34.2a mostra a influência dos valores tomados à 1:00 h no perfil diário. Quando se retira essa influência, ou quando ela é corrigida segundo os valores na faixa de 2:00 h – 24:00 h como mostra a Figura 6.28.1a (PCA, p286), e Figura 6.34.1 (PARAFAC), o perfil que emerge é o das emissões veiculares. Isto pode ser confirmado quando se observa, na Figura 6.34.2, o perfil semanal, no qual a cor verde, em b (e marrom em d)

representa o perfil automotivo, e a cor azul representa a outra fonte de emissão. Uma inspeção na Figura 6.34.2c permite confirmar o aumento de amplitude durante o inverno (curva verde). Essa curva é exatamente igual a curva da Figura 6.34.1c para a variação anual, quando se consideram apenas os dados coletados durante o dia no intervalo de 2:00 h – 24:00 h. É possível constatar também para a segunda fonte, uma variação anual cuja curva guarda grande semelhança com a obtida para os dados coletados à 1:00 h como mostram as curvas em azul ao alto (Figura 6.28.3b, p289, para os dados originais corrigidos; e Figura 6.28.5a, p291, para os dados originais corrigidos considerando-se somente os dias correspondentes ao período afetado).

É instrutivo lembrar, nesse ponto, o poder de modelagem do modelo PARAFAC por meio de sua comparação com PCA. Nesse caso, isto pode ser feito construindo-se, em separado, os gráficos correspondentes às duas componentes que aparecem na Figura 6.34.2a, acima.

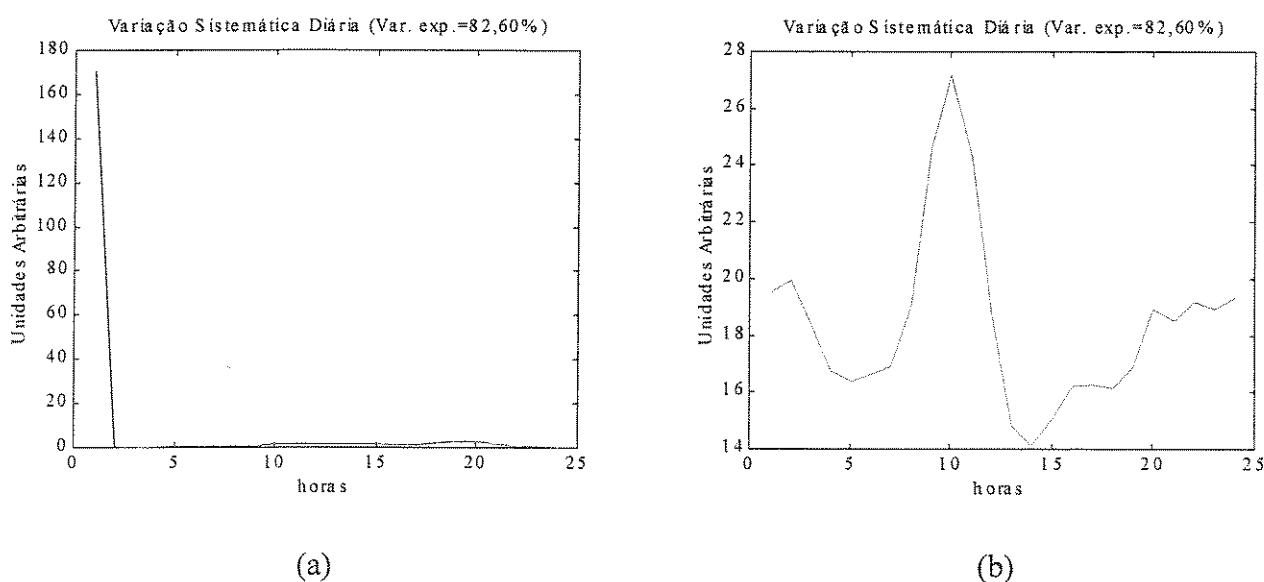


Fig.6.34.3: Modelo PARAFAC para o SO₂ na Estação de Medição Ibirapuera em 1999 (incluídos os dados à 1:00 h no período afetado). a) PC₁ para as variações sistemáticas com periodicidade diária (2ª fonte: indústrias). b) PC₂ para as variações sistemáticas com periodicidade diária (fonte automotiva).

A Figura 6.34.3a põe em destaque a primeira componente (em azul) que aparece na Figura 6.34.2a, e a Figura 6.34.3b o faz para a segunda componente naquela mesma figura.

Uma comparação entre as análises PCA (Figuras 6.28.1c para PC₁, e 6.28.1d para PC₂, à p286) e PARAFAC (Figuras 6.34.3a para PC₁, e 6.34.3b para PC₂, ambas acima) mostra que ambos os modelos capturam igualmente as variações sistemáticas descritas por PC₁. Entretanto, enquanto PCA encontra dificuldades para capturar as variações sistemáticas descritas por PC₂ (Figura 6.28.1d, p286), o modelo de ordem superior consegue separar melhor as variações sistemáticas descritas por meio de cada uma das componentes. Isto pode ser verificado comparando-se as extremidades esquerdas de cada uma das figuras acima mencionadas. Ao contrário de PCA, na análise com o modelo PARAFAC o valor para 1:00 h, em PC₂, está apenas ligeiramente baixo, sendo, por isto, mais consistente com o perfil para as fontes automotivas.

Os resultados e discussões sobre a aplicação de PCA e do modelo PARAFAC à variável SO₂ permitem afirmar a existência de mais de uma fonte de emissão desse poluente na atmosfera da RMSP nas três localidades mencionadas (Parque D. Pedro II, São Caetano do Sul e Ibirapuera) durante o ano de 1999. Uma delas é, sem dúvida, a fonte de emissões automotivas, a outra, proveniente de indústrias.

Para a segunda fonte, no caso do Ibirapuera (modeladas por PC₁), dadas as características locais, os resultados indicam que as emissões sejam provenientes de indústrias, e que atingem aquele local por meio de transporte atmosférico.

Para os três locais, o fato de que os valores em torno de 1:00 h estejam muito abaixo daqueles medidos à 1:00 h poderia ser explicado agregando-se os perfis diários para a temperatura e umidade a estudos de formação de aerossóis e nuvens na troposfera, para o poluente. Permanece, entretanto, a dúvida sobre porque daquelas emissões “pontuais” (isto é, sempre e apenas naquele horário). Talvez isto possa ser atribuído a limpeza de filtros (compostos de enxofre acumulados durante operação diária), ou a um grupo específico de indústrias cujas características de operação impliquem naquele tipo de emissão, coadjuvadas por condições climáticas específicas ocorridas durante o ano de 1999.

Esse assunto é um pouco mais explorado no final desse capítulo, quando são incorporados os resultados obtidos com as estruturas multimodo em quatro ou mais modos.

Modelo PARAFAC para a Temperatura

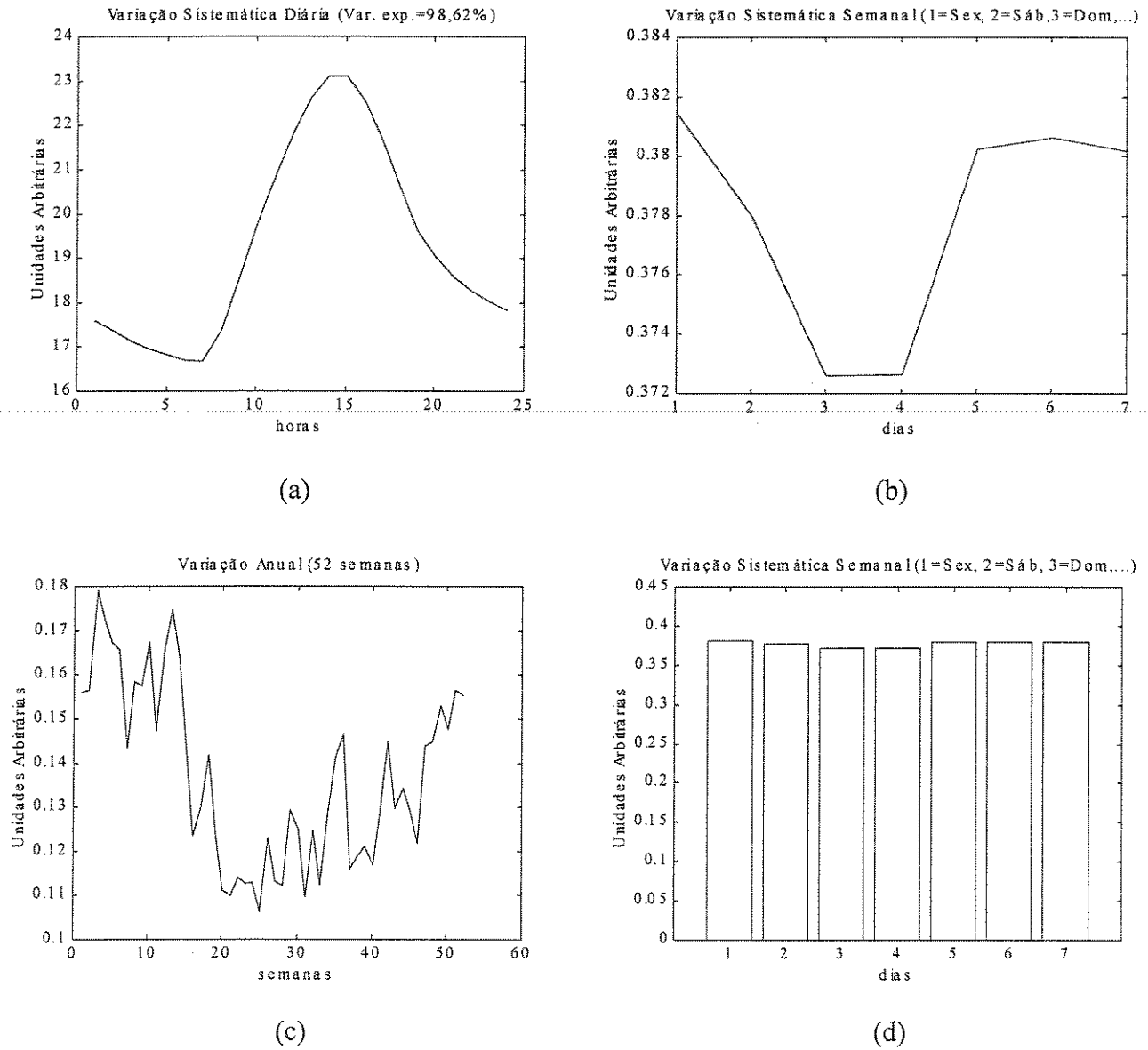


Fig.6.35: a) Variações sistemáticas com periodicidade diária. b) e d) Variações sistemáticas com periodicidade semanal para a temperatura. c) Variação anual.

A Figura 6.35a mostra, no perfil diário, concordância com as localidades anteriores, e a Figura 6.35c mostra que a variação anual também concorda com os resultados obtidos para os outros locais. As Figuras 6.35b e 6.35d (aproximadamente constantes) mostram que a temperatura não varia de forma significativa durante a semana.

Modelo PARAFAC para a Umidade

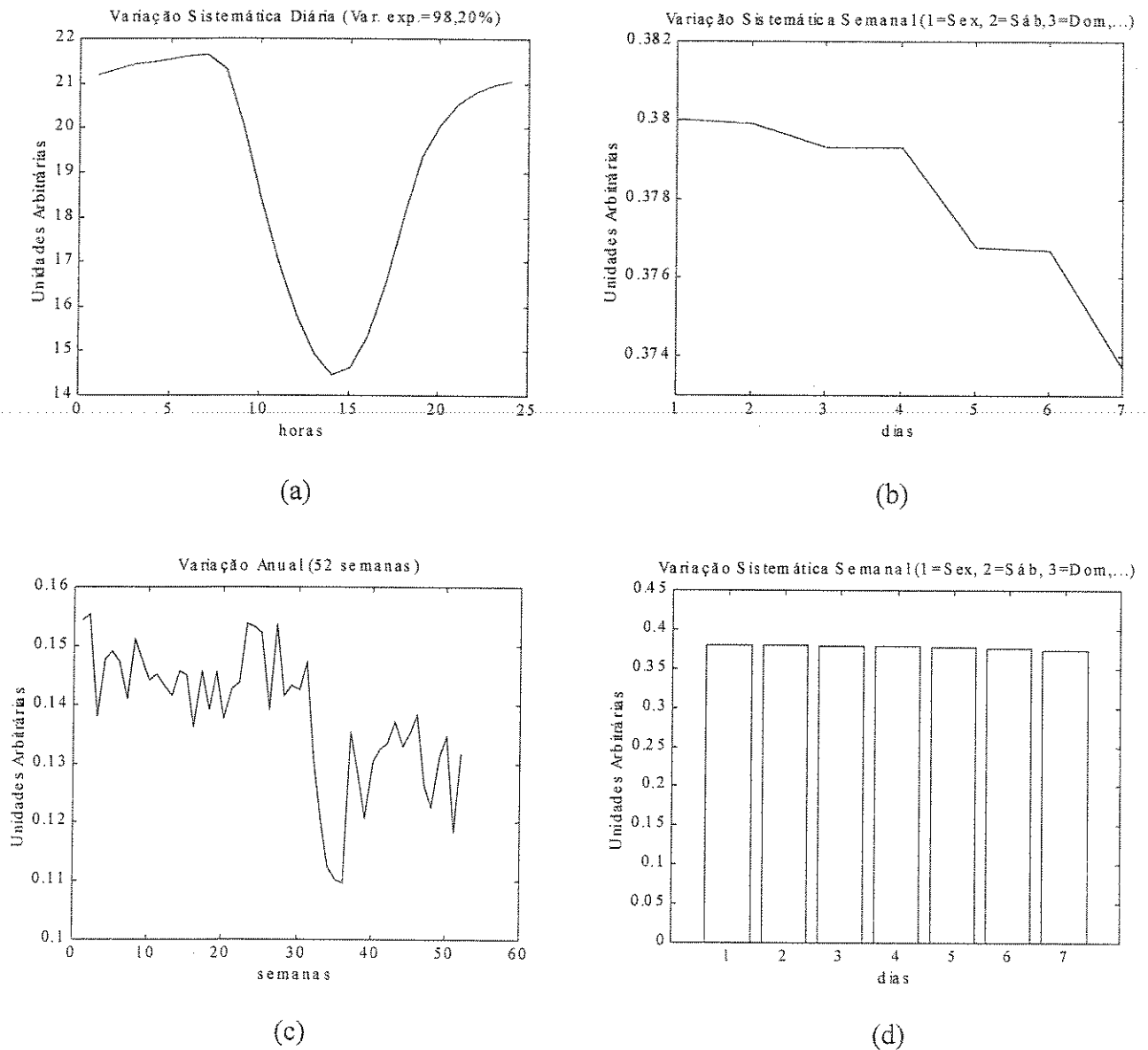


Fig.6.36: a) Variações sistemáticas com periodicidade diária para a umidade. b) Variações sistemáticas com periodicidade semanal (também em d). (c) Variação anual.

Como no caso da temperatura, a umidade no Ibirapuera segue o mesmo padrão que a variável possui nas outras localidades. O perfil diário (Figura 6.36a) possui um mínimo em torno das 14:00 h. É interessante notar a queda na variação anual após a 32ª semana (Figura 6.36c), fato ocorrido também no P. D. Pedro II e em São Caetano. As Figuras 6.35b e 6.35d não mostram variação significativa durante a semana para a umidade.

Os resultados apresentados até aqui referem-se às variáveis CO, NO, NO₂, O₃, SO₂, Temp e Umid, registrados pela instrumentação das Estações de Medição Parque D. Pedro II, São Caetano do Sul e Ibirapuera, durante o ano de 1999.

As variáveis CH₄ e NMHC não foram registradas (de acordo com os relatórios da CETESB) na Estação Ibirapuera nos períodos estudados. Assim, o estudo das influências de CH₄ e NMHC na atmosfera da RMSP só pode ser considerado para o Parque D. Pedro II e São Caetano do Sul. São tomados todos os resultados para as sete variáveis citadas inicialmente nessas duas Estações de Medição (dados do Quadro VI.1 para Pedro II e São Caetano) incluindo-se os dois últimos poluentes conforme indicado no Quadro VI.2.

O bloco de análise de dados citado acima (Quadros VI.1 e VI.2) torna-se completo a partir da inclusão das análises PCA e PARAFAC para o CH₄ e NMHC. Esses resultados são mostrados nas Figuras 6.37 a 6.40 e 6.41 a 6.44, respectivamente, para as Estações de Medição P. D. Pedro II e São Caetano do Sul.

CH₄ e NMHC: P. D. Pedro II – 1999

6.3.1.19 – Análise PCA para o CH₄

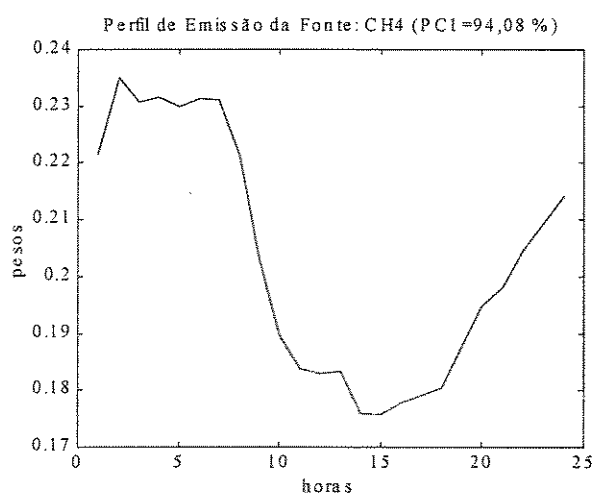


Fig.6.37: Pesos para a primeira componente principal, PC₁.

O perfil diário apresentado pelo CH₄ difere daqueles obtidos para as fontes automotivas. Nota-se pela Figura 6.37 que ocorre uma queda a partir de aproximadamente 7:00 h até alcançar um mínimo às 15:00 h, passando a crescer continuamente enquanto se aproxima das 24:00 h. A partir daí os valores permanecem constantes até cerca de 7:00 h.

6.3.1.20 – Análise PARAFAC (Pedro II): Modelo para o CH₄

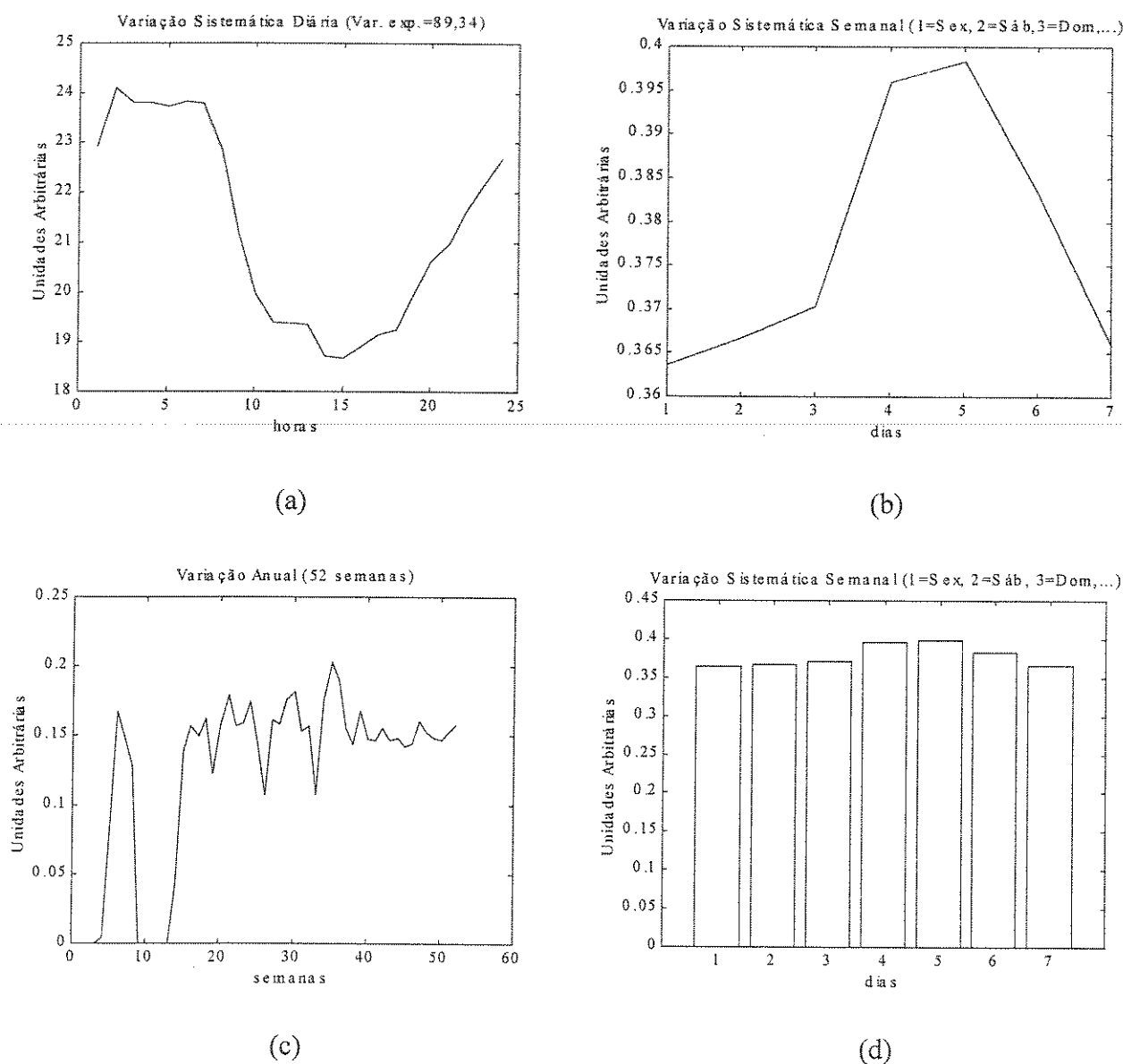


Fig.6.38: a) Variações sistemáticas com periodicidade diária. b) e d) Variações sistemáticas com periodicidade semanal. c) Variação anual. Os gráficos b e d mostram valores mais elevados às 2^a, 3^a e 4^a feiras e valores mais baixos nos demais dias. Os valores iguais a zero nas primeiras (e em torno da décima) semanas são originalmente nulos. Em c, pode-se notar maior acumulação nas semanas correspondentes ao período de inverno.

A Figura 6.38a ratifica o resultado obtido para o perfil diário na análise PCA. Uma inspeção na Figura 6.38b (e 6.38d) mostra valores mais elevados de 2^a a 4^a feira, e queda nos demais dias. A Figura 6.38c não mostra uma maior acumulação no inverno.

6.3.1.21 – Análise PCA para o NMHC

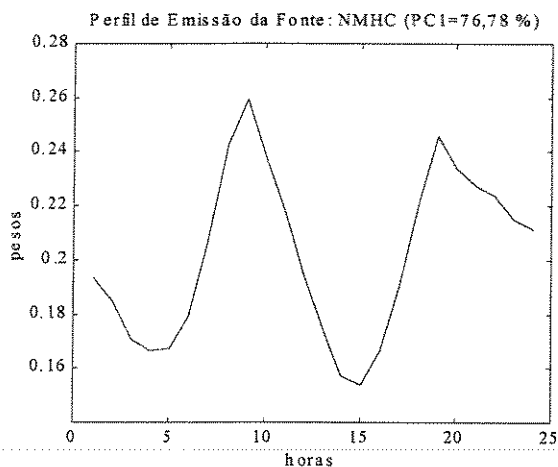
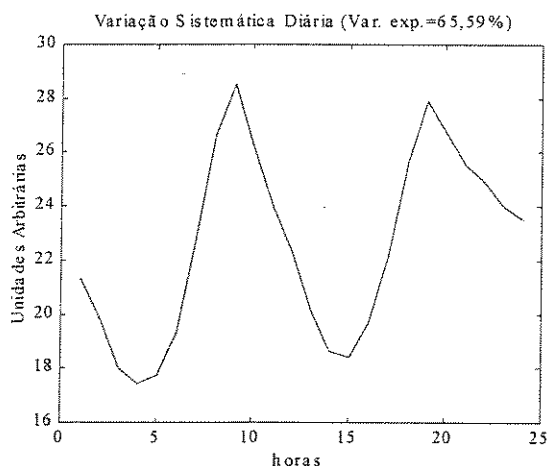


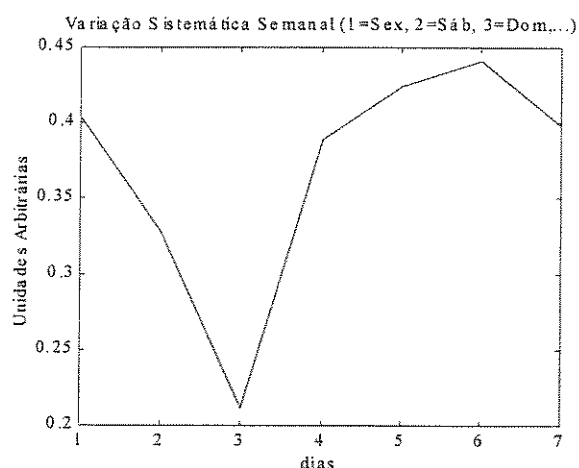
Fig.6.39: Pesos para a primeira componente principal, PC₁.

O perfil diário para os hidrocarbonetos não metano, NMHC, mostra a formação de um pico em torno das 9:00 h, um vale aproximadamente às 15:00 h, e um segundo pico em torno das 19:00 h. Esse resultado mostra que a principal fonte de emissão desse poluente constitui-se de veículos automotores, tal a sua concordância com o perfil apresentado pelos outros poluentes (monóxido de carbono e óxidos de nitrogênio) emitidos pela mesma fonte.

6.3.1.22 – Análise PARAFAC (Pedro II): Modelo para o NMHC



(a)



(b)

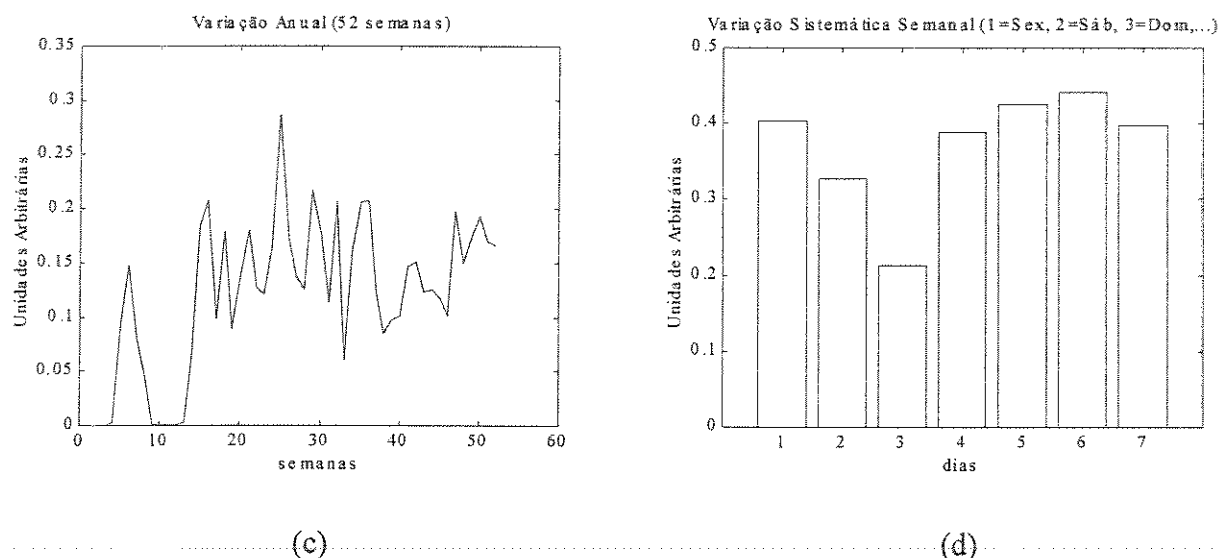


Fig.6.40: a) Variações sistemáticas com periodicidade diária. b) e d) Variações sistemáticas com periodicidade semanal. c) Variação anual.. Os gráfico b e d mostram queda nas emissões nos finais de semana, com valores mais elevados nos dias úteis. Os valores iguais a zero nas primeiras (e em torno da décima) semanas são originalmente nulos Em c, pode-se notar maior acumulação nas semanas correspondentes ao período de inverno.

Como nos casos dos outros poluentes, há concordância com os resultados obtidos na análise PCA para o perfil diário. Embora existam outras fontes de emissão para os NMHC, as fontes automotivas são de longe as mais importantes como mostram as análises PCA (Fig.9.39) e PARAFAC (Fig.6.40).

Os resultados obtidos com o modelo PARAFAC para os NMHC mostram que o modelo capta as variações sistemáticas diárias e semanais, e a variação anual, como já mostrado para os outros poluentes. A Figura 6.40a mostra as variações sistemáticas diárias. Tal perfil mantém a semelhança que os perfis dos poluentes primários emitidos por fontes automotivas possuem com o do tráfego de veículos. As Figuras 6.40b e 6.40d mostram as variações sistemáticas semanais. Nota-se aí a queda nos finais de semana com os valores mais elevados nos chamados dias úteis. A Figura 6.40c mostra a variação anual, com o aumento de amplitude no período de inverno. Do mesmo modo que nos casos anteriores os resultados são consistentes com a análise PCA.

6.3.1.23 – Análise PCA para o CH₄

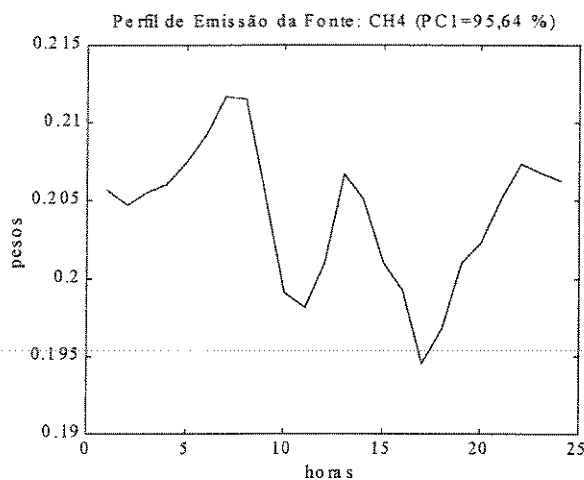
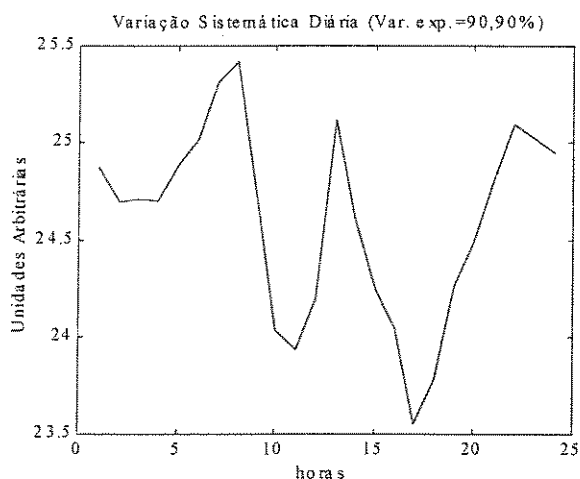


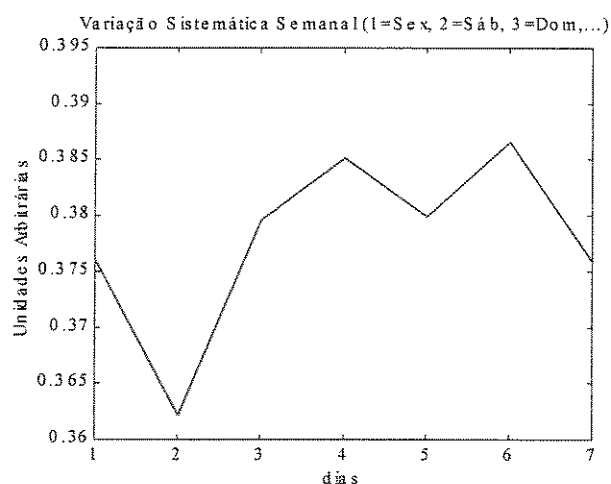
Fig.6.41: Pesos para a primeira componente principal, PC₁.

A Figura 6.41 mostra o perfil diário para a emissão de CH₄ registrada no São Caetano. Aqui o perfil possui um “traço” de emissões automotivas. Há crescimento nas primeiras horas da manhã, com um pico em torno das 7:00-8:00 h, queda até cerca de 10:00-11:00 h, crescimento com outro pico às 13:00 h, nova queda com mínimo às 17:00 h, para voltar a crescer enquanto se aproxima de um terceiro pico em torno de 22:00 h.

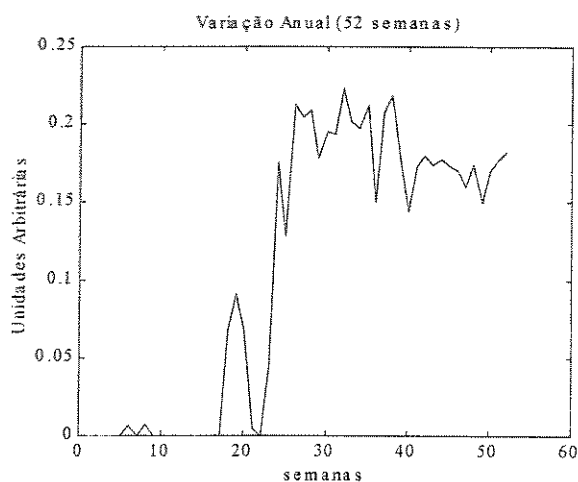
6.3.1.24 – Análise PARAFAC (São Caetano): Modelo para o CH₄



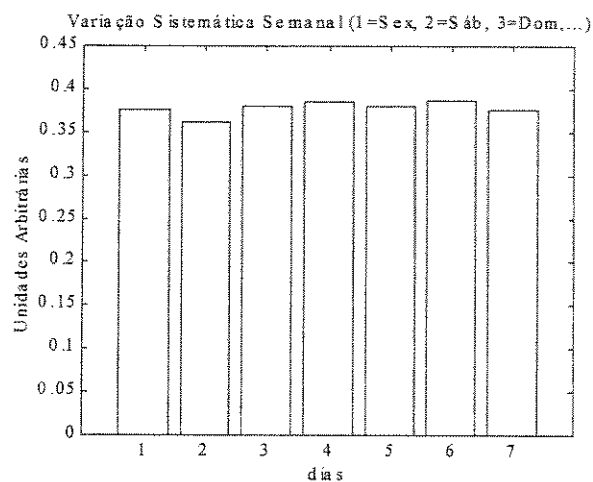
(a)



(b)



(c)



(d)

Fig.6.42: a) Variações sistemáticas com periodicidade diária. b) e d) Variações sistemáticas com periodicidade semanal. c) Variação anual. Os gráficos b e d mostram valores aproximadamente constantes com ligeira queda no sábado. Em c, (não obstante valores nulos até ~18ª semana) pode ser observada maior acumulação no período de inverno.

Os resultados mostrados na Figura 6.42 mostram concordância com a PCA (Figura 6.41) e diferenças em relação àqueles obtidos para o CH₄ no Parque D. Pedro II para as variações sistemáticas diárias e semanais e concordância para a variação anual.

6.3.1.25 – Análise PCA para o NMHC

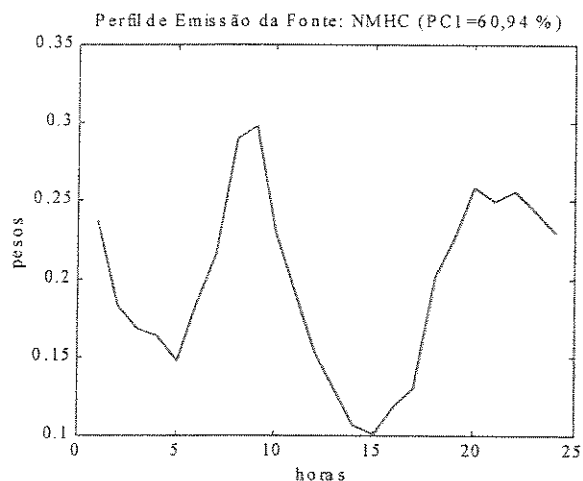


Fig.6.43: Pesos para a primeira componente principal, PC₁.

A Figura 6.43 mostra um pico em torno das 9:00 h, queda contínua com um vale às 15:00 h, com outro pico em torno das 20:00h, como no caso do P. D. Pedro II.

6.3.1.26 – Análise PARAFAC (São Caetano): Modelo para o NMHC

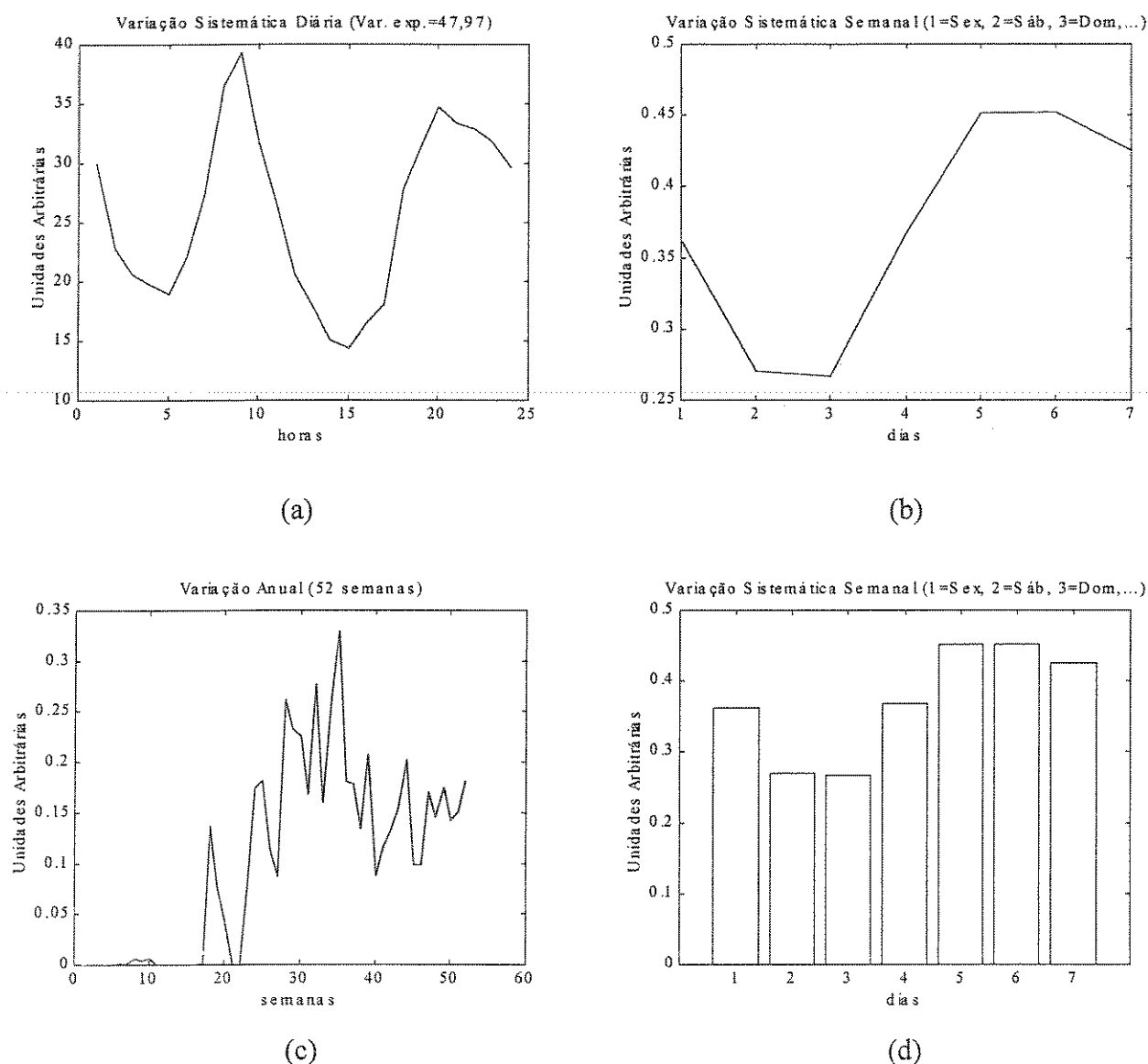


Fig.6.44: a) Variações sistemáticas com periodicidade diária. b) Variações sistemáticas com periodicidade semanal. c) Variação anual. Como nos casos dos outros poluentes, há concordância com os resultados obtidos na análise PCA para o perfil diário. O gráfico b mostra uma queda nas emissões nos finais de semana, com valores mais elevados nos outros dias. Esse mesmo resultado é mostrado em d. Em c, pode-se notar maior acumulação nas semanas correspondentes ao período de inverno.

Os resultados na Figura 6.44 são semelhantes àqueles obtidos no Parque D. Pedro II para essas emissões sugerindo, portanto, que a principal fonte emissora de NMHC em São Caetano também é automotiva.

6.4 – Análise Multimodo

Na seção 6.3 foram apresentadas análises por meio do modelo PARAFAC. Entretanto, naquela seção, tal arranjo de ordem superior foi usado para se estudar os perfis diários e semanais de uma única variável por vez. Não obstante o modelo PARAFAC se aplique a estruturas multimodo, as citadas análises foram apresentadas juntamente com PCA para que se pudesse comparar os perfis diários (PCA e PARAFAC) e semanais (PARAFAC) extraídos em cada análise para cada variável individualmente. Tais resultados devem ser, na prática, consistentes tanto no plano das discussões em nível teórico quanto no desenvolvimento da metodologia aplicada.

Na presente seção são apresentados os resultados extraídos das análises multimodo considerando-se não somente uma, mas um grupo de variáveis por análise. Enquanto nas primeiras análises (seção 6.3) procurou-se descobrir o comportamento de cada variável individualmente, na seção 6.4 procura-se descobrir o comportamento das variáveis à luz das influências que cada uma delas sofre pelas demais. Assim, as análises devem levar em conta não apenas as influências internas a cada ciclo (o do monóxido de carbono e o do metano), mas também as influências entre os ciclos. Adicionalmente, variáveis não pertencentes aos referidos ciclos (por exemplo, SO₂) são introduzidas para se estudar como os ciclos se comportam sob a influência das mesmas.

O Quadro VI.4 contém o conjunto de variáveis envolvidas nos diferentes casos estudados nessa seção. São estudadas combinações dessas variáveis de acordo com o ciclo a qual pertencem. Basicamente, a análise considera, para o ano de 1999, o ciclo do monóxido de carbono (CO, NO, NO₂ e O₃), o ciclo do metano (CH₄, NO, NO₂ e O₃) e a junção dos dois ciclos em cada Estação de Medição, Parque D. Pedro II (no corpo do texto) e São Caetano do Sul (Apêndice C1). Posteriormente são introduzidas as variáveis SO₂ e NMHC e refaz-se a análise dos ciclos à luz dessas duas variáveis.

Nesse ponto deve-se chamar a atenção para um fato. A inclusão dos NMHC nos dois ciclos acima revelou um aspecto novo: os NMHC possuem um comportamento típico tal que o seu estudo sugeriu a criação de um outro ciclo, o “*Ciclo dos Hidrocarbonetos Não Metano*”.

Alguns dos resultados das análises realizadas com os dados registrados nos outros períodos (1997-1999) são apresentados no Capítulo V. Destes, aos mais importantes,

imprescindíveis para uma melhor compreensão do conjunto das análises, far-se-á menção no capítulo que discute os principais resultados para o ano de 1999, os quais aparecem no corpo do texto.

Quadro VI.4: Análises Multimodo (PARAFAC e Tucker)

Estação	Ano	CO	NO	NO ₂	O ₃	SO ₂	CH ₄	NMHC	Modelos
P. D. Pedro II	1999	x	x	x	x	x	x	x	PARAFAC e Tucker

Das análises de cada poluente individualmente foi possível identificar variáveis com elevado grau de correlação. Por exemplo, o perfil diário para a temperatura é uma imagem invertida do perfil para a umidade. Por outro lado, o perfil para a temperatura é muito semelhante ao do ozônio, tal que os níveis para as variáveis temperatura e ozônio apresentam correlação positiva entre si e negativa com os níveis para a umidade. Assim, a inclusão das variáveis temperatura e umidade nas análises das estruturas multimodo não contribui para melhorar a interpretação dos resultados. Pelo contrário, a variável ozônio as substitui com vantagens (correlação positiva com uma delas e negativa com a outra).

A variável Partícula Inalável, PI, enseja o estudo dos fenômenos de adsorção e absorção com influência na formação de aerossóis e, assim, no equilíbrio entre as fases gasosa e líquida das espécies envolvidas, escapando ao escopo dessa tese. No entanto, uma breve discussão sobre o seu comportamento na atmosfera da RMSP é apresentado separadamente no Apêndice C2.

A partir desse ponto são apresentados os resultados obtidos com os modelos PARAFAC e Tucker nas análises das variáveis CO, NO, NO₂, O₃, SO₂, CH₄ e NMHC para a Estação de Medição Parque D. Pedro II.

PARAFAC : CO, NO, NO₂ e O₃ (P. D. Pedro II – 1999)

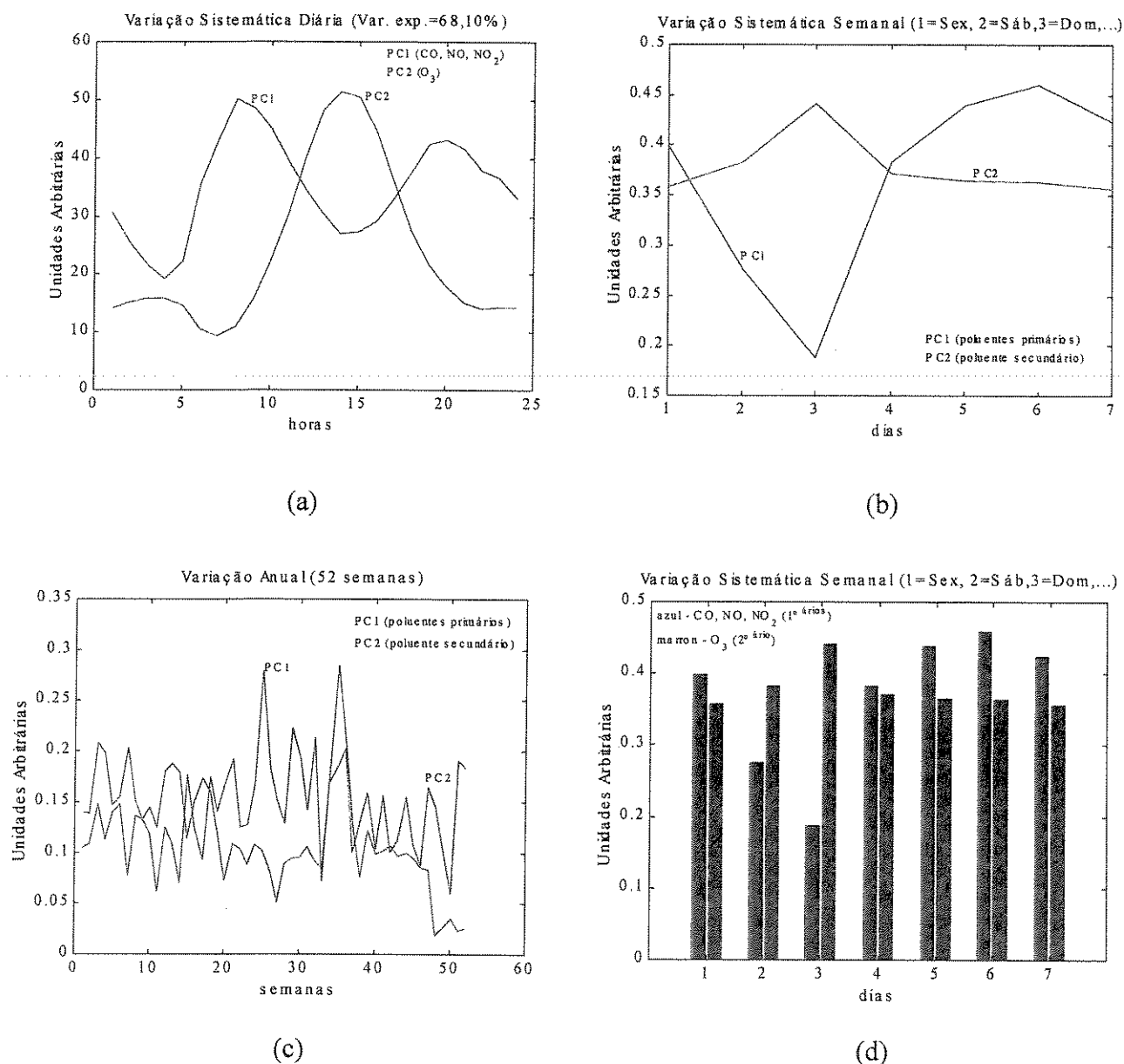


Fig.6.45: a) Variações sistemáticas diárias. b) Variações sistemáticas semanais, c) Variação anual. d) Representação em barras para a variação sistemática semanal.

Em a, PC₁ descreve o perfil da fonte de emissão devido ao tráfego de veículos e PC₂ representa o perfil para o poluente secundário, O₃. Os gráficos b e d mostram que nos finais de semana as emissões de veículos caem enquanto o O₃ cresce. O gráfico c mostra que no inverno os poluentes primários (PC₁) acumulam mais do que o secundário (PC₂).

PARAFAC : CH₄, NO, NO₂ e O₃ (P. D. Pedro II – 1999)

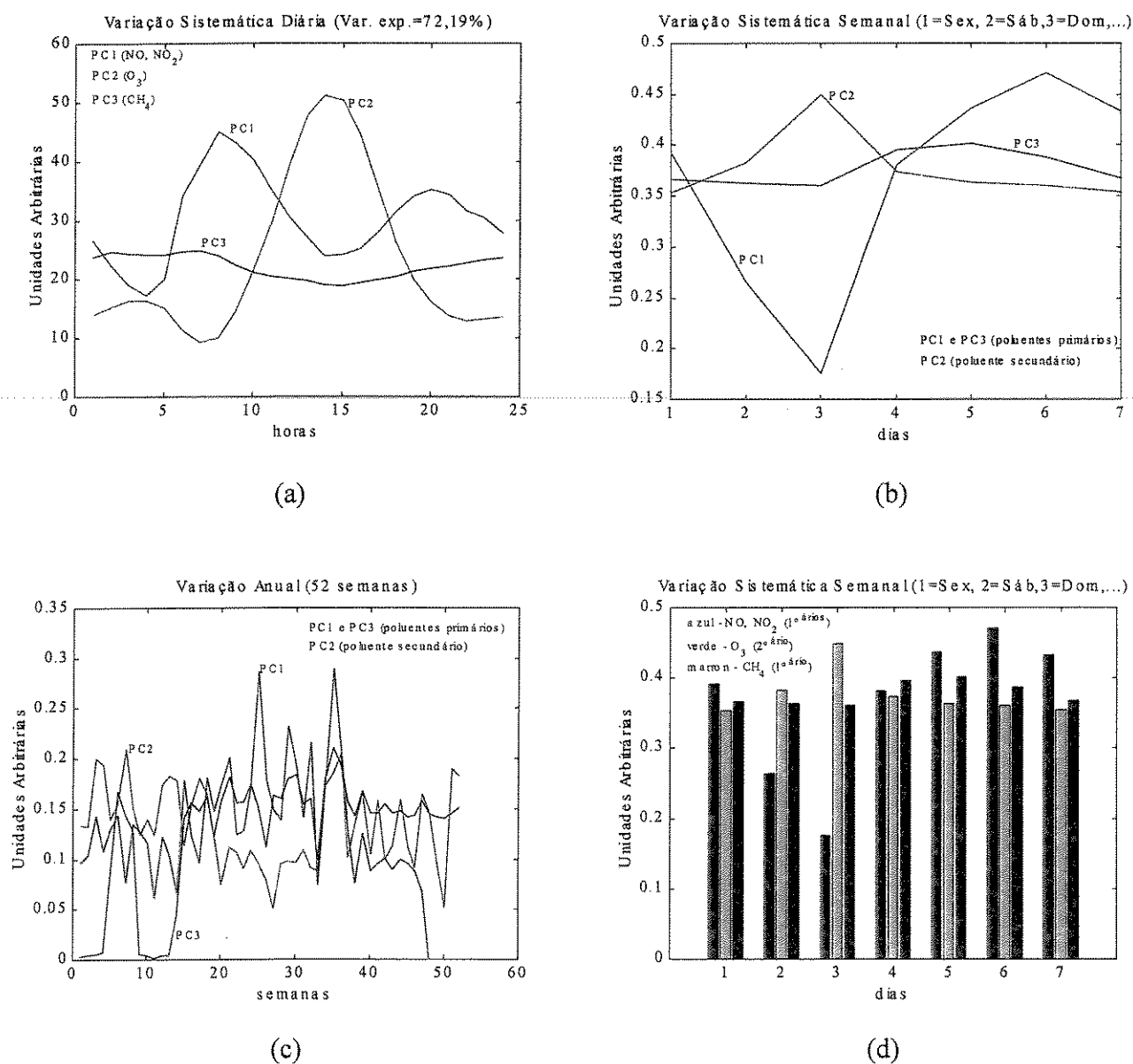


Fig.6.46: a) Variações sistemáticas diárias. b) Variações sistemáticas semanais, c) Variação anual. d) Representação em barras para a variação sistemática semanal.

A substituição do monóxido de carbono pelo metano para formar seu ciclo básico (CH₄, NO, NO₂ e O₃) mostra a necessidade da inclusão de uma terceira componente para representar o perfil do CH₄. Assim, na Figura 6.46, em a, PC₁ descreve o perfil diário da fonte de emissão devido ao tráfego de veículos (aqui devido somente a NO e NO₂), PC₂ representa o perfil diário para o poluente secundário, O₃, e PC₃ o perfil diário para o

metano. Do mesmo modo, o gráfico b mostra essas variações sistemáticas com periodicidade semanal. Aqui, PC_1 representa o perfil de emissão das fontes automotoras ao longo da semana e, PC_2 e PC_3 as variações sistemáticas semanais, respectivamente, para o ozônio e para o metano. Observa-se que ocorre queda das emissões veiculares nos finais de semana, enquanto para o CH_4 os valores apresentam uma leve queda nesse mesmo período, voltando a subir na segunda-feira, mantendo-se na terça e voltando a cair na quarta-feira. Esses resultados são também mostrados no gráfico de barras, em d. O gráfico c mostra as variações anuais para as variáveis acima mencionadas. PC_1 e PC_3 referem-se aos poluentes primários, isto é, representam as variações anuais para os compostos emitidos, respectivamente, pelos veículos e pelas fontes não automotoras, e PC_2 representa a variação anual para o ozônio. Nota-se que ocorre maior acumulação para os poluentes primários durante o inverno, ao contrário do que ocorre com o poluente secundário.

Do resultado acima pode-se inferir que os níveis de concentração encontrados para o metano, também um poluente primário, se devem, principalmente, à emissões de outras fontes que não as não automotivas. É provável que exista metano proveniente de fontes automotoras (por exemplo, de veículos movidos a gás natural), porém tais níveis, se puderam ser modelados, estarão contidos no perfil de emissão das fontes automotivas, portanto, em PC_1 , a menos que pertençam a algum tipo específico de fonte automotiva que exija uma componente adicional para capturar suas variações sistemáticas.

Particularmente, o resultado acima indica que os valores mais elevados para o ozônio nos finais de semana, exatamente quando caem as emissões automotivas, se devem, em alguma medida e adicionalmente à formação de HONO (e outros fatores), às reações do ciclo do metano. Esses resultados são discutidos mais adiante nesse capítulo, à luz das reações dos ciclos do monóxido de carbono e do metano.

PARAFAC: NMHC, NO, NO₂ e O₃ (P. D. Pedro II – 1999) – (1)

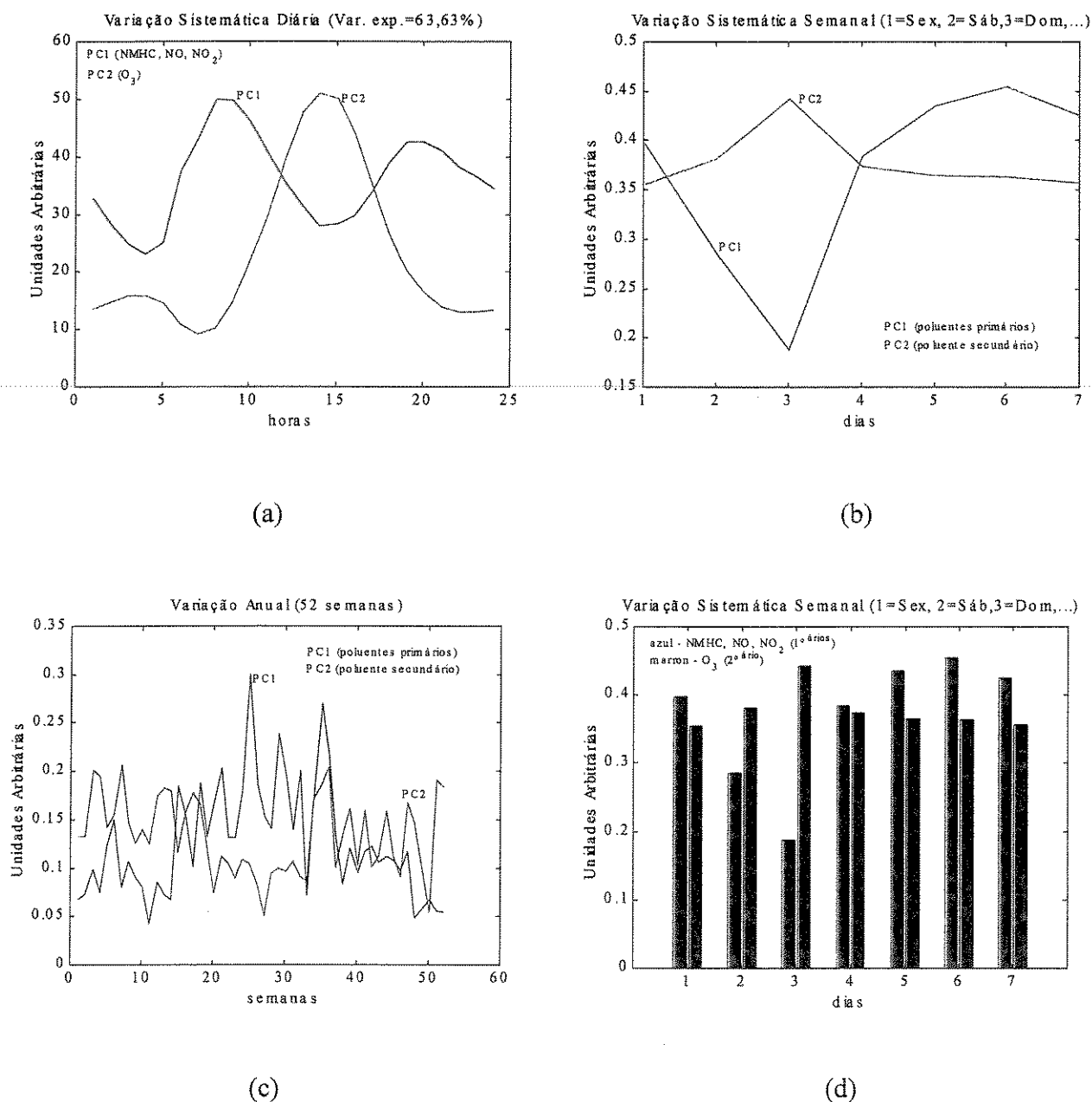


Fig.6.47.1: a)Variações sistemáticas diárias. b)Variações sistemáticas semanais, c)Variação anual. d) Representação em barras para a variação sistemática semanal.

Quando os NO_x e o O₃ são acrescentados dos NMHC (Figura 6.47.1), o modelo PARAFAC concentra as variações em duas componentes. O resultado é semelhante ao do ciclo básico do CO, indicando que boa parte das emissões de hidrocarbonetos não metano é proveniente de emissões veiculares. Porém, devem existir outras fontes de NMHC.

PARAFAC : NMHC, NO, NO₂ e O₃ (P. D. Pedro II – 1999) – (2)

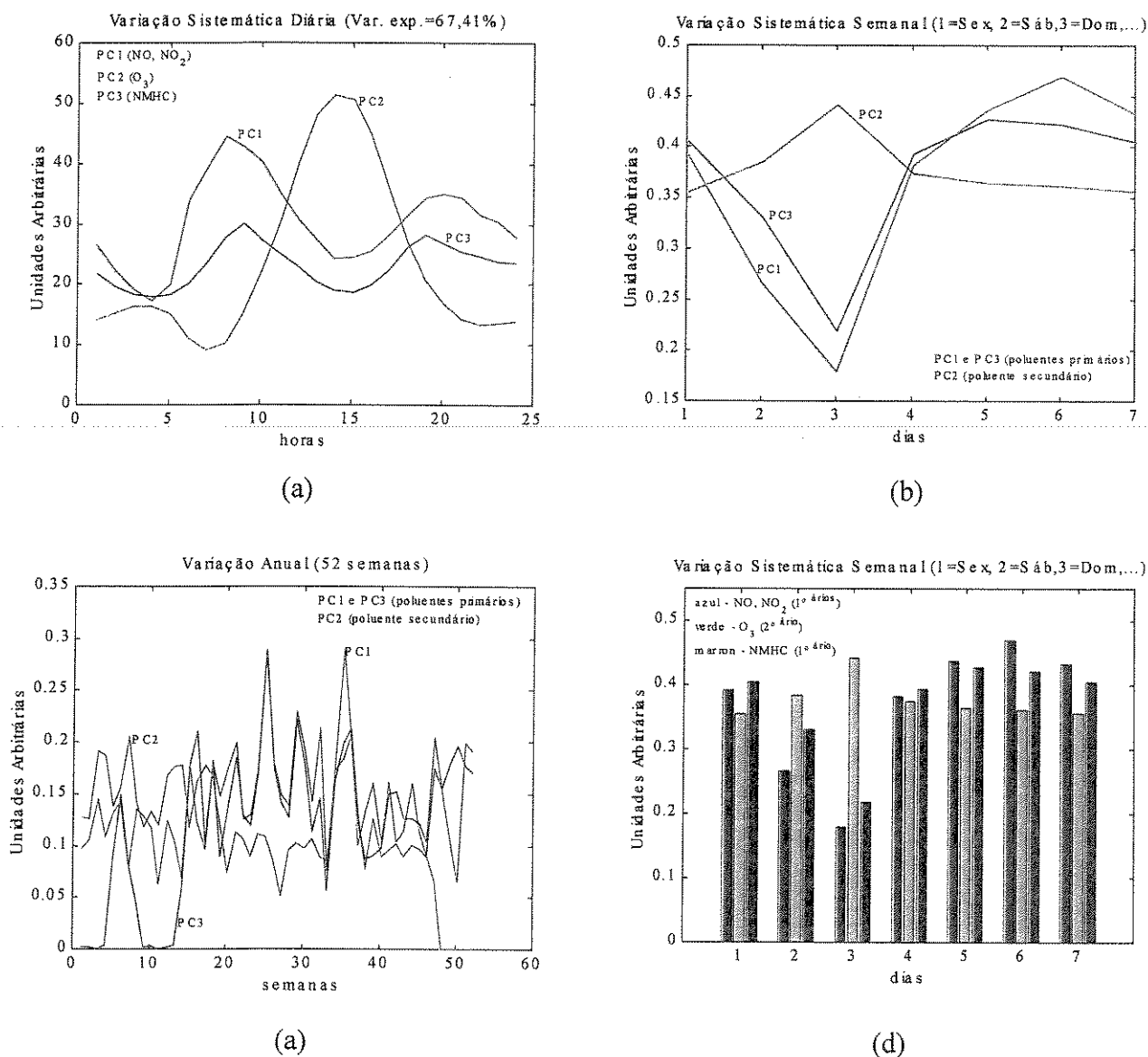


Fig.6.47.2: a)Variações sistemáticas diárias. b)Variações sistemáticas semanais, c)Variação anual. d) Representação em barras para a variação sistemática semanal.

Repetindo-se a análise com três fatores, o modelo PARAFAC passa a concentrar as variações em três componentes. O resultado ainda é semelhante ao do ciclo básico para o CO indicando, porém, que, além de boa parte das emissões de hidrocarbonetos não metano se originarem de emissões automotivas, há outra(s) fonte(s) de NMHC. Essas fontes estão de alguma forma relacionadas com a fonte automotiva, como mostra a Figura 6.47.2.

PARAFAC : SO₂, NO, NO₂ e O₃ (P. D. Pedro II – 1999)

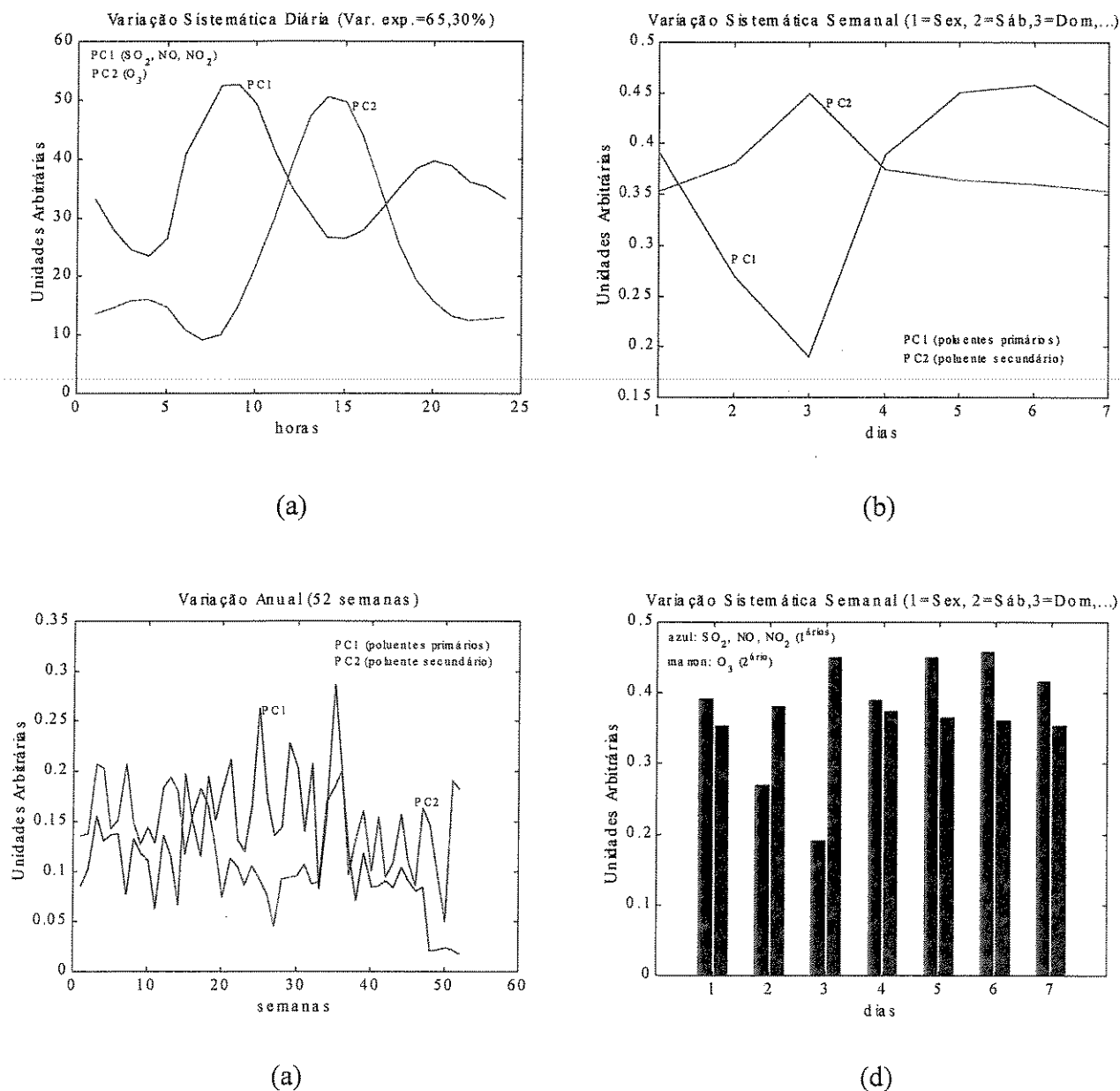


Fig.6.48 a) Variações sistemáticas diárias. b) Variações sistemáticas semanais, c) Variação anual. d) Representação em barras para a variação sistemática semanal.

Quando a variável introduzida é o SO₂, o modelo PARAFAC também concentra as variações em apenas duas componentes. O resultado é semelhante ao do ciclo básico para o CO indicando que as emissões de dióxido de enxofre são principalmente provenientes de veículos automotores como mostra a Figura 6.48.

PARAFAC : CO, NO, NO₂, O₃ e CH₄ (P. D. Pedro II – 1999)

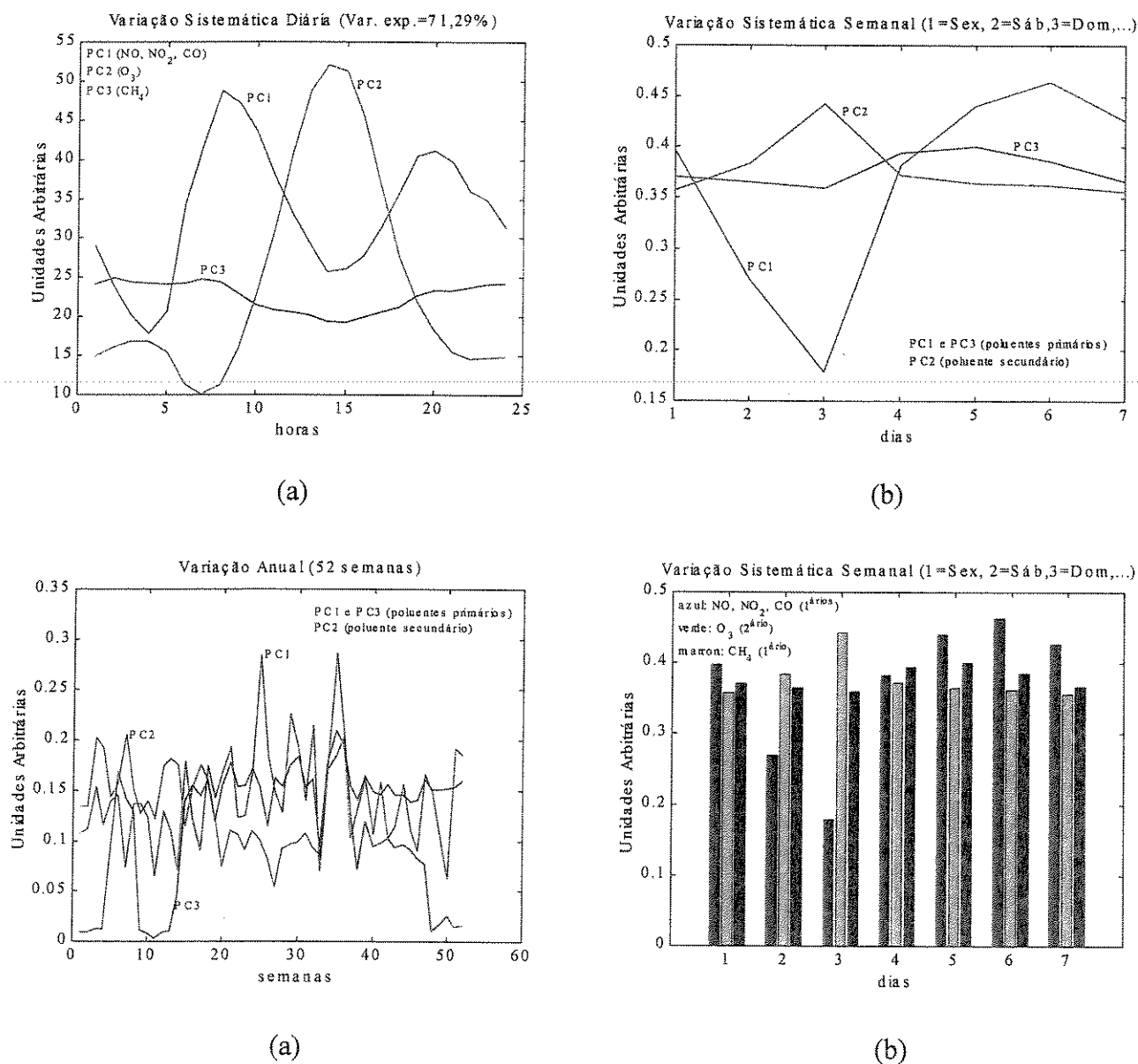


Fig.6.49: a) Variações sistemáticas diárias. b) Variações sistemáticas semanais, c) Variação anual. d) Representação em barras para a variação sistemática semanal.

A inclusão do metano no ciclo básico do monóxido de carbono também exige do modelo três componentes para representar os perfis das variáveis consideradas. Assim, na Figura 6.49, em a, PC₁ é o perfil diário da fonte de emissão devido ao tráfego de veículos, PC₂ representa o perfil diário para o poluente secundário, O₃, e PC₃ o perfil diário para o metano. O gráfico b mostra essas variações sistemáticas com periodicidade semanal. Aqui

PC₁ representa o perfil de emissão das fontes automotoras e, PC₂ e PC₃ as variações sistemáticas semanais, respectivamente, para o ozônio e para o metano. Tal como na Figura 6.46 (p317), observa-se queda acentuada das emissões veiculares nos finais de semana, enquanto para o CH₄ os valores apresentam uma leve queda no mesmo período, voltando a subir na segunda-feira, mantendo-se na terça e voltando a cair na quarta-feira. Esses resultados são também mostrados no gráfico de barras, em d. O gráfico c mostra as variações anuais, nas quais PC₁ e PC₃ referem-se aos poluentes primários e PC₂ representa a variação anual para o ozônio. Do mesmo modo que nos casos anteriores, nota-se maior acumulação para os poluentes primários durante o inverno, ao contrário do que ocorre com o poluente secundário.

Tais resultados são semelhantes àqueles mostrados na Figura 6.46 (p317) e confirmam a breve discussão apresentada na página 318. As variações para o monóxido de carbono estão na primeira componente PC₁, juntamente com as dos óxidos de nitrogênio, enquanto aquelas para o metano ficam contidas em PC₃. Como em 7.46, as variações para o ozônio estão representadas na componente PC₂.

Como se verá adiante, quando as variáveis consideradas são CO, NO, NO₂, O₃ e NMHC, o modelo PARAFAC volta acomodar as variações sistemáticas em apenas duas componentes, sugerindo que as emissões de hidrocarbonetos não metano são originárias principalmente de veículos. A aparente obviedade dessa afirmação deixa de existir quando se considera que uma dada fonte pode ser constituída de emissores diferentes, por exemplo, fonte automotiva desmembrada veículos pesados (ônibus e caminhões) e veículos leves, os quais utilizam combustíveis diferentes.

PARAFAC : CO, NO, NO₂, O₃ e NMHC (P. D. Pedro II – 1999)

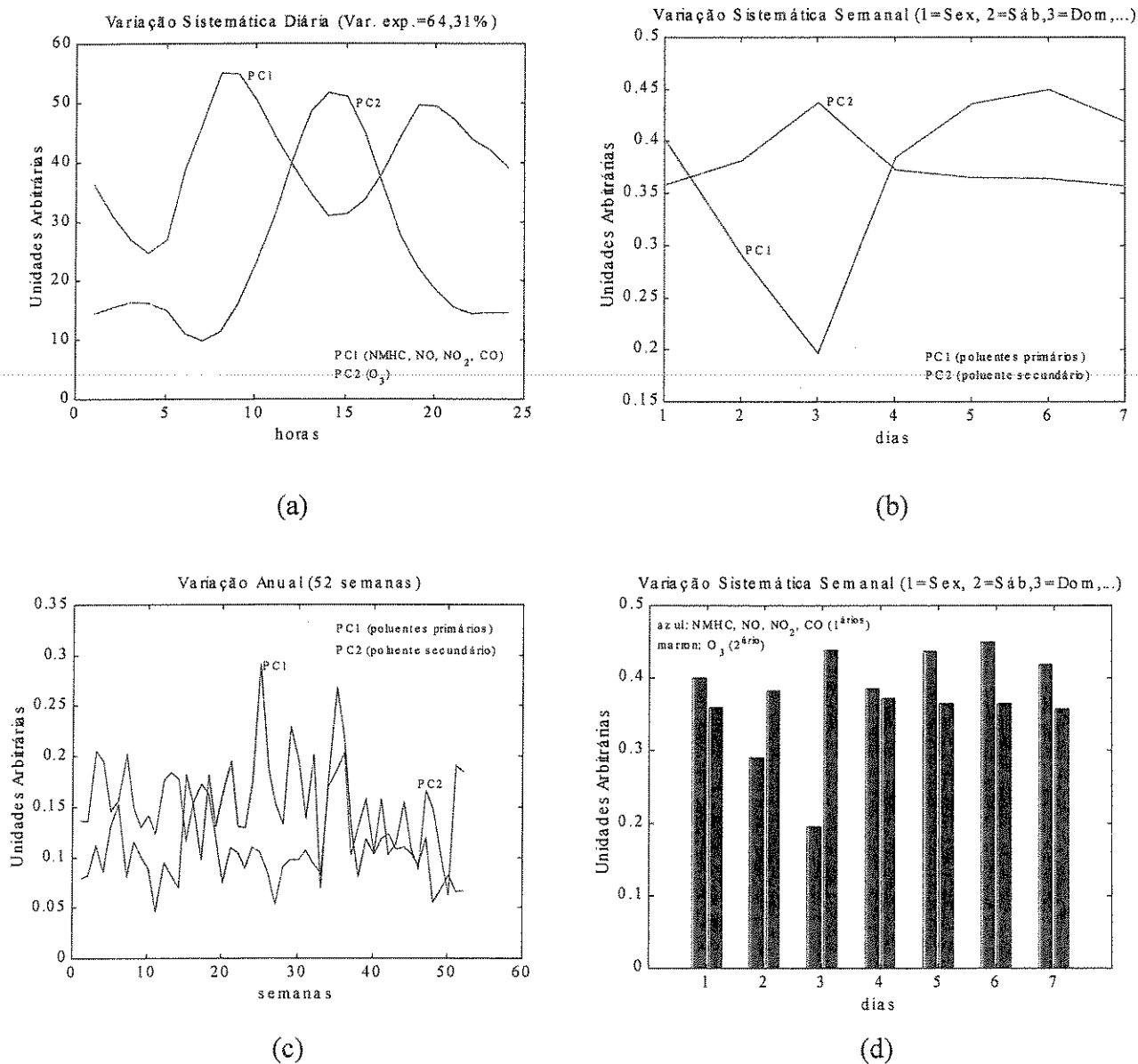


Fig.6.50: a) Variações sistemáticas diárias. b) Variações sistemáticas semanais, c) Variação anual. d) Representação em barras para a variação sistemática semanal.

Quando NMHC é incluído no ciclo básico do monóxido de carbono, as variações sistemáticas daquele podem ser acomodadas em PC₁, enquanto o perfil para o ozônio é representado por PC₂, como mostram os gráficos a, b, c e d da Figura 6.50.

PARAFAC : NMHC, NO, NO₂, O₃ e CH₄ (P. D. Pedro II – 1999)

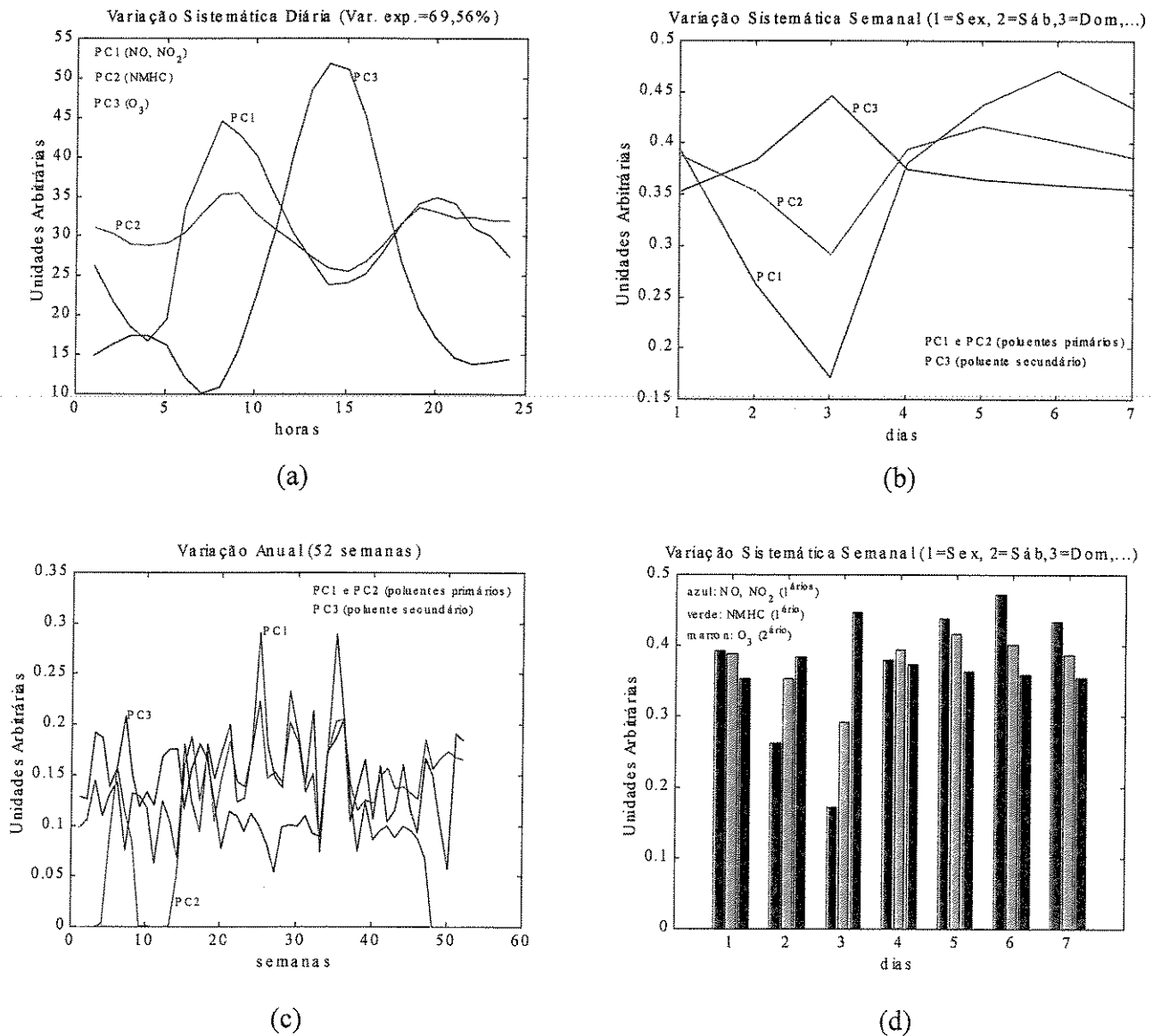


Fig.6.51: a) Variações sistemáticas diárias. b) Variações sistemáticas semanais, c) Variação anual. d) Representação em barras para a variação sistemática semanal.

Quando NMHC é incluído no ciclo básico do metano, as variações sistemáticas daquele não são totalmente acomodadas em PC₁ juntamente com NO e NO₂ como esperado (são emissões principalmente automotoras), mas sim em PC₂, enquanto o perfil para o ozônio é representado por PC₃, como mostram os gráficos a, b, c e d da Figura 6.51. Isto pode estar associado a emissões de um tipo específico de veículo, queda no transporte de combustível, e queda na utilização de postos de abastecimento nos fins de semana.

PARAFAC : CO, NO, NO₂, O₃ e SO₂ (P. D. Pedro II – 1999)

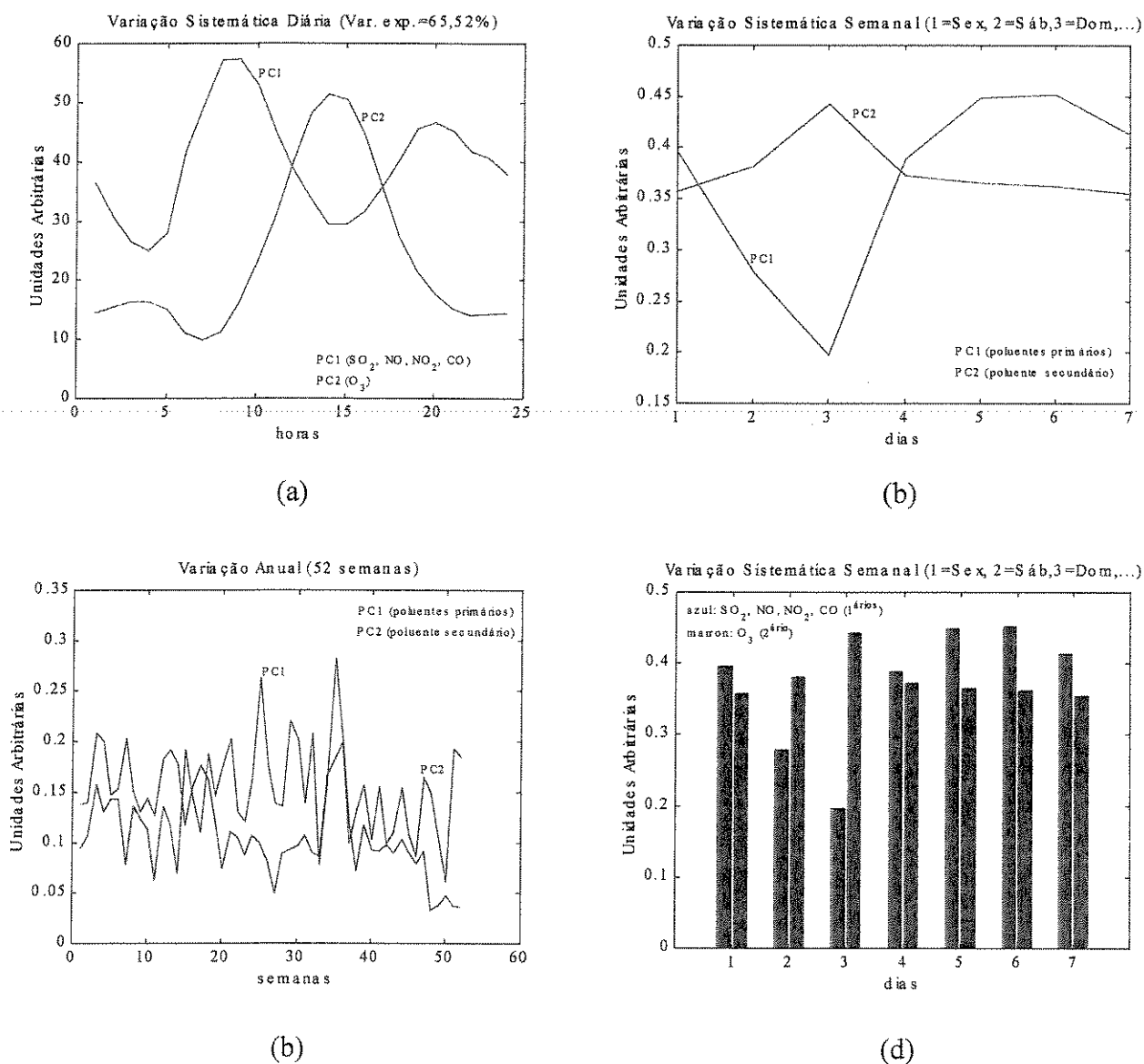


Fig.6.52: a) Variações sistemáticas diárias. b) Variações sistemáticas semanais, c) Variação anual. d) Representação em barras para a variação sistemática semanal.

Quando SO₂ é incluído no ciclo básico do monóxido de carbono, as variações sistemáticas do primeiro são acomodadas em PC₁ juntamente com CO, NO e NO₂ como esperado, pois suas emissões são provenientes de fontes automotivas. Por sua vez, o ozônio é acomodado em PC₂. Tais variações sistemáticas diárias assim como as semanais, e a variação anual são mostradas nos gráficos a, b, c e d da Figura 6.52.

PARAFAC : CH₄, NO, NO₂, O₃ e SO₂ (P. D. Pedro II – 1999)

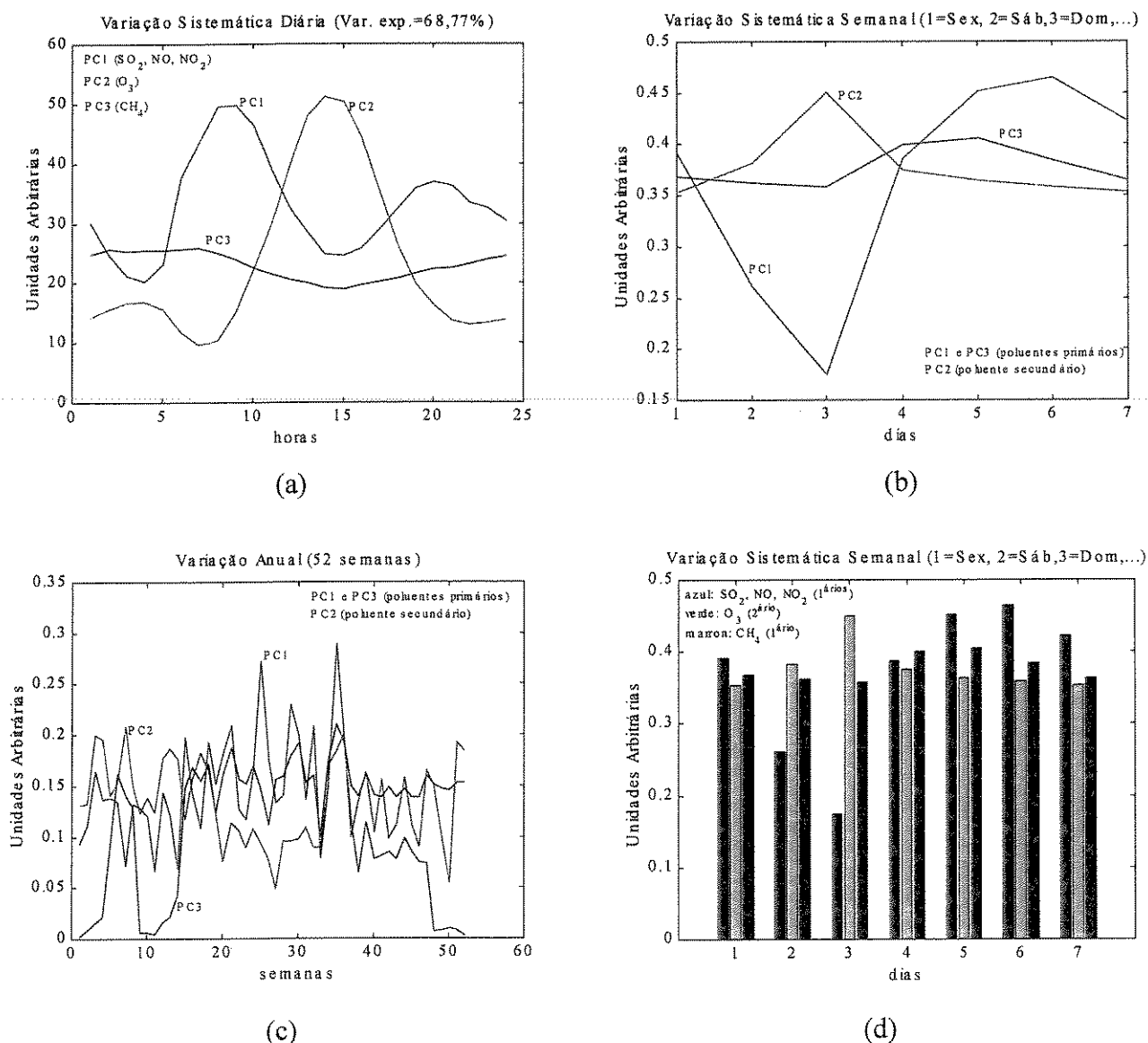


Fig.6.53: a) Variações sistemáticas diárias. b) Variações sistemáticas semanais, c) Variação anual. d) Representação em barras para a variação sistemática semanal.

A Figura 6.53 mostra o ciclo básico do metano acrescido do SO₂. Em a, PC₁ representa o perfil diário de emissão das fontes automotivas, PC₂ o poluente secundário, O₃, e PC₃ outras fontes primárias (CH₄). Em b e d pode-se observar a queda nos finais de semana para as fontes automotivas, valores mais elevados para o O₃ nesse período, e a variação sistemática para o CH₄. Em c aparecem as variações anuais para cada variável.

PARAFAC : NMHC, NO, NO₂, O₃ e SO₂ (P. D. Pedro II – 1999)

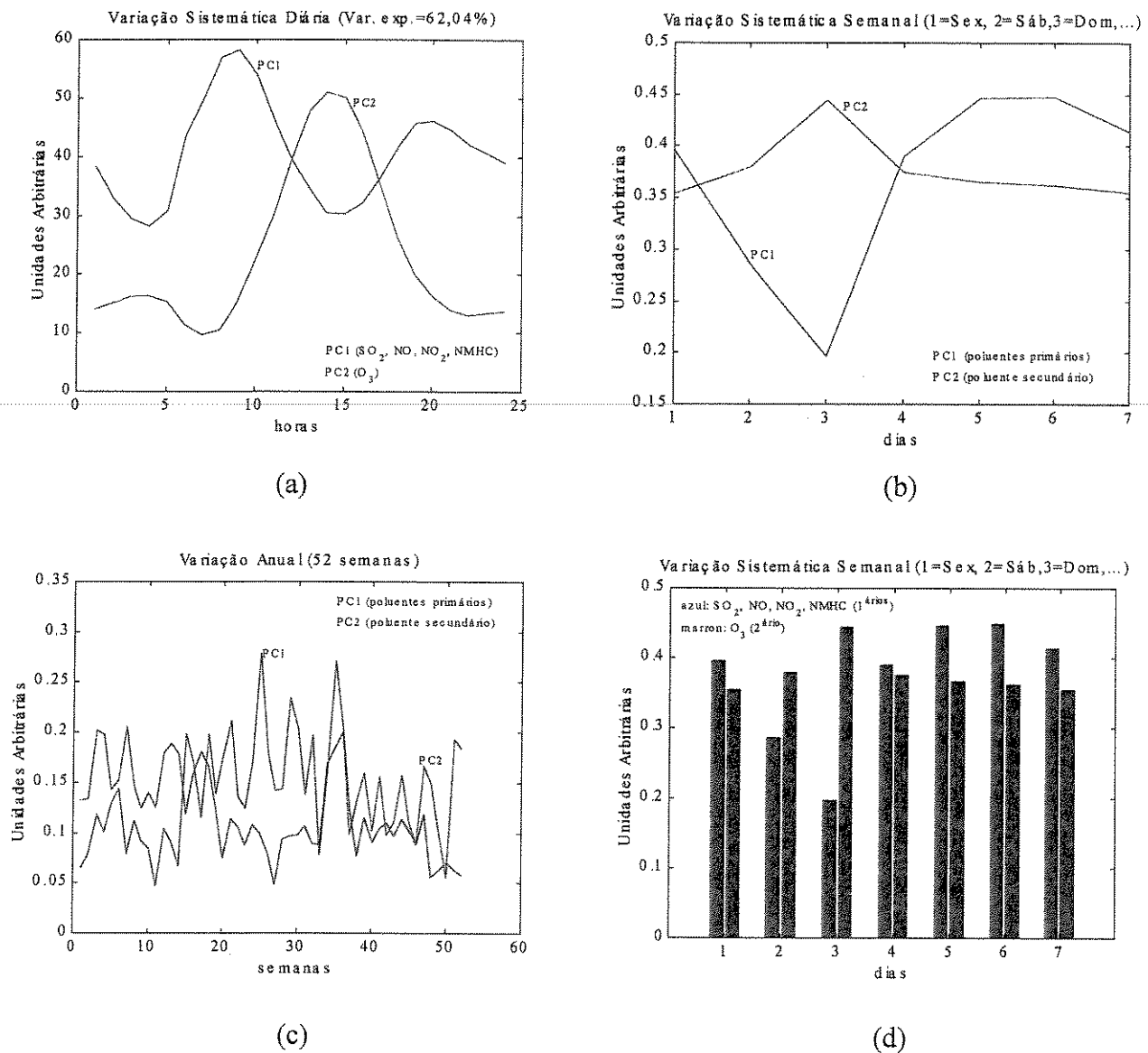


Fig.6.54: a) Variações sistemáticas diárias. b) Variações sistemáticas semanais, c) Variação anual. d) Representação em barras para a variação sistemática semanal.

A Figura 6.54 mostra que o modelo PARAFAC requer apenas duas componentes para descrever as variações sistemáticas diárias, semanais, e a variação anual dos compostos envolvidos. Em a, PC₁ registra o perfil de emissão das fontes automotivas, em b e d pode-se observar a queda dessas emissões nos finais de semana, e valores mais elevados para o ozônio no mesmo período, enquanto c mostra valores mais elevados para os poluentes primários durante o inverno e um quadro inverso para o secundário.

PARAFAC : CO, NO, NO₂, O₃, SO₂ e CH₄ (P. D. Pedro II – 1999)

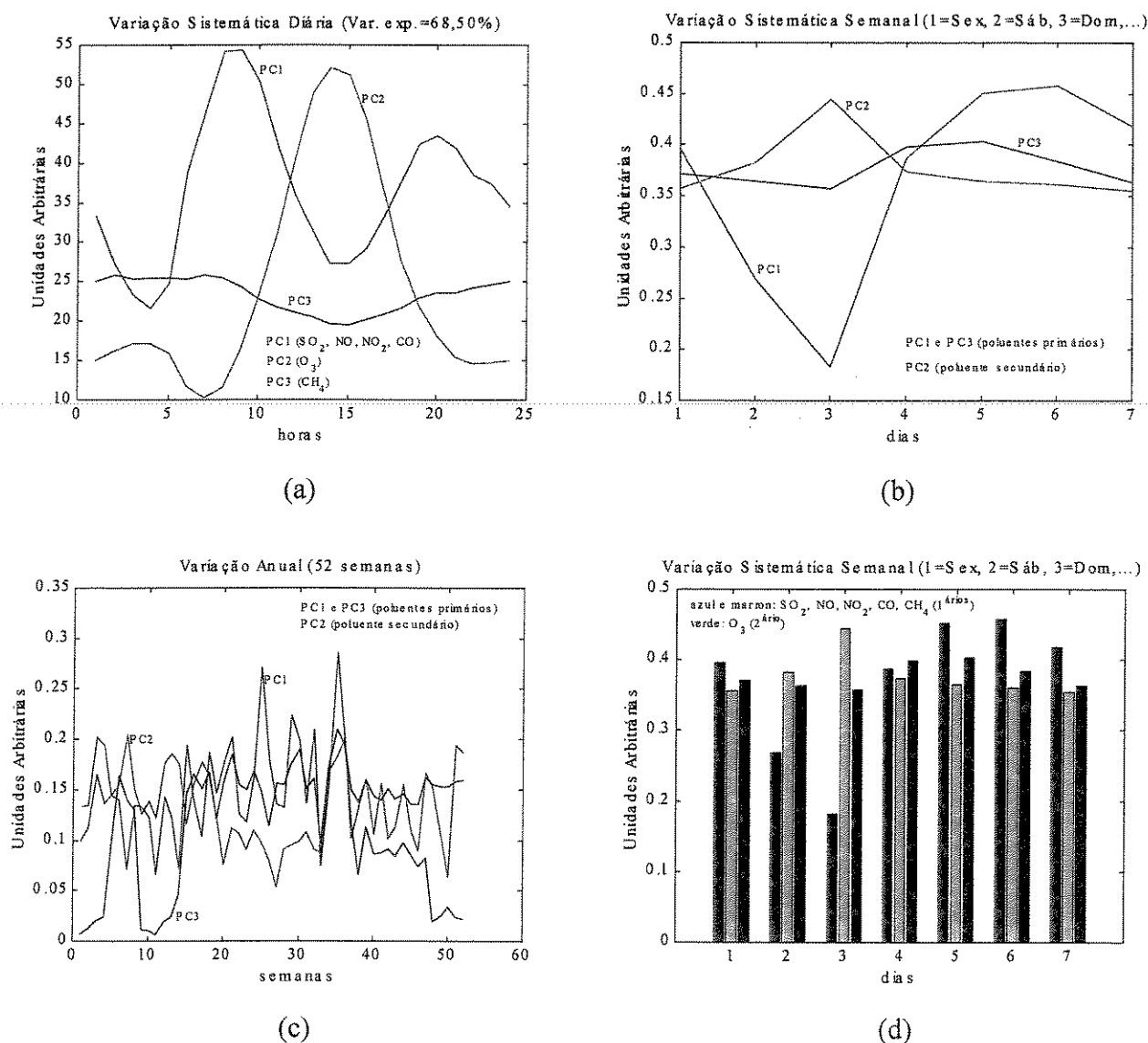


Fig.6.55: a) Variações sistemáticas diárias. b) Variações sistemáticas semanais, c) Variação anual. d) Representação em barras para a variação sistemática semanal.

A Figura 6.55 mostra que os perfis são acomodados em três PCs. Em a PC₁ registra o perfil diário de emissão das fontes automotivas, PC₂, o do O₃, e PC₃ o do CH₄. Em b e d observa-se queda das emissões automotoras nos finais de semana, valores mais elevados para o O₃ no mesmo período, e ligeira queda para o metano, o qual cresce na 2^a feira, tendo breve queda 4^a e 5^a feiras. A Figura 6.55c mostra valores mais elevados para os poluentes primários durante o inverno ao contrário do que ocorre para o poluente secundário.

PARAFAC : CO, NO, NO₂, O₃, SO₂ e NMHC (P. D. Pedro II – 1999)

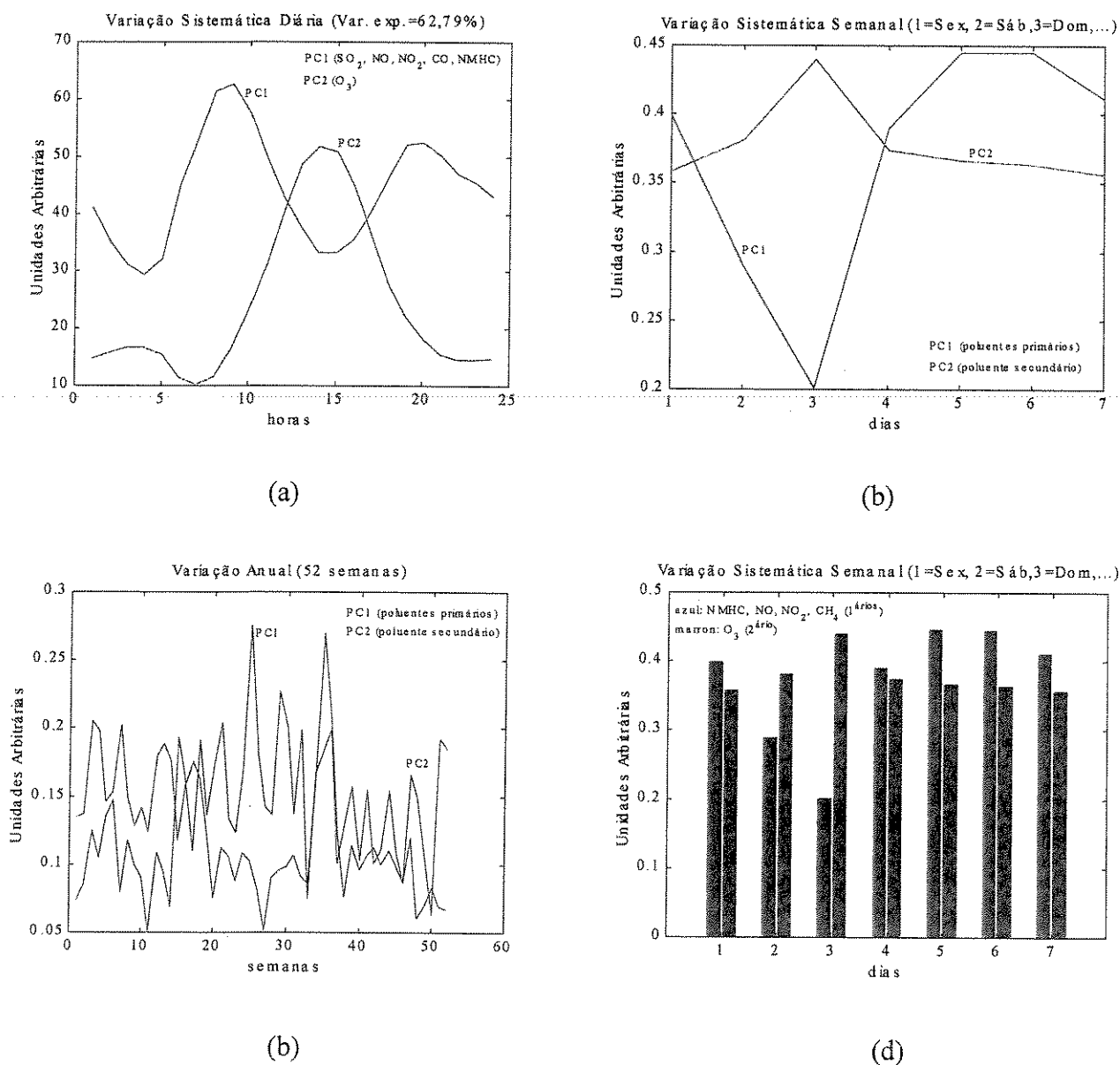


Fig.6.56: a) Variações sistemáticas diárias. b) Variações sistemáticas semanais, c) Variação anual. d) Representação em barras para a variação sistemática semanal.

A Figura 6.56 é semelhante à 6.55. Em a, PC₁ registra o perfil diário de emissão das fontes automotivas, e PC₂, o do O₃. Em b e d pode-se observar a queda das emissões automotoras nos finais de semana, e valores mais elevados para o ozônio no mesmo período, enquanto c mostra valores mais elevados para os poluentes primários durante o inverno ao contrário do que ocorre para o poluente secundário.

PARAFAC : CO, NO, NO₂, O₃, CH₄ e NMHC (P. D. Pedro II – 1999)

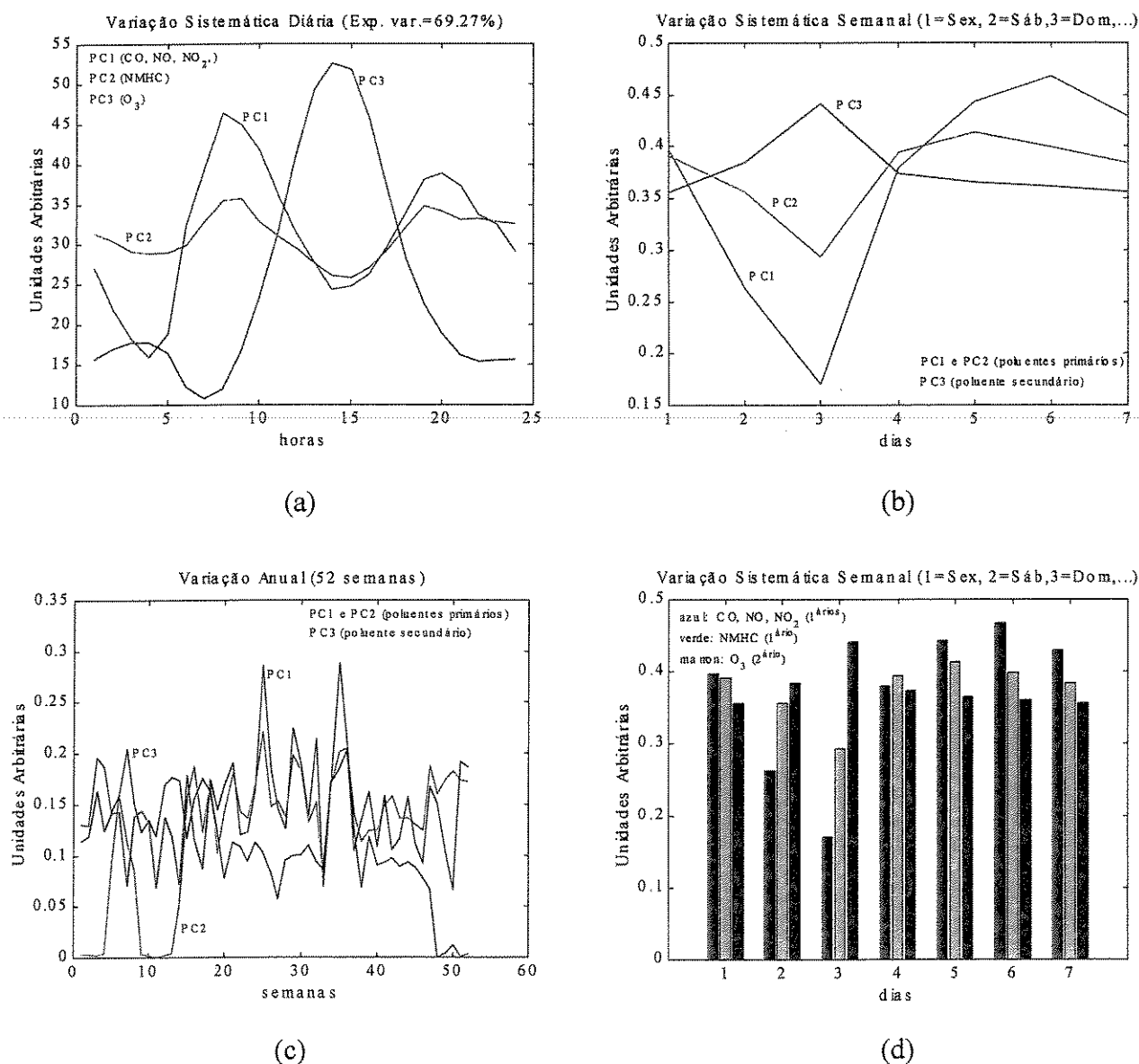


Fig.6.57: a) Variações sistemáticas diárias. b) Variações sistemáticas semanais. c) Variação anual. d) Representação em barras para a variação sistemática semanal.

Quando ao ciclos do CO acrescentam-se CH₄ e NMHC, o modelo acomoda em PC₁ as emissões automotivas. Entretanto, como na Figura 6.51 (p.341), modela NMHC em PC₂ e as variações para o O₃ estão representadas em PC₃, como mostram os gráficos a, b, c e d da Figura 6.57. Como se observa, novamente aqui NMHC, embora possua perfil característico de emissões automotoras, precisa de um fator exclusivamente para modelar essa fração de suas emissões.

PARAFAC : SO₂, NO, NO₂, O₃, CH₄ e NMHC (P. D. Pedro II – 1999)

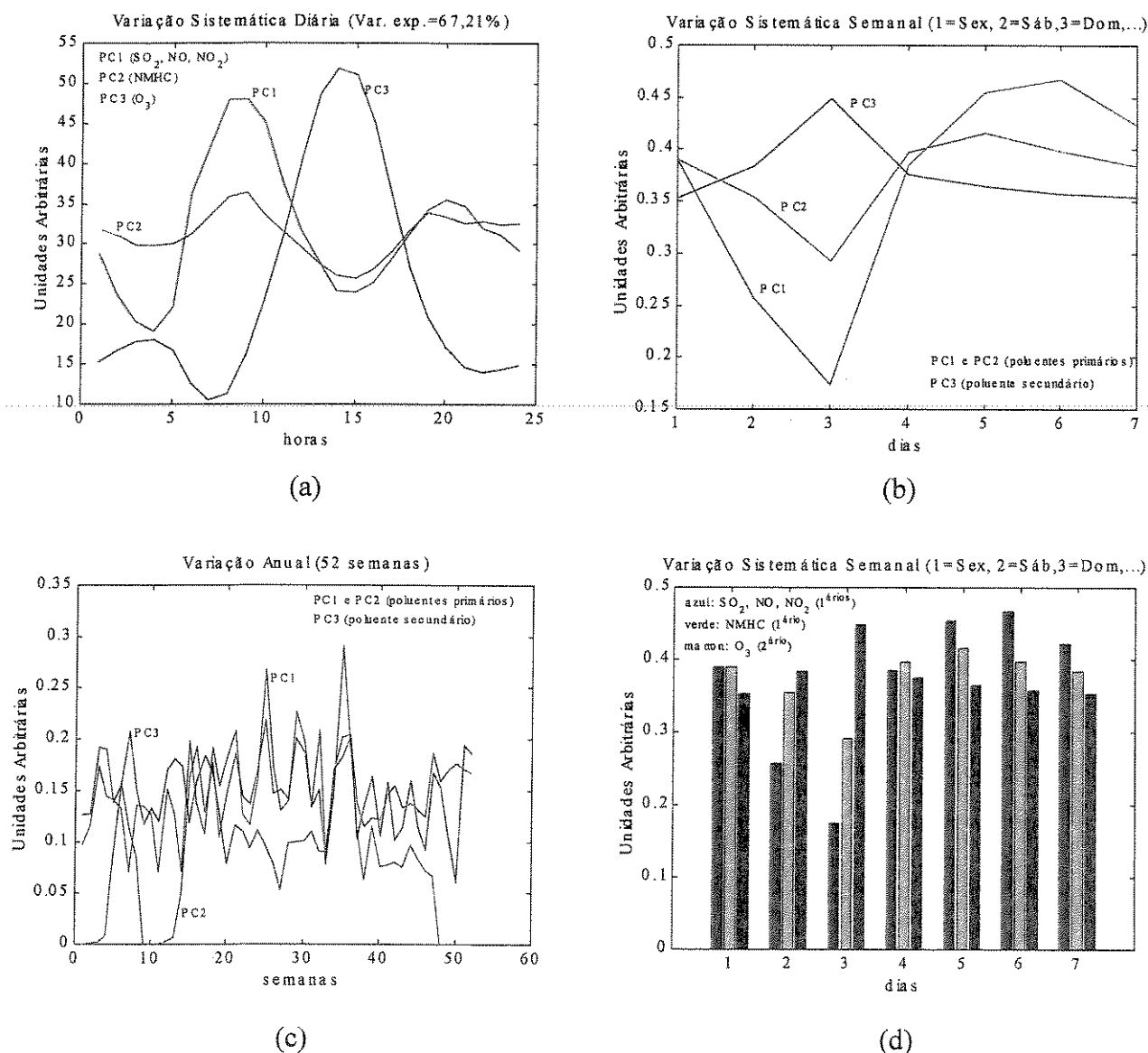


Fig.6.58: a) Variações sistemáticas diárias. b) Variações sistemáticas semanais. c) Variação anual. d) Representação em barras para a variação sistemática semanal.

Quando ao ciclo do CH₄ acrescentam-se SO₂ e NMHC, o modelo acomoda em PC₁ as emissões automotivas, entretanto, como na Figura 6.57, modela NMHC em PC₂ e as variações para o O₃ estão representadas em PC₃, como mostram os gráficos a, b, c e d da Figura 6.58. As Figuras 6.57 e 6.58 são muito parecidas. Os NMHC, também requerem componente adicional para modelar a fração de suas emissões automotivas.

PARAFAC : CO, NO, NO₂, O₃, SO₂, CH₄ e NMHC (P. D. Pedro II – 1999)

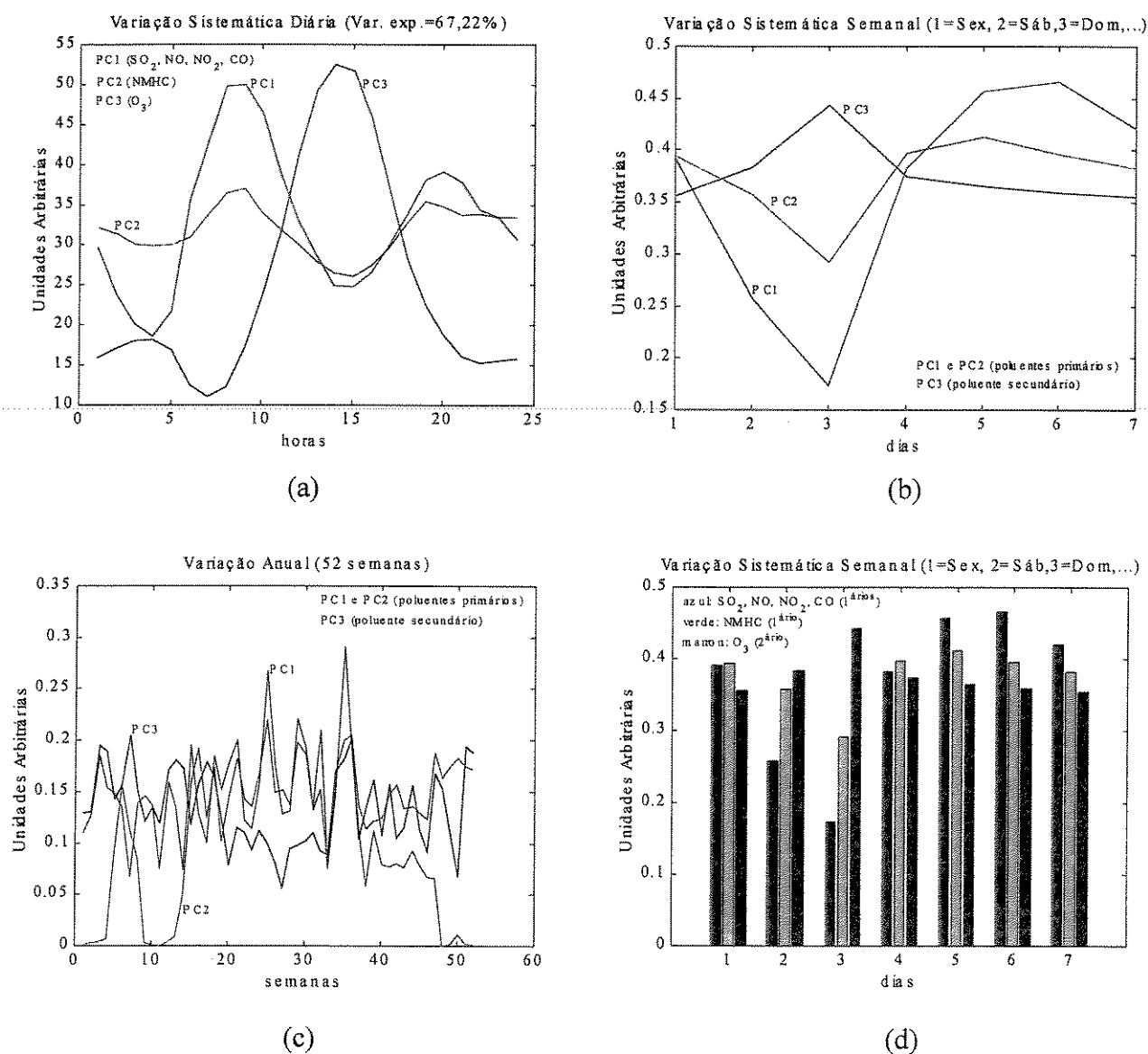


Fig.6.59: a) Variações sistemáticas diárias. b) Variações sistemáticas semanais. c) Variação anual. d) Representação em barras para a variação sistemática semanal.

A Figura 6.59 é semelhante às Figuras 6.57 e 6.58. Em a, PC₁ capta o perfil diário de emissões automotivas, PC₂ a fração de emissões de NMHC proveniente dos veículos, e PC₃ o perfil do O₃. Em b e d observa-se a queda das emissões automotoras nos finais de semana, e valores mais elevados para o ozônio no período. Em c nota-se a maior acumulação para os poluentes primários no inverno.

Tucker : CO, NO, NO₂ e O₃ (P. D. Pedro II – 1999)

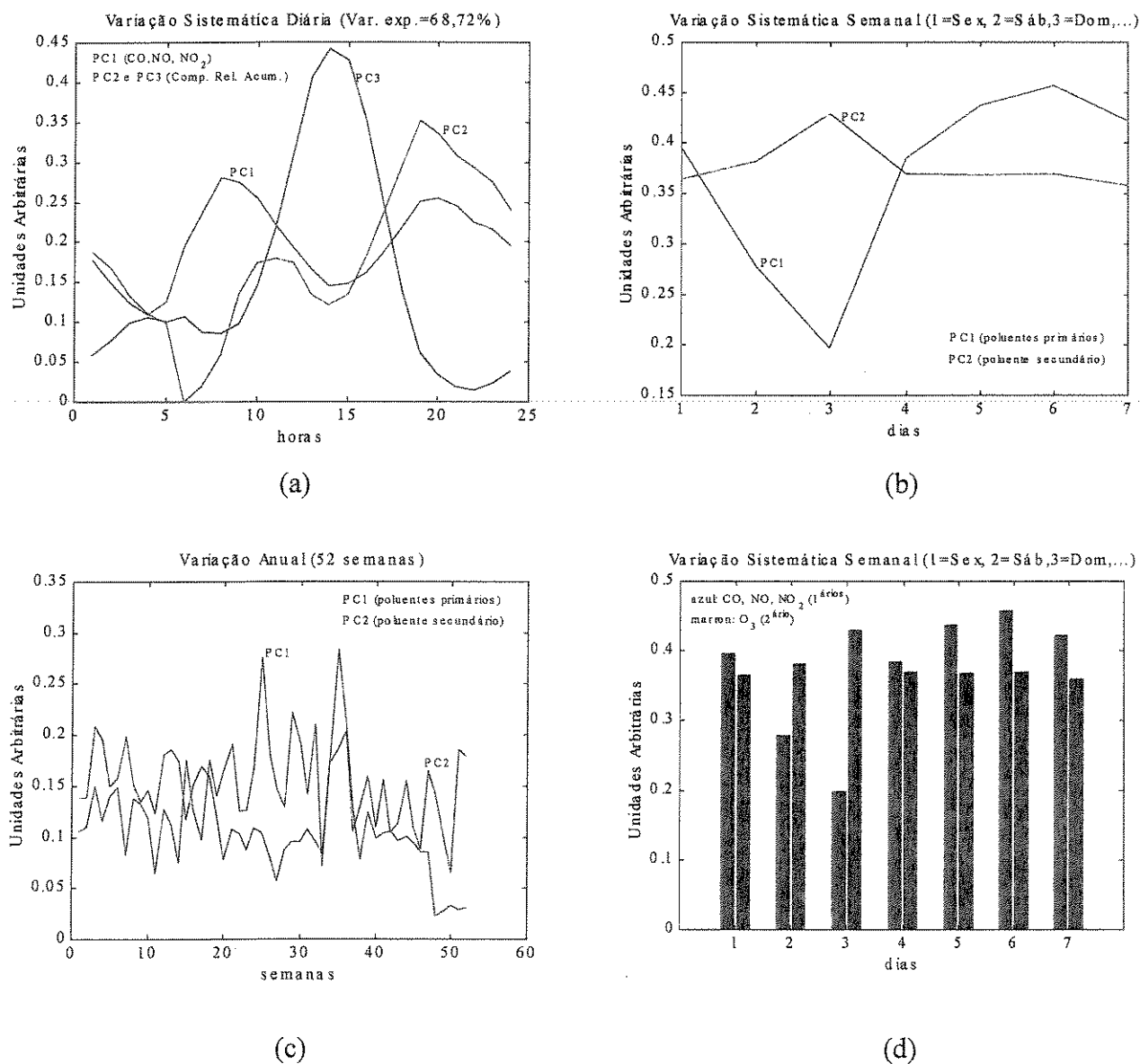


Fig.6.60: a)Variações sistemáticas diárias. b)Variações sistemáticas semanais. c)Variação anual. d)Representação em barras para a variação sistemática semanal.

A Figura 6.60a mostra, em PC₁ o perfil de emissão de fontes automotivas, e PC₂ e PC₃ são componentes relacionadas com a acumulação. Os gráficos b e d mostram a queda das emissões automotoras nos finais de semana e o aumento dos valores para o O₃ nesses dias. Em c notam-se os valores mais elevados (PC₁) no inverno, exatamente o período em que os valores para o O₃ são mais baixos.

A primeira grande diferença entre resultados fornecidos pelos modelos PARAFAC e Tucker já é visível quando se compara a quantidade de componentes que se extrai em cada modo. Enquanto no PARAFAC aquela quantidade em cada modo é sempre a mesma, no Tucker (para o caso acima), o modo A (variações sistemáticas diárias) foi construído com três componentes, e os modos B (variações sistemáticas semanais) e C (variação anual) possuem apenas duas componentes. A base dessa escolha está fundamentada, em parte, na teoria apresentada no Capítulo III, Seção 3.3. Entretanto, é na Subseção 3.3.3 do Capítulo III, onde se discute a construção da metodologia, que são postos os argumentos para orientar a escolha do número de componentes que cada modo deve possuir quando da utilização do modelo de Tucker.

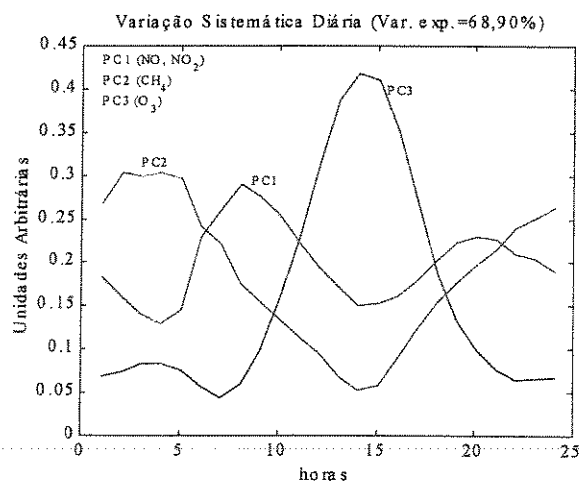
Cabe lembrar que, de acordo com aquela discussão, o modo A deveria possuir três componentes. Uma para cada fonte em potencial, sendo o terceiro para outras fontes (ou uma fonte, e uma combinação de reação e dispersão, denominada acumulação). O modo B deveria possuir um fator associado às emissões veiculares e, eventualmente um segundo associado a outras fontes. Porém, o modo C deveria possuir duas componentes as quais estão associadas com a variação sazonal.

O resultado mostrado na Figura 6.60 para o que foi denominado de ciclo básico do monóxido de carbono, passou a ser a base que orientou os resultados subseqüentes. Resultados preliminares com o modelo PARAFAC para o ciclo, indicaram que no modo B (perfil semanal), de fato, existe uma segunda componente a qual está associada com a formação do poluente secundário, O_3 (cujo comportamento difere daquele dos primários). Assim foi possível construir o núcleo, G , do modelo de Tucker de modo a se considerar três componentes no modo A, duas no modo B e duas no modo C (ver p370 e Apêndice A8.7) .

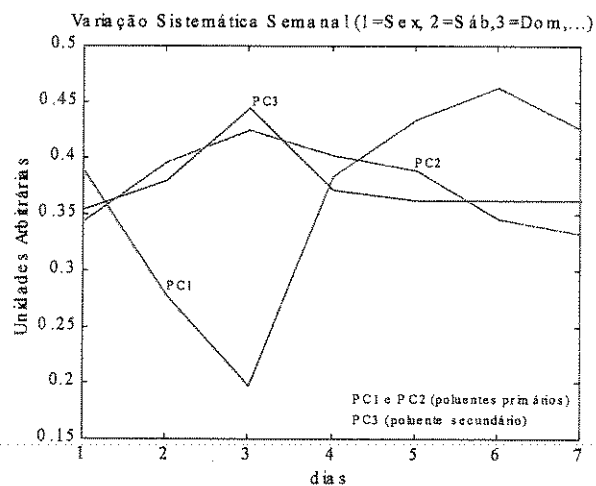
O resultado para o ciclo básico do CO é exaustivamente discutido adiante, à luz das reações químicas que ocorrem na atmosfera envolvendo os poluentes CO, NO, NO_2 e O_3 . Nessa discussão é incluída a importante participação da reação de formação de HONO e de sua decomposição fotoquímica, além de se apontar para outras possibilidades de reações químicas na atmosfera que justificam a existência daquele perfil para o ozônio.

A discussão em torno do ciclo básico do metano mostrou, como será visto nos resultados à frente, a necessidade de outra formulação para o núcleo, G , do modelo de Tucker, para que se pudesse explicar os diferentes perfis existentes (p370).

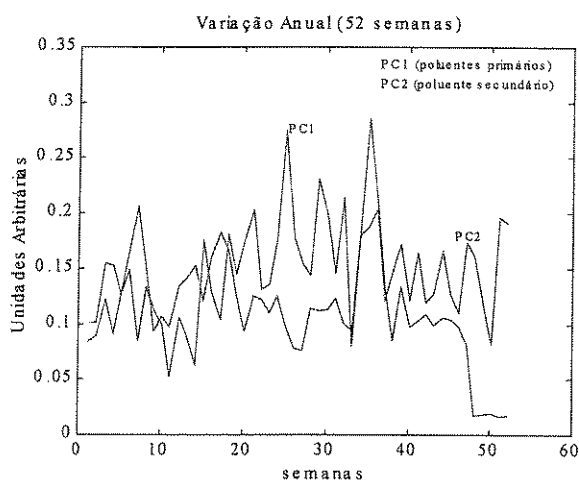
Tucker : CH₄, NO, NO₂ e O₃ (P. D. Pedro II – 1999)



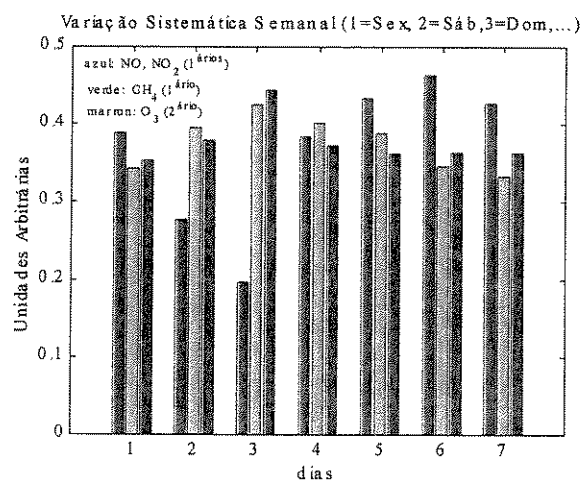
(a)



(b)



(c)



(d)

Fig.6.61: a)Variações sistemáticas diárias. b)Variações sistemáticas semanais. c)Variação anual. d)Representação em barras para a variação sistemática semanal.

A Figura 6.61a para o ciclo básico do CH₄ mostra, em PC₁, o perfil de emissão de fontes automotivas. Nesse caso essas emissões são de NO e NO₂ pois aqui leva-se em conta o comportamento do CH₄ sob concentrações elevadas daqueles poluentes, sem a contribuição do CO. PC₂ é o perfil de outras fontes primárias de CH₄ e PC₃ é o perfil para o O₃. Os gráficos b e d mostram a queda de NO e NO₂ nos finais de semana e o aumento dos valores para o O₃ nesse mesmo período. Mostram também um aumento menos pronunciado

de CH_4 no mesmo período. Em c notam-se os valores mais elevados (PC_1) no inverno, exatamente o período em que os valores para o O_3 são mais baixos.

Para o caso presente, nota-se uma diferença significativa quando a Figura 6.61a é comparada com a Figura 6.60a. Nos dois casos, PC_1 representa o perfil de emissão de fontes automotivas (NO e NO_2 no primeiro caso e CO , NO e NO_2 no segundo). Entretanto, enquanto PC_2 e PC_3 são componentes relacionadas com a acumulação no ciclo básico do monóxido de carbono (Figura 6.60a), PC_2 representa o perfil do CH_4 e PC_3 o do O_3 , no ciclo básico do metano (Figura 6.61a).

Quando se compara a Figura 6.61b (e obviamente também 6.61d) com a Figura 6.60b (e também 6.60d), a diferença é ainda mais significativa. Para o ciclo do CO , o perfil semanal apresenta apenas duas componentes, uma para representar as emissões veiculares e outra para representar o ozônio. Para o ciclo do CH_4 , entretanto, existem três componentes. Essa terceira componente está descrevendo as variações sistemáticas semanais para o metano. Acompanhado-a ao longo dos sete dias da semana percebe-se que os seus valores apresentam crescimento de sexta-feira para sábado, continuam a crescer, porém com menor intensidade ao “caminhar” para o domingo, passando, então a decrescer ao longo dos demais dias. O ciclo básico do metano exigiu, então, a construção de um novo núcleo, **G**, no modelo de Tucker para captar essas diferenças.

Não obstante valores mais elevados para o ozônio possam ser parcialmente explicados pelas reações de formação do HONO , os fatos acima mencionados parecem estar estritamente relacionados com aquele aumento para o ozônio. Ou seja, a acumulação desse poluente nos finais de semana parece dever-se à efeitos combinados dos dois ciclos, o do monóxido de carbono e o do metano. Adicionalmente, os hidrocarbonetos não metano, NMHC, também parecem ter participação nos efeitos observados para o comportamento do ozônio, como se pode observar no próximo resultado. Isto remete a discussão para o final das análises com o modelo de Tucker (ciclos do CO e dos NMHC).

É notável que, o núcleo, **G**, receba apenas dois tipos de Algarismos: *zero* e *um*. Suas posições estratégicas em **G** obedecem a: 1) n° de componentes em cada modo; 2) produto de Kronecker. Descobre-se, assim, a configuração a ser usada na função *FactorsG* e no núcleo, **G**, do modelo. Para levar a efeito as análises com o modelo de Tucker, foram necessárias apenas duas configurações (detalhes na p370 e Apêndice A8.7).

Tucker : NMHC, NO, NO₂ e O₃ (P. D. Pedro II – 1999)

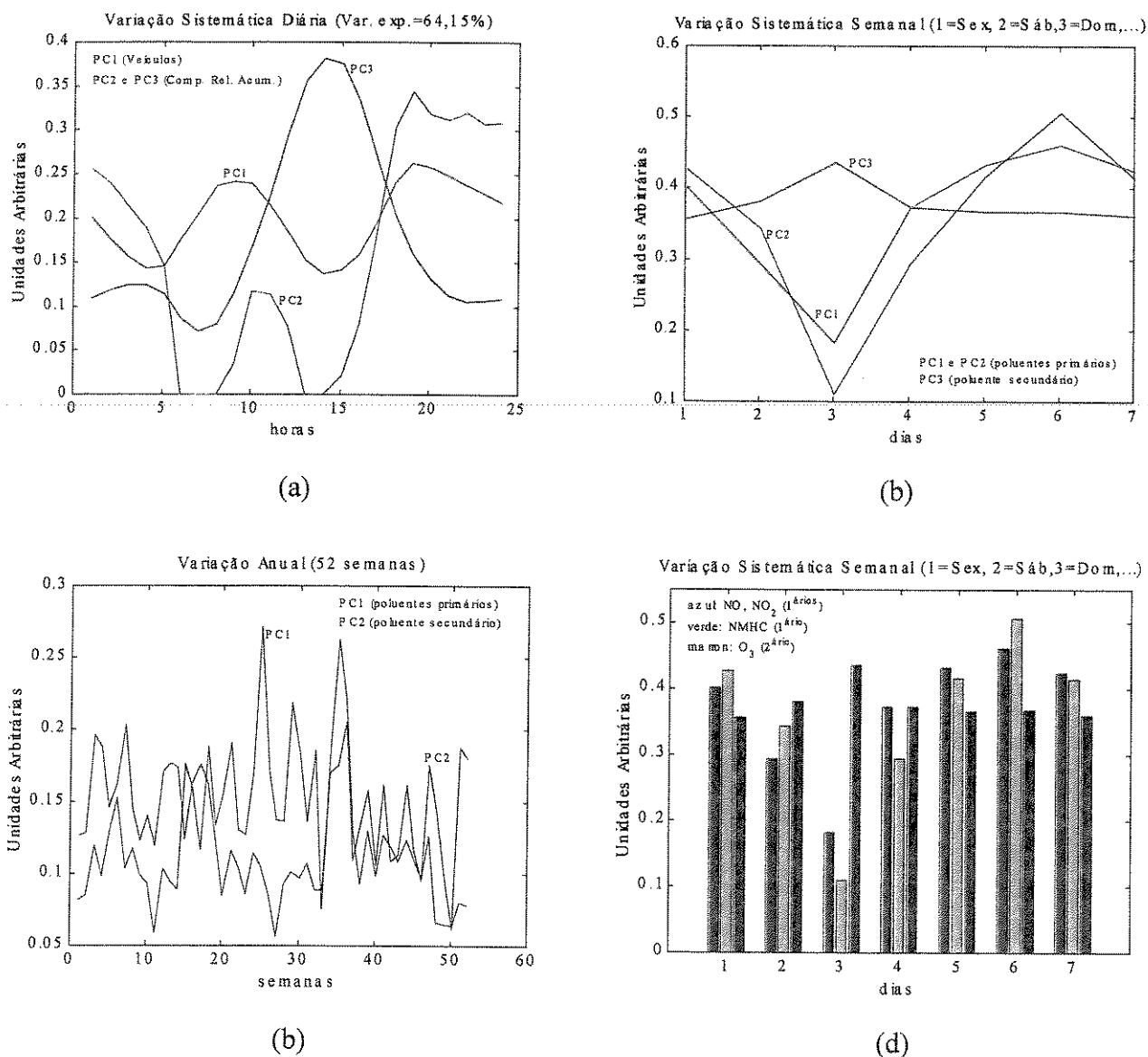
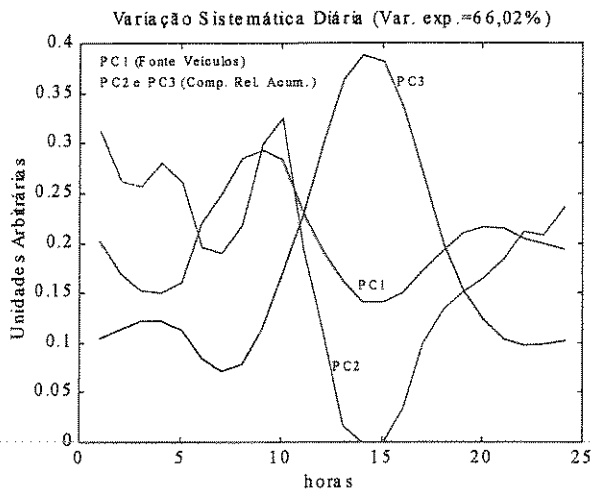


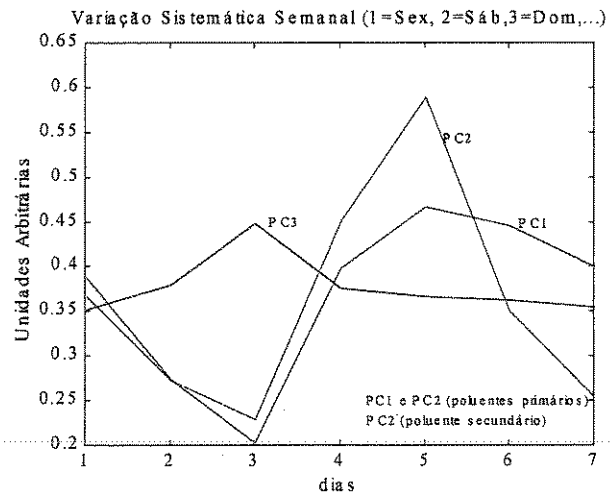
Fig.6.62: a)Variações sistemáticas diárias. b)Variações sistemáticas semanais. c)Variação anual. d)Representação em barras para a variação sistemática semanal.

Aqui o NMHC é introduzido em substituição ao CO ou (CH₄) em seu respectivo ciclo básico. A Figura 6.62a mostra, em PC₁, o perfil de emissão de fontes automotivas (NO, NO₂ e parte do NMHC). PC₂ e PC₃ são componentes relacionadas com a acumulação. Em 6.62b (e 6.62d) nota-se queda nas emissões veiculares (PC₁) nos finais de semana e valores elevados para o O₃ no mesmo período. Nota-se também queda em PC₂, e em 6.62c, valores mais elevados para os poluentes primários no inverno, em relação aos secundários.

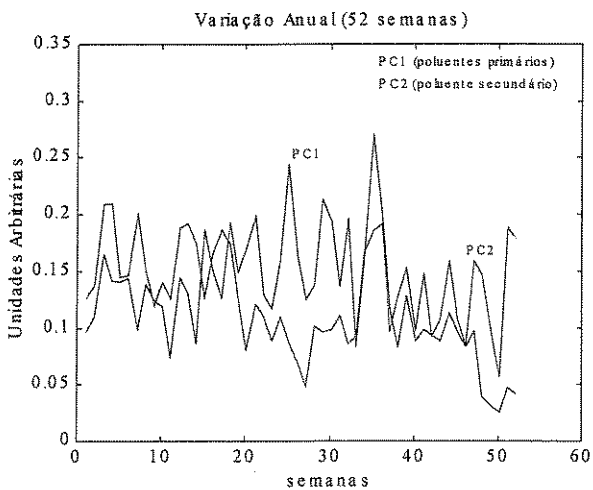
Tucker : SO₂, NO, NO₂ e O₃ (P. D. Pedro II – 1999)



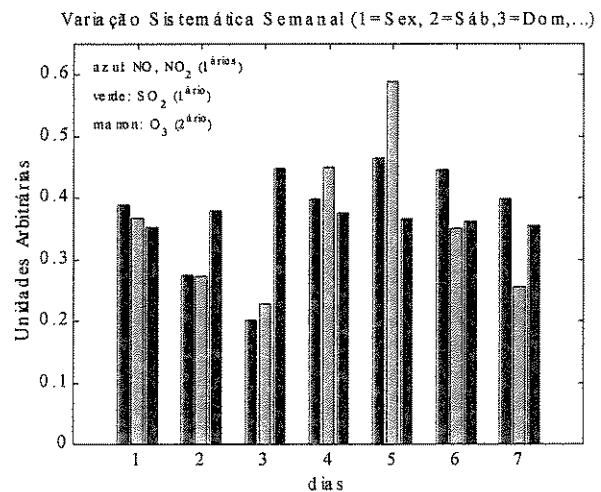
(a)



(b)



(c)



(d)

Fig.6.63: a)Variações sistemáticas diárias. b)Variações sistemáticas semanais. c)Variação anual. d)Representação em barras para a variação sistemática semanal.

Nesse caso, SO₂ é introduzido em substituição ao CO ou (CH₄) em seu respectivo ciclo básico. A Figura 6.63a mostra, em PC₁, o perfil de emissão de fontes automotivas (NO, NO₂ e de SO₂). PC₂ e PC₃ são componentes relacionadas com a acumulação. Em 6.63b (e 6.63d) nota-se queda nas emissões veiculares (PC₁) nos finais de semana e valores elevados para o O₃ no mesmo período. Nota-se também queda do poluente primário SO₂, conforme mostra PC₂. Em 6.63c observam-se valores mais elevados para os poluentes

primários no inverno, quando comparados aos secundários. Essa análise mostra outra diferença em relação ao ciclo básico do monóxido de carbono. Na Figura 6.63a, PC₁ representa o perfil de emissão das fontes automotivas, englobando aí o NO, NO₂ e o SO₂. As componentes relacionadas com a acumulação, PC₂ e PC₃, contém informações associadas, respectivamente, aos poluentes primários e ao O₃. Ou seja, associadas à reação e dispersão dos poluentes emitidos pelas fontes (ambas automotoras, mas uma delas não emite SO₂, ou ele não é mensurável), e à formação de ozônio.

Ao se analisar a Figura 6.63b (ou 6.63d), observa-se que também são três as componentes que descrevem as variações sistemáticas semanais. PC₁ e PC₂ possuem um perfil bastante parecido e poder-se-ia indagar, como no caso da Figura 6.62 para os NMHC, porque essas variações não são acomodadas em um mesmo fator. Não obstante possam, na verdade, não o são porque de tipos diferentes (veículos leves e pesados) da mesma fonte (automotiva), ou uma combinação desses tipos distintos da referida fonte. De fato, como se verá adiante, existe uma correlação entre as emissões de SO₂ e um tipo específico de NMHC.

Embora o modelo PARAFAC tenha captado apenas dois perfis para essa mesma análise (o modo B só modela, com significado físico, duas das três componentes), o fato de que o SO₂ pudesse possuir características de acumulação em escala regional e também local o colocaria, nesse aspecto, em condições de igualdade com os hidrocarbonetos não metano. Assim, na análise com o modelo de Tucker construiu-se o núcleo, *G*, tal que o modo B possuísse três fatores. Isto parece ter sentido pois o SO₂ possui características de comportamento na atmosfera (formações do ácido e sal correspondentes) que o distingue do CO e do CH₄.

Ainda na Figura 6.63b pode-se notar o valor bastante elevado (3ª feira) para o perfil semanal representado por PC₂. Este fato parece estar relacionado com várias reações da química atmosférica, especialmente aquelas que envolvem o NO, o NO₂ e o próprio O₃ entre outros. Esses aspectos são discutidos também mais adiante nesse capítulo.

A seguir são apresentados os resultados para cinco, seis e sete poluentes distribuídos em várias análises. Ainda aqui, tomou-se por base, como nas aplicações do PARAFAC, os ciclos básicos do monóxido de carbono e do metano, embora não se tenha perdido de vista os NMHC.

Tucker : CO, NO, NO₂, O₃ e CH₄ (P. D. Pedro II – 1999)

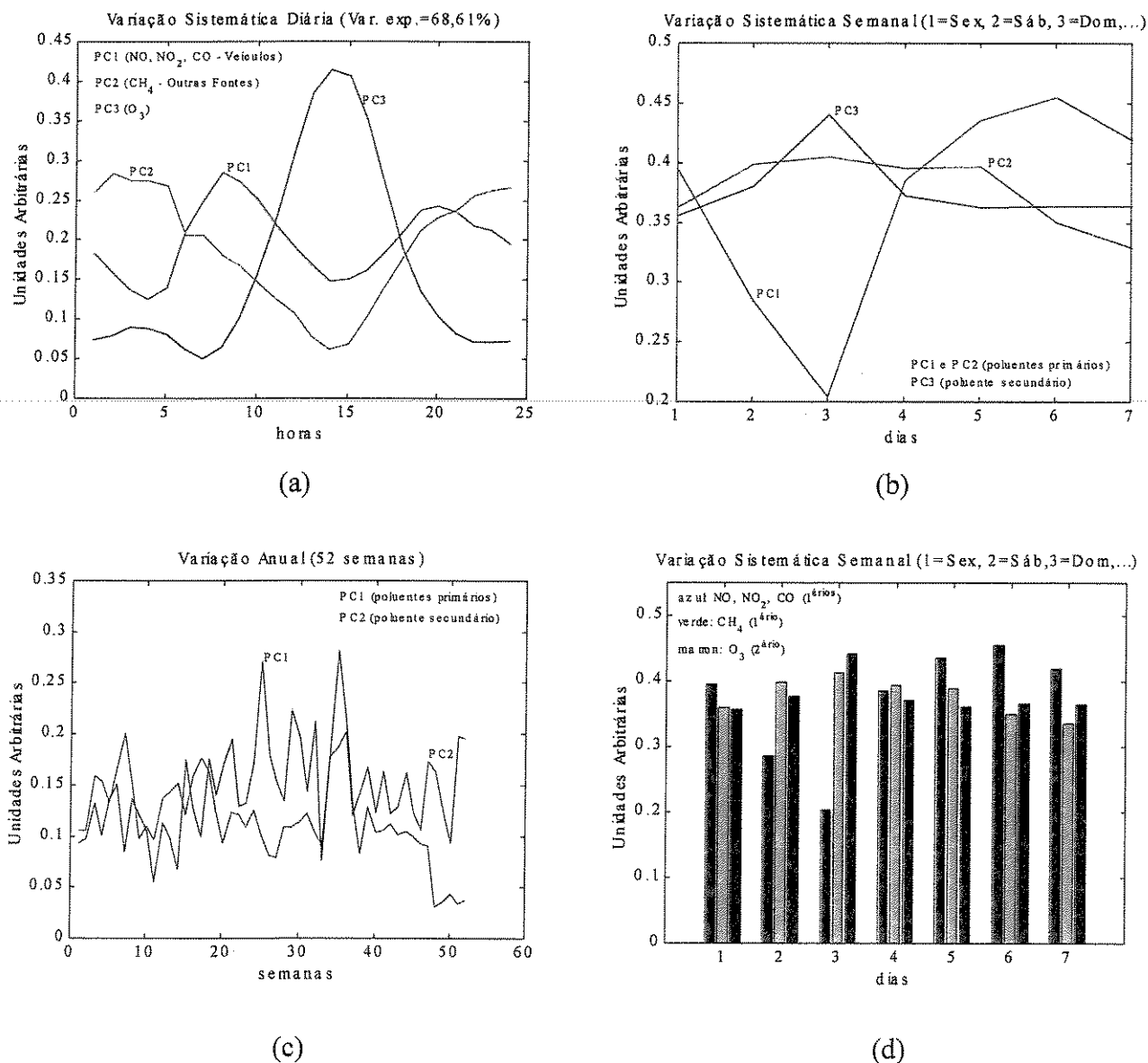


Fig.6.64: a)Variações sistemáticas diárias. b)Variações sistemáticas semanais. c)Variação anual. d)Representação em barras para a variação sistemática semanal.

Aqui aparece a influência inter-ciclos. CH₄ é introduzido no ciclo do CO ou vice-versa. A Figura 6.64a mostra, em PC₁, o perfil de emissão de fontes automotivas (CO, NO e NO₂), PC₂ é o perfil diário para o CH₄ e PC₃ é o perfil para o O₃. Em 6.64b (e 6.64d) nota-se queda nas emissões veiculares (PC₁) nos finais de semana, e valores mais elevados para o O₃ no mesmo período. Nota-se também que a curva para as variações do metano

apresenta crescimento de sexta-feira para sábado, um crescimento em menor extensão de sábado para domingo e, em seguida, queda até a quinta-feira. Em 6.64c observa-se maior acumulação para os poluentes primários no inverno, em relação aos secundários.

Nesse ponto cabe perguntar novamente porque o ozônio também não “cai” nos fins de semana se ele é um poluente secundário que se forma a partir de reações entre os poluentes primários, os quais têm queda de emissões nesse período? A seqüência de resultados até aqui mostra que, de fato, os valores mais elevados para o ozônio nos finais de semana provém de outras reações que não só não permitem que o O_3 acompanhe o perfil dos poluentes primários, como causam aumento de sua acumulação no referido período.

Deve-se chamar a atenção para a importância que possui a correta interpretação para as variações sistemáticas semanais na explicação do fenômeno. Ademais, ressalte-se também que o processo de construção da metodologia (e não apenas os métodos e modelos utilizados) é que permite inferências dessa natureza.

Mais uma vez, todos esses resultados são discutidos de forma mais aprofundada no próximo capítulo.

Tucker : CO, NO, NO₂, O₃ e NMHC (P. D. Pedro II – 1999)

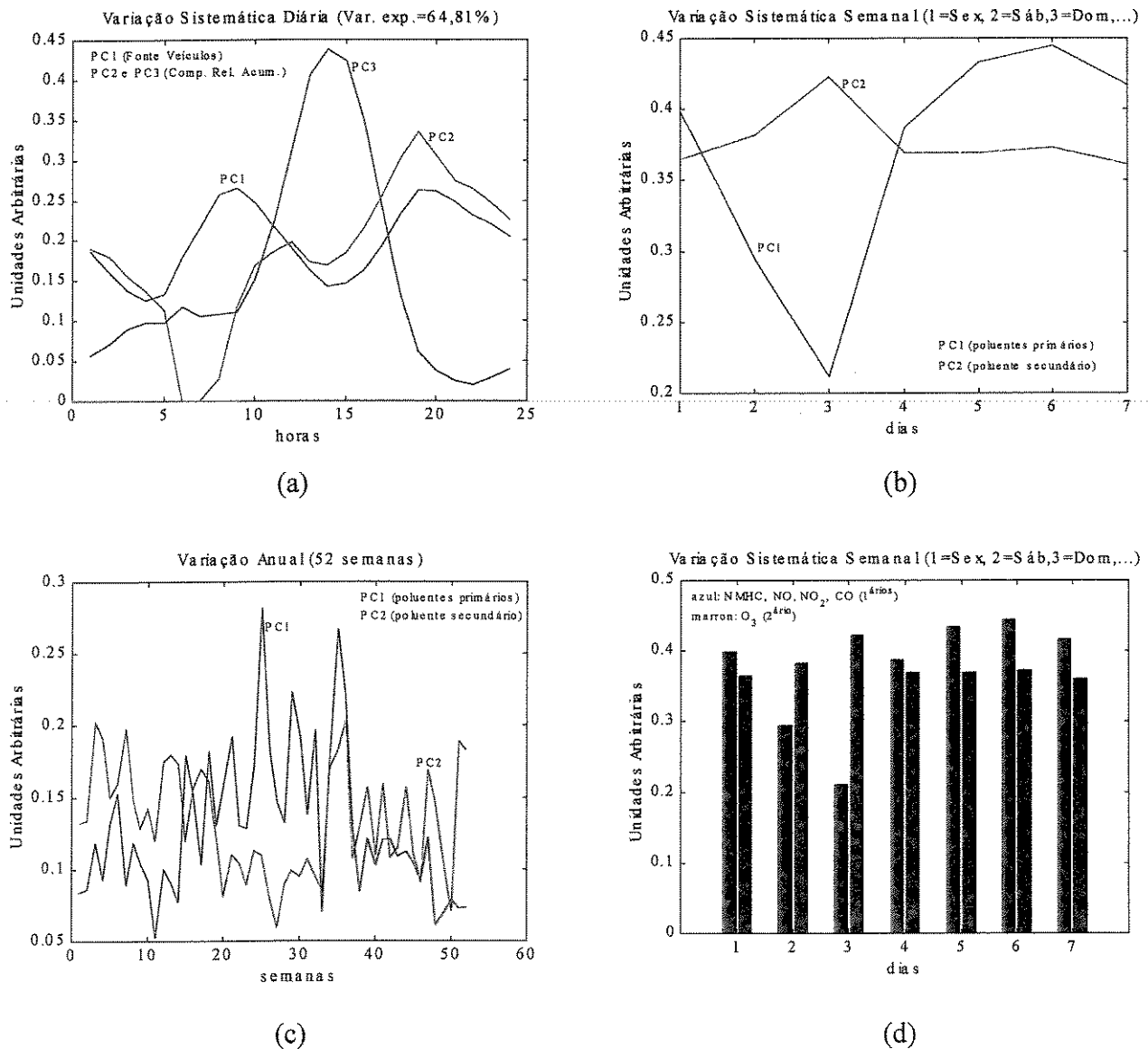
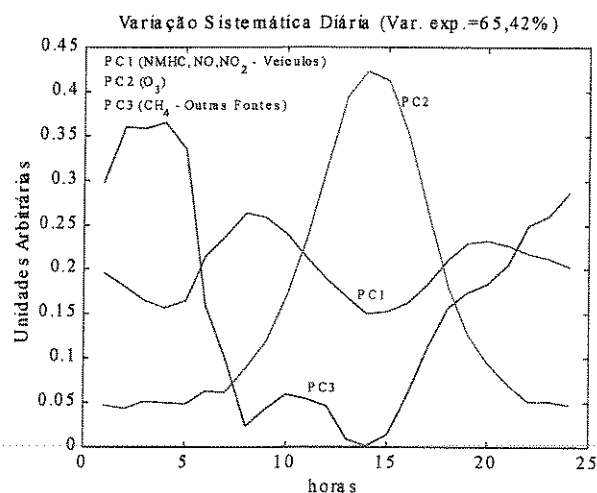


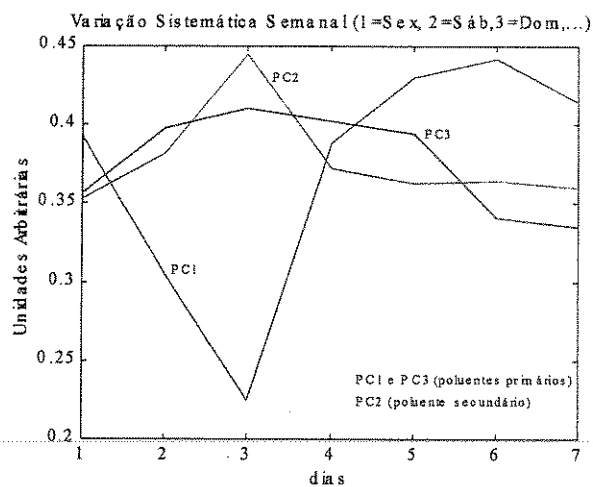
Fig.6.65: a)Variações sistemáticas diárias. b)Variações sistemáticas semanais. c)Variação anual. d)Representação em barras para a variação sistemática semanal.

Aqui é mostrado o ciclo do CO acrescido dos NMHC. A Figura 6.65a mostra, em PC₁, o perfil de emissão de fontes automotivas (CO, NO, NO₂ e NMHC) e PC₂ e PC₃ são componentes relacionadas com a acumulação. Em 6.65b (e 6.65d) nota-se queda nas emissões veiculares (PC₁) nos finais de semana e valores mais elevados para o O₃ no mesmo período. Em 6.65c notam-se as curvas características para os poluentes primários (com maior acumulação no inverno) e secundários.

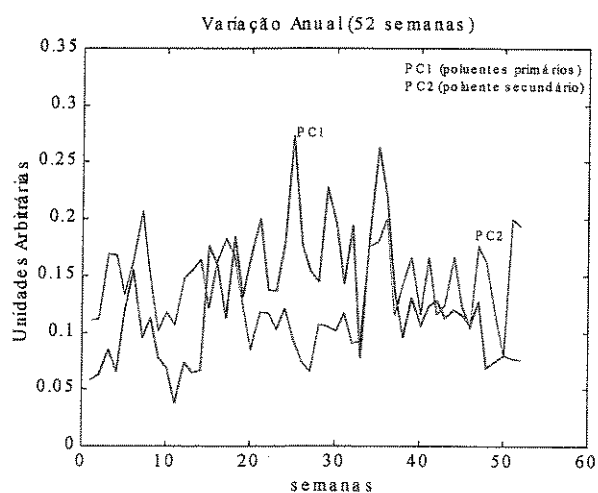
Tucker : CH₄, NO, NO₂, O₃ e NMHC (P. D. Pedro II – 1999)



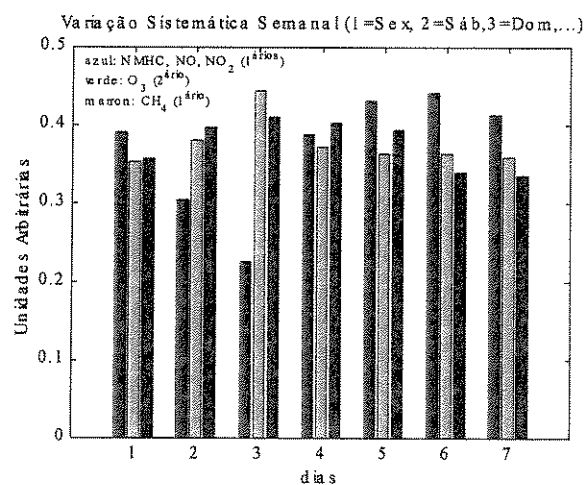
(a)



(b)



(c)



(d)

Fig.6.66: a)Variações sistemáticas diárias. b)Variações sistemáticas semanais. c)Variação anual. d)Representação em barras para a variação sistemática semanal.

A inclusão do NMHC no ciclo do CH₄ exige três componentes nos modos A e B. A Figura 6.66a mostra, em PC₁, o perfil de emissão de fontes automotivas (NO, NO₂ e NMHC), PC₂ é o perfil do O₃, e PC₃ o do CH₄. Em 6.66b (e 6.66d) nota-se queda nas emissões veiculares (PC₁) nos finais de semana, valores mais elevados para o O₃ no período, e comportamento semelhante ao da Figura 6.64b para o CH₄. Em 6.66c notam-se as curvas características para os poluentes primários e secundários.

Tucker : CO, NO, NO₂, O₃ e SO₂ (P. D. Pedro II – 1999)

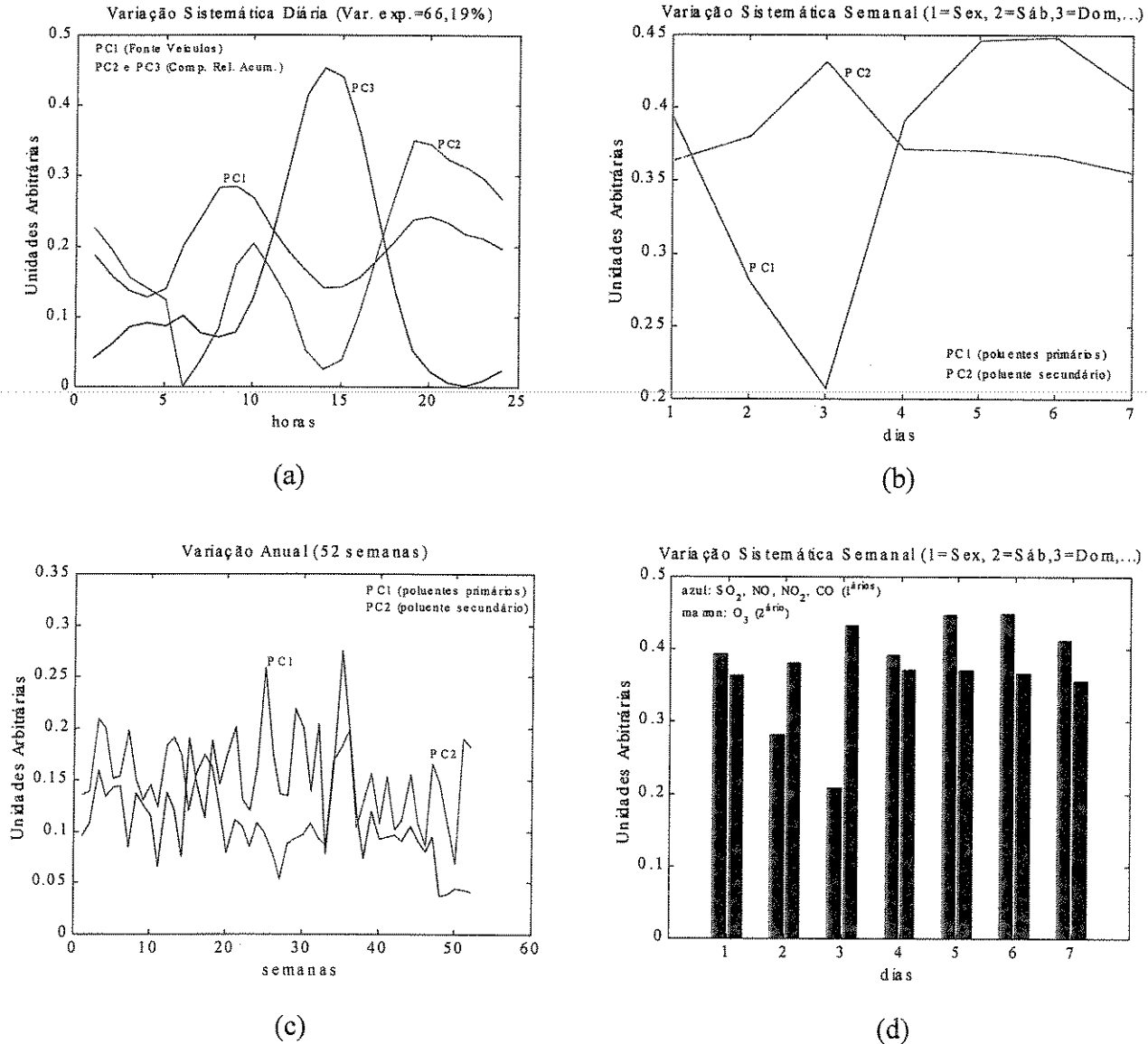
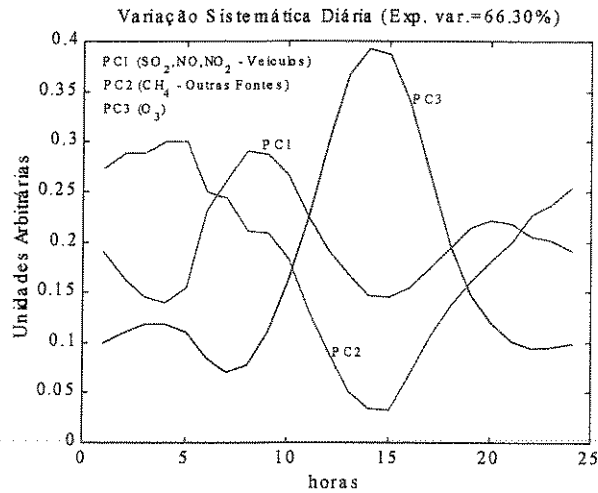


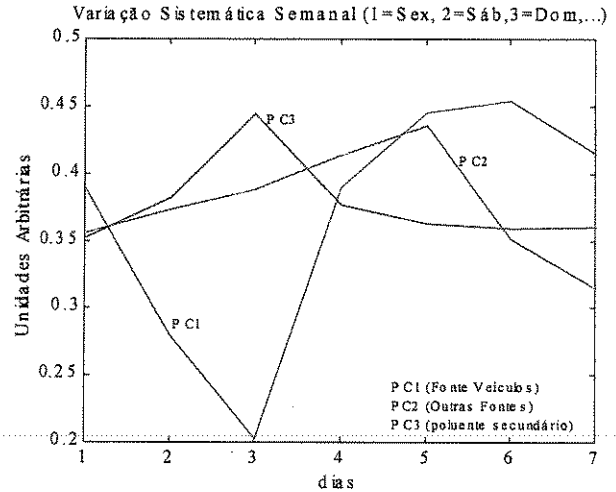
Fig.6.67: a)Variações sistemáticas diárias. b)Variações sistemáticas semanais. c)Variação anual. d)Representação em barras para a variação sistemática semanal.

Nesse caso, o SO₂ é incluído no ciclo do CO. A Figura 6.67a mostra, em PC₁, o perfil de emissão de fontes automotivas (CO, NO, NO₂ e SO₂), e PC₂ e PC₃ são componentes relacionadas com a acumulação. Em 6.67b (e 6.67d) nota-se queda nas emissões automotivas (PC₁) nos finais de semana e valores mais elevados para o O₃ no mesmo período. Em 6.67c notam-se as curvas características para os poluentes primários (com maior acumulação no inverno) e o poluente secundário.

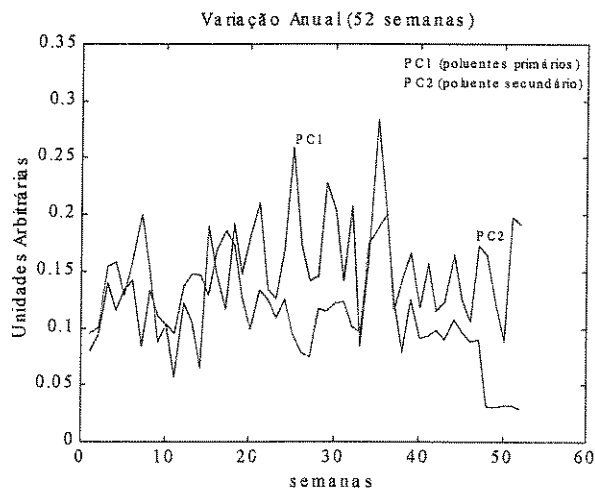
Tucker : CH₄, NO, NO₂, O₃ e SO₂ (P. D. Pedro II – 1999)



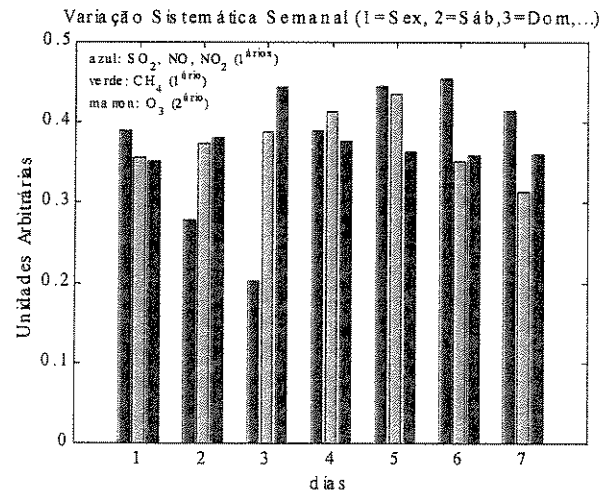
(a)



(b)



(b)



(d)

Fig.6.68: a)Variações sistemáticas diárias. b)Variações sistemáticas semanais. c)Variação anual. d)Representação em barras para a variação sistemática semanal.

Nesse caso, o SO₂ é incluído no ciclo do CH₄. A Figura 6.68a mostra, em PC₁, o perfil de emissão veicular (NO, NO₂ e SO₂). PC₂ e PC₃ mostram as variações sistemáticas, respectivamente, para o CH₄ e o O₃. Em 6.68b (e 6.68d) nota-se queda nas emissões automotoras (PC₁) nos finais de semana, valores mais elevados para o O₃ no período, e elevação de CH₄ até 3ª feira, seguida de queda. Em 6.68c notam-se as curvas características para os poluentes primários (com maior acumulação no inverno) e o poluente secundário.

Tucker : NMHC, NO, NO₂, O₃ e SO₂ (P. D. Pedro II – 1999)

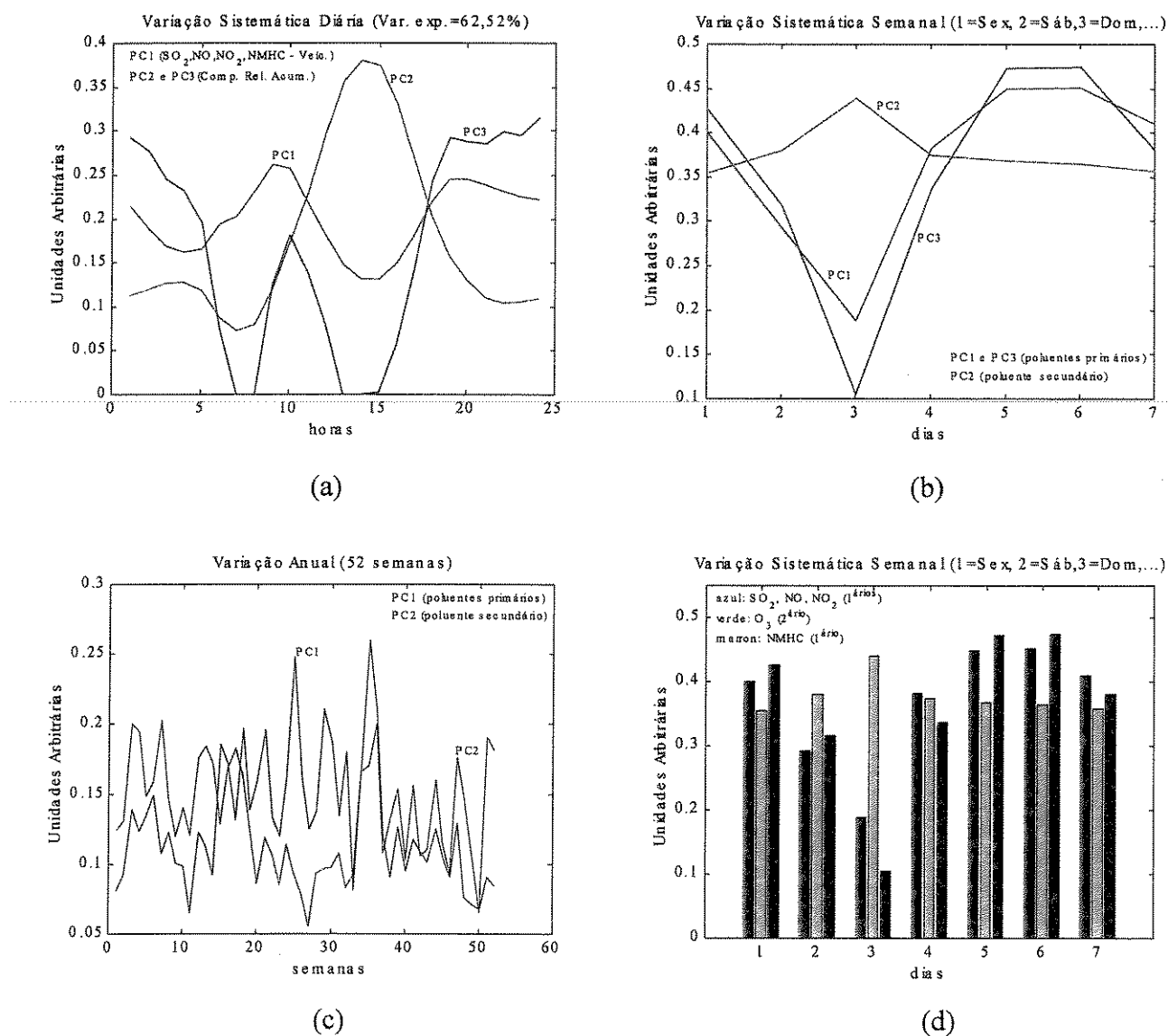
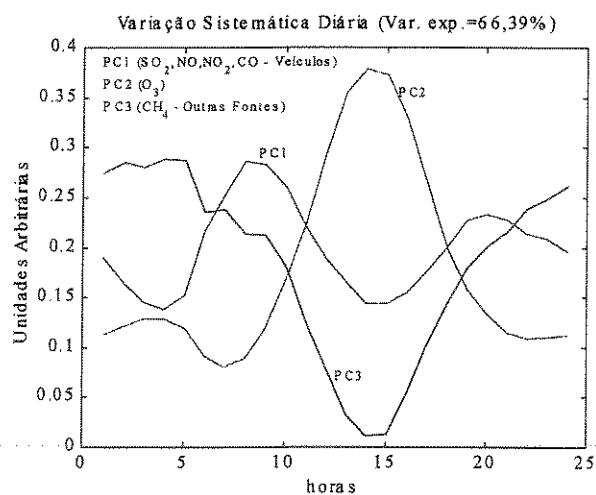


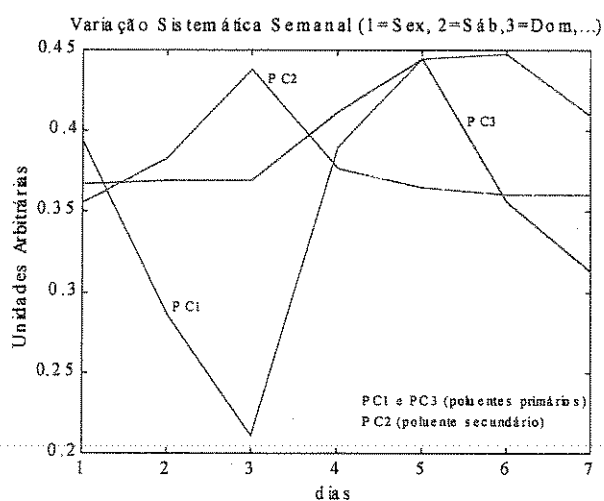
Fig.6.69: a)Variações sistemáticas diárias. b)Variações sistemáticas semanais. c)Variação anual. d)Representação em barras para a variação sistemática semanal.

A Figura 6.69a mostra, em PC₁, o perfil de emissão automotiva (NO, NO₂, SO₂ e parte do NMHC). PC₂ e PC₃ são componentes relacionadas com a acumulação. Em 6.69b (e 6.69d) nota-se queda nas emissões de veículos (PC₁) nos finais de semana, queda nos valores de NMHC, e valores mais elevados para o O₃ no período. Em 6.69c notam-se as curvas características para os poluentes primários (com maior acumulação no inverno) e para o poluente secundário.

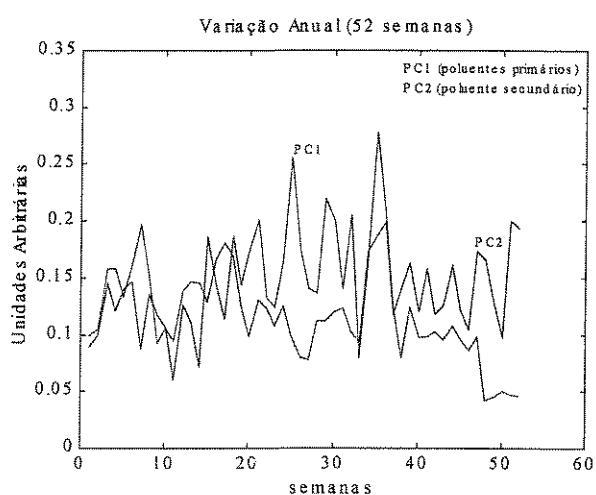
Tucker : CO, NO, NO₂, O₃, CH₄ e SO₂ (P. D. Pedro II – 1999)



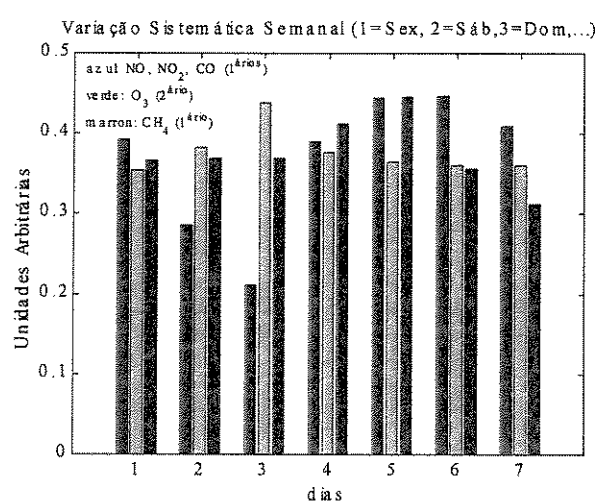
(a)



(b)



(c)



(d)

Fig.6.70: a)Variações sistemáticas diárias. b)Variações sistemáticas semanais. c)Variação anual. d)Representação em barras para a variação sistemática semanal.

A Figura 6.70a mostra, em PC₁, o perfil de emissão veicular (CO, NO, NO₂ e SO₂). PC₂ é o perfil para o O₃ e PC₃ para o CH₄. Em 6.70b (e 6.70d) nota-se queda nas emissões veiculares (PC₁) nos finais de semana, valores mais elevados para o O₃ no período, valores constantes para o CH₄ entre 6^afeira e domingo, elevação até 3^afeira e, em seguida, queda. Em 6.70c notam-se as curvas características para os poluentes primários (com maior acumulação no inverno) e para o poluente secundário.

Tucker : CO, NO, NO₂, O₃, SO₂ e NMHC (P. D. Pedro II – 1999)

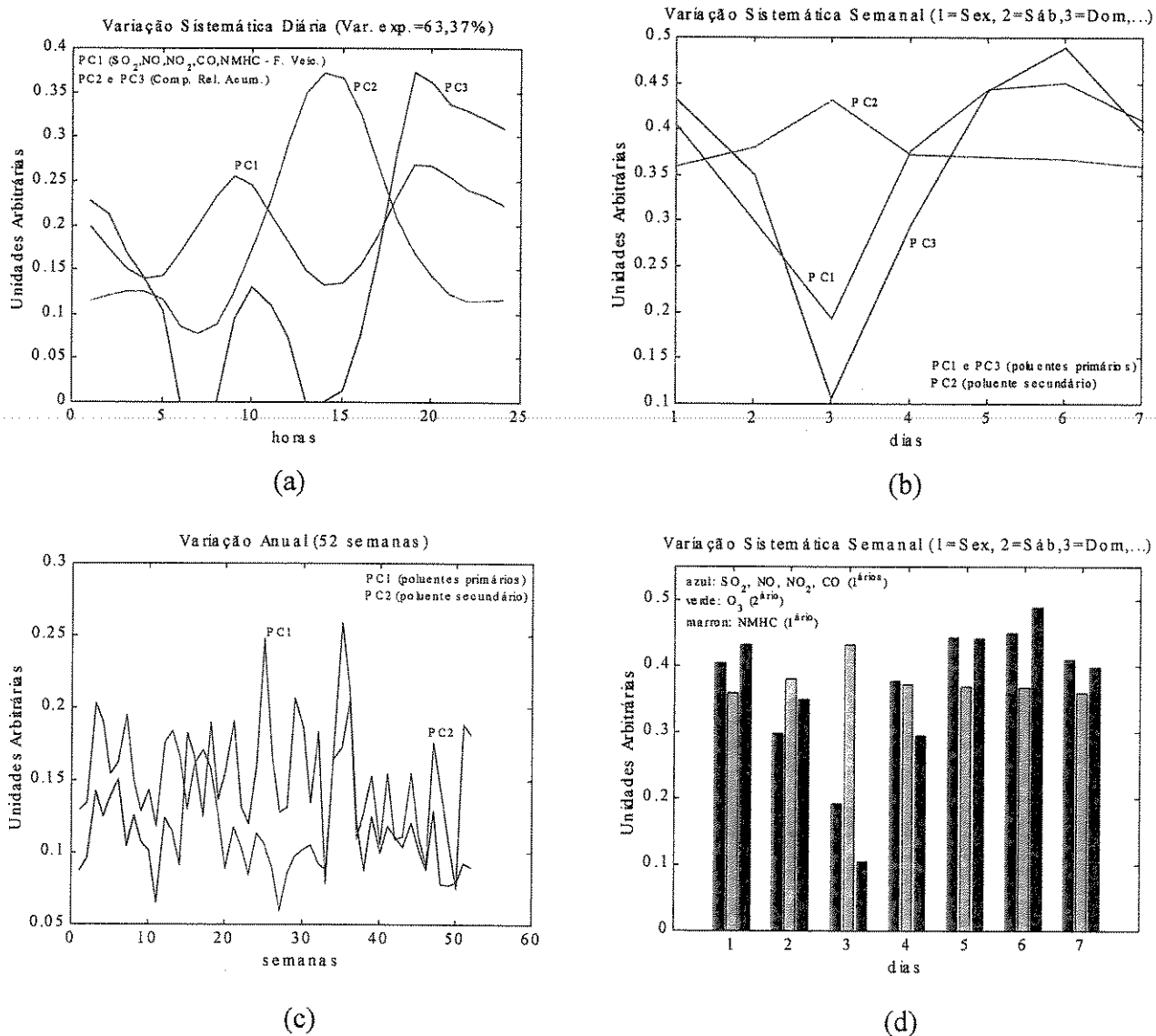
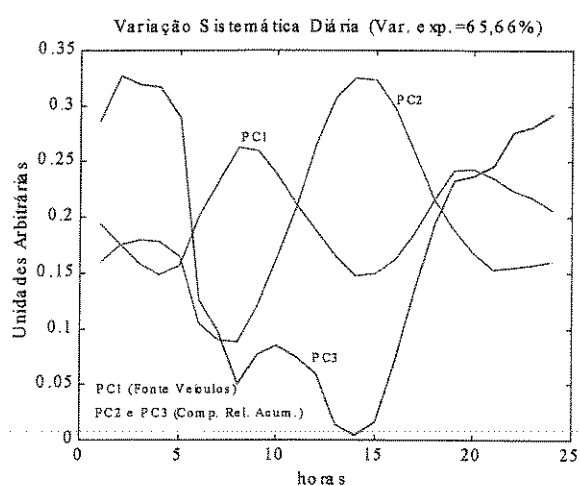


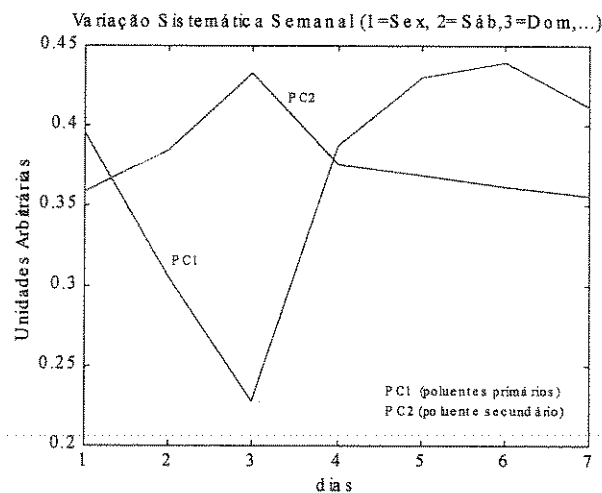
Fig.6.71: a)Variações sistemáticas diárias. b)Variações sistemáticas semanais. c)Variação anual. d)Representação em barras para a variação sistemática semanal.

A Figura 6.71a mostra, em PC₁, o perfil de emissão veicular (CO, NO, NO₂, SO₂ e parte do NMHC), e PC₂ e PC₃ são componentes relacionadas com a acumulação. Em 6.71b (e 6.71d) nota-se queda nas emissões veiculares (PC₁) nos finais de semana, queda no perfil dos NMHC, e valores mais elevados para o O₃ no período. Em 6.71c notam-se as curvas características para os poluentes primários (com maior acumulação no inverno), e para o poluente secundário com valores mais baixos nesse período.

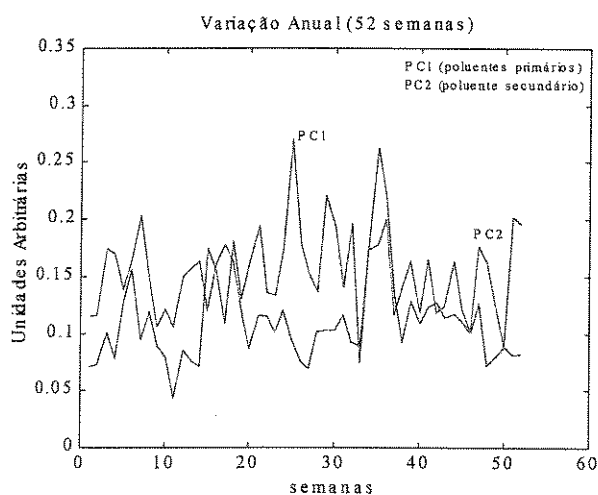
Tucker : CO, NO, NO₂, O₃, CH₄ e NMHC (P. D. Pedro II – 1999)



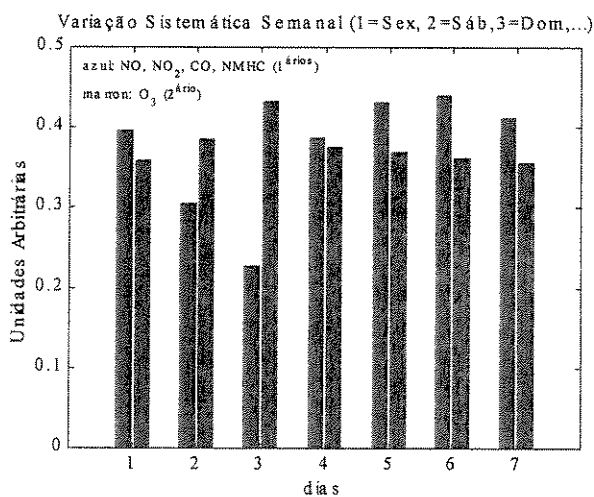
(a)



(b)



(c)



(d)

Fig.6.72: a) Variações sistemáticas diárias. b) Variações sistemáticas semanais. c) Variação anual. d) Representação em barras para a variação sistemática semanal.

A Figura 6.72 a mostra, em PC₁, o perfil de emissão das fontes automotivas. PC₂ e PC₃ são componentes relacionadas com a acumulação. Nos gráficos b e d notam-se, em azul, a queda nas emissões automotivas nos finais de semana, enquanto a outra curva mostra o aumento de O₃ no período. Em c (cor azul) pode-se observar a maior acumulação dos poluentes primários no inverno, e em verde, a curva para o poluente secundário.

Tucker : SO₂, NO, NO₂, O₃, CH₄ e NMHC (P. D. Pedro II – 1999)

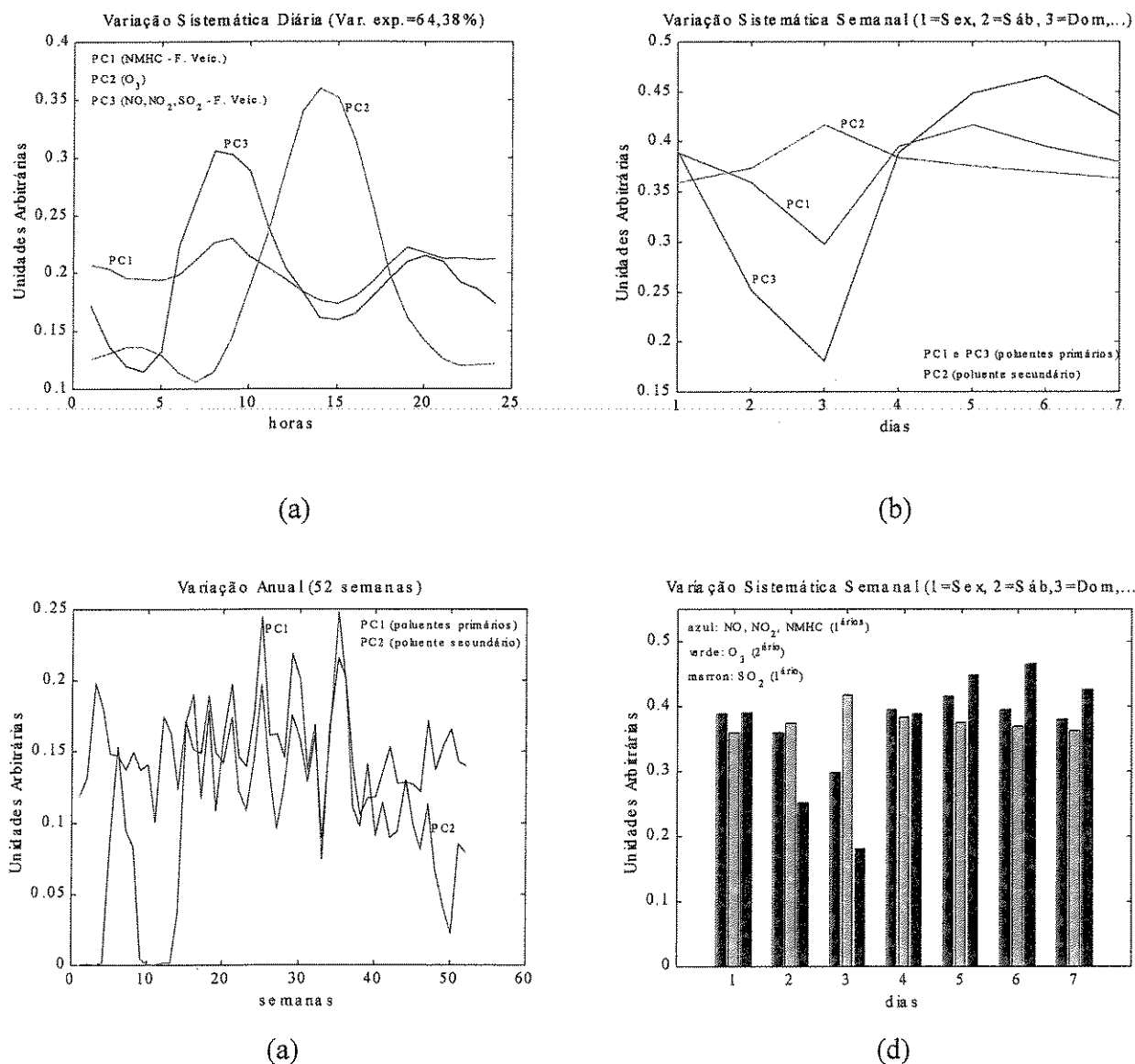
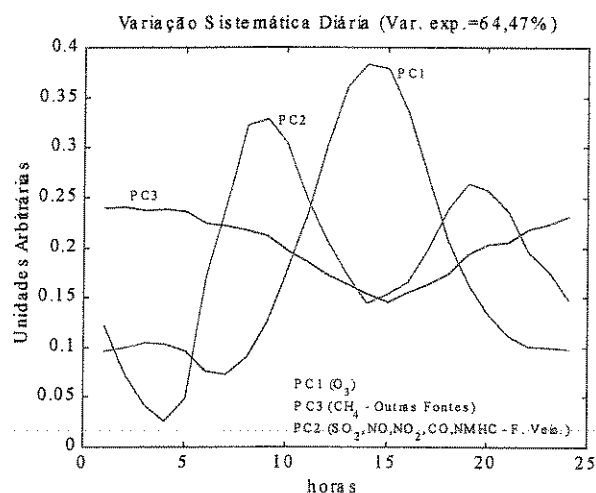


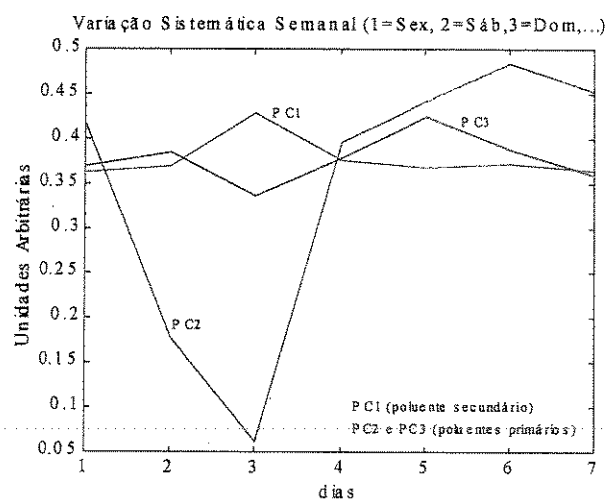
Fig.6.73: a) Variações sistemáticas diárias. b) Variações sistemáticas semanais. c) Variação anual. d) Representação em barras para a variação semanal.

A figura 6.73a mostra, em PC₁, o perfil dos NMHC, enquanto PC₃ capta, as emissões automotivas para os outros poluentes primários, e PC₂ mostra o perfil para o O₃. Em b e d observa-se a queda dos poluentes primários (PC₁ e PC₃) nos fins de semana, e valores mais elevados para o ozônio no período. Em c, a acumulação para os poluentes primários (curva azul) não é tão maior do que a do poluente secundário como nos demais casos.

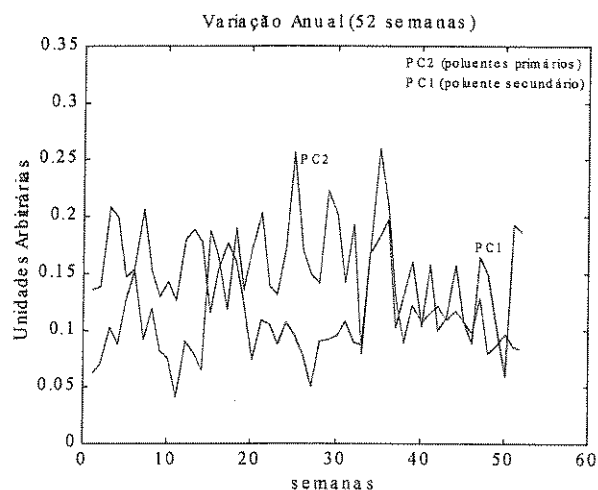
Tucker : CO, NO, NO₂, O₃, CH₄, SO₂ e NMHC (P. D. Pedro II – 1999)



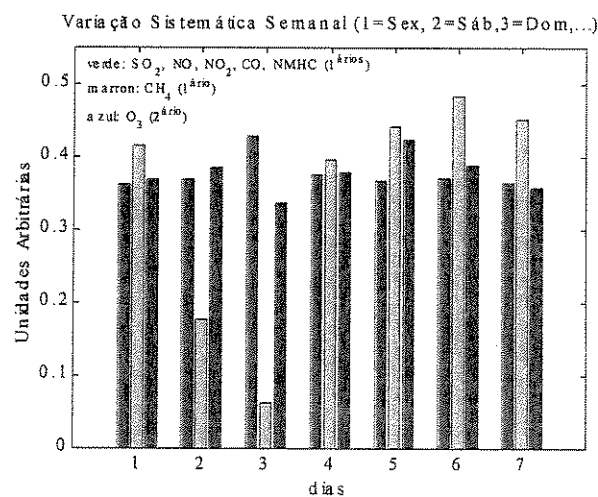
(a)



(b)



(c)



(d)

Fig.6.74: a) Variações sistemáticas diárias. b) Variações sistemáticas semanais. c) Variação anual. d) Representação em barras para a variação semanal.

A Figura 6.74a mostra, em PC₂, o perfil de emissões automotivas, em PC₁ o perfil para o O₃ e em PC₃ o perfil para o CH₄. Em b e d nota-se queda nas emissões de veículos nos finais de semana (PC₂), valores mais elevados para o O₃ no período, ligeira elevação para o CH₄ entre 6^a feira e Sábado, queda até Domingo, elevação até 3^a feira, em seguida

queda. Em 6.74c notam-se a curva (em verde) característica para os poluentes primários (com maior acumulação no inverno), e a curva para o poluente (em azul) secundário.

6.4.1 – Arranjos do Núcleo (“core”) no Modelo de Tucker

As análises por meio do modelo de Tucker apresentam um grau de dificuldade a mais em relação ao modelo PARAFAC devido à necessidade de utilização de um arranjo denominado núcleo que não se reduz, como no modelo PARAFAC, à “superdiagonal”, I_{sd} , como visto na parte teórica da tese que discute as estruturas multimodo.

Vale lembrar que o modelo PARAFAC exige o mesmo número de componentes em todos os modos. Isto implica necessariamente na extração de três componentes no modo C (variação anual). Nesse caso, a terceira componente irá descrever a variação anual do(s) poluente(s) que exerce(m) a menor influência no conjunto de dados. Por outro lado, é possível se assim desejado (ou necessário) o for, extrair um maior número de componentes no modo B quando da aplicação do modelo de Tucker. Entretanto, considerou-se nos casos sob investigação a variação sazonal dividida em dois períodos de maior influência dos efeitos climáticos e meteorológicos (o período de chuvas e o de seca).

Assim, nas análises com o modelo de Tucker, mostradas nas Figuras 6.60 a 6.74 (às páginas 334 a 352), os núcleos dos arranjos foram construídos em função da necessidade de se explicar um dado fenômeno. Para a variação anual restringiu-se o modo que contém as componentes que a descreve (o modo C) à duas, as quais estão relacionadas com a variação sazonal. Desse modo, uma das componentes no modo C refere-se ao período de chuvas (outubro a abril) e a outra, ao período de inverno (ou de seca) que ocorre entre maio e setembro (ver relatório CETESB 1998/1999 para confirmação desses períodos).

Manteve-se fixo o modo A em três componentes. Estas estão relacionadas com as variações sistemáticas das principais fontes de emissão de poluentes (os primários) na atmosfera da RMSP, e com aquelas do principal poluente secundário formado, o ozônio.

As quedas bruscas nas emissões de poluentes de fonte automotiva nos finais de semana são um divisor importante na identificação e distinção desse tipo de emissão daquelas originárias de outras fontes. Isto se deve ao fato de que tais quedas ocorrem sistematicamente ao longo do ano sendo, portanto, de importância fundamental para a caracterização e associação dos perfis às fontes. São essas variações portanto que permitem inferir sobre as possíveis reações que ocorrem na atmosfera e, a partir daí, sugerir possíveis mecanismos para explicar a acumulação de ozônio e suas correlações com os poluentes primários (Rotinas incluindo formulações para o núcleo nos Apêndices A8.6 e A8.7).

CAPÍTULO VII

DISCUSSÃO

7.1 – Breve Retropecto

Após ultrapassar a árdua Seção 6.4, chega-se, finalmente, à parte que trata da discussão dos resultados obtidos nessa tese. Considerou-se importante apresentar a seção precedente daquela forma para que se pudesse avaliar o comportamento da atmosfera da RMSP em sua plenitude, sob o ponto de vista das variáveis estudadas.

A imensa quantidade de “respostas” expressas graficamente nas seções 6.3 e 6.4 exigiu que a discussão sobre os principais resultados fosse organizada em um capítulo à parte, o Capítulo VII. Neste, faz-se a discussão tendo em mente os principais resultados obtidos por meio da Análise de Componentes Principais, do modelo PARAFAC e do modelo de Tucker. Tal discussão tomou por base o ciclo básico do monóxido de carbono, o qual considera-se constituído pelas variáveis CO, NO, NO₂ e O₃, do ciclo básico do metano (CH₄, NO, NO₂ e O₃) e do recém criado ciclo básico dos hidrocarbonetos não metano, no qual NMHC substitui o CO ou o CH₄. Depois estudou-se a interação entre os ciclos e, finalmente, a eles, foram agregadas outras variáveis.

As análises PCA com as variáveis tomadas individualmente mostram em todos os casos, um perfil que pode ser associado às emissões provenientes da frota de veículos. Excetuam-se as variáveis temperatura e umidade, e o metano cujo perfil mostra que o gás é originário principalmente de outras fontes.

Todas as variáveis cujos resultados apresentaram um perfil diário com dois picos, um pela manhã em torno das 8-9 horas e outro à noite aproximadamente às 20 horas, estão

em estrita concordância com o perfil de tráfego de veículos na RMSP, como mostra um gráfico da densidade da densidade de tráfego de veículos nas principais áreas da RMSP, construído pela Companhia de Engenharia de Tráfego de São Paulo (CET, 2000). Nele verifica-se a existência de dois picos de movimentação de veículos nessas áreas, os quais são muito próximos daqueles apresentados pelos poluentes (Figura 7.1).

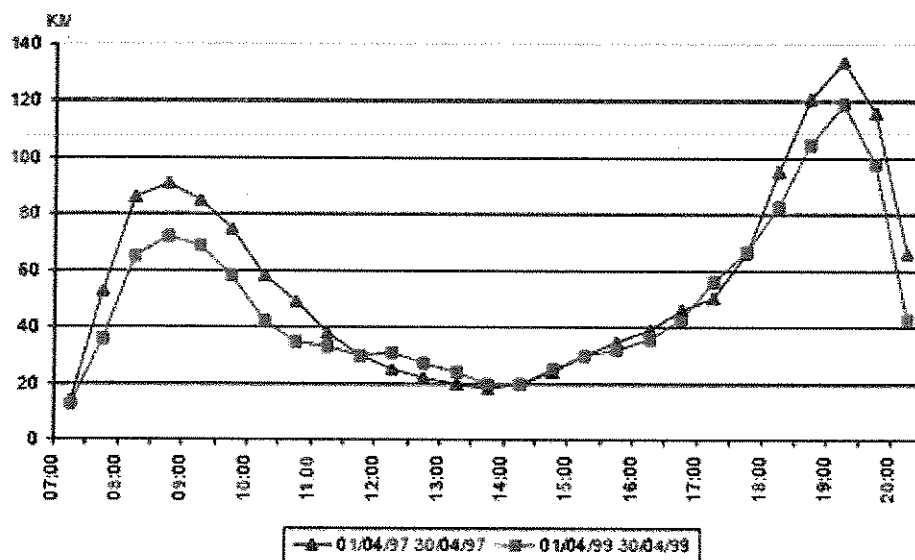


Fig.7.1: Densidade de tráfego para algumas áreas da RMSP. A abscissa varia no intervalo de 7:00 h às 20:00 h. O período coberto encontra-se entre 01/04 e 30/04, em vermelho para 1997, e verde para 1999.

Por outro lado, resultados de Baldasano *et alli* (1998), Statheropoulos *et alli* (1998), Alsberg (1989), Costa e Baldasano (1996), também indicaram as emissões automotivas como a principal fonte de poluição em ambientes urbanos.

Os resultados para o dióxido de enxofre foram parcialmente discutidos na seção 6.3. Para o ano de 1999, observou-se uma diferença em termos de fontes de emissão em relação aos poluentes CO, NO e NO₂. Os resultados mostram a existência de uma fonte adicional com uma característica bastante peculiar. Durante um determinado período de 1999,

verificou-se, à 1:00 h da manhã, a ocorrência de valores bem mais elevados do que para os outros horários. Isto foi verificado nas Estações de Medição Parque D. Pedro II, São Caetano do Sul e Ibirapuera, sendo tal efeito menos pronunciado e por um período menor na primeira localidade, e muito mais pronunciado e por um período maior na terceira.

Cabe resgatar (Figura 7.2), antes de passar à discussão desses casos, os gráficos contendo os dados originais para que se possa realçar o alcance das correções realizadas em cada matriz. Tal figura deve ser comparada com as Figuras 7.3, 7.4 e 7.5 adiante.

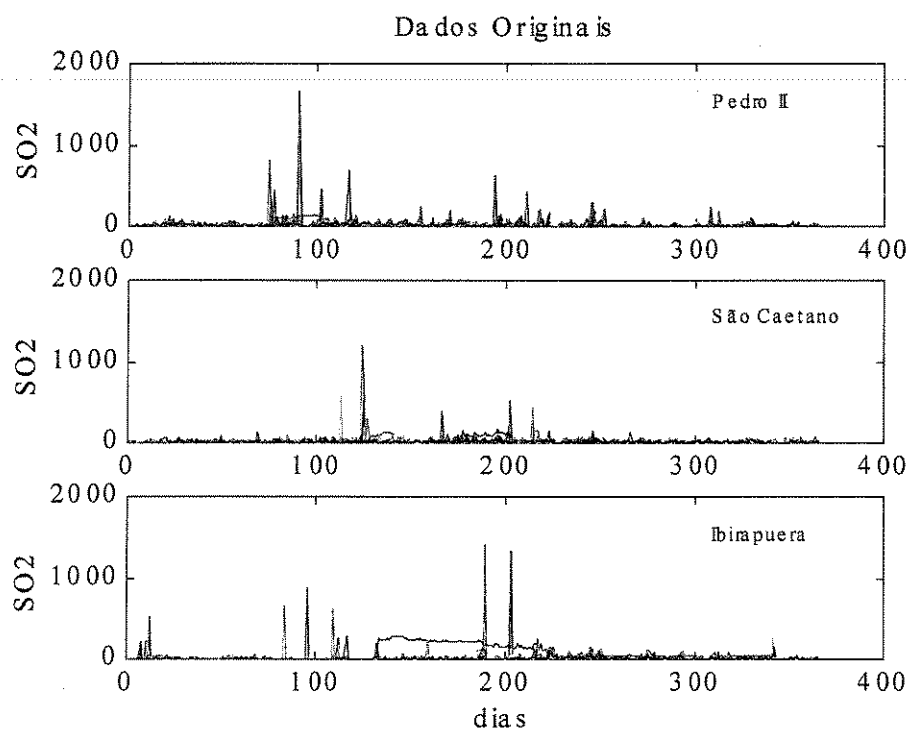


Fig.7.2: Dados originais para as Estações de Medição do P. D. Pedro II, São Caetano do Sul e Ibirapuera, para o poluente SO_2 .

Na Figura 7.2 é possível identificar (curva em azul) os períodos nos quais ocorrem as emissões “descoladas” dos outros valores, para cada uma das estações mencionadas. Após as devidas correções dos picos, os dados são ajustados para as duas situações (uma ou duas fontes), de acordo com a discussão iniciada na Seção 6.3.

Para a estação Pedro II os efeitos acima foram mostrados nos gráficos de pesos por meio das Figuras 6.4a e 6.4b (p245), considerando-se no primeiro caso, uma fonte de

emissão, e no segundo, duas fontes. Para essa localidade, os dados originais corrigidos são mostrados na Figura 7.3 adiante. Nela observa-se a curva para os dados coletados à 1:00 h em determinado período do ano, totalmente isolada do restante dos dados. Para o Parque D. Pedro II o período afetado possui 27 dias (76° ao 102° dias do ano), com início em 17/03/1999 e estendendo-se até 12/04/1999. Para o presente caso, poder-se-ia explicar o desaparecimento do “pico pontual” (um valor elevado apenas naquele horário – no caso, exatamente à 1:00 h) por meio de vários processos diferentes, em função do período no qual essas emissões ocorrem.

Na Figura 7.2 observa-se, em cada caso (Pedro II, São Caetano e Ibirapuera), picos isolados. Estes sofreram correções para dados anômalos e também correções adicionais para variações súbitas. Tais procedimentos resultaram nas Figuras 7.3 a 7.5 (dados originais corrigidos) apresentados à frente.

Ainda na Figura 7.2, para Pedro II e São Caetano, há certa dificuldade para se observar a curva azul referida acima, a qual é mais visível no Ibirapuera. Isto não deve se constituir em problema, aliás, o objetivo a alcançar com a apresentação da Figura 7.2 é o de mostrar a influência que os dados anômalos possuem no conjunto de dados. Após a correção destes, o espectro adquire características mais adequadas à realidade reinante nas localidades sob investigação como mostram as Figuras 7.3 a 7.5.

Deve ser observado que as Figuras 7.3 a 7.5 referem-se aos dados da Figura 7.2 com as correções sobre os picos. Entretanto, considera-se, para cada localidade dois casos: a) uma fonte (sem a influência do período afetado), e b) duas fontes (com a influência do período afetado).

Quando houver referência a uma fonte isto deverá ser interpretado como dados nos quais foram realizadas correções sobre os valores tomados à 1:00 h no período afetado, isto é, apenas nos pontos nos quais os valores para 1:00 h apresentam discrepância em relação ao conjunto de dados. Obviamente, após correções prévias para dados anômalos (e variações súbitas). Por outro lado, menção à duas fontes deve ser interpretado como dados que sofreram correções apenas para dados anômalos (e variações súbitas).

As Figuras 7.3 a 7.5 a seguir consolidam as explanações acima.

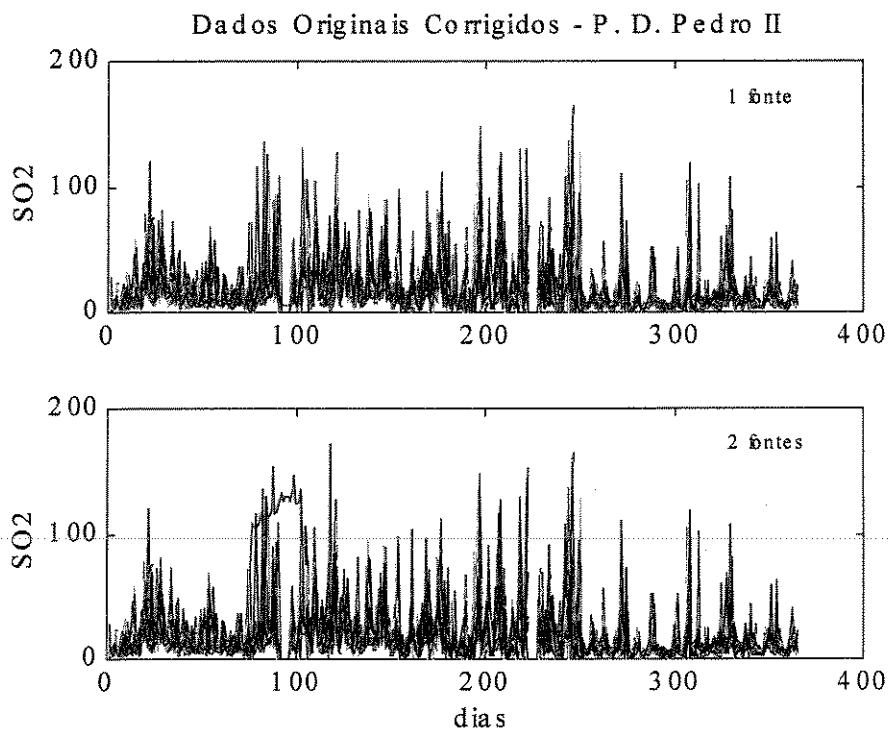


Fig.7.3: Dados originais corrigidos para a Estação de Medição Pedro II. Acima o espectro está livre da influência dos dados coletados à 1:00 h, e abaixo é visível aquela influência.

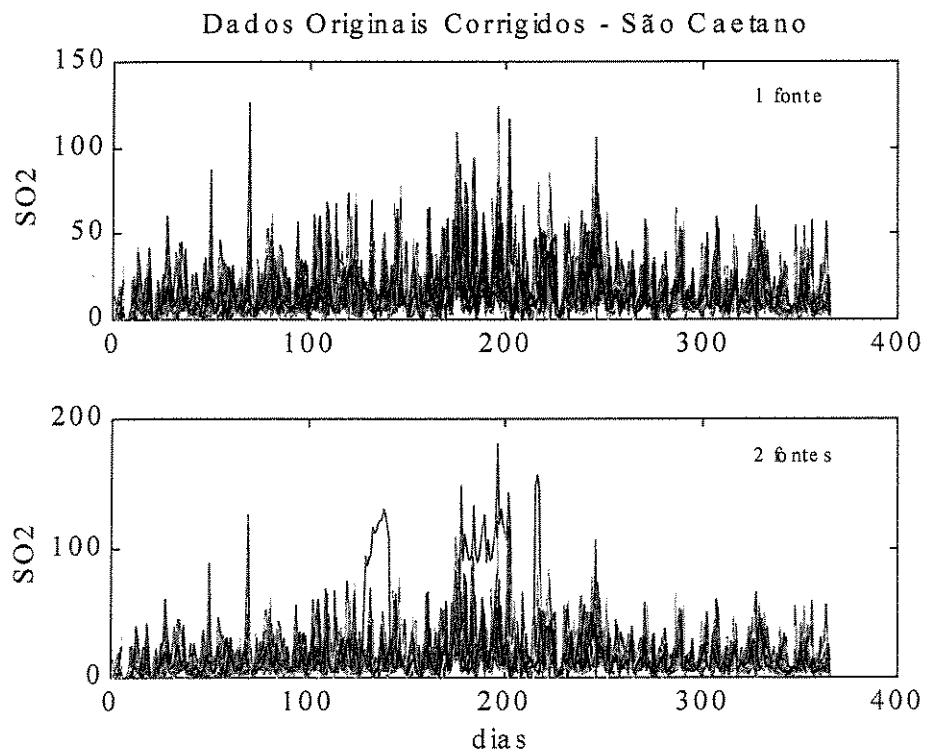


Fig.7.4: Dados originais corrigidos (Estação de Medição S. Caetano). Acima a ausência da influência dos dados coletados à 1:00 h, e abaixo é visível e mais intensa aquela influência.

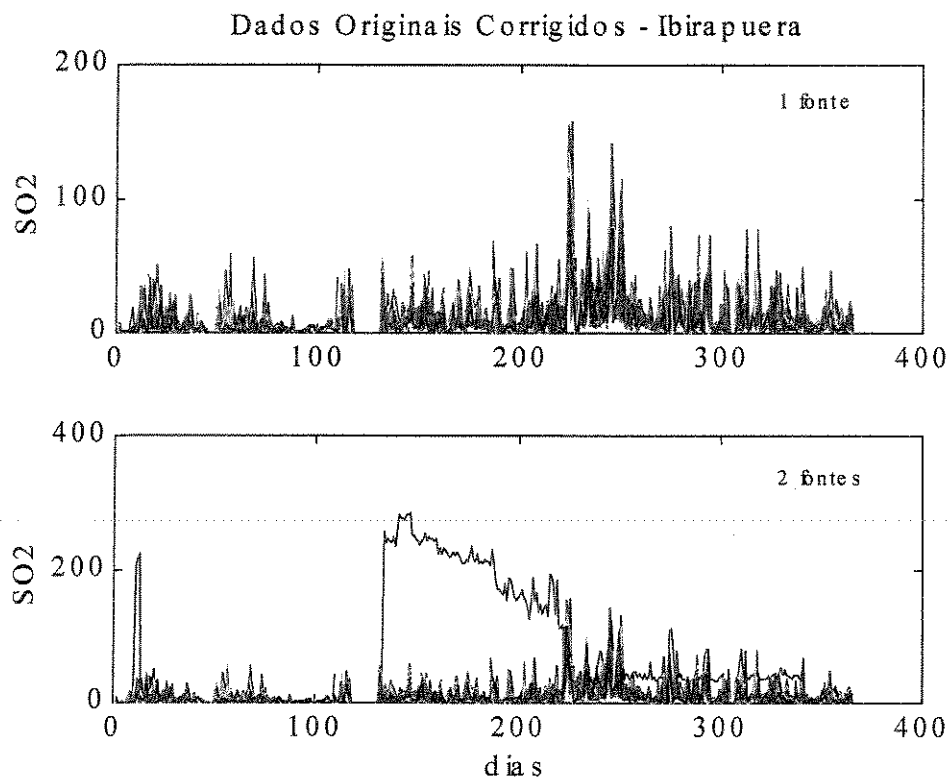


Fig.7.5: Dados originais corrigidos (Estação de Medição S. Caetano). Acima a ausência da influência dos dados coletados à 1:00 h, e abaixo é muito mais intensa aquela influência.

Uma comparação entre as Figuras 7.3, 7.4 e 7.5 mostra o crescimento da extensão do problema quando se “caminha” da primeira localidade (Pedro II) para a terceira (Ibirapuera).

Não obstante fosse possível tentar explicar cada caso em separado, o tipo de problema sugere uma mesma explicação para todos. Como mencionado anteriormente, vários processos ou uma combinação deles poderia ser a causa dos dois fenômenos (registro de um valor muito alto em um horário, e registro de valores uma hora antes e uma hora depois muito mais baixos) envolvidos relativos à medição. As possíveis explicações para o desaparecimento daqueles valores poderiam ser: 1) falha no equipamento somente no momento daquela medição, o que parece pouco provável; 2) precipitação pluviométrica no momento da emissão o que possibilitaria apenas o primeiro registro em níveis elevados. Tal coincidência tem probabilidade de ocorrência tão baixa que para acontecer teria que obedecer a ações deliberadas o que, em princípio, deveria ser descartado; 3) transporte atmosférico sob regime de ventos de alta velocidade no momento da medição tal que

valores elevados pré e pós 1:00 h não seriam captados pelo instrumento. Também parece pouco provável; 4) formação de aerossóis acarretando rápida oxidação do SO_2 a H_2SO_4 e a sulfatos, tais que o registro daqueles níveis de SO_2 só seriam possíveis no momento da emissão. Dentre outras possibilidades, as hipóteses 2, 3 e 4 merecem atenção. Assim, dados sobre frequências de precipitação pluviométrica, de inversões térmicas e de condições desfavoráveis à dispersão podem ser úteis.

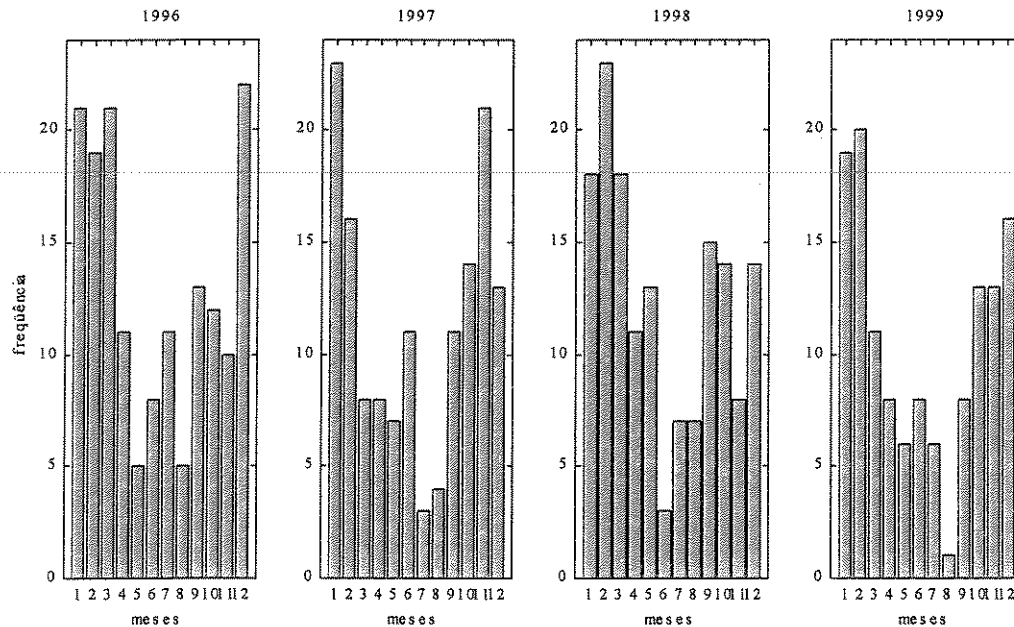


Fig.7.6: Frequência de precipitação pluviométrica no período de 1996 a 1999 na RMSP.

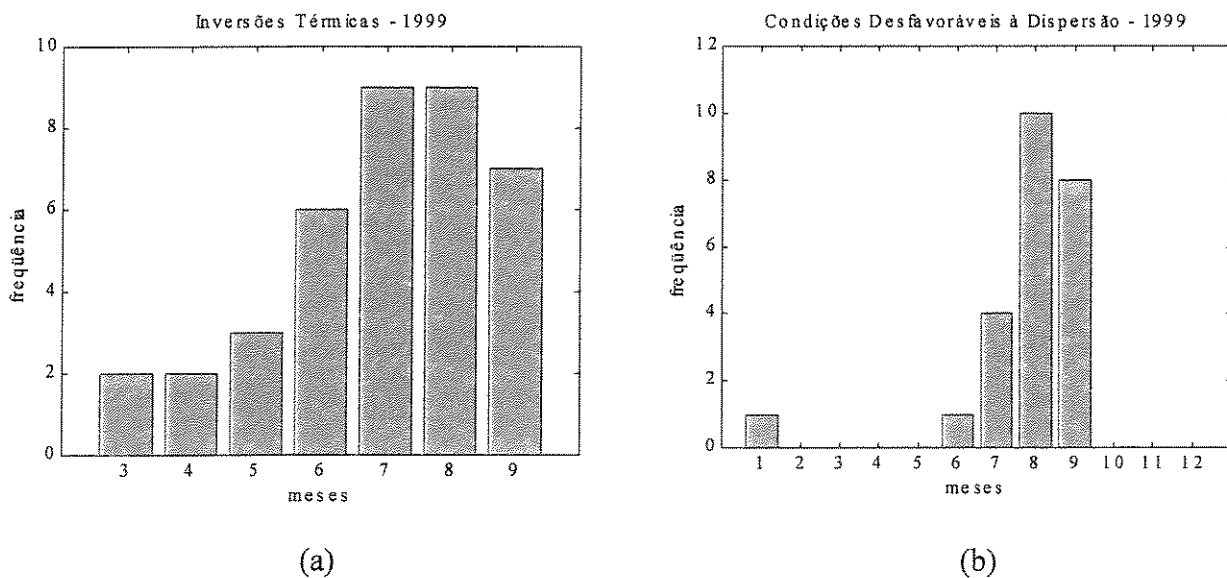


Fig.7.7: Frequências a)de inversões térmicas, e b)de condições desfavoráveis à dispersão.

A Figura 7.6 mostra a incidência de chuvas no período de 1996 a 1999 na RMSP. Em todos os anos, a faixa mais crítica situa-se entre maio e setembro. O nível mais baixo foi registrado em agosto de 1999. Assim, os dados sobre precipitação pluviométrica tornam, de fato, pouco provável a segunda hipótese.

Transporte atmosférico só poderia ser concebido se uma massa compacta de ar contendo altas concentrações de SO_2 passasse pelas áreas em questão. Admitindo-se que isto seja possível, deveria haver registros nas Estações de Medição Cerqueira César e Centro, locais muito próximos (e numa provável rota de deslocamento da massa de ar) do Ibirapuera, no qual os fenômenos se registaram com maior intensidade. No Centro tal fato não se verificou, porém, na localidade Cerqueira César houve o registro de dados cuja análise levou a resultados semelhantes ao ocorrido nas estações citadas anteriormente.

As Figuras 7.7a e 7.7b mostram, respectivamente, as frequências de inversões térmicas e de condições desfavoráveis à dispersão de poluentes na RMSP. O resultado mostra que houve maior ocorrência de inversões de junho a setembro, período no qual a dispersão de poluentes foi menos favorável. Se todas as hipóteses anteriores forem descartadas, isto remeteria para a quarta hipótese. À 1:00 h da manhã tem-se umidade relativa mais elevada favorecendo a estabilidade de equilíbrio entre fases gasosa e aquosa, o que poderia levar alguma espécie química possuindo maior afinidade pelo SO_2 a oxidá-lo. Tal fato poderia explicar o problema da queda súbita dos valores de concentração de SO_2 no curto intervalo de tempo entre três medições consecutivas (0:00 h, 1:00 h e 2:00 h). Entretanto, permaneceria o “mistério” sobre a origem das emissões.

Retornando à hipótese de transporte atmosférico, em uma análise das posições geográficas das estações situadas no raio de ação daquelas que manifestaram o fenômeno (Fig. 4.1, Cap. I; e Fig. B1, Apêndice B1), verifica-se que P. D. Pedro II, Centro, Cerqueira César e Ibirapuera são locais relativamente próximos entre si. São Caetano fica um pouco mais afastado. Porém, sob o ponto de vista de direção e velocidade do vento, todas as cinco localidades podem ser consideradas como receptoras de poluentes originários de locais como Santo Amaro, Diadema, São Bernardo do Campo (especialmente do Distrito Industrial de São Bernardo do Campo), por meio de frentes frias provenientes do sul.

O fato da localidade Osasco não ter apresentado o fenômeno poderia ser atribuído à sua localização, um pouco fora da direção de ventos provenientes do sul do país. Todavia,

tal argumento não explica a ausência do fenômeno no Centro. Provavelmente, o que explicaria o fato seria a distância entre a fonte de emissão e a localidade. Isto faz sentido se for considerado que o efeito perde força à medida que se desloca do Ibirapuera (local mais próximo da fonte emissora) para Pedro II (local mais afastado daquelas emissões). De todo modo há que se considerar a velocidade do vento alta o suficiente para que o efeito só possa ser registrado naquele horário simultaneamente nas quatro localidades.

As análises PARAFAC nas quais é incluído o SO_2 não revelam nada além do fato de que o gás tem como principal fonte de emissão os veículos automotivos, como pode ser observado nas Figuras 6.48, 6.52, 6.53, 6.54, 6.55, 6.56, 6.58 e 6.59 (p321 a p333). Em todos esses casos o modelo captou duas componentes. Uma delas revela o perfil da fonte de emissão (automotiva), portanto, associado aos poluentes primários. A outra mostra o perfil para o poluente secundário, o ozônio. A tentativa de se extrair mais fatores no modelo PARAFAC causa degenerescência, tal que o algoritmo não consegue convergir, ou, quando converge, apenas duas das componentes no modo B possuem significado físico. Uma delas descreve as variações para a fonte automotiva, e a outra, as do ozônio.

Quando a análise PARAFAC é conduzida sobre os dados considerando-se o período afetado sem correções para os dados à 1:00 h, o modelo utiliza três componentes para descrever as variações sistemáticas presentes nos dados. Além dos perfis para a fonte automotiva e para o ozônio, o modelo capta também as variações sistemáticas originárias da indústria.

Por outro lado, quando o SO_2 (com correção no período afetado) é introduzido em substituição ao CO (ou CH_4 , ou NMHC) no respectivo ciclo (Figura 6.63, p339), e a análise é conduzida por meio do modelo de Tucker, são extraídos, como no PARAFAC, três componentes no modo A (PC_1 para a fonte de emissão, e PC_2 e PC_3 como componentes relacionadas com a acumulação). Porém, com adequada construção do núcleo do modelo, é possível extrair também três componentes no modo B. Uma delas refere-se ao perfil semanal do poluente secundário, o ozônio. Das outras duas, uma está relacionada com as variações sistemáticas semanais para os poluentes primários NO e NO_2 , e a outra está relacionada com o também poluente primário SO_2 . Isto é, uma das componentes representa o perfil semanal de emissão da fonte, e a outra representa o perfil semanal de emissão exclusivamente do SO_2 , porém, da mesma fonte, a automotiva. Este fato está

provavelmente relacionado com os parâmetros determinantes das reações a que estão sujeitas, na atmosfera, as diferentes substâncias emitidas pela fonte. Como se verá adiante, essa característica para o SO_2 o aproxima, em termos de comportamento, da variável NMHC.

Analogamente à análise com o modelo PARAFAC conduzida sobre o período afetado sem correções para os dados à 1:00 h, o modelo de Tucker pode ser ajustado para extrair três componentes no modo A (perfis diários), três no modo B (perfis semanais) e duas no modo C (variação anual). Entretanto, para a Estação de Medição Ibirapuera, o modelo exige três componentes no modo C. Uma delas para modelar as emissões à 1:00 h, e as outras duas para descrever as variações anuais para os poluentes primários da fonte automotiva e para o poluente secundário, O_3 .

Observe-se que é possível (no caso do período afetado com correções para os dados à 1:00 h) impor restrições para que o modo B, para as variações sistemáticas semanais, possua apenas duas componentes, já que as curvas apresentam o mesmo padrão de variação, como indicado pela análise PARAFAC. Todavia, em alguns casos, e o Tucker permite que se faça isto, é desejável desmembrar o perfil de determinada fonte (no caso a automotiva) em seus respectivos tipos de veículos emissores. Esta providência pode facilitar a associação do perfil com as reações químicas e com as condições de dispersão a que estão sujeitos os poluentes.

Assim do resultados das análises para o SO_2 pode-se depreender que durante o ano de 1999, emissões provenientes de indústrias também tiveram papel importante na queda da qualidade do ar na RMSP. Para as localidades P. D. Pedro II e São Caetano, as emissões automotivas constituíram-se como a principal fonte para o poluente SO_2 . Para o Ibirapuera a situação se inverte, assumindo as emissões industriais o papel de principal fonte de poluição daquele composto naquele local da RMSP.

Para os anos de 1997 e 1998, as emissões automotivas constituíram-se como a fonte principal de poluição por SO_2 , cujos níveis registrados foram sempre muito inferiores aos do ano de 1999.

Após a discussão para o SO_2 , não foi possível encontrar, sem sombra de dúvidas, uma explicação para o problema das emissões à 1:00 h.

Para o metano, suspeita-se que possa haver pelo menos quatro fontes importantes de emissão. Uma delas poderia se concentrar em veículos movidos a gás natural, outra em indústria(s) que também utiliza(m) esse combustível, outra em estações de tratamento de esgotos, e a última em “aterros sanitários” (“lixões”). Nos dois locais estudados, os resultados apresentam diferenças, as quais não os tornam excludentes em termos de fontes de emissão.

As figuras 6.37 (p307) e 6.41 (p311) mostram os perfis de emissão da fonte obtidos nas análises PCA. Para o Parque D. Pedro II (Figura 6.37), os valores se mantêm aproximadamente constantes até cerca de 7:00 h da manhã, quando os níveis começam a cair rapidamente atingindo um mínimo às 15:00 h, e voltando a crescer enquanto se aproxima das 24:00 h. Esse perfil não tem semelhança com as variações sistemáticas diárias de fontes de emissão automotivas.

Quando se observam os resultados pela análise PARAFAC (Figura 6.38, p308) para esse local, nota-se concordância no perfil de emissão da fonte (Figura 6.38a). A Figura 6.38b (e também d) mostra que os valores para segunda, terça e quarta feiras são mais elevados do que para os demais dias da semana. Ou seja, juntamente com os valores dos pesos para Sexta-feira e Quinta-feira, aqueles para os finais de semana encontram-se na parte inferior do gráfico. Isto poderia levar à conclusão que a principal fonte de emissões poderia ser automotiva se as variações sistemáticas diárias (PCA e PARAFAC) não contrariassem tal assertiva. Assim, se tais variações existirem, não foram captadas pelos modelos ou ficaram acomodadas nas curvas correspondentes às variações sistemáticas automotivas. Nesse caso, a principal fonte de emissão de metano nessa localidade não é automotora.

Para São Caetano do Sul (Figura 6.41, p311), o perfil de emissão da fonte obtido na análise PCA difere daquele obtido para a Estação Pedro II. Uma inspeção na Figura 6.41 mostra a existência de três picos. O primeiro por volta das 9:00 h, o segundo às 13:00 h e o terceiro aproximadamente às 22:00 h. É interessante notar que o pico central, de menor intensidade que o primeiro, ocorre exatamente às 13:00 h, horário no qual as emissões dos veículos estão quase no seu nível mais baixo, de onde passam a crescer novamente, e no qual os níveis de ozônio estão se aproximando do seu pico.

Os resultados PARAFAC (Figura 6.42, p311-312) confirmam o perfil para as variações sistemáticas diárias (Figura 6.42a). Entretanto, para as variações sistemáticas semanais (Figura 6.42b) são, à primeira vista, no mínimo, intrigantes. A queda nos níveis ocorre mais pronunciadamente no sábado, voltando a crescer, atingindo valores elevados, já no domingo, e mantendo-os nos chamados dias úteis. Esse perfil semanal difere daquele obtido para o metano no P. D. Pedro II (Figura 6.38, gráficos b e d), onde os fins de semana ocupam a parte inferior do gráfico.

A explicação para isto não é simples, e não pode ser obtida pelas análises dos poluentes individualmente, a menos que sejam introduzidas outras informações (isto é, análise de outros poluentes). Por outro lado, quando as variáveis são analisadas sob a influência das demais é possível extrair informações capazes de conduzir a uma melhor interpretação dos resultados. Ou seja, na atmosfera os poluentes estão sujeitos a diversos tipos de interação, tal que a sua análise individual, embora necessária, não é suficiente para fazer emergir o real comportamento subjacente aos dados, expresso pelos valores registrados para cada variável. Assim, torna-se prudente mais um pouco de discussão no sentido de tornar mais sólidos os argumentos que porventura possam esclarecer tais aparentes discrepâncias.

Para os hidrocarbonetos não metano, as análises PCA sobre os dados coletados nas duas localidades em discussão foram mostrados nas Figuras 6.39 (Pedro II, p309) e 6.41 (São Caetano, p311). Nos dois casos os perfis diários são semelhantes e apresentam os dois picos característicos de emissões automotoras.

Os resultados obtidos nas análises PARAFAC, expressos nas Figuras 6.40 (p309-310) e 6.44 (p313), respectivamente, para Pedro II e São Caetano, também são semelhantes tanto nas variações sistemáticas diárias quanto nas semanais, e na variação anual.

Quando os NMHC substituem o CO (ou CH₄) no ciclo do monóxido de carbono (ou do metano), o modelo PARAFAC acomoda as variações sistemáticas diárias, semanais e a variação anual em até três componentes. Este fato coloca um dado novo porque com duas componentes (Figura 6.47.1, p319) o modelo está captando as variações sistemáticas da fonte de emissão e as variações para o poluente secundário. Nesse caso, o perfil de emissão diário para a fonte automotiva (Figura 6.47.1a) é representado por PC₁, enquanto PC₂ descreve as variações para o ozônio. Na Figura 6.47.1b (e d, p319), percebe-se que o perfil

semanal para os poluentes primários é representado por PC_1 , enquanto PC_2 representa o perfil para o poluente secundário. Nota-se a queda das emissões automotivas nos finais de semana (PC_1) e os valores elevados para o O_3 nesse mesmo período. Entretanto, quando solicitado a processar os dados com três componentes, o PARAFAC acomoda em PC_1 e em PC_3 as emissões da fonte automotiva, e em PC_2 as variações sistemáticas diárias para o O_3 (Figura 6.47.2a, p320). Indo além, no modo B, as variações sistemáticas semanais para a fonte automotiva são também acomodadas em PC_1 e PC_3 , enquanto PC_2 representa o perfil semanal para o ozônio, como mostra a Figura 6.47.2b (p320). E mais, enquanto o ozônio reproduz seu comportamento típico (PC_2) durante a semana, PC_1 e PC_3 possuem um mesmo padrão de variação, isto é, ambos apresentam queda nos finais de semana. Isto, a primeira vista “soou” de modo estranho porque se a fonte de emissão é automotiva, o modelo PARAFAC deveria usar uma das componentes para essa fonte. No entanto, duas componentes são utilizadas para a mesma. Se a alta correlação entre esses dois fatores não foi suficiente para causar degenerescência, então existem, de fato, dois tipos diferentes de emissores vinculados à mesma fonte de emissão. E ambos são importantes. Essa constatação inspirou a criação do ciclo dos NMHC.

Os demais resultados, incluindo aqueles para a temperatura e umidade foram parcialmente discutidos na Seção 6.3. As Figuras correspondentes construídas para expressar aqueles resultados são, em sua maioria, auto-explicativas. Assim, a sua discussão nessa seção torna-se supérflua e somente quando necessário far-se-á menção àqueles resultados.

De acordo com os resultados apresentados na Seção 6.4, as análises por meio dos modelos PARAFAC e de Tucker foram conduzidas organizando inicialmente os poluentes nos dois ciclos básicos em grupos inicialmente com quatro variáveis. Todavia, o comportamento dos NMHC emergentes das análises PARAFAC (Figuras 6.47.1 e 6.47.2, p319 e p320) sugeriu a criação de um ciclo básico específico para essa variável, pois isto facilitaria a interpretação dos resultados. De fato, a alta reatividade dos hidrocarbonetos não metano em relação ao monóxido de carbono e ao metano (Birks, 1998, p.249; Sherwood e Craig, 2003) confere-lhes características próprias. Assim, as discussões nesse capítulo tomam por base a existência de três ciclos básicos “reguladores” das reações químicas e da dispersão dos poluentes na atmosfera da RMSP.

Os três ciclos têm em comum a presença dos óxidos de nitrogênio e do ozônio. Isto porque parece ser NO o controlador de todos os processos químicos ocorrendo na baixa troposfera, e é o NO₂ (Baird, 1995, p.126; Madronich, 1993) a principal fonte de oxigênio atômico que, em reação com moléculas de oxigênio, vai gerar o O₃ nessa região da atmosfera. Assim, mantidos fixos esses três poluentes, as outras variáveis são introduzidas, segundo um critério de avaliação de suas características potenciais fornecidas pelos modelos PARAFAC e Tucker, até finalizar com a junção dos ciclos contendo todos os poluentes.

7.2 – Ciclo do Monóxido de Carbono

A discussão dos primeiros resultados ficará restrita ao ciclo tríplice (Jacob, 2000) mínimo, denominado nessa tese de *Ciclo Básico do Monóxido de Carbono* (Figura 7.8).

Conforme já discutido no Capítulo II, a presença do ozônio na troposfera pode ter origem em duas fontes (Baird, 1995, p.82-83; Birks, 1998, p.235; Cárdenas *et alli*, 1998; Jenkin e Hayman, 1999; Tsutsumi e Matsueda, 2000; Jacob, 2000): 1) transporte de ozônio a partir da estratosfera, e 2) produção fotoquímica direta na própria troposfera. As condições necessárias para a ocorrência do segundo caso são: uma fonte de carbono (CO e/ou hidrocarbonetos), óxidos de nitrogênio (NO, NO₂) e a luz solar.

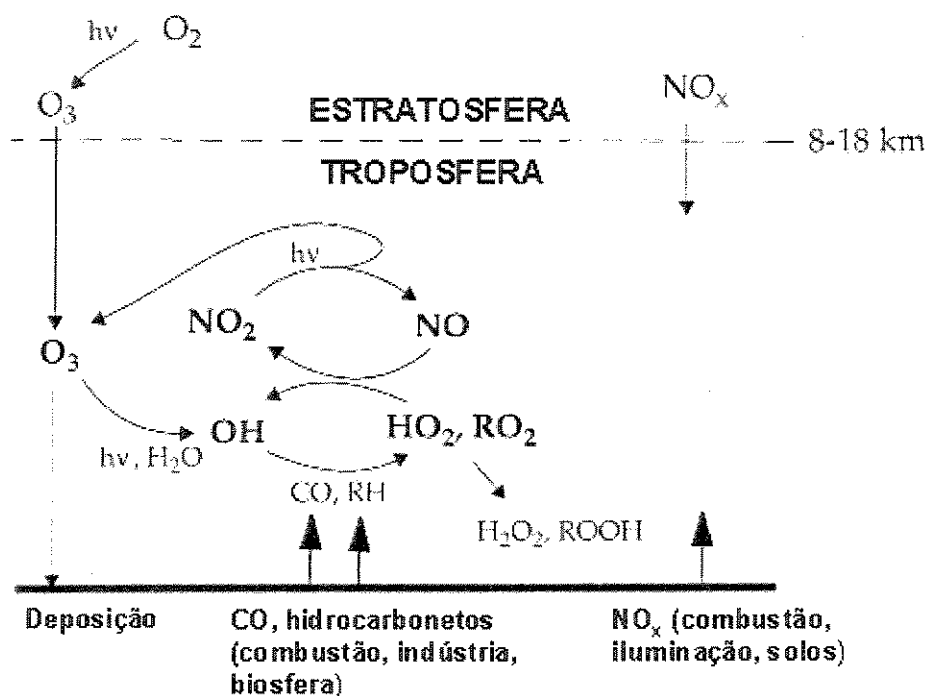


Fig.7.8: Ciclo do monóxido de carbono (Jacob, 2000)

Os perfis para os poluentes primários e secundários pertencentes ao ciclo do monóxido de carbono foram mostrados na Seção 6.3. As Figuras 6.1, 6.2a, 6.2b e 6.3 (p241 a p244) referem-se às variações sistemáticas diárias, respectivamente, para o CO, o NO, o NO₂ e o O₃. Seria interessante resgatar aqui os gráficos de escores de tais variáveis para que

se possa comparar suas variações ao longo do ano. Isto é, construir em uma única figura os gráficos de escores (“scores”) das variáveis desse ciclo para que se possa avaliar melhor o porque dos resultados PARAFAC (e Tucker) para o perfil semanal dos poluentes primários e secundários.

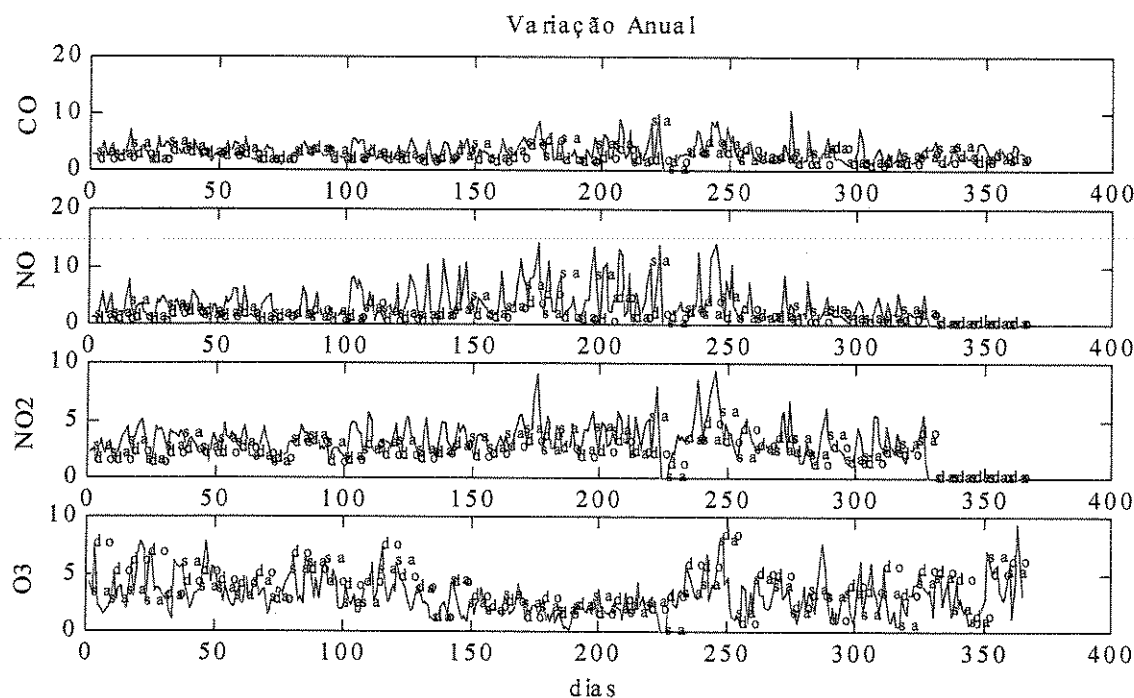


Fig.7.9: Variação anual para os poluentes do ciclo básico do monóxido de carbono. Estão assinalados apenas os finais de semana (sa = Sábado e do = Domingo).

A Figura 7.9 mostra, como já visto anteriormente, que os poluentes primários possuem baixos “scores” durante os fins de semana, enquanto o poluente secundário possui valores mais elevados nesse mesmo período. É possível captar tais informações por meio de PCA, apenas mediante modificação adequada na matriz de dados originais, o que não é necessário quando são usadas estruturas de ordem superior. Isto é, a matriz convencional com um dos modos para os 365 dias e o outro para as 24 horas, é ela própria, a fonte de dados de entrada para os modelos multimodo. Apenas essa simplificação já seria um forte indicador para o uso de arranjos multimodo.

De fato, tais variações semanais distintas para os poluentes primários e secundários foram captadas por meio de estruturas multimodo. Por suas características, o modelo

PARAFAC foi utilizado primeiramente para “rastrear” e confirmar a existência de determinados perfis para os poluentes. Conhecidas as suas formas, utilizou-se o modelo de Tucker para conquistar o “terreno” mais adiante, imperfeito demais para o primeiro, e, finalmente, tentar consolidar os resultados apontados pelo modelo PARAFAC.

Não obstante a importância de todo o trabalho desenvolvido até aqui, ele não seria de muita utilidade se não se pudesse associar tais resultados à realidade da atmosfera, espaço privilegiado de um quase sem número de reações químicas no qual vários pesquisadores têm-se debruçado na tentativa de desvendar parte considerável de seus processos. Assim, tentar-se-á, a partir desse ponto, estabelecer conexão entre os resultados que os modelos utilizados nas análises permitiram obter e as próprias reações químicas ocorrendo na atmosfera.

Para enfatizar o caráter reativo das espécies químicas envolvidas, serão usadas as estruturas de Lewis (Baird, 1995, p.34-35) para representá-las, na discussão que segue desse ponto em diante.

A produção de O₃ na troposfera pode ser descrita como uma cadeia de reações de oxidação de CO e de hidrocarbonetos, catalisados por radicais hidroxila e suas formas peroxi, na presença de óxidos de nitrogênio (Jacob, 2000).

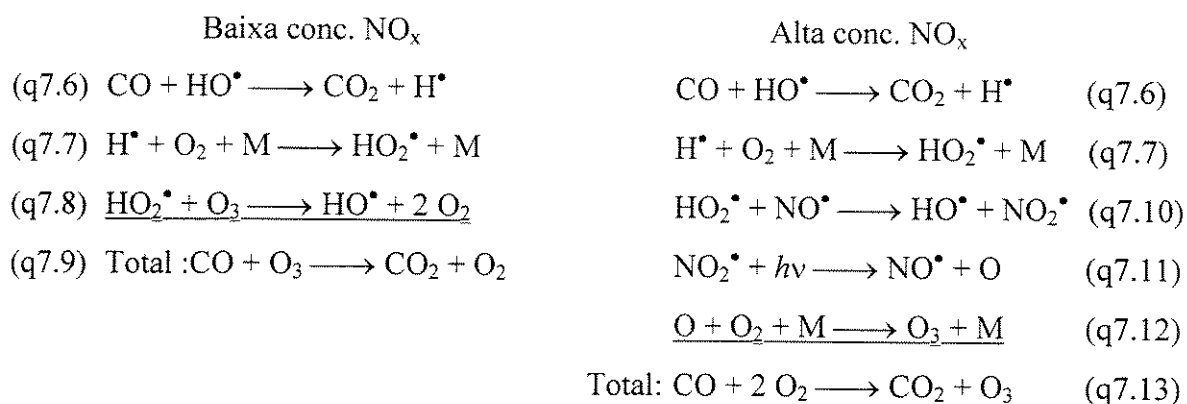
De forma resumida, mesmo na ausência de poluição (no ar troposférico limpo), pode-se dizer que o início do processo se dá com a decomposição fotoquímica de quantidades em níveis traço de ozônio presentes na atmosfera, formando moléculas de oxigênio e átomos excitados de oxigênio. Estes, por sua vez, reagem com moléculas de água no estado gasoso, produzindo radicais livres hidroxila. Essa última espécie é a responsável pelo início de todo o complexo conjunto de reações na troposfera (Baird, 1998; Jacob, 2000; Madronich, 1993).



A formação das espécies peróxido e hidroperóxido dá continuidade ao processo, por meio da reação com o CO e com metano e outros hidrocarbonetos, em presença de óxidos de nitrogênio. O caso mais simples envolve a oxidação de CO (Jacob, 2000):



Uma série de outras reações passa a ocorrer, formando um mini ciclo para o ozônio, outro para os radicais hidroxila e suas formas peróxido, e um terceiro para os óxidos de nitrogênio (Jacob, 2000). É de particular interesse os dois diferentes caminhos que as reações podem tomar dependendo das concentrações de óxidos de nitrogênio (especialmente de NO) no meio. Abaixo são apresentadas essas duas possíveis vias (Birks, 1998, p.235). Se as concentrações dos óxidos de nitrogênio são baixas não ocorre a formação de ozônio conforme mostra a reação global (q7.9) abaixo (baixa conc. NO_x). Porém, se as concentrações desses óxidos são altas, há formação de ozônio, como se vê no segundo conjunto de reações e na reação global (q7.13) para esse caso (alta conc. NO_x).



A análise do conjunto de reações e da reação global para a situação em que há elevada concentração de NO_x, mostra que o ozônio é produzido às custas da oxidação de CO a CO₂. Esse processo, como se observa, é catalisado por HO_x (radicais hidroxila e suas formas "peroxi") na presença de NO_x (Jacob, 2000).

É interessante comparar essas equações químicas com os resultados obtidos através dos modelos quimiométricos convencional e multimodo.

O acompanhamento da discussão que se faz doravante, exige, inicialmente, o acesso às Figuras 6.1, 6.2 e 6.3, respectivamente às p241, 242 e 244. No decorrer da explanação outros gráficos deverão ser analisados. Assim, quando necessário, remeter-se-á o leitor às páginas contendo as respectivas figuras.

Para a Estação de Medição Parque D. Pedro II, as Figuras 6.1, 6.2 e 6.3 (p241 a p244) mostram a relação entre CO e os óxidos de nitrogênio e suas correlações com o ozônio. Observando-se as reações q7.8 e q7.10 (p372) em cada bloco acima, verifica-se que o O_3 e o $NO^\bullet$ competem pelo $HO_2^\bullet$. Quando a concentração de $NO^\bullet$ é suficientemente baixa a reação ocorre preferencialmente com o O_3 , destruindo-o no processo (Birks, 1998, p.235). Caso contrário a reação ocorre com o $NO^\bullet$, provocando a formação de O_3 . Uma inspeção nos gráficos da variação diária de $NO^\bullet$ (Figura 6.2a) e de O_3 (Figura 6.3) permite verificar, de fato, uma correlação negativa entre esses compostos, indicando que a produção de ozônio se deu às custas do consumo de $NO^\bullet$.

Quando se faz essa mesma comparação com o gráfico que mostra a variação diária para o CO (Figura 6.1), observa-se, do mesmo modo, que a produção de O_3 também se deve ao consumo de CO. Para perceber isto melhor basta se concentrar nos horários de pico para o CO e o $NO^\bullet$, os quais ocorrem aproximadamente às 8:00–9:00 h, e no horário de pico para o O_3 que ocorre às 14:00 h, exatamente nos vales de CO e $NO^\bullet$. Entretanto, deve-se chamar a atenção para o fato de que o decréscimo de amplitude nesse mesmo horário não se deve exclusivamente ao consumo de CO, $NO^\bullet$ e $NO_2^\bullet$ para a produção de O_3 . A queda na densidade de tráfego no mesmo período, por exemplo, também é responsável por esse resultado. Afinal, as curvas representam os perfis de emissão da fonte, porém como o O_3 não possui uma fonte de emissão, ao contrário, é formado por meio de reações químicas entre poluentes primários (estes sim provenientes de fontes emissoras), pode-se, pela análise das figuras, associar a formação de O_3 às custas do consumo daqueles.

Além disto, observa-se também que enquanto os níveis de O_3 (Fig.6.3, p244) caem quando se “caminha” para as 20:00 h, os níveis de CO (Fig. 6.1, p241) e de $NO^\bullet$ (Fig. 6.2a, p242) voltam a subir. Entretanto, os níveis de CO nesse horário atingem valores mais elevados do que os do pico da manhã. É natural que isto ocorra (dadas as condições de tráfego local), porque nesse horário ainda há tráfego intenso, mas não luz solar suficiente para favorecer o processo, acumulando mais CO à noite.

Fatores tais como a reatividade do CO e outros compostos presentes na atmosfera (por exemplo, SO₂) em relação ao radical hidroxila, também parecem ser determinantes no nível de acumulação de CO na respectiva localidade.

Da mesma forma deve ser analisado o segundo pico do NO[•]. Diferentemente do que ocorre com o CO, para o NO[•] o pico das 20:00 h é menor do que o da manhã. Também é natural que isto aconteça (dadas as condições de tráfego local) pois nesse horário, no qual ocorre também a queda de O₃, o pico do NO[•] não se eleva tanto quanto o do CO devido a uma combinação de pelo menos três fatores:

a) parte do NO[•] continua se convertendo a NO₂[•], como mostra a equação q7.10 (p372);

b) parte do NO[•] reage com uma grande quantidade de O₃ acumulado. Esse processo não somente não permite que o NO[•] acumule mais, como contribui para causar uma rápida queda das concentrações de O₃; e

c) parte do NO[•] reage com HO[•] para formar HONO (Baird, 1995, p.131). Apesar de Calvert *et alli* (1994) terem proposto um mecanismo para a formação de HONO noturno (ver também trabalho de Northolt *et alli*, 1992) envolvendo reação de NO[•] com NO₂[•], resultados anteriores de Jenkin *et alli* (1988), e também de Bambauer *et alli* (1994) e de Mertes e Wahner (1995), mostram que a formação de HONO não envolve reação direta com NO[•], mas sim com o NO₂[•]. Assim, no item c está sendo considerado que parte do NO₂[•] formado na reação q6.10 (à noite) leva à formação de HONO. Como na reação q7.10 o NO₂[•] é formado às custas do consumo de NO[•], poder-se-ia considerar a formação de HONO noturno por meio de consumo indireto de NO[•].

Continuando a explanação, o NO₂[•] produzido na equação q6.10 não encontra mais a luz solar (hν) para se transformar em NO[•] + O e, desse modo, prosseguir o ciclo.

De fato, isto pode ser constatado ao se observar o gráfico de variação diária de NO₂[•] (Fig.6.2b, p242). Nele verifica-se que o pico das 19:00 h é aproximadamente igual ao da manhã, mostrando que o NO₂[•] não mais se decompõe fotoquimicamente, acumulando-se mais do que o NO[•] à noite.

O decréscimo de NO₂[•] após 20:00 h pode ser devido à produção de nitrato de peroxiacetila (PAN) e outros nitratos de peroxiacila (PANs), a partir do NO₂[•], e mesmo

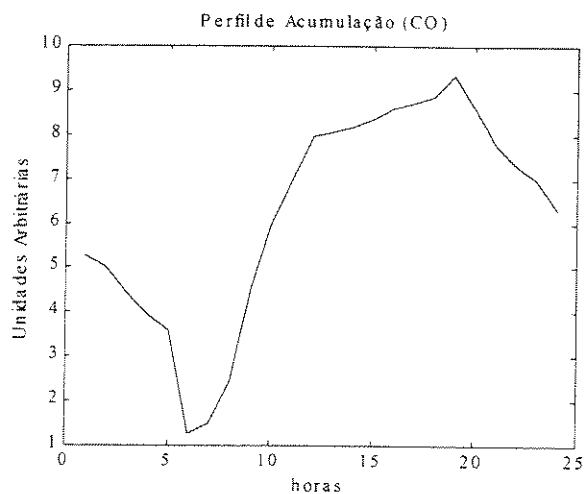
outras reações como aquelas produzindo o radical $\text{NO}_3^\bullet$, o qual é convertido a HNO_3 (Baird, 1995, p.129; Birks, 1998, p.244; Madronich, 1993).

Quando essas variações são examinadas por meio do modelo PARAFAC, verifica-se que a reação global para o caso de altos níveis de óxidos de nitrogênio, está representada em apenas duas componentes (PC_1 e PC_2 na Figura 6.45a, p306), sendo que o modelo ainda distingue os poluentes primários do secundário, O_3 . É importante salientar que o modelo capta o perfil da fonte (PC_1) caracterizado pela emissão de poluentes primários, e o perfil associado com a formação do poluente secundário (PC_2).

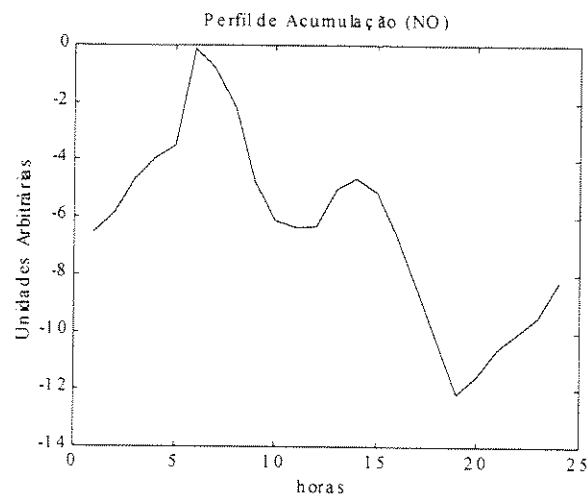
Quando os dados foram submetidos ao modelo de Tucker, os resultados evidenciaram a existência de um perfil relacionado com a fonte de emissão (PC_1 na Figura 6.60a, p334), outro com a acumulação/consumo de poluentes primários (PC_2 na Figura 6.60a), e um terceiro com a produção/acumulação de poluentes secundários (PC_3 na Figura 6.60a). Além disso, é possível observar o perfil semanal (Figura 6.60b, e d, p334) onde há queda das emissões originárias dos veículos nos finais de semana, e o inverso para o ozônio. Adicionalmente, nota-se a variação anual (Figura 6.60c, p334), na qual se verifica a maior acumulação para os poluentes primários durante o inverno. Não obstante o termo acumulação usado nessa tese englobe *reações entre e dispersão de* poluentes, os termos acumulação/consumo e produção/acumulação foram usados aqui para enfatizar as relações estritas entre reações (de consumo e de formação) e acumulação.

Nesse ponto é interessante introduzir os resultados referentes à acumulação dos poluentes do ciclo do monóxido de carbono para que se possa interpretar os resultados obtidos à luz das reações químicas ocorrendo na baixa troposfera.

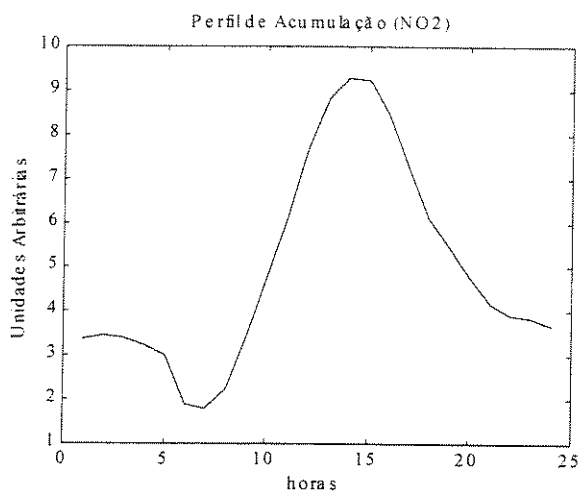
Apenas para o ciclo básico do monóxido de carbono é apresentada uma discussão detalhada associando os perfis encontrados pelo modelo de Tucker às reações químicas que têm lugar na atmosfera.



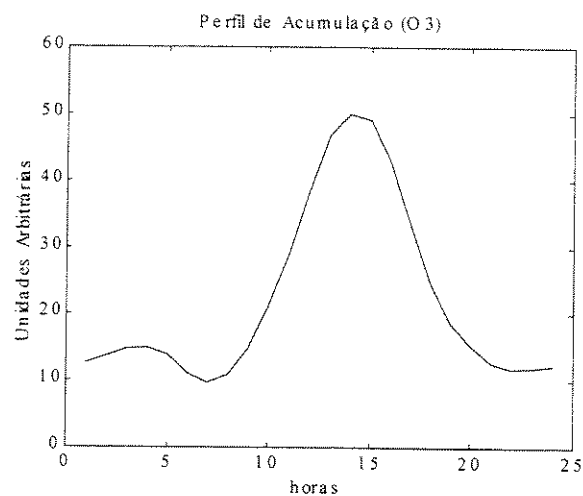
(a)



(b)



(c)



(d)

Fig.7.10: Perfis de acumulação: a) CO, b) NO, c) NO₂, d) O₃.

A Figura 7.10a mostra o perfil de acumulação para o CO (soma dos efeitos de PC₂ e PC₃, Figura 6.60a, p334, pesados pelo CO). Observe-se que até cerca de 15:00 h esse perfil recebe uma dominante contribuição da componente PC₂ mostrada na Figura 6.60a e, logo após, a contribuição de PC₃ predomina. PC₂ possui alta correlação com PC₁ nesse período. Depois desse ponto, embora a correlação exista, PC₂ cresce mais rápido do que PC₁. Isto significa que as emissões continuam mas algo está afetando o consumo de CO. Esse fato é caracterizado pelo rápido crescimento de PC₂ após 15:00 h. Aqui, as emissões de CO

possuem um peso maior em PC_2 , do que o $NO^\bullet$ e o $NO_2^\bullet$. Deve-se ter em mente que PC_2 leva em conta contribuições de CO, $NO^\bullet$ e $NO_2^\bullet$, mas nesse período aquelas relativas a $NO^\bullet$ e $NO_2^\bullet$ são menores. Este fato pode ser visto na Figura 6.60a (p334) onde, à tarde, PC_2 cresce enquanto PC_3 decresce como resultado da reação entre $NO^\bullet$ e O_3 , e outras reações tais como perda indireta de $NO^\bullet$ assinalada pelo processo citado na página 374 (Jenkin *et alli*, 1988; Bambauer *et alli*, 1994; Mertes e Wahner, 1995), e perdas de $HO^\bullet$ (Jacob, 2000; Madronich, 1993). Desse modo, a queda nas concentrações de $NO^\bullet$ e $HO^\bullet$ provoca uma queda no consumo de CO tal que a sua acumulação aumenta como se pode ver na Figura 7.10a (p376). Assim, a acumulação de CO deve-se, em grande extensão, à contínua emissão automotiva.

Após 20:00 h, cai a acumulação de CO. Durante o dia ainda existe tráfego de veículos entre os dois horários de pico tal que existe uma pequena mas contínua emissão de CO em torno de 12:00-15:00 h. Ao contrário, após cerca de 20:00 h cai drasticamente o tráfego de veículos provocando, desse modo, uma queda na acumulação de CO.

A Figura 7.10b mostra o perfil de acumulação para o $NO^\bullet$ (soma dos efeitos de PC_2 e PC_3 , pesados pelo $NO^\bullet$). A curva é uma imagem invertida da componente PC_2 na Figura 6.60a (p334). Seus valores negativos significam que o $NO^\bullet$ controla todos os processos na troposfera. Isto é, $NO^\bullet$ tem uma remarcável “identidade” com a componente PC_2 . Isto pode ser confirmado observando-se a equação q7.10 (p372). Nela pode-se constatar que quando os níveis de emissão de NO_x são elevados ocorre o “chaveamento” da destruição para a produção de O_3 . Assim que os níveis de $NO^\bullet$ atingem um dado valor mínimo inicia-se o seu consumo em reações com radicais $HOO^\bullet$ levando ao final à formação de O_3 (Birks, 1998; Jacob, 2000; Madronich, 1993). Quando o ozônio formado alcança níveis elevados, ele próprio reage com o $NO^\bullet$ reduzindo a concentração troposférica desse último (Madronich, 1993).

Não obstante os dados analisados nessa tese refiram-se a registros em Estações de Medição fixas e relativamente próximas do nível solo (isto é da superfície), a curva para a acumulação de $NO^\bullet$ mostrada na Figura 7.10b reflete, à luz das reações do ciclo para o CO, e dos resultados que mostram a reação entre O_3 e $NO^\bullet$, essa característica do $NO^\bullet$ como regulador de boa parte dos processos ocorrendo na troposfera.

A despeito de tudo isto, os resultados nessa tese para essa localidade também indicam aumento nos níveis de ozônio durante os finais de semana, e a ocorrência um pequeno pico noturno (aproximadamente 3:00-4:00 h) para o poluente. Esses resultados também necessitam explicação plausível.

Vários fatores afetam a acumulação de $\text{NO}^\bullet$. A análise de sua acumulação deve necessariamente ser conduzida tendo em mãos as Figuras 6.60a (p334) e 7.10b (p476). Além disto, é conveniente, para a análise, dividir o dia nos seguintes períodos:

- 1) ~20:00 h até ~6:00 h;
- 2) ~6:00 h até 10:00 h;
- 3) ~10:00 h até ~14:00 h; e
- 4) ~14:00 h até 20:00 h.

No primeiro período (~20:00 até ~6:00 h), diminuem os níveis de $\text{NO}^\bullet$, o qual é produzido por meio da decomposição fotoquímica de $\text{NO}_2^\bullet$ (reação q7.11, p372). Uma série de reações noturnas passa a ocorrer:

1) reação de produção de HONO por meio do processo c assinalado na página 374 (Jenkin *et alli*, 1988; Bambauer *et alli*, 1994; Mertes e Wahner, 1995),

2) o $\text{NO}_2^\bullet$ acumulado na ausência de luz solar passa a se constituir, juntamente com o O_3 residual, como a mais significativa fonte de radicais $\text{NO}_3^\bullet$ (Dimitroulopoulou e Marsh, 1997),

3) reação entre $\text{NO}_3^\bullet$ recém produzido e o $\text{NO}_2^\bullet$ para formar N_2O_5 dentro de um equilíbrio dependente da temperatura, cuja queda favorece o deslocamento para a formação de N_2O_5 . Esse processo constitui-se em sumidouro para o $\text{NO}_3^\bullet$ somente perante a existência de sumidouro para o N_2O_5 (Dimitroulopoulou e Marsh, 1997),

4) reação de N_2O_5 com H_2O para formar 2 HNO_3 em fase gasosa, aquosa, ou em superfícies de aerossóis, ou ainda deposição seca no solo (Dimitroulopoulou e Marsh, 1997). Medições para $\text{NO}_3^\bullet$ (Plat *et alli*, 1981, 1984) mostraram decréscimo no tempo de vida do $\text{NO}_3^\bullet$ com o aumento da umidade relativa. Assim, uma queda nos níveis de $\text{NO}_3^\bullet$ implica em remoção de N_2O_5 por meio de reação com H_2O , indicando que essa reação atua como mecanismo de perda de NO_x e como caminho para a formação de HNO_3 (Dimitroulopoulou e Marsh, 1997),

5) reação entre o radical $\text{NO}_3^\bullet$ e o radical $\text{HO}_2^\bullet$ para formar radicais hidroxila, $\text{HO}^\bullet$, (Mellouki *et alli*, 1988, 1993; Plat *et alli*, 1990).

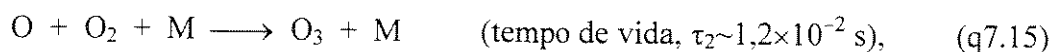
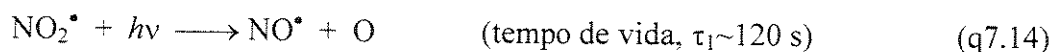
A segunda etapa acima indica que no primeiro período (~20:00 até ~6:00 h) diminui a remoção de $\text{NO}^\bullet$ pelo O_3 . Por causa da ausência de luz solar, tem início a reação indicada na primeira etapa para formar HONO. O efeito dessa combinação causa menor consumo de $\text{NO}^\bullet$ como mostram as inclinações das curvas PC_3 após 15:00 h e PC_2 após 20:00 h na Figura 6.60a (comparar a inclinação da curva PC_3 em queda, com aquela para a queda em PC_2 , nos respectivos horários). Desse modo, cai o consumo de $\text{NO}^\bullet$, o que provoca aumento em sua acumulação à despeito da possibilidade de reação com O_3 , a espécie oxidante noturna mais importante depois do $\text{NO}_3^\bullet$ (Dimitroulopoulou e Marsh, 1997). O aumento na acumulação de $\text{NO}^\bullet$ pode ser visto na Figura 6.84b onde, no referido período, o perfil de acumulação torna-se menos negativo.

No segundo período (~6:00 h até 10:00 h), uma combinação de dois entre vários fatores determina o comportamento do $\text{NO}^\bullet$. Primeiramente, na presença de luz solar, o HONO formado à noite é decomposto formando $\text{HO}^\bullet$ e $\text{NO}^\bullet$ (Jacob, 2000). Segundo, tem início o tráfego de veículos com emissões ao ar de seus compostos característicos. O resultado disto traduz-se como um consumo de $\text{NO}^\bullet$ por $\text{HO}_2^\bullet$ (reação q7.4, p372), combinado com o consumo de vários outros hidrocarbonetos pelo $\text{HO}^\bullet$ (formado nas primeira e quinta etapas), o qual regenera mais radicais $\text{HO}_2^\bullet$ para reagir com o $\text{NO}^\bullet$. Esse processo faz cair a acumulação de $\text{NO}^\bullet$ como pode ser visto na Figura 7.10b, na qual o perfil de acumulação torna-se mais negativo.

No terceiro período (~10:00 h até ~14:00 h), as emissões começam a cair provocando uma queda na efetividade das reações acima, causando queda no consumo de $\text{NO}^\bullet$. Desse modo, como mostra a Figura 7.10b, onde o perfil de acumulação torna-se menos negativo, aumenta a acumulação de $\text{NO}^\bullet$.

No último período (~14:00 h até ~20:00 h), as emissões devidas ao tráfego de veículos começam a aumentar novamente. Nesse período, as concentrações de O_3 atingem valores elevados o suficiente para causar, junto com os radicais peróxido, um rápido consumo de $\text{NO}^\bullet$. Isto pode ser visto na Figura 6.84b, na qual o perfil de acumulação torna-se mais negativo, indicando um decréscimo na acumulação de $\text{NO}^\bullet$.

As Figuras 7.10c (soma dos efeitos de PC₂ e PC₃, pesados por NO₂[•], p376) e 7.10d (soma dos efeitos de PC₂ e PC₃, pesados por O₃, p376) mostram, respectivamente, os perfis de acumulação de NO₂[•] e de O₃. Essas componentes são muito correlacionadas. Isto deve ser atribuído à ocorrência das seguintes reações (Birks, 1998, p.236-237; Jacob, 2000):



Assim, pode-se considerar que a fotólise do NO₂[•] resulta instantaneamente na produção de O₃. Assim, seus perfis de acumulação são muito parecidos (Figuras 7.10c e 7.10d, p376). A queda na acumulação de NO₂[•] nas primeiras horas do dia (antes das 6:00 h) deve ser atribuída à reações de formação de PANs (Baird, 1995, p.132; Birks, 1998, p.243; Jacob, 2000) e de HNO₃ (Dimitroulopoulou e Marsh, 1997; ver item 4 na p378).

O início do crescimento de NO₂[•] após 7:00 h é devido à reação q7.10, isto é, a reação entre HO₂[•] e NO[•] para formar NO₂[•]. Como já mencionado antes, após as 15:00 h, a queda de O₃ é devida principalmente à sua reação com NO[•]. Por outro lado, a queda de NO₂[•] na parte da tarde (após 15:00 h) como ocorre também com o O₃, deve-se à reação de auto destruição quando as concentrações de ambos, NO₂[•] e O₃, são simultaneamente altas (Baird, 1995, p.127).

A maior importância do NO₂[•] está no fato de que a sua fotólise é a única fonte capaz de levar à formação de ozônio troposférico. Nenhum outro composto pode sofrer fotólise a comprimentos de onda maiores do que 320 nm para formar os átomos de oxigênio necessários à sua produção (Madronich, 1993; Pisano *et alli*, 1997).

O primeiro pico do O₃ (~3:00-4:00 h) obtido em todos os casos analisados nessa tese não parece estar relacionado com inversão noturna ou de radiação, pois isto exigiria inversão para o necessário deslocamento vertical descendente. Isto é, considerando-se uma massa de ar a uma dada altitude contendo ozônio e o ar não saturado com vapor d'água (processo *adiabático seco*), a hipótese de que algum O₃ pudesse descer implicaria em elevação da temperatura durante a descida (Boubel *et alli*, 1994, p.252-253), fato não ocorrido como mostram as Figuras 6.5a (p251), 6.11a (p261). Nelas e em todas os casos

não há indícios de aumento da temperatura no referido período (~3:00-4:00 h). Então a formação do O₃ precisa ser explicada em termos de uma fonte noturna de oxigênio atômico. Durante o dia, a luz solar se encarrega de fornecê-lo por meio da reação q7.14, porém à noite, portanto na ausência de luz solar, essa reação não ocorre por aquela via. Desse modo, é necessário encontrar alguma fonte alternativa de oxigênio atômico capaz de fornecer essa espécie química necessária à produção noturna de O₃.

Uma delas poderia ser a decomposição de NO₂[•] causada por ftons provenientes da iluminação pública. Outra seria a decomposição de 1-octeno a qual gera oxigênio atômico (Paulson e Seinfeld, 1992). Entretanto, as duas possibilidades acima podem apresentar restrições. A primeira depende de haver iluminação local, e a segunda, depende da emissão do referido composto, o qual poderia ser evitado nos processos que o produzem. Assim, procura-se uma fonte que esteja intrinsecamente relacionada com a emissão de NO_x. Isto é, que seja dependente das próprias emissões de óxidos de nitrogênio, quer de fontes não naturais, tais como a automotiva e a de indústrias, ou de fontes naturais.

Essa possibilidade remete ao estudo da baixa troposfera noturna, a qual por sua vez, remete às pesquisas que procuram explicar persistência do ozônio na baixa troposfera.

Medições sobre o perfil vertical (desde o solo até 800 m) para o ozônio e variáveis meteorológicas (temperatura, pressão, umidade, velocidade e direção do vento) no vale superior de Rhine (Upper Rhine Valley) próximo a Strasbourg (Alemanha), em balão com ascendência e descendência controladas do solo, mostram a existência de inversão noturna, cuja altura pode ser determinada (Güsten *et alli*, 1998). Adicionalmente, mostram que, dentro da camada de inversão noturna, os níveis de ozônio decrescem rapidamente durante a noite por deposição seca no solo, e por reações químicas com óxidos de nitrogênio (Güsten *et alli*, 1998). Seus resultados mostram que os níveis verticalmente crescentes de O₃ a partir do solo, correlacionam-se positivamente com a temperatura e negativamente com a umidade. Mostram também o início de inversão noturna de superfície ~3:00 h, seu desaparecimento à tarde, e restabelecimento à noite (~20:00 h).

Não obstante a ausência de medições entre 20:00 h–3:00 h, concluem que a inversão de superfície não se altera no período noturno. Assim, atribuem a perda química às reações: $O_3 + NO^{\bullet} \rightarrow NO_2^{\bullet} + O_2$; e $NO_2^{\bullet} + O_3 \rightarrow NO_3^{\bullet} + O_2$. Até 50-70% da concentração máxima de O₃ no dia seguinte pode se originar da camada noturna residual (Neu *et alli*, 1994).

Pisano *et alli* (1997) utilizando o mesmo tipo de balão, mediram concentrações verticais (desde o solo até 1000 m) de NO_2^* e O_3 (além de temperatura e umidade) no vale inferior de Fraser (Lower Fraser Valley). Seus resultados também mostram aumento da concentração de O_3 com a altitude no período entre ~7:30 h e ~18:00 h. Por outro lado, os níveis para o NO_2^* apresentaram, após ~9:30 h, crescimento desde nível do solo, com altura da camada constante até uma dada altitude, após a qual ocorria a inversão com os níveis caindo a ~ zero. Esse comportamento prosseguiu até ~11:30 h, quando os níveis de concentração caíram, porém mantendo um valor constante com a altitude, para apresentar indícios de retorno ao ciclo ~18:00 h. Os perfis para a temperatura, após ~9:30 h, apresentaram sempre valores decrescentes com a altitude, portanto em correlação negativa com os níveis para o O_3 , contrariamente ao encontrado por Güsten *et alli* (1998).

Afirmaram também a ocorrência consumo de O_3 por NO^* para formar NO_2^* (reação conhecida como titulação de O_3).

Observando-se as figuras com os perfis para a temperatura nada impede que ao invés de se afirmar que os valores sempre decresceram com a altitude, seja afirmado que os valores para a temperatura cresceram com o decréscimo da altitude. Isto é correto, porque uma massa de ar se aquece quando desce (Boubel, 1994, p.252-253).

Em resumo, ambas as pesquisas consideram que parte do O_3 é retirado da troposfera por deposição seca e parte por reações químicas com óxidos de nitrogênio. Além disso, de acordo com Güsten *et alli* (1998), quando a inversão noturna é rompida nas primeiras horas da manhã devido ao aquecimento da superfície (do solo), movimento descendente rápido de mistura restaura os altos níveis de ozônio típicos do dia anterior.

Deve ser enfatizado que, não obstante os dados usados nessa tese refiram-se a concentrações registradas no nível solo, boa parte das reações que ocorrem nessa região deve, provavelmente, ocorrer também em maiores altitudes na baixa troposfera.

Os resultados encontrados para o ozônio nessa tese não são contraditórios com aqueles obtidos por Güsten *et alli* (1998) e Pisano *et alli* (1997) no que diz respeito à perda de O_3 em reações com NO^* . Os perfis encontrados por meio das análises PCA e dos modelos multimodo (PARAFAC e Tucker), mostram isto. Entretanto os resultados vão além, pois mostram um pequeno pico para o O_3 (3:00 h–4:00 h) que não pode ser explicado satisfatoriamente por meio de inversão noturna pois isto implicaria em aumento da

temperatura da massa de ar descendente contendo o poluente, fato não observado nos resultados encontrados para os casos sob investigação nessa tese. Isto está de acordo com a afirmação de que difusão descendente de massas de ar ricas em ozônio, são inibidas durante a noite devido à estabilidade atmosférica (Güsten *et alli*, 1998).

Para dar prosseguimento à interpretação dos resultados obtidos para o ozônio algumas considerações devem ser levantadas:

1) A reação de formação de O_3 (q7.12, p372) é exotérmica. Ela é a mesma que ocorre na estratosfera (Madronich, 1993), a qual gera (junto com a fotólise do O_3) calor suficiente para determinar a temperatura nessa região da atmosfera (Baird, 1995, p.28). Considerando-se que uma concentração de O_3 da ordem de 8 ppm, acarreta uma variação na temperatura em cerca de $50\text{ }^{\circ}\text{C}$ (Baird, 1995, p.17), o mesmo efeito na troposfera, à atmosferas com concentrações da ordem de 80 ppb, provocaria um aumento de aproximadamente $0,1\text{ }^{\circ}\text{C}$ na temperatura local.

Desse modo, as concentrações de ozônio encontradas usualmente hoje na baixa troposfera são oriundas de quantidades tais que não geram calor suficiente para causar aumentos significativos na temperatura nessa região da atmosfera. Nessa tese, os resultados encontrados para os perfis da temperatura em todas as localidades estudadas mostram que o efeito de aquecimento direto da atmosfera provém de uma única fonte: a radiação solar.

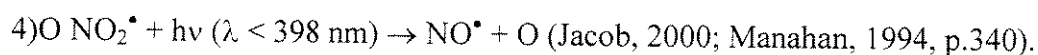
2) No início da noite, quando boa parte da concentração de $NO^{\bullet}$ cai devido ao seu consumo pelo O_3 , um excesso de $NO_2^{\bullet}$ (que se acumula na ausência de luz solar) pode reagir com o O_3 para formar o radical $NO_3^{\bullet}$ ($NO_2^{\bullet} + O_3 \rightarrow NO_3^{\bullet} + O_2$) sendo esta a principal via para a formação do mesmo (Dimitroulopoulou e Marsh, 1997).

Durante o dia o tempo de vida do $NO_3^{\bullet}$ é limitado por meio de fotólise de retorno a $NO_2^{\bullet}$ (caminho favorecido) ou $NO^{\bullet}$, a comprimentos de onda menores do que 700 nm. O outro processo importante (Dimitroulopoulou e Marsh, 1997) é a reação com $NO^{\bullet}$ dando origem a 2 $NO_2^{\bullet}$ ($NO_3^{\bullet} + NO^{\bullet} \rightarrow 2 NO_2^{\bullet}$). Com o sol “à pino”, o tempo de vida do $NO_3^{\bullet}$ é de $\sim 5\text{ s}$ (Wayne *et alli*, 1990).

Observe-se que essa segunda via mais importante de consumo diurno de $NO_3^{\bullet}$ dá origem a dois $NO_2^{\bullet}$ por $NO_3^{\bullet}$ que reage.

3) Uma outra participação importante do radical $\text{NO}_3^\bullet$ é a sua reação com $\text{NO}_2^\bullet$ para formar N_2O_5 , o qual, em reação com H_2O , origina 2 HNO_3 , constituindo-se este como a via de perda de $\text{NO}_3^\bullet$ quando a umidade relativa aumenta (Dimitroulopoulou e Marsh, 1997).

Das considerações acima depreende-se que o radical $\text{NO}_3^\bullet$ desempenha um papel importante na troposfera noturna. Entretanto, essa espécie química parece estar envolvida em outras reações durante a noite, como mostram os processos abaixo:



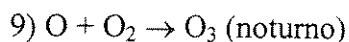
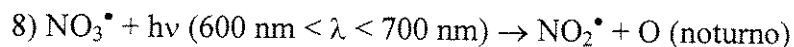
5) O $\text{NO}_2^\bullet$ também pode interagir com fotons de comprimentos de onda maiores do que 430 nm dando origem a espécies NO_2^* excitadas: $\text{NO}_2^\bullet + h\nu (\lambda > 430 \text{ nm}) \rightarrow \text{NO}_2^*$ (Manahan, 1994, p.340).

6) O NO_2^* também pode ser formado em processos que recebem a denominação de *quimiluminescência*, como a reação: $\text{O}_3 + \text{NO}^\bullet \rightarrow \text{NO}_2^* + \text{O}_2$ (Manahan, 1994, p.283-285).

7) As espécies excitadas NO_2^* podem perder o excesso de energia por meio de vários processos diferentes. Um deles, denominado *luminescência*, implica no retorno à espécie $\text{NO}_2^\bullet$ e em emissão de energia: $\text{NO}_2^* \rightarrow h\nu (\lambda > 398 \text{ nm}) + \text{NO}_2^\bullet$. A intensidade da luz emitida situa-se entre 600 e 3000 nm, dependendo da concentração de $\text{NO}_2^\bullet$ (Manahan, 1994, p.730).

Considerando-se que os processos 5, 6 e 7 podem funcionar como mecanismos de “retro-alimentação” de espécies NO_2^* , se apenas 10% (aproximadamente a relação entre os dois picos de ozônio nos resultados obtidos nessa tese) de todo o $\text{NO}_3^\bullet$ formado levar, efetivamente, à formação de oxigênio atômico, obtém-se a procurada fonte.

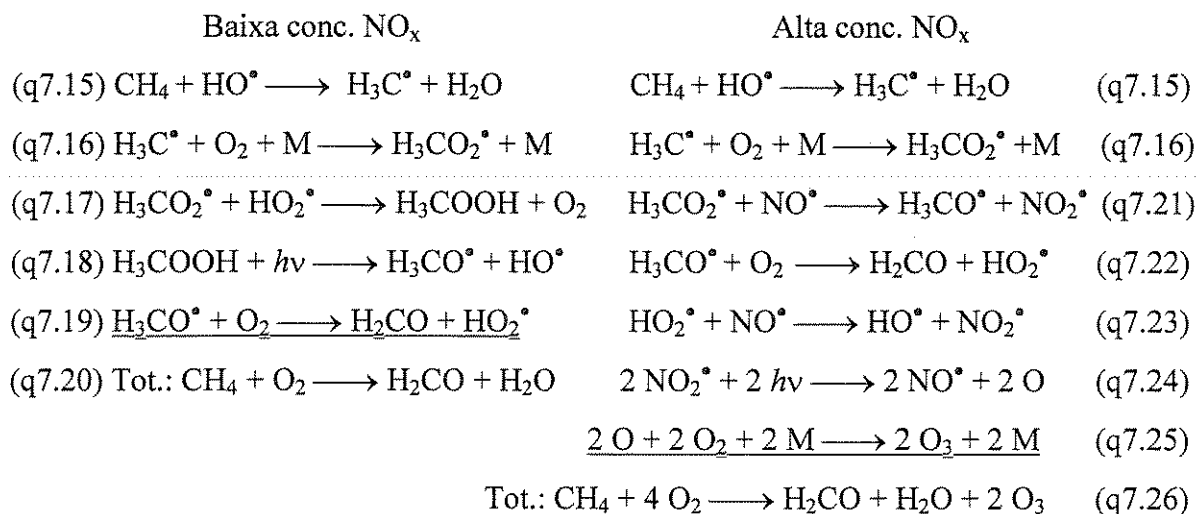
Do exposto acima abre-se uma possibilidade de formação de oxigênio atômico e, conseqüentemente, formação de ozônio, por meio dos processos noturnos:



Nesse ponto, deve-se chamar a atenção para o fato de que essa possibilidade possui cunho apenas qualitativo. Ela depende de vários fatores cujas informações até aqui não permitem que se assegure sem questionamentos, a sua viabilidade. Por outro lado ela pode explicar a formação de ozônio noturno independentemente da ocorrência de inversão.

6.5.2 – Ciclo do Metano

O ciclo do metano, como já visto, fica constituído pela introdução do metano em substituição ao monóxido de carbono no ciclo desse último. Da mesma forma que no caso do CO, o ciclo também deve se estudado sob condições de baixa e de alta concentração de óxidos de nitrogênio na atmosfera. As reações que ocorrem em cada uma das situações (baixa ou alta concentração de NO_x) estão representadas abaixo (Birks, 1998, p.236):



Semelhantemente ao que acontece no ciclo do monóxido de carbono, também nesse caso, existe um ponto no qual a seqüência de reações conduz à destruição ou à produção de O₃. Pelas reações acima, nota-se que HO₂[•] e NO[•] competem pelo H₃CO₂[•] como mostram as equações q7.17 e q7.21. Sob baixa concentração de NO_x prevalece a reação representada pela equação q7.17 levando à formação de água e formaldeído como mostra a equação para a reação total q7.20. Nesse caso, embora a destruição de ozônio não esteja evidente nas equações sob baixa concentração de NO_x, subsequente decomposição fotoquímica do aldeído, sob tais condições, provoca a destruição de O₃ (Birks, 1998, p.236).

A seqüência de reações que pode conduzir à destruição do ozônio é mostrada na Figura 7.11 adiante. Na verdade, a decomposição fotoquímica do aldeído só pode levar à destruição direta de ozônio por uma única via: reação deste com HO₂[•] (via nº 1) como mostra a Figura 7.11. Entretanto, sob tais condições (baixa concentração de NO_x) a concentração de O₃ deve ser muito pequena, de modo que essa via não deve ser a mais provável. Por outro lado, cada molécula do aldeído que sofre decomposição forma dois

radicais $\text{HO}_2^\bullet$ e uma molécula de CO. Esses radicais $\text{HO}_2^\bullet$ são altamente reativos e podem reagir mutuamente formando H_2O_2 , a qual se decompõe por ação da luz dando origem a dois radicais $\text{HO}^\bullet$ (via nº 2). A partir daí estão dadas as condições para a formação do ciclo do CO independentemente de existirem emissões diretas de alguma fonte desse gás. As reações do ciclo (sob as condições dadas acima) estão assinaladas na Seção 7.2.

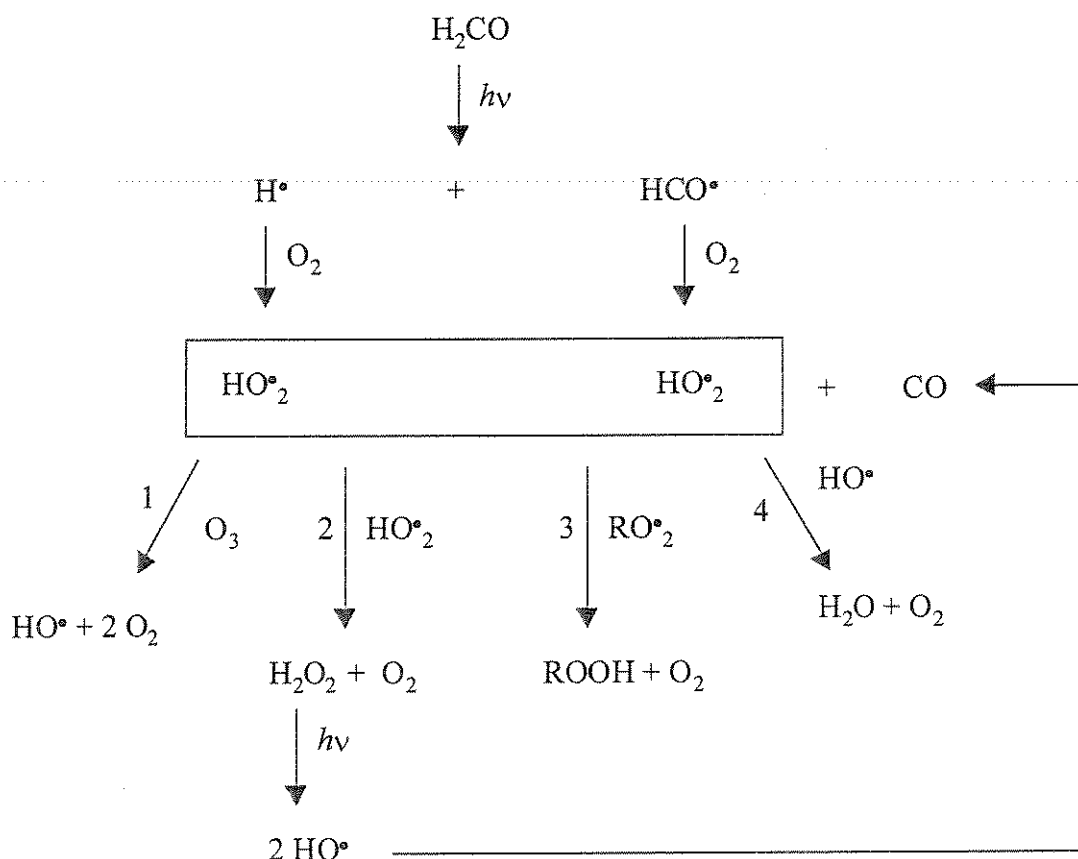


Fig.: 7.11: As possíveis vias de destruição de O_3 , com início no ciclo do metano.

A Figura 7.11 mostra as possibilidades de destruição de ozônio a partir do ciclo do metano sob condições de baixa concentração de óxidos de nitrogênio. A via mais provável, nas condições dadas parece ser a da reação mútua de radicais $\text{HO}_2^\bullet$, criando as condições para que se estabeleça o ciclo do CO, por meio do qual o O_3 pode ser destruído na atmosfera (sob baixas concentrações de NO_x).

É importante observar que esta deve ser, de fato, a via mais provável porque, por meio dela, sobra um radical $\text{HO}^\bullet$, o qual penetra novamente no ciclo do metano, consumindo-o e gerando outra molécula de aldeído a qual é decomposta, e assim sucessivamente. A razão para se acreditar que este caminho é o mais provável está assentada em dados históricos.

Antes de 1750 a concentração de metano era constante no valor de 0,75 ppm (Baird, 1995, p.168-169; EPA, 2003), mas nos anos 1980 ela cresceu aproximadamente 0,6 % ao ano, alcançando 1,7 ppm (Baird, 1995, p.168-169; EPA, 2003; Akimoto *et alli*, 2001). Nos anos 1970 a taxa de crescimento era quase duas vezes tão grande quanto esta, e não se sabe ao certo porque a taxa decresceu e hoje é aproximadamente igual a zero (Baird, 1995, p.168-169). Ainda de acordo com Baird (1995, p.123), “A etapa inicial (abstração pelo radical $\text{HO}^\bullet$ de um átomo de hidrogênio do metano) é um processo relativamente lento necessitando, em média, uma década para ocorrer. Uma vez concretizada, entretanto, as etapas subseqüentes que levam à formação de aldeído ocorrem muito rapidamente. A lentidão da etapa inicial na oxidação do metano e as crescentes quantidades do gás liberadas para a superfície da Terra, têm resultado em um aumento na concentração atmosférica de CH_4 nos tempos recentes”. Cabe aqui a pergunta. Se a taxa decresceu e hoje é aproximadamente igual a zero, como existe um aumento crescente na concentração atmosférica de CH_4 ultimamente?

Observe-se que o autor refere-se a taxa de crescimento das concentrações de metano e não de taxa de emissão do gás. São processos diferentes porque as concentrações podem aumentar mesmo com taxa de emissão constante, bastando para isto que o consumo se dê a uma taxa inferior à de emissão.

Continuando, Baird (1995, p.171) acrescenta, “Alguns cientistas especulam que a concentração de $\text{HO}^\bullet$ na atmosfera pode estar diminuindo por causa das concentrações crescentes de CO e CH_4 , (...), assim algum aumento observado na concentração de metano pode ser atribuído a uma redução na taxa de sua reação com o radical hidroxila.”

Se não existisse a era atual ou se o mundo estivesse vivendo a sua era pré-industrial, as emissões seriam pequenas e a concentração de metano estaria na faixa de 0.75 ppm. Isto pode ser constatado pelas reações do ciclo do CH_4 (a oxidação do metano vai gerar dois radicais hidroxila e um CO. Um dos radicais vai consumir CO e o outro vai consumir outro

metano emitido para a atmosfera). Nesse processo também é consumido um ozônio, proveniente, por exemplo da estratosfera, fechando, portanto, o ciclo. Esse quadro é o de um estado estacionário (ou quase estacionário).

Sob baixa concentração de NO_x deve prevalecer um estado estacionário como mencionado acima. Assim, se for considerado apenas o ciclo do metano não procederia a especulação de que estaria caindo a concentração do radical hidroxila na atmosfera por causa de aumentos nas concentrações de CH_4 e de CO . Isto porque a oxidação do metano gera, ela própria, dois radicais hidroxila.

Por outro lado, o interesse no metano está circunscrito ao seu potencial como “gás de efeito estufa” (Hansen, 2001; American Geophysical Union, 2002; Smith e Bogner, 1997; Madronich, 1993). Desse modo várias pesquisas com enfoques distintos têm sido desenvolvidas com o objetivo de identificar os mecanismos que conduzem à acumulação de metano, ozônio, e de outros compostos com efeito semelhante na atmosfera.

Alguns pesquisadores “elegem” uma classe de substâncias e se dedicam a estudar o seu comportamento sob determinadas condições na atmosfera (Hansen *et alli*, 2000; Rind, 2000; Hansen, 2001; Betts, 2000; Cox *et alli*, 2000). Outros concentram sua atenção em possíveis alterações na formação de nuvens e sua interação com a radiação solar (Farrar, 2000; Christensen, 2000; Marsh e Svensmark, 2000).

Algumas pesquisas avaliam a ocorrência de mudanças climáticas causadas por agentes precursores poluentes na troposfera (Fuglestad *et alli*, 2003), enquanto outras avaliam o próprio processo de medição como meio para a convergência de resultados mais acurados (Ladstatter-Weissenmayer *et alli*, 2003).

Uma outra classe de pesquisas concentra a atenção em mudanças na química da troposfera em função de emissores específicos e de suas respostas à radiação solar (Madronich, 1993). Outras pesquisas concentram-se nas relações entre as emissões com o objetivo de explicar variações nas concentrações de certos poluentes na troposfera (Akimoto *et alli*, 2001).

Pesquisas mais recentes procuram explicações para a resposta climática às emissões por meio da “força ou potência radiativa” de alguns “gases de efeito estufa” (Manne e Richels, 2001; Hansen *et alli*, 2000; Hansen, 2000).

Alguns pesquisadores baseiam-se em fontes específicas de emissão de poluentes para avaliar o comportamento da troposfera à luz das variações das concentrações desses compostos (Smith e Bogner, 1997).

Para o caso do metano, os resultados encontrados para a localidade Pedro II nessa tese indicam que sob certas condições, esse poluente tem participação na acumulação de ozônio em nível local. Tais resultados, obtidos por meio das análises PCA e multimodo, mostram ser o CH_4 proveniente de outras fontes que não a automotiva. Desse modo, um esforço foi despendido no sentido de se explicar as variações sistemáticas diárias e semanais para o metano.

As emissões de metano registradas no P. D. Pedro II devem ser provenientes de fontes tais como, estações de tratamento de esgotos domésticos e industriais, e de depósitos de resíduos sólidos.

Assim, voltando a analisar o ciclo do metano, desta vez, sob condições de alta concentração de NO_x , verifica-se que a rota se modifica a partir da equação q7.21. Aqui, o radical $\text{H}_3\text{CO}_2^\bullet$ reage com $\text{NO}^\bullet$, de modo que, ao final do ciclo, ocorre a formação de ozônio, como mostra a equação total q7.26. As etapas até a formação do aldeído são representadas pelas equações q7.15 a q7.16 e q7.21 a q7.26 (p385).

Todavia, no Capítulo II foi discutida a oxidação de hidrocarbonetos para o caso da atmosfera poluída. Para o metano, tal processo está representado pelo conjunto de reações q2.22 a q2.36, indicando, ao final, a formação de seis moléculas de O_3 na atmosfera.

A comparação das reações químicas no ciclo do metano com os resultados obtidos com os modelos quimiométricos mostra consistência entre si. Para a Estação de Medição Parque D. Pedro II, a Figura 6.37 (p307) mostra que de zero hora até aproximadamente 7:00 h os loadings para o CH_4 possuem valores elevados. A partir daí ocorre uma rápida queda atingindo um valor mínimo às 15:00 h e, passando a crescer continuamente na medida em que se aproxima novamente das 24:00 h. É necessário saber que ou quais fatores estariam causando a queda nas concentrações de metano naquele período. O perfil de emissão da fonte (Figura 6.37, p307) mostra que o metano é produzido durante a noite, mas essa produção cai durante o dia. A queda na emissão deve estar associada tanto com a emissão propriamente dita quanto com o consumo do gás por alguma espécie química (naturalmente o $\text{HO}^\bullet$) quando este é liberado para a atmosfera.

Do ponto de vista da fonte, o metano deve ser formado pela via anaeróbia. De fato, relatório recente da CETESB (2002) apresentando estimativas das emissões de metano decorrentes da disposição de resíduos sólidos e do tratamento de águas residuárias no Brasil (1990-1994), com base nas Diretrizes Revisadas de 1996 do Painel Intergovernamental Sobre Mudança do Clima – IPCC, revela que as duas maiores fontes de produção de metano pela via anaeróbia são os aterros de lixo e o tratamento de esgoto e águas residuárias.

Os tratamentos de águas residuárias e de efluentes da indústria têm seguido a tendência dos últimos anos para a utilização de reatores anaeróbios, cuja maioria no Brasil é do tipo UASB (CETESB, 2002). Embora esses sistemas sejam indicados para a recuperação eficiente do metano para fins energéticos, grande quantidade do gás é liberada para a atmosfera (CETESB, 2002), contribuindo para o agravamento do quadro climático local e regional.

No caso dos resíduos sólidos, a disposição e tratamento no país se distribui da seguinte forma: 76,5% depositados em “lixões” a céu aberto, 21,8% em aterros controlados e aterros sanitários, e 2,1% têm outros destinos (IBGE, 1992) tais como a compostagem e a incineração (CETESB, 2002). No Estado de São Paulo, por exemplo, a CETESB (2002) avalia que as condições gerais dos locais de disposição de resíduos sólidos até 1994 eram precárias.

Em função dos dados acima pode-se afirmar que boa parte do metano emitido para a atmosfera na RMSP é proveniente de processos anaeróbios. Além disto, os “lixões” devem ser os principais responsáveis pela parcela mais significativa dessas emissões. Assim, nessa tese, o enfoque é dirigido para via anaeróbia de produção de metano, restringindo-o em alguma medida ao tratamento de esgotos e, em grande medida à disposição de resíduos sólidos.

Metano biogênico é um termo usado para descrever o processo de obtenção do gás natural oriundo da redução de CO₂ por meio de processos biogeoquímicos (Rice e Claypoal, 1981). Processos geoquímicos têm sido identificados ocorrendo em conjunto com os processos biológicos os quais incluem metanogênese e hidrogenogênese como os determinantes na obtenção do metano (Schoell *et alli*, 2001). Em ambientes naturais a produção de metano ocorre em pântanos, solo, sedimentos de rios, lagos e mares, assim

como nos aparelhos digestivos de animais ruminantes (Chernicharo, 1997, vol.5, p.23). No primeiro estágio da digestão anaeróbia de compostos orgânicos complexos, um grupo de bactérias facultativas e anaeróbias (formadoras de ácidos ou fermentativas) convertem aqueles (carboidratos, proteínas e lipídios) em materiais orgânicos mais simples, principalmente ácidos voláteis. No segundo estágio, ocorre a conversão dos ácidos orgânicos, dióxido de carbono e hidrogênio em metano e dióxido de carbono, por um grupo especial de bactérias estritamente anaeróbias denominadas formadoras de metano ou metanogênicas (Chernicharo, 1997, vol.5, p.23; Snape *et alli*, 1995, p.350-372; Micales e Skog, 1997; Noble e Henk, 1998).

Dos três importantes grupos de microrganismos (as bactérias fermentativas, as acetogênicas e as metanogênicas) que participam da decomposição anaeróbia, o último grupo (metanogênicas), constitui dois sub-grupos, um dos quais (metanogênicas hidrogenotróficas) degrada o acetato transformando-o em metano e dióxido de carbono, enquanto o outro (metanogênicas acetoclásticas) produz metano por meio da redução do dióxido de carbono (Chernicharo, 1997, vol.5, p.24-25).

Considerando-se que parte expressiva do metano gerado na RMSP ocorre por processos anaeróbios, tomar-se-á como base os parâmetros utilizados em unidades de tratamento controladas (reatores anaeróbios). Considerando-se também que parte expressiva do metano produzido na localidade Pedro II é proveniente de “lixões”, tomar-se-á também como base critérios relacionados com esse tipo de descarte de rejeitos.

Vários fatores influenciam a decomposição em aterros de lixo, os quais incluem: a) manejo e composição do rejeito; b) fatores que influenciam o crescimento bacteriano (mistura, disponibilidade de nutrientes, pH e temperatura); c) planejamento (tipo de tratamento dos gases formados); d) operação do aterro (Micales e Skog, 1997).

Considerando-se as características dos aterros no Brasil, apenas os fatores que influenciam o crescimento bacteriano contribuem para os níveis de CH₄ encontrados na localidade sob investigação.

A cinética da digestão anaeróbia pode ser considerada como um processo em três etapas, as quais envolvem: 1) hidrólise de compostos orgânicos complexos; 2) produção de ácidos; e 3) produção de metano (Chernicharo, 1997, vol.5, p.45). As variações sistemáticas diárias encontradas para o CH₄ no P. D. Pedro II (Figura 6.37, p.307) sugerem uma

avaliação da cinética de crescimento e decaimento de microrganismos, especialmente a inibição do crescimento. Isto remete para os requisitos ambientais visto que não há acompanhamento das reações que se desenvolvem nos aterros. Nesse aspecto, a temperatura parece exercer papel fundamental na cinética de degradação anaeróbia dos materiais encontrados nos “lixões”.

Dos fatores físicos que afetam o crescimento microbiano, a temperatura ambiente externa é um dos mais importantes na seleção das espécies, já que os microrganismos não possuem mecanismos de controle de sua própria temperatura (Chernicharo, 1997, vol.5, p.61). Três faixas de temperatura estão associadas ao crescimento microbiano na maioria dos processos biológicos (Lettinga *et alli*, 1996, apud Chernicharo, 1997; Manahan, 1994, p.150):

- bactérias psicrófilas: entre 0 e 20 °C;
- bactérias mesófilas: entre 20 e ~45 °C; e
- bactérias termófilas: entre 45 e 70 °C, ou mais

Sob o ponto de vista dos efeitos da temperatura sobre os parâmetros cinéticos, os dados indicam que a constante de saturação, K_s , e o coeficiente de produção de biomassa, Y , decrescem com o aumento da temperatura, enquanto o coeficiente de respiração endógena, K_d , de culturas de baixo crescimento é pouco afetado pela temperatura (Grady e Lim, 1980, apud Chernicharo, 1997). A quantificação dos efeitos da temperatura sobre as reações bioquímicas é usualmente realizada por meio da equação de Arrhenius:

$$K = K_0 \text{EXP}(-E/RT), \quad (i)$$

onde K é a taxa de reação, K_0 uma constante, E a energia de ativação, R a constante dos gases, e T a temperatura absoluta (Chernicharo, 1997, vol.5, p.63).

A equação acima pode ser expressa em termos da taxa máxima de crescimento específico, $\mu_{\text{máx}}$, em função da temperatura, a qual deve descrever a síntese e o decaimento bacteriano. Assim obtém-se a equação:

$$K_{\text{liq}} = K_1 \text{EXP}(-E_1/RT) - K_2 \text{EXP}(-E_2/RT), \quad (ii)$$

onde K_{liq} representa a taxa líquida de crescimento, K_1 a taxa de síntese bacteriana, e K_2 a taxa de decaimento bacteriano. A equação representa, então, a competição entre dois processos: o de síntese bacteriana e o de decaimento bacteriano (Chernicharo, 1997, vol.5, p.63-64).

Observações experimentais indicam que a taxa máxima de crescimento bacteriano decresce 11% por °C para digestores anaeróbios operados à temperaturas abaixo de 30 °C (Henze e Harremões, 1983, apud Chernicharo, 1997).

No caso dos aterros de lixo à céu aberto, nenhum tipo de controle sobre os processos anaeróbios é exercido, tal que são as próprias condições ambientais que determinam a cinética da digestão anaeróbia. Um outro aspecto a considerar é que a carga de resíduos que abastece o aterro não é de fluxo contínuo no curto prazo, isto é, para efeito do perfil diário do metano produzido localmente. Em geral a carga é adicionada ao final do dia, ou durante parte da noite, e em muitos casos não existe seletividade nos materiais recolhidos após a operação de coleta. Isto significa uma grande variedade de substratos a cada dia entregues aos microrganismos para o processo de decomposição.

Um outro fator a ser considerado é que pode se estabelecer um mecanismo de competição entre espécies, típico de um processo de fluxo biológico contínuo, o qual interfere na dinâmica da população. Esse tipo de interação, *substrato-presa-predador*, pode ser formulado também em termos de cinética de reação denominada cinética de interação microbiológica. Nesses casos, uma das espécies (a de menor taxa de crescimento específico, μ) é, após algum tempo lixiviada e sai do processo. Reatores anaeróbios podem operar, teoricamente, desse modo, quando a taxa de diluição, D , se iguala a taxa de interseção μ_{int} , situação na qual ambas as espécies podem coexistir no reator. Sabe-se que o sistema *presa-predador* exhibe condições oscilatórias estáveis. Esta é, entretanto, uma situação metaestável que não pode existir no reator se $D < \mu_{int}$ ou $D > \mu_{int}$, situações nas quais uma ou outra espécie é, respectivamente, lixiviada (Snape, 1997, p.373-377).

No caso dos aterros de lixo, parece ser possível o estabelecimento desse tipo de interação, na qual a variação da concentração do substrato com o tempo assumia oscilações com amplitude constante (o que acarretaria em oscilações estáveis), em função do carregamento intermitente característico dos “lixões”, diferentemente do que ocorre nos reatores controlados.

Da discussão acima pode-se inferir que os procedimentos envolvidos no problema da disposição de resíduos sólidos, devem interferir na cinética de degradação de várias maneiras tal que a produção do metano não ocorre modo uniforme ao longo do dia. Talvez ocorra inclusive a transição de um tipo de microrganismo para outro em função do processo intermitente de carregamento e da variação de temperatura durante o dia.

Sob tais condições, a dinâmica existente nos aterros de lixo deve passar pelas seguintes etapas:

1) desenvolvimento de microrganismo anaeróbios nas camadas inferiores do aterro, em regiões com temperaturas mais baixas, cujo período noturno favoreça o crescimento de bactérias psicrófilas,

2) ao amanhecer, o aumento da temperatura externa pela incidência de radiação solar passa a inibir o crescimento microbiano (equação *ii* acima), causando queda na efetividade das reações que levam à formação de metano,

3) ao final da tarde, quando a temperatura volta a cair devido a perda radiativa da superfície, a cultura microbiana passa a recuperar a taxa de crescimento tornando novamente efetivas as reações de formação de metano.

É importante reconhecer que se o sistema entrar em colapso (os dois termos do segundo membro da equação *ii* se igualarem ou ficarem próximos demais), causando a morte das bactérias, o sistema necessitaria de algum tempo para restabelecer o crescimento microbiano. Isto causaria paralisação na produção de metano.

Por outro lado, o fato de se acreditar que o processo pode alcançar estabilidade e ter continuidade mesmo com carregamento intermitente de substrato está relacionado ao fato de que reação de fermentação (primeira etapa) na formação do metano envolve uma variação de energia livre da ordem de $-23,0 \text{ kJ mol}^{-1}$ de elétrons (Manahan, 1994, p.154).

Das considerações acima pode-se inferir que uma parte da queda nos níveis do metano (Figura 6.37, p307) deve ser atribuída à dinâmica de formação de metano nos termos da discussão acima.

Por outro lado as Figuras 6.1 e 6.2 (p241 e 242) mostram que as concentrações de CO , NO^* e NO_2^* iniciam seu rápido crescimento a partir das 5:00 h com seus picos característicos da parte da manhã (8:00–9:00 h para o CO e o NO^* , e ~10:00 h para o NO_2^*) e da noite (~20:00h, CO e o NO^* , e ~19:00 h, NO_2^*) e seus vales ~14:00–15:00 h.

O consumo de CH_4 pelo radical $\text{HO}^\bullet$ leva à formação de $\text{H}_3\text{COO}^\bullet$. Este, em atmosfera poluída, passa a consumir o $\text{NO}^\bullet$, mostrando que a queda nos níveis de CH_4 está, de algum modo relacionada ao consumo de $\text{NO}^\bullet$. Como já discutido para o ciclo do CO, após as 15:00 h os níveis de CH_4 voltam a crescer, a despeito de também haver crescimento de CO, $\text{NO}^\bullet$ e $\text{NO}_2^\bullet$. Nesse período, o $\text{NO}^\bullet$ reage também com o ozônio formado, o qual apresenta queda nesse mesmo período (Figura 6.3, p244). Como acontece no ciclo do CO, tais resultados indicam que no ciclo do metano o ozônio é produzido, em parte, à custa do consumo de CH_4 .

Assim, a queda nos níveis de CH_4 observada na Figura 6.37 (p307) deve ser atribuída ao conjunto dos efeitos causados (em grande extensão) pela dinâmica da produção anaeróbica de metano e (em muito menor extensão) pelo consumo de CH_4 pelo radical $\text{HO}^\bullet$ em atmosferas poluídas, nas quais são elevados os níveis de concentração de poluentes provenientes das emissões automotivas (CO e NO_x).

A análise por meio do modelo PARAFAC revelou, para o ciclo do metano, a existência de três perfis. Um para a fonte de emissão automotiva, outro para o ozônio, e um terceiro para o metano, este, originário de outras fontes que não as automotoras como mostra a Figura 6.46a (p317). Entretanto, surge aqui um fato novo. A despeito de ser o CH_4 também um poluente primário e seu consumo estar relacionado às emissões automotivas, portanto à poluentes primários, as variações sistemáticas semanais para o metano não são acomodadas na mesma componente que descreve as variações sistemáticas daqueles. O modelo precisa de uma terceira componente (PC3) para representar o perfil do metano como pode ser visto na Figura 6.46b (p317).

Esse resultado é de extrema importância porque, juntamente com as reações de formação de HONO e as do ciclo dos hidrocarbonetos não metano, explicam a ausência de queda, durante os finais de semana, nos níveis encontrados para o ozônio, um poluente secundário formado às custas do consumo dos primários.

A análise com o modelo de Tucker confirma esse resultado. Porém foi necessário construir um novo núcleo, **G**, levando em conta a existência de uma terceira componente também no modo B para mostrar as variações sistemáticas semanais para o CH_4 . A Figura 6.61a (p336) a mostra os perfis diários, enquanto a Figura 6.61b (p336) mostra os perfis semanais. Nesta última pode-se observar que durante os finais de semana, quando ocorre

queda nas emissões dos veículos automotores, a queda dos níveis de metano é modesta se comparada com os demais dias da semana. Se não houvesse consumo de metano para sustentar a formação de ozônio nos finais de semana, os níveis para o primeiro seriam mais elevados nesse período como pode ser visto pela inclinação da curva em seu primeiro trecho (de sexta-feira para sábado). De sábado para domingo a inclinação é menor ainda, indicando que o consumo de CH_4 foi maior do que no dia anterior, e de domingo para segunda-feira a curva muda de inclinação indicando que os níveis de O_3 alcançados no domingo o foram às custas do consumo de CH_4 . A partir desse ponto os níveis de metano continuam caindo levemente, provavelmente em função do “time lag” provocado pelo crescimento das emissões automotivas a partir de Segunda-feira. É importante notar que na Quinta-feira muda novamente a inclinação da curva para o CH_4 , quando o perfil para as emissões automotivas também muda a sua inclinação (Figura 6.61b, p336). Essas variações devem estar relacionadas com a razão de mistura próximo ao solo.

Por outro lado, a interpretação dos resultados obtidos em função das emissões, portanto, relacionado às fontes, deve levar em consideração os seguintes aspectos: Para uma dada fonte que emite durante todos os dias da semana devem ocorrer oscilações no período. Entretanto se a referida fonte em um dado dia da semana deixar de emitir, isto causaria reflexos diretos no seu perfil de emissão, com as devidas repercussões na participação da substância emitida em suas reações na atmosfera. Este é exatamente o caso das emissões automotivas, as quais exibem um perfil diário (característico do tráfego diário), e um perfil semanal (característico do tráfego semanal). Nesse segundo caso ocorre uma queda drástica nas emissões por um período mais prolongado (o fim de semana), com reflexos diretos na participação dos poluentes emitidos pela fonte em suas reações químicas na atmosfera.

Desse modo é fundamental a correta interpretação das variações que ocorrem nos perfis de emissão durante os finais de semana, ou seja, quando ocorre súbita queda nas emissões de uma das fontes (a automotiva) que participa do conjunto de emissões no local sob investigação.

7.4 – Ciclo dos Hidrocarbonetos Não Metano

Dentre os compostos orgânicos voláteis (VOCs), estavam disponíveis para análise nessa tese apenas os hidrocarbonetos. Dos poluentes orgânicos encontrados na atmosfera, os hidrocarbonetos, devido a sua grande utilização em combustíveis, constituem a classe predominante. Eles podem ser emitidos diretamente para a atmosfera, ou como subprodutos da combustão parcial de outros hidrocarbonetos (Manahan, 1994, p.357).

Considerando a importância que possuem os NMHCs na atmosfera, e os recentes resultados das pesquisas em torno do papel desempenhado por um conjunto de poluentes tanto na baixa troposfera quanto em maiores altitudes da atmosfera, uma breve menção ao tema do aquecimento global pode contribuir para enriquecer a discussão que é apresentada ao longo dessa subseção.

Parece existir uma polêmica nos meios científicos em torno da melhor estratégia para a redução de gases de efeito estufa. Alguns defendem uma nova abordagem que leve em conta uma estimativa de cálculo para a força radiativa (Hansen, 2000; Hansen et alli, 2000). Nessa abordagem, propõe-se que os esforços sejam concentrados na redução das emissões de outros gases de efeito estufa que não o CO₂ (non-CO₂ greenhouse). O principal alvo dessa corrente científica é o metano, considerado o segundo mais importante gás de efeito estufa.

A força radiativa é conhecida pelos colaboradores do Intergovernmental Panel on Climate Change (IPCC) como Potencial de Aquecimento Global (Global Warming Potential – GWP). O conceito de GWP foi desenvolvido para comparar a habilidade que cada gás de efeito estufa possui para reter calor na atmosfera em relação a um outro gás considerado como referência, “status” de que goza hoje o CO₂. De acordo com o conceito, o GWP de um gás de efeito estufa é a razão de aquecimento global (direto e indireto) entre uma unidade de massa do gás de efeito estufa e uma unidade de massa de dióxido de carbono, em um período de tempo (EPA, 2003).

Uma das correntes contrárias àquele movimento citado acima (liderado por Hansen do NASA Goddard Institute for Spaces Studies), tem a sua expressão nos vários documentos elaborados pelo Painel Intergovernmental Sobre Mudança Climática (IPCC), nos quais a redução direta de produção de CO₂ deve ser a meta à perseguir.

Uma outra face dessa discussão está na preocupação em se relacionar as emissões naturais com as provocadas pela ação do homem. Nesse sentido, pesquisa recente sobre interação planta-CO₂ submeteu três plantações de uma cultura a diferentes atmosferas de dióxido de carbono (Rosentiel *et alli*, 2003). Seus resultados mostraram uma redução de até 41% na produção de isopreno, um hidrocarboneto não metano que é emitido em grandes quantidades principalmente pelas árvores (dentre a vegetação) e que é responsável pela produção de grandes quantidades de ozônio (Chameides *et alli*, 1988, Harley *et alli*, 1999). Calcula-se que os níveis de emissão desse NMHC sejam maiores do que o dobro de todos os outros NMHCs combinados, e com capacidade para aumentar as concentrações de ozônio na superfície, em 50-60% sobre o solo e em cerca de 40% sobre a camada limite marinha (Poisson *et alli*, 2000). Estima-se ainda que as emissões biogênicas (isopreno predominantemente) possuam um papel fundamental na química troposférica do metano, aumentando seu tempo de vida atmosférico em 14% (Poisson *et alli*, 2000).

Em relação ao metano, desde o início das medições com instrumentação moderna, a taxa de crescimento de sua concentração troposférica passou a diminuir nos anos 1990, tornando-se negativa em 2000 (Simpson *et alli*, 2002). Esses resultados, sugerem que qualquer que seja a causa desse declínio (que pode incluir outros processos além daqueles causados por reduções nas emissões de isopreno pelas plantas, por indução de CO₂) é tão poderoso que poderá levar a um decréscimo anual na concentração atmosférica de metano em futuro não muito distante (Sherwood e Craig, 2003).

Entretanto, em relação ao ozônio, a queda nas emissões de isopreno dos vegetais, induzida por CO₂, não são tão evidentes porque as concentrações de ozônio troposférico continuam a crescer em consequência da expansão das atividades antrópicas (Sherwood e Craig, 2003).

Toda a polêmica em torno dos problemas levantados acima (gases de efeito estufa – diretos e indiretos e das fontes de emissão – naturais e antrópicas) está fundamentada, em grande parte, na realidade dos países mais industrializados. Assim, por sua própria origem, a metodologia do IPCC adotada pela Convenção tem como referência pesquisas realizadas e metodologias elaboradas por especialistas de países desenvolvidos, nos quais as emissões são, em sua maioria, provenientes da queima de combustíveis fósseis (CETESB, 2002). Independentemente do IPCC, grande parte da construção do conhecimento em torno desses

problemas toma por base os trabalhos desenvolvidos por agências e institutos de pesquisas que usam como base de dados aqueles levantados invariavelmente acima da linha do equador. Desse modo, todo o esforço realizado pelas nações ditas em desenvolvimento passa necessariamente por duas vias complementares: adaptação daquele, e construção do conhecimento fundamentado em sua própria realidade.

Considerando a realidade brasileira e, especialmente a da localidade P. D. Pedro II, verificou-se que, diferentemente do que ocorre em vários países, a fonte de emissão de poluentes primários em centros urbanos como a cidade de São Paulo (onde coexistem habitações, comércio e indústrias), é predominantemente automotiva ($PC_1 = 83,63\%$ para o CO; $PC_1 = 74,27\%$ para o NO; $PC_1 = 91,79\%$ para o NO₂; $PC_1 = 75,78\%$ para o SO₂; $PC_1 = 94,08\%$ para o CH₄; $PC_1 = 76,78\%$ para os NMHC).

Para efeito de comparações, em cidades como Barcelona (Espanha) 55,9% das emissões são de fonte automotiva contra 31,9% de indústrias (Montserrat e Baldasano, 1996); em Atenas (Grécia), PC_1 para o CO, NO, NO₂ e SO₂ foram encontrados, respectivamente, 32,01%, 20,90%, 18,28% e 1,32% (Statheropoulos *et alli*, 1998); em Martorell (Espanha), $PC_1 = 35,6\%$ para emissões do tráfego de veículos (+ evaporação em postos de combustíveis) e $PC_1 = 35,1\%$ para indústrias (Baldasano *et alli*, 1998).

Com relação aos resultados encontrados nessa tese para os hidrocarbonetos não metano, a discussão dos resultados leva em conta a polêmica existente sobre essas emissões no contexto da localidade Pedro II.

As emissões automotivas dessa classe de hidrocarbonetos estão distribuídas em cerca de um terço para os alcanos, um terço para os alcenos mais reativos, e um terço para os aromáticos (Manahan, 1994, p.357).

Essas emissões foram analisadas por meio de PCA. Os resultados mostraram que a principal fonte de emissão desses compostos é automotiva. Para a Estação de Medição Parque D. Pedro II (1999), a Figura 6.39 (p309) mostra as variações sistemáticas diárias para os NMHC. Como se vê, o perfil encontrado segue o mesmo padrão dos outros compostos oriundos da fonte automotora.

A análise PARAFAC para os NMHC individualmente, na mesma localidade, mostrou, como PCA, para o perfil semanal, um rápido crescimento a partir das 5:00 h até alcançar o primeiro pico às 9:00 h. Em seguida, queda brusca com um mínimo às 14:00-

15:00 h, nova elevação rápida atingindo outro pico aproximadamente às 19:00 h, para em seguida cair continuamente até próximo das 4:00 h, como pode ser visto na Figura 6.40a (p309). Para as variações sistemáticas semanais (Figura 6.40b, p309) nota-se a queda típica das emissões automotivas nos finais de semana.

Considerando-se o ciclo dos hidrocarbonetos não metano (NMHC, NO, NO₂ e O₃), quando o mesmo é analisado por meio do modelo PARAFAC, surge um novo aspecto em relação ao número de fatores que o modelo utiliza para descrever as variações sistemáticas diárias e semanais e a variação anual.

Segundo as informações obtidas preliminarmente por meio de PCA e PARAFAC para o poluente isolado das demais variáveis, esperava-se que ao analisar os NMHC no seu ciclo, isto é, sob a influência do NO^{*}, do NO₂^{*} e do O₃, as variações sistemáticas diárias e semanais para os hidrocarbonetos ficassem acomodadas na mesma componente que continha os óxidos de nitrogênio, ou seja, na componente que captura as variações para os poluentes primários. A outra componente conteria as variações para o ozônio. Partindo-se dessa informação, na análise do ciclo com o PARAFAC foram utilizadas duas componentes. A Figura 6.47.1a (p319) mostra que PC₁ contém o perfil diário característico dos poluentes primários provenientes de fontes automotivas, enquanto PC₂ representa o perfil diário para o O₃. Para o perfil semanal (Figura 6.47.1b, p319), também estava confirmada a queda dos poluentes primários nos finais de semana, período no qual o poluente secundário atinge valores mais elevados. Para a variação anual (Figura 6.47.1c, p319) também se confirmava a maior acumulação para os poluentes primários no período de inverno.

O teste para verificação do número correto de componentes, entretanto, indicou que o modelo PARAFAC não só não degenerava como não perdia significado físico no modo B (perfil semanal), quando solicitado a processar os dados com três componentes. Assim, o modelo também consegue modelar as variáveis com três componentes, como mostra a Figura 6.47.2 (p320). Nessa figura, em a, PC₁ representa o perfil diário de emissão da fonte automotiva, PC₂, o perfil diário para o ozônio, e PC₃, o perfil de emissão diário também para a fonte automotiva. Então, esse perfil deve representar as emissões de um outro tipo de veículo, porém da mesma fonte, isto é, a automotiva. Isto, porque o padrão captado pelo modelo assemelha-se bastante ao das emissões automotivas, com os seus dois picos

característicos coincidentes com os horários do “hush”. De fato, análise multivariada aplicada a hidrocarbonetos aromáticos policíclicos permitiram distinguir entre emissões de veículos à gasolina e veículos movidos à óleo diesel (Alsberg, 1989).

Quando se observam as variações sistemáticas semanais, o resultado é, mais uma vez, surpreendente. Se o poluente é emitido pela mesma fonte (a automotiva), então o seu perfil semanal deve ser o mesmo, com a queda característica nos finais de semana. O que se vê, no entanto, são dois perfis (PC_1 e PC_3 na Figura 6.47.2b, p320) com o mesmo padrão, isto é, com um alto grau de correlação entre eles. Se não ocorreu degenerescência no modelo nem perda de significado físico em qualquer dos modos, isto é uma forte indicação da existência de um tipo específico de emissões que, mesmo pertencendo a mesma fonte, apresenta alguma característica que a distingue das demais emissões. A hipótese mais provável é a de que sejam emissões de fonte automotiva, mas de um tipo diferente de veículo automotor. Uma outra hipótese, não menos provável que a primeira, é a de que essas emissões estejam associadas com a reatividade dos NMHC. Esse comportamento também foi, com pequenas diferenças (no plano algébrico), encontrado para o SO_2 .

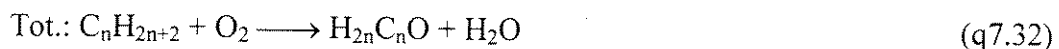
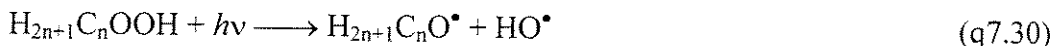
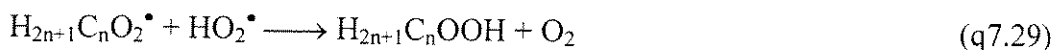
Sob o modelo de Tucker, o ciclo dos NMHC ofereceu os resultados concordantes com as preliminares do PARAFAC. A Figura 6.62a (p338) mostra, em PC_1 , o perfil de emissão da fonte, e PC_2 e PC_3 são as componentes relacionadas com a acumulação. Para as variações sistemáticas semanais (Figura 6.62b e d, p338), PC_1 e PC_2 representam os perfis para os poluentes primários, e PC_3 para o poluente secundário, O_3 . Deve-se chamar a atenção para o fato da existência de um valor mais elevado para PC_2 na 4ª feira. Este também foi um padrão encontrado para o SO_2 (porém com um máximo na 3ª feira).

A interpretação dos resultados fornecidos pela quimiometria deve estar associado, de algum modo, com as reações químicas da atmosfera. Diferentemente dos ciclos para o CO e o CH_4 , nos quais as reações possuem como ponto de partida apenas uma substância (CO ou CH_4 , dependendo do ciclo), no caso do ciclo dos NMHC, várias substâncias diferentes podem funcionar como precursores dos produtos das reações no e fora do ciclo. Assim, seguir-se-á a classificação mencionada no início dessa seção. Os poluentes de partida são divididos em hidrocarbonetos com ligações simples (alifáticos: parafinas), com ligações duplas (alifáticos: olefinas), e compostos aromáticos.

De estudos realizados em câmaras de “smog”, encontram-se, respectivamente, para alcanos e alcenos, frações molares iguais a 0,556 mol e 0,444 mol. Os alcanos mais presentes possuem no máximo oito átomos de carbono por molécula. Alcanos com seis ou menos átomos de carbono na molécula possuem pressão de vapor elevada tal que se encontram na fase gasosa. Aqueles possuindo entre sete e vinte átomos de carbono podem ser encontrados na fase gasosa ou em partículas dependendo das condições (Manahan, 1994, p.357). Para efeito das reações abaixo, assumir-se-á que compostos com até oito átomos de carbono encontram-se na fase gasosa, especialmente durante o dia.

Como nos demais casos, o estudo deve ser conduzido também sob condições de baixa e de alta concentração de óxidos de nitrogênio. Assim, para os alcanos têm-se a seguinte seqüência de reações:

Alcanos: Baixa conc. NO_x



Para o conjunto de reações q7.27 a q7.32 considera-se que o alcano genérico, representado por C_nH_{2n+2} , encontra-se na fase gasosa.

Das reações acima pode-se observar que a oxidação de alcanos sob baixas concentrações de NO_x leva a formação do aldeído representado por $H_{2n}C_nO$. Este pode sofrer decomposição fotoquímica segundo as etapas mostradas na Figura 7.12.

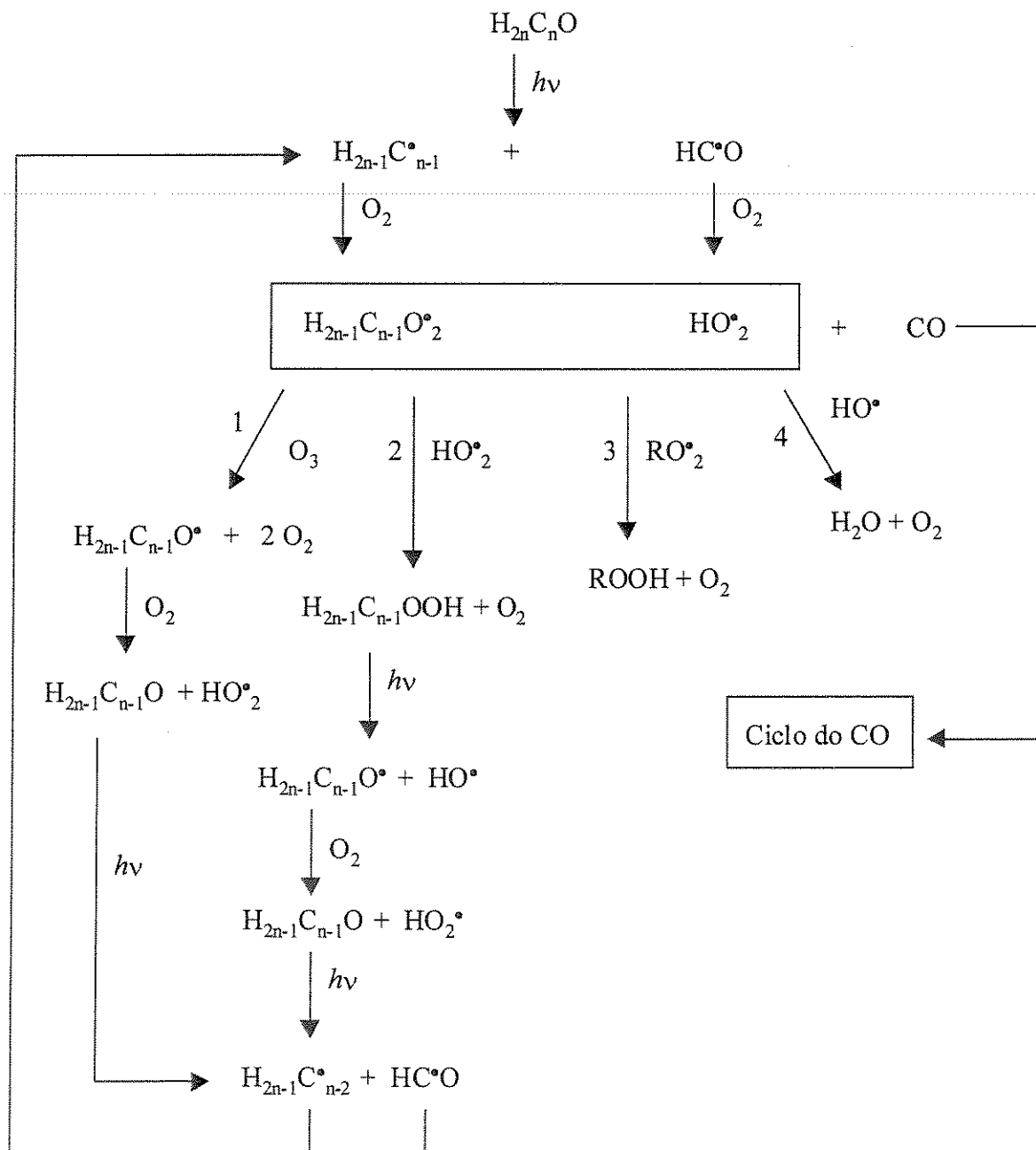
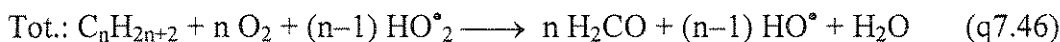
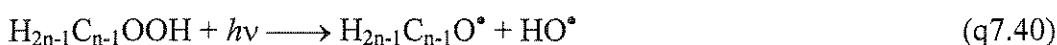
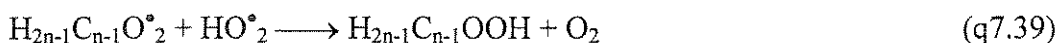
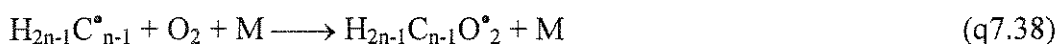
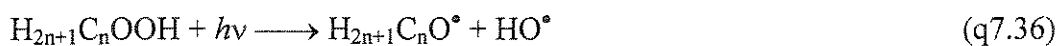
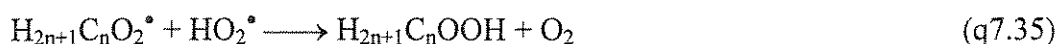
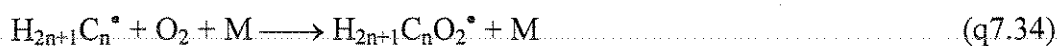


Fig.7.12: Mecanismo para a decomposição de aldeídos na atmosfera (baixa conc. NO_x).

Para o esquema da decomposição de aldeídos mostrado na Figura 7.12, se o grupamento CH₃ for representado por R, o mecanismo envolve a diminuição do número de átomos de carbono na cadeia (CH₃) via oxidação e decomposição fotoquímica em série, resultando na eliminação do radical formil, HC°O. Assim, denominou-se a via em vermelho na Figura 7.12 de *ciclo de eliminação formil*.

Pode-se construir um mecanismo alternativo para as reações de alcanos sob baixas concentrações de NO_x (ver explanação em alcenos sob alta conc. NO_x, na p407).

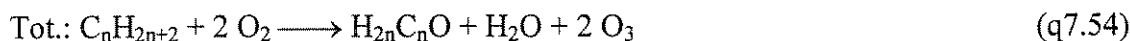
Alcanos (rota alternativa): Baixa conc. NO_x



A diferença entre os dois casos está na equação q7.31 do primeiro mecanismo e nas equações q7.37 e q7.41 do mecanismo alternativo. Nesse último caso obtém-se ao final, tantas moléculas de formaldeído e, consequentemente de CO, quantos forem os átomos de

carbono do alcano inicial. No primeiro caso, obtém-se o mesmo número de moléculas de aldeídos (e de CO), porém, em cada etapa ocorre a eliminação do radical formil, como mostrado na via em vermelho (*ciclo de eliminação formil*) da Figura 7.12. Em qualquer dos mecanismos haverá, via CO, a destruição do mesmo número n de moléculas de O₃.

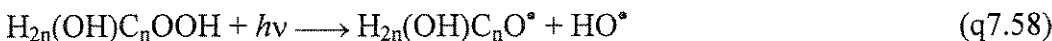
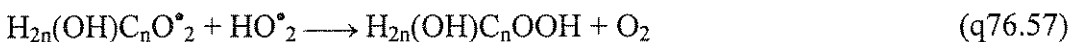
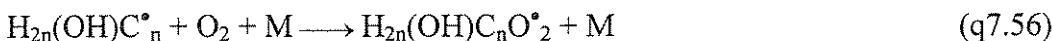
Alcanos: Alta conc. NO_x



O aldeído formado na equação q7.54 é decomposto fotoquimicamente e segue as etapas do ciclo de eliminação formil da Figura 7.12, com a conseqüente formação de CO. A diferença fundamental é que nesse caso ocorre a formação de duas moléculas de O₃ no primeiro bloco de reações (q7.47 a q7.53). As subseqüentes reações no ciclo de eliminação formil geram tantos CO (gerador de O₃, nesse caso, pois sob altas concentrações de O₃), e O₃ quantos forem os radicais formil eliminados, isto é, tantos quantos forem os átomos de carbono no alcano inicial (excluídos os dois citados inicialmente – até a reação q7.53).

Para os alcenos o resultado não é muito diferente

Alcenos: Baixa conc. NO_x



No bloco de reações q7.55 a q7.61 observa-se a formação de dois aldeídos. Se o alceno possui n átomos de carbono na cadeia, um dos compostos carbonílicos é o formaldeído, o outro é o aldeído com $(n - 1)$ átomos de carbono, de fórmula $H_{2n-1}C_{n-1}O$. A Figura 7.13 mostra as reações subsequentes para esse caso.

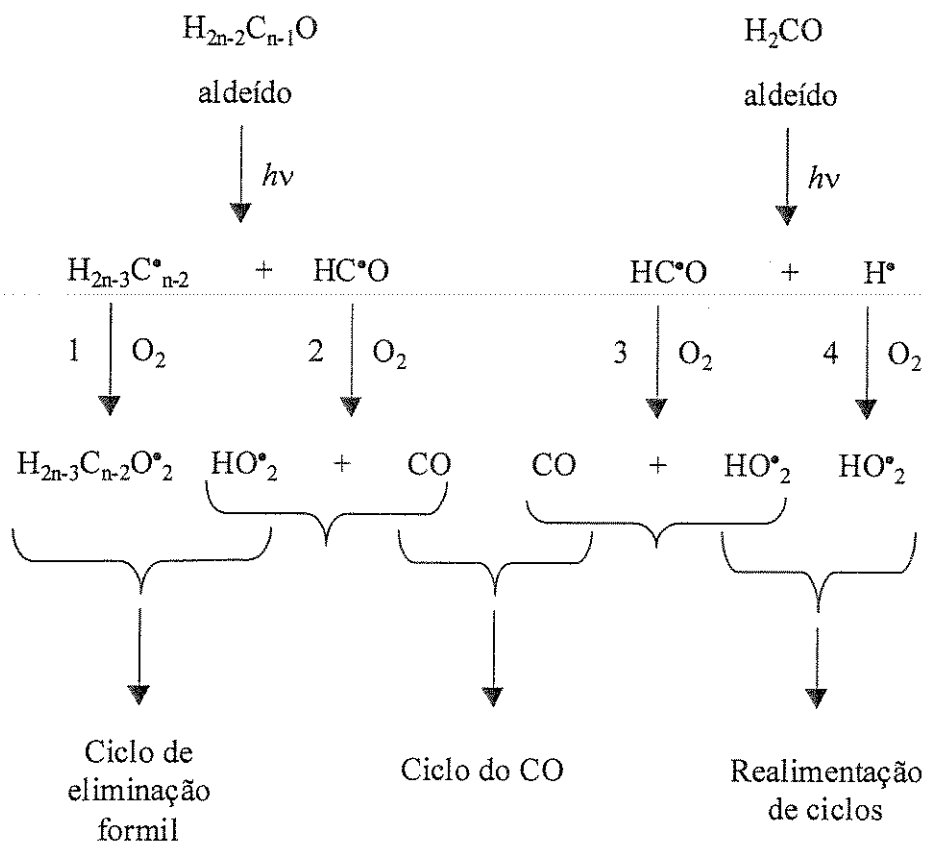


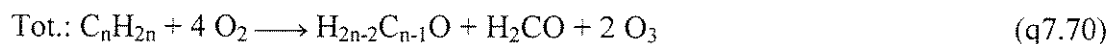
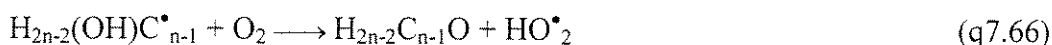
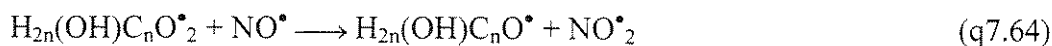
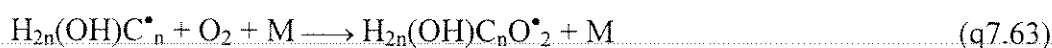
Fig.7.13: Mecanismo para decomposição de aldeídos na atmosfera (baixo NO_x).

A Figura 7.13 mostra as principais vias para a decomposição de aldeídos produzidos na atmosfera sob condições de baixas concentrações de NO_x . A rota número um leva à diminuição do tamanho da cadeia de carbonos do aldeído por meio da sua reação com o O_2 e posterior reação com o radical $HO_2^{\bullet}$ fornecido pela reação do radical formil da rota número dois. Este, por meio da rota número três também fornece outro radical $HO_2^{\bullet}$. A rota número quatro fornece mais um radical $HO_2^{\bullet}$. Os radicais produzidos pelas rotas três e quatro podem seguir realimentando ciclos ou reagirem mutuamente formando O_2 e H_2O_2 . Essa última espécie pode sofrer decomposição fotoquímica produzindo dois radicais $HO^{\bullet}$.

Finalmente, as rotas dois e três produzem duas moléculas de CO, as quais entram no ciclo do monóxido de carbono causando sob as condições dadas (baixas concentrações de NO_x) a destruição de O₃. Nesse caso, como o bloco inicial de reações (q7.55 a q7.61) produz dois aldeídos, sendo um deles o formaldeído, haverá destruição de n+1 moléculas de O₃.

Para os hidrocarbonetos com ligações duplas, sob altas concentrações de óxidos de nitrogênio, adicionalmente aos dois aldeídos que se formam no primeiro bloco de reações, são produzidas duas moléculas de ozônio, como será visto em seguida.

Alcenos: Alta conc. NO_x



Nesse ponto deve-se chamar a atenção para as reações q7.31, q7.37, q7.41, q7.59 e q7.65. No mecanismo para alcanos sob baixa concentração de NO_x, a reação sugerida na equação q7.31 difere daquela utilizada nas equações q7.37 e q7.41 (mecanismo alternativo). Para as reações q7.59 e q7.65 sugeridas para o mecanismo no caso dos alcenos, respectivamente, sob baixas e altas concentrações de NO_x, foi utilizado o mesmo caminho das equações q7.37 e q7.41, diferentemente daquele seguido pela reação q7.31.

No caso das equações q7.59 e q7.65, Baird (1995, p.127) sugere que ao invés de ocorrer perda de H via abstração por O₂, os radicais representados pelas equações q7.59 e q7.65, sofrem decomposição espontânea por clivagem da ligação carbono-carbono (C-C), dando origem a uma molécula não radical contendo uma ligação dupla C=O, e um outro radical RHC[•]OH. A justificativa para tal percurso alternativo seria porque a reação não requereria entrada de energia, isto é, ΔH é próximo de zero, porque nesse caso a formação de uma ligação C=O a partir de uma C-O compensaria, do ponto de vista energético, a

perda da ligação C–C. Entretanto, para as equações q7.31, q7.37 e q7.41, não existe nenhum grupamento OH “enfraquecendo” a ligação C–C do grupo carbonílico, o que causa uma queda na energia de ativação tornando ΔH muito pequeno. Assim, para as três reações acima, pode-se alcançar o “status” de dupla ligação compensadora do ponto de vista energético, sem ocorrer qualquer perda de energia com o rompimento de uma ligação C–C, e ainda ocorrer a formação de um outro radical (HO^*_2). Este foi o argumento que deu preferência a reação q7.31 em relação ao mecanismo alternativo que passa pelas reações q7.37 e q7.41 para o caso dos alcanos.

Retornando aos alcenos sob altas concentrações de NO_x , as reações subsequentes
..... são mostradas na Figura 7.14 a seguir.

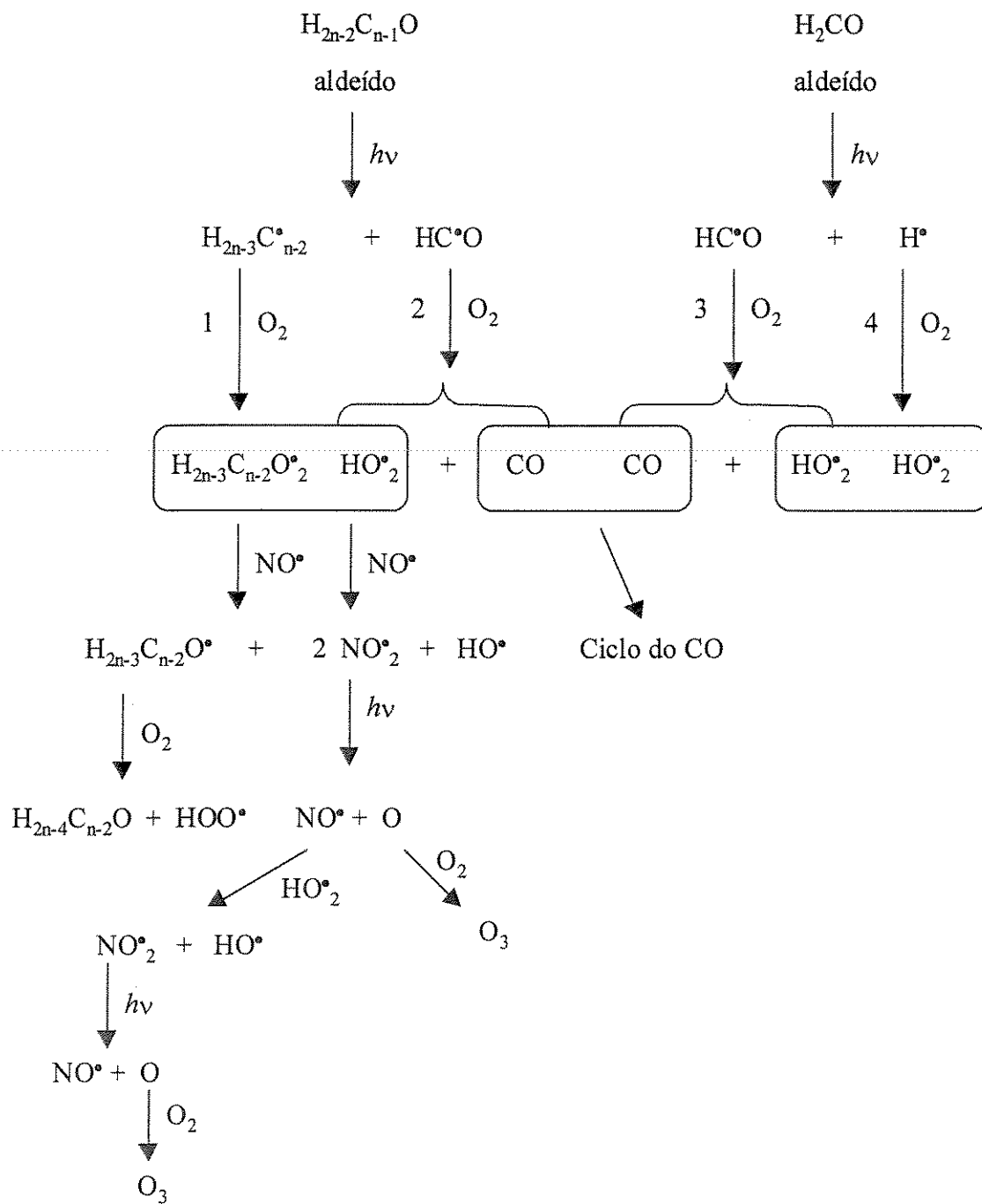
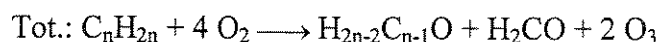


Fig.7.14: Esquema para o mecanismo de reação para os aldeídos no ciclo dos alkenos sob altas conc. NO_x .

O papel desempenhado pelos hidrocarbonetos não metano com ligações duplas (e com até oito átomos de carbono na molécula) é tão importante que vale a pena discutir um pouco mais sobre as reações do ciclo sob condições de altas concentrações de óxidos de nitrogênio.

Tomando-se como ponto de partida a reação global do primeiro bloco para esse caso (equação q7.70) reproduzida abaixo, observa-se que a partir do alceno inicial C_nH_{2n} obtém-se um aldeído expresso como $H_{2n-2}C_{n-1}O$ e um formaldeído.

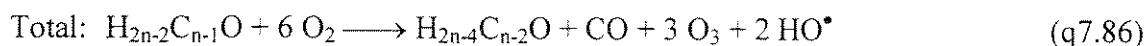
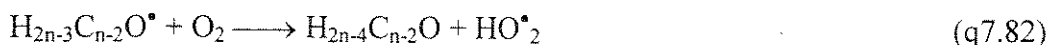
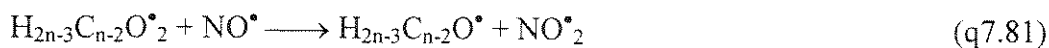
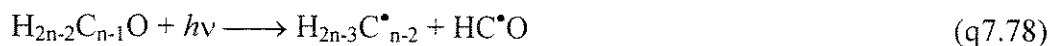


Considerando-se os desdobramentos da equação global em termos da decomposição fotoquímica sofrida pelos aldeídos pode-se calcular a relação entre as quantidades de compostos envolvidos no ciclo sob as condições dadas. Tomando-se as reações para o formaldeído tem-se:



Para o bloco de reações acima, se n (número de átomos de carbono no alceno inicial) for igual a dois, o hidrocarboneto de partida é o etileno. Este, leva à formação de duas moléculas de formaldeído. Cada uma destas, produz um CO, dois radicais $HO^{\bullet}$ e duas moléculas de ozônio. Assim, cada molécula do etileno inicial produzirá o dobro das quantidades expressas na equação q7.77, isto é, 2 CO, 4 $HO^{\bullet}$ e 4 O_3 .

Estendendo-se o cálculo para os casos em que o hidrocarboneto possui até oito átomos de carbono, obtém-se como resultado do primeiro bloco de reações, a equação q7.70. Assim, tomando-se as reações para o aldeído de fórmula $H_{2n-2}C_{n-1}O$ tem-se:



Os resultados obtidos segundo os mecanismos acima, permitem calcular as quantidades de monóxido de carbono, de radicais hidroxila e de ozônio que se formam na atmosfera poluída com NMHC's da classe dos alcenos. Assim, para **n** igual a **dois** são produzidos **1n** CO, **2n** HO[•] e **2n** O₃. Por outro lado, quando **n** é maior do que dois, são gerados **3n** CO, **2n** HO[•] e **3n** O₃.

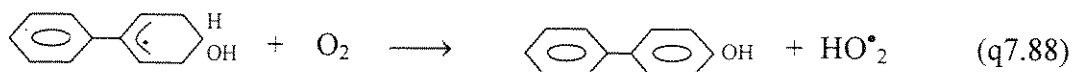
Para os compostos aromáticos, é de particular interesse nas emissões automotivas o composto bifenil emitido pelos veículos movidos à diesel (Manahan, 1994, p.359-361).

Entretanto, Baldasano et alli (1998) identificaram altos níveis de emissão de alquilbenzenos (especialmente tolueno e m+p-xileno) e hexano nos veículos movidos à gasolina. Seus resultados mostraram também que os veículos diesel são responsáveis pela emissão de alcanos superiores (especialmente decano, dodecano e tetradecano).

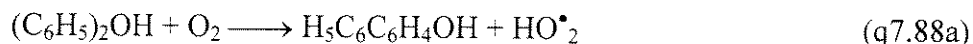
Por outro lado, resultados de Alsberg (1989) em PAHs, identificaram emissões de aromáticos simples (benzeno, tolueno e xilenos), benzo(ghi)perileno e coroneno, associadas aos veículos à gasolina. Os compostos fenantreno, fluoranteno, pireno e dibenzotiofeno, por seu lado, estavam associados aos veículos diesel.

Assim, as reações para os compostos aromáticos sob baixas e altas concentrações desses compostos serão representadas pelo composto bifenil. Da mesma maneira que a maioria dos hidrocarbonetos atmosféricos, a reação mais provável do benzeno e seus derivados é com o radical hidroxila. A adição de HO[•] ao anel benzênico conduz à formação de radicais instáveis, os quais em reação com O₂ formam espécies estáveis e HO₂[•].

Aromáticos: Baixa conc. NO_x

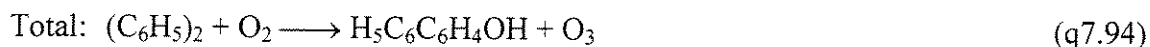
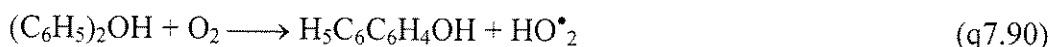


As equações acima podem ser escritas de forma simplificada como:



As equações q7.87 e q7.88 (ou q7.87a e q7.88a) mostram que sob baixas concentrações de NO_x o composto bifenil formará o composto estável bifenol e radical HO₂[•].

Aromáticos: Alta conc. NO_x



Sob concentrações elevadas de óxidos de nitrogênio, ocorre a formação do composto bifenol e de ozônio. Assim, para cada molécula de bifenil emitida para a atmosfera nas condições dadas, ocorre a produção de uma molécula de O₃. O raciocínio pode ser estendido para os outros aromáticos emitidos tal que, de posse de suas quantidades relativas, pode-se estimar a quantidade de O₃ produzido pela emissão de aromáticos.

As discussões até aqui sobre os resultados encontrados nessa tese levam em conta que os poluentes existentes na atmosfera da RMSP (nas localidades investigadas) são oriundos de emissões predominantemente automotivas. Isto está expresso em todas as análises, bastando inspecionar a variância da primeira componente principal para cada poluente individualmente nas análises PCA.

Em seguida, o modelo PARAFAC evidenciou a existência de um perfil para o ozônio que contrastaria, em princípio, com o fato de que este, sendo um poluente secundário cujos níveis de concentração dependem das emissões de NO_x , deveria ter tais níveis compatíveis com uma concomitante queda nas emissões de NO_x . Mas isto não foi observado nos resultados encontrados. Ou a queda era insignificante quando comparada à queda dos NO_x , ou ocorria ligeiro aumento nos níveis de concentração de O_3 nos finais de semana.

Sob o modelo de Tucker foi possível analisar os ciclos e comparar os efeitos provocados nos resultados pela inclusão de cada poluente em cada ciclo. Em todos os casos foi confirmada a aparente discrepância apontada pelo modelo PARAFAC. Entretanto, ao se juntar os ciclos (Figura 6.74, p352), submetendo-se os poluentes analisados à influência uns dos outros, foi possível extrair informações fundamentais para explicar tal comportamento para o O_3 na principal localidade investigada nessa tese. Adicionalmente, foi possível avaliar as participações do CO , CH_4 nos níveis de O_3 encontrados, em função das emissões de NO_x .

Assim, sob o título “Conexão dos Ciclos” segue uma discussão que tem como objetivo mostrar dependência dos poluentes com a reatividade que cada um possui em relação aos radicais presentes na atmosfera e em relação à decomposições fotoquímicas (quando for o caso). E, mais adiante, sob o título, “Discussão Final”, a apresentação de evidências conclusivas para a explicação dos resultados encontradas nas análises.

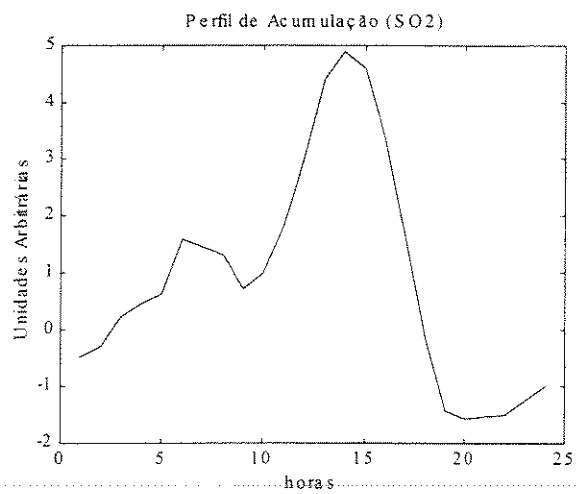
7.5 – Conexão dos ciclos

Os resultados obtidos por meio das análises com o modelo de Tucker para o estudo das influências dos poluentes em seus respectivos ciclos permite que sejam encontrados os perfis de acumulação de cada poluente.

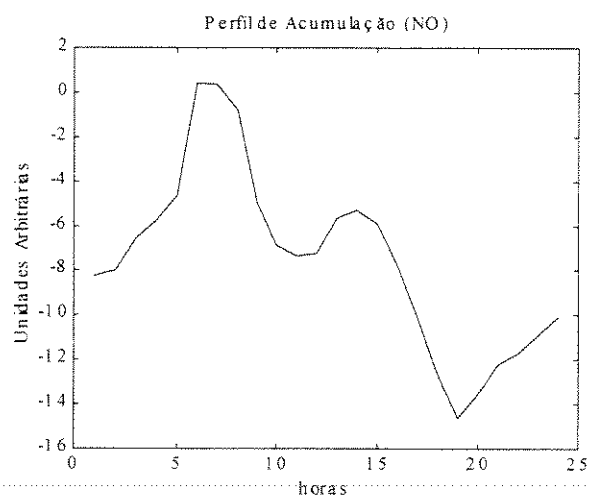
De acordo com a breve discussão na Seção 6.3, e com a discussão mais aprofundada na Seções 7.2 desse capítulo (ciclo do CO), especial atenção deve ser dada ao núcleo do modelo para que se obtenham resultados consistentes com os pressupostos utilizados na construção da metodologia discutida no Capítulo III. Decorre dessa observação que o núcleo usado para a análise dos poluentes no ciclo básico do monóxido de carbono difere, por exemplo, daquele usado para o ciclo básico do metano. Nesse último caso, como foi visto, análises preliminares com PCA e PARAFAC mostraram que a principal fonte de CH₄ não é automotiva, exigindo, na análise com o Tucker, a construção de um núcleo capaz de levar em conta esse comportamento diferente para a principal fonte de metano.

Para o ciclo básico do monóxido de carbono fez-se uma discussão detalhada sobre os resultados encontrados para os perfis de acumulação dos poluentes no ciclo. Ou seja, fez-se uma interpretação daqueles resultados à luz das reações químicas que exercem influência no comportamento das variáveis do ciclo. Tal discussão mostrou o papel fundamental do óxido nítrico nas reações da química da atmosfera. Ademais, tais resultados obtidos por meio dos modelos utilizados mostraram-se consistentes com as reações químicas ocorrendo na baixa troposfera.

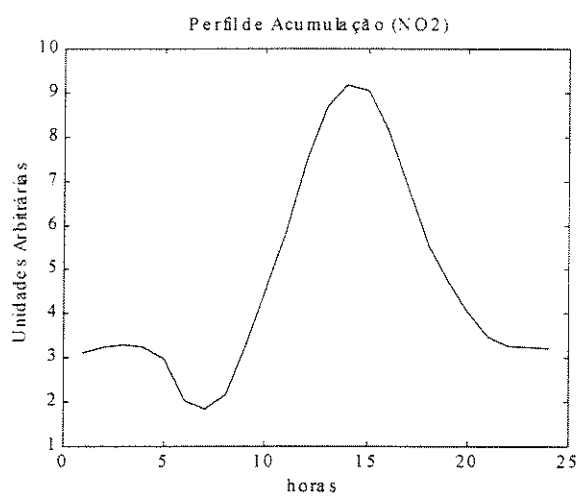
Os respostas encontradas para o ciclo básico do CO devem ser válidas para os outros ciclos e para o caso limite da junção dos ciclos. Assim, partindo dos resultados encontrados para o ciclo básico do monóxido de carbono, e daqueles para os demais ciclos e suas inter relações expressos nas Figuras 6.62 (ciclo básico dos NMHC, p338), 6.65 (ciclo básico do CO com NMHC incluído, p343), 6.66 (ciclo básico do CH₄ com NMHC incluído, p344), 6.67 (ciclo básico do CO com SO₂ incluído, p345), 6.69 (ciclo básico do NMHC com SO₂ incluído, p347), 6.71 (ciclos básicos do CO e NMHC com SO₂ incluído, p349), foram encontrados os perfis de acumulação para os demais poluentes na atmosfera da RMSP, como mostra a Figura 7.15 adiante.



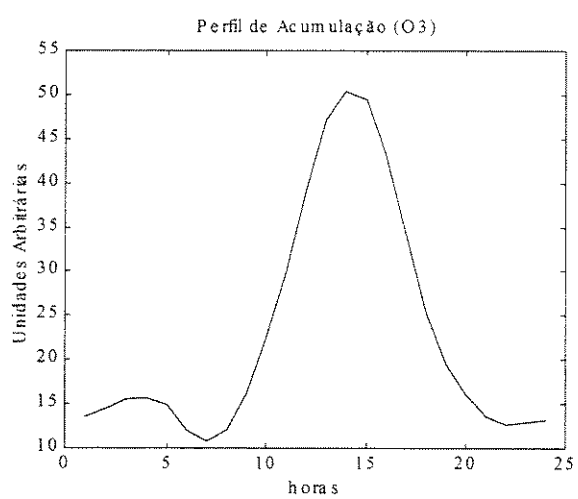
(a)



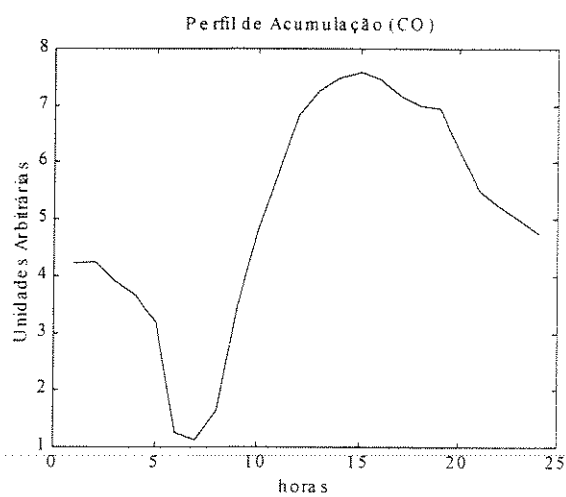
(b)



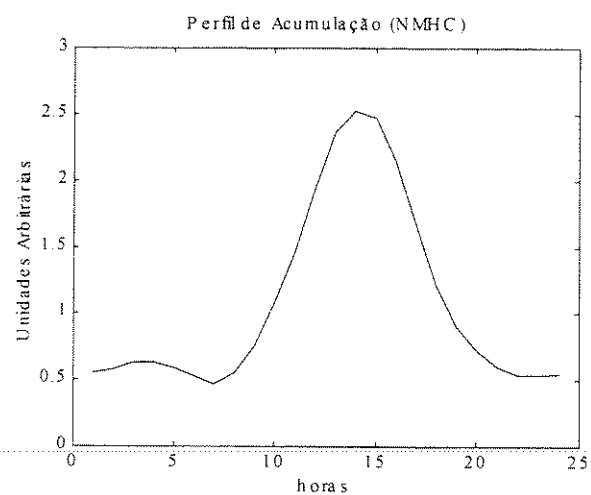
(c)



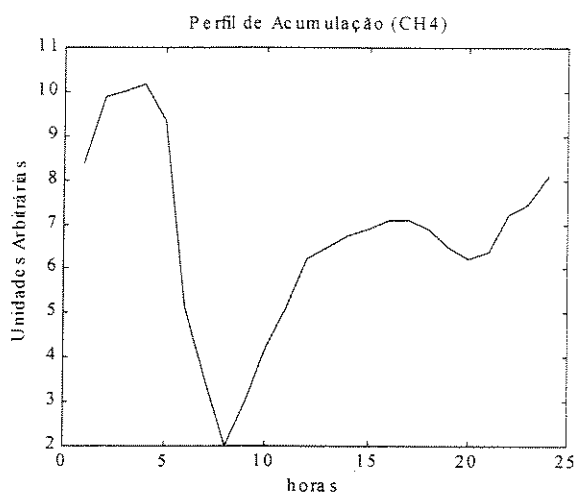
(d)



(e)



(f)



(g)

Fig.7.15: Perfis de acumulação para os poluentes nos ciclos básicos do monóxido de carbono, metano, e dos hidrocarbonetos não metano.

A Figura 7.15 mostra os perfis de acumulação para os poluentes estudados nos diferentes ciclos básicos. Uma comparação entre os gráficos das Figuras 7.10 (p376) e 7.15 (p.415-416), mostra a consistência entre os perfis para o CO, NO, NO₂ e O₃ da Figura 7.10 com aqueles para esses respectivos poluentes quando do estudo da interação entre os ciclos.

Como já discutido para o ciclo do monóxido de carbono, parece ser o dióxido de nitrogênio que controla todo o processo na atmosfera. Quando se estuda o ciclo do metano e o dos hidrocarbonetos não metano, o perfil de acumulação do NO[•] apresenta as mesmas características daquelas apresentadas no ciclo do CO. Do mesmo modo, os perfis para os outros dois poluentes básicos em todos os ciclos (NO₂[•] e O₃), e o perfil para o CO, para os referidos ciclos, mantém as mesmas características daquelas encontradas no ciclo do CO. Assim, toda a discussão realizada sobre o ciclo básico do monóxido de carbono pode ser extrapolada para o caso da interação entre os ciclos.

As Figuras 7.10a e 7.15e mostram os perfis de acumulação de CO para os casos, respectivamente, do ciclo do CO, e dos três ciclos. O mesmo pode ser visto para o NO, comparando-se as Figuras 7.10b (p376) com a Figura 7.15b (p415). Para o NO₂ a comparação deve ser feita entre as Figuras 7.10c (p376) e 7.15c (p415), e para o O₃, deve-se comparar as Figuras 7.10d (p376) com a Figura 7.15d (p415).

Entretanto, há que se emprestar interpretação do ponto de vista das reações químicas, para os perfis encontrados por meio dos modelos quimiométricos, para as outras três variáveis, SO₂, CH₄ e NMHC.

O metano, e o monóxido de carbono, possuem características que, do ponto de vista da acumulação, tornam seus perfis semelhantes. Os dois poluentes, embora sejam emitidos por fontes diferentes, acabam acumulando em escala global devido à sua menor reatividade com o radical HO[•]. Assim, quando são analisadas as diversas etapas das reações nos respectivos ciclos para esses dois poluentes pode-se assumir que a primeira etapa, em cada caso, é a mais lenta sendo, portanto, a que determina a taxa de consumo de CO e de CH₄ na atmosfera. Então o perfil de acumulação para o CH₄ não deve diferir do perfil de acumulação do CO. De fato, os dois perfis possuem a mesma tendência de acumulação. Isto pode ser verificado comparando-se as Figuras 7.10a (p376) com a Figura 7.15e (p416).

A interpretação dos perfis do SO₂ e dos NMHC deve ser considerada sob dois aspectos. O primeiro leva em conta que o SO₂ deve acumular em escala regional, enquanto os NMHC devem fazê-lo em escala local. O segundo aspecto remete para as propriedades e para as reações químicas desses dois poluentes. Tomando-se os conjuntos de reações q7.47 a q7.94 (p405 a 412) para os NMHC (alcanos, alcenos e aromáticos) e as reações q7.89 a q7.94 (p412) para o SO₂, pode-se assumir que, em se tratando de emissões principalmente automotivas, as reações na primeira etapa em cada um desses casos não se constitui da etapa mais lenta, tal que a acumulação deles não deve ser função desse primeiro passo na cadeia de reações. Todos esses compostos são muito mais reativos do que o CO e o CH₄. Isto torna diferente os seus processos de acumulação.

Observando-se as diversas etapas em cada caso verifica-se que todos os passos são marcados por reações entre radicais livres ou entre um radical e um composto muito reativo. Tomando-se, por exemplo, as reações para o ciclo dos alcenos (reações q7.62 a q7.70, p407), pode-se assumir que, sob atmosfera poluída, os passos até a decomposição fotoquímica do NO[•]₂ são muito rápidos e, tal que a fotólise do NO[•]₂ pode ser considerada como a etapa que determina a taxa de consumo dos NMHC.

A estimativa para a aproximação acima foi feita com base no cálculo do tempo de vida, τ , da espécie química considerada em relação a sua reação com HO[•], o qual é dado pela expressão (Birks, 1998, p.240):

$$\tau = 1 / k [\text{HO}^{\bullet}]_{\text{média}} \quad (iii)$$

onde k é uma constante de taxa de reação de segunda ordem. Considerou-se para o cálculo uma concentração média para o radical HO[•] igual a $6,5 \times 10^5$ radicais cm⁻³ (Birks, 1998, p.241). Observe-se que essa média é dez vezes menor do que a usualmente assumida em algumas das pesquisas envolvendo o radical (ver, por exemplo, Wayne *et alli*, 1991).

Assumindo-se o valor para o tempo de vida dos NMHC, τ_{NMHC} , igual a 5 dias (comparável a valores de Birks, 1998, p.241), encontram-se valores para os tempos de vida dos NMHC_{médias}, τ_{NMHC} , e do NO₂, τ_{NO_2} , de modo que a sua relação fica:

$$\tau_{\text{NO}_2} / \tau_{\text{NMHC}} \cong 3,3 \times 10^{-4} \quad (iv)$$

Por outro lado, a relação entre os tempos de vida para o O₃ e o NO₂

$$\tau_{\text{O}_3} / \tau_{\text{NO}_2} = 1,0 \times 10^{-4} \quad (v)$$

é da mesma ordem de grandeza da equação (iv), indicando que, de fato, a razão dos fatores de tempo para que os NMHCs alcancem a etapa de fotólise, é a mesma que existe entre esta e a etapa de formação de O₃.

Assim, todas os passos anteriores a essa reação ocorrem rapidamente, tal que o tempo total entre a primeira etapa e a formação de NO[•]₂ é apenas um pouco maior do que 120 s. O próximo passo, a reação entre o átomo de oxigênio produzido pela decomposição de NO[•]₂, e a molécula de oxigênio presente na atmosfera, é, todavia, muito mais rápido ($\tau \sim 12 \times 10^{-3}$ s, ver reações q7.14 e q7.15, p380), tal que a decomposição de NO[•]₂ e, por extensão, a do NMHC, pode ser considerada como a formação de O₃. Assim, o perfil de acumulação do NMHC toma uma forma semelhante a do NO[•]₂, como mostra a Figura 7.15f.

Esse mesmo raciocínio pode ser aplicado às reações para o SO₂ (equações q7.89 a q7.94, p412). Todos os passos até a produção de NO[•]₂ são rápidos, tal que a taxa de consumo de SO₂ pode ser atribuída à reação de decomposição de NO[•]₂.

Para o SO₂ os cálculos levam aos seguintes resultados:

$$\tau_{\text{NO}_2} / \tau_{\text{SO}_2} \cong 1,0 \times 10^{-4} \quad (vi)$$

A equação vi tem mesma ordem de grandeza da equação v, tal que a reação do SO₂ com radicais hidroxila, embora menos veloz (por um fator de ~3) do que a dos NMHCs, também deve levar à formação de O₃ em tempo apenas um pouco maior do que o caso dos NMHCs.

Do mesmo modo que no caso anterior, o perfil para a acumulação do SO₂ apresenta alguma semelhança com o do NO[•]₂, com mostra a Figura 7.15a (p415).

7.6 – Discussão Final

Os resultados encontrados nessa tese evidenciaram fatos que poderão ser de utilidade em pesquisas relacionadas ao tema. Assim sendo, devem ser considerados como contribuições aos esforços realizados pelas diversas frentes de pesquisa que estudam o problema da mudança climática. Sob essa ótica, o resgate de alguns dos resultados obtidos (o que não significa prescindir dos demais), ajudará a delimitar o caminho que conduz às conclusões.

Formulou-se uma hipótese para explicar o pico noturno para o ozônio. Para esse caso a discussão está nas Seções 7.2 e 7.3. Entretanto, aquela discussão não levou em conta a possibilidade de que o mesmo mecanismo de reação usado para explicar os níveis de ozônio próximo ao solo, possa explicar a persistência nas concentrações de ozônio noturno em maiores altitudes. Nesse caso, a ocorrência da sequência de reações (abaixo) no período noturno, implicaria na manutenção de certos níveis de concentrações para o O_3 na baixa troposfera:



O $NO^{\bullet}_2$ formado pode reagir com ozônio para formar o radical $NO^{\bullet}_3$,



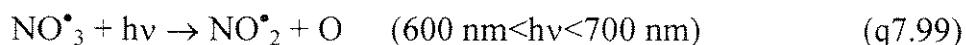
Se uma fração da reação q6.95 ocorrer como quimiluminescência,



A espécie $NO^{\bullet}_2$ pode perder energia por luminescência emitindo fótons de comprimentos de onda maiores do que 398 nm,



Os fótons podem interagir com radicais $\text{NO}_3^\bullet$ formados na reação q7.96, produzindo oxigênio atômico, o qual, em reação com oxigênio molecular, leva à formação de ozônio,



A seqüência de reações acima tem alguns pontos de “estrangulamento”. Dentre eles podem ser citados, por exemplo:

- 1) reações dos radicais $\text{NO}_3^\bullet$ com VOCs, especialmente alcenos, desde que a constante de taxa para essa reação com o $\text{NO}_3^\bullet$ é cerca de duas a quatro ordens de grandeza maior do que a reação com o O_3 (Dimitroulopoulou e Marsh, 1997),
- 2) reações dos radicais $\text{NO}_3^\bullet$ com $\text{NO}_2^\bullet$ gerando N_2O_5 , o qual pode, em reação com água formar HNO_3 (Dimitroulopoulou e Marsh, 1997),
- 3) deposição seca na superfície (Günsten *et alli*, 1998).

O primeiro deles remete para as reações com VOCs, especialmente os NMHCs, com as conseqüências já apontadas nas Seções 7.4 e 7.5 (produção de ozônio). O segundo, pode funcionar como sumidouro, mas é fortemente dependente da temperatura reinante. Por último, acredita-se que o rápido desaparecimento do ozônio aprisionado abaixo da camada de inversão noturna seja devido à deposição seca no solo (Colbeck e Harrison, 1985; Shepson *et alii*, 1992), embora não se possa desprezar a ocorrência de reações químicas com o ozônio se existem quantidades suficientes de outros poluentes nessa região da atmosfera (Harrison *et alli*, 1978).

Considerando-se as reações q7.95 a q7.100, pode-se verificar a instauração de um ciclo, no qual não há perda líquida de O_3 . Desse modo, na ausência de sumidouros eficazes, o ciclo acima mantém o O_3 em níveis de concentração tais que, a continuada emissão de poluentes formadores de ozônio devem impor um pequeno mas paulatino crescimento dos níveis de concentração de O_3 na troposfera.

Há um outro aspecto importante a considerar, especialmente para a localidade P. D. Pedro II que engloba maior parte dos resultados. Essa região é bem menos arborizada quando comparada, por exemplo, com o Parque do Ibirapuera. Assim, pode-se considerar que as emissões biogênicas, não obstante sua importância, têm uma influência bem menor nas reações locais e, em consequência, os NMHCs (especialmente o isopreno) emitidos pela vegetação, devem possuir um peso menor no conjunto das emissões que afetam o local.

Desse modo, sem desprezar a contribuição das outras fontes de VOCs tais como a evaporação de solventes, as emissões automotivas constituem-se, de fato como a principal fonte de emissão de poluentes no P. D. Pedro II.

Na discussão realizada para justificar a ausência de queda nas concentrações do ozônio durante os finais de semana, formulou-se a hipótese de que, dentre outras espécies (por exemplo NMHCs), o metano deveria ter alguma participação no fato.

Os resultados indicaram uma variação no perfil semanal para o metano que pode ser atribuída à sua participação na manutenção dos níveis elevados de O_3 nos finais de semana. Isto está expresso de forma especial nas análises com os modelos PARAFAC e de Tucker.

Algumas das pesquisas envolvendo a identificação de fontes de poluição atmosférica partem de informações *a priori*, as quais associam um certo conjunto de poluentes à determinada fonte.

Alsberg *et alli* (1989), valendo-se de um dos métodos de reconhecimento de padrões (SIMCA), identificou dois tipos de fontes automotivas (veículos diesel e veículos à gasolina), associando as emissões de cada tipo da mesma fonte, à “assinatura” que certos compostos possuem nos diferentes tipos da referida fonte. Adicionalmente, usou um composto como traçador com “assinatura” em apenas um dos tipos de fonte com o objetivo de confirmar os resultados.

Por meio de Análise de Componentes Principais (PCA) e Regressão Linear Múltipla (MLR) aplicadas a VOCs, Baldasano *et alli* (1998) identificaram emissões oriundas de veículos a gasolina e de posto de combustíveis como a principal fonte de

poluição em Martorell (Espanha). A segunda em importância, revelou-se ser fonte vinculada a indústrias, sendo que as fontes biogênicas contribuíram com apenas 6,0% do total das emissões.

Costa e Baldasano (1996), determinaram emissões de CO, NO, NO₂, SO₂, CH₄, partículas e VOCs (alcanos, alcenos, aromáticos e aldeídos) de várias fontes em Barcelona (Espanha). As emissões automotivas constituíram-se como a principal fonte de emissões de poluentes naquela cidade. As indústrias ocuparam a segunda posição, ficando as fontes biogênicas responsáveis por apenas 7,80% do total de emissões. Mais uma vez os resultados ficaram restritos à associação fonte × poluentes característicos de dada fonte. Isto é, fonte × “assinatura” do poluente àquela fonte (× = *versus*).

Statheropoulos *et alli* (1998) aplicaram Análise de Componentes Principais (PCA) e Análise de Correlação Canônica (CCA) a concentrações de CO, NO, NO₂, O₃, SO₂ e fumaça registrados em Estação de Monitoração na cidade de Atenas (Grécia). Seus resultados também indicaram emissões automotivas como a principal fonte de poluição na cidade. Nesse caso, também ocorreu a associação fonte × “assinatura” do poluente àquela fonte.

Esses resultados são importantes contribuições ao tema da mudança climática provocada por diferentes fontes de emissão. Porém é necessário que se identifique uma característica comum neles: em todas as cidades, a principal fonte de emissões foi sempre a automotiva, seguida de fonte industrial. Adicionalmente, as fontes biogênicas adquiriram importância apenas secundária no total das emissões.

Essas características reinantes naquelas cidades, são compartilhadas pelas localidades da RMSP analisadas nessa tese, particularmente Pedro II (local mais intensivamente investigado), onde as emissões automotivas assumem o papel de fonte de poluição mais importante. Além disto, a fonte automotiva adquire esse “status” com quase exclusividade, dados os valores encontrados para a primeira componente principal para todos os poluentes analisados na localidade. Isto é válido para outras localidades como Cerqueira César, Centro e Lapa, em menor extensão

para São Caetano do Sul. Por outro lado, há que se reconhecer um outro aspecto importante. Quando as fontes biogênicas comparecem em quantidades mais expressivas como no Parque do Ibirapuera, as emissões automotivas, embora significativas, caem em importância, perdendo o caráter de “quase exclusividade” adquirido em outros locais.

De qualquer modo, com base na predominância das emissões automotivas sobre as outras, maior preocupação foi dedicada à tentativa de se explicar a sua contribuição nos processos ocorrendo na baixa troposfera local e, conseqüentemente a sua participação na qualidade do ar na RMSP.

Não obstante as variáveis estudadas nessa tese sejam as mesmas usadas em pesquisas semelhantes, foi a metodologia empregada que permitiu, dar um passo além das considerações finais, isto é, chegar, de fato, a algumas conclusões.

Por meio de metodologia semelhante, Reis e Ferreira (2002), usando o modelo PARAFAC combinado com análise funcional, ajustaram variações sistemáticas diárias para verificar a presença de efeitos sazonais em emissões de monóxido de carbono na cidade de São Paulo durante os anos de 1997 e 1999. Para contornar o problema causado por variações súbitas, os autores trabalharam com os cinco chamados dias úteis da semana (2^a a 6^a feiras).

O que era indesejável na investigação de Reis e Ferreira (2002), poderia não sê-lo em outros estudos. De fato, as variações súbitas causadas por quedas nos finais de semana não foram consideradas como tal nessa tese porque as emissões sempre foram marcadas pela fonte automotiva. Nesse caso variações bruscas causadas por quedas nas emissões sempre nos fins de semana constituem-se em variações sistemáticas, quando avaliadas por longos períodos. E é exatamente a inconfundível queda nas emissões, causada pela fonte automotiva nos finais de semana, que permite distinção em relação a outras variações súbitas.

Assim, foi exatamente essa tendência de queda nos níveis para o metano nos finais de semana que indicou ser esse poluente um provável responsável, não só pela ausência de queda, como em ligeira elevação nos níveis de O₃ no referido período.

Esse fato está registrado nas Figuras 6.46 (p317), 6.49 (p322), 6.53 (p327) e 6.55 (p329), por meio das análises PARAFAC. Quando a análise é conduzida sob o modelo de Tucker o resultado é apontado nas Figuras 6.61 (p336), 6.64 (p341) 6.66 (p344), 6.68 (p346), 6.70 (p348) e 6.74 (p352).

Na análise com o modelo PARAFAC, é possível observar que a ligeira tendência de queda nos níveis de CH_4 está associada com uma grande queda nos níveis das emissões automotivas. Entretanto quando se faz a análise com o modelo de Tucker é possível verificar essas correlações com maior facilidade. As Figuras 6.61 (p336), 6.64 (p341) e 6.66 (p344) mostram as mudanças de inclinação para o perfil do CH_4 nos finais de semana. Aqui os perfis são semelhantes.

Entretanto, especial atenção deve ser dada, em primeiro lugar, às Figuras 6.68 (p346) e 6.70 (p348). A primeira delas (6.68) considera o ciclo do metano (CH_4 , NO , NO_2 e O_3) com o SO_2 incluído. Aqui a mudança de inclinação nos fins de semana para o CH_4 quase não é percebida. A explicação para isto deve estar relacionada com a maior reatividade do SO_2 em relação aos radicais $\text{HO}^\bullet$ quando comparada à do CH_4 . A segunda figura (6.70) considera a situação anterior, porém é introduzido o CO no ciclo. Nesse caso, é bastante evidente a queda nos níveis de CH_4 nos fins de semana, em relação a situação na qual o CO não estava presente.

Em segundo lugar, a atenção deve se voltar para a Figura 6.74 (p352), na qual se considera o ciclo do metano (CH_4 , NO , NO_2 e O_3) acrescido de SO_2 , NMHC e do CO . É bastante evidente a mudança de inclinação do perfil para o metano durante os finais de semana.

Os resultados acima indicam que a queda das emissões automotivas nos finais de semana estão, de algum modo, relacionadas com a participação do metano nos níveis de concentração de ozônio na localidade Pedro II. Observe-se que todos os ciclos contém três poluentes básicos (NO , NO_2 e O_3). Assim, quando se considera a queda nas emissões automotivas, a correlação positiva existente entre os níveis dos poluentes básicos NO e NO_2 , e os outros poluentes emitidos também por fonte

automotiva, mascara o fato caracterizado pela presença simultânea dos poluentes primários, tal que torna-se difícil distinguir o real comportamento deles.

A metodologia empregada nessa tese permitiu vislumbrar tais distinções sutis no comportamento dos poluentes. Os resultados evidenciam o fato de que a queda nos níveis de CO (CH_4 , NMHCs e SO_2), NO e NO_2 contribuem para diminuir os níveis de O_3 na atmosfera. Evidenciaram também o fato de que ao se introduzir o CO no ciclo para o metano, os níveis desse último caem quando comparados aos níveis alcançados na ausência do CO.

Akimoto *et alli* (2001), desenvolveram um método para analisar as respostas de O_3 e CH_4 a variações na quantidade de emissões de poluentes, usando um modelo tridimensional global de transporte químico (CTM), o qual combina um modelo global para o clima e um modelo de reações fotoquímicas. Os resultados obtidos em cálculos preliminares demonstraram que, embora a produção de ozônio caia quando somente as emissões de NO_x são reduzidas, a concentração de metano cresce, o que contribui para acelerar o aquecimento global no longo prazo. Entretanto, quando as emissões de NO_x e de CO são reduzidas simultaneamente, abate-se o crescimento de metano, tornando a medida efetiva para amenizar o aquecimento global (Akimoto *et alli*, 2001).

De fato, a única espécie efetiva na eliminação de CO e de CH_4 na atmosfera é o radical $\text{HO}^\bullet$. Assim, o metano e o dióxido de carbono devem competir em reações com o radical. Como a reatividade do CH_4 em relação ao $\text{HO}^\bullet$, é menor do que a do CO, e este é emitido em quantidades muito maiores em ambientes urbanos, somente uma queda nas emissões de CO (associada à queda simultânea nas emissões de NO_x) pode resultar em uma queda correspondente nos níveis de concentração de CH_4 na baixa troposfera, com reflexos tanto na queda dos níveis de concentração de ozônio, quanto da acumulação de ambos, CH_4 de médio e longo prazos, nessa camada da atmosfera.

CONCLUSÕES

Conclusões

Os resultados obtidos permitem concluir que a metodologia adotada e os modelos de ajuste aos dados:

- 1) são perfeitamente adequados para identificar fontes de poluição atmosférica,
- 2) tornaram possível determinar as variações sistemáticas diárias, as variações sistemáticas semanais, e a variação anual, para cada variável estudada durante os anos de 1997 a 1999 em várias localidades da RMSP, e apontar as emissões automotivas como a principal fonte de poluentes atmosféricos na mesma,
- 3) tornaram possível organizar as diversas reações químicas que ocorrem na baixa troposfera em três ciclos básicos tornando mais fácil a interpretação do comportamento dos poluentes,
- 4) tornaram possível determinar os perfis de acumulação de cada poluente estudado, à luz de suas reações químicas em fase gasosa na atmosfera,
- 5) tornaram possível sugerir as rotas preferenciais dos mecanismos possíveis para as várias reações químicas atmosféricas,
- 6) tornaram possível estimar, a partir do número de átomos de carbono na molécula do poluente sob investigação (para o caso dos hidrocarbonetos), as quantidades de ozônio produzidas sob condições de baixas ou altas concentrações de óxidos de nitrogênio emitidos para a atmosfera,

7) tornaram possível inferir sobre a principal fonte de metano na localidade P. D. Pedro II, na RMSP,

8) tornaram possível identificar a causa da formação de ozônio noturno (pico ~3:00–4:00 h) próximo à superfície,

9) tornaram possível identificar a causa da ausência de queda nas concentrações de ozônio durante os finais de semana,

10) tornaram possível associar a ausência de queda nas concentrações de ozônio, simultaneamente ao consumo de metano por meio de radicais hidroxila, e à queda nos níveis de concentração de monóxido de carbono.

11) tornaram possível associar uma queda nas concentrações de ozônio e de metano, à reduções simultâneas nas concentrações de óxidos de nitrogênio e de monóxido de carbono.

12) tornaram possível ampliar o espectro de possibilidades na construção de modelos preditivos para aplicação em poluição atmosférica.

APÊNDICES

APÊNDICE A1

A.1 – Métodos, Modelos, Algoritmos

• “The Power Method” – esse método é, na verdade, um procedimento para calcular os autovalores de uma matriz. O método opera com base no mesmo princípio de uma equação de diferença. Ele começa com uma suposição inicial, $\mathbf{u}_0$, e então forma sucessivamente $\mathbf{u}_1 = \mathbf{A} \mathbf{u}_0$, $\mathbf{u}_2 = \mathbf{A} \mathbf{u}_1$, e em geral $\mathbf{u}_{k+1} = \mathbf{A} \mathbf{u}_k$. Cada etapa é uma multiplicação matriz-vetor que, ao final de k etapas, produz $\mathbf{u}_k = \mathbf{A}^k \mathbf{u}_0$, embora a matriz $\mathbf{A}^k$ nunca apareça. Assumindo que $\mathbf{A}$ seja de posto completo e, portanto, possua um conjunto completo de autovetores $\mathbf{x}_1, \dots, \mathbf{x}_n$, o vetor $\mathbf{u}_k$ será dado pela fórmula usual para a equação da diferença (Strang, 1988, p.370)

$$\mathbf{u}_k = c_1 \lambda_1^k \mathbf{x}_1 + \dots + c_n \lambda_n^k \mathbf{x}_n \quad (\text{A.1.1})$$

Aplicado ao cálculo da variabilidade explicada pelas PCs o método parte da variabilidade explicada pela primeira PC, a qual é, por exemplo, $\mathbf{u}_1 \mathbf{u}_1^T$, e a subtrai da matriz de covariância, $\mathbf{S}$. Então as variâncias e covariâncias explicadas pela primeira componente principal, PC_1 são (Jackson, 1991, p.27)

$$\mathbf{S} - \mathbf{u}_1 \mathbf{u}_1^T = \mathbf{S}_1 \quad (\text{A.1.2})$$

onde $\mathbf{S}_2$ é a matriz de variâncias e covariâncias residuais. Tomando $\mathbf{u}_2 \mathbf{u}_2^T$ e subtraindo de $\mathbf{S}_1$ obtém-se nova matriz de covariâncias residuais

$$\mathbf{S}_1 - \mathbf{u}_2 \mathbf{u}_2^T = \mathbf{S}_2 \quad (\text{A.1.3})$$

onde $\mathbf{S}_2$ é a nova matriz de variâncias e covariâncias residuais.

Para n autovetores o resultado seria

$$\mathbf{S}_k - \mathbf{u}_n \mathbf{u}_n^T = \mathbf{S}_n \quad (\text{A.1.4})$$

Procede-se desse modo até que as variâncias e covariâncias sejam reduzidas aos menores valores possíveis. A combinação das equações (A.1.2), (A.1.3) e (A.1.4) leva a

$$\mathbf{S} - \mathbf{u}_1 \mathbf{u}_1^T - \mathbf{u}_2 \mathbf{u}_2^T - \dots - \mathbf{u}_n \mathbf{u}_n^T = \mathbf{S}_k \quad (\text{A.1.5})$$

a qual expressa de modo mais evidente a maneira como o método trabalha.

• Solução do Problema de Autovalor-Autovetor (eigenvalue analysis)

Um sistema pode ser descrito por suas variáveis, x_i , na base originalmente ocupada por elas. Uma transformação linear transforma o sistema descrito por essas variáveis em um sistema mais simples (um subespaço), descrito por novas variáveis, as quais são combinações lineares das variáveis originais, x_i (Noble e Daniel, 1988, p.280-281).

Isto pode ser realizado por meio da solução da equação fundamental para o problema do autovalor-autovetor (ver abordagem diferente no apêndice B),

$$\mathbf{S} \mathbf{p} = \lambda \mathbf{p} , \quad (\text{A.1.6})$$

onde λ e $\mathbf{p}$ são, respectivamente, um autovalor e seu correspondente autovetor, de $\mathbf{S}$.

A equação é não linear (envolve o produto de λ e $\mathbf{p}$ desconhecidos). Se o valor de λ puder ser encontrado, ela torna-se linear em $\mathbf{p}$ (Strang, 1988, p.245), tornando-se diretamente aplicável às transformações lineares de interesse.

As propriedades importantes de uma estrutura de autovalor-autovetor podem ser sumarizadas (Jobson, 1992, p.356) em um resultado chave da álgebra de matrizes (Jackson, 1991, p.7), válido tanto para duas como para $n > 2$ variáveis: uma matriz simétrica $n \times n$, não singular, tal como a matriz de covariância, $\mathbf{S}$, pode ser reduzida a uma matriz diagonal, Λ , pré e pós multiplicando-a por uma matriz particular ortonormal, $\mathbf{P}$, tal que:

$$\mathbf{P}^T \mathbf{S} \mathbf{P} = \Lambda \quad (\text{A.1.7})$$

Os elementos da diagonal da matriz Λ ($\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n$) são chamados *raízes características*, *raízes latentes* ou *autovalores* de $\mathbf{S}$. As colunas de $\mathbf{P}$ ($\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \dots, \mathbf{p}_n$) são chamadas *vetores característicos*, *vetores latentes* (termo pouco utilizado) ou *autovetores* de $\mathbf{S}$. Os autovalores podem ser obtidos através da solução de uma equação determinantal, chamada *equação característica*, a qual produz um polinômio do n -ésimo grau em λ , de onde os valores $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n$ são obtidos (Jackson, 1991, p.7-8).

Introduzindo-se a matriz identidade, $\mathbf{I}$, a equação (A.1.6) pode ser escrita como:

$$(\mathbf{S} - \lambda \mathbf{I}) \mathbf{p} = 0 \quad (\text{A.1.8})$$

Se $\mathbf{p} = 0$, qualquer valor atribuído a λ satisfaz a equação (A.1.8). Porém, interessam apenas aqueles valores particulares de λ para os quais existe um correspondente autovetor não nulo, $\mathbf{p}$ (Noble e Daniel, 1988, p.281-289; Strang, 1988, p.246). Para ser útil, o espaço de $\mathbf{S} - \lambda\mathbf{I}$ deve conter algum vetor não nulo (Strang,), ou seja, a matriz $\mathbf{S} - \lambda\mathbf{I}$ deve ser singular (Noble e Daniel, 1988, p.281-289; Strang, 1988, p.246), e a condição para λ ser um autovalor de $\mathbf{S}$, com um correspondente autovetor não nulo, ou seja, $\mathbf{S} - \lambda\mathbf{I}$ ser singular, é que (Noble e Daniel, 1988, p.281-289; Strang, 1988, p.246 ; Jobson, p.368):

$$\det (\mathbf{S} - \lambda\mathbf{I}) = 0 \quad (\text{A.1.9})$$

Da equação (A.1.9) calculam-se os autovalores, λ , e de (A.1.8) calculam-se os autovetores, $\mathbf{p}$, os quais devem ser normalizados.

Vale a pena explorar alguns resultados importantes decorrentes da equação (A.1.7). Na equação (A.1.6), por exemplo, a matriz $\mathbf{S}$ contém, na diagonal, as variâncias das variáveis originais, cuja soma é dada por $\text{tr}\mathbf{S} = s_{11} + s_{22} + \dots + s_{rr}$, onde r é o posto (número de colunas linearmente independentes) de $\mathbf{S}$. Então no subespaço caracterizado por $\lambda\mathbf{p}$, a matriz $\lambda\mathbf{I}$ deve conter as variâncias das componentes principais, cuja soma, deve, naturalmente, ser dada por $\text{tr}\mathbf{\Lambda}$.

De fato, a soma dos r autovalores é igual a soma das entradas na diagonal de $\mathbf{S}$, isto é:

$$s_{11} + s_{22} + \dots + s_{rr} = \lambda_1 + \lambda_2 + \dots + \lambda_r \quad (\text{A.1.10})$$

É interessante observar também que a matriz de covariância, $\mathbf{S}$ (de posto r), pode ser expressa em termos de $\mathbf{X}^t \mathbf{X}$, para $r = \min(m,n)$, onde $\mathbf{X}$ é matriz de dados originais (corrigidos pela média ou autoescalados), tal que a expressão (A.1.8) fica:

$$(\mathbf{X}^t \mathbf{X} - \lambda\mathbf{I}) \mathbf{p} = 0 \quad (\text{A.1.11})$$

$$\text{e, portanto, } \text{tr} \mathbf{S} = \text{tr} \mathbf{\Lambda} = \text{tr} \mathbf{X}^t \mathbf{X} = \sum_{j=1}^r \lambda_j \quad (\text{A.1.12})$$

Considerando que $\mathbf{S} = \mathbf{X}^T \mathbf{X}$ pode ser expresso como $\mathbf{t}_1^T \mathbf{t}_1 + \mathbf{t}_2^T \mathbf{t}_2 + \dots + \mathbf{t}_r^T \mathbf{t}_r$, onde $\mathbf{t}_i$ são os scores das observações para as variáveis originais, então a matriz $\lambda \mathbf{I} = \mathbf{\Lambda}$, pode ser expressa como $\mathbf{t}_{sc1}^T \mathbf{t}_{sc1} + \mathbf{t}_{sc2}^T \mathbf{t}_{sc2} + \dots + \mathbf{t}_{scr}^T \mathbf{t}_{scr}$, onde $\mathbf{t}_{sci}$ são os scores das observações para as componentes principais. Pode-se depreender daí então que a matriz transformação possui pares de autovetores tais que um dos vetores do par torna máxima a variação na direção de uma componente principal, enquanto o outro vetor torna mínima a variação na direção da outra componente, ortogonal a primeira.

Além disso, independentemente de λ ser um autovalor, a maioria dos vetores $\mathbf{p}$ não irá satisfazer a equação (A.1.6). Um vetor $\mathbf{p}$ típico muda de direção quando é multiplicado por $\mathbf{S}$, tal que o produto $\mathbf{Sp}$ não é um múltiplo de $\mathbf{p}$. Isto significa que somente certos números especiais λ são autovalores, e somente certos vetores especiais $\mathbf{p}$ são autovetores (Strang, 1988, p.249).

• **Non Iterative Partial Least Squares – NIPALS**

É oportuno acrescentar que esta ainda é uma ferramenta bastante utilizada na computação de PCA. É o NIPALS (Nonlinear Iterative Partial Least Squares).

Pode-se entender PCA como um método que permite escrever uma Matriz $\mathbf{X}$ de posto r como uma soma de r matrizes de posto 1 (Geladi and Kowalski, 1986; Wold et alli, 1987):

$$\mathbf{X} = \mathbf{M}_1 + \mathbf{M}_2 + \mathbf{M}_3 + \dots + \mathbf{M}_r, \quad (\text{A.1.13})$$

onde denomina-se posto a um número que expressa a dimensionalidade verdadeira subjacente a uma dada matriz (Geladi and Kowalski, 1986; Wold *et alli*, 1987):

O desdobramento de $\mathbf{X}$ nas matrizes $\mathbf{M}_i$, $i = 1, 2, \dots, r$, decorre do resultado obtido na equação (50), $\mathbf{T} = \mathbf{X} \mathbf{P}$.

De fato, pós-multiplicando-se por $\mathbf{P}^T$ a equação (14), $\mathbf{X} \mathbf{P} = \mathbf{T}$, obtém-se:

$$\mathbf{X} \mathbf{P} \mathbf{P}^T = \mathbf{T} \mathbf{P}^T \quad (\text{A.1.14})$$

Como $\mathbf{P}$ é ortogonal, $\mathbf{P} \mathbf{P}^T = \mathbf{I}$, portanto,

$$\mathbf{X} = \mathbf{T} \mathbf{P}^T \quad (\text{A.1.15})$$

As matrizes de posto 1, $\mathbf{M}_h$, podem ser escritas como produtos externos (Geladi and Kowalski, 1986; Wold et alli, 1987;), também referidas como relações externas (Brown, 1995), de dois vetores, um de scores, $\mathbf{t}_h$, e um autovetor (loading) $\mathbf{p}_h^T$,

$$\mathbf{X} = \mathbf{t}_1 \mathbf{p}_1^T + \mathbf{t}_2 \mathbf{p}_2^T + \mathbf{t}_3 \mathbf{p}_3^T + \dots + \mathbf{t}_r \mathbf{p}_r^T, \quad (\text{A.1.16})$$

resultado equivalente ao da equação (A.1.15).

Em duas dimensões, como já visto, $\mathbf{p}_h^T$ é um vetor linha 1×2 , cujos elementos p_1 e p_2 são os cossenos diretores, ou as projeções, sobre a base original, de um vetor unitário na

direção da componente principal. O vetor de scores, $\mathbf{t}_h$, é um vetor coluna $n \times 1$, cujos elementos são as coordenadas dos respectivos pontos (observações) sobre o eixo da componente principal (Geladi and Kowalski, 1986).

O algoritmo NIPALS (esquema abaixo) calcula $\mathbf{t}_1$ e $\mathbf{p}_1^T$ da matriz $\mathbf{X}$; em seguida o produto externo, $\mathbf{t}_1 \mathbf{p}_1^T$, é subtraído de $\mathbf{X}$ e é calculado o resíduo, $\mathbf{E}_1$. Esse resíduo é usado para calcular $\mathbf{t}_2 \mathbf{p}_2^T$, o qual é subtraído de $\mathbf{E}_1$, e assim sucessivamente até a iteração convergir para um valor suficientemente pequeno previamente estipulado.

O algoritmo NIPALS é fornecido abaixo (Geladi and Kowalski, 1986; Martens e Naes, 1996, p.111-112):

$$(1) \text{ tome um vetor } \mathbf{x}_j \text{ de } \mathbf{X} \text{ e denomine-o } \mathbf{t}_h: \mathbf{t}_h = \mathbf{x}_j \quad (\text{A.1.17})$$

$$(2) \text{ calcule } \mathbf{p}_h^T: \mathbf{p}_h^T = (\mathbf{t}_h^T \mathbf{X}) / (\mathbf{t}_h^T \mathbf{t}_h) \quad (\text{A.1.18})$$

$$(3) \text{ normalize } \mathbf{p}_h \text{ ao comprimento 1: } (\mathbf{p}_{h\text{ novo}})^T = (\mathbf{p}_{h\text{ anterior}})^T / \|(\mathbf{p}_{h\text{ anterior}})^T\| \quad (\text{A.1.19})$$

$$(4) \text{ calcule } \mathbf{t}_h: \mathbf{t}_h = (\mathbf{X} \mathbf{p}_h) / (\mathbf{p}_h^T \mathbf{p}_h) \quad (\text{A.1.20})$$

(5) compare $\mathbf{t}_h$ usado no passo (2) com aquele obtido no passo (4). Se eles são o mesmo, pare (a iteração convergiu). Se eles ainda diferem, vá para o passo (2). Observe que após a primeira componente ser calculada, deve-se trocar $\mathbf{X}$ nas etapas (2) e (4) por seu resíduo (passo seguinte).

(6) calcule a matriz de resíduos: $\mathbf{E}_1 = \mathbf{X} - \mathbf{t}_h \mathbf{p}_h^T$. Retorne ao passo (1) usando a matriz de resíduos, $\mathbf{E}_1$, para o cálculo dos vetores de “scores” e de “loadings” da próxima componente principal.

Levando-se em conta a equação (A.1.11) e as considerações no seu entorno, é possível ter uma melhor compreensão de como funciona o método NIPALS. Observe que nas equações (A.1.18), (A.1.19) e (A.1.20), respectivamente, $\mathbf{t}_h^T \mathbf{t}_h$, $\|\mathbf{p}_h\|$ e $\mathbf{p}_h^T \mathbf{p}_h$ são escalares, os quais podem ser combinados em uma constante geral C (Geladi and Kowalski, 1986). Escrevendo as equações (A.1.13) e (A.1.14), acima, como

$$\mathbf{t}_h^T (\mathbf{t}_h^T \mathbf{t}_h) = \mathbf{t}_h^T \mathbf{X} \quad (\text{A.1.21})$$

$$\mathbf{t}_h (\mathbf{p}_h^T \mathbf{p}_h) = \mathbf{X} \mathbf{p}_h \quad (\text{A.1.22})$$

Tomando-se a transposta de (A.1.22) e substituindo-se em (A.1.21) obtém-se:

$$\mathbf{p}_h^T \mathbf{C} = \mathbf{p}_h^T \mathbf{X}^T \mathbf{X}, \quad (\text{A.1.23})$$

cujas transpostas resultarão em:

$$(\mathbf{C} \mathbf{I}_m - \mathbf{X}^T \mathbf{X}) \mathbf{p}_h = 0 \quad (\text{A.1.24})$$

Remanejando a equação (A.1.23), obtém-se:

$$\mathbf{p}_h^T (\mathbf{X}^T \mathbf{X}) = C_0 \mathbf{p}_h^T, \quad (\text{A.1.25})$$

cujas transpostas fornecerá, finalmente:

$$(\mathbf{X}^T \mathbf{X} - C_0 \mathbf{I}_m) \mathbf{p}_h = 0 \quad (\text{A.1.26})$$

Observe a semelhança entre as equações (A.1.26) e (A.1.11). C_0 encontrado na equação (A.1.26) corresponde a λ (ou a um múltiplo seu) na equação (A.1.11).

Por um procedimento análogo pode-se encontrar também que:

$$(\mathbf{X} \mathbf{X}^T - C_0 \mathbf{I}_n) \mathbf{t}_h = 0 \quad (\text{A.1.27})$$

Essas são as equações clássicas de autovalor-autovetor para $\mathbf{X}^T \mathbf{X}$ e $\mathbf{X} \mathbf{X}^T$, nas quais $\mathbf{I}_m$ é a matriz identidade de tamanho $m \times m$, e $\mathbf{I}_n$ é a matriz identidade de tamanho $n \times n$ (Geladi and Kowalski, 1986).

APÊNDICE A2

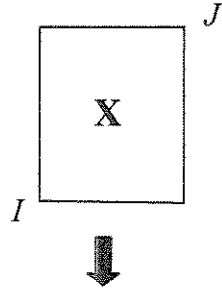
Aqui os termos "modes" e "ways" são traduzidos da literatura corrente, respectivamente, como "modos" e "vias". Esses termos são usados com um mesmo significado, ou seja, o termo "modo" é usado como tradução tanto para "mode" quanto para "way", e o contexto deverá indicar ao leitor em que sentido ele deverá interpretá-lo ao fazer leituras comparativas em estudos orientados. Deve ser enfatizado que a adoção dessa terminologia não traz qualquer prejuízo para a compreensão do assunto.

A título de esclarecimento, segue uma explicação para a existência de distinção entre aqueles termos.

Alguns autores fazem distinção entre os termos "mode" e "way". O termo "via" (way) deverá ser entendido como uma coleção de índices através dos quais os dados podem ser classificados, enquanto o termo "modo" (mode) referir-se-á às entidades que compõem cada via do arranjo de dados (Kroonenberg, 1994). Nesse caso, se os três modos pertencem a três diferentes conjuntos de entidades (três modos independentes, objetos $\times$ variáveis $\times$ localidades), o arranjo é denominado de "dados em três-modos três-vias" (three-mode three-way data) ou simplesmente "dados em três-vias" (three-way data). Se dois dos modos pertencem ao mesmo conjunto de entidades (caso de correlações entre variáveis), então somente dois modos diferentes estão envolvidos passando o arranjo a se chamar "dados em dois-modos três-vias" (two-mode three-way data). Finalmente, quando os três modos pertencem a um mesmo conjunto de entidades o arranjo é denominado de "dados em um-modo três-vias" (one-mode three-way data) (Kiers, 2000).

APÊNDICE A3

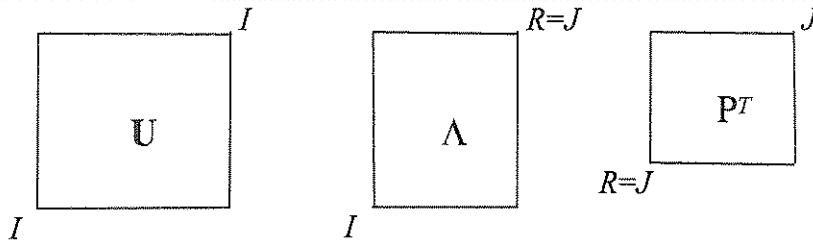
$\mathbf{X}(I \times J)$, $J < I$, posto completo ($R = \min\{I, J\}$)



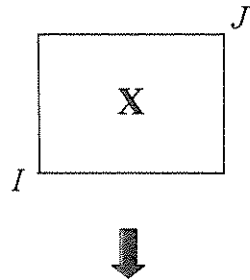
$$\mathbf{X}_{(I,J)} = \mathbf{U}_{(I,R)} \mathbf{\Lambda}_{(R,R)} \mathbf{P}_{(R,J)}^T$$

$$\mathbf{X}_{(I,J)} = \mathbf{U}_{(I,R)} \mathbf{\Lambda}_{(R)} \mathbf{P}_{(R)}^T$$

$$\mathbf{X} = \mathbf{U}_I \mathbf{\Lambda}_R \mathbf{P}_R^T$$



$\mathbf{X}(I \times J)$, $I < J$, posto completo ($R = \min\{I, J\}$)



$$\mathbf{X}_{(I,J)} = \mathbf{U}_{(I,R)} \mathbf{\Lambda}_{(R,J)} \mathbf{P}_{(J,J)}^T$$

$$\mathbf{X}_{(I,J)} = \mathbf{U}_{(I,R)} \mathbf{\Lambda}_{(R,J)} \mathbf{P}_{(J,J)}^T$$

$$\mathbf{X} = \mathbf{U}_R \mathbf{\Lambda}_R \mathbf{P}_J^T$$

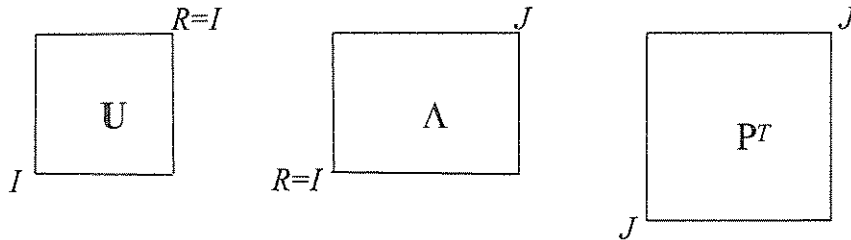
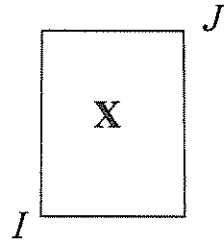


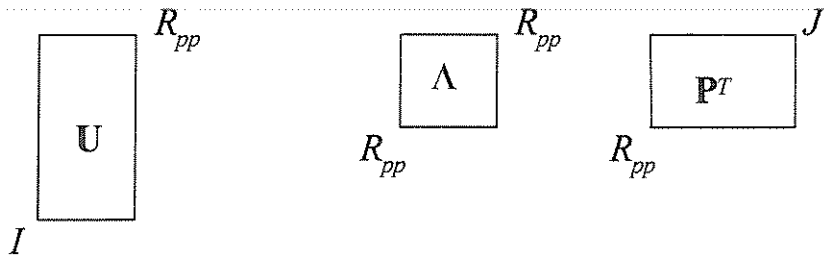
Fig.A.3.1: SVD para uma matriz $\mathbf{X}(I \times J)$ considerando-se posto completo ($R = \min\{I, J\}$, com $J < I$, acima, e $I < J$, abaixo. (R = posto de $\mathbf{X}$).

$\mathbf{X}(I \times J)$, $J < I$, deficiente de posto ($R < \min\{I, J\}$)

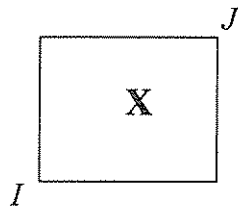


$$\mathbf{X} = \mathbf{U}_{(I,R)} \mathbf{\Lambda}_{(R,R)} \mathbf{P}_{(R,J)}^T$$

$$\mathbf{X} = \mathbf{U}_R \mathbf{\Lambda}_R \mathbf{P}_R^T$$



$\mathbf{X}(I \times J)$, $I < J$, deficiente de posto ($R < \min\{I, J\}$)



$$\mathbf{X} = \mathbf{U}_{(I,R)} \mathbf{\Lambda}_{(R,R)} \mathbf{P}_{(R,J)}^T$$

$$\mathbf{X} = \mathbf{U}_R \mathbf{\Lambda}_R \mathbf{P}_R^T$$

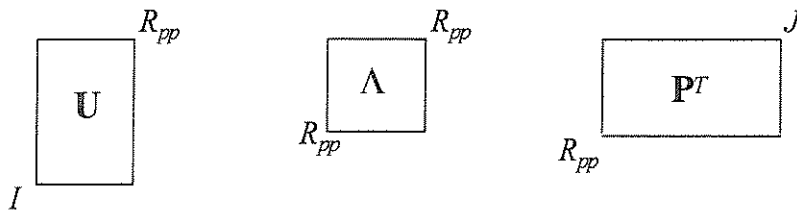


Fig.A.3.2: SVD para uma matriz $\mathbf{X}(I \times J)$ considerando-se matriz deficiente de posto com ($R_{pp} < \min\{I, J\}$, com $R_{pp} < J < I$, acima, e $R_{pp} < I < J$, abaixo. (R_{pp} = pseudo posto de $\mathbf{X}$).

APÊNDICE A4

No corpo do texto supõe-se uma coleta de dados constituída de I objetos, J variáveis e K localidades, os quais poderiam ser analisados via PCA convencional de duas maneiras:

a) Organizando-se os dados justapostos verticalmente (ver Figura 3.13 no texto da tese) tal que sejam empilhadas K matrizes de dimensões $I \times J$. Nesse caso, a nova matriz obtida constitui-se de diferentes conjuntos de objetos (um conjunto para cada uma das K localidades) sob as mesmas variáveis.

b) Organizando-se os dados justapostos horizontalmente (ver Figura 3.13 no texto da tese) tal que sejam reunidas lado a lado K matrizes de diferentes conjuntos de variáveis (um conjunto para cada uma das K localidades).

Quando as matrizes são organizadas horizontalmente (variáveis projetadas no subespaço dos objetos) obtém-se informações sobre a influência de cada variável sobre cada objeto “pesadas” pela localidade, isto é, a inter-relação dos objetos “pesados” pela localidade. Quando são organizadas verticalmente (objetos projetados no subespaço das variáveis), obtém-se a inter-relação das variáveis “pesadas” pela localidade.

Todavia, quando o índice K representa o período no qual cada conjunto $I \times J$ foi levantado em uma mesma localidade, a justaposição horizontal dá informações sobre a variação sazonal. Nesse caso, a justaposição vertical continua fornecendo informações sobre a inter-relação das variáveis “pesadas” pela localidade, independentemente dos conjuntos $I \times J$ terem sido tomados em períodos diferentes.

Os casos são denominados conjuntos múltiplos de dados (multiple data sets) e não podem ser considerados como dados de ordem superior porque os três modos (localidades, objetos e variáveis) não são totalmente cruzados na análise (Kiers, 1991).

APÊNDICE A5

Operadores Matemáticos:

$\text{tr}(\mathbf{X})$ = soma dos elementos da diagonal de $\mathbf{X}$

$\text{Vec}(\mathbf{X})$ = estrutura obtida quando os vetores de uma matriz são organizados um sobre o outro, isto é, são "enfileirados verticalmente" ("vertical stringout"). Por exemplo:

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_J] \rightarrow \text{Vec}(\mathbf{X}) = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_J \end{bmatrix}$$

$\mathbf{X} \otimes \mathbf{Y}$ = tensor ou produto de Kronecker de $\mathbf{X}$ por $\mathbf{Y}$, com $\mathbf{X}(I \times J)$, é definido como:

$$\mathbf{X} \otimes \mathbf{Y} = \begin{bmatrix} x_{11}\mathbf{Y} & \cdots & x_{1J}\mathbf{Y} \\ \vdots & & \vdots \\ x_{I1}\mathbf{Y} & \cdots & x_{IJ}\mathbf{Y} \end{bmatrix}$$

$$\|\cdot\|_F^2 = \text{quadrado da norma de Frobenius} = \sum_i \sum_j \mathbf{X}_{ij}^2 = \text{tr}(\mathbf{X}^T \mathbf{X})$$

Desdobramento de uma estrutura de Dados de Ordem Superior em termos de suas matrizes de componentes e do seu núcleo para o modelo de Tucker.

Seja dado $\underline{\mathbf{X}}(I \times J \times K)$, com $\mathbf{A}(I \times P)$, $\mathbf{B}(J \times Q)$ e $\mathbf{C}(K \times R)$. Cada elemento da estrutura é representado por:

$$x_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R a_{ip} b_{jq} c_{kr} g_{pqr} + e_{ijk}$$

Escrito de forma algébrica, o arranjo fica:

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

O desdobramento da estrutura de ordem superior em suas matrizes de componentes, representado na forma algébrica acima, $\mathbf{X}$, é mostrado na próxima página.

Continuação Apêndice A5

Desdobramento de uma estrutura de Dados de Ordem Superior em termos de suas matrizes de componentes e do seu núcleo para o modelo de Tucker.

$$\mathbf{X} = \begin{bmatrix} a_{11}a_{12} \cdots a_{1P} \\ a_{21}a_{22} \cdots a_{2P} \\ \vdots \\ a_{n1}a_{12} \cdots a_{1P} \end{bmatrix} \begin{bmatrix} g_{111}g_{112} \cdots g_{11R} & g_{211}g_{212} \cdots g_{21R} & \vdots & g_{P11}g_{P12} \cdots g_{P1R} \\ g_{121}g_{122} \cdots g_{12R} & g_{221}g_{222} \cdots g_{22R} & \vdots & g_{P21}g_{P22} \cdots g_{P2R} \\ \vdots & \vdots & \vdots & \vdots \\ g_{1Q1}g_{1Q2} \cdots g_{1QR} & g_{2Q1}g_{2Q2} \cdots g_{2QR} & \vdots & g_{PQ1}g_{PQ2} \cdots g_{PQR} \end{bmatrix} \begin{bmatrix} c_{11}^T \begin{bmatrix} b_{11}b_{21} \cdots b_{J1} \\ \vdots \\ b_{1Q}b_{2Q} \cdots b_{JQ} \end{bmatrix} \\ c_{21}^T \begin{bmatrix} b_{11}b_{21} \cdots b_{J1} \\ \vdots \\ b_{1Q}b_{2Q} \cdots b_{JQ} \end{bmatrix} \\ \vdots \\ c_{KR}^T \begin{bmatrix} b_{11}b_{21} \cdots b_{J1} \\ \vdots \\ b_{1Q}b_{2Q} \cdots b_{JQ} \end{bmatrix} \end{bmatrix}$$

APÊNDICE A6

A discussão que conduziu à proposta de formulação de um modelo para os dados, considerou a existência de um perfil diário e um semanal, os quais adquirem importância específica no comportamento dos poluentes na atmosfera.

Para a discussão deve-se ter em mente que os dados referentes aos valores de concentração usados nessa tese não se constituem de medições realizadas por meio de instrumentação instalada na própria fonte de emissão (por exemplo, saída dos sistemas de descarga dos veículos, ou nas chaminés das indústrias). A rede telemétrica da CETESB está instalada em determinados locais na RMSP por meio de um critério de escolha. Desse modo, excetuando-se as unidades móveis (também funcionam sob o mesmo critério), os pontos de amostragem são fixos. Isto significa que as variáveis medidas devem percorrer uma dada distância entre o momento da emissão e o do registro dessa emissão. Para as fontes fixas (indústrias), tais distâncias são aproximadamente constantes. Para fontes móveis (automotivas) as distâncias variam. Isto funciona como um obstáculo ao registro fiel dos valores de concentrações dos poluentes, portanto, dificultando a medição, a qual está, desse modo, sujeita à variações nas condições climáticas. O modelo deve considerar esse fato, ou seja, que as condições climáticas afetam as medições.

Assumiu-se, então que os dados recebem dois tipos de contribuição: 1) a da fonte de emissão (colocada nas colunas da matriz **B**), e 2) a do grau de dificuldade de se realizar a medição (colocada nas colunas da matriz **A**).

A contribuição da fonte de emissão associa as concentrações medidas às fontes de emissão. O grau de dificuldade em se realizar a medição associa as concentrações às variações nas condições climáticas.

1) contribuição da fonte de emissão: **B**(horas $\times$ fontes)

2) grau de dificuldade de se realizar a medição **A**(dias $\times$ fontes)

Para que PCA possa captar as variações sistemáticas diárias sob a influência da fonte e das condições climáticas, a decomposição da matriz de dados resulta em:

$$\mathbf{X}(365 \text{ dias} \times 24 \text{ horas}) = \mathbf{A}(\text{dias} \times \text{fontes}) \mathbf{B}^T(\text{fontes} \times \text{horas})$$

A decomposição acima captura o perfil diário sob a influência de mudanças nas condições climáticas a cada dia durante o ano. Posta dessa maneira, PCA não pode fornecer informações sobre as variações sistemáticas semanais, porque isto exige que se imponha na decomposição a influência de mudanças nas condições climáticas, não a cada dia, mas sim, a cada semana. Para isto, modificação adequada deve ser realizada na matriz de dados. Nesse caso, a decomposição conduz a:

$$\mathbf{X}_{ps}(52 \text{ semanas} \times 7 \text{ dias}) = \mathbf{A}(\text{semanas} \times \text{fontes}) \mathbf{B}^T(\text{fontes} \times \text{dias})$$

onde $\mathbf{X}_{ps}$ corresponde a matriz modificada de modo a fornecer o perfil semanal na análise PCA.

Quando os modelos multimodo são usados, a decomposição da matriz de dados conduz simultaneamente às variações sistemáticas diárias e às variações sistemáticas semanais. Adicionalmente, fornece a variação anual. Toma-se como ponto de partida a mesma matriz, $\mathbf{X}(365 \text{ dias} \times 24 \text{ horas})$, usada em PCA, porém deve-se ajustar os modos adequadamente para que a decomposição reproduza corretamente os perfis procurados.

APÊNDICE A7

Os dados fornecidos pela CETESB apresentam as características da tabela abaixo. As matrizes foram construídas considerando-se quatro colunas: 1)data do registro; 2)hora do registro; 3)valores de concentrações; e 4)nome da Estação de Medição.

```
"01/01/1999";"01:00";27,39;"804000000";"VALID";1,00;"P. D. Pedro II"  
"01/01/1999";"02:00";44,77;"804000000";"VALID";1,00;"P. D. Pedro II"  
"01/01/1999";"03:00";33,03;"804000000";"VALID";1,00;"P. D. Pedro II"  
"01/01/1999";"04:00";32,39;"804000000";"VALID";1,00;"P. D. Pedro II"  
"01/01/1999";"05:00";24,61;"804000000";"VALID";1,00;"P. D. Pedro II"  
"01/01/1999";"06:00";19,20;"80400200";"NOT VALID";1,00;"P. D. Pedro II"  
"01/01/1999";"07:00";17,85;"804000000";"VALID";1,00;"P. D. Pedro II"  
"01/01/1999";"08:00";24,74;"804000000";"VALID";1,00;"P. D. Pedro II"  
"01/01/1999";"09:00";34,11;"804000000";"VALID";1,00;"P. D. Pedro II"  
"01/01/1999";"10:00";32,99;"804000000";"VALID";1,00;"P. D. Pedro II"  
"01/01/1999";"11:00";34,56;"804000000";"VALID";1,00;"P. D. Pedro II"  
"01/01/1999";"12:00";21,15;"804000000";"VALID";1,00;"P. D. Pedro II"  
"01/01/1999";"13:00";22,72;"804000000";"VALID";1,00;"P. D. Pedro II"  
"01/01/1999";"14:00";23,23;"804000000";"VALID";1,00;"P. D. Pedro II"  
"01/01/1999";"15:00";19,08;"804000000";"VALID";1,00;"P. D. Pedro II"  
"01/01/1999";"16:00";24,26;"804000000";"VALID";1,00;"P. D. Pedro II"  
"01/01/1999";"17:00";37,96;"804000000";"VALID";1,00;"P. D. Pedro II"  
"01/01/1999";"18:00";55,29;"804000000";"VALID";1,00;"P. D. Pedro II"  
"01/01/1999";"19:00";73,08;"804000000";"VALID";1,00;"P. D. Pedro II"  
"01/01/1999";"20:00";74,83;"804000000";"VALID";1,00;"P. D. Pedro II"  
"01/01/1999";"21:00";77,45;"804000000";"VALID";1,00;"P. D. Pedro II"  
"01/01/1999";"22:00";79,36;"804000000";"VALID";1,00;"P. D. Pedro II"  
"01/01/1999";"23:00";79,50;"804000000";"VALID";1,00;"P. D. Pedro II"  
"01/01/1999";"24:00";74,14;"804000000";"VALID";1,00;"P. D. Pedro II"  
"02/01/1999";"01:00";66,39;"804000000";"VALID";1,00;"P. D. Pedro II"  
"02/01/1999";"02:00";58,52;"804000000";"VALID";1,00;"P. D. Pedro II"
```

APÊNDICE A8

O Apêndice 8 está dividido em seis partes. Cada uma delas refere-se a uma rotina desenvolvida. Inserido em algumas das rotinas encontram-se comandos próprios de certos programas utilizados nas análises.

APÊNDICE A8.1

Rotina desenvolvida para construção de uma matriz contendo quatro colunas:

1)data da medição; 2)hora da medição; 3)concentração do poluente; e 4)Estação de Medição.

```
%Início
cd c:\path; %Caminho para o arquivo
fid=fopen('co99.txt','rt'); % Abre o arquivo
line=fgets(fid); % Toma a primeira linha
Ok=1; %Inicia o "loop"
cont=1;
while Ok

%Organiza Estações de Medição (lab)

    ind=find(line==' ');
    lab(cont,1:size([line(ind(size(ind,2))+1:size(line,2))],2))=line(ind(size(ind,2))+1:size(line,2));
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Organiza data da medição (data)

    data_tmp=line(2:ind(1)-2);
    ind_data=find(data_tmp==' ');
    data(cont,:)= [sscanf([data_tmp(1:ind_data(1)-1)],'%f'),...
                   sscanf([data_tmp(ind_data(1)+1:ind_data(2)-1)],'%f'),...
                   sscanf([data_tmp(ind_data(2)+1:size(data_tmp,2))],'%f') ];

    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Organiza horário da medição (hora)

    hora_tmp=line(ind(1)+2:ind(2)-2);
    ind_hora=find(hora_tmp==' ');
    hora(cont,:)= [sscanf([hora_tmp(1:ind_hora(1)-1)],'%f'),...
                   sscanf([hora_tmp(ind_hora(1)+1:size(hora_tmp,2))],'%f') ];

    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Organiza valor da concentração (nome e ano da variável , por ex: CO99)

    co99_tmp= line(ind(2)+1:ind(3)-1);
    ind_conc=find(co99_tmp==' ');
    co99_tmp(ind_conc)='.';
    co99(cont,1)= sscanf(co99_tmp,'%f');

    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Verifica a última linha do arquivo original (fornecido pela CETESB)

```

```

        if all(lab(cont,2:10)=='[Cerqueira]') & all(data(cont,:)==[09 11 1999]) &
all(hora(cont,:)==[17 0])
            Ok=0;
        end;
        cont=cont+1;
        clear line
        line=fgets(fid);
        if all(size(line)==[1 1]) & Ok~=0
            while all(size(line)==[1 1])
                line=fgets(fid);
            end;
        end;
end;
end;

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```

%Para salvar os dados em arquivo:

```

```

cd c:\path\co\co99 %Indica o arquivo para salvar os dados

```

```

save co99.dat co99 -ascii
save data.dat data -ascii
save hora.dat hora -ascii
save lab lab

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```

%Para carregar os dados para análise

```

```

cd c:\path\co\co99 %Carrega os dados para início da análise

```

```

load co99.dat
load data.dat
load hora.dat
load lab

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

APÊNDICE A8.2

Rotina desenvolvida para acessar os dados da matriz central (rotina 1). Uma função denominada “estac” resgata, sob o nome da Estação de Medição, os dados de uma dada localidade para dar início à análise. A função leva o nome “estac” porque os outros títulos são comuns a todas as localidades (data, horário e concentração). Um vetor contendo o número de dias de cada mês (num_dia_mes) organiza todas as horas do dia dentro de um dado mês. Outro vetor organiza os dias na semana.

%Vetor para os organizar os dias dentro do mês (esse vetor deve preceder a “chamada” da %função “estac”:

```
num_dia_mes=[31 28 31 30 31 30 31 31 30 31 30 31];
```

%Comando para “chamar” a função “estac” (para cada Estação de Medição):

%”Cerqueira Cesar”

```
[Cesar,Ces_nan,ind_ces]=estac(co99,data,hora,lab,['Cerq'],num_dia_mes);
```

%”Centro”

```
[Centro,Cent_nan,ind_cent]=estac(co99,data,hora,lab,['Cent'],num_dia_mes);
```

%”Lapa”

```
[Lapa,Lap_nan,ind_lap]=estac(co99,data,hora,lab,['Lapa'],num_dia_mes);
```

%”Osasco”

```
[Osasco,Osc_nan,ind_osc]=estac(co99,data,hora,lab,['Osas'],num_dia_mes);
```

%”P. D. Pedro II”

```
[PedroII,Ped_nan,ind_ped]=estac(co99,data,hora,lab,['P. D'],num_dia_mes);
```

%”São Caetano do Sul”

```
[Sao_Caetano,Sao_nan,ind_sao]=estac(co99,data,hora,lab,['São'],num_dia_mes);
```

%”Ibirapuera”

```
[Ibir,Ibir_nan,ind_ibir]=estac(co99,data,hora,lab,['Ibir'],num_dia_mes);
```

%%%

%Vetor para organizar os dias na semana (em 1999, o primeiro dia do ano foi uma 6ª feira).
%Rótulos utilizados: s = Sexta-feira; sa = Sábado; do = Domingo; se = Segunda-feira;
%te = Terça-feira; qa = Quarta-feira; qi = Quinta-feira)

```
dias=['s','sa','do','se','te','qa','qi']; %Vetor contendo os dias da semana
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
%Rotina para grafar os rótulos referentes aos dias da semana nos gráficos (pode ser usado  
%no momento da construção dos gráficos correspondentes. Ver Apêndice A8.4):
```

```
cont=1;  
for i=1:52  
    for j=1:7  
        dias_ano(cont,:)=dias(j,:);  
        cont=cont+1;  
    end;  
end;
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
%Função “estac”:
```

APÊNDICE A8.3

Rotina construída para verificação da influência causada por dados faltantes:

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
%Dados Faltantes (Ex: Pedro II):
```

```
X=Pedro_99(1:364,:);
```

```
[T,P,Xest] = mdpca(X,3,999999999999);
```

```
cd c:\path\Dados_Falt
```

```
load ind_miss.dat %Matriz contendo a distribuição aleatória ao longo dos 364 dias do ano:
```

```
for i=1:10
```

```
X=Pedro_99(ind_miss(:,i),1:2:24);
```

```
[t,p,Xest] = mdpca(X,3,999999999999);
```

```
    t1(:,i)=t(:,1);
```

```
    t2(:,i)=t(:,2);
```

```
    t3(:,i)=t(:,3);
```

```
    p1(:,i)=p(:,1);
```

```
    p2(:,i)=p(:,2);
```

```
    p3(:,i)=p(:,3);
```

```
X=Pedro_99(ind_miss(:,i),2:2:24);
```

```
[t,p,Xest] = mdpca(X,3,999999999999);
```

```
    t21(:,i)=t(:,1);
```

```
    t22(:,i)=t(:,2);
```

```
    t23(:,i)=t(:,3);
```

```
    p21(:,i)=p(:,1);
```

```
    p22(:,i)=p(:,2);
```

```
    p23(:,i)=p(:,3);
```

```
end;
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
%Matriz ind_miss (Ao final dos Apêndices, mas pode ser gerada aleatoriamente):
```

APÊNDICE A8.4

%Construção de gráficos (matriz de dados originais e sua transposta)

% Subplot (Pedro II e Pedro II transposta: dados originais):

```
figure
set(gcf,'Position',[10,100,750,300])
set(gcf,'DefaultAxesFontSize',9)
subplot(1,2,1)
plot(PedroII)
title('CO (dados originais: Pedro II)')
ylabel('scores')
xlabel('dias')
```

```
subplot(1,2,2)
plot(PedroII')
title('CO (dados originais: Pedro II transposta)')
ylabel('scores')
xlabel('horas')
```

%%%

%Construção dos gráficos de “loadings” considerando-se três PCs:

```
figure
set(gcf,'DefaultAxesFontSize',13)
plot(p)
xlabel('horas')
ylabel('loadings')
title('Loadings para as três primeiras PCs')
gtext('PC1')
gtext('PC2')
gtext('PC3')
```

%%%

%Construção do gráfico de “loadings” (PC_1):

```
figure
set(gcf,'DefaultAxesFontSize',13)
plot(p(:,1))
xlabel('horas')
ylabel('loadings')
title('Perfil de Emissão da Fonte ( $PC_1=84.48\%$ )')
```


%%

%Construção dos gráficos de “scores” considerando-se a primeira PC:

```
dias=[' ',' ','sa','do',' ',' ',' ']; %(Rótulos somente para os Sábados e Domingos)
cont=1;
for i=1:52
    for j=1:7
        dias_ano(cont,:)=dias(j,:);
        cont=cont+1;
    end;
end;
figure
%set(gcf,'Position',[10,100,750,400])
set(gcf,'DefaultTextFontSize',9)
set(gcf,'DefaultAxesFontSize',13)
plot(t(:,1))
for i=1:364
    text(i,t(i,1),dias_ano(i,:))
end;
xlabel('dias')
ylabel('scores')
title('Variação anual de CO (Sáb e Dom - PC1=84.48 %)')
```

%%

%Construção dos gráficos para dados faltantes:

```
figure
A=P(:,1);
A1=p11;
A2=p21;
set(gcf,'DefaultAxesFontSize',13)
plot(A./mean(A))
hold
for i=1:10
    plot([2:2:24],A2(:,i)./mean(A2(:,i)),'r')
    plot([1:2:24],A1(:,i)./mean(A1(:,i)),'b')
end;
hold
xlabel('horas')
ylabel('loadings')
title('Influência Dados Faltantes')
gtext('CO (Pedro II - 1999)')
```

%%

APÊNDICE A8.5

%%

%Rotina para processar os dados sob o modelo PARAFAC (PedroII: uma variável):

X1=[Pedro_99(1:364,:)./mean(Pedro_99(:))];

Options=[1e-10,2,0,0,20,5000];

[Factors] = parafac(X1,[24,7,52],1,Options,[2 2 2]);

[A,B,C]=fac2let(Factors,[24,7,52],1);

%%

.....%Rotina para processar os dados sob o modelo PARAFAC (PedroII: quatro variáveis,
%duas componentes):

Xped=[co99pedn(:,1:364)./mean(co99ped(:)),...

no99pedn(:,1:364)./mean(no99ped(:)),...

no299pedn(:,1:364)./mean(no299ped(:)),...

o399pedn(:,1:364)./mean(o399ped(:)),...

];

Options=[1e-10,2,0,0,20,5000];

[Factor] = parafac(Xped,[24,7,52,4],2,Options,[2 2 2 0]);

[A,B,C,D]=fac2let(Factor,[24,7,52,4],2);

APÊNDICE A8.6

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
%Rotina para processar os dados sob o modelo de Tucker (PedroII: quatro variáveis):
```

```
%modo A: 3 componentes
```

```
%modo B: 2 componentes
```

```
%modo C: 2 componentes
```

```
%modo D: 3 componentes
```

```
%Matriz de dados:
```

```
Xped=[co99pedn(:,1:364)./mean(co99ped(:)),...
```

```
no99pedn(:,1:364)./mean(no99ped(:)),...
```

```
no299pedn(:,1:364)./mean(no299ped(:)),...
```

```
o399pedn(:,1:364)./mean(o399ped(:)),...
```

```
];
```

```
%Restrições ao núcleo (posições estratégicas no núcleo):
```

```
ConstG_1=[1 0 0 0  
0 0 0 0  
0 0 0 0];
```

```
ConstG_2=[0 0 0 0  
0 0 0 1  
0 0 0 0];
```

```
ConstG_3=[0 0 0 0  
0 0 0 0  
0 0 0 1];
```

```
ConstG=[ConstG_1 ConstG_2 ConstG_3];
```

```
Options=[1e-10,2,0,0,20,1000];
```

```
[Factors,G] = tucker(Xped,[24,7,52,4],[3,2,2,3],Options,[1 1 1 2],ConstG);
```

```
[A,B,C,D]=fac2let(Factors,[24,7,52,4],[3 2 2 3]);
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

APÊNDICE A8.7

São apresentados os arranjos para os núcleos do modelo de Tucker utilizados em cada análise. Nos casos em que não houver mudança da estrutura do núcleo para diferentes análises, é mostrado somente o número de componentes que compõe cada modo no núcleo.

1º caso – ciclo básico do CO (CO, NO, NO₂ e O₃ – Fig.6.60, p334):

%A seguinte notação é utilizada para o ciclo do CO:

%Xped – matriz contendo as entradas de dados para os poluentes do ciclo,

%co99ped e co99pedn – matrizes contendo as entradas de dados para o CO na Estação de
%Medição Pedro II,

.....%no99ped e no99pedn – matrizes contendo as entradas de dados para o NO na Estação de
%Medição Pedro II,

%no299ped e no299pedn – matrizes contendo as entradas de dados para o NO₂ na Estação
%de Medição Pedro II,

%o399ped e O399pedn – matrizes contendo as entradas de dados para o O₃ na Estação de
%Medição Pedro II,

%ConstG_core – restrições impostas ao núcleo (core = 1, 2, 3, ...)

%Options, Factors e factlet (ver instruções na referência correspondente, ou Apêndice A8)

Xped=[co99pedn(:,1:364)./mean(co99ped(:)),...

no99pedn(:,1:364)./mean(no99ped(:)),...

no299pedn(:,1:364)./mean(no299ped(:)),...

o399pedn(:,1:364)./mean(o399ped(:))];

ConstG_1=[1 0 0 0; 0 0 0 0; 0 0 0 0];

ConstG_2=[0 0 0 0; 0 0 0 1; 0 0 0 0];

ConstG_3=[0 0 0 0; 0 0 0 0; 0 0 0 1];

ConstG=[ConstG_1 ConstG_2 ConstG_3];

Options=[1e-10,2,0,0,20,1000];

[Factors,G] = tucker(Xped,[24,7,52,4],[3,2,2,3],Options,[1 1 1 2],ConstG);

[A,B,C,D]=fac2let(Factors,[24,7,52,4],[3 2 2 3]);

As posições estratégicas estão assinaladas com o nº 1 nas restrições (ConstG), e os números de componentes em cada modo [3,2,2,3] estão assinaladas em [Factors,G].

Nos casos seguintes a notação é análoga ao primeiro e não será indicada.

2º caso – ciclo básico do CH₄ (CH₄, NO, NO₂ e O₃ – Fig.6.61, p336):

```
Xped=[ch499pedn(:,1:364)./mean(ch499ped(:)),...
      no99pedn(:,1:364)./mean(no99ped(:)),...
      no299pedn(:,1:364)./mean(no299ped(:)),...
      o399pedn(:,1:364)./mean(o399ped(:))];
ConstG_1=[1 0 0 0 0 0; 0 0 0 0 0 0; 0 0 0 0 0 0];
ConstG_2=[0 0 0 0 0 0; 0 0 0 0 1 0; 0 0 0 0 0 0];
ConstG_3=[0 0 0 0 0 0; 0 0 0 0 0 0; 0 0 0 0 0 1];
ConstG=[ConstG_1 ConstG_2 ConstG_3];
[Factors,G] = tucker(Xped,[24,7,52,4],[3,3,2,3],Options,[1 1 1 2],ConstG);
[A,B,C,D]=fac2let(Factors,[24,7,52,4],[3 3 2 3]);
```

As posições estratégicas estão assinaladas com o nº 1 nas restrições (ConstG), e os números de componentes em cada modo [3,3,2,3] estão assinaladas em [Factors,G].

Deve ser observado que para esse ciclo, há duas fontes diferentes de poluentes primários. A matriz de componentes A contém três componentes. Uma delas descreve as variações sistemáticas diárias para um dos poluentes primários (CH₄), a outra descreve essas variações para os outros dois também poluentes primários (NO e o NO₂), e a terceira, as variações sistemáticas diárias para o poluente secundário (O₃). Assim, o modo A pode ser mantido com três componentes.

A existência de uma outra fonte de poluentes primários exige, por seu turno, a inclusão de uma terceira componente no modo que descreve as variações sistemáticas semanais (o modo B). Isto acarreta uma modificação na estrutura do núcleo como mostrado nas restrições ([ConstG_1, ConstG_2, ConstG_3]). Como se observa, não obstante a manutenção de um total de três posições no núcleo, estas estão “estrategicamente” distribuídas de modo diferente daquelas para o núcleo do ciclo do CO.

Para os outros dois modos, o modo C (variação anual) e o modo D (poluentes), fica mantida a composição anterior. Duas componentes no modo C, e três no modo D.

Doravante, para a formulação de ConstG, basta examinar a função *FactorsG*.

3º caso – ciclo básico do NMHC (NMHC, NO, NO₂ e O₃ – Fig.6.62, p338):

```
Xped=[nmhc99pedn(:,1:364)./mean(nmhc99ped(:)),...
```

```
no99pedn(:,1:364)./mean(no99ped(:)),...
```

```
no299pedn(:,1:364)./mean(no299ped(:)),...
```

```
o399pedn(:,1:364)./mean(o399ped(:))];
```

```
[Factors,G] = tucker(Xped,[24,7,52,4],[3,3,2,3],Options,[1 1 1 2],ConstG);
```

```
[A,B,C,D]=fac2let(Factors,[24,7,52,4],[3 3 2 3]);
```

Aqui foi usada a mesma estrutura anterior, com três componentes no modo B. Esse arranjo já fora apontado pelo modelo PARAFAC, o qual mostrou que os NMHCs mesmo quando emitidos por uma mesma fonte (por exemplo, a automotiva), podem pertencer a tipos diferentes dessa mesma fonte (ver discussão na ver discussão nas páginas 335, 340, 354, 370, e na Seção 6.5).

4º caso – substituição do CO ou CH₄ pelo SO₂, no ciclo (SO₂, NO, NO₂ e O₃ – Fig.6.63, p339):

```
Xped=[so299pedn(:,1:364)./mean(so299ped(:)),...
```

```
no99pedn(:,1:364)./mean(no99ped(:)),...
```

```
no299pedn(:,1:364)./mean(no299ped(:)),...
```

```
o399pedn(:,1:364)./mean(o399ped(:))];
```

```
Options=[1e-10,2,0,0,20,1800];
```

```
[Factors,G] = tucker(Xped,[24,7,52,4],[3,3,2,3],Options,[1 1 1 2],ConstG);
```

```
[A,B,C,D]=fac2let(Factors,[24,7,52,4],[3 3 2 3]);
```

Também aqui o modo B possui três componentes. Semelhantemente ao que acontece com os NMHCs, o dióxido de carbono também pode apresentar características de emissão que indiquem um tipo específico de uma mesma fonte (ver discussão nas páginas 335, 340, 354, 370, e na Seção 6.5).

5º caso – ciclos básicos do CO e do CH₄ (CO, CH₄, NO, NO₂ e O₃ – Fig.6.64, p341):

```
Xped=[co99pedn(:,1:364)./mean(co99ped(:)),...  
      no99pedn(:,1:364)./mean(no99ped(:)),...  
      no299pedn(:,1:364)./mean(no299ped(:)),...  
      o399pedn(:,1:364)./mean(o399ped(:)),...  
      ch499pedn(:,1:364)./mean(ch499ped(:))];
```

```
[Factors,G] = tucker(Xped,[24,7,52,5],[3,3,2,3],Options,[1 1 1 2],ConstG);  
[A,B,C,D]=fac2let(Factors,[24,7,52,5],[3 3 2 3]);
```

Também aqui o modo B possui três componentes. Esse resultado simples e aparentemente sem importância, parece ser fundamental para explicar o comportamento dos poluentes CO, CH₄ e O₃, em função dos níveis de emissão de NO_x em certos centros urbanos (ver explicação nas páginas 335, 340, 342, 354, 370, e na Seção 6.5).

6º caso – inclusão do NMHC no ciclo básico do CO (NMHC, CO, NO, NO₂ e O₃ – Fig.6.65, p343):

```
Xped=[nmhc99pedn(:,1:364)./mean(nmhc99ped(:)),...  
      no99pedn(:,1:364)./mean(no99ped(:)),...  
      no299pedn(:,1:364)./mean(no299ped(:)),...  
      o399pedn(:,1:364)./mean(o399ped(:)),...  
      co99pedn(:,1:364)./mean(co99ped(:))];
```

```
Options=[1e-10,2,0,0,20,1000];  
[Factors,G] = tucker(Xped,[24,7,52,5],[3,2,2,3],Options,[1 1 1 2],ConstG);  
[A,B,C,D]=fac2let(Factors,[24,7,52,5],[3 2 2 3]);
```

Nesse caso o modo B volta a possuir duas componentes. Quando, na análise PARAFAC, os dados são processados com três componentes, ocorre degenerescência (não há convergência ou no caso do modelo convergir, uma das componentes não possui significado físico). Isto também ocorre no modelo de Tucker.

7º caso – inclusão do NMHC no ciclo básico do CH₄ (NMHC, CH₄, NO, NO₂ e O₃ – Fig.6.66, p344):

```
Xped=[nmhc99pedn(:,1:364)./mean(nmhc99ped(:)),...
      no99pedn(:,1:364)./mean(no99ped(:)),...
      no299pedn(:,1:364)./mean(no299ped(:)),...
      o399pedn(:,1:364)./mean(o399ped(:)),...
      ch499pedn(:,1:364)./mean(ch499ped(:))];
```

```
[Factors,G] = tucker(Xped,[24,7,52,5],[3,3,2,3],Options,[1 1 1 2],ConstG);
[A,B,C,D]=fac2let(Factors,[24,7,52,5],[3 3 2 3]);
```

Novamente, o modo B volta a possuir três componentes. O perfil para o CH₄ difere tanto dos demais que é necessário a inclusão de uma componente para descrever suas variações sistemáticas.

8º caso – inclusão do SO₂ no ciclo básico do CO (SO₂, CO, NO, NO₂ e O₃ – Fig.6.67, p345):

```
Xped=[so299pedn(:,1:364)./mean(so299ped(:)),...
      no99pedn(:,1:364)./mean(no99ped(:)),...
      no299pedn(:,1:364)./mean(no299ped(:)),...
      o399pedn(:,1:364)./mean(o399ped(:)),...
      co99pedn(:,1:364)./mean(co99ped(:))];
```

```
[Factors,G] = tucker(Xped,[24,7,52,5],[3,2,2,3],Options,[1 1 1 2],ConstG);
[A,B,C,D]=fac2let(Factors,[24,7,52,5],[3 2 2 3]);
```

O modo B possui duas componentes. As variações sistemáticas dos poluentes primários são descritas por uma das componentes, enquanto a outra os faz para o poluente secundário, O₃.

9º caso – inclusão do SO₂ no ciclo básico do CH₄ (SO₂, CH₄, NO, NO₂ e O₃ – Fig.6.68, p346):

```
Xped=[so299pedn(:,1:364)./mean(so299ped(:)),...  
      no99pedn(:,1:364)./mean(no99ped(:)),...  
      no299pedn(:,1:364)./mean(no299ped(:)),...  
      o399pedn(:,1:364)./mean(o399ped(:)),...  
      ch499pedn(:,1:364)./mean(ch499ped(:))];  
  
[Factors,G] = tucker(Xped,[24,7,52,5],[3,3,2,3],Options,[1 1 1 2],ConstG);  
[A,B,C,D]=fac2let(Factors,[24,7,52,5],[3 3 2 3]);
```

O modo B volta a possuir três componentes. A exigência é imposta pela presença do CH₄.

10º caso – inclusão do SO₂ no ciclo básico do NMHC (SO₂, NMHC, NO, NO₂ e O₃ – Fig.6.69, p347):

```
Xped=[so299pedn(:,1:364)./mean(so299ped(:)),...  
      no99pedn(:,1:364)./mean(no99ped(:)),...  
      no299pedn(:,1:364)./mean(no299ped(:)),...  
      o399pedn(:,1:364)./mean(o399ped(:)),...  
      nmhc99pedn(:,1:364)./mean(nmhc99ped(:))];  
  
[Factors,G] = tucker(Xped,[24,7,52,5],[3,3,2,3],Options,[1 1 1 2],ConstG);  
[A,B,C,D]=fac2let(Factors,[24,7,52,5],[3 3 2 3]);
```

Mantém-se o modo B com três componentes devido à existência de tipos diferentes de emissores da mesma fonte de emissão (a automotiva: veículos leves e veículos pesados).

11º caso – inclusão do SO₂ no ciclo do CO e CH₄ (SO₂, CO, CH₄, NO, NO₂ e O₃ – Fig.6.70, p348):

```
Xped=[so299pedn(:,1:364)./mean(so299ped(:)),...
      no99pedn(:,1:364)./mean(no99ped(:)),...
      no299pedn(:,1:364)./mean(no299ped(:)),...
      o399pedn(:,1:364)./mean(o399ped(:)),...
      co99pedn(:,1:364)./mean(co99ped(:)),...
      ch499pedn(:,1:364)./mean(ch499ped(:))];
```

```
[Factors,G] = tucker(Xped,[24,7,52,6],[3,3,2,3],Options,[1 1 1 2],ConstG);
```

```
[A,B,C,D]=fac2let(Factors,[24,7,52,6],[3 3 2 3]);
```

Mantém-se o modo B com três componentes.

12º caso – inclusão do SO₂ no ciclo do CO e NMHC (SO₂, CO, NMHC, NO, NO₂ e O₃ – Fig.6.71, p349):

```
Xped=[so299pedn(:,1:364)./mean(so299ped(:)),...
      no99pedn(:,1:364)./mean(no99ped(:)),...
      no299pedn(:,1:364)./mean(no299ped(:)),...
      o399pedn(:,1:364)./mean(o399ped(:)),...
      co99pedn(:,1:364)./mean(co99ped(:)),...
      nmhc99pedn(:,1:364)./mean(nmhc99ped(:))];
```

```
[Factors,G] = tucker(Xped,[24,7,52,6],[3,2,2,3],Options,[1 1 1 2],ConstG);
```

```
[A,B,C,D]=fac2let(Factors,[24,7,52,6],[3 2 2 3]);
```

Mantém-se o modo B com três componentes (presença de SO₂ e NMHC)

13º caso – ciclos do CO, CH₄ e NMHC (CO, CH₄, NMHC, NO, NO₂ e O₃ – Fig.6.72, p350):

```
Xped=[co99pedn(:,1:364)./mean(co99ped(:)),...
      no99pedn(:,1:364)./mean(no99ped(:)),...
      no299pedn(:,1:364)./mean(no299ped(:)),...
      o399pedn(:,1:364)./mean(o399ped(:)),...
      nmhc99pedn(:,1:364)./mean(nmhc99ped(:)),...
      ch499pedn(:,1:364)./mean(ch499ped(:))];
```

```
[Factors,G] = tucker(Xped,[24,7,52,6],[3,2,2,3],Options,[1 1 1 2],ConstG);
[A,B,C,D]=fac2let(Factors,[24,7,52,6],[3 2 2 3]);
```

Curiosamente, apesar de aparecer o CH₄, quando se processa incluindo três componentes no modo B, uma delas perde o significado físico.

14º caso – inclusão do SO₂ nos ciclos do CH₄ e NMHC (SO₂, CH₄, NMHC, NO, NO₂ e O₃ – Fig.6.73, p351):

```
Xped=[so299pedn(:,1:364)./mean(so299ped(:)),...
      no99pedn(:,1:364)./mean(no99ped(:)),...
      no299pedn(:,1:364)./mean(no299ped(:)),...
      o399pedn(:,1:364)./mean(o399ped(:)),...
      ch499pedn(:,1:364)./mean(ch499ped(:)),...
      nmhc99pedn(:,1:364)./mean(nmhc99ped(:))];
```

```
[Factors,G] = tucker(Xped,[24,7,52,6],[3,3,2,3],Options,[1 1 1 2],ConstG);
[A,B,C,D]=fac2let(Factors,[24,7,52,6],[3 3 2 3]);
```

Nesse caso também ocorre um fato curioso. Quando processado com esse arranjo para o núcleo, o modelo não consegue modelar as variações sistemáticas do CH₄. Pode-se incluir mais componentes no modo B para tentar capturar o perfil do CH₄, porém quando isto foi realizado, a componente adicional introduzida passou a descrever as variações para

um dos poluentes primários da fonte automotiva. Isto é, houve o desmembramento do perfil de emissão da fonte, separando um dos poluentes e acomodando-o na nova componente introduzida. A adição de um número maior de componentes no modo B acaba exigindo a introdução de componentes adicionais no modo A, sendo que tal providência não implicou no surgimento de um perfil inequívoco para o CH₄ no modo B.

15º caso – inclusão do SO₂ nos ciclos do CO, do CH₄ e dos NMHCs (SO₂, CO, CH₄, NMHC, NO, NO₂ e O₃ – Fig.6.74, p352):

```
Xped=[so299pedn(:,1:364)./mean(so299ped(:)),...
```

```
no99pedn(:,1:364)./mean(no99ped(:)),...
```

```
no299pedn(:,1:364)./mean(no299ped(:)),...
```

```
o399pedn(:,1:364)./mean(o399ped(:)),...
```

```
co99pedn(:,1:364)./mean(co99ped(:)),...
```

```
nmhc99pedn(:,1:364)./mean(nmhc99ped(:)),...
```

```
ch499pedn(:,1:364)./mean(ch499ped(:))];
```

```
[Factors,G] = tucker(Xped,[24,7,52,7],[3,3,2,3],Options,[1 1 1 2],ConstG);
```

```
[A,B,C,D]=fac2let(Factors,[24,7,52,7],[3 3 2 3]);
```

Nesse caso, estão todos os poluentes do caso anterior (no qual não foi possível capturar o perfil para o CH₄), mais o CO. Entretanto, não houve qualquer dificuldade em se extrair o perfil semanal para o CH₄ mantendo-se o modo B com três componentes (e três no modo A, duas no modo C, e três no modo D).

Esse resultado, não percebido inicialmente, mostra a existência de uma relação entre os comportamentos do CO e do CH₄ em presença de certos níveis de concentrações de NO_x e de NMHCs. O Capítulo VII aborda esse significativo resultado, o qual implica em conclusão importante nessa tese.

Não obstante as informações contidas na exposição desses 15 casos analisados por meio do modelo de Tucker permitam a reprodução daqueles resultados, a determinação das posições estratégicas no núcleo, constituído dos Algarismos zero (“0”) e um (“1”), se faz por meio do produto de Kronecker como mostrado no Apêndice A5.

Para o ciclo básico do CO tem-se: **A**(24 × 3), **B**(7 × 2), **C**(52 × 2) e **D**(4 × 3), o que leva a um núcleo **G**(3 × 12). Quando não se impõe restrições ao núcleo obtém-se:

$$\mathbf{G} = \begin{pmatrix} 113,7218 & 33,0759 & \mathbf{92,7562} & -42,4204 & | & 70,1174 & -69,7735 & \mathbf{154,1540} & -43,8815 & | \\ -5,2569 & 45,6265 & -28,9120 & -3,1346 & | & -31,1978 & -31,7857 & -74,9797 & 40,7255 & | \\ 50,1921 & -19,9311 & -35,2645 & 3,6243 & | & -58,7264 & 29,3515 & 14,1877 & -9,7155 & | \\ \\ 8,0257 & -1,3478 & \mathbf{-49,9987} & 26,9252 & \\ -3,2093 & -34,5894 & 32,6247 & -0,5620 & \\ 15,8900 & 6,0988 & 13,3588 & -3,7550 & \end{pmatrix}$$

Para a matriz de componentes dos poluentes tem-se:

$$\mathbf{D} = \begin{pmatrix} 0,4859 & -0,0895 & 0,5621 \\ 0,5443 & -0,5353 & -0,6422 \\ 0,5037 & -0,1008 & 0,4265 \\ 0,4625 & 0,8338 & -0,2996 \end{pmatrix} \begin{matrix} \text{CO} \\ \text{NO} \\ \text{NO}_2 \\ \text{O}_3 \end{matrix}$$

No núcleo, **G**, tem-se como valores de maior importância aqueles destacados em negrito e todos na primeira linha da matriz. Isto significa que a primeira coluna da matriz **A** (matriz que dá informações sobre as fontes de poluentes primários, e sobre o poluente secundário) contém informações as mais importantes. Todavia aparecem os valores assinalados em vermelho, os quais são de importância mediana, e apenas uns poucos valores em **G** poderiam ser considerados desprezíveis. Por outro lado, ao se analisar a matriz **D** observa-se que na sua primeira coluna todos os poluentes possuem aproximadamente o mesmo peso, e todos com o mesmo sinal. Isto significa uma alta correlação entre os poluentes, resultando em um perfil irreal, especialmente porque o O_3 é um poluente secundário devendo ser, necessariamente, acomodado em outra componente.

Quando são impostas restrições ao núcleo obtém-se:

$$\mathbf{G} = \left[\begin{array}{cccc|cccc|cccc} 181,1626 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 105,5456 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 56,3378 \end{array} \right]$$

E para a matriz de componentes dos poluentes:

$$\mathbf{D} = \left[\begin{array}{ccc} 0,3996 & 0,1108 & 0,4343 \\ 0,7947 & -0,0092 & -0,6092 \\ 0,4509 & 0,1661 & 0,2228 \\ -0,0735 & 0,9798 & 0,6250 \end{array} \right] \begin{array}{l} \text{CO} \\ \text{NO} \\ \text{NO}_2 \\ \text{O}_3 \end{array}$$

As restrições impostas ao núcleo permitem atribuir uma dada importância para cada coluna da matriz $\mathbf{A}$, portanto, destacando a importância de cada fonte de poluentes primários e do poluente secundário. Uma análise da matriz $\mathbf{D}$ mostra, em sua primeira coluna, que os poluentes primários, CO, NO e NO₂, possuem pesos elevados, quando comparados ao O₃, o qual apresenta um peso muito baixo, além de negativo. Isto significa que a primeira componente modela apenas os poluentes primários.

As outras duas colunas da matriz $\mathbf{D}$ mostram o O₃ com um peso maior que os outros poluentes. Isto ocorre porque são necessários apenas duas componentes para modelar os dados. Porém, foi imposto inicialmente que a matriz $\mathbf{A}$ deveria levar em conta pelo menos duas fontes em potencial (veículos e indústrias, por exemplo) e uma componente para o poluente secundário. Assim, capturada a informação referente aos poluentes primários pela primeira componente, o restante fica distribuído nas outras duas componentes.

É importante notar que apenas a primeira posição no núcleo, $\mathbf{G}(1,1,1,1)$ é comum aos dois casos (113,7218 sem e 181,1626 com restrições). As outras posições no núcleo diferem em cada caso levando, portanto a resultados diferentes. Além disso, para o caso com restrições o valor é mais elevado. O mesmo vale para as outras duas posições no núcleo com restrições quando comparado ao caso sem restrições.

Para o ciclo básico do CH₄ tem-se: **A**(24 × 3), **B**(7 × 3), **C**(52 × 2) e **D**(4 × 3), o que leva a um núcleo **G**(3 × 18). Quando não se impõe restrições no núcleo obtém-se:

$$\mathbf{G} = \begin{pmatrix} \mathbf{126,8880} & -3,8266 & 18,4244 & 50,4722 & 16,1113 & -6,6098 \\ 75,6163 & 1,6940 & -25,3336 & -70,6576 & 0,0691 & 14,9621 \\ 16,8083 & -17,1177 & -0,8295 & 0,8732 & 1,9153 & 0,8334 \\ -22,3028 & 21,0314 & -0,6346 & \mathbf{114,3141} & 21,3280 & 3,7750 \\ -60,7281 & -0,2164 & 22,2640 & 0,3842 & -21,8982 & -17,8856 \\ -38,4970 & 12,0896 & -5,8441 & -21,0744 & 3,2469 & 4,3094 \\ -4,0584 & 14,1321 & 1,6156 & -13,9019 & -14,9925 & -2,5801 \\ 13,8356 & 1,7245 & 8,0314 & 41,1007 & 3,2928 & -4,2687 \\ -\mathbf{43,2771} & 11,9782 & 2,6195 & 2,2232 & -2,1641 & -1,3682 \end{pmatrix}$$

E para a matriz de componentes dos poluentes:

$$\mathbf{D} = \begin{pmatrix} 0,4641 & -0,0358 & 0,8204 \\ 0,5481 & -0,5625 & -0,4887 \\ 0,5081 & -0,1246 & 0,0503 \\ 0,4755 & 0,8166 & -0,2924 \end{pmatrix} \begin{matrix} \text{CH}_4 \\ \text{NO} \\ \text{NO}_2 \\ \text{O}_3 \end{matrix}$$

Os resultados aqui são análogos àqueles obtidos para o ciclo do CO. O núcleo, **G**, possui valores de maior importância (em negrito), sendo dois deles na primeira linha da matriz. Do mesmo modo, aparecem os valores assinalados em vermelho (de importância mediana), e poucos valores em **G** poderiam ser considerados desprezíveis. Por outro lado, ao se analisar a matriz **D** observa-se também que na sua primeira coluna todos os poluentes possuem aproximadamente o mesmo peso, e todos com o mesmo sinal. Logo o resultado é semelhante ao do ciclo para o CO.

Quando, no ciclo do CH₄, são impostas restrições no núcleo obtém-se:

$$\mathbf{G} = \begin{pmatrix} 165,5424 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 55,4631 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 120,3672 \end{pmatrix}$$

E para a matriz de componentes dos poluentes:

$$\mathbf{D} = \begin{pmatrix} 0,2292 & 0,8918 & 0,1806 \\ 0,8351 & -0,3234 & -0,0905 \\ 0,4946 & 0,0952 & 0,1984 \\ -0,0734 & 0,3017 & 0,9591 \end{pmatrix} \begin{matrix} \text{CH}_4 \\ \text{NO} \\ \text{NO}_2 \\ \text{O}_3 \end{matrix}$$

As restrições impostas ao núcleo permitem atribuir uma dada importância para cada coluna da matriz **A**, portanto, destacando a importância de cada fonte de poluentes primários e do poluente secundário. Uma análise da matriz **D** mostra, em sua primeira coluna, que os poluentes primários, NO e NO₂, possuem pesos maiores, que o do outro também poluente primário, porém de fonte diferente, CH₄, enquanto o O₃ apresenta um peso muito baixo, além de negativo.

A segunda coluna da matriz **D** possui o CH₄ com valor bem maior do que os outros poluentes, enquanto sua terceira coluna apresenta o O₃ com um peso maior em relação aos outros poluentes.

Isto significa que a primeira componente modela os poluentes primários, NO e NO₂, de uma das fontes, a segunda componente modela o também poluente primário, porém de outra fonte, CH₄, e a terceira componente modela o poluente secundário, O₃. Então, no ciclo do metano, são necessários pelo menos três componentes para representar os dados.

Novamente apenas a primeira posição no núcleo é comum aos dois casos (126,8880 sem e 165,5424 com restrições). As outras posições no núcleo diferem em cada caso levando, portanto a resultados diferentes. Do mesmo modo que no caso do ciclo para o CO, aqui também, para o caso com restrições o valor é mais elevado.

Cabe observar ainda que a ordem em que aparecem as entradas para os poluentes na matriz de componentes **D**, é a mesma da matriz de dados **X** (no caso acima: CH₄, NO, NO₂ e O₃). Se a ordem for modificada, por exemplo, para O₃, CH₄, NO e NO₂, os valores no núcleo **G**, passam de 165,5424, 55,4631 e 120,3672, respectivamente, para 165,5310, -120,1708 e -55,6587. As diferenças entre os valores (165,5310 e 165,5424; 120,3672 e -120,1708; -55,6587 e 55,4631) são desprezíveis, porém, além de ocorrer mudança no sinal dos dois últimos valores, também ocorre inversão de posição entre o segundo e terceiro valores em consequência da modificação da posição dos vetores no espaço de dados. Assim, a segunda componente passa a ser a terceira e vice-versa. É claro que isto não modifica os perfis correspondentes aos respectivos poluentes.

Desse modo, é preferível consumir um pouco mais de tempo orientando corretamente os tipos de “cruzamento” entre os vetores de interesse, do que correr o risco de não conseguir encontrar a informação, de fato, significativa.

Cabe lembrar ainda que, as diferenças obtidas para os casos sem e com restrições ocorrem porque o produto de Kronecker realiza os “cruzamentos” possíveis entre as matrizes de componentes **B**, **C** e **D**, tal que na ausência de restrições ao núcleo, o Tucker tende a modelar as variações sistemáticas de forma redundante, distribuindo os pesos nas outras posições do núcleo, o que dificulta a escolha das componentes mais importantes e a sua associação com o comportamento dos poluentes tanto na periodicidade diária quanto na semanal, além da variação anual.

APÊNDICE B1

Tabela B1: Configuração da Rede Automática

Estação	Localização das Estações	P	A	R	A	M	E	T	R	O	S					
Nº		PI	SO ₂	NO	NO ₂	NO _x	CO	CH ₄	HCNM	O ₃	UR	TEMP	VV	DV	P	RAD
01	Parque D. Pedro II	x	x	x	x	x	x	x	x	x	x	x	x	x		
02	Santana	x								x			x	x		
03	Moóca	x								x			x	x		
04	Cambuci	x														
05	Ibirapuera	x	x	x	x	x	x			x	x	x	x	x	x	x
06	Nossa Senhora do Ó	x														
07	São Caetano do Sul	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
08	Congonhas	x	x	x	x	x	x			x						
09	Lapa	x		x	x	x	x			x			x	x		
10	Cerqueira César	x	x	x	x	x	x									
11	Penha	x														
12	Centro	x	x	x	x	x	x									
13	Guarulhos	x											x	x		
14	S ^{to} André-Centro	x					x						x	x		
15	Diadema	x								x						
16	Santo Amaro	x					x						x	x		
17	Osasco	x	x	x	x	x	x			x			x	x		
18	S ^{to} André-Capuava	x											x	x		
19	São Bernardo do C ^{po}	x											x	x		
20	Taboão da Serra	x														
21	São Miguel Paulista	x								x	x	x	x	x		
22	Mauá	x		x	x	x				x						
24	Cubatão-Centro	x	x	x	x	x		x	x	x	x	x			x	
25	Cubatão-Vila Parisi	x	x										x	x		
27	Pinheiros	x	x	x	x	x				x						
49	Lab. Volante I	x	x	x	x	x	x			x	x	x	x	x		
50	Lab. Volante II	x	x	x	x	x	x	x	x	x	x	x	x	x		

Fonte: Relatório CETESB, 1999

PI – Partículas Inaláveis; HCNM – Hidrocarbonetos Não Metano; VV – Velocidade do Vento; DV – Direção do Vento; UR – Umidade Relativa; P – Pressão; TEMP – Temperatura; RAD – Radiação Total e Ultravioleta.

A Tabela B1 mostra a configuração da rede automática da CETESB. Foram objeto de investigação apenas as Estações de Medição assinaladas em cinza escuro (01, 05, 07, 09, 10, 12 e 17), e as variáveis CO, NO, NO₂, O₃, SO₂, CH₄, HCNM, TEMP e UR.

A Figura B₁, abaixo, mostra a localização das Estações de Medição da CETESB. Os locais numerados de 01 a 25 fazem parte da rede automática. A Estação de Medição Pinheiros não constava do relatório de qualidade do ar de 1998, porém já aparece no de 1999 sob o número 27.

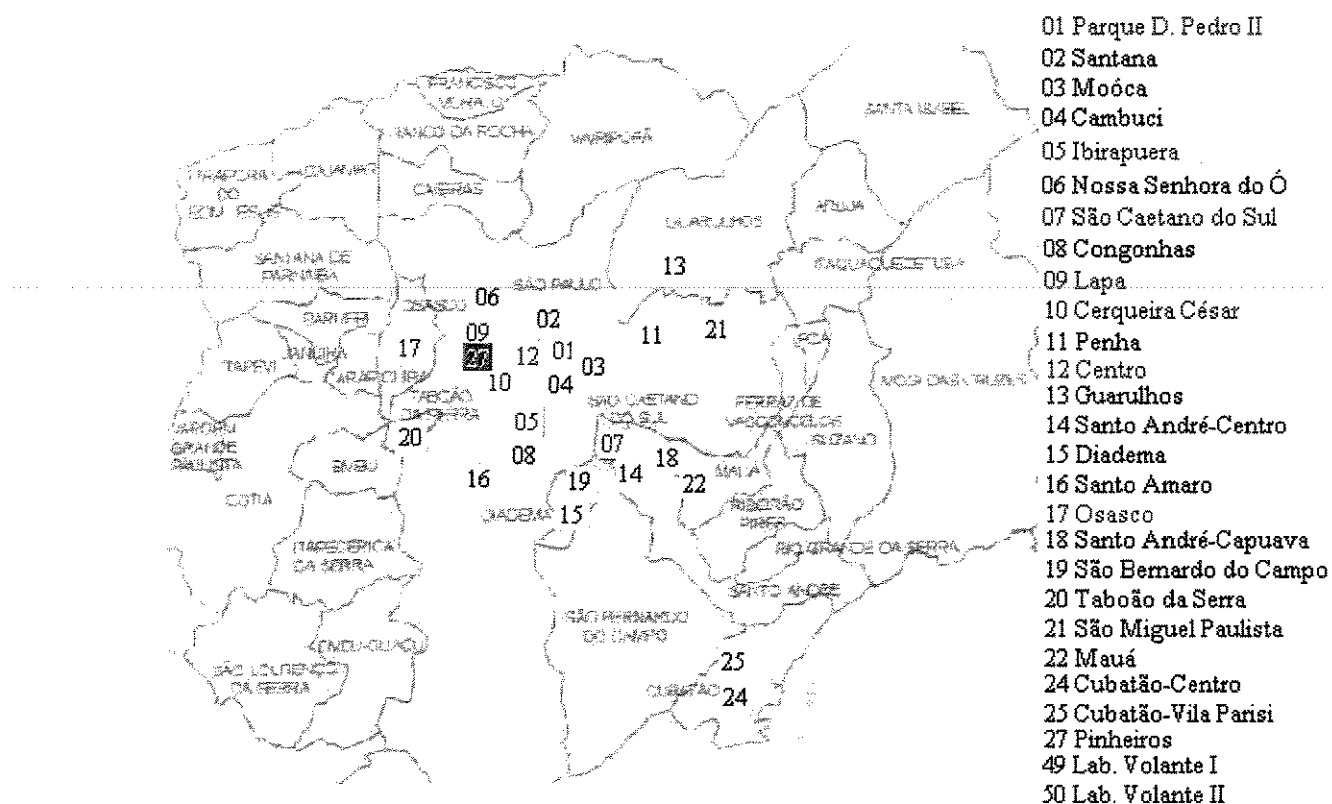


Fig. B₁: Região coberta pela instrumentação da CETESB. A Estação n°01 (em vermelho), Parque D. Pedro II, e aquelas na cor lilás (05, 07, 09, 10, 12 e 17) foram alvo de investigação sobre as variáveis CO, NO, NO₂, O₃, SO₂, CH₄, NMHCs, Temperatura e Umidade Relativa, durante os anos de 1996 a 1999. (Adaptado do Relatório de Qualidade do Ar da CETESB – 1999).

APÊNDICE B2

Abaixo segue um breve resumo sobre alguns instrumentos de medição utilizados pela CETESB.

1) Monóxido de Carbono – o CO é quase sempre medido por espectroscopia não dispersiva. O limite de detecção (LOD) para instrumentos convencionais, inclusive os montados em aeronaves, é de cerca de 50 ppb, com tempo de resolução de 10 s, e precisão e exatidão desde aproximadamente 10% à valores maiores do que 10 vezes o LOD (Keith, 1988).

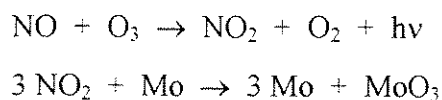
O CO vem sendo monitorado pela CETESB pela técnica de correlação com filtro de gás (GFC) e infra vermelho (IR) não dispersivo. A espectroscopia GFC é baseada na comparação da detalhada estrutura do espectro de absorção no IR do gás sendo medido, em relação àqueles outros também presentes na amostra. A técnica permite sensibilidade abaixo de 1 ppm, com LOD de 0,1 ppm e vazão de 0,5 a 2,0 L/min.

2) Óxidos de Nitrogênio (NO e NO₂) – as medições são mais freqüentemente realizadas com técnicas baseadas na reação quimiluminescente de ozônio com o óxido nítrico (Fontijn *et alli*, 1970, 575-579). Existem instrumentos para análise em tempo real de NO e NO_x, porém com LOD de apenas cerca de 2 ppb. Estes não são adequados para aplicações em regiões não urbanas, onde as concentrações de NO_x são freqüentemente menores do que 2,0 ppb.

A determinação de NO₂ por meio de quimiluminescência com ozônio é problemática porque o catalisador usado, o Mo, converte vários óxidos e oxiácidos a NO junto com o NO₂. Para contornar o problema, tem sido usado o sulfato ferroso para a conversão do NO₂ a NO (nesse caso parece ocorrer adsorção associada com umidade na presença de PANs (Tanner *et alli*, 1983). A outra solução é a conversão fotolítica de NO₂ a NO (Kley *et alli*, 1981).

Uma outra determinação em tempo real de NO₂ inteiramente diferente é baseada na observação da quimiluminescência de NO₂ na presença de uma solução aquosa de luminol (Wendell *et alli*, 937-940, 1983).

A CETESB utiliza um analisador quimiluminescente para NO/NO₂/NO_x capaz de medir óxidos de nitrogênio desde subppb até 20 ppm. As reações são:



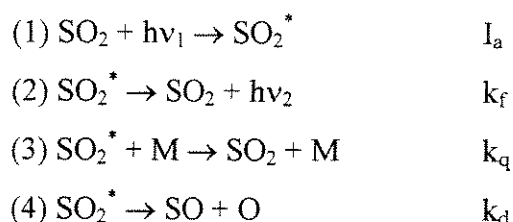
A conversão de NO_2 para NO é realizada sob o Mo como catalisador, podendo induzir a erros nos resultados.

3) Dióxido de Enxofre – a maioria das medições em tempo real pode ser realizada por um detetor fotométrico de chama (FDP) modificado, ou por fluorescência de pulso (Keith, 1988).

A CETESB usa em suas medições de SO_2 o detetor por fluorescência de pulso. A luz ultravioleta pulsante é filtrada em um passa-banda e focalizada em uma câmara de fluorescência, na qual as moléculas de SO_2 são excitadas a estados superiores de energia. Quando decaem, emitem radiação característica. Um filtro permite que apenas essa radiação atinja o tubo fotomultiplicador.

O SO_2 absorve luz em três regiões primárias: a) 390-340 nm – região de fraca absorção e grande perda de radiação fluorescente; b) 320-250 nm – onde as moléculas de SO_2 excitadas são fortemente seqüestradas por oxigênio e nitrogênio do ar; 3) 250-220 nm – onde o seqüestro é mínimo para a maioria das espécies no ar além de O_2 e N_2 .

A fluorescência é, portanto, localizada nessa região. A fonte de luz emite radiação à frequências ν_1 , a qual é absorvida por moléculas de SO_2 produzindo espécies excitadas, SO_2^* . As reações abaixo ilustram o processo:



Em (1), $h\nu_1$ é a energia absorvida pelas moléculas de SO_2 . A intensidade de luz absorvida por elas a partir de um feixe de luz incidente é dada por:

$$I_a = I_0 [1 - \text{EXP} (a b c)]$$

onde a é o coeficiente de absorção de SO_2 .

As equações (2), (3) e (4) representam os decaimentos, respectivamente, por fluorescência, por seqüestro e por dissociação.

As equações acima podem ser combinadas em uma única expressão representando a intensidade de fluorescência, F , em um detector. Assim,

$$F = G k_f I_0 [1 - \text{EXP} (a b c)] / (k_f + k_d + k_q [M])$$

onde G representa uma característica geométrica da câmara de fluorescência. Quando a concentração de SO_2 é relativamente baixa e o comprimento do caminho da luz excitante é pequeno, a expressão acima pode ser aproximada para:

$$F = (G k_f I_0 a b c) / (k_f + k_d + k_q [M])$$

Os valores de k_f , k_d e k_q permanecem relativamente constantes sobre uma grande faixa de temperatura e atmosfera de fundo. I_0 pode, por projeto da célula, ser tomado como constante. G e b dependem somente do projeto mecânico da câmara, sendo também constantes. Assim, a equação fica:

$$F = K c, \text{ ou } F = K [\text{SO}_2]$$

A última equação mostra que a radiação fluorescente atingindo o detector é diretamente proporcional à concentração de SO_2 o que, em última análise, é a base dos instrumentos desse tipo.

4) Ozônio – as medições de O_3 podem ser realizadas em tempo real por meio da quimiluminescência com etileno (Nederbragt *et alli*, 1965). Precisão e exatidão de $\pm 10\%$ ou melhor, na faixa de 20 a 200 ppb são usualmente adequadas para a maioria das aplicações, e um tempo de resolução de 1-3 s é excepcionalmente bom (Keith, 1988).

A CETESB usa, na sua rede telemétrica, para a medição de O_3 , um analisador fotométrico UV, o qual determina concentrações do gás medindo a atenuação de luz devida ao O_3 na célula de absorção, a um comprimento de onda de 254 nm. Variações de ± 1 ppb na concentração de ozônio causam variações na relação entre os feixes da amostra e de referência, I/I_0 , de 0,002%. Assim, a fonte deve ter estabilidade melhor do que 0,002% no tempo (o que conduz à medidas de I e I_0 tipicamente em torno de 10 s), para que o fotômetro tenha uma sensibilidade de ± 1 ppb.

APÊNDICE C1

Os resultados das análises para os ciclos do CO e do CH₄ em São Caetano são mostrados nas figuras C1.1 e C1.2 abaixo.

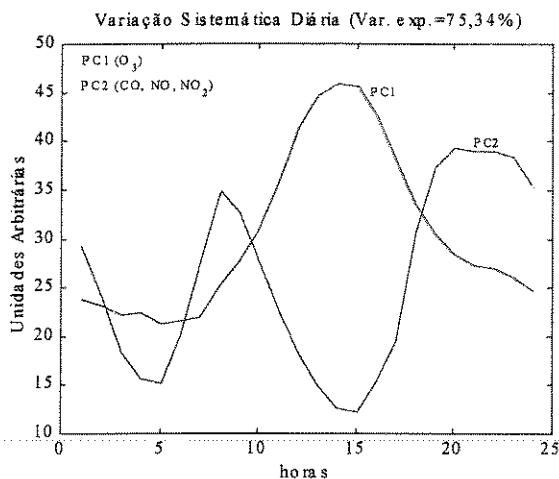
Para o ciclo do CO, observa-se que, diferentemente do que ocorre no P. D. Pedro II, quando os níveis de concentração dos poluentes primários caem nos finais de semana, o ozônio também apresenta ligeira queda. Isto poderia estar relacionado com uma menor dispersão de CO no local. Uma queda nas emissões automotivas significa queda nos níveis de concentração de NO_x, o que causaria queda nos níveis do O₃. Entretanto, se o CO também tem origem principalmente em fonte automotiva, seus níveis de concentração deveriam cair também nos finais de semana, o que deixaria o CH₄ em condições de reagir com os radicais HO^{*}, o que, em contrapartida, não deixaria que os níveis de concentração de O₃ caíssem. Outro aspecto a considerar é o pico noturno para o O₃ (~3:00–4:00 h), o qual é bem mais discreto do que no Pedro II.

Para o ciclo do CH₄, isto é, quando os poluentes CO e CH₄ são analisados tendo-se em conta sua mútua influência, nota-se que a queda nos níveis do ozônio é acompanhada de queda para o metano, queda nas emissões automotivas e, portanto, queda de CO. Quando as emissões automotivas voltam a crescer, na 2^a feira, os níveis de concentração de O₃ voltam a subir.

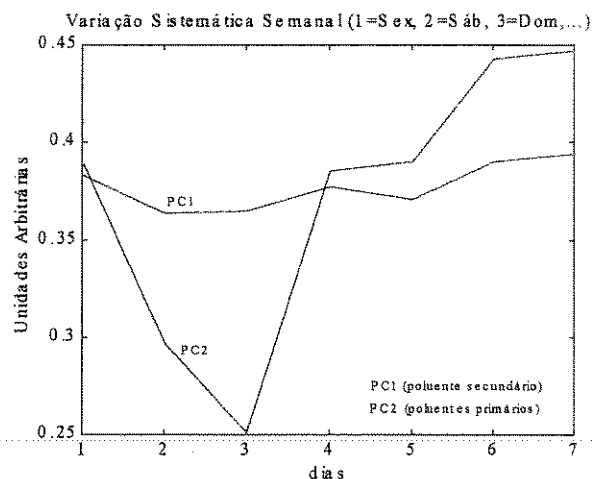
Esses resultados não são contraditórios pois uma queda brusca nos níveis de NO_x deveria causar queda correspondente nos níveis de O₃, e como tal não acontece, os altos níveis de concentração de O₃ devem ser atribuídos a outras reações. Nesse caso, a queda nos níveis do CH₄ o coloca como uma provável espécie responsável pela manutenção dos altos níveis de concentração de O₃.

Esses resultados mostram que em locais nos quais as emissões automotivas são a principal fonte de poluição atmosférica, a persistência do O₃ na atmosfera não pode ser investigada somente em função das emissões dos NO_x, mas sim como um efeito sinérgico, isto é, como um efeito causado pelas influências simultâneas das emissões de CH₄ e de CO sob determinados níveis de emissão de NO_x.

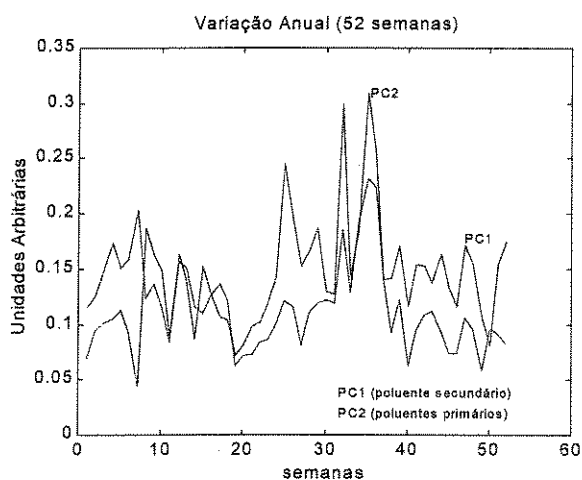
Análise PARAFAC para São Caetano (1999): ciclo do CO



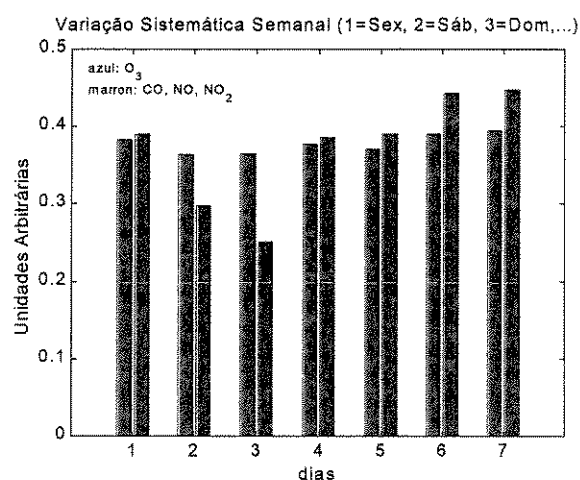
C1a



C1b



C1c



C1d

Fig.C1.1: a)Variações sistemáticas com periodicidade diária. b) e d)Variações sistemáticas com periodicidade semanal. c)Variação anual.

Em C1.1a, PC₁ representa o perfil para o O₃ e PC₂ o perfil de emissões automotivas. Em C1.1b e C1.1d, PC₁ e PC₂ representam, respectivamente, os perfis semanais para o O₃ e para as emissões automotivas. Em C1.1c, PC₁ representa a variação anual para o ozônio e PC₂ a variação anual para a fonte automotiva.

Análise PARAFAC para São Caetano (1999): ciclo do CH₄ (CO incluído)

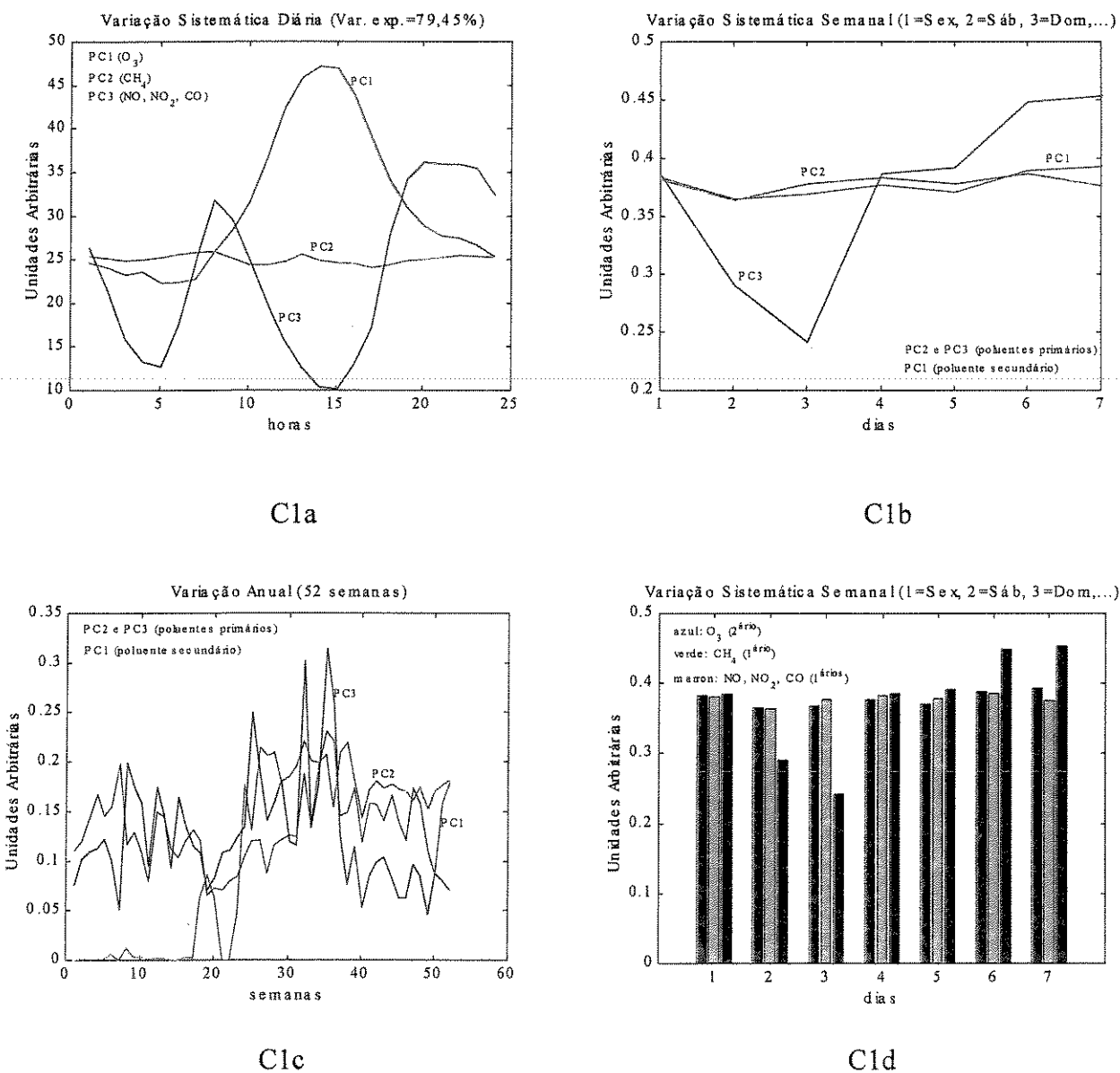


Fig.C1.2: a)Variações sistemáticas com periodicidade diária. b) e d)Variações sistemáticas com periodicidade semanal. c)Variação anual.

Em C1.2a, PC₁ representa o perfil para o O₃, PC₂ o do CH₄, e PC₃ o perfil de emissões automotivas. Em C1.2b e C1.2d, PC₁ representa o perfil semanal para o O₃, e PC₂ e PC₃ descrevem, respectivamente, as variações sistemáticas semanais para o CH₄ e para as emissões automotivas. Em C1.2c, PC₁ representa a variação anual para o ozônio, PC₂ para o CH₄, e PC₃ a variação anual para a fonte automotiva.

APÊNDICE C2

Os resultados obtidos por meio da aplicação do modelo PARAFAC à partícula inalável, PI (Pedro II – 1999), são apresentados na Figura C2 abaixo.

Nota-se em C2a existência de dois picos. Um deles aproximadamente 8:00–9:00 h e outro em cerca de 20:00–21:00 h. Esta é uma característica de emissões automotivas. Para o perfil semanal (Figura C2b e C2d), verifica-se queda nos níveis de PI nos finais de semana, uma característica também de emissões automotivas. Em C2c nota-se maior acumulação no período de inverno. Um fato interessante é a constatação de que os níveis mais elevados encontram-se na 3ª feira, fato registrado também para as emissões de SO₂ nessa mesma localidade como mostrado nas Figuras 6.10.1b e d (p273) e Figuras 6.10.2b e d (p274). Nesse últimos caso (duas fontes) deve ser considerada a curva em azul para emissões de veículos.

Tais resultados mostram que essas emissões devem ser originárias de fonte automotiva. De fato, relatório de 1998 da CETESB indica que 40% do total de emissões têm origem no escapamento de veículos (8,0% veículos leves e 32% diesel), sendo que as indústrias contribuem com 10% do total. Assim, os níveis registrados para PI devem ser atribuídos também a fontes automotivas.

Análise PARAFAC para São Caetano (1999): Partícula Inalável (PI)

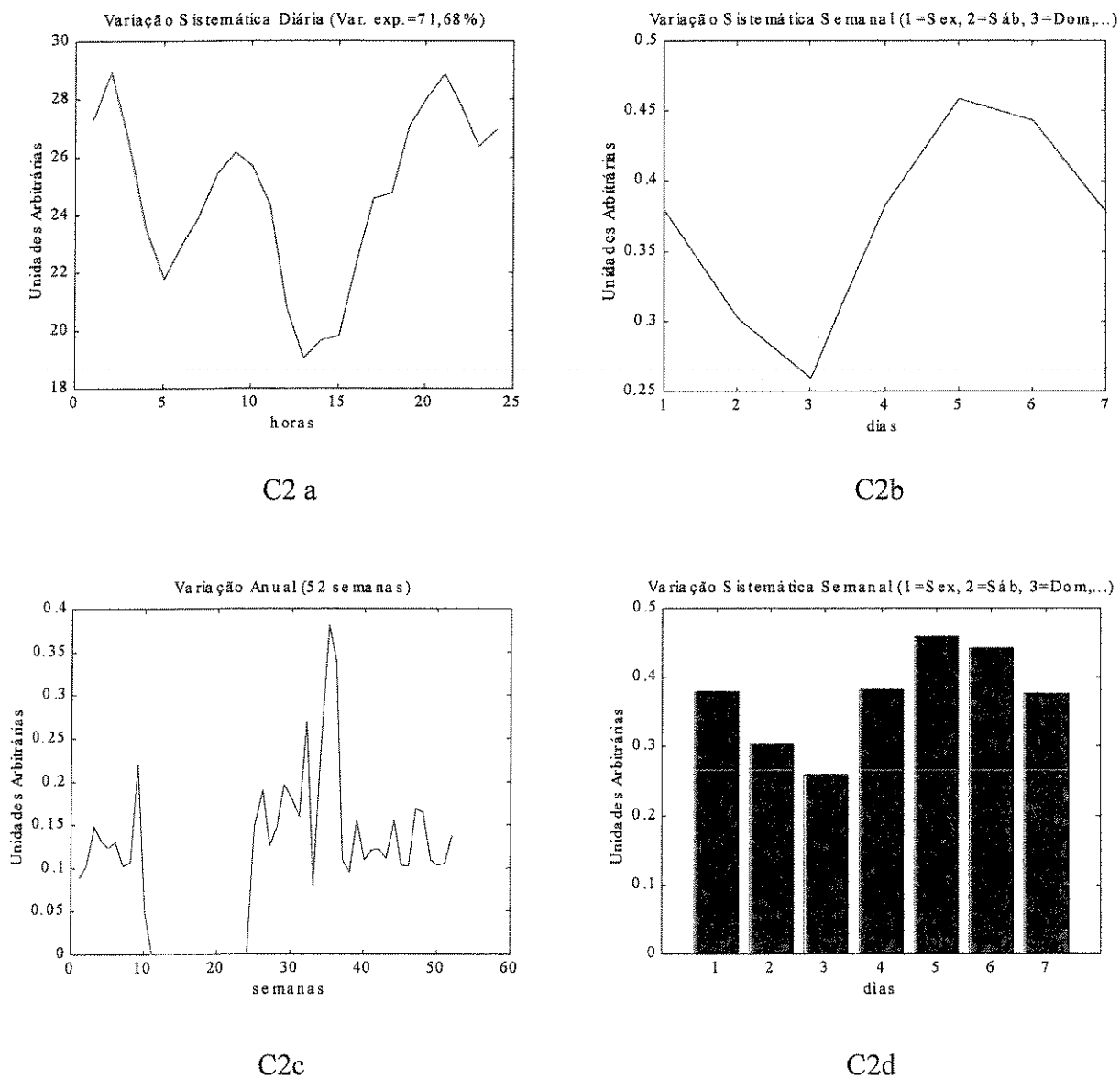


Fig.C2: a)Variações sistemáticas com periodicidade diária. b) e d)Variações sistemáticas com periodicidade semanal. c)Variação anual. Dados para Pedro II – 1999.

Referências Bibliográficas

Akimoto, H.; Wild, J. O.; Prather, J. M. Geophysical Research Letters, 28, 2001, 1719-1725

Alsberg, T.; Hakansson, S.; Strandell, M.; Westerholm, R.. Chemometrics and Intelligent Laboratory Systems, 7, 1989, 143-182

American Geophysical Union. <http://www.sciencedaily.com/releases/2002.htm>

Andersson, C. A.; Bro, R.. Chemometrics and Intelligent Laboratory Systems, 42, 1998, 93-103

Andersson, C. A.; Bro, R.. <http://newton.nli.kvl.dk/foodtech.html>

Appellof, C. J.; Davidson, E. R.. Analytica Chimica Acta, 146, 1983, 9-14

Appellof, C. J.; Davidson, E. R.. Chemometrics and Intelligent Laboratory Systems, 42, 1998, 105-113

Baird, C.. Environmental Chemistry. W. H. Freeman and Company, New York, 1995, Caps. 2-4

Baldasano, J. M.; Delgado, R.; Calbo, J.. Environmental. Science and Technology., 32, 1998, 405-412

Bambauer, A.; Brantner, B.; Paige, M.; Novakov, T.. Atmospheric Environment 1994, 20, 3225-3332

Beebe, K. R.; Pell, R. J.; Seasholtz, M. B.. Chemometrics: A Pratical Guide. Willey, New York 1998

Beebe, K. R.; Kowalski, B. R.. An Introduction to Multivariate Calibration and Analysis. Analytical Chemistry, 59, 1987, 1007A-1017A

Betts, R. A.. Nature, 408, 2000, 187-190

Birks, J.W. In Perspectives in Environment Chemistry; Macalady, D. L., Ed.; Oxford University Press, Inc., New York, **1998**

Bloxom, B.. Psychometrika, 33, **1968**, 347-350

Booksh, K.; Kowalski, B. R.. Journal of Chemometrics, 8, **1994**, 45-63

Boubel, R. W; Fox, Donald, L.; Turner, D. B.; Stern, A. C.. Fundamentals of Air Pollution. 3rd ed., Academic Press, Inc., **1994**, Caps.12, 17-19.

Bro, R.; Workman Jr.; J. J.; Mobley, P. R.; Kowalski, B. R.. Applied Spectroscopy Reviews, 32, **1997**, 237-261

Bro, R.. Chemometrics and Intelligent Laboratory Systems, 38, **1997**, 149-171

Bro, R.. "Multi-way Analysis in the Food Industry: Models, Algorithms and Applications". ACADEMISH PROEFSCRIPT, Universiteit van Amsterdam, **1998**.

Brown, S. D. Applied Spectroscopy, 49, **1995**, A14-A31

Brune, W. H. In Stratospheric Chemistry – Perspectives in Environment Chemistry; Macalady, D. L., Ed.; Oxford University Press, Inc., New York, **1998**

Burdick, D. S.. Chemometrics and Intelligent Laboratory Systems, 28, **1995**, 229-237

Calvert, J. G.; Yarwood, G.; Dunker, A. M.. Research on Chemical Intermediates, 20, **1994**, 463-502

Carrol, J. D.; Chang, J. J.. Psychometrika **1970**, 35, 283-319

Cárdenas, L. M.; Austin, J. F.; Burgess, R. A.; Clemitshaw, K. C.; Dorling, S.; Penkett, S. A.; Harrison, R. M... Atmospheric Environment, 32, **1998**, 3339-3351

Cattell, R. B.. Psychometrika, 9, **1944**, 267-283

Chameides, W. L.; Lindsay, R. W.; Richardson, J.; Kiang, C. S.. Science, 241, **1998**, 1473-1481

Chernicharo, C. A. de L.. Princípios do Tratamento Biológico de Águas Residuárias. Vol. 5, Departamento de Engenharia Sanitária e Ambiental – DESA/UFMG, Belo Horizonte, **1997**

Colbeck, I.; Harrison, R. M.. Atmospheric Environment, 19, **1985**, 1807-1818

Costa, M.; Baldasano, J. M.. Atmospheric Environment, 30, **1996**, 309-318

Cox, P. M.; Betts, R. A.; Jones, C. D.; Spall, S. A.; Totterdell, I. J.. *Nature*, 408, **2000**, 184-187

CETESB. Relatório de Qualidade do Ar no Estado de São Paulo. **1998**

Companhia de Engenharia de Tráfego de São Paulo (CET). <http://www.cetesp.com.br> (Julho/2000).

Dimitroulopoulou, C; Marsh, A. R. W.. *Atmospheric Environment*, 31, **1997**, 3041-3057

Datta, B.. *Numerical Linear Algebra and Applications*. Books/Cole Publishing Company, Pacific Grove, **1995**

Dunn III, W. J.. *Quantitative Structure-Activity Relationships (QSAR) Chemometrics and Intelligent Laboratory Systems*, 6, **1989**, 181-190

Einax, J. W.; Swanziger, H. W.; Geiß, S.. *Chemometrics in Environmental Analysis*. Verlagsgesellschaft VCH, Weinheim, **1997**.

Faber, K.. *Journal of Chemometrics*, 11, **1997**, 87-91

Faber, N. M.; Buydens, L. M. C.; Kateman, G.. *Journal of Chemometrics*, 7, **1993**, 495-526

Faber, N. M.; Buydens, L. M. C.; Kateman, G.. *Journal of Chemometrics*, 8, **1994a**, 147-154

Faber, N. M.; Buydens, L. M. C.; Kateman, G.. *Journal of Chemometrics*, 8, **1994b**, 181-203

Farrar, P. D.. *Climatic Change*, 47, **2000**, 7-15

Ferreira, M. M. C.; Antunes, A. M.; Melgo, M. S.; Volpe, P. L. O.. *Química Nova*, 22, **1999**, 724-731

Fontijn, J. N.; Sabadel, A. J.; Ronco, R. J.. *Analytical. Chemistry*, 42, **1970**, 575-579

Friis-Christensen, E.. *Climatic Change*, 47, **2000**, 1-5

Fuglestad, J. S.; Berntsen, T. K.; Godal, O.; Sausen, R.; Shine, K.P.; Skodvin, T.. *Climatic Change*, 58, **2003**, 267-331

Fukuyama, F.. *The end of the history and the last man*. *American Historical Review*, 94, **1989**, 654 – 698.

Geladi, P.; Kowalski, B. R.. *Partial Least Squares Regression: A Tutorial*. *Analytica Chimica Acta*, 185, **1986**, 1-17

- Geladi, P.. Chemometrics and Intelligent Laboratory Systems, 7, **1989**, 11-30
- Golub, G. H.; van Loan, C. F.. Matrix Computation; The John Hopkins University. Press, London, **1989**
- Grady, C. P. L.; Lim, H.. Biological Wastewater Treatment: Theory and Application. Marcel Decker, New York, **1980**
- Green, P. E.; Carrol, J. D.. Mathematical Tools For Applied Multivariate Analysis. Academic Prees, Inc., San Diego, **1978**
- Güsten, H.; Heinrich, G.; Sprung, D.. Atmospheric Environment, 32, **1998**, 1195-1202
- Haaland, D. M.; Easterling, R. G.; Vopicka, D. A.. Applied Spectroscopy, 39, **1985**, 73-83
-
- Haaland, D. M.; Thomas, E. V.. Analytical Chemistry, 60, **1988**, 1193-1202
- Haaland, D. M.; Thomas, E. V.. Analytical Chemistry, 60, **1988**, 1202-1208
- Haaland, D. M.. Analytical Chemistry, 60, **1988**, 1208-1217
- Hansen, J. E.. Space Science Review, 94, **2000**, 349-356
- Hansen, J. E.; Sato, M.; Ruedy, R; Lacis, A.; Oinas, V.. Proceedings National Academy of Siences, 97, **2000**, 9875-9880
- Hansen, J. E.; Sato, M.. Proceedings National Academy of Sciences, 98, **2001**, 14778-14783
- Harley, P. C.; Monson, R. K.; Lerdau, M. T.. Oecologia, 118, **1999**, 109-123
- Harrison, R. M.; Holmann, C. D.; McCartney, H. A.; McIlvern, J. F. R.. Atmospheric Environment , 12, **1978**, 2021-2026
- Harshman, R. A., Lundy, M. E.. Computational Statistcs & Data Analysis, 18, **1994**, 39-72
- Henrion, R.. Chemometrics and Intelligent Laboratory Systems, 25, **1994**, 1-23
- Henze, M.; Harremoës, P.. Water Science and Technology, 15, **1983**, 1-101
- Ho, C.; Christian, G. D.; Davidson, E. R.. Analytical Chemistry, 50, **1978**, 1108-1113
- Ho, C.; Christian, G. D. Davidson, E. R.. Analytcal Chemitry, 52, **1980**, 1071-1079
- Hopke, Ph. K.. Chemometrics and Intelligent Laboratory Systems, 6, **1989**, 7-19

Howard, C. J. e Evenson, K. M., 1977, apud Birks, J.W. In Perspectives in Environment Chemistry; Macalady, D. L., Ed.; Oxford University Press, Inc., New York, **1998**

I Salau, J. S., Tauler, R., Bayona, J. M., Tolosa, I.. Environment Science Technology, 31, **1997**, 3482-3490

Jackson, J. E.. A User's Guide To Principal Components. John Wiley & Sons, Inc., New York, **1991**

Jacob, D. J. Atmospheric Environment, 34, **2000**, 2131-2159

Jenkin, M.; Cox., R. A., Williams, D. J.. Atmospheric Environment, 22, **1988**, 487-498

Jenkin, M. E.; Hayman, G. D. Atmospheric Environment, 33, **1999**, 1275-1293

Jobson, J. D.. Multivariate Analysis. 6th ed. Lewis Publishers, CRC Press, Inc., **1994**. Caps. 9–10.

Jolliffe, I. T.. Principal Component Analysis. Springer–Verlag Inc., New York, **1986**

Johnston & Podolske, apud Brune W. H. In Stratospheric Chemistry – Perspectives in Environment Chemistry; Macalady, D. L., Ed.; Oxford University Press, Inc., New York, **1998**

Keith, L. H. (ed.). Principles of Environment Chemistry, American Chemical Society, ACS Professional Reference Book, New York, **1988**

Kiers, H. A. L.. Psychometrika, 56, **1991**, 449-470

Kiers, H. A. L... Journal of Chemometrics 24, **2000**, 105-122

Kley, D.; Drumond, J. W.; McFarland, M.; Liu, S. C.. J. Geophys. Res. 86, 1981, 3153-3161, apud Keith, L. H. (ed.). Principles of Environment Chemistry, American Chemical Society, ACS Professional Reference Book, New York, **1988**

Kroonenberg, P. M., De Leeuw, J.. Psychometrika, 45, **1980**, 69-97

Kroonenberg, P. M.. “Three-Mode Principal Component Analysis”. DSWO, Leiden, **1989**

Kroonenberg, P. M.;Computational Statistics & Data Analysis, 18, **1994**, 73-96

Ladstatter-Weissenmayer, A.; Heland, J.; von Kuhlmann, R.; Lawrence, M. G.;Meyer-Arnek, J.; Richter, A.; Ziereis, H.; Burrows, J. P.. Atospheric Chemistry and Physics, 3, **2003**, 1887-1902

Lettinga, G.; Hulshof Pol, L. W.; Zeeman, G.. Biological Wastewater Treatment. Part I: Anaerobic Wastewater Treatment. Lecture Notes. Wageningen Agricultural University, ed. January, **1996**

Levin, J.. Tech. Rept., Dept. Psychol. Univ. Illinois, 1963, in Tucker, L. R.. Psychometrika, 31, **1966**, 279-311

Li S.. Hamilton, J. C., Gemperline, P. J.. Analytical Chemistry, 64, **1992**, 599-607

Lorber, A.. Anal. Chem., 57, **1985**, 2395-2397

Macalady, D. L.. Perspectives in Environmental Chemistry. Oxford University Press, New York, Oxford, **1998**

Madronich, S.. In the role of the Stratosphere in global change. Vol 18. NATO-ASI Series, ed. M-L. Chanin, 437-461. Amsterdam, Springer-verlag, **1993**

Maeder, M., Zilian, A.. Evolving Factor Analysis, a New Multivariate Technique in Chromatography. Chemometrics and Intelligent Laboratory Systems, 3, **1988**, 205-213

Malinowski, E. R.. Factor Analysis in Chemistry. A Wiley-Interscience Publication. John Wiley & Sons, Inc., **1991**

Manahan, E. S.. Environment Chemistry. 6th ed. Lewis Publishers, CRC Press, Inc., **1994**, Caps. 1-2, 9-14.

Manly, B. F. J.. Multivariate Statistical Methods. A Primer. 2 nd ed. Chapman & Hall, St Edmundsbury Press, London, **1997**

Manne, A. S.; Richels, R. G.. Nature, 5, **2001**, 675-677

Marsh, N.; Svensmark, H.. Space Science Reviews, 94, **2000**, 215-230

Mardia, K. V., Kent, J. T., Bibby, J. M.. Multivariate Analysis. Academic Press, London, **1979**

Martens, H.; Naes, T.. Multivariate Calibration. John Wiley e Sons, New York, **1989**

Massart, D. L., Vandeginst, B. G. M., Deming, S. M., Michotte, Y., Kaufman, L.. Chemometrics: a textbook. Elsevier Science Publishing Company, Inc., New York, **1990**

Morrison, F. D.. Multivariate Statistical Methods. 3rd ed. McGraw-Hill, Inc., **1990**

Max, K..O Capital. Ed. Civilização Brasileira S. A., Rio de Janeiro, **1975**, Livro 1, Cap. 1.

Mellouki, A.; Le Bras, G.; Poulet, G.. Journal of Physical Chemistry, 92, **1988**, 2229-2234

- Mellouki, A.; Talukdar, R. K.; Bopegedera, A. M. R. P.; Howard, C. J.. International Journal of Chemical Kinetics, 25, **1993**, 25-39
- Mertes, S.; Wahner, A.. Journal of Physical Chemistry, 99, **1995**, 14000-14006
- Micales, J. A.; Skog, K. E.. International Biodeterioration & Biodegradation, 39, **1998**, 145-158
- Nederbragt, G. W.; Van der Horst, A.; Van Duijin, J. Nature (London), 206, **1965**, 87, apud Keith, L.H., Keith, L. H. (ed.). Principles of Environment Chemistry, American Chemical Society, ACS Professional Reference Book, **1988**
-
- Neu, U.; Künzle, T.; Wanner, H.. Boundary Layer Meteorology, 69, **1994**, 221-247
- Noble, B.. Daniel, J. W.. Applied Linear Algebra. 3rd ed., Prentice Hall, New Jersey, **1988**
- Noble, R. A.; Henk Jr., F. N.. Organic Chemistry, 29, **1998**, 301-314
- Notholt, J.; Hjorth, J. Raes, F.. Atmospheric Environment, 26, **1992**, 211-217
- Paatero, P.. Journal of Computational and Graphical Statistics 8, **1999**, 1-35
- Paulson, S.; E., Seinfeld, J. H.. Environment Science and Technology, 26, **1992**, 1165-1173
- Parkhurst, D. F.. Aritmetic Versus Geometric Means for Environmental Concentration Data. Environment Science and Technology, 1, **1998**, 92A-98A
- Pisano, J. T.; McEndry, I.; Steyn, D. G.; Hastie, D. R... Atmospheric Environment, 31, **1997**, 2071-2078
- Platt, U; Perner, D.; Schroeder, J.; Kessler, C; Toennissen, A.. Journal of Geophysical Researches, 86, **1981**, 11965-11970
- Platt, U; Winer, A. M.; Biermann, H. W.; Atkinson, R; Pitts, J. N.. Environment Science and Technology, 18, **1984**, 365-369
- Platt, U; Le Bras, G.; Burrows, J. P.; Moortgat, G.. Nature, 348, **1990**, 147-149
- Poisson, N.; Kanabidon, M.; Crutzen, P. J.. Journal of Atmospheric Chemistry, 36, **2000**, 157-230
- Reis, M. M.; Ferreira, M. M. C.; Química Nova **1999**, 22, 11-17
- Reis, M. M.; Ferreira, M. M. C.; Journal of Chemometrics, 16, **2002**, 444-450
- Rice, D. D; Claypoal, G. E.. AAPG Bulletin, 65, **1981**, 5-25

- Rind, D.. Nature, 3, **2000**, 1-3
- Rosentiel, T. N.; Potosnak, M.J.; Griffin, K. L.; Fall, R.; Monson, R. K.. Nature, advance online publication, 5 January **2003** (doi: 10.1038/nature 01312)
- Russel, J. B..Química Geral. 2a. ed. Makron Books, São Paulo, **1994**, Vol. 2
- Simpson, I. J.; Blake, D. R.; Rowland, F. S.; Chen, T. Y.. Geophysical Research Letters, 29, 10, **2002**, Art. N° 1479
- Smith, K. A.; Bogner, J.. IGACTivities Newsletter, 10, **1997**, 1-5
- Sanches, E., Kowalski, B. R.. Analytical Chemistry, 58, **1986**, 496-499
- Sanches, E., Kowalski, B. R.. Journal of Chemometrics, 2, **1988**, 247-263
- Sanches, E.. Journal of Chemometrics, 2, **1988**, 265-280
- Sanches, E.. Journal of Chemometrics, 4, **1990**, 29-45
- Schoell, M.; Muehlenbachs, K.; *et alli*.. Submitted to AAPG **2001** National Conference, Denver.
- Sharaf, M. A., Ilman D. L., Kokalski, B. R.. Chemometrics. John Willey & Sons, **1986**
- Shepson, P. B.; Bottenheim, J. W.; Hastie, D. R.; Venkatram, A.. Geophysical Research Letters, 19, **1992**, 1121-1124
- Sherwood, K., Craig, I.. <http://www.co2science.org/edit/v6n3edit.htm>
- Smilde, A. K.. Chemometrics and Intelligent Laboratory Systems, 15, **1992**, 143-157
- Snape, J. B.; Dunn, I. J.; Ingham, J.; Prenosil, J. E.. Dynamics of Environmental Bioprocesses: Modelling and Simulation. VCH Verlagsgesellschaft mbH, D-69451 Weinheim (FDA), **1998**, p.350-372
- Statheropoulos, M., Vassiliadis, N., Pappa, A.. Atmospheric Environment , 32, **1998**, 1087-1095
- Strang, G..Linear Algebra and its Applications. H. B. Jovanovich, Inc. Orlando, **1988**
- Tanner, R. L.; Lee, Y. -N; Kelly, T. J.; Gaffney, J. J.. Presented at the 25th Rocky Mountain Conference, Denver, CO, August, 1983; Paper n. 102, apud Keith L. H. (ed.), Principles of Environmental Sampling, American Chemical Society, ACS Professional Reference Book, **1988**

- Thomas, E. V.. A Primer on Multivariate Calibration. *Analytical Chemistry* 66, **1994**, 795A-804A
- Tsutsumi, Y., Matsueda, H.. *Atmospheric Environment*, 34, **2000**, 553-561
- Tucker, L. R. *Psychometrika*, 31, **1966**, 279-311
- Watkins, S. D.. *Fundamentals of Matrix Computations*. John Wiley & Sons, New York, **1991**
- Wayne, R. P.; Barnes, I.; Biggs, P.; Burrows, J. P.; Conosa-Mos, C. E.; Hjorth, J.; Le Bras, G.; Moortgat, G. K.; Perner, D.; Poulet, G.; Restelli, G.; Sidebottom, H.. *Atmospheric Environment*, 25A, **1991**, 1-203
- Wendell, D.; Stedman, D. H.; Cantrell, C. A.; Dramraver, L. D. *Anal. Chem.* 55, **1983**, 937-940
- Wilson, B. E., Sanches E., Kowalski, B. R.. *Journal of Chemometrics*, 3, **1989**, 493-498
- Wold, H.. in *Research Paper in Statistics*. David F. (ed.) Wiley, New York, **1966**, 411-444
- Wold, S.; Albano, C.; Dunn, W. J.; Esbensen, K.; Hellberg, S.; Johansson, E.; Sjöström, M.. in *Food Research and Data Analysis*, ed. by Martens H. And Russwurm H, Applied Science Publishers, London, **1983**, 147-188
- Wold, S.. *Principal Component Analysis. Chemometrics and Intelligent Laboratory Systems*, 2, **1987**, 37-52
- Wold, S.; Esbensen, K.; Geladi, P.. *Principal Component Analysis. Chemometrics and Intelligent Laboratory Systems* 2, **1987**, 37-52
- Wold, W., Geladi, P., Esbensen, K. H. , Öhman, J.. *Journal of Chemometrics*, 1, **1987**, 41-56
- Zeng, Y., Hopke, P. K.. *Journal of Chemometrics*, 6, **1992**, 65-83