



UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE QUÍMICA

PÓS-GRADUAÇÃO EM FÍSICO-QUÍMICA

FABIANA ALVES DE LIMA RIBEIRO

**APLICAÇÃO DE MÉTODOS DE ANÁLISE MULTIVARIADA NO ESTUDO DE
HIDROCARBONETOS POLICÍCLICOS AROMÁTICOS**

Dissertação apresentada ao Instituto de Química
como parte dos requisitos para obtenção
do título de Mestre em Química

Orientadora: Professora Doutora Márcia Miguel Castro Ferreira

CAMPINAS - SÃO PAULO

Fevereiro - 2001

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

R354a	Ribeiro, Fabiana Alves de Lima Aplicação de métodos de análise multivariada no estudo de hidrocarbonetos policíclicos aromáticos / Fabiana Alves de Lima Ribeiro. – Campinas, SP: [s.n], 2001. Orientadora: Márcia Miguel de C. Ferreira. Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Química. 1. Quimiometria. 2. PAH. 3. Química Ambiental. 4. QSPR. I. Ferreira, Márcia Miguel de Castro. II. Universidade Estadual de Campinas, Instituto de Química. III. Título.
-------	--

Agradecimentos

Na época em que eu estava elaborando este texto, encontrei uma colega dos tempos da graduação que tinha concluído o mestrado há pouco. Lembro dela ter me dito algo como:

“- Este período em que estamos escrevendo é tão ruim, de muita solidão. Apenas você, o micro, os artigos e os livros, e muito... muito trabalho para ser feito...”

Finalizar uma dissertação não é realmente uma tarefa trivial, mas também não é um processo tão solitário assim. Com certeza ele não teria sido possível se não fosse o apoio, a colaboração e a amizade de muitas pessoas, a quem eu gostaria de agradecer.

Entre elas minha orientadora, Prof^a. Márcia M. C. Ferreira, pelo incentivo e pelas contribuições durante a elaboração deste trabalho. Os colegas do grupo de trabalho, Luciana, Thais, Luci, Aline, Edilton, Cristiano, Marlon, Ciriaco, Rudolf e Marcos, e o pessoal da Teórica, que mesmo no dia-a-dia corrido que a vida acadêmica exige sempre proporcionaram um ambiente de companheirismo e cooperação. À Luci, também pela realização dos cálculos iniciais.

À CAPES e ao FAEP, pelo apoio financeiro. Ao CENAPAD, pela utilização do polo computacional.

Aos meus pais e meus irmãos, que acompanharam de longe, e sempre foram compreensivos com as minhas ausências. Ao Mauro, por todo o carinho e pelo auxílio durante a editoração do trabalho.

Resumo

Neste trabalho foram feitas algumas aplicações de métodos de análise multivariada no estudo de hidrocarbonetos aromáticos policíclicos (PAHs). O texto está organizado em três partes:

A primeira parte contém uma breve introdução aos métodos de análise multivariada, com ênfase nas duas técnicas utilizadas neste projeto: análise de componentes principais (PCA) e regressão por mínimos quadrados parciais (PLS). Esta introdução não tem a pretensão de ser minuciosa e extensiva, abordando exaustivamente cada um dos inúmeros métodos de análise multivariada. Foi colocada no texto apenas para situar o leitor menos experiente na área, a respeito dos métodos aqui utilizados.

Na segunda parte foi realizada uma análise exploratória dos efeitos do derramamento de óleo do petroleiro *Exxon Valdez* sobre sedimentos marinhos *em Prince William Sound*, Alasca. Os dados para este estudo foram retirados da literatura, em que os autores analisaram o teor de 10 tipos diferentes de PAHs em 82 amostras de sedimento marinho coletadas em três profundidades diferentes em duas baías atingidas pelo óleo. Este é um exemplo interessante e muito ilustrativo das possibilidades da aplicação de métodos quimiométricos no tratamento de dados ambientais.

Na terceira parte foi realizado um estudo das relações quantitativas estrutura-propriedades dos PAHs por meio da utilização de modelos de previsão de algumas propriedades físico-químicas relacionadas ao impacto ambiental destes compostos. As propriedades estudadas foram: ponto de ebulição, coeficiente de partição octanol-água, índice de retenção e fototoxidez. Para a construção dos modelos, a geometria das moléculas de PAH foi otimizada, e alguns descritores termodinâmicos, eletrônicos, estéricos e topológicos foram calculados: energia dos orbitais de fronteira, dureza molecular, momento de dipolo, eletronegatividade, polarizabilidade, cargas, propriedades termodinâmicas, volume, área, além de 151 índices topológicos. Utilizamos o método dos mínimos quadrados parciais com preprocessamento autoescalado e validação cruzada para elaboração de modelos de regressão de um conjunto de amostras cujas propriedades experimentais eram conhecidas. Uma vez validados estes modelos, tais propriedades foram previstas para um conjunto de moléculas cujos resultados experimentais não foram ainda medidos experimentalmente.

Abstract

This work encompasses some applications on multivariate analysis in order to study aspects related to the polycyclic aromatic hydrocarbon (PAH). It is organized in three parts.

Part I – this part was dedicated to a brief introduction of multivariate analyses methods, emphasizing two techniques used in this project: principal components analysis (PCA) and partial least squares regression (PLS). This introductory part does not intend to be detailed, extensive and exhaustive approach of each of these methods. It was included in the text just to guide the reader not so familiar about the methods used in this area.

Part II – an exploratory data analysis on the effects of the spill oil of the tanker Exxon Valdez over marine sediments in Prince William Sound, Alaska, was made. The data used here was extracted from literature. The authors analysed the content of 10 different types of PAHs in 82 samples of marine sediment collected at three different depths in two bays that were affected by the spill. This is an interesting and very illustrative chemometric application in the treatment of environmental data.

Part III – Quantitative Structure-Property Relationship study of PAHs using prediction models of some physicochemical properties of these compounds related to their environmental impact was made. The following properties were studied: boiling point, octanol-water partition coefficient, retention index, and phototoxicity. Geometry of all PAHs were optimized and used to calculate thermodynamic, electronics, sterics and topologic descriptors. Such descriptors were: HOMO and LUMO energies, molecular hardness, polarizability, charges, thermodynamics properties, volume, area and additionally, 151 topologic index. The data was autoscaled and partial least squares with cross-validation was used for building the regression models. After the models were validated, such properties were predicted for a set of molecules in which experimental measurements were not performed yet.

Fabiana Alves de Lima Ribeiro

Formação Acadêmica

Mestrado em Físico – Química

Laboratório de Quimiometria Teórica e Aplicada

Instituto de Química da Universidade Estadual de Campinas

Título da Dissertação: “Aplicação de Análise Multivariada no Estudo de Hidrocarbonetos Aromáticos Policíclicos”. Defendida em 02/02/2001.

Bacharelado em Química Tecnológica

Instituto de Química da Universidade Estadual de Campinas

Conclusão: Dezembro de 1997

Publicações

F. A. L. Ribeiro & M. M. C. Ferreira, “QSAR Models of the Boiling Point, Octanol-Water Partition Coefficient and Retention Index of Polycyclic Aromatic Hydrocarbons” - submetido

F. A. L. Ribeiro & M. M. C. Ferreira, “QSAR Model of the Phototoxicity of Polycyclic Aromatic Hydrocarbon” - submetido

R. C. R. Figueiredo, F. A. L. Ribeiro, E. Sabadini, “Ciências de Espumas – Aplicação na Extinção de Incêndios”, Química Nova, volume 22, número 1, 1999, 126 – 128.

Trabalhos em Anais de Congresso

F. A. L. Ribeiro, L. R. Cirino, M. M. C. Ferreira, “Estimativa de Índice de Retenção para Análises Cromatográficas de PAHs”, Poços de Caldas, MG, BRASIL, 24^a. Reunião Anual da Sociedade Brasileira de Química (2001).

F. A. L. Ribeiro, E. R. Pereira-Filho, M. M. C. Ferreira, “Utilização de Análise de Componentes Principais na Visualização de Dados Ambientais: Um Estudo de Caso”, Poços de Caldas, MG, BRASIL, 23^a. Reunião Anual da Sociedade Brasileira de Química (2000).

L. R. Cirino, F. A. L. Ribeiro, M. M. C. Ferreira, “Correlação Quantitativa Estrutura-Propriedades de Hidrocarbonetos Aromáticos Policíclicos”, Poços de Caldas, MG, BRASIL, 23^a. Reunião Anual da Sociedade Brasileira de Química (2000).

Experiência Profissional

Rhodia do Brasil

Estágio no Centro de Documentação e Informação – CEDOC

Fevereiro de 1998 - Julho de 1998

CPQBA - UNICAMP

Centro Pluridisciplinar de Pesquisas Químicas, Biológicas e Agrícolas

Estágio supervisionado no Laboratório de Óleos e Essências

Agosto de 1997 – Dezembro de 1997

SIEMENS S. A. – Divisão de Transformadores

Estágio na Área de Controle de Qualidade.

Fevereiro de 1995 - Julho 1995

Sumário

<i>Lista de Quadros.....</i>	<i>xvii</i>
<i>Lista de Figuras.....</i>	<i>xix</i>
 <i>Introdução.....</i>	 <i>1</i>
 <i>Capítulo 1- Métodos de Análise Multivariada.....</i>	 <i>4</i>
A Organização dos dados.....	4
Tipos de Preprocessamento.....	7
Dados Centrados na Média.....	7
Escalamento pela Variância.....	8
Autoescalamento.....	9
Os Métodos de Análise Multivariada.....	9
Análise Exploratória.....	9
Análise de Componentes Principais (PCA).....	10
Análise Hierárquica de Agrupamentos (HCA).....	17
Modelos de Regressão Multivariada.....	19
Calibração e Previsão Multivariada.....	19
Relação Quantitativa Estrutura-Propriedade ou Estrutura-Atividade.....	20
Etapa de Modelagem.....	21
Etapa de Validação.....	22
Algumas Considerações Sobre A Escolha do Modelo.....	24
Os Métodos de Regressão Multivariada.....	28
Métodos dos Mínimos Quadrados Clássicos (CLS).....	28
Métodos dos Mínimos Quadrados Inverso (ILS).....	29
Programas Computacionais.....	35

<u>Capítulo 2 - Utilização de Análise de Componentes Principais para o Estudo dos Efeitos do Derramamento de Óleo sobre Sedimentos em Prince William Sound, Alasca.....</u>		36
O acidente.....		36
O perfil geográfico da região.....		38
Os Trabalhos de Restauração.....		39
Características do óleo ANS (<i>Alaska North Slope</i>) ou <i>Exxon Valdez Oil</i> (EVO)		40
As fontes de hidrocarbonetos da região		41
Apresentação dos dados		43
Preprocessamento dos Dados		44
Tipos de Tratamento		47
Tratamento 1		47
Tratamento 2		47
Resultados		47
Análise Preliminar.....		47
Análise dos Dados.....		49
Tratamento 1: Análise de Componentes Principais das Duas Baías.....		50
Tratamento 2A: Análise de Componentes Principais em Bay of Isles.....		59
Tratamento 2B: Análise de Componentes Principais de Drier Bay.....		61
Discussão dos resultados.....		66
Tratamento 1: Bay of Isles vs. Drier Bay		67
Tratamento 2: A distribuição dos PAHs ao longo das três profundidades em cada baía.....		72
 <u>Capítulo 3 - Relação Quantitativa Estrutura-Propriedades de Hidrocarbonetos Aromáticos Policíclicos</u>		75
Prevedendo as Propriedades Químicas - Materiais e Métodos.....		78
PAH – As Moléculas Estudadas		79
A Pesquisa dos Compostos.....		84
Propriedades Moleculares		94
Ponto de Ebulição (pe)		94

Coeficiente de Partição Octanol-Água (Log K _{ow} ou Log P)	95
Índice de Retenção (IR).....	97
Fototoxidez	99
Descritores Moleculares.....	102
Os descritores eletrônicos.....	103
Os descritores topológicos.....	107
Os descritores geométricos.....	108
O Cálculo dos Descritores Moleculares	110
Apresentação dos Dados	111
Análise dos Dados	112
Seleção de Variáveis	112
QSPR para ponto de ebulição (p.e.).....	114
Apresentação dos Resultados para Ponto de Ebulição.....	114
Discussão dos Resultados para Ponto de Ebulição	118
QSPR para coeficiente de partição octanol-água (Log de K _{ow}).....	122
Apresentação dos Resultados para Log K _{ow}	122
Discussão dos resultados para Log K _{ow}	125
QSPR para Índice de Retenção (IR).....	128
Apresentação dos Resultados para índice de Retenção.....	128
Discussão dos Resultados para Índice de Retenção	133
QSPR para Fototoxidez	136
Apresentação dos Resultados para Fototoxidez.....	136
Discussão dos Resultados para Fototoxidez	141
<u>Conclusão</u>	<u>146</u>
<u>Apêndice I.....</u>	<u>150</u>
<u>Apêndice II.....</u>	<u>165</u>
<u>Referências Bibliográficas</u>	<u>166</u>

Lista de Quadros

Quadro 1: Tabela de variância descrita pelas PC para uma PCA.	14
Quadro 2: Teores de cada grupo de PAH nas profundidades amostradas, em ppb.	45
Quadro 3: Porcentagem de variância descrita pelas componentes principais no tratamento 1, retirando-se apenas a amostra 13 (R101).	50
Quadro 4: Porcentagem de variância descrita pelas componentes principais no tratamento 1 após a retirada da amostra 1.	51
Quadro 5: Porcentagem de variância descrita pelas componentes principais no tratamento 1 após a retirada das amostras 1, 45 e 48.	52
Quadro 6: Comparação das variâncias descritas e respectivos números de fatores utilizados com a retirada das amostras 1 (F011), 45 (F071) e 48 (R031).	53
Quadro 7: Peso de cada variável nas três primeiras componentes principais após a retirada das amostras 1, 13, 45 e 48.	54
Quadro 8: Porcentagem de variância descrita pelas componentes principais retirando-se as variáveis C14P, Per, Petro e C2P uma de cada vez.	57
Quadro 9: Peso de cada variável nas três primeiras componentes principais após a retirada das amostras 1, 13, 45 e 48 e da variável Per.	58
Quadro 10: Porcentagem de variância descrita pelas componentes principais no tratamento 2A (Bay of Isles).	60
Quadro 11: Peso de cada variável na PC1 e PC2, para o tratamento 2A (Bay of Isles).	61
Quadro 12: Porcentagens de variância descrita pelas componentes principais no tratamento 2B.	62
Quadro 13: Peso de cada variável nas três primeiras componentes principais, para o tratamento 2B.	64
Quadro 14: Porcentagem de variância descrita pelas componentes principais após a retirada das amostras F041, R081, R091 e R032.	65
Quadro 15: Peso de cada variável nas duas primeiras componentes principais, para o tratamento 2B, após a retirada das amostras F041, R081, R091 e R032.	66
Quadro 16: Lista dos hidrocarbonetos aromáticos policíclicos utilizados neste trabalho.	86
Quadro 17: Lista das Propriedades em Estudo. Os dados experimentais utilizados para a construção dos modelos foram retirados da literatura.	111
Quadro 18: Coeficientes das variáveis no vetor de regressão para ponto de ebulição.	114
Quadro 19: Resultados dos modelos de QSPR utilizando PLS com validação cruzada para ponto de ebulição (°C).	114
Quadro 20: Erros de validação para o ponto de ebulição (° C).	116
Quadro 21: Valores de Ponto de Ebulição medidos experimentalmente e previstos pelo modelo. (os valores previstos encontram-se em negrito).	117

Quadro 22: Resultados dos modelos de QSPR utilizando PLS com validação cruzada para ponto de ebulição com as variáveis energia do LUMO, volume, PM e R.....	121
Quadro 23: Coeficientes das variáveis no vetor de regressão para Log K _{ow}	122
Quadro 24: Resultados dos modelos de QSPR utilizando PLS com validação cruzada para Log K _{ow} utilizando as variáveis volume molecular, PM e R.	122
Quadro 25: Erros de validação para Log K _{ow}	123
Quadro 26: Valores para Log K _{ow} determinados experimentalmente e previstos pelo modelo (em negrito).	124
Quadro 27: Resultados dos modelos de QSPR utilizando PLS com validação cruzada para Log K _{ow} utilizando as variáveis energia do LUMO, volume molecular, PM e R.....	128
Quadro 28: Coeficiente das variáveis no vetor de regressão para índice de retenção.....	129
Quadro 29: Resultados dos modelos de QSPR utilizando PLS com validação cruzada para Índice de Retenção.	129
Quadro 30: Erros de validação para Índice de Retenção.....	131
Quadro 31: Valores para Índice de Retenção determinados experimentalmente e previstos pelo modelo (em negrito).	132
Quadro 32: Coeficientes de cada variável no vetor de regressão para Fototoxidez.	138
Quadro 33: Resultados dos modelos de QSPR utilizando PLS com validação cruzada para Fototoxidez.	138
Quadro 34: Erros de Validação para Fototoxidez.....	139
Quadro 35: Valores para Fototoxidez determinados experimentalmente e previstos pelo modelo (em negrito).	140

Lista de Figuras

Figura 1: Matriz de dados X.....	4
Figura 2: Representação da amostra x nos eixos da variáveis 1, 2 e 3.....	4
Figura 3: Gráfico dos dados num sistema bidimensional, com os eixos das variáveis 1 e 2 e das duas primeiras componentes principais (PC1 e PC2).	11
Figura 4: Gráficos dos escores (a) e dos pesos (b) de um sistema de dados bidimensional.	12
Figura 5: Representação da decomposição da matriz de dados X.	15
Figura 6: Dendograma de uma análise de agrupamento hierárquico	17
Figura 7: Representação de um processo de calibração multivariada	20
Figura 8: Representação de um estudo de QSAR/QSPR.....	21
Figura 9: Representação gráfica da equação 23.....	31
Figura 10: Representação gráfica da equação 24.....	32
Figura 11: Representação da equação 26.	33
Figura 12: Representação da equação 27.	34
Figura 13: Mapa da região atingida pelo derramamento de óleo do petroleiro Exxon Valdez, mostrando a localização de Prince William Sound (adaptado para o português). Cortesia de Exxon Valdez Oil Spill Trustee Council (site oficial).	37
Figura 14: Prince William Sound, mostrando Knight Island no detalhe, a localização de Bay of Isles e Drier Bay, e a Corrente Costeira do Alasca, retirado do artigo de Boehm <i>et alli</i> (1998).	39
Figura 15: Representação da três regiões em que a amostragem foi realizada.	43
Figura 16: Gráfico de teor de perileno versus C2P mostrando o comportamento da amostra 1 (F011).....	48
Figura 17: Gráfico de teor de perileno versus pirogênicos mostrando o comportamento diferenciado das amostras 45 (F071) e 48 (R031).	48
Figura 18: Gráfico de teor de pirogênicos versus TPAH, mostrando o comportamento diferenciado das amostras 45 (F071) e 48 (R031).	49
Figura 19: Gráfico da variância descrita pelas PCs para o tratamento 1, excluindo a amostra 13.....	51
Figura 20: Gráfico dos escores mostrando-se o comportamento anômalo das amostras 1 (F011), 45 (F071) e 48 (R031).	51
Figura 21: Gráfico da variância descrita em cada componente principal para o tratamento 1, excluindo a amostra 1 (F011) e 13 (R101).	52
Figura 22: Gráfico dos escores mostrando o comportamento das amostras 45 (F071) e 48 (R031).....	52
Figura 23: Gráfico da variância descrita pelas componentes principais para o tratamento 1, excluindo a amostra 1 (F011), 45 (F071) e 48 (R031).	53

Figura 24: Gráficos dos escores em PC1, PC2 e PC3, para o tratamento 1, após a retirada das amostras 1, 13, 45 e 48.	55
Figura 25: Gráficos dos pesos em PC1, PC2 e PC3, para o tratamento 1, após a retirada das amostras 1, 13, 45 e 48.	55
Figura 26: Gráfico dos escores para o tratamento 1, após a retirada das amostras 1, 13, 45 e 48, e da variável Per.	58
Figura 27: Gráfico dos pesos para o tratamento 1, após a retirada das amostras 1, 13, 45 e 48, e da variável Per.	58
Figura 28: Variância descrita pelas componentes principais para o tratamento 2A.	59
Figura 29: Gráficos dos escores para o tratamento 2A (Bay of Isles).	60
Figura 30: Gráficos dos pesos para o tratamento 2A (Bay of Isles).	60
Figura 31: Variância descrita pelas componentes principais para o tratamento 2B.	62
Figura 32: Gráficos dos escores em PC1, PC2 e PC3, para o tratamento 2B.	63
Figura 33: Gráficos dos pesos em PC1, PC2 e PC3, para o tratamento 2B.	63
Figura 34: Gráfico dos escores em PC1 e PC2, para o tratamento 2B, após a retirada das amostras F041, R081, R091 e R032.	65
Figura 35: Gráfico dos pesos em PC1 e PC2, para o tratamento 2B, após a retirada das amostras F041, R081, R091 e R032.	65
Figura 36: Detalhe de Bay of Isles mostrando os braços leste e oeste da baía.	68
Figura 37: Detalhe de Drier Bay mostrando a localização das ruínas de Port Audrey e (Barnes Cove)	71
Figura 38: Estruturas de alguns dos PAHs mais abundantes: (a) benzo[a]pireno, (b) benzo[a]antraceno, (c) criseno, (d) benzo[e]pireno e (e) benzo[j]fluoranteno.	80
Figura 39: Conversão do benzo[a]pireno (I) em 7,8-óxido de benzo[a]pireno (II).	82
Figura 40: Conversão do benzo[a]pireno (II) em benzo[a]pireno-dihidro-7,8-diol-7,8 (III).	82
Figura 41: Formação do dihidroxi-epóxido (IV)	83
Figura 42: Conversão do composto (IV) no composto (V).	83
Figura 43: Gráfico de valor medido vs. valor previsto para ponto de ebulição.	115
Figura 44: Estruturas do antraceno (a), naftaceno (b) e trifenileno (c).	119
Figura 45: Estruturas do acenafteno (a), fluoreno (b), benzo[b]fluoreno (c), benzo[c]fluoreno (d) e benzo[a]fluoreno (e).	120
Figura 46: Gráfico de valor medido vs. valor previsto para Log K _{ow}	123
Figura 47: Estruturas do fluoranteno (a), benzo[ghi]fluoranteno (b), ciclopenta[cd]pireno (c), benzo[j]fluoranteno (d), benzo[k]fluoranteno (e) e benzo[b]fluoranteno (f).	127
Figura 48: Gráfico de valor medido vs. valor previsto para índice de retenção.	130
Figura 49: Estruturas do antraceno (a) e do naftaceno (b).	133
Figura 50: Estruturas do naftaceno (a), benz[a]antraceno (b), criseno (c), trifenileno (d) e benzo[c]fenantreno (e).	134

Figura 51: Estrutura do benzo[c]fenantreno mostrando a não planaridade da molécula: (a) visão inicial; (b) rotação vertical; (c) rotação horizontal.	134
Figura 52: Gráfico de Fototoxidez vs. Energia do HOMO.....	137
Figura 53: Gráfico de Fototoxidez vs. Energia do LUMO.....	137
Figura 54: Gráfico de Fototoxidez vs. GAP.....	137
Figura 55: Gráfico de Fototoxidez vs. Energia do HOMO, após a linearização.....	137
Figura 56: Gráfico de Fototoxidez vs. Energia do LUMO, após a linearização.....	137
Figura 57: Gráfico de Fototoxidez vs. GAP, após a linearização.....	137
Figura 58: Gráfico de valor medido vs. valor previsto para Fototoxidez.....	139
Figura 59: Gráfico de Fototoxidez vs. Energia do HOMO para todos os PAHs.....	144
Figura 60: Gráfico de Fototoxidez vs. Energia do LUMO para todos os PAHs.....	144
Figura 61: Gráfico de Fototoxidez vs. GAP para todos os PAHs.....	144

Introdução

A quimiometria é o emprego de estatística multivariada no tratamento de dados químicos, e em geral suas aplicações enfocam dois temas centrais: a escolha das condições experimentais ótimas para uma investigação, também conhecido como planejamento experimental; e a análise de dados multivariados, de forma a extrair deles o máximo possível de informação química.

Até a década de 60, a maioria dos problemas químicos eram resolvidos através de métodos experimentais simples, como titulação, precipitação, pesagem e reações com reagentes específicos. Um pouco mais tarde, foram surgindo inúmeras técnicas instrumentais de análise, que eram realizadas através da utilização de sofisticados equipamentos: os métodos espectrais como ressonância magnética nuclear (RMN), infravermelho (IV), infravermelho próximo (NIR¹), ultravioleta (UV), fluorescência de raio-X (XRF²) e as várias técnicas cromatográficas, além é claro dos métodos hifenados, como cromatografia gasosa acoplada a espectrofotômetro de massa (CG-EM), cromatografia líquida de alta eficiência com arranjo de diodo UV-Vis (CLAE-UV-Vis), análise por injeção em fluxo (FIA³) (Kowalski, 1983).

Com o desenvolvimento tecnológico destas técnicas experimentais analíticas, e a crescente utilização de computadores, tornou-se possível obter uma quantidade de dados até há bem pouco tempo atrás impensável num tempo tão reduzido. A velocidade dos processos aumentou, permitindo também o aumento da quantidade de dados gerados em um tempo menor, e lógico, com muito menos trabalho manual.

Mas os olhos e a mente humana têm suas limitações, e à medida que cresce a quantidade de dados e as correlações entre eles, cresce também a dificuldade de se obter informações interessantes, e a avaliação das variáveis que mais contribuem para se extrair a informação desejada torna-se uma tarefa cada vez mais árdua. Em geral, a escolha destas variáveis é feita após um minucioso exame individual de cada uma e

¹ NIR – Utiliza-se a sigla do nome em inglês, *Near Infrared*.

² XRF – idem, *X - Ray Fluorescence*.

³ FIA – idem, *Flow Injection Analysis*

depende muito da experiência do profissional que está lidando com elas, sendo muitas vezes realizada de maneira intuitiva, e portanto muito mais suscetível à erros.

A utilização de estatística multivariada permite que a avaliação da influência das variáveis na informação que se deseja extrair dos dados seja feita de forma global, sendo possível observar as correlações entre elas.

Desta forma, o emprego de técnicas multivariadas tem encontrado inúmeras aplicações no tratamento de dados químicos, como por exemplo na discriminação de diferentes variedades de café de acordo com o conteúdo de metais, em trabalhos de química de alimentos (Martin *et alli*, 1998 e 1999); ou na alocação de fontes de contaminação, em trabalhos de química ambiental (Burns *et alli*, 1997; Salau *et alli*, 1997); na caracterização de regiões costeiras e suas variações sazonais à partir da composição química das águas, em trabalhos de oceanografia (Ferreira *et alli*, 1999); e muitos outros trabalhos, até mesmo de áreas não ligadas à química, como economia, (Weldscholte *et alli*, 1998) e psicologia (Taylor & Cox, 1998; Barratt *et alli*, 1999).

Neste trabalho, algumas aplicações de análise multivariada serão utilizadas no estudo de hidrocarbonetos policíclicos aromáticos (PAH^4), um grupo de compostos que se caracterizam por possuir mais de dois anéis aromáticos condensados, e que têm despertado o interesse da comunidade científica devido ao seu impacto ambiental e às características carcinogênicas apresentadas por algumas estruturas. Estes compostos são produzidos pela queima de matéria orgânica, e são introduzidos no meio ambiente por meio de atividades humanas, contaminando solo, água, sedimento e alimentos (Pereira Netto *et alli*, 2000; Lehner *et alli*, 1996).

No estudo destes compostos serão empregadas duas técnicas de estatística multivariada: *análise de componentes principais* (PCA^5), geralmente utilizada como método de análise exploratória ou como método de classificação; e o método dos *mínimos quadrados parciais* (PLS^6) mais usado em problemas de calibração ou estudos

⁴ Polycyclic Aromatic Hydrocarbon.

⁵ Principal Component Analysis

⁶ Partial Least Square

de relação quantitativa estrutura-propriedades (QSPR⁷), ou estrutura-atividades (QSAR⁸).

No primeiro capítulo deste trabalho serão introduzidas algumas breves informações a respeito de análise multivariada, com ênfase nos dois métodos utilizados. Esta pequena introdução não tem a pretensão de ser extensiva e abordar minuciosamente cada método, pois para tanto existem excelentes publicações, e algumas serão citadas e sugeridas ao longo do texto caso o leitor deseje se aprofundar mais no assunto. Este capítulo tem por objetivo apenas introduzir algumas noções básicas para que o leitor que não tenha familiaridade com o assunto se situe em relação aos métodos utilizados neste trabalho.

O segundo capítulo ilustra um dos métodos apresentados no capítulo 1, a PCA, que representa a base para o entendimento de muitas das técnicas de análise multivariada, por meio de um estudo de caso envolvendo a contaminação do ambiente marinho por PAHs.

A PCA foi aplicada no estudo do derramamento de óleo do petroleiro *Exxon Valdez* sobre sedimentos bênticos em *Prince William Sound*, Alasca, cujos dados foram retirados da literatura. A escolha deste conjunto de dados para demonstrar a aplicação de um método multivariado em química ambiental se deve em parte por se referir à um desastre ambiental que pela sua dimensão impulsionou a pesquisa sobre a recuperação de áreas atingidas por derramamentos de óleo, e também por se tratar de análises de PAHs (os compostos escolhidos como tema para este trabalho de pesquisa).

No terceiro capítulo serão apresentados estudos da relação estrutura-propriedade e estrutura-atividade de um conjunto de 70 PAHs. Neste estudo, técnicas de estatística multivariada serão empregadas para construir modelos de regressão para prever algumas propriedades moleculares e atividades biológicas para estes compostos, cujos valores não tenham sido determinados experimentalmente. Para a construção destes modelos de regressão serão testados alguns descritores moleculares de natureza geométrica, eletrônica e molecular. Após construídos, a validade destes modelos será testada para avaliar a sua capacidade preditiva.

⁷ *Quantitative Structure Property Relationship*

⁸ *Quantitative Structure Activity Relationship*

Capítulo 1

Métodos de Análise Multivariada

A Organização dos dados

Antes de iniciar um tratamento os dados devem ser organizados e uma avaliação prévia deve ser feita, pois dependendo da sua natureza pode ser necessário um tratamento prévio, ou a transformação dos dados, ou até mesmo para saber o tipo de método de análise multivariada mais adequado.

A primeira coisa a fazer é organizar as informações na forma de uma matriz $\mathbf{X}$ ⁹, chamada *matriz de dados*, em que as linhas representam as n -amostras e as colunas representam as m -variáveis, como mostra o esquema apresentado na figura 1. Na matriz, cada variável é representada por um vetor coluna, e cada amostra é representada por um vetor linha, sendo que é possível descrever cada vetor linha, ou vetor resposta, como um ponto no espaço de dimensão m , isto é, em que os eixos das coordenadas são descritos pelas variáveis, como pode ser representado na figura 2.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$$

Figura 1: Matriz de dados $\mathbf{X}$.

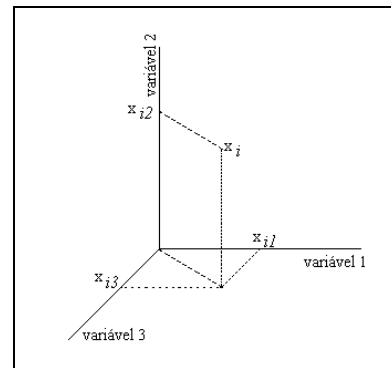


Figura 2: Representação da amostra x nos eixos da variáveis 1, 2 e 3.

⁹ A notação utilizada em álgebra linear utiliza letras maiúsculas em negrito para representar matrizes.

Por exemplo, na determinação da concentração de 5 diferentes PCBs¹⁰ em amostras de solo contaminado coletadas em 10 pontos diferentes, organizaríamos a matriz de dados de forma que cada linha representasse um ponto de coleta e cada coluna representasse o teor de um tipo de PCB em cada amostra. Esta matriz é o ponto de partida para qualquer método de análise multivariada.

Gráficos como aquele da figura 2 são bastante ilustrativos, e podem fornecer informações interessantes sobre o conjunto de dados, de uma maneira mais eficiente do que a observação da matriz **X**. Mas análises baseadas neste tipo de gráficos têm uma limitação: somente três variáveis podem ser utilizadas simultaneamente em cada gráfico tratado. Sendo assim, não é possível analisar a influência de todas as variáveis envolvidas de uma única vez. Esta é uma das limitações da estatística univariada. A utilização de métodos estatísticos de análise multivariada contorna esta limitação.

Uma adequada organização e avaliação dos dados é essencial para que a análise multivariada seja feita de forma correta. Para tanto, alguns fatores devem ser levados em consideração. Deve-se investigar a forma como os dados foram gerados, os métodos de medidas utilizados e a confiabilidade destes dados.

Outro fator importante é a representatividade do grupo de amostras disponíveis que se deseja analisar. De nada adianta um conjunto extenso de dados com informações interessantes se o grupo de amostras não for representativo o suficiente para fornecer as informações adequadas ao tratamento que se propõe fazer.

Por exemplo, suponha que se deseja utilizar um método multivariado para resolver o seguinte problema: uma indústria produz diacetona álcool (DAA) através da condensação da acetona em meio alcalino e temperatura ambiente e que outros subprodutos também sejam formados durante a reação. Em algumas condições ainda desconhecidas, ocorre uma inversão na direção da reação, fazendo com que no final do processo se obtenha um teor de reagente muito alto para as especificações exigidas para que o produto final seja comercializado, acarretando em atrasos no fornecimento de DAA e conseqüentes prejuízos para a indústria. O laboratório de controle de qualidade dispõe de um enorme número de análises do produto formado, as variações de

¹⁰ PCB – Abreviação para *polychlorinated biphenyls*, um grupo de compostos de conhecida ação pesticida (Tolosa *et alli*, 1995).

temperatura ao longo do tempo em que o problema vem ocorrendo e uma série de outras informações que possam ser organizadas como as *variáveis* (as colunas de nossa matriz de dados) que poderiam estar contribuindo para a ocorrência ou não da reversão. Nas linhas da matriz seriam colocadas as *amostras* de DAA formado em cada dia de processo, e estas amostras seriam divididas em duas *classes*: uma em que o fenômeno não ocorreu (vamos chamá-la de classe 1), e uma outra em que tenha ocorrido a reversão (a classe 2).

Um método que poderia ser utilizado com muita eficiência para tentar identificar quais as variáveis estariam influenciando na ocorrência ou não do problema da reversão, ou seja, quais variáveis estariam relacionadas a esta ou aquela *classe* de amostras, seria a *análise de componentes principais*. Supondo que as variáveis escolhidas contenham a informação desejada, a única exigência do método seria então conseguir um conjunto de amostras que tivesse mais ou menos o mesmo número de amostras da classe 1 e 2, para que a população de amostras estivesse bem representada.

Como dito anteriormente, os químicos responsáveis pelo processo não sabem as causas desta reversão, e portanto não têm como controlar sua ocorrência. Agora suponha que o acaso estivesse contra a aplicação da PCA e durante todo o tempo em que as amostras foram coletadas e analisadas (vamos supor 3 meses, sendo uma alíquota por dia, portanto uma amostra, uma linha de nossa matriz de dados por dia), o problema da reversão teria ocorrido somente 4 vezes. Fazendo as contas, em três meses de coleta e análises químicas diárias teríamos 90 (vamos arredondar o mês para 30 dias) amostras das quais 86 pertenceriam à classe 1 e somente 4 pertenceriam à classe 2. As amostras não seriam representativas das duas classes, então a análise não seria possível de ser realizada com sucesso.

Mas este foi só um exemplo de como os dados devem ser averiguados antes de se propor esta ou aquela análise, para avaliar a sua viabilidade.

Outra etapa prévia importante num tratamento de dados é a etapa do *preprocessamento*. Muitas vezes os dados são expressos em grandezas diferentes, ou apresentam muito ruído, e dependendo do tipo de informação que se deseja extrair, isto poderá ou não comprometer a análise. Um tratamento prévio dos dados, para ajustar as

grandezas em valores equivalentes ou de possível comparação resolve este tipo de problema.

Tipos de Preprocessamento

Segundo Beebe *et alli* (1998), “antes de se iniciar qualquer tipo de análise multivariada, é necessária a realização de uma manipulação matemática prévia do conjunto de dados para adequação, ou as vezes até mesmo remoção de possíveis fontes de variação”. Por exemplo, muitas vezes as variáveis possuem diferentes dimensões e amplitudes e é necessário um tratamento prévio para expressar cada observação em dimensões e amplitudes equivalentes, sem perda de informações. Há várias maneiras de se realizar este preprocessamento, os mais comuns são *centrar os dados na média*, *escalamento pela variância* e o *autoescalamento*. Todos estes métodos são sensíveis à presença de amostras anômalas¹¹, que são amostras que apresentam um comportamento muito diferente do restante do conjunto.

Dados Centrados na Média

Este caso é utilizado quando há presença de sinais caracterizados como ruído e é normalmente utilizado na construção de modelos de calibração. Neste caso a média de cada variável é subtraída de seus respectivos elementos, como na equação 1. Desta forma, a origem dos eixos no qual os dados se encontram é deslocada de modo a colocar os dados numa forma mais conveniente à análise e visualização.

$$x_{ij(cm)} = x_{ij} - \bar{x}_j \quad (1)$$

Em que:

$x_{ij(cm)}$ = valor centrado na média para a variável j na amostra i ;

¹¹ É muito comum a utilização do termo em inglês *outlier*, para designar as amostras com comportamento anômalo.

x_{ij} = valor da variável j amostra i ;

$\bar{x}_j$ = média dos valores das amostras na coluna j , calculado pela equação 2;

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad (2)$$

Na equação 2, n é o número de amostras.

Escalamento pela Variância

Neste tipo de preprocessamento, cada valor da variável j é dividido pelo seu desvio padrão, como na equação 3. É utilizado quando as variáveis possuem dimensões muito discrepantes entre si. Desta forma, o peso de variáveis em diferentes escalas é considerado equivalente, minimizando o risco de perda de informações relevantes.

$$x_{ij(ev)} = \frac{x_{ij}}{S_j} \quad (3)$$

Em que:

$x_{ij(ev)}$ = valor escalado pela variância para a variável j na amostra i ;

x_{ij} = valor da variável j na amostra i ;

S_j = desvio padrão dos valores da variável j , calculado a partir da variância S^2 dada pela equação 4.

$$S_j^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \quad (4)$$

Na equação 4:

n = número de amostras;

x_{ij} = valor da variável j amostra i ;

$\bar{x}_j$ = média dos valores das amostras na coluna j , calculado pela equação 2;

Autoescalamento

O autoescalamento aplica ambas as técnicas descritas anteriormente de uma só vez, de modo que a transformação realizada sobre o conjunto original dos dados permite que cada variável apresente média zero e variância igual a um (autoescalamento). Desta forma será dada a mesma importância para todas as variáveis, independente da sua dimensão (equação 5).

$$x_{ij(as)} = \frac{\bar{x}_j}{s_j} \quad (5)$$

$x_{ij(as)}$ = valor autoescalado da variável j para a amostra i ;

$\bar{x}_j$ = média dos valores das amostras na coluna j , calculado pela equação 2;

s_j = desvio padrão dos valores da variável j , calculado a partir da variância

s^2 dada pela equação 4.

Os Métodos de Análise Multivariada

Existem diversos tipos de técnicas de estatística multivariada, com as mais variadas aplicações. Num primeiro instante estes métodos podem ser classificados em dois tipos principais: os métodos de *análise exploratória* e os métodos de *calibração multivariada*. A aplicação de um ou outro método, ou até mesmo da combinação dos dois depende da natureza do problema que se deseja resolver por meio da utilização de análise multivariada, ou do tipo de informação que se deseja obter.

Análise Exploratória

Antes de iniciar o tratamento de um conjunto de dados, principalmente aqueles muito extensos, é necessário uma análise prévia para se avaliar a qualidade das

informações que se tem em mãos. O objetivo dos métodos de análise exploratória é justamente fazer a avaliação inicial dos dados para descobrir que tipo de informação é possível extrair deles, e assim definir as diretrizes para um tratamento mais aprofundado. Isto é feito através da utilização de algoritmos que permitem reduzir a dimensão dos dados, ou organizá-los numa estrutura que facilite a visualização de todo o conjunto, de forma global. Para realizar esta análise inicial, os dois métodos estatísticos mais conhecidos são a *Análise de Componentes Principais (PCA)*¹² e a *Análise Hierárquica de Agrupamentos (HCA)*¹³.

Este capítulo irá tratar mais detalhadamente da PCA, que foi utilizada neste trabalho. A abordagem da HCA será mais superficial, somente para que o leitor que ainda não tenha familiaridade com esta área possa se situar e tomar conhecimento da existência e aplicação deste método¹⁴.

A análise exploratória permite ainda avaliar a possibilidade da construção de modelos de regressão ou de classificação.

Análise de Componentes Principais (PCA)

A análise de componentes principais é uma técnica de estatística multivariada que tem por objetivo reduzir a dimensão dos dados originais permitindo a fácil visualização das informações mais importantes em um número menor de fatores, ou *componentes principais*. A PCA é o ponto de partida para vários métodos de análise multivariada, e tem tido as mais variadas aplicações em: química de alimentos (Haswell *et alli*, 1998; Martin *et alli*, 1996, 1998 e 1999); na alocação de fontes de contaminação, em trabalhos de química ambiental (Burns *et alli*, 1997; Salau *et alli*, 1997); oceanografia (Ferreira *et alli*, 1999); e até mesmo em trabalhos de outras áreas não ligadas à química, como economia, (Weldscholte *et alli*, 1998), pelo Sistema de Informações Geográficas (SIG)¹⁵ (Zhang & Selinus, 1998) e em psicologia (Taylor & Cox, 1998; Barratt *et alli*, 1999).

¹² A sigla é originária do termo em inglês *Principal Component Analysis*.

¹³ A sigla é originária do termo em inglês *Hierarchical Cluster Analysis*.

¹⁴ Cf. Beebe *et alli*, 1998.

¹⁵ GIS – *Geographical Information System*.

Malinowski foi quem primeiro utilizou a PCA para o tratamento de dados químicos, ainda com o nome de *análise de fatores principais* (Malinowski, 1991), isto ainda no final da década de 60. Poucos anos depois inúmeros trabalhos sobre as diferentes aplicações deste método na resolução de problemas químicos foram publicados (Wold *et alli*, 1987).

Voltemos à ilustração das figuras 1 e 2, em que tínhamos uma matriz de dados $\mathbf{X}$ formada por n objetos (as amostras, ou as linhas, ou ainda, os n vetores linha) descritas por m variáveis (os m vetores colunas). A maior limitação para se analisar o conjunto de dados era a impossível visualização de espaços com a dimensão m maior que três (impossível traçar gráficos com mais de três coordenadas).

Na PCA, as coordenadas das amostras são reescritas num novo sistema de eixos, mais conveniente à análise dos dados, chamados de *componentes principais* ou *fatores*¹⁶. Neste novo sistema de eixos, cada uma das componentes principais é gerada a partir da combinação linear das m -variáveis originais, e são ortogonais entre si, ou seja, são completamente não correlacionadas (figura 3).

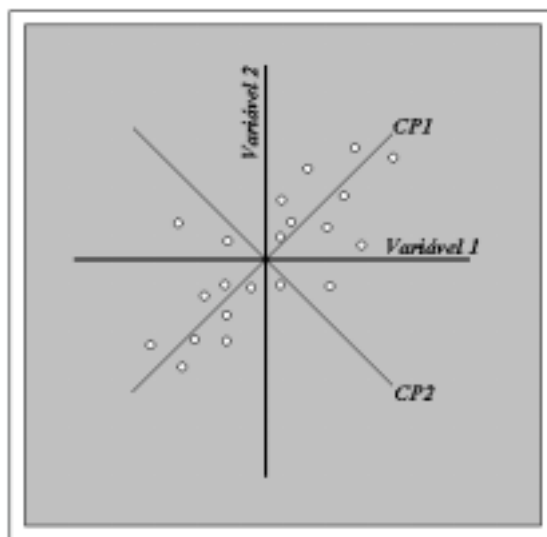


Figura 3: Gráfico dos dados num sistema bidimensional, com os eixos das variáveis 1 e 2 e das duas primeiras componentes principais (CP1 e CP2).

¹⁶ É comum a utilização da sigla PC, do termo em inglês *principal component*, para designar estes fatores.

De uma forma geral e bem simplificada, as combinações lineares das m -variáveis originais que geram cada componente principal podem ser representadas pela equação 6.

$$PC_i = a_{i1}v_1 + a_{i2}v_2 + \dots + a_{im}v_m \quad (6)$$

Nesta equação, v_j (para $j = 1, 2, \dots, m$) são as variáveis originais e a_{ij} (para $j = 1, 2, \dots, m$) são os coeficientes que medem a importância de cada variável na i -ésima componente principal (CP_i), ou seja, o *peso* que cada variável tem naquela combinação linear. Estes pesos nada mais são do que o cosseno do ângulo entre o eixo da componente principal e o eixo da variável original, e portanto seu valor estará sempre entre 1 e -1 . Quanto mais próximo de ± 1 , maior a influência que esta determinada variável tem na descrição desta componente principal *i.e.*, quanto mais próximo de zero este coeficiente estiver, menor a influência da variável naquela componente principal (Beebe *et alli*, 1998; Thomas, 1994).

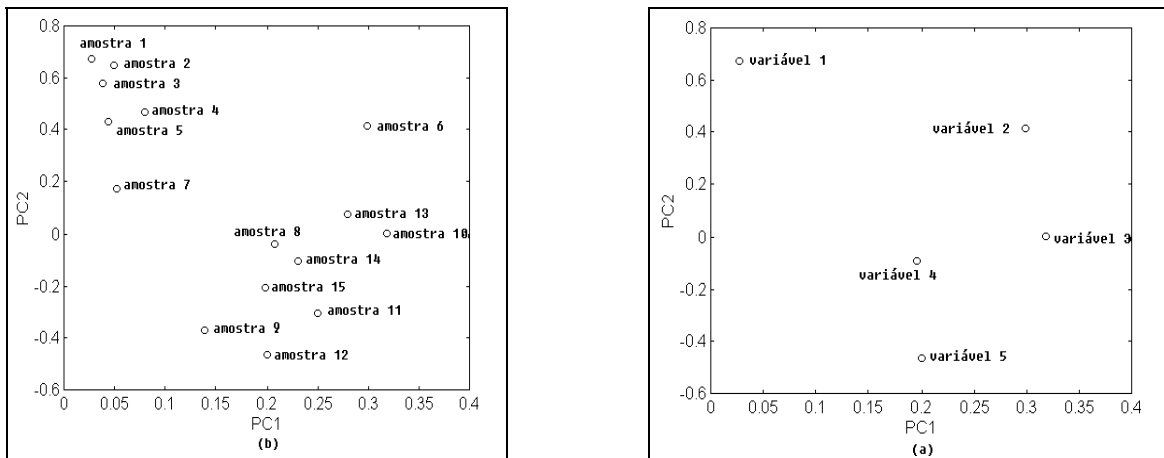


Figura 4: Gráficos dos escores (a) e dos pesos (b) de um sistema de dados bidimensional.

A projeção de cada amostra i ($i = 1, 2, \dots, n$) neste novo sistema de eixos fornece os *escores*¹⁷, e cada amostra terá então um valor de *escore* para cada um dos novos

¹⁷ O termo em inglês (*scores*) também é muito utilizado.

eixos (as CPs), que podem ser organizados na forma de um gráfico como o da figura 4a. (Wold *et alli*, 1987).

O gráfico dos escores poderá revelar agrupamentos ou tendências das amostras analisadas, que poderiam ser de difícil visualização no caso das variáveis originais. Pode ainda destacar amostras cujo comportamento seja muito diferente do restante do conjunto, por exemplo no caso dos PCBs citado anteriormente, uma determinada amostra de solo que apresente um teor do PCB do tipo 1 muito mais alto do que as outras amostras¹⁸. Anomalias assim somente poderão ser detectadas através do exame minucioso de todo o conjunto de dados, e da comparação entre os valores de concentração um a um, o que pode ser um trabalho demorado e muito sensível ao erro humano, dependendo da quantidade de dados que se está analisando, da experiência do analista etc. A PCA torna esta visualização muito mais óbvia.

Como dito anteriormente, os *pesos*¹⁹ (os coeficientes a_{ij}) são capazes de mostrar quais variáveis originais tem maior importância na combinação linear de cada componente principal (Ferreira *et alli*, 1999). Através disto pode-se saber por exemplo, que variáveis contribuem para a descrição deste ou daquele conjunto de amostras. A visualização da importância dos *pesos* para esta ou aquela componente principal, ou para este ou aquele agrupamento de amostras torna-se mais fácil se os *pesos* forem colocados num gráfico, chamado *gráfico dos pesos* (figura 4b).

Um procedimento básico para melhorar a análise de modo a reduzir a dimensão dos dados é a escolha do número de componentes principais a serem utilizadas na descrição do sistema. Voltemos à figura 3, onde é possível observar que os dados estão mais espalhados ao longo da primeira componente principal, que agrupa as variáveis que estão mais correlacionadas. Isto significa que as variáveis originais com maior peso na combinação linear das primeiras componentes principais descrevem a direção da maior variação dos dados. Isto quer dizer que a primeira componente principal descreve uma maior quantidade de informação que a segunda, que descreve mais que a terceira e assim por diante. A escolha do número de componentes principais permite descrever até

¹⁸ As amostras anômalas, ou *outliers*.

¹⁹ O termo em inglês (*loadings*) também é comumente utilizado.

mais de 90% da variância dos dados com um número muito menor de fatores do que das variáveis originais (Ferreira *et alli*, 1999; Moita Neto & Moita, 1998).

A escolha das componentes principais a serem utilizadas na descrição dos dados é feita levando-se em conta a porcentagem de variância descrita pelas CPs e a variância residual. Por exemplo, observe o quadro 1, em que estão as variâncias descritas pelas componentes para uma PCA.

Quadro 1: Tabela de variância descrita pelas CP para uma PCA

	Variância Descrita	Variância (%)	Variância Acumulada (%)
CP1	264.42	83.94	83.94
CP2	32.9	10.44	94.39
CP3	10.73	3.41	97.79

A CP1 descreve somente 83,9% da variância total dos dados, o que é insuficiente, pois a variância residual ainda é muito grande; alguma informação poderá estar sendo perdida, por exemplo, a contribuição de variáveis com alto peso na CP2 poderia estar sendo desprezada, o que prejudicaria a descrição dos dados. É necessário então a utilização de mais uma componente. A CP2 descreve 10,44% da variância total dos dados, que somada à CP1 descreverão juntas 94,39%. A CP3 descreve somente 3,41% da variância total dos dados, o que aumentaria para 97,79% de variância total descrita. Apesar deste aumento na variância total, a quantidade de informação descrita pela CP3 é pequena (apenas 3.41%), e sua inclusão poderia estar inserindo um resíduo desnecessário a descrição dos dados. Pode-se dizer então para este caso, que 2 componentes principais seriam suficientes.

No entanto cada caso deve ser avaliado com cautela, pois cada conjunto de dados tem suas particularidades. Mas de um modo geral pode-se dizer que a escolha do número de componentes principais deve permitir a descrição do sistema com um número mínimo de fatores que descrevam a variação máxima, sem introduzir informações desnecessárias. Mas isto não significa que o número de variáveis originais utilizado para descrever o sistema será menor. Todas as variáveis originais serão utilizadas, a diferença é que serão escolhidas as combinações destas variáveis que

melhor descrevam o conjunto com o máximo de variação, sem introduzir ruído, ou informações irrelevantes.

a) Conceitos Matemáticos

Como dito anteriormente, a equação 6²⁰ representa de forma simplificada como são calculadas as combinações lineares das variáveis que dão origem às componentes principais. Em termos matemáticos a PCA corresponde à decomposição da matriz de dados $\mathbf{X}$, de dimensão $n \times m$, no produto de duas matrizes: a matriz dos *escores* $\mathbf{T}$ e a transposta da matriz dos *pesos* $\mathbf{L}^T$, como mostra a equação 7, que pode ser representada pelo esquema da figura 5.

$$\mathbf{X} = \mathbf{T} * \mathbf{L}^T \quad (7)$$

Os vetores linha da matriz dos escores $\mathbf{T}$ correspondem às projeções das n -amostras, ou objetos, da matriz original $\mathbf{X}$, nos novos eixos formados pelas m -combinações lineares das variáveis originais, ou seja, cada vetor coluna da matriz dos escores corresponde às coordenadas das amostras em cada *componente principal* gerada.

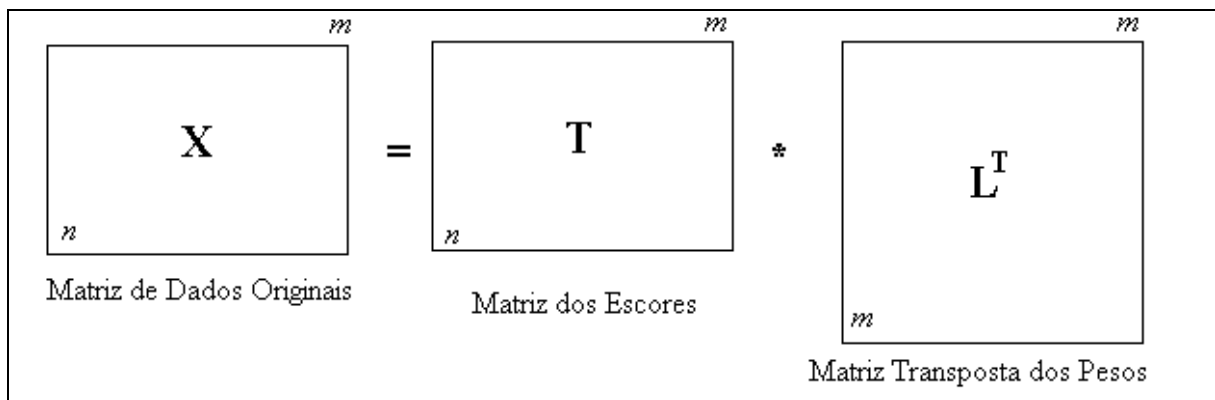


Figura 5: Representação da decomposição da matriz de dados $\mathbf{X}$.

²⁰ Cf. página 12.

Desta forma, o número n de linhas da matriz original é igual ao número de linhas da matriz dos escores e o número de colunas corresponde ao número m de componentes geradas (Wold *et alli*, 1987).

Na matriz dos pesos $\mathbf{L}$, cada vetor coluna corresponde aos pesos que cada variável possui na combinação linear das m componentes geradas.²¹ Então esta matriz possui m linhas (as variáveis originais) e m colunas (as componentes geradas) (Wold *et alli*, 1987).

b) Sobre a Escolha do Número Ótimo de Componente Principais

Como dito anteriormente, cada componente principal gerada descreve uma certa quantidade de informação através da combinação linear das variáveis originais na direção da maior variância dos dados (figura 3), mas não descreve a variância total (exemplo do quadro 1). A cada componente modelada uma certa quantidade de informação permanece sem ser explicada, ou seja, uma quantidade de variância continua não descrita. A esta quantidade de variância não descrita chamamos de *resíduos* e podem ser organizados na forma da matriz dos resíduos $\mathbf{E}$. Podemos então reescrever a equação 7 como descrito na equação 8.

$$\mathbf{X} = \mathbf{T}_k * \mathbf{L}_k^T + \mathbf{E} \quad (8)$$

Este resíduo é formado pela subtração de cada elemento da matriz original $\mathbf{X}$ dos produtos da matriz dos escores e da matriz transposta dos pesos, como na equação 9. Este resíduo será utilizado na escolha do número ótimo de componentes principais (k) a ser utilizado na análise.

$$\mathbf{E} = \mathbf{X} - \mathbf{T}_k * \mathbf{L}_k^T \quad (9)$$

A dimensão da matriz de dados originais necessária à completa descrição do sistema correspondia ao número de colunas, ou seja, ao número de variáveis (m). Após

²¹ i.e., corresponde aos coeficientes a_{ij} da equação 6 da página 12.

a PCA, a dimensão do sistema corresponde ao número de colunas da matriz dos pesos, ou seja, ao número de componentes principais k escolhidas, que é menor que o número de variáveis originais.

O cálculo das matrizes dos escores e dos pesos para uma análise de componentes principais pode ser feito por uma série de algoritmos, sendo que os mais conhecidos são a decomposição de valor singular (SVD) e o NIPALS.²²

Análise Hierárquica de Agrupamentos (HCA)

Esta outra técnica de análise exploratória tem como objetivo observar a formação de agrupamentos naturais das amostras a partir das suas similaridades, mas numa estrutura hierárquica em que as amostras são colocadas num sistema de grupos e subgrupos dependendo do grau desta similaridade. Para entender melhor como isto acontece pode-se fazer uma analogia com a forma como os seres vivos são classificados em biologia segundo suas similaridades: reino, filo, classe, ordem, família, gênero e espécie. Esta classificação agrupa os seres vivos em diferentes classes segundo uma ordem hierárquica baseada nos aspectos morfológicos comuns.

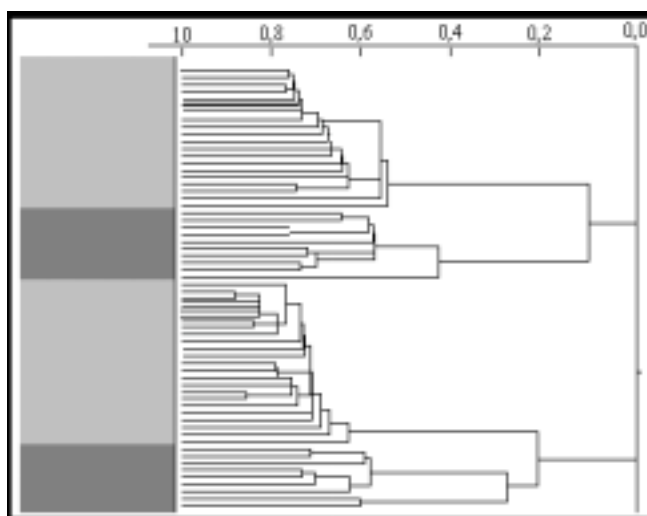


Figura 6: Dendrograma de uma análise de agrupamento hierárquico

Na HCA, esta estrutura é organizada na forma de um gráfico bidimensional chamado *dendograma* onde é possível observar as correlações e similaridades entre as amostras (figura 6). A HCA pode ser igualmente aplicada nas variáveis, para avaliar as similaridades e correlações (Ferreira *et alli*, 1999).

A similaridade entre as amostras é avaliada medindo-se inicialmente as distâncias entre os pares de amostras e colocando num mesmo agrupamento aquelas amostras com menores distâncias entre si. A seguir a distância entre estes pequenos agrupamentos é medida e estabelece-se então novos agrupamentos e assim por diante até que todas as amostras tenham sido enquadradas neste ou naquele grupo (segundo diferentes graus de similaridades). Desta forma, a verificação dos agrupamentos é iniciada a partir de cada amostra individualmente, em direção ao conjunto total, e não ao contrário, do conjunto para o indivíduo. Isto permite observar os diferentes graus de similaridades entre as amostras, desde os mais gerais até aqueles mais específicos.

A similaridade entre as amostras é colocada numa escala, que é construída a partir da equação 10, em que d_{ab} é a distância entre as amostras a e b , e d_{max} é a distância máxima entre todas as amostras consideradas (Moita Neto & Moita, 1998).

$$similaridade_{ab} = 1 - \frac{d_{ab}}{d_{max}} \quad (10)$$

Existem vários métodos para se medir a distância entre os pares das amostras e agrupamentos, entre eles estão a *distância Euclidiana* e a *distância de Mahalanobis* etc.²³ Uma vez medidas as distâncias entre os pares de amostras inicia-se o agrupamento daquelas com uma maior similaridade. Este agrupamento pode ser feito utilizando-se várias técnicas de conexão: *simples*, *completa*, *centróide*, *incremental*, *medianas*, *médias de grupo* ou *flexível*.²⁴

²² Existem várias publicações que tratam em detalhe estes algoritmos. Cf. Wold *et alli* (1987) e Geladi & Kowalski (1986).

²³ Existem vários (e muito bons) livros e artigos que descrevem com mais detalhes o método HCA e cada um destes tipos de conexão. Cf. Beebe *et alli* (1998), Sharaf *et alli* (1986) e Kowalski (1983).

²⁴ Idem.

Algumas aplicações interessantes da análise hierárquica de agrupamentos têm sido publicadas, como o trabalho de Ferreira *et alli* (1999), em que os autores empregam esta técnica na caracterização da costa nordeste do estado de São Paulo, a partir de um estudo das variações sazonais da composição química e de alguns parâmetros físico-químicos das águas marinhas; ou em Moita Neto & Moita (1998) em que a técnica é empregada num estudo da composição de algumas frutas; e ainda em Salau *et alli* (1997), que utiliza HCA para caracterizar uma região de acordo com as diferentes fontes de poluentes existentes.

Modelos de Regressão Multivariada

Os métodos de regressão multivariada têm por objetivo modelar algum dado medido independentemente, com algumas propriedades das amostras. As aplicações mais frequentes dos métodos de regressão são em calibração e previsão multivariada, e estudos das relações estrutura-propriedade ou estrutura-atividade (QSPR e QSAR, respectivamente).

Calibração e Previsão Multivariada

A *calibração* é o processo de construção de um modelo matemático que relacione a resposta de um instrumento a certas propriedades das amostras. Por exemplo, na determinação da concentração de um composto químico em diferentes amostras, muitas vezes o instrumento utilizado para se determinar o teor do composto não fornece uma resposta direta. É necessária a construção de um modelo de regressão que correlacione a resposta instrumental ao teor do analito, ou outra informação qualquer que se deseja obter. Este modelo poderá então ser utilizado para prever a concentração do analito em amostras desconhecidas, processo conhecido como *previsão* (Ferreira *et alli*, 1999).

Quando existe apenas uma resposta instrumental por amostra, por exemplo um valor de absorvância para o analito que está sendo investigado, os métodos de calibração e previsão são ditos *univariados*, isto é, dependem de apenas uma variável

(no caso, a resposta instrumental) que será relacionada em termos matemáticos ao teor de apenas um analito. A calibração univariada tem sido utilizada com sucesso em técnicas onde a lei de *Beer* pode ser aplicada e que não há a presença de interferentes, como por exemplo em determinações de um único analito através de métodos espectroscópicos (Thomas, 1994).

A calibração e a previsão *multivariadas* são processos utilizados quando mais de uma resposta instrumental são utilizadas, como por exemplo, nos métodos espectrais, como análises de infravermelho próximo (Rohrbach, 1991) e fluorescência de raio-X (Yap, 1992). Neste caso, o instrumento fornece mais de uma resposta por amostra e também é possível analisar mais de um analito simultaneamente, o que não seria possível utilizando um método univariado. De uma forma geral, estas correlações podem ser representadas de acordo com o esquema da figura 7 (Ferreira *et alli*, 1999).

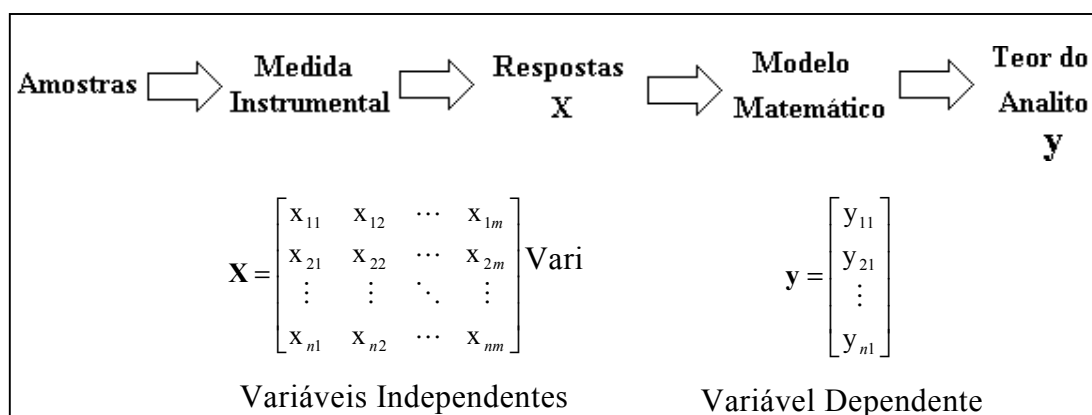


Figura 7: Representação de um processo de calibração multivariada .

Relação Quantitativa Estrutura-Propriedade ou Estrutura-Atividade

Os métodos de QSAR e QSPR utilizam descritores moleculares referentes às características estruturais, eletrônicas ou físico-químicas de um grupo de moléculas homólogas para construir modelos de regressão que correlacionem estes descritores à alguma propriedade, ou atividade biológica, cujo valor já tenha sido determinado experimentalmente. Este modelo pode ser então utilizado para prever as propriedades ou atividades cujos valores são desconhecidos, para compostos similares. De uma forma

geral, estas correlações podem ser representadas de acordo com o esquema da figura 8, em que X é a matriz de variáveis independentes e y o vetor da variável dependente, semelhante ao esquema da figura 7.

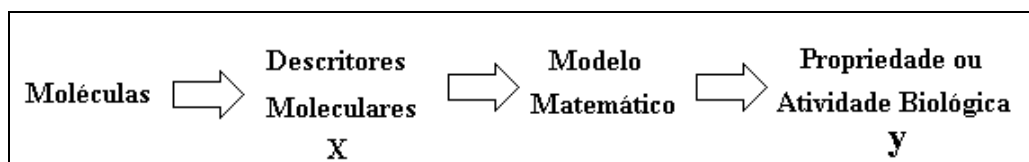


Figura 8: Representação de um estudo de QSAR/QSPR.

Nos dois casos (calibração e QSAR/QSPR), os dados são organizados na forma de matrizes, em que as linhas representam cada amostra, ou molécula, e as colunas representam as propriedades, ou descritores moleculares. Na matriz X são colocadas aquelas propriedades que descrevem a molécula, ou amostra, e que serão utilizadas para construir o modelo de regressão. Estas variáveis são chamadas de *variáveis independentes*.

No vetor y são colocadas as propriedades que se deseja modelar: o teor do analito, ou a propriedade biológica. Estas variáveis são chamadas de *variáveis dependentes*. Em alguns casos, são modeladas várias variáveis dependentes de uma só vez, então ao invés de um vetor, têm-se uma matriz de *variáveis dependentes*, representadas por Y . Para simplificar as demonstrações, utilizaremos o caso em que se trabalha com apenas uma variável dependente, e portanto, os valores para a propriedade de interesse (ou concentração do analito) será organizado na forma de um vetor.

O processo de construção dos modelos de calibração, ou correlação estrutura-propriedade envolve duas etapas principais: a *modelagem* e a *validação do modelo*. Após estas duas etapas serem concluídas, o modelo está apto a ser utilizado na previsão de propriedades ou concentrações desconhecidas para amostras similares.

Etapas de Modelagem

A modelagem é a construção do modelo matemático que melhor descreva a correlação entre a matriz X e o vetor y . Primeiramente, é feita a seleção do conjunto de

amostras a ser utilizado. Isto porque amostras com comportamento ou características muito distintas do restante do conjunto podem interferir de forma negativa e dependendo do caso, devem ser excluídas. Por exemplo: na construção de um modelo para previsão de ponto de ebulição de hidrocarbonetos aromáticos policíclicos, a inclusão do benzeno (única molécula com apenas 1 anel aromático, e portanto sem similares no conjunto) poderia estar forçando o modelo a descrever uma amostra muito diferente do restante do conjunto, num ajuste que poderia introduzir algum erro, ou alguma distorção do modelo.

Em seguida é feito o préprocessamento dos dados, se necessário, similarmente à análise de componentes principais, e a coleta dos dados experimentais (propriedades do vetor y).

Após escolher as amostras que melhor represente os dados que se deseja modelar e efetuar o préprocessamento, a próxima etapa é a construção do modelo matemático, que pode ser encontrado através do *vetor de regressão* β como representado pela equação 11.

$$y = X * \beta \quad (11)$$

A construção deste modelo pode ser feita por vários métodos, e a escolha de cada um deles está relacionada ao perfil do conjunto de dados. Alguns dos métodos mais comuns para a construção de modelos de regressão multivariada, serão citados posteriormente no texto.

Etapa de Validação

Uma vez construído o modelo, é necessário que ele seja testado, ou *validado*, para garantir que os valores das variáveis dependentes estimados sejam iguais, ou muito próximos dos valores obtidos experimentalmente. Isto pode ser feito utilizando um *conjunto de validação externo* ou o *método de validação cruzada*.

a) Conjunto de Validação Externo

Neste caso, um grupo de amostras de propriedades conhecidas (*conjunto de validação*) é separado do restante do conjunto e o modelo é construído com as amostras restantes (*conjunto de modelagem*). Este modelo é então utilizado para prever as propriedades das amostras deste conjunto de validação.

As propriedades previstas são então comparadas com as propriedades obtidas experimentalmente e o *erro de previsão* é calculado (equação 12). A eficiência do modelo, ou seja, sua capacidade preditiva, é avaliada de acordo com a dimensão deste erro. Se o erro for pequeno, o modelo pode ser utilizado com confiança para prever as propriedades desconhecidas de outras moléculas. Se este erro for grande, significa que o modelo não é eficaz.

$$erro = (\hat{y}_i - y_i) \quad (12)$$

Em que

$\hat{y}_i$ é o valor previsto para a amostra i utilizando o modelo;

y_i é o valor medido para a amostra

b) Validação Cruzada

Este método é muito similar ao método de validação externo. Primeiramente se retira um pequeno conjunto de amostras, constrói-se o modelo com as amostras restantes e faz-se a previsão para estas amostras. A seguir, estas amostras são incluídas novamente *no conjunto de modelagem*, outras são retiradas e se repete o processo até que todas as amostras do *conjunto de modelagem* sejam testadas (Ferreira *et alli*, 1999).

A validação cruzada pode ser feita retirando-se uma amostra de cada vez, ou duas amostras de cada vez, ou até mais, dependendo do conjunto de dados que está sendo estudado, da quantidade de amostras, do número de amostras de cada subconjunto etc.

Os valores previstos são então comparados com os valores obtidos experimentalmente e o *erro de previsão* é calculado, como descrito anteriormente pela equação 12. Calcula-se a seguir a soma dos quadrados dos erros de previsão ($PRESS^{25}$), dada pela equação 13; ou a raiz quadrada dos erros de previsão ($RMSEP^{26}$), que nada mais é do que o desvio padrão (equação 14) (Thomas, 1994).

$$PRESS = \sum_i (\hat{y}_i - y_i)^2 \quad (13)$$

$$RMSEP = \sqrt{\frac{\sum_i (\hat{y}_i - y_i)^2}{n}} \quad (14)$$

Em que

$\hat{y}_i$ é o valor previsto para a amostra i utilizando o modelo;

y_i é o valor medido para a amostra i

n = número de amostras do conjunto de calibração

A validação do modelo permite testar sua eficiência, ou seja, se é realmente capaz de prever as propriedades (variáveis dependentes) descritas pelas variáveis independentes, de forma confiável. Uma vez validado o modelo, ele poderá ser utilizado para fazer a previsão de propriedades (ou concentrações) em amostras desconhecidas.

Algumas Considerações Sobre A Escolha do Modelo

Na escolha de um bom modelo de calibração, alguns fatores devem ser levados em consideração: a escolha do número de componentes principais (em PCR) ou variáveis latentes (no PLS) e a remoção de amostras anômalas.

A escolha do número de componentes principais, ou variáveis latentes deve levar em consideração a capacidade preditiva do modelo, e deve ser avaliada pelo seu erro de previsão e pelo coeficiente de correlação. O modelo ótimo deverá ter um pequeno erro de previsão, e um bom coeficiente de correlação (aquele que mais se aproxime de 1).

²⁵ A sigla tem origem na língua inglesa: *prediction error sum of squares*.

²⁶ A sigla tem origem na língua inglesa: *root mean square error prediction*.

Este coeficiente é a medida do quanto o modelo se aproxima de um comportamento linear; quanto mais perto de 1.0, mais linear ele será.

Em trabalhos de química analítica, como a construção de curvas de calibração para se determinar a concentração de um analito, por exemplo, somente serão aceitos coeficientes de correlação acima de 0.99. Em trabalhos de QSAR, no entanto, é muito difícil se obter correlações com coeficientes tão altos. Em geral, coeficientes acima de 0.85 são considerados satisfatórios.

Algumas vezes, é necessário a remoção de certas amostras que apresentem baixo poder de modelagem. Estas amostras geralmente têm um comportamento diferenciado, apresentam características peculiares em relação ao restante do conjunto, e podem distorcer ou contribuir negativamente para a construção do modelo. Um exemplo deste tipo de comportamento já foi citado anteriormente: imagine que se deseja construir um modelo de previsão para determinada propriedade de PCBs e no conjunto de moléculas (que representam as amostras) houvesse um hidrocarboneto aromático policíclico, o naftaleno, por exemplo. Apesar de o naftaleno ter algumas propriedades semelhantes aos PCBs (é um hidrocarboneto aromático policíclico), não possui átomos de cloro na sua estrutura, e pertence a um outro grupo de moléculas diferente deste conjunto. A inclusão do naftaleno acarretaria numa distorção do modelo, levando a um erro de previsão alto e num conseqüente abaixamento do coeficiente de correlação.

É claro que este exemplo é muito óbvio, mas há casos em que o comportamento anômalo de uma substância não é tão facilmente explicado, ou até mesmo detectado, então faz-se necessário o uso de recursos estatísticos para detectar este tipo de amostra, e até mesmo para decidir se ela deve ser removida ou não.

Como dito anteriormente, o gráfico dos escores pode indicar o comportamento diferenciado de uma amostra, mas a influência da amostra para a construção do modelo somente poderá ser avaliada pela dimensão de duas grandezas complementares: os resíduos de *Student* e a *leverage*.

Leverage

A leverage é a medida da influência de uma amostra no modelo de regressão. Quando uma amostra apresenta um valor pequeno de *leverage* significa que a amostra influencia pouco na construção do modelo de calibração, e uma leverage alta significa que a amostra influencia muito. A leverage pode ser geometricamente interpretada como a distância de uma amostra ao centróide do conjunto de dados e pode ser calculada pela equação 15 (Ferreira *et alli*, 1999).

$$h_{ii} = \frac{1}{n} + (\mathbf{x}_i - \bar{\mathbf{x}})^T (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) \quad (15)$$

Em que:

h_{ii} é a *leverage*

n é o número de amostras (ou de moléculas) no conjunto de calibração

$\mathbf{x}_i$ é o espectro da i -ésima amostra, ou, para o caso de QSAR, o conjunto de descritores para a i -molécula

$\bar{\mathbf{x}}$ é o espectro médio, ou, o valor médio para os descritores

$\mathbf{X}^T \mathbf{X}$ é a matriz de variância-covariância

Uma outra forma de calcular a *leverage* é através da distância da amostra ao centro do conjunto através da estimativa da distância Euclidiana no espaço das componentes principais, como demonstra a equação 16.

$$\mathbf{h} = \mathbf{T}^* \mathbf{T}^T \quad (16)$$

Em que:

$\mathbf{h}$ representa o valor da leverage

$\mathbf{T}$ é a matriz dos escores

$\mathbf{T}^T$ é a transposta de $\mathbf{T}$

Mas nem sempre uma grande influência na construção do modelo poderá ser positiva. Amostras anômalas, por exemplo, costumam apresentar valor de leverage alto, pois são capazes de forçar o modelo a descrever características que podem não refletir o perfil do restante do conjunto de dados. Isto não significa que amostras com alto valor de leverage devem ser excluídas. Cada caso é único, e deve ser avaliado cautelosamente.

Resíduo de Student

Uma forma de avaliar se uma amostra com alto valor de leverage deve ser retirada ou não, é avaliar a dimensão dos resíduos dos valores previstos na etapa de validação, para a variável dependente para cada amostra. Isto pode ser feito calculando-se o *resíduo de Student* para cada amostra, segundo a equação 17 (Ferreira *et alli*, 1999).

$$resíduo = \frac{(y - y_i)}{Lresc \sqrt{1 - h_i}} \quad (17)$$

Em que:

y é o valor previsto para a concentração ou propriedade da i -ésima amostra ou molécula

y_i é a concentração, ou o valor da propriedade para a i -ésima amostra ou molécula

h_i é o valor da leverage para a i -ésima amostra

$Lresc$ é o resíduo do valor previsto para a amostra i corrigido pela *leverage*, de acordo com a equação 18.

$$Lresc = \sqrt{\frac{(c - c_i)^2}{(n - 1)(1 - h_i)}} \quad (18)$$

Construindo um gráfico do resíduo *vs.* a leverage é possível detectar amostras anômalas e avaliar se sua retirada é conveniente ou não. Amostras que apresentam alto resíduo e alto valor de leverage devem sempre ser excluídas do tratamento. Para os casos de amostras que apresentam alto resíduo e pequeno valor de leverage, ou o contrário, uma avaliação cautelosa deve ser feita construindo-se o modelo inicialmente com a amostra e depois após a sua exclusão e comparando os resultados obtidos.

Os Métodos de Regressão Multivariada

A construção dos modelos de regressão multivariada pode ser feita por vários métodos, e a escolha de cada um deles está relacionada ao perfil do conjunto de dados. Basicamente, os métodos de regressão multivariados são classificados em dois tipos principais: os métodos dos mínimos quadrados clássicos (*CLS – classical least square*) e os métodos dos mínimos quadrados inverso (*ILS – inverse least square*).

Métodos dos Mínimos Quadrados Clássicos (CLS)

Segundo Beebe *et alli* (1998), “os métodos dos mínimos quadrados clássicos devem ser utilizados quando o sistema é simples e todos os analitos são conhecidos”. A grande restrição deste método é justamente a exigência do conhecimento do valor atribuído a cada variável, ou a concentração de cada espécie presente no sistema, o que é considerado uma grande limitação, pois na maioria das vezes isto é impossível, mas o método tem sido aplicado com sucesso na resolução de sistemas mais simples.

Existem dois tipos de métodos de regressão do tipo CLS: o direto (*DCLS – direct classical least square*) e o indireto (*ICLS – indirect classical least square*). Estes métodos não serão abordados, por não terem sido utilizados neste trabalho.²⁷

²⁷ para uma leitura mais aprofundada Cf. Beebe *et alli* (1998).

Métodos dos Mínimos Quadrados Inverso (ILS)

Caso o sistema seja mais complexo, e nem todos os analitos sejam conhecidos, é recomendável a utilização dos métodos dos mínimos quadrados inverso, que são os métodos de regressão linear múltipla (MLR^{28}), regressão por componentes principais (PCR^{29}) e mínimos quadrados parciais (PLS^{30}).

MLR

A *regressão linear múltipla (MLR)* é utilizada quando o número de variáveis é pequeno, e o objetivo do tratamento é justamente reduzir o número de variáveis do conjunto de dados. Neste caso, é necessária a utilização de uma etapa prévia que utilize algum método de seleção de variáveis.

De uma forma geral este método pode ser entendido como sendo a construção de uma relação matemática linear (ou de primeira ordem) entre as variáveis independentes (ou seja, as colunas da matriz **X**) e a variável dependente (o vetor **y**) para cada amostra (cada linha da matriz de **X**). Para uma amostra apenas, esta relação pode ser melhor entendida pela representação da equação 19. Para um conjunto de *n* amostras isto pode ser ilustrado pela equação 20 (Geladi & Kowalsky, 1986).

$$y = \beta_1 var_1 + \dots + \beta_m var_m + e \quad (19)$$

Em que:

y é a variável dependente para uma amostra

m corresponde ao número de variáveis utilizadas

var = variáveis independentes

β é o coeficiente de regressão para cada *var*

e corresponde ao resíduo

$$y = X * \beta + E \quad (20)$$

²⁸ *multiple linear regression*

²⁹ *principal components regression*

³⁰ *partial least square*

Em que:

y é o vetor das variáveis dependentes

X é a matriz das variáveis independentes

β é o vetor de regressão

E é a matriz de resíduos

Este método não foi utilizado neste trabalho e portanto não será tratado neste texto.³¹

Como mencionado anteriormente, este método é restrito à um conjunto de dados com poucas variáveis, e na maioria das vezes é necessário uma prévia seleção de variáveis. Mas nem sempre a etapa de seleção de variáveis é capaz de reduzir o número de variáveis sem introduzir algum erro, ou sem excluir variáveis relevantes ao modelo. Quando isto ocorre, é necessária a utilização de algum método que suporte um número maior de variáveis, como é o caso do PCR e do PLS. Estes dois métodos são muito eficientes no tratamento de ruídos experimentais, e na resolução de problemas de colinearidade, que é uma restrição do método MLR. Nestes casos, a seleção das variáveis importantes é feita com a utilização dos próprios métodos PCR e PLS. São métodos robustos, isto é, a inclusão de novas amostras não altera tanto o modelo (Geladi & Kowalski, 1986).

PCR

A *regressão por componentes principais (PCR)* utiliza etapas semelhantes às da PCA para decompor a matriz de dados X de variáveis independentes e então relaciona os resultados da PCA com as variáveis independentes do vetor y .

A primeira etapa, que consiste na decomposição da matriz das variáveis independentes utilizando *análise de componentes principais* pode ser representada pelas transformações descritas pelas equações que se seguem. Primeiramente a equação 7³² é

³¹ Cf. Beebe *et alli* (1998) e Geladi & Kowalsky, (1986).

³² Cf. página 15.

reescrita em função da matriz dos escores. Isto é feito multiplicando-se à direita em ambos os lados por $\mathbf{L}$, como na equação 21.

$$\mathbf{X} * \mathbf{L} = \mathbf{T} * \mathbf{L}^T * \mathbf{L} \quad (21)$$

Como $\mathbf{L}$ é uma matriz ortonormal, então:

$$\mathbf{L}^T * \mathbf{L} = \mathbf{I} \quad (22)$$

E portanto a equação 7 poderá finalmente ser escrita em função dos escores:

$$\mathbf{T} = \mathbf{X} * \mathbf{L} \quad (23)$$

Para as equações 21, 22 e 23:

$\mathbf{T}$ é a matriz dos escores

$\mathbf{L}$ é a matriz dos pesos

$\mathbf{L}^T$ é a transposta de $\mathbf{L}$

$\mathbf{I}$ é a matriz identidade

A equação 23 pode ser representada pelo esquema da figura 9.

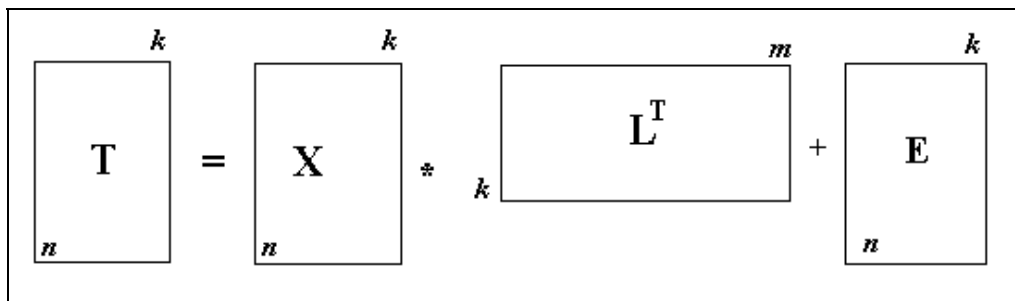


Figura 9: Representação gráfica da equação 23.

Na outra etapa é feita a correlação entre a matriz dos escores transformada, contendo as informações das variáveis independentes, e o vetor das variáveis

dependentes como descrito pelas equações 24 cuja solução pode ser vista na equação 25 e representada graficamente pela figura 10 (Geladi & Kowalski, 1986).

$$\mathbf{y} = \mathbf{T} * \boldsymbol{\beta} + \mathbf{E} \quad (24)$$

$$\boldsymbol{\beta} = (\mathbf{T}^T * \mathbf{T})^{-1} * \mathbf{T}^T * \mathbf{y} \quad (25)$$

Neste caso a inversão de $\mathbf{T}^T * \mathbf{T}$ não irá causar problemas devido à ortogonalidade entre as linhas da matriz dos escores.

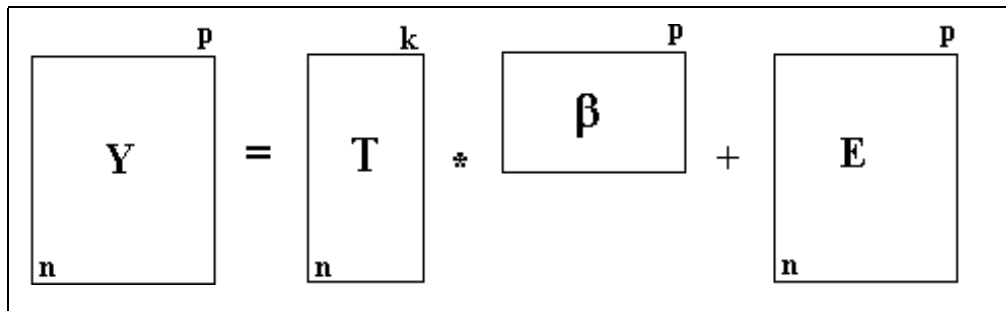


Figura 10: Representação da equação 24.

Segundo Ferreira *et alli* (1999), “o fato da PCR primeiramente decompor a matriz $\mathbf{X}$ e somente depois relacioná-la ao vetor $\mathbf{y}$ pode constituir uma fragilidade do método nos casos em que o analito de interesse, ou a propriedade que se deseja modelar, possuem valores muito pequenos e portanto não influenciam nas primeiras componentes principais, fazendo com que um número maior delas seja necessário para a construção do modelo”.

PLS

Nos casos em que esta limitação da PCR tem que ser contornada, uma alternativa interessante é o método dos *mínimos quadrados parciais (PLS)* que também decompõe as variáveis independentes, mas com uma variação: passa-se informação extraída das variáveis independentes ($\mathbf{X}$) para as variáveis dependentes ($\mathbf{Y}$) e vice-versa. Isto é possível porque as informações contidas nas variáveis dependentes também são

utilizadas na obtenção dos fatores, ou variáveis latentes, que nada mais são do que combinações lineares das componentes principais obtidas pela PCR. No método PLS, tanto a matriz das variáveis independentes (**X**) quanto a das variáveis dependentes (**Y**) é representada pelos seus escores que são calculados utilizando a análise de componentes principais, como demonstra as equações 26 e 27 (Ferreira *et alli*, 1999). Estas transformações são chamadas por Geladi & Kowalski (1986) de *relações externas*.

$$\mathbf{X} = \mathbf{T} * \mathbf{L}^T + \mathbf{E} = \sum_{i=1}^k \mathbf{t}_i \mathbf{l}_i^T + \mathbf{E} \quad (26)$$

X é a matriz de dados originais

T é a matriz dos escores de **X**

L^T é a matriz transposta dos pesos
de **X**

E é a matriz de resíduos de **X**

t são os escores para a *i*-ésima PC

l são os pesos para a *i*-ésima PC

i é cada componente principal

k é o número de componentes
principais utilizado na
modelagem

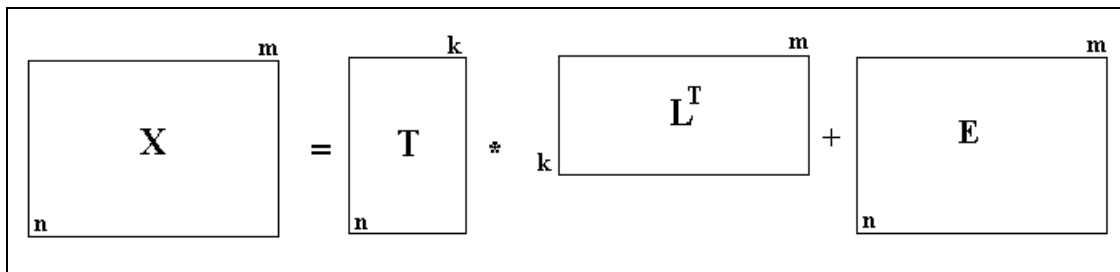


Figura 11: Representação da equação 26.

$$\mathbf{y} = \mathbf{U} * \mathbf{Q}^T + \mathbf{F} = \sum_{i=1}^k \mathbf{u}_i \mathbf{q}_i^T + \mathbf{F} \quad (27)$$

Y é a matriz de dados originais

U é a matriz dos escores de **Y**

Q^T é a matriz transposta dos pesos
de **Y**

F é a matriz de resíduos de **Y**

u são os escores para a *i*-ésima PC.

q^t são os pesos para a *i*-ésima PC.

i é cada componente principal

k é o número de componentes
principais utilizado na modelagem

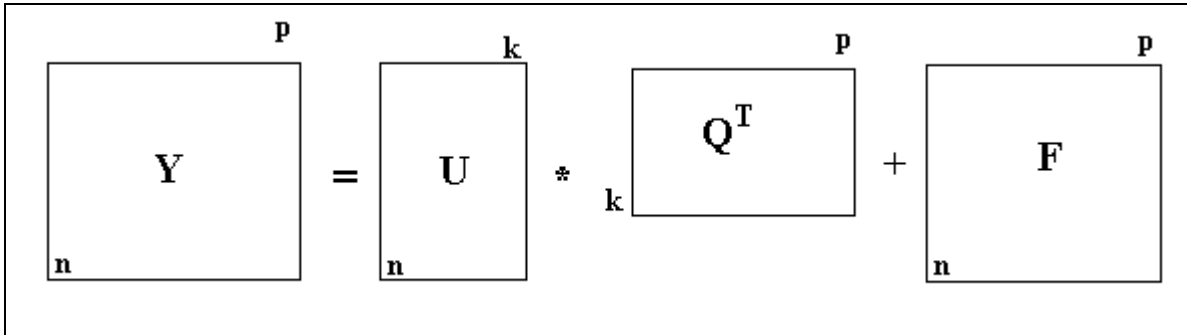


Figura 12: Representação da equação 27.

Estas transformações são feitas para cada componente principal, ou seja, de 1 a k , até que o resíduo E seja igual ou muito próximo ao resíduo F . O intuito do modelo é encontrar a melhor relação entre X e Y e que ao mesmo tempo resulte num valor do resíduo F mais próximo de zero possível.

A correlação entre os escores do bloco Y com os escores do bloco X , chamada por Geladi & Kowalski (1986) de *relações internas*, pode ser feita através da equação 28, para uma componente de cada vez, que fornece o modelo mais simples para esta correlação.

$$u_i = \beta * t_i \quad (28)$$

Em que o coeficiente β pode ser descrito pela equação 29.

$$\beta = \frac{u_i' t_i}{t_i' t_i} \quad (29)$$

Uma das limitações deste modelo é que os escores são calculados separadamente para X e Y , e somente depois é construída a relação entre eles, o que pode resultar num modelo insatisfatório, por exemplo, numa relação não linear entre X e Y . Existe um tratamento alternativo que permite calcular os escores dos dois blocos de forma a produzir a melhor relação linear entre eles. Isto é feito fazendo-se uma rotação dos

eixos das componentes principais. Como os escores são as projeções dos pontos nestes eixos, conseqüentemente seus valores são mudados, de maneira a permitir que a melhor relação linear entre os escores de **X** e de **Y** seja construída (Geladi & Kowalski, 1986).

Ambos os métodos, PCR e PLS, desenvolvem modelos separados para cada variável dependente. O método PLS utiliza o algoritmo NIPALS (Geladi & Kowalski, 1986; Geladi, 1988; Ferreira *et alli*, 1999; Beebe *et alli*, 1998).

Programas Computacionais

Existem no mercado vários pacotes computacionais de estatística que trazem rotinas prontas para a aplicação das técnicas de estatística multivariada. Alguns deles encontram-se listados abaixo:

1. PLS Toolbox, que roda no ambiente computacional do *software* MATLAB (MATLAB[®] for Windows, MathWorks Inc.
2. Pirouette Multivariate Data Analysis for IBM PC Systems, Infometrix, Seattle, WA, 1997.
3. Scilab 2.5 – de domínio público, pode ser encontrado no site <http://www-rocq.inria.fr/scilab>
4. SAS – Statistical Analysis System for Windows, StatSoft, 1997.
5. Unscramble for IBM PC systems, CAMO ASA, Oslo, Norway.

Capítulo 2

Utilização de Análise de Componentes Principais para o Estudo dos Efeitos do Derramamento de Óleo sobre Sedimentos em Prince William Sound, Alasca

O acidente

Em 24 de Março de 1989 o petroleiro *Exxon Valdez* encalhou em *Bligh Reef*, em *Prince William Sound* (PWS), Alasca, espalhando parte de sua carga no mar. Foi o maior acidente deste tipo já ocorrido nos Estados Unidos, e o 34º maior no mundo, mas ficou conhecido como o acidente que maiores estragos ambientais provocou, devido à quantidade de óleo derramado (aproximadamente 41 milhões de litros)³³, pela rapidez com que o óleo se espalhou (figura 13), pela peculiar geografia do local atingido, os milhares de quilômetros de costa contaminados e principalmente pela espetacular biodiversidade da fauna e da flora da região atingida (site do *Exxon Valdez Oil Spill Trustee Council*; Schnoor, 1991).

O petroleiro deixou o terminal *Alyesca Pipeline* (Alasca) com destino a *Long Beach* (Califórnia) carregando 1.264.155 barris³⁴ de óleo bruto proveniente de *Alaska North Slope* (óleo ANS) quando encalhou em *Bligh Reef*, na região noroeste de PWS (figura 13), derramando 257.000 barris³⁵ no mar. O óleo rapidamente se espalhou, atingindo em poucos dias a região sudoeste, de PWS até o Arquipélago de *Kodiak* e a Península do Alasca, contaminando 1.300 milhas (2.093 Km) de linha costeira. Para se ter uma idéia da dimensão da área atingida, a região contém mais de 9000 milhas (14490 Km) só de linha costeira (site do *Exxon Valdez Oil Spill Trustee Council*).

³³ Para uma análise comparativa da quantidade de óleo derramada, o vazamento de óleo na Baía de Guanabara em janeiro de 2000 foi estimado em 1 milhão 292 mil litros (site da Petrobrás).

³⁴ Um barril de petróleo equivale a 158,98 litros (Perry & Green, 1984).

³⁵ Equivalente a 10,8 milhões de galões ou 40857,86 litros (site do *Exxon Valdez Oil Spill Trustee Council*)

Estima-se que o acidente tenha matado cerca de 100.000 a 300.000 pássaros e milhares de animais, afetando a fauna e a flora ao longo da linha costeira de PWS (Schnoor, 1991; Kelso & Kendziorek, 1991).

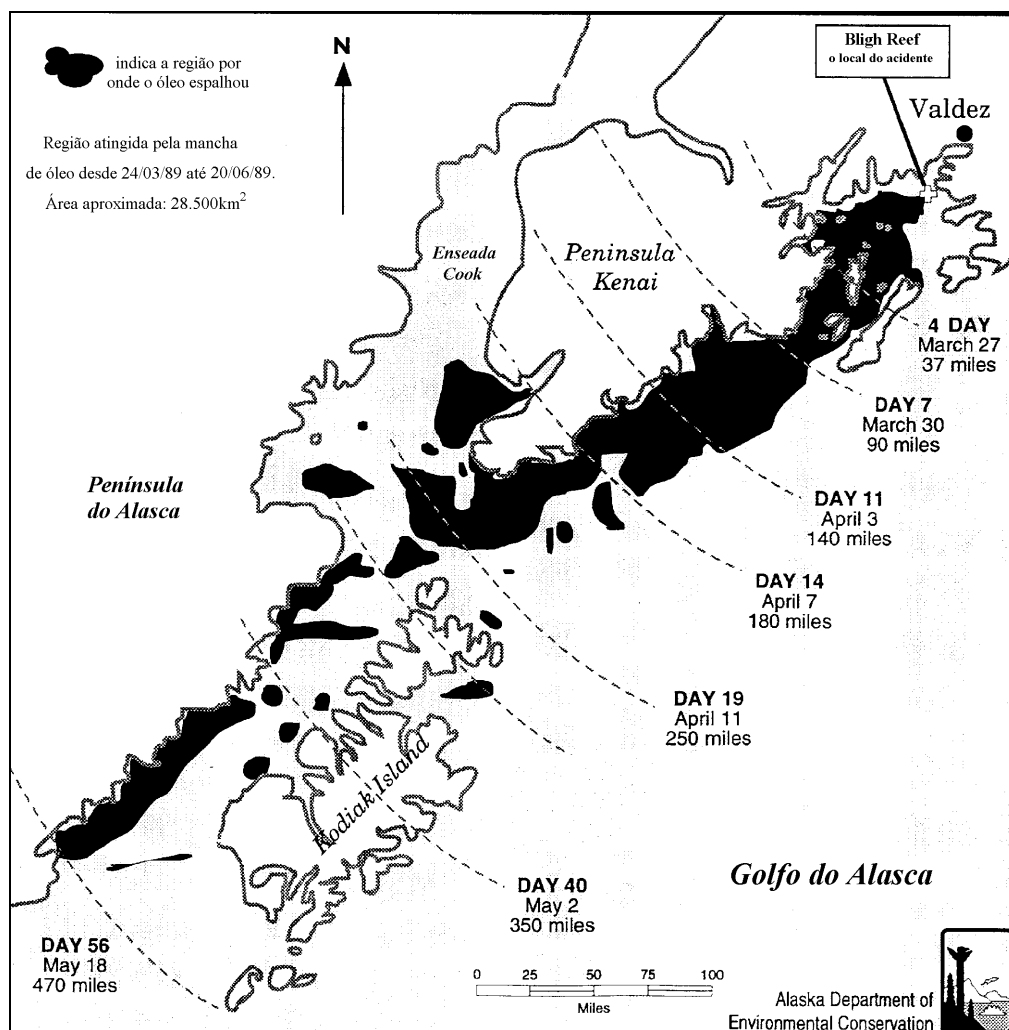


Figura 13: Mapa da região atingida pelo derramamento de óleo do petroleiro Exxon Valdez, mostrando a localização de Prince William Sound (para o português). Cortesia de Exxon Valdez Oil Spill Trustee Council (site oficial).

Após 10 anos do acidente, algumas espécies ainda sofrem as conseqüências da contaminação pelo óleo, enquanto outras já se recuperaram completamente. Das 23 espécies atingidas, somente 2 têm sido consideradas totalmente recuperadas (*site do Exxon Valdez Oil Spill Trustee Council*).

Embora não se tenha registro de nenhuma morte humana ocorrida durante o desastre, quatro casos de mortes associadas às operações de limpeza da área atingida foram reportadas (Final Report, 1990).

O perfil geográfico da região

Para entender a dimensão do estrago ambiental causado por este acidente é necessário conhecer um pouco sobre os aspectos geográficos, sociais, e a biodiversidade da região atingida.

Prince William Sound (PWS), como pode ser visto na figura 14 localiza-se no Golfo do Alasca e possui uma complexa estrutura geográfica formada por uma grande ilha principal e várias ilhas menores, com uma vasta linha costeira toda recortada, cheia de passagens e baías irregulares, formando várias áreas propícias à deposição de toda matéria que circula pelas águas que banham a região, trazidas pelas correntes marítimas locais, principalmente pela Corrente Costeira do Alasca. A temperatura das águas superficiais da região varia de 0 a 16°C, resultando numa taxa de reprodução da fauna e da flora muito baixa, o que torna o ecossistema da região extremamente delicado (Boehm *et alli*, 1998; Schnoor, 1991; Short & Heintz, 1997).

Perto de *Knight Island*, a maior parte das correntes marítimas locais caminham numa velocidade de 15 a 25 Km/dia obedecendo a um padrão cíclico anti-horário ao redor de *Montague Island* (Galt *et alli*, 1991). As velocidades e direções deste fluxo contribuíram para contaminar a região leste de *Knight Island*, onde está localizada *Bay of Isles*, uma baía que recebeu diretamente o impacto do derramamento do óleo *Exxon Valdez*, e que foi sítio de vários estudos sobre o acidente.

Do outro lado de PWS, está localizada *Drier Bay*, que por causa da sua posição geográfica não foi tão contaminada quanto a costa leste e por isto foi utilizada em vários trabalhos como baía de referência (Burns *et alli*, 1997).

As comunidades existentes na área atingida pelo derramamento tinham uma economia baseada na agricultura, pesca e caça de subsistência, além de atividades de recreação e turismo. O acidente atingiu drasticamente a vida destas comunidades não só

pelos seus efeitos imediatos mas também pela dúvida sobre os efeitos do derramamento ao longo do tempo, sobre a possibilidade do desastre ter afetado a saúde e a produtividade do delicado ecossistema da região (Kelso & Kendziorek, 1991).

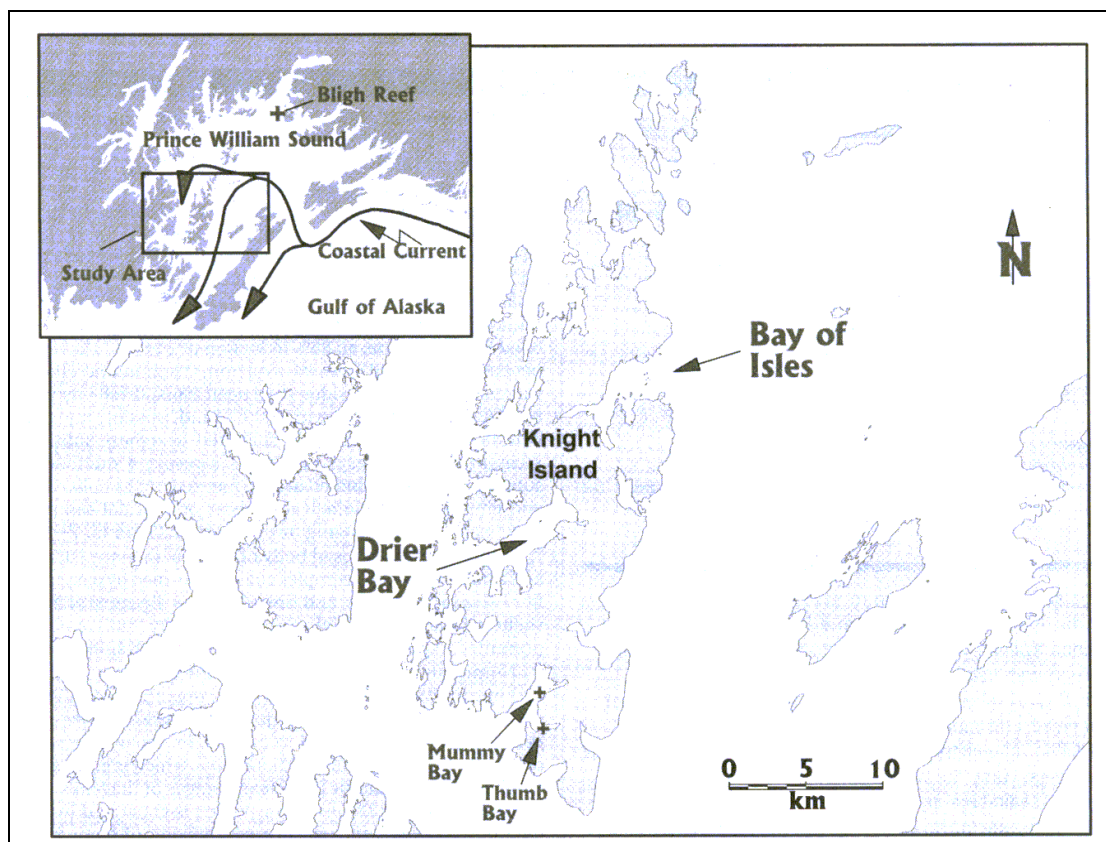


Figura 14: Prince Willian Sound, mostrando Knight Island no detalhe, a localização de Bay of Isles (BOI) e Drier Bay (DB), e a Corrente Costeira do Alasca, extraído do artigo de Boehm et alli (1998).

Os Trabalhos de Restauração

A urgência dos trabalhos de reparação da área afetada mobilizou e promoveu a completa cooperação entre agências estaduais, federais e indústria, e cerca de 2.1 bilhões de dólares foram gastos na recuperação da área (Maki, 1991).

Estiveram envolvidos nos esforços de limpeza cerca de 10.000 pessoas, 1.000 barcos e 100 aviões e helicópteros, que ficaram conhecidos como o Exército, a Marinha e a Força Aérea da *Exxon* (*site* oficial de *Exxon Valdez Oil Spill Trustee Council*).

No início, os esforços de restauração se concentraram na contenção imediata e remoção do óleo da água do mar, por meio da utilização de barragens e dispersantes químicos, com o intuito de diminuir os prejuízos ao ecossistema da região minimizando o contato da fauna e da flora com o óleo. Em seguida foram iniciados os trabalhos de remoção do óleo da linha costeira, utilizando uma variedade de técnicas desde pás e materiais absorventes até jatos de água quente a alta pressão. Cessados os esforços de recuperação imediata, foi iniciado um tratamento a longo prazo da área atingida, pela utilização de técnicas de biorremediação (Kelso & Kendziorek, 1991; Westermeyer, 1991; Pritchard & Costa, 1991; Coffin *et alli*, 1997).

O local atingido transformou-se num imenso laboratório natural para inúmeras pesquisas que se seguiram sobre o impacto de derramamento de óleo e tornou-se referência para trabalhos de prevenção para este tipo de acidente (Schnoor, 1991).

Características do óleo ANS (Alaska North Slope) ou Exxon Valdez Oil (EVO)

“O petróleo é uma complexa mistura de milhares de compostos orgânicos formados a partir de uma variedade imensa de matéria orgânica que foi transformada ao longo de milhões de anos, sob diferentes condições geológicas. O óleo cru contém primariamente carbono e hidrogênio, formando uma infinidade de hidrocarbonetos desde os mais leves até os mais pesados, mas também pode conter pequenas quantidades de enxofre, oxigênio, nitrogênio e alguns metais, como níquel, vanádio e ferro. A natureza variável destes fatores resultam em diferenças químicas bem distintas entre os óleos. Os produtos refinados são frações derivadas da destilação do óleo cru. Por causa destas diferenças na composição do óleo cru, os produtos refinados também diferem em suas composições químicas. Então, todo óleo cru e derivados têm composições químicas que diferem de um para o outro, e esta variabilidade na composição química resulta num perfil único para cada óleo e providencia a base para a identificação de fontes de contaminação de óleo derramado” (Wang *et alli*, 1999).

O óleo que o *Exxon Valdez* carregava foi produzido nos depósitos de petróleo localizados em *Alaska North Slope*, sendo originário de um período geológico recente. Este tipo de óleo contém uma mistura de compostos alifáticos dominada por uma série homóloga de *n*-alcanos que diminuem em concentração relativa à medida que se aumenta o peso molecular e varia em número de átomos de carbono de *n*-C11 até *n*-C40 (Hostettler & Kvenvolden, 1994).

O perfil de hidrocarbonetos aromáticos policíclicos (PAH) deste óleo contém principalmente naftaleno e fenantreno e seus derivados alquilados, sendo que a fração de naftalenos predomina (Hostettler & Kvenvolden, 1994).

As fontes de hidrocarbonetos da região

Como dito anteriormente, a composição de derivados de petróleo varia quando exposta aos efeitos do intemperismo, e ao longo do tempo. Os compostos mais voláteis tendem a evaporar rapidamente, aqueles menos voláteis tendem a se depositar. Dependendo das características moleculares (tamanho, peso, aromaticidade etc.), os hidrocarbonetos podem ser mais ou menos solúveis, ou se acumular no sedimento, ou serem mais suscetíveis à fotodegradação, ou à ação microbiológica. De um modo geral, os processos de degradação por intemperismo do óleo derramado são: espalhamento, evaporação, dissolução, dispersão e formação de emulsão.

Os PAHs são muito mais resistentes à degradação por intemperismo do que os hidrocarbonetos alifáticos, e por isto a distribuição de PAHs em uma região é muito utilizada para se avaliar o impacto da descarga de derivados de petróleo no meio ambiente (Boehm *et alli*, 1998). No entanto, existem inúmeras outras fontes de contaminação de PAHs no meio ambiente, como descargas urbanas e industriais, deposição atmosférica e processos biológicos, e estimar por meio de análises químicas qual a contribuição destas fontes tem sido um dos grandes desafios da química ambiental (Short *et alli*, 1999).

Existem inúmeras fontes de PAHs em *PWS*, de origem tanto natural quanto antropogênicas, sendo que os hidrocarbonetos identificados podem ser agrupados em três tipos: biogênicos, petrogênicos e pirogênicos. Os PAHs biogênicos são aqueles

gerados por processos biológicos ou nos estágios iniciais da diagênese em sedimentos marinhos recentes. As fontes biológicas são as plantas terrestres, o fitoplâncton, animais, bactérias, macroalgas e microalgas (Wang *et alli*, 1999); os PAHs petrogênicos são aqueles relacionados ao petróleo e seus derivados; e os PAHs pirogênicos são aqueles gerados nos processos a altas temperaturas, ou pela combustão de combustíveis fósseis (carvão ou óleo) e materiais orgânicos recentes como a madeira (Page *et alli*, 1996).

Dois tipos de pirogênicos têm sido documentados no sedimento bêntico de *PWS*. Um deles está presente em pequenas quantidades e possui uma distribuição de hidrocarbonetos mais uniforme ao longo de todo o sedimento, e é atribuído à contaminação por deposição atmosférica; e um outro em maior concentração, localizado nas áreas com um histórico de atividade humana, e que podem ser originários da queima de combustíveis como madeira, carvão e óleo (Wang *et alli*, 1999).

As maiores fontes antropogênicas de pirogênicos encontram-se localizadas em Drier Bay, no lado oeste de *PWS*, onde até a década de 40 existiam fábricas de conserva e alguns vilarejos com intensa atividade de pesca e mineração. Hoje em dia estas instalações estão em ruínas, mas os resíduos da queima de carvão e óleo remanescentes das antigas máquinas escoaram para o mar, introduzindo considerável carga de pirogênicos nos sedimentos (Burns *et alli*, 1997; Boehm *et alli*, 1998).

Numerosos depósitos naturais de óleo e carvão têm sido identificados ao longo da costa do Golfo do Alasca. Os produtos destes depósitos aderem aos sedimentos terrestres e marinhos da região, que são transportados pela Corrente Costeira do Alasca até *PWS*.

Estas são as maiores contribuições para o perfil de PAHs petrogênicos e pirogênicos nos sedimentos marinhos do Alasca (Short *et alli*, 1999; Page *et alli*, 1996). O óleo naturalmente presente no sedimento das áreas costeiras do Golfo do Alasca e trazidos até *Prince William Sound* pela Corrente Costeira do Alasca apresenta baixos teores de dibenzotiofeno alquilado.

A maior fonte antropogênica de PAHs é o óleo *Exxon Valdez* proveniente do derramamento, que possui um alto teor deste hidrocarboneto (Short & Heintz, 1997; Short *et alli*, 1999; Boehm *et alli*, 1998).

Apresentação dos dados

Os dados deste trabalho foram retirados de um estudo de Boehm *et alli* (1998) em que os autores utilizam análise de variância³⁶ para identificar as diferenças na distribuição de PAHs nas duas baías citadas: *Bay of Isles* (BOI), que por causa da sua posição geográfica foi mais atingida pelo derramamento de óleo do petroleiro *Exxon Valdez*, e *Drier Bay* (DB), a baía localizada do outro lado da ilha, que recebeu pouco impacto do derramamento e por isto foi utilizada como baía de referência.

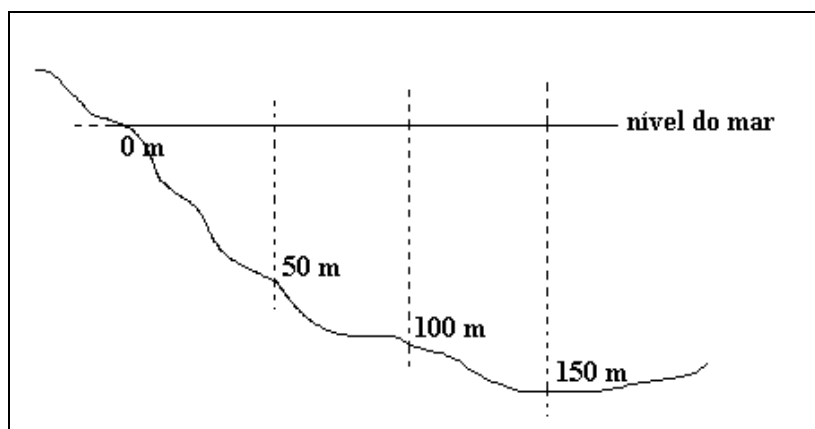


Figura 15: Representação da três regiões em que a amostragem foi realizada.

Os autores coletaram e analisaram amostras de sedimento marinho ao longo de três profundidades das duas baías: 0-50 m, 50-100 m e 100-150 m, como representado na figura 15. Cada amostra corresponde à um ponto de coleta. Eles detectaram o teor em ppb de 38 diferentes PAHs neste sedimento, que foram agrupados em 10 tipos principais, utilizados no tratamento apresentado neste trabalho como sendo as variáveis detectadas nas amostras.

- C14N – soma dos teores de todos os isômeros do naftaleno contendo um grupo alquil nas posições C1, C2, C3 e C4.
- C14P – soma dos teores de todos os isômeros do fenantreno contendo um grupo alquil nas posições C1, C2, C3 e C4.

- C13D – soma dos teores de todos os isômeros do dibenzotiofeno com um grupo alquil nas posições C1, C2 e C3.
- C14C – soma dos teores de todos os isômeros do criseno contendo um grupo alquil nas posições C1, C2, C3 e C4.
- Per – teor de perileno
- Pyro – soma dos teores de todos os analitos pirogênicos, de 3, 4 e 5 anéis
- TPAH – soma dos teores de todos os PAHs analisados
- Petro - teor total de PAHs de origem petrogênica, calculado como total de PAHs detectado menos os teores de perileno e pirogênicos.
- D2/P2 – razão entre os teores C2 – dibenzotiofeno/C2 - fenantreno
- C2P – soma dos teores de todos os isômeros do fenantreno com contendo um grupo alquil na posição C2.

Os dados foram dispostos em uma matriz $\mathbf{X} = (n, m)$, com $n = 82$ linhas (*i. e.*, número de pontos de coleta) e $m = 10$ colunas (*i. e.*, as variáveis correspondem aos teores dos PAHs em cada ponto), como pode ser observado no quadro 2.

Preprocessamento dos Dados

Ainda no quadro 2, pode-se notar que os teores de PAH no sedimento seco oscilam entre 0 e 400 ppb, exceto para os teores de PAHs de origem pirogênica (Pyro) e petrogênica (Petro), e total de PAHs (TPAH), sendo que para algumas amostras estes valores são extremamente altos, da ordem de 1200 ppb. Devido à discrepância entre estes valores optou-se pelo preprocessamento autoescalado, que coloca os valores numa escala equivalente, minimizando o risco de perder informações por causa da variação da grandeza dos resultados.

³⁶ Cf. Kowalski, 1983.

Quadro 2: Teores de cada grupo de PAH nas profundidades amostradas, em ppb

	<u>Estação</u>	<u>C14N</u>	<u>C14P</u>	<u>C13D</u>	<u>C14C</u>	<u>Per</u>	<u>Pyro</u>	<u>TPAH</u>	<u>Petro</u>	<u>D2/P2</u>	<u>C2P</u>	<u>Baía</u>	<u>Profundidade</u>
1	F011	47	38	0	7	480	52	673	140	0	22	BOI	1
2	F021	32	53	32	29	57	38	273	178	0,67	18	BOI	1
3	F031	31	38	12	17	29	91	248	128	0,46	11	BOI	1
4	R011	18	24	10	23	24	20	144	100	0,39	8	BOI	1
5	R021	52	65	25	53	72	52	381	257	0,44	22	BOI	1
6	R031	16	32	13	28	29	26	171	117	0,44	11	BOI	1
7	R041	35	72	43	76	46	38	346	262	0,73	22	BOI	1
8	R051	25	31	11	21	16	24	151	111	0,37	10	BOI	1
9	R061	56	42	12	21	20	30	237	187	0,33	15	BOI	1
10	R071	19	25	7	20	28	22	141	91	0,31	9	BOI	1
11	R081	23	29	11	15	16	21	135	98	0,47	9	BOI	1
12	R091	8	10	7	10	8	7	54	39	0,8	3	BOI	1
13	R101	0	0	0	0	5	2	9	3	-	0	BOI	1
14	R111	9	18	9	8	7	20	83	56	0,62	6	BOI	1
15	R121	4	7	3	4	1	4	28	23	0,53	2	BOI	1
16	R131	43	74	38	37	77	50	394	267	0,5	28	BOI	1
17	R141	51	102	54	96	98	86	568	385	0,68	28	BOI	1
18	R151	3	4	1	2	1	4	20	15	0,34	2	BOI	1
19	F042	177	170	46	81	50	121	802	631	0,28	58	BOI	2
20	F052	128	119	41	58	59	91	605	455	0,37	38	BOI	2
21	R012	50	63	17	46	20	63	312	230	0,3	20	BOI	2
22	R022	118	106	24	65	22	79	511	409	0,26	35	BOI	2
23	R032	101	106	30	53	26	70	469	374	0,31	35	BOI	2
24	R042	89	87	28	62	30	71	41	339	0,32	29	BOI	2
25	R052	37	50	19	33	15	32	226	179	0,36	16	BOI	2
26	R062	89	117	33	75	100	93	614	421	0,31	39	BOI	2
27	R072	72	72	22	45	41	53	386	292	0,33	24	BOI	2
28	R082	80	82	25	48	70	61	448	318	0,33	30	BOI	2
29	R092	60	53	12	25	9	39	243	195	0,27	18	BOI	2
30	R102	161	172	34	105	62	124	802	615	0,22	55	BOI	2
31	R013	151	173	43	90	55	115	770	600	0,28	58	BOI	3
32	R023	171	159	28	88	45	119	750	585	0,21	53	BOI	3
33	R033	132	120	28	53	28	82	546	436	0,24	41	BOI	3
34	R043	134	135	34	66	26	85	599	487	0,26	47	BOI	3
35	R053	94	80	19	45	13	55	378	310	0,24	29	BOI	3
36	R063	179	175	38	93	44	120	803	639	0,24	58	BOI	3
37	R073	147	130	30	68	14	84	587	489	0,27	41	BOI	3
38	R083	67	63	17	48	19	43	330	267	0,27	20	BOI	3
39	R093	202	175	41	83	49	115	828	664	0,26	58	BOI	3
40	R103	159	141	29	61	25	99	634	510	0,23	47	BOI	3
41	F021	22	29	5	8	85	73	249	92	0,25	10	DB	1
42	F041	28	52	7	22	41	293	510	176	0,19	16	DB	1
43	F051	10	7	0	3	12	6	43	25	0	4	DB	1
44	F061	31	32	5	8	74	29	205	102	0,23	11	DB	1
45	F071	48	112	10	82	120	843	1364	401	0,13	35	DB	1

	<u>Estação</u>	<u>C14N</u>	<u>C14P</u>	<u>C13D</u>	<u>C14C</u>	<u>Per</u>	<u>Pyro</u>	<u>TPAH</u>	<u>Petro</u>	<u>D2/P2</u>	<u>C2P</u>	<u>Baía</u>	<u>Profundidade</u>
46	R011	57	50	3	21	24	124	340	192	0	20	DB	1
47	R021	32	31	3	9	25	33	161	102	0,17	11	DB	1
48	R031	46	126	17	70	130	939	1497	429	0,2	35	DB	1
49	R041	40	36	2	10	16	27	159	116	0	13	DB	1
50	R051	16	17	2	5	19	39	114	56	0,17	6	DB	1
51	R061	11	15	1	8	23	42	112	47	0	4	DB	1
52	R071	27	35	4	16	39	82	231	110	0,15	13	DB	1
53	R081	25	53	3	15	27	477	656	152	0,09	13	DB	1
54	R091	30	46	3	25	66	289	514	160	0,12	15	DB	1
55	R101	11	12	0	4	10	14	58	34	0	5	DB	1
56	R111	14	18	0	6	21	45	115	49	0	6	DB	1
57	R121	17	17	3	5	21	49	129	60	0,23	6	DB	1
58	R131	26	32	4	8	34	53	182	94	0,18	10	DB	1
59	R141	29	24	0	3	31	23	123	69	0	10	DB	1
60	R151	51	49	8	17	15	32	218	171	0,29	18	DB	1
61	R012	37	33	6	10	25	40	187	122	0,2	11	DB	2
62	R022	63	59	8	18	38	66	308	204	0,18	19	DB	2
63	R032	77	89	10	27	140	222	644	282	0,13	32	DB	2
64	R042	46	60	7	20	53	64	294	177	0	21	DB	2
65	R052	116	101	10	17	28	89	453	336	0,15	35	DB	2
66	R062	47	49	8	19	43	74	288	171	0,2	17	DB	2
67	R072	124	97	15	23	23	87	466	355	0,17	35	DB	2
68	R082	89	97	15	35	54	112	495	329	0,18	32	DB	2
69	R092	26	27	2	9	21	52	160	87	0	9	DB	2
70	R102	85	70	9	15	28	77	349	244	0,18	25	DB	2
71	F013	91	52	8	4	9	44	264	210	0,18	20	DB	3
72	F033	115	83	3	13	23	61	375	291	0	28	DB	3
73	R013	171	170	25	65	72	166	815	577	0,18	55	DB	3
74	R023	152	156	20	52	100	177	800	524	0,18	55	DB	3
75	R033	69	66	11	18	29	56	304	219	0,18	22	DB	3
76	R043	160	134	16	24	62	134	647	451	0,15	43	DB	3
77	R053	164	147	24	28	92	148	736	496	0,18	52	DB	3
78	R063	116	73	4	8	18	67	357	272	0	28	DB	3
79	R073	112	102	13	30	36	119	516	361	0,19	35	DB	3
80	R083	143	132	19	32	73	124	645	448	0,19	44	DB	3
81	R093	34	30	3	9	10	20	136	106	0,17	12	DB	3
82	R103	96	93	14	31	74	94	482	314	0,19	32	DB	3

- BOI significa *Bay of Isles* e DB significa *Drier Bay*.
- profundidade 1: 0-50m; profundidade 2: 50-100m; profundidade 3: 100-150m.
- no nome dos pontos de amostragem, F significa *pontos fixos* e R significa *pontos aleatórios*.
- Dados extraídos de Boehm *et alli*, 1998.

Tipos de Tratamento

O tratamento dos dados foi feito de dois modos diferentes:

Tratamento 1

A matriz contendo os dados das duas baías foi dividida em duas classes, referentes às duas baías estudadas: *Bay of Isles* (BOI) e *Drier Bay* (DB). O objetivo foi verificar a discriminação dos pontos de amostragem em dois grupos distintos, correspondentes às duas baías, à partir dos teores de PAHs.

Tratamento 2

A matriz foi dividida em dois subconjuntos de dados, um para a baía oleada e outro para a baía de referência. Cada um destes subconjuntos possui uma divisão em três classes correspondentes às três profundidades de amostragem. O objetivo foi observar e comparar a discriminação das três profundidades de cada uma das duas baías, a partir dos teores de PAHs, para se avaliar a contaminação do ambiente marinho. Estes dois subconjuntos foram denominados tratamento 2A (para BOI) e tratamento 2B (para DB).

Resultados

Análise Preliminar

Inicialmente os teores de PAHs de cada amostra foram checados, no intuito de detectar amostras que apresentassem comportamento muito diferente do restante do conjunto, como por exemplo valor muito alto, ou muito baixo em alguma variável. Isto pode ser resultado de uma série de fatores, desde um resultado analítico errado ou

valores faltantes, até mesmo um ponto de amostragem que possua um perfil diferente do restante, como por exemplo uma fonte de determinado tipo de composto.

Para procurar amostras com comportamento anômalo, primeiramente foram observados os teores de PAHs constantes no quadro 2, e depois foram construídos gráficos de variável *versus* resposta, amostras *versus* resposta e variável *versus* variável. Dependendo do quanto algumas amostras diferem do restante do conjunto, elas foram excluídas do tratamento. A amostra 13 (R101), pertencente à profundidade intermediária de *Bay of Isles*, apresenta um valor faltante: o teor de D2/P2, então foi excluída.

Observando os gráficos de variável *versus* variável nas figuras 16, 17 e 18, foram detectadas 3 amostras com comportamento distinto: as amostras 1 (F011), 45 (F071) e 48 (R031).

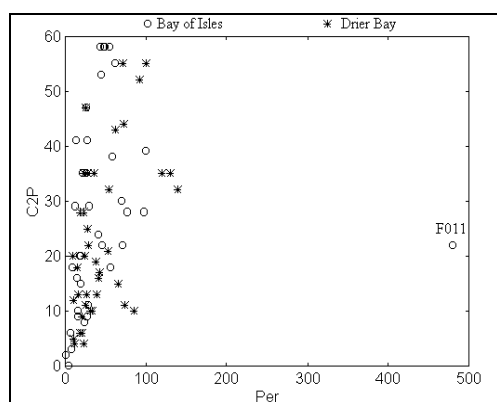


Figura 16: Gráfico de teor de perileno versus C2P mostrando o comportamento da amostra 1 (F011).

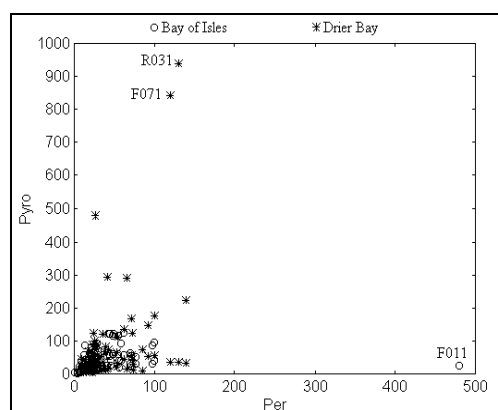


Figura 17: Gráfico de teor de perileno versus pirogênicos mostrando o comportamento diferenciado das amostras 45 (F071) e 48 (R031).

A amostra 1 (F011), pertencente à região mais superficial de *Bay of Isles*, possui alto valor na variável Per (480 ppb), se comparado aos valores do restante do conjunto de dados (1-140 ppb). Este comportamento pode ser observado no quadro 2 apresentado anteriormente e nas figuras 16 e 17.

As amostras 45 (F071) e 48 (R031), pertencentes à área mais superficial de Drier Bay e apresentam alto teor de Pyro (843 e 939 ppb) e de TPAH (1364 e 1497 ppb) em

relação ao restante do conjunto de dados, como pode ser observado no quadro 2 apresentado anteriormente, e nas figuras 17 e 18.

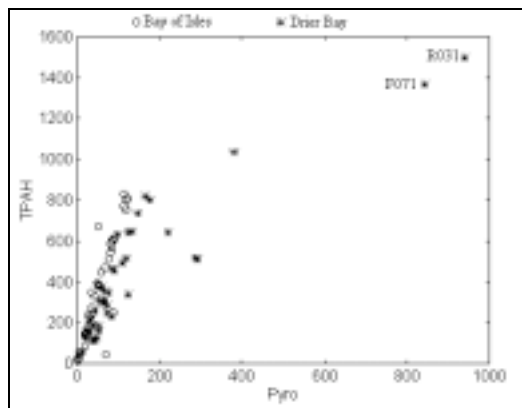


Figura 18: Gráfico de teor de pirogênicos versus TPAH, mostrando o comportamento diferenciado das amostras 45 (F071) e 48 (R031).

As análises foram feitas inicialmente excluindo-se apenas a amostra 13, e depois excluindo-se as amostras 1 (F011), 45 (F071) e 48 (R031), para comparar quais apresentaram os melhores resultados.

Análise dos Dados

Durante a análise dos dados, dois itens são levados em consideração: o número de componentes principais (ou fatores) que são necessários para descrever o sistema, e quanto da variância destes dados cada uma destas componentes é capaz de descrever.

A variância descrita pelas componentes principais indica a quantidade de informação que aquele fator contém. Um dos objetivos de uma análise de componentes principais é justamente descrever o conjunto de dados com o menor número de fatores (ou componentes principais) possível.

Uma forma de encontrar as condições ótimas que obedecem a estes critérios, é selecionando um conjunto de amostras e de variáveis que melhor descreva o conjunto de dados, tendo sempre em mente que amostras com um comportamento anômalo muitas vezes podem estar relacionadas a algum fenômeno, e contribuem para sua

explicação. Estas amostras não devem ser simplesmente excluídas do conjunto de dados e ter seu comportamento ignorado.

Tratamento 1: Análise de Componentes Principais das Duas Baías

O objetivo deste tratamento foi verificar a distribuição de hidrocarbonetos policíclicos aromáticos ao longo de duas baías: *Bay of Isles*, a baía que recebeu o maior impacto do derramamento; e *Drier Bay*, a baía que por causa da sua localização geográfica mais protegida não ficou tão exposta aos efeitos do derramamento.

Detecção de Amostras Anômalas

Inicialmente foi feita uma análise exploratória excluindo a amostra 13 (R101), como descrito anteriormente, e mantendo as amostras 1 (F011), 45 (F071) e 48 (R031).

Quadro 3: Porcentagem de variância descrita pelas componentes principais no tratamento 1, retirando-se apenas a amostra 13 (R101)

	Variância Descrita	Variância (%)	Variância Acumulada (%)
PC1	484.88	60.61	60.61
PC2	137.78	17.22	77.83
PC3	90.03	11.25	89.09
PC4	61.83	7.73	96.81

A primeira componente principal descreveu 60,61% da variância total dos dados. Pelos valores da porcentagem de variância descrita pelas componentes principais, escolheu-se 4 fatores como número ótimo para a descrição do sistema. No quadro 3 e na figura 19 podem ser observadas as variâncias descritas pelas CPs.

Pelo gráfico dos escores apresentados na figura 20 pode-se observar a perfeita distinção entre as duas classes de amostras, que ocorre ao longo da segunda componente principal. Pode-se notar que as amostras 1 (F011), 45 (F071) e 48 (R031) aparecem separadas do restante do conjunto. Esta separação ocorre por causa dos valores de algumas variáveis, como descrito anteriormente.

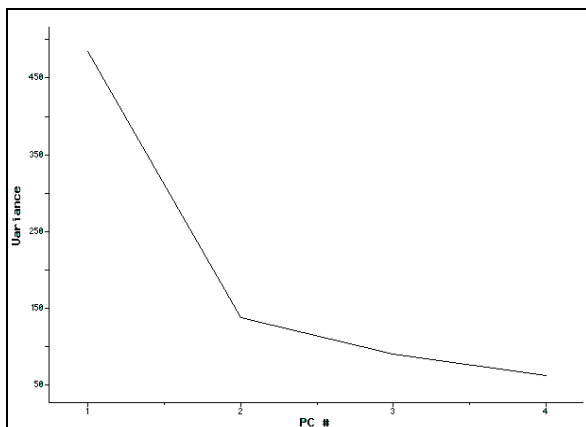


Figura 19: Gráfico da variância descrita pelas componentes principais para o tratamento 1, excluindo a amostra 13.

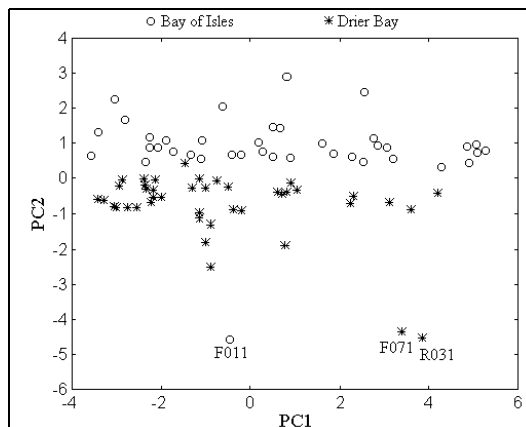


Figura 20: Gráfico dos escores mostrando-se o comportamento anômalo das amostras 1 (F011), 45 (F071) e 48 (R031).

Uma nova análise foi feita, retirando-se somente a amostra 1 (F011). Após refazer a análise, a variância descrita pela primeira componente principal aumentou de 60,61% para 63,34%. Isto significa que o mesmo fator descreve uma quantidade maior de informação. O número de componente principais diminuiu para 3 e as porcentagens de variância que cada uma destas componentes descreve podem ser observadas no quadro 4 e no gráfico da figura 21.

Quadro 4: Porcentagem de variância descrita pelas componentes principais no tratamento 1 após a retirada da amostra 1

	Variância Descrita	Variância (%)	Variância Acumulada (%)
PC1	500.41	63.34	63.34
PC2	132.73	16.8	80.14
PC3	98.89	12.52	92.66

A discriminação entre as duas classes pode ser observada no gráfico dos escores, na figura 22. Ainda assim, duas amostras aparecem separadas do restante do conjunto: as amostras 45 (F071) e 48 (R031), como era esperado por causa dos teores de algumas variáveis, diferente do restante do conjunto.

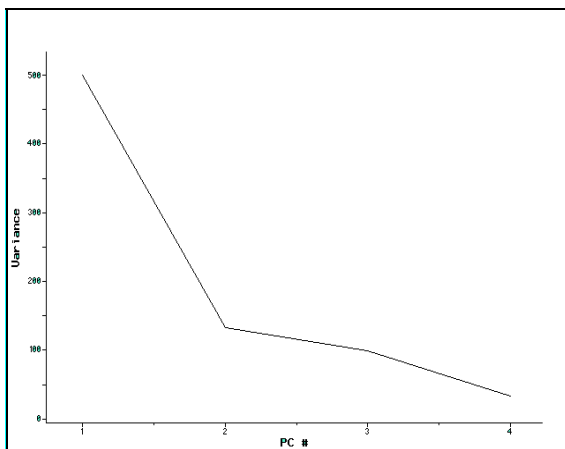


Figura 21: Gráfico da variância descrita em cada componente principal para o tratamento 1, excluindo a amostra 1 (F011) e 13 (R101).

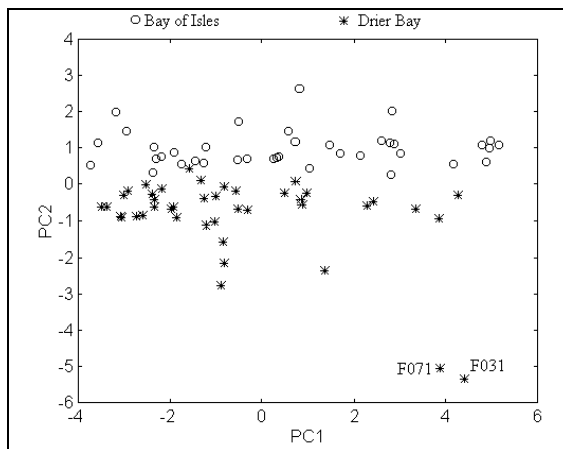


Figura 22: Gráfico do escores mostrando o comportamento das amostras 45 (F071) e 48 (R031).

Uma nova análise foi feita, desta vez excluindo-se as amostras 45 (F071) e 48 (R031), e a variância descrita pela primeira componente principal aumentou de 63,34% para 65.82%. Isto significa que o mesmo fator descreve uma quantidade maior de informação.

Quadro 5: Porcentagem de variância descrita pelas componentes principais no tratamento 1 após a retirada das amostras 1, 45 e 48

	Variância Descrita	Variância (%)	Variância Acumulada (%)
PC1	506.82	65.82	65.82
PC2	122.73	15.94	81.76
PC3	75.38	9.79	91.55

O número de componentes principais manteve-se em 3 e a variância que cada uma destas componentes descreve podem ser observadas no quadro 5 e no gráfico da figura 23.

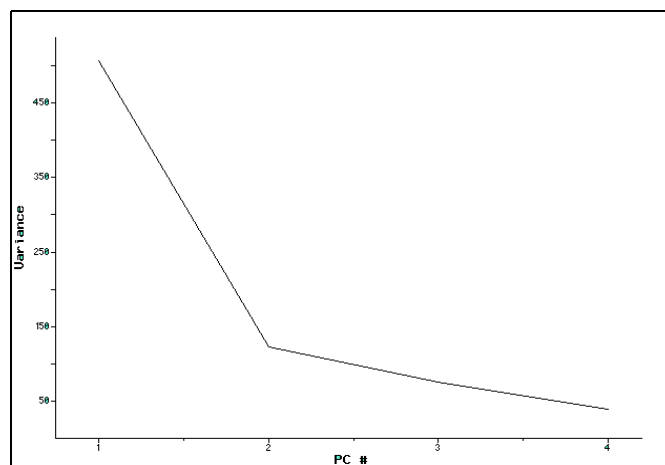


Figura 23: Gráfico da variância descrita pelas componentes principais para o tratamento 1, excluindo a amostra 1 (F011), 45 (F071) e 48 (R031).

Os resultados obtidos após a retirada de cada uma destas amostras podem ser observados no quadro 6. A variância descrita pela primeira componente principal aumentou inicialmente de 60,61% para 63,34%, após a retirada da amostra 1; após a retirada das outras duas amostras anômalas (45 e 48) este valor subiu para 65,82%. O número de componentes principais se manteve em relação à retirada somente da amostra 1, sendo descrita uma variância total de 91,55%.

Quadro 6: Comparação das variâncias descritas e respectivos números de fatores utilizados com a retirada das amostras 1 (F011), 45 (F071) e 48 (R031)

Amostras Excluídas	Número de Fatores	% de variância Descrita pela PC 1	% de Variância Acumulada
13	4	60.61	96.81
1, 13	3	63.34	92.66
1, 13, 45 e 48	3	65.82	91.55

Para se ter uma visão melhor do comportamento do conjunto de dados como um todo, observa-se que seria mais adequado retirar as amostras 1, 13, 45 e 48.

Análise de Escores e Pesos

A discriminação entre as duas baías está relacionada à segunda componente principal, como pode ser observado na figura 24. Nota-se perfeitamente que as amostras de cada uma das duas baías formam agrupamentos distintos ao longo do eixo da segunda componente principal, sendo que as amostras de BOI possuem valores mais altos nesta componente do que as amostras de DB. As variáveis que permitem esta separação, ou seja, as variáveis que caracterizam uma ou outra baía, estão associadas portanto a esta componente, e podem ser observadas no gráfico dos pesos (figura 25). Estas variáveis possuem pesos diferenciados na segunda componente principal, que podem ser observados no quadro 7.

Quadro 7: Peso de cada variável nas três primeiras componentes principais após a retirada das amostras 1, 13, 45 e 48

Variável	PC1	PC2	PC3
C14N	0.355	-0.112	-0.3327
C14P	0.384	-0.037	-0.1347
C13 D	0.319	0.409	0.0609
C14C	0.338	0.27	-0.0123
Per	0.217	-0.089	0.6647
Pyro	0.177	-0.46	0.4949
TPAH	0.372	-0.1667	0.1148
Petro	0.384	0.005	-0.1668
D2/P2	0.0187	0.702	0.3353
C2P	0.38	-0.055	-0.1643

Pelo gráfico dos pesos (na figura 25) e no quadro 7, pode-se verificar claramente que as variáveis C13D, C14C e D2/P2 possuem alto *peso positivo* na componente 2.

Estes pesos estão associados às amostras pertencentes à classe BOI, que apresentam alta concentração destes compostos. No gráfico dos escores, na figura 24 pode-se observar que as amostras da classe BOI possuem escores positivos na segunda componente principal.

Por sua vez, a variável Pyro possui um alto *peso negativo* nesta mesma componente. Estes pesos estão associados às amostras pertencentes à classe DB, que apresentam alta concentração de compostos pirogênicos.

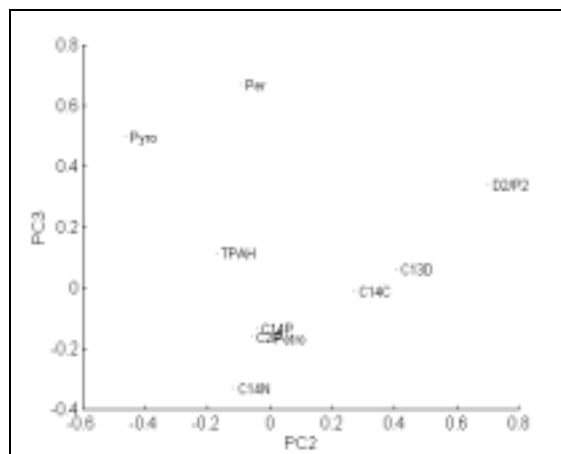
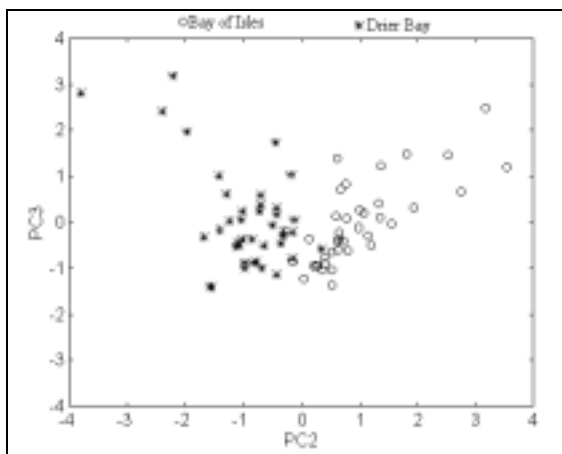
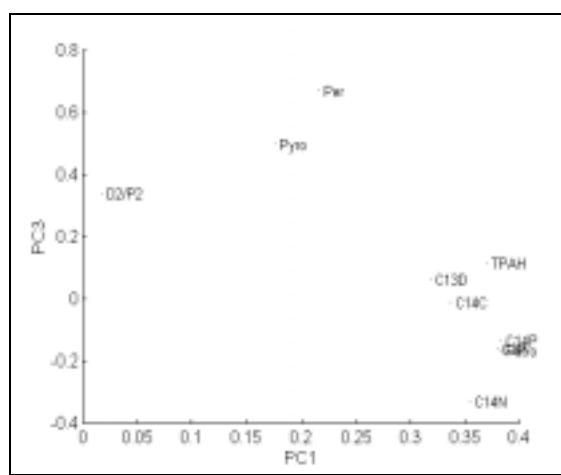
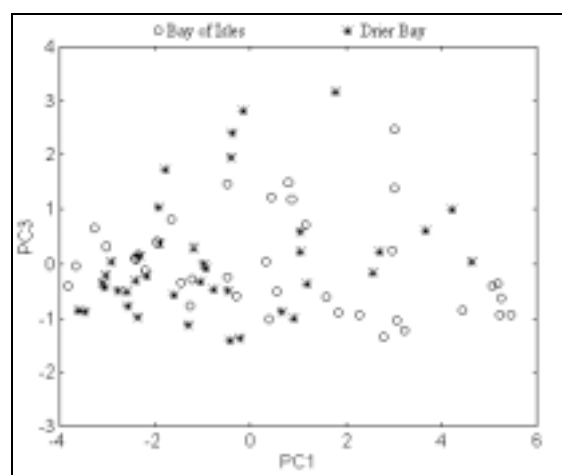
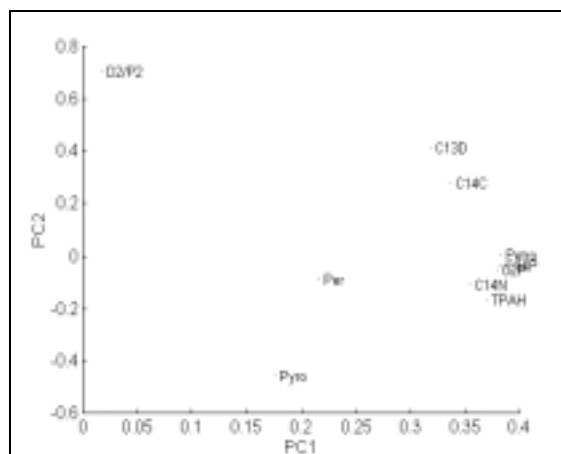
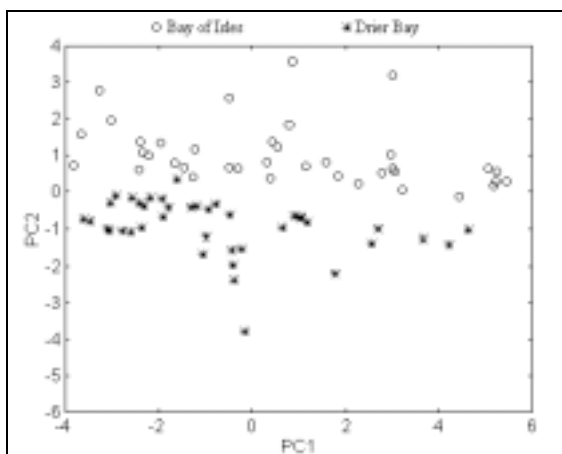


Figura 24: Gráficos dos escores em PC1, PC2 e PC3, para o tratamento 1, após a retirada das amostras 1, 13, 45 e 48.

Figura 25: Gráficos dos pesos em PC1, PC2 e PC3, para o tratamento 1, após a retirada das amostras 1, 13, 45 e 48.

Ainda no quadro 7 é possível notar que algumas variáveis possuem baixos pesos na segunda componente principal, que é a responsável pela separação entre as duas

baías. Isto quer dizer que estes grupos de hidrocarbonetos praticamente não interferem na diferenciação entre uma baía e a outra. Estas variáveis são C14P (isômeros do C1-C4 fenantreno), Per (isômeros do perileno), Petro (isômeros dos PAHs petrogênicos) e C2P (isômeros do C2-fenantreno).

Para checar a influência que cada uma destas variáveis tem na descrição dos dados, foi necessário refazer a análise de componentes principais, desta vez retirando-se uma de cada vez e comparando a mudança que ocorre nos valores de variância descrita em cada fator com os valores obtidos anteriormente, até se chegar nas condições otimizadas.

Análise das Variáveis

Os dados foram analisados novamente, desta vez retirando-se uma a uma as variáveis que possuem valores de pesos próximos de zero na segunda componente principal e comparando os valores de variância descrita em cada componente principal com os valores anteriores, bem como o número destes valores necessários à descrição dos dados.

No quadro 8 pode-se observar que o número de componentes principais para cada caso permanece o mesmo, no entanto as porcentagens de variância descritas em cada uma delas é diferente. Com a retirada da variável Per, houve um aumento na porcentagem de variância descrita em cada componente principal, o que resulta numa melhora do tratamento. O aumento foi de 65.82% para 70.07% na primeira componente principal e de 91.55% para 96.09% do total da variância descrita, o que pode ser considerado um aumento expressivo. Isto significa que o mesmo número de componentes principais está sendo utilizado para descrever o mesmo conjunto de dados, mas contém mais informação.

A retirada das variáveis C14P, Petro e C2P provocou uma diminuição nestes valores. Ao observarmos os valores dos pesos de cada uma das variáveis nas componentes principais (quadro 8), veremos que as variáveis C14P, Per, Petro e C2P apresentam baixo peso na segunda componente principal, que é a componente responsável pela discriminação das duas baías. Se observarmos os pesos destas

variáveis na primeira componente principal, veremos que todas elas apresentam valores acima de 0.3, com exceção de Per. Isto significa que apesar de não contribuírem para descrever as informações contidas na segunda componente principal, no caso a discriminação entre as duas baías, estas três variáveis são responsáveis por alguma informação contida na primeira componente principal, e que não está sendo considerada no momento, e portanto não devem ser excluídas do tratamento³⁷.

Quadro 8: Porcentagem de variância descrita pelas componentes principais retirando-se as variáveis C14P, Per, Petro e C2P uma de cada vez

Retirada	Número de Fatores		Variância Descrita	Variância (%)	Variância Acumulada (%)
nehumas	3	PC1	506.82	65.82	65.82
		PC2	122.73	15.94	81.76
		PC3	75.38	9.79	91.55
C14P	3	PC1	432.39	62.39	62.39
		PC2	122.39	17.68	80.07
		PC3	73.74	10.64	90.71
Per	3	PC1	485.57	70.07	70.07
		PC2	122.21	17.63	87.7
		PC3	58.11	8.38	96.09
Petro	3	PC1	432.57	62.43	62.43
		PC2	122.73	17.71	80.13
		PC3	72.85	10.51	90.64
C2P	3	PC1	434.26	62.66	62.66
		PC2	122.31	17.65	80.31
		PC3	72.98	10.53	90.84

Per, Pyro e D2/P2 apresentam os menores pesos (abaixo de 0.3) na primeira componente principal. Todas as outras variáveis apresentam pesos maiores do que 0.3. Das três, somente Per apresenta baixo peso na segunda componentes principal, um indicativo de que somente as variáveis Pyro e D2/P2 são importantes para a descrição dos dados. Escolheu-se então retirar a variável Per e manter as outras variáveis.

³⁷ De fato, a primeira componente principal está relacionada às profundidades amostradas, como será apresentado posteriormente nos tratamentos 2A e 2B.

Pelo gráfico dos escores (figura 26), observa-se que a discriminação entre as amostras das duas baías ocorre ao longo da segunda componente principal, como descrito anteriormente. Esta separação está relacionada às variáveis D2/P2, C13D, C14C e Pyro, como pode ser observado no gráfico dos pesos (figura 27), cujos valores em cada componente principal podem ser verificados no quadro 9.

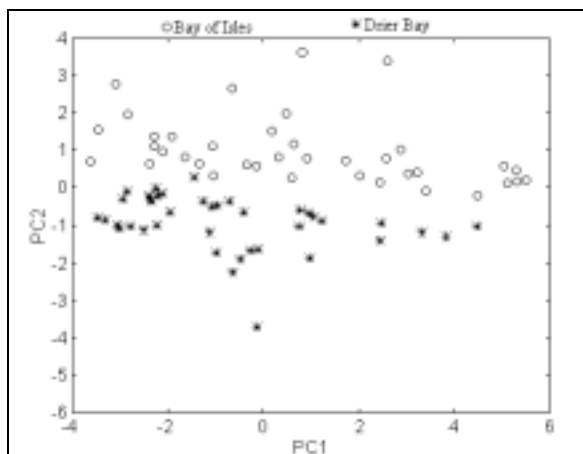


Figura 26: Gráfico dos escores para o tratamento 1, após a retirada das amostras 1, 13, 45 e 48, e da variável Per.

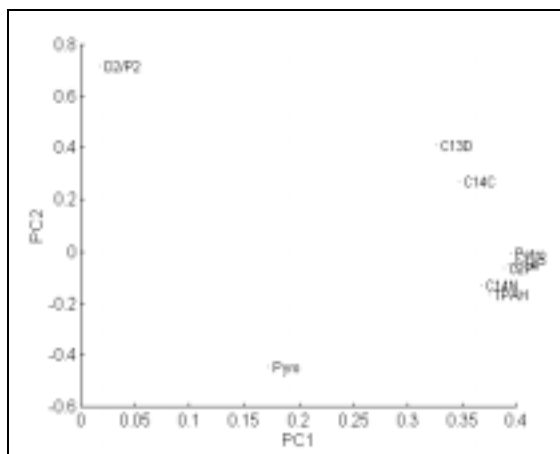


Figura 27: Gráfico dos pesos para o tratamento 1, após a retirada das amostras 1, 13, 45 e 48, e da variável Per.

Nenhuma mudança foi observada na terceira componente principal, portanto nas figuras 26 e 27 são mostradas somente as componentes 1 e 2.

Quadro 9: Peso de cada variável nas três primeiras componentes principais após a retirada das amostras 1, 13, 45 e 48 e da variável Per

Variável	PC1	PC2	PC3
C14N	0.3686	-0.1324	-0.3074
C14P	0.3942	-0.0489	-0.1057
C13 D	0.3272	0.407	0.0794
C14C	0.3481	0.2661	0.0597
Pyro	0.1741	-0.4482	0.8016
TPAH	0.3764	-0.1677	0.1974
Petro	0.3951	-0.0009	-0.127
D2/P2	0.0183	0.71409	0.4011
C2P	0.3899	-0.0668	-0.1609

A amostra 1 (F011), que apresenta alto teor de perileno e portanto apresenta comportamento anômalo até então, poderia ser incluída no tratamento, uma vez que esta

variável foi excluída. O teor de perileno no entanto é considerado no cálculo do teor total de PAH (variável TPAH) (Boehm *et alli*, 1998). Conseqüentemente o valor desta variável para a amostra 1 continua sendo alto, contribuindo para que esta amostra continue apresentando um comportamento anômalo em relação às amostras da sua classe. Decidiu-se manter esta amostra fora do tratamento.

Tratamento 2A: Análise de Componentes Principais em Bay of Isles

O objetivo deste tratamento foi verificar a distribuição de hidrocarbonetos aromáticos policíclicos ao longo de três profundidades distintas de *Bay of Isles* (BOI), a baía que recebeu o maior impacto do derramamento: uma mais superficial perto da costa (de 0 a 50 m), outra intermediária (de 50 a 100 m) e outra mais profunda e mais distante da linha da praia (de 100 a 150 m).

Como decidido anteriormente, as amostras 1 (F011), 13 (R101), 45 (F071) e 48 (R031) e a variável Per foram retiradas. O gráfico da figura 28 descreve quanto da variância total dos dados é explicado pela incorporação de cada componente principal adicional.

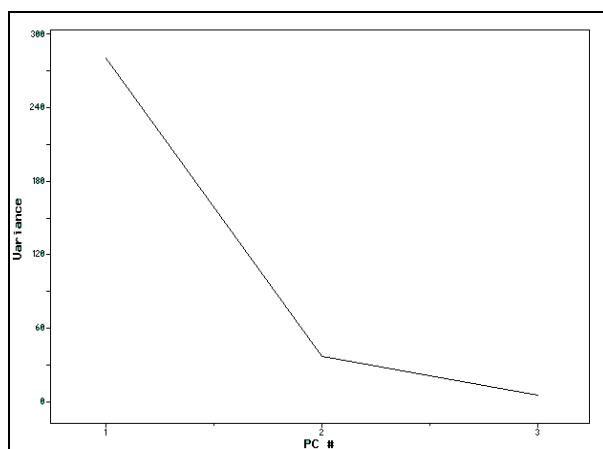


Figura 28: Variância descrita pelas componentes principais para o tratamento 2A.

Na tabela do quadro 10 é possível observar que a primeira componente principal descreve 84% da variância dos dados. Considerando que 95% da informação dos dados originais está contida nas duas primeiras componentes principais e que a terceira

componente contém apenas 1,67% de informação, considera-se que duas componentes principais são suficientes para a descrição dos mesmos.

Quadro 10: Porcentagem de variância descrita pelas componentes principais no tratamento 2A (Bay of Isles)

	Variância Descrita	Variância (%)	Variância Acumulada (%)
PC1	280.31	84.18	84.17
PC2	37.34	11.21	95.39
PC3	5.55	1.67	97.06

Análise de Escores e Pesos

Pelos gráficos dos escores, na figura 29, vê-se que a profundidade 1 forma um agrupamento perfeitamente separado ao longo da primeira componente principal. As profundidades 2 e 3, apesar de predominantemente ocuparem regiões diferentes, apresentam uma maior homogeneidade entre os perfis das amostras e não é possível distinguir a região limite entre as duas classes. Elas praticamente se misturam.

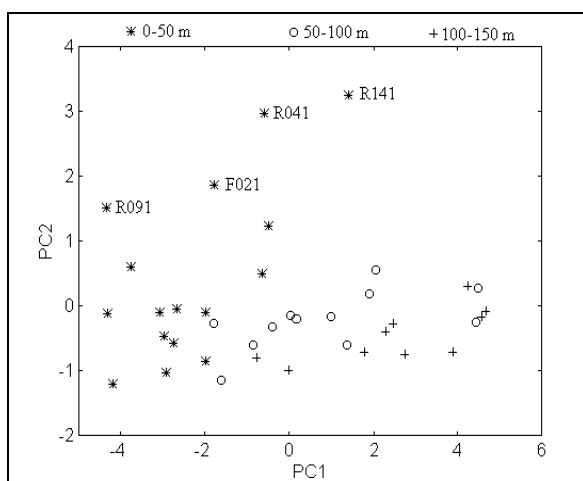


Figura 29: Gráficos dos escores para o tratamento 2A (Bay of Isles).

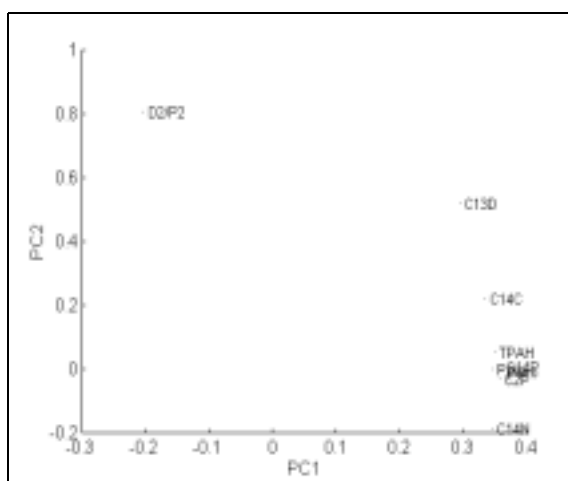


Figura 30: Gráficos dos pesos para o tratamento 2A (Bay of Isles).

A discriminação entre as três profundidades, seja ela evidente (como no caso da profundidade 1) ou não (profundidades 2 e 3, com perfis semelhantes) ocorre ao longo

da primeira componente principal, mas com a contribuição da componente 2. Pelos gráficos dos pesos (figura 30) pode-se verificar claramente que a única variável que apresenta alguma contribuição para a separação das amostras da profundidade 1 em *Bay of Isles* é a variável D2/P2. Ainda no gráfico dos pesos pode-se observar que as variáveis C14N, C14P, C14C, Pyro, TPAH, Petro e C2P caracterizam as amostras de sedimento coletadas nas regiões mais profundas da baía.

Ao longo da segunda componente principal é possível notar que algumas amostras da classe 1 aparecem ligeiramente separadas do agrupamento principal. Estas amostras foram coletadas nas estações F021, R041, R091 e R141 e pelo gráfico dos escores e dos pesos (figuras 29 e 30) é possível observar que as variáveis responsáveis por esta separação são as variáveis D2/P2 e C13D. No quadro 2 é possível observar que as razões D2/P2 no sedimento das amostras da classe BOI variam de 0.23 a 0.62, e que para estas 4 amostras esta razão é bem mais alta (0.67, 0.73, 0.80 e 0.68 respectivamente). No quadro 11 podem ser observados os pesos que cada variável tem em cada uma das duas componentes principais que foram utilizadas na descrição do sistema.

Quadro 11: Peso de cada variável na PC1 e PC2, para o tratamento 2A (Bay of Isles)

Variável	PC1	PC2
C14N	0.3479	-0.1908
C14P	0.3621	0.0025
C13D	0.2972	0.5156
C14C	0.3373	0.2177
Pyro	0.3475	-0.0038
TPAH	0.3515	0.0508
Petro	0.3623	-0.0101
D2/P2	-0.201	0.804
CP2	0.3601	-0.0344

Tratamento 2B: Análise de Componentes Principais de Drier Bay

Drier Bay foi a baía que recebeu pouco impacto do derramamento e por isto foi utilizada como referência para comparar com a contaminação de *Bay of Isles*. A amostragem também foi feita em três profundidades distintas, como explicado

anteriormente. Como decidido no tratamento anterior, as amostras 45 (F071) e 48 (R031), e a variável Per foram retiradas.

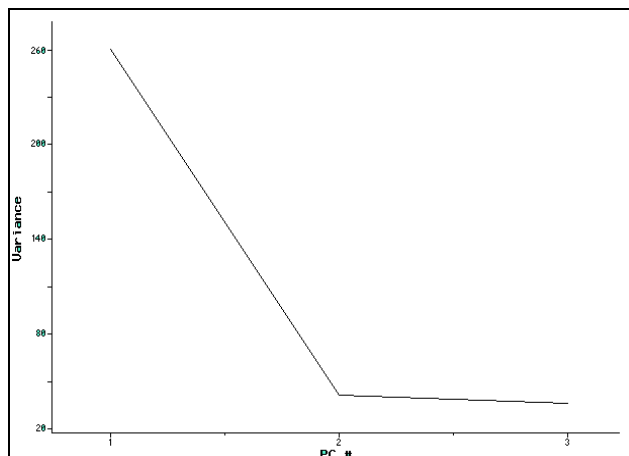


Figura 31: Variância descrita pelas componentes principais para o tratamento 2B.

O gráfico da figura 31 e o quadro 12 descrevem quanto da variância total dos dados é explicada pela incorporação de cada componente principal adicional.

Quadro 12: Porcentagens de variância descrita pelas componentes principais no tratamento 2B

	Variância Descrita	Variância (%)	Variância Acumulada (%)
PC1	260.47	74.21	74.21
PC2	41.74	11.89	86.1
PC3	36.22	10.32	96.42

A primeira componente principal descreve 74.21% da variância total dos dados (Quadro 12). Pelos valores da porcentagem de variância descrita pelas componentes principais, escolheu-se 3 componentes principais como número ótimo para a descrição do conjunto.

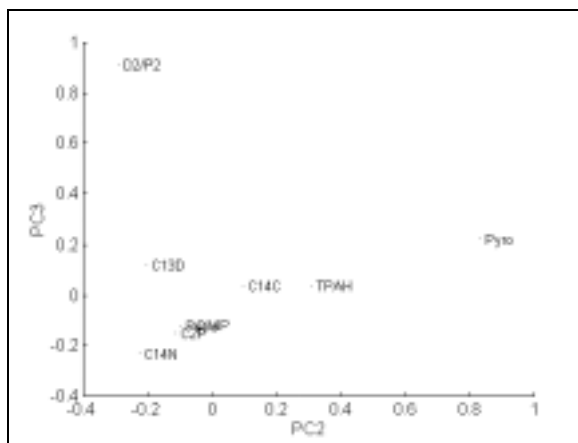
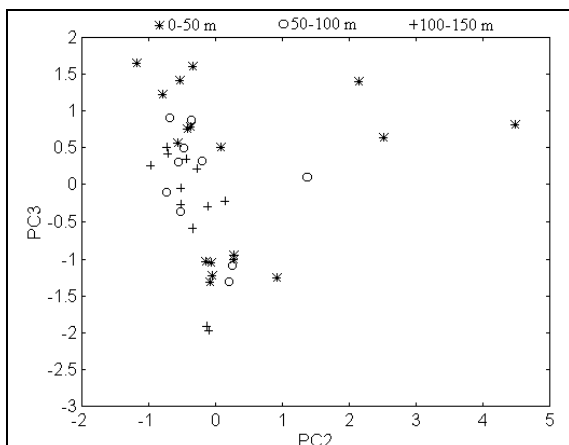
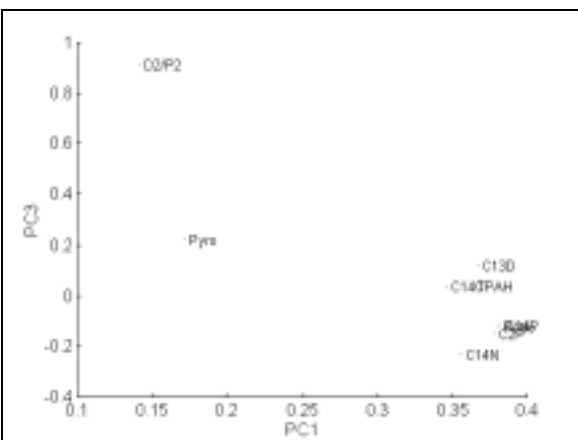
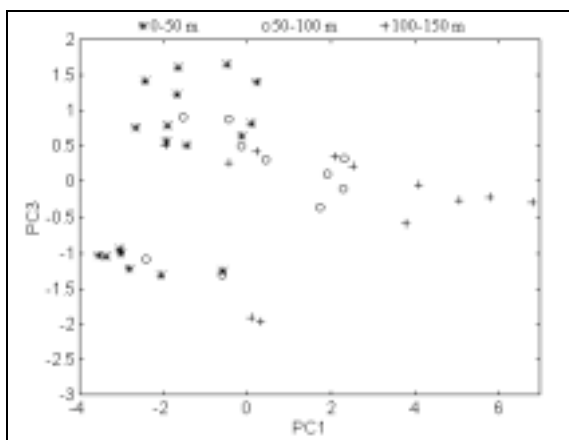
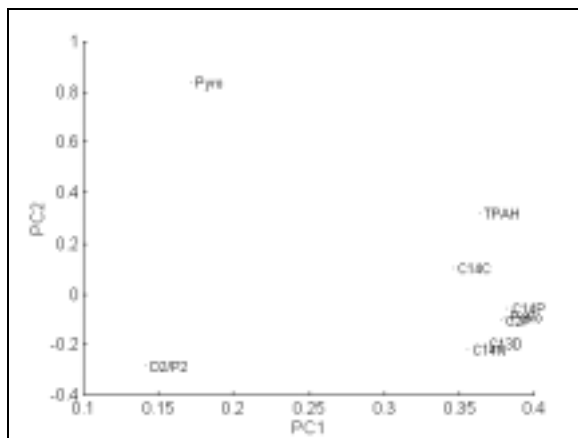
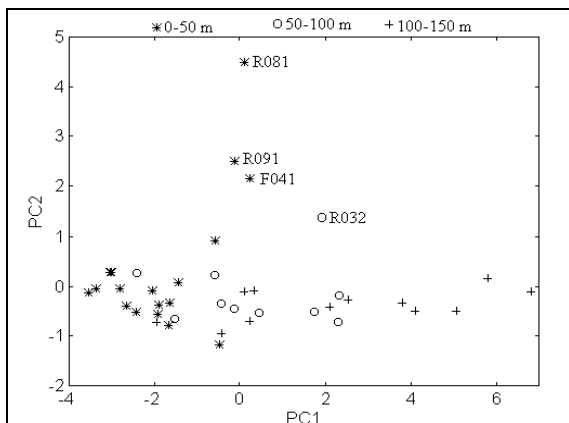


Figura 32: Gráficos dos escores em PC1, PC2 e PC3, para o tratamento 2B.

Figura 33: Gráficos dos pesos em PC1, PC2 e PC3, para o tratamento 2B.

Nos gráficos dos escores e dos pesos (figuras 32 e 31) é possível notar que as amostras das classes referentes às três profundidades estão bem misturadas. É possível distinguir a região espacial que cada classe ocupa, mas não existe uma linha limite entre

as amostras de uma ou outra profundidade. As amostras estão dispersas ao longo da primeira componente principal, de acordo com a classe a que pertencem. Isto acontece porque a baía em questão não foi muito contaminada pelo derramamento, então as três profundidades possuem um perfil relativamente parecido. Diferente do caso anterior, em que a região mais contaminada pelo derramamento, no caso os sedimentos marinhos mais próximos à costa de *Bay of Isles*, apresentam um perfil extremamente diferenciado das outras regiões.

Observando os valores dos pesos no quadro 13 nota-se que as variáveis Pyro e D2/P2 têm um valor de peso pequeno na primeira componente principal, e estão relacionadas à área mais superficial da baía (profundidade 1).

Quadro 13: Peso de cada variável nas três primeiras componentes principais, para o tratamento 2B

Variável	PC1	PC2	PC3
C14N	0.3564	-0.2222	-0.2354
C14P	0.3826	-0.0606	-0.1204
C13D	0.3673	-0.1981	0.1153
C14C	0.3467	0.107	0.0368
Pyro	0.1723	0.8342	0.2191
TPAH	0.3642	0.3126	0.0343
Petro	0.3815	-0.0912	-0.1236
D2/P2	0.1422	-0.2881	0.9098
CP2	0.3786	-0.1105	-0.152

Na segunda componente principal, estas amostras também têm uma alta contribuição, sendo que a variável Pyro é responsável pela discriminação das amostras das estações F041, R081, R091 e R032 do restante do conjunto. Estes pontos de amostragem estão localizados na região mais superficial de Drier Bay (exceto R032 que pertence à segunda profundidade), e apresentam altos teores de hidrocarbonetos de origem pirogênica, como pode ser observado no quadro 2.

As variáveis C14N, C14P, C13D, C14C, TPAH, Petro e C2P caracterizam as regiões mais profundas de *Drier Bay*.

Retirada das amostras F041, R081, R091 e R032

Para testar a influência da variável Pyro na descrição do sistema, as amostras com altos teores de pirogênicos foram retiradas e a análise foi feita novamente. O número de componentes principais necessários para se descrever o conjunto de dados caiu de 3 para 2, como pode ser observado no quadro 14, e a variável Pyro passa a contribuir somente para explicar as amostras das regiões mais profundas de Drier Bay, como pode ser observado no gráfico dos pesos e dos escores (figuras 34 e 35) e na tabela dos pesos (quadro 15).

Quadro 14: Porcentagem de variância descrita pelas componentes principais após a retirada da amostras F041, R081, R091 e R032

	Variância Descrita	Variância (%)	Variância Acumulada (%)
PC1	264.42	83.94	83.94
PC2	32.9	10.44	94.39
PC3	10.73	3.41	97.79

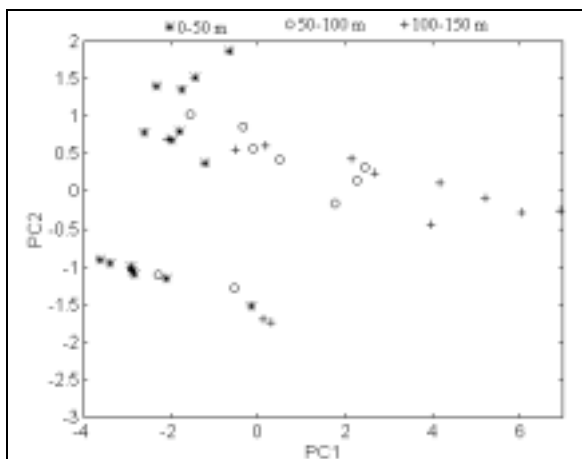


Figura 34: Gráfico dos escores em PC1 e PC2, para o tratamento 2B, após a retirada das amostras F041, R081, R091 e R032.

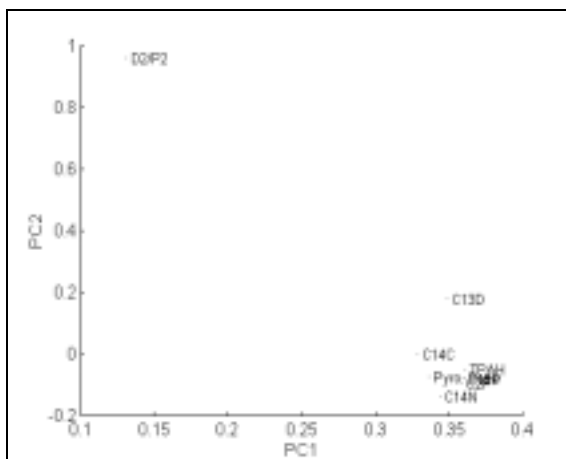


Figura 35: Gráfico dos pesos em PC1 e PC2, para o tratamento 2B, após a retirada das amostras F041, R081, R091 e R032.

A variável D2/P2 continua relacionada às amostras pertencentes à área mais superficial da baía, pois é a única variável com baixo peso na primeira componente principal. Na figura 34 é possível observar que ao longo da segunda componente

principal ocorre a formação de dois agrupamentos formados por amostras pertencentes às três profundidades. Esta separação ocorre por causa da razão D2/P2. Nas amostras com baixo peso na segunda componente principal a razão D2/P2 é zero, como pode ser observado no quadro 2³⁸.

Quadro 15: Peso de cada variável nas duas primeiras componentes principais, para o tratamento 2B, após a retirada das amostras F041, R081, R091 e R032

Variável	PC1	PC2
C14N	0.3443	-0.1402
C14P	0.3607	-0.0827
C13D	0.3495	0.1774
C14C	0.3292	-0.0039
Pyro	0.3368	-0.0781
TPAH	0.3621	-0.0519
Petro	0.3603	-0.0788
D2/P2	0.1321	0.9577
CP2	0.3592	-0.099

Discussão dos resultados

O petróleo e seus derivados contêm uma complexa mistura de milhares de compostos orgânicos, entre eles os PAHs. Os estudos a respeito do impacto ambiental destes compostos em ambientes contaminados por descarga de petróleo ou derivados não é tarefa simples, pois estes compostos podem ser introduzidos no ambiente através de outras fontes, como por exemplo descargas industriais e urbanas, deposição atmosférica, processos diagenéticos naturais. Neste trabalho, a análise de componentes principais (PCA) foi utilizada para observar a discriminação entre duas baías localizadas em *Knight Island, Prince William Sound*, que receberam o impacto do derramamento de óleo do petroleiro *Exxon Valdez*, ocorrido em março de 1989, através do perfil de distribuição de hidrocarbonetos aromáticos policíclicos (PAH) nos sedimentos marinhos amostrados em três profundidades ao longo de cada baía.

³⁸ Cf. página 45.

Tratamento 1: Bay of Isles vs. Drier Bay

Na primeira parte da análise (tratamento 1) as amostras do conjunto formaram dois agrupamentos ao longo da segunda componente principal, caracterizando cada uma das duas baías. Neste agrupamento, as variáveis D2/P2, C13D e C14C apresentaram alto peso nesta componente (figuras 24 e 25, e quadro 7), sendo responsáveis pela discriminação de *Bay of Isles*, no caso a baía que recebeu diretamente o impacto do derramamento. A variável Pyro está relacionada à baía que sofreu menos os efeitos do desastre, e por isto foi utilizada como baía de referência.

A influência da variável Per na descrição dos dados

A variável Per, que fornece o teor de perileno, apresenta um peso muito baixo na PC2, então foi retirada da análise por não favorecer a separação de nenhuma das classes (como dito anteriormente, esta separação acontece na PC2). O perileno é um PAH de origem diagenética, pois se forma na etapa da diagênese em processos biológicos, e portanto costuma estar presente onde há acúmulo de matéria orgânica (Wang *et alli*, 1999).

As duas baías apresentam um perfil parecido quanto à presença deste PAH, exceto para a amostra 1, coletada na estação F011. Esta amostra possui um teor de perileno muito alto, 480 ppb, em comparação não somente com as outras amostras da profundidade 1 em *Bay of Isles* (1-98 ppb de sedimento), como também em relação a todas as amostras do conjunto, cujo valor mais alto de perileno é de 140 ppb de sedimento. A estação onde o sedimento foi amostrado está localizada na entrada do braço oeste de *Bay of Isles* (figura 36), onde existe uma depressão que atua como área de deposição de matéria orgânica de origem tanto marinha quanto terrestre, que favorece a formação do perileno. A amostra 1 também apresenta alto teor de TPAH, cujo cálculo é feito levando-se em consideração também o teor de perileno (Boehm *et alli*, 1998).

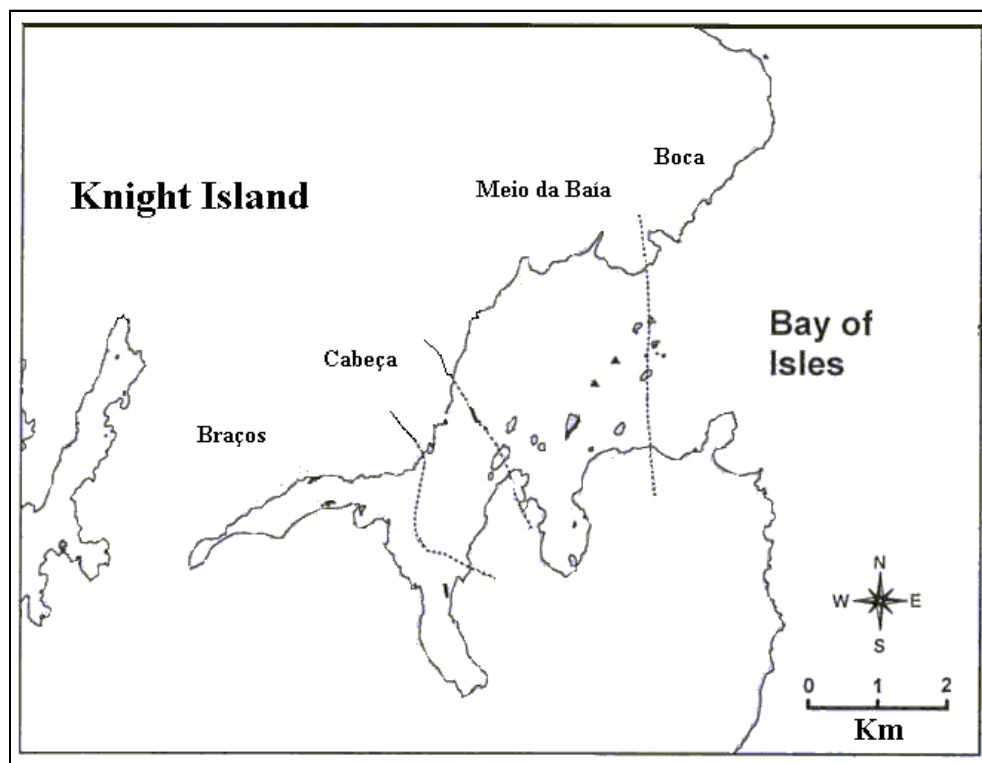


Figura 36: Detalhe de Bay of Isles mostrando os braços leste e oeste da baía (adaptado de Burns et alli, 1997).

A influência da variável Petro na descrição dos dados

As frações de hidrocarbonetos petrogênicos são formadas pelas famílias do naftaleno, fluoreno, fenantreno, dibenzotiofeno e crisenos, e seus homólogos alquilados, sendo que estes últimos são mais abundantes do que os não-substituídos. Frações pirogênicas, por exemplo, têm uma maior quantidade de PAHs não substituídos do que os petrogênicos (Page et alli, 1996).

Assim como a variável Per, a variável Petro também possui um peso baixo ao longo da PC2 no tratamento 1, pois está presente em toda a extensão da baía. Porém, ao contrário de Per, Pyro apresenta uma alta contribuição na PC1, então foi mantida no tratamento.

Em *Prince William Sound* existem fontes naturais e antropogênicas de PAHs petrogênicos. Parte destes PAHs vêm do Golfo do Alasca, região rica em depósitos de carvão e petróleo, e chegam até *PWS* aderidos ao sedimento trazido pelas correntes

marítimas. Estes PAHs petrogênicos têm um baixo teor de alquil dibenzotiofenos. Outra parte destes PAHs foram introduzidos em *PWS* pelo óleo *Exxon Valdez*, e possuem alto teor de alquil-dibenzotiofenos (Short & Heintz, 1997; Short *et alli*, 1999; Boehm *et alli*, 1998). Então, apesar de estar presente ao longo das duas baías, esta variável auxilia na caracterização do óleo proveniente do derramamento (petrogênicos com altos teores de alquil-dibenzotiofeno) e também os petrogênicos naturalmente presentes na baía.

Caracterização de Bay of Isles

Esta baía localiza-se na parte nordeste de *Knight Island* e ficou mais exposta aos efeitos do derramamento por causa da sua posição geográfica (figura 14).

Na análise de componentes principais a variável D2/P2 está associada às amostras da classe BOI. Esta variável fornece a proporção entre os teores dos alquil-PAHs C2-dibenzotiofeno e C2-fenantreno, dois hidrocarbonetos que possuem velocidades de degradação por intemperismo muito parecidas, o que significa que esta razão tende a se manter constante ao longo do tempo, e por este motivo costumam ser utilizadas em estudos de derramamento de óleo no meio ambiente (Page *et alli*, 1996). Se observarmos o quadro 2, é possível verificar que a razão entre estes compostos tende a ser maior para as amostras de *Bay of Isles* do que para *Drier Bay*. A utilização da PCA permitiu relacionar esta variável com o sedimento contaminado pelo óleo *Exxon Valdez*, sem necessidade de checar estes valores um a um, para cada amostra.

Um estudo publicado em 1997 por Burns *et alli* mostrou as seguintes razões D2/P2: 0.14 em amostras coletadas em camadas bem abaixo da superfície do sedimento, de períodos anteriores ao derramamento; 0.15 e 0.23 na região não costeira de *Yakataga* (região não atingida) e *Chalmers* (região próxima às áreas atingidas); e valores de 0.88 até 1.26 em amostras do óleo *Exxon Valdez* fresco e após degradação por intemperismo nos mais variados graus.

Comparando estes valores é possível notar que a razão D2/P2 encontrada no sedimento da região é significativamente menor do que no óleo *Exxon Valdez* e com o uso de análise de componentes principais foi possível detectar esta diferença e utilizá-la na caracterização das amostras provenientes de cada uma das duas baías estudadas de

uma só vez, sem necessidade de checar estes teores um a um, tornando a análise mais rápida e menos suscetível a erros.

Estudos feitos em 1999 por Short *et alli* confirmam valores mais baixos da razão D2/P2 para amostras não contaminadas. Eles encontraram uma razão D2/P2 de 0.18 em amostras de óleo e carvão de *Katalla*, 0.11-0.21 em amostras de sedimentos bênticos em *Yakutat Bay*, e 0.07-0.31 em amostras terrestres próximos aos rios *Bering* e *Duktoth*. Estas regiões localizam-se no Golfo do Alasca, numa área que não foi atingida pelo derramamento *Exxon Valdez*.

Page *et alli* (1996) encontraram uma razão D2/P2 de 0.15 ± 0.02 em amostras de sedimento coletadas em camadas 10 cm abaixo da superfície água/sedimento. Esta profundidade corresponde ao sedimento depositado 10 anos antes do derramamento *Exxon Valdez*. Para o óleo *Exxon Valdez*, eles encontraram uma razão de 1.07 ± 0.02 , ou seja, um teor de homólogos de fenantreno consideravelmente maior.

A variável C13D também apresenta uma alta contribuição na segunda componente principal, responsável pela discriminação das amostras das duas classes. Esta variável é a soma de todos os alquil-PAHs (C1+C2+C3)-dibenzotiofenos, e inclui o teor de C2-dibenzotiofeno, que é o numerador responsável pela variação na razão D2/P2, que contribui para a separação das duas baías.

A variável C14C apresenta uma pequena contribuição na segunda componente principal. Esta variável é a soma de todos aos alquil-PAHs (C1+C2+C3+C4)-crisenos e também caracteriza o óleo proveniente do derramamento e está relacionada à degradação por intemperismo sofrida pelo óleo *ANS*. O óleo *Exxon Valdez* contém altos teores de crisenos alquilados (Burns *et alli*, 1997).

Caracterização de Drier Bay

Esta baía localiza-se na parte noroeste de *Knight Island* (figura 14), e apesar de também ter sido atingida pelo óleo do derramamento, por causa de sua posição geográfica mais protegida não sofreu tanto os efeitos quanto *Bay of Isles*, então foi utilizada nos estudos de Boehm *et alli* (1998) como baía de referência.

Na verdade, PAHs de origem pirogênica foram encontrados nos sedimentos provenientes das duas baías, mas em DB o teor é bem maior devido as atividades daquele local. Por isto o peso da variável Pyro tem uma considerável contribuição na PC2, mas não é da mesma ordem de grandeza que para a variável D2/P2, que possui peso muito alto na PC2 (a alta razão destes dois tipos de PAH é característica somente da BOI).

As amostras 45 (F071) e 48 (R031), pertencentes à área mais superficial de Drier Bay, apresentam os maiores teores de Pyro (843 e 939 ppb) e de TPAH (1364 e 1497 ppb). Estes pontos de amostragem situam-se em *Port Audrey* (figura 37), perto das ruínas de antigas fábricas de conserva (Boehm *et alli*, 1998).

Outras variáveis que também apresentam uma certa contribuição na PC2, porém bem pequena, e portanto pode-se dizer que ajudam na caracterização de Drier Bay são as variáveis TPAH e C14N, referentes ao teor total de PAH e de naftalenos alquilados.

Tratamento 2: A distribuição dos PAHs ao longo das três profundidades em cada baía

Numa segunda parte do estudo buscou-se traçar o perfil de hidrocarbonetos de cada baía separadamente, em amostras de sedimento coletadas ao longo de três profundidades diferentes: 0-50, 50-100 e 100-150 m.

Para as duas baías (tratamento 2A e 2B), as amostras de cada profundidade se espalharam ao longo da primeira componente principal. Apesar de ser possível distinguir a região da PC1 que cada grupo ocupa, não existe uma linha limite entre cada classe, mostrando que a distribuição de PAHs entre as regiões amostradas é mais ou menos a mesma, exceto para as amostras coletadas na parte mais superficial de *Bay of Isles*, que formam um agrupamento bem distinto do restante do conjunto.

A parte superficial de Bay os Isles

Esta separação da profundidade 1 em *Bay of Isles* está relacionada à variável D2/P2, como pode ser observado no gráfico dos escores e dos pesos (figuras 29 e 30) no

tratamento 2A. Segundo Boehm *et alli* (1998), “ao se misturar à água, o óleo formou uma emulsão que pela ação das correntes marítimas e dos ventos fortes da região foi aos poucos sendo arrastada para as regiões mais costeiras, depositando-se no sedimento marinho”. Isto explica o perfil diferenciado das amostras da classe 1 em Bay of Isles, e a homogeneidade entre as amostras das classes 2 e 3. Como o óleo teria se depositado na região mais costeira de *Bay of Isles*, a razão D2/P2 encontrada para as amostras coletadas nesta região também são maiores.

A parte superficial de Drier Bay

No tratamento 2B, com as amostras de *Drier Bay*, a região mais superficial não aparece tão separada quanto no tratamento 1A. O motivo é que a contaminação pelo óleo, caracterizada pela variável D2/P2 é menor, então as amostras desta classe aparecem mais misturadas às amostras da classe adjacente. Esta região mais superficial é caracterizada pelos teores de D2/P2 e de Pyro, devido à contaminação proveniente do continente. Esta separação pode ser visualizada nos gráficos dos escores e dos pesos (figura 32 e 33) no tratamento 2B.

As regiões mais profundas

Nos tratamentos 2A e 2B as variáveis C14N, C14P, C13D, C14C, Pyro (esta somente para BOI), TPAH, Petro e C2P possuem altos valores na caracterização as regiões mais profundas.

Além da estação F011 descrita anteriormente, as estações F021, R041, R091 e R141 na profundidade 1 de *Bay of Isles* também apresentam alta razão D2/P2 e alta concentração da variável C13D e este comportamento pode ser melhor visualizado através do tratamento 2A. Estas estações ficam localizadas no final do braço oeste de *Bay of Isles*, um local bem protegido e apropriado para a acumulação da matéria trazida pelas correntes marítimas, e portanto, com maiores resíduos da contaminação do derramamento (Boehm *et alli*, 1998).

Em *Drier Bay*, além das amostras excluídas do tratamento (F071 e R031) outras quatro amostras também apresentaram altos teores de pirogênicos: as amostras F041, R081 e R091 (da profundidade 1) e a amostra R032 (da profundidade 2). No gráfico dos escores do tratamento 2B (figura 32) é possível observar a separação destas amostras ao longo da segunda componente principal, e pelo gráfico dos pesos (figura 33) a variável Pyro aparece como responsável por esta separação. Estas estações estão localizadas próximas às ruínas de *Port Audrey* (Boehm *et alli*, 1998). Com a retirada destas amostras, o teor de pirogênicos passa a caracterizar as regiões mais profundas de *Drier Bay*.

Algumas variáveis apresentam ainda certas contribuições para a terceira componente principal. É o caso de variável D2/P2, que possui um alto peso nesta componente (0.908) e caracteriza algumas amostras pertencentes as classes 1 e 2 que possuem um alto valor para esta variável. A razão D2/P2 está em torno de 0.1 a 0.19 para as amostras de *Drier Bay*, exceto as amostras das estações F021, R121, R151 (da profundidade 1) e R012 e R062 (da profundidade 2) que cujas razões D2/P2 encontram-se entre 0.2 e 0.25. A estação R151 se localiza na entrada de *Drier Bay*, e pode ter sido mais contaminada.

Capítulo 3

Relação Quantitativa Estrutura-Propriedades de Hidrocarbonetos Aromáticos Policíclicos

Existem mais de 60.000 compostos químicos em produção atualmente. Destes, cerca de 500 são de interesse ambiental, devido à toxidez, tendência à bioacumulação e persistência no meio ambiente (Mackay *et alli*, 1998).

A existência destes produtos interferiu de maneira significativa na vida do planeta, iniciando um processo com duas conseqüências de caráter totalmente oposto: se por um lado permitiu um padrão de vida excepcional, e nunca visto até então, por outro iniciou um processo de desequilíbrio e degradação ambiental que pode culminar na extinção da vida. Recentemente a busca de um meio termo, ou seja, de uma forma de continuar usufruindo dos benefícios que a produção destes produtos pode oferecer à vida humana, mas minimizando as conseqüências danosas e sem agredir o meio ambiente, fez surgir inúmeras pesquisas na área ambiental, e neste contexto surgiu a química ambiental, que trata dos processos que dão origem aos produtos químicos, e as formas como eles podem contaminar o ar, a água, o sedimento e o solo, além claro, do tratamento dos resíduos gerados pelos inúmeros processos industriais utilizados para manter o padrão de vida humana atual.

Para entender e avaliar o impacto ambiental dos produtos químicos, é necessário saber por exemplo, quais os processos químicos e bioquímicos que geram ou consomem estes produtos, quanto tempo eles persistem no meio ambiente, como eles serão metabolizados pelos organismos. Todos estes efeitos dos compostos sobre o meio ambiente dependem de dois fatores fundamentais (Mackay *et alli*, 1998):

1. das condições do ambiente, como por exemplo, a temperatura, os fluxos e acumulação de ar, água e matéria sólida e a composição do ar, água, sedimento e solo;
2. e das propriedades físico-químicas dos poluentes, que influenciarão na forma como estes compostos irão se dispersar ou acumular no ambiente. Por exemplo, se um

determinado composto evapora ou se acumula no sedimento, se o composto é destruído, ou transformado em outra espécie química que pode ou não ser ativa.

Todas estas características despertam grande interesse no estudo de compostos com potencial poluente, com a finalidade justamente de se obter informações detalhadas sobre o comportamento dos mesmos no meio. Este entendimento é essencial para que estes produtos, que apresentam inúmeras aplicações, possam ser utilizados de forma consciente e responsável pela sociedade, sem ameaçar a qualidade de vida do planeta.

Uma técnica que vem sendo muito utilizada no estudo das propriedades moleculares é a *relação quantitativa estrutura-propriedade (QSPR³⁹)*, ou *estrutura-atividade (QSAR⁴⁰)*, que utiliza o cálculo de descritores moleculares para prever as propriedades físico-químicas e atividades dos compostos e assim estimar seu potencial tóxico e poluente.

Nas últimas décadas, muito tem sido estudado sobre as relações entre a estrutura molecular e as propriedades físico-químicas, ou a atividade biológica de algumas substâncias, principalmente na química medicinal e farmacêutica. Os estudos destas relações constituem uma importante e eficiente ferramenta para o desenvolvimento de novas drogas. Da mesma forma como na química ambiental o conhecimento das propriedades físico-químicas de um composto leva ao entendimento de como ele se dispersa no meio ambiente, o conhecimento dos parâmetros estruturais, e até mesmo destas propriedades físico-químicas, leva ao entendimento de como a droga atuará no combate a uma doença, ou na forma como um composto será metabolizado pelo organismo etc. (Kubinyi, 1993).

Este trabalho visa obter as relações quantitativas estrutura-propriedade de um conjunto de PAHs por meio da construção de modelos de previsão de ponto de ebulição (pe), coeficiente de partição octanol-água (Log Kow), índice de retenção (IR) e fototoxidez. Estas propriedades foram escolhidas por estarem relacionadas diretamente com o potencial tóxico e poluente destes compostos, e cada uma delas será tratada com detalhes mais adiante.

³⁹ QSPR – *Quantitative Structure Property Relationship*

⁴⁰ QSAR – *Quantitative Structure Activity Relationship*

Desde o seu surgimento, a quimiometria tem sido usada principalmente em problemas de química analítica, especialmente em calibração. Nos anos recentes, com o crescimento da importância das bases de dados e aplicações em química medicinal e computacional, o domínio da aplicação da quimiometria tem aumentado significativamente. Especialmente a relação entre estrutura e funções moleculares tem se tornado de especial interesse (Wehrens *et alli*, 1999).

Existem atualmente inúmeros bancos de dados disponíveis à comunidade científica, listando as substâncias químicas e suas propriedades físico-químicas e atividades farmacológicas. O potencial de informação destas bases de dados é ilimitada, mas requer métodos racionais para se fazer uso apropriado destes dados (Wehrens *et alli*, 1999).

No começo do século, a pesquisa sobre as propriedades e atividades dos compostos era feita com base na experimentação. Com o aparecimento da equação de Hammett⁴¹, que foi utilizada durante três décadas para descrever a reatividade de compostos orgânicos de uma maneira quantitativa, um novo caminho para a investigação de novos compostos foi dando lugar aos inúmeros e dispendiosos experimentos. Por meio de cálculos teóricos começaram a se tornar possíveis as comparações entre as diferenças estruturais das moléculas e a se correlacionar estas diferenças às propriedades moleculares, possibilitando descobrir, ou melhor dizendo, calcular as atividades e propriedades moleculares. Com o aprimoramento destas relações surgiram na década de 60 os estudos *de relação quantitativa estrutura-atividade (QSAR)*, e em seguida os estudos *de relação quantitativa estrutura-propriedades (QSPR)* (Leitão *et alli*, 2000; Kubinyi, 1993; van de Waterbeemd, 1995).

As análises de QSAR e QSPR utilizam análise de regressão para a elaboração de modelos matemáticos à partir de dados estruturais de compostos, e a partir destes modelos prever suas propriedades físico-químicas e atividades biológicas.

⁴¹ Para uma leitura mais detalhada sobre o histórico dos métodos de QSAR/QSPR sugiro o livro de H. Kubinyi (1993).

Estes métodos têm sido largamente empregados na estimativa de propriedades como densidade (Karelson & Perkson, 1999), ponto de ebulição (Arupjyoti and Iragavarapu, 1998; Wiener, 1947), solubilidade (Krop *et alli*, 1997), pressão de vapor (Burkhard *et alli*, 1985), coeficiente de partição octanol-água (Krop *et alli*, 1997), além de fatores como potencial tóxico de compostos (Ramos *et alli*, 1998; Zhao *et alli*, 1998; Verhaar *et alli*, 1996), constante de *Henry* (Krop *et alli*, 1997), de compostos pertencentes à uma classe com estruturas similares, como por exemplo, os bifenis policlorados (PBCs⁴²), dibenzofuranos, clorobenzenos, dioxinas policloradas, hidrocarbonetos policíclicos aromáticos (PAH), ácidos graxos e muitos outros.

Os estudos de QSAR/QSPR são muito empregados nos casos em que a determinação experimental das propriedades e atividades biológicas dos compostos é difícil, ou quase impossível, podendo envolver riscos de toxidez, ou talvez porque o composto em questão não é muito estável. Esta estimativa permite ainda reduzir o custo da determinação experimental destes parâmetros.

Previendo as Propriedades Químicas - Materiais e Métodos

As diferenças de comportamento dos compostos no meio ambiente ou até mesmo a forma como eles são metabolizados nos organismos dos seres vivos está relacionada às suas propriedades moleculares. O estudo destas propriedades envolve o conhecimento de fatores como a estrutura dos compostos, a disposição dos átomos, o tamanho, a forma e a simetria da molécula, além de fatores eletrônicos, como a energia dos orbitais de fronteira. Justamente por estarem relacionadas a estes fatores, estas propriedades moleculares costumam ser similares para compostos da mesma classe, e variam dentro de séries homólogas (Kubinyi, 1993). Ponto de ebulição, coeficiente de partição octanol-água, índice de retenção e fototoxidez são alguns bons exemplos de algumas propriedades moleculares que ilustram este tipo de correlação entre as propriedades e os fatores estruturais.

Neste trabalho foram utilizados descritores moleculares referentes às características estruturais, topológicas, termodinâmicas e eletrônicas de um conjunto de

⁴² PCB – *Polychlorinated Biphenyls*

70 PAHs, para a construção de modelos de previsão de algumas propriedades físico-químicas e atividade para as moléculas deste conjunto que não possuem estas propriedades determinadas experimentalmente.

Para a construção dos modelos de regressão foi utilizado o *método dos mínimos quadrados parciais – PLS*, com *preprocessamento autoescalado e validação cruzada*, com a ajuda dos programas Pirouette 2.02 (Infometrix) e MATLAB (MathWorks Inc).

PAH – As Moléculas Estudadas

A preocupação com o crescente aporte de hidrocarbonetos policíclicos aromáticos no meio ambiente impulsionou o estudo das propriedades e atividades biológicas destes compostos, uma vez que estas propriedades estão intimamente relacionadas com seu potencial poluente e tóxico.

Estes compostos predominam nas regiões mais urbanizadas, ou em locais onde existe a prática das queimadas, e são conhecidos contaminantes de corpos aquáticos. Eles são produzidos a partir de hidrocarbonetos saturados sob condições deficientes de oxigênio, ou seja, através da degradação ou combustão incompleta de materiais orgânicos (carvão, tabaco, petróleo e derivados). Estas condições favorecem o consumo do hidrogênio, em relação ao consumo do carbono. Os carbonos excedentes se arranjarão na forma termodinamicamente mais favorável que consiste em sistemas de anéis aromáticos condensados, os *hidrocarbonetos policíclicos aromáticos*, ou mais comumente, os PAHs (Manahan, 1994).

Os processos de produção de PAHs geram uma variedade imensa de compostos similares em estrutura e propriedades, o que dificulta a análise e controle destes poluentes, que são conhecidos por suas características mutagênicas e carcinogênicas (Lehner *et alli*, 1996). O perfil de distribuição de PAHs gerado através dos processos de combustão varia de acordo com a fonte emissora e também devido à uma série de outros fatores, como o tipo de matéria orgânica utilizada como combustível e a extensão do aquecimento, portanto é muito difícil determinar exatamente as concentrações relativas de cada PAH gerado (Pereira Netto *et alli*, 2000).

Para estimar o potencial tóxico, assume-se que para um caso ideal, todos os perfis são idênticos e a medida da concentração de um analito pode ser usada como a medida do potencial carcinogênico de todo o conjunto. Na maioria dos casos, a avaliação das concentrações de PAHs no ar é baseada na quantidade de benzo[a]pireno. Alguns dos PAHs mais abundantes são o benzo[a]pireno, benzo[a]antraceno, criseno, benzo[e]pireno, e benzo[j]fluoranteno, cujas estruturas podem ser visualizadas na figura 38 (Manahan, 1994; Ullman, 1989).

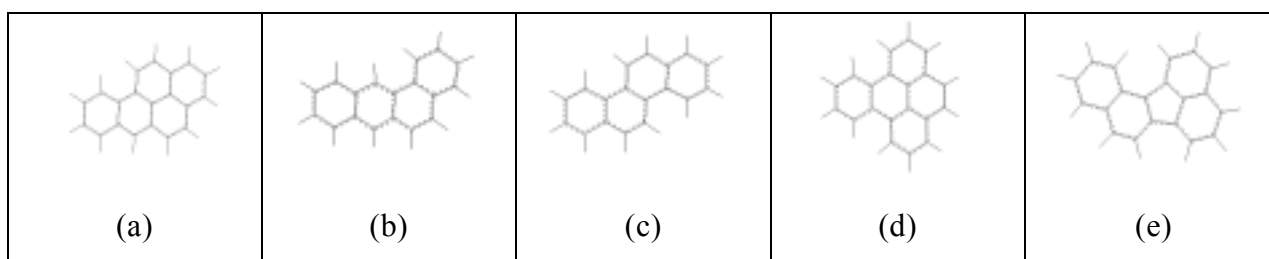


Figura 38: Estruturas de alguns dos PAHs mais abundantes: (a) benzo[a]pireno, (b) benzo[a]antraceno, (c) criseno, (d) benzo[e]pireno e (e) benzo[j]fluoranteno.

No ambiente aquático, os PAHs podem contaminar tanto a água quanto o sedimento, dependendo da sua hidrofobicidade, isto é, de sua afinidade com compostos polares ou apolares. Os PAHs mais hidrofóbicos tendem a aderir nas partículas do sedimento, ou nos tecidos biológicos, enquanto aqueles menos hidrofóbicos tendem a ficar menos retidos. (Hatch & Burton Jr., 1999).

A contaminação do solo por esta classe de compostos se dá principalmente por deposição atmosférica e despejo de águas e sedimentos contaminados, e o transporte ocorre por difusão. Portanto, a permanência no solo está intimamente relacionada à solubilidade de cada composto em água. A eliminação dos PAHs do solo pode ocorrer por diversas vias, entre elas: a degradação biológica, a degradação fotolítica, a oxidação química e a volatilização, no caso dos compostos menores. A velocidade destes processos varia bastante, de acordo com a estabilidade e o peso molecular de cada composto, sendo em geral mais lenta para os compostos com um número maior de anéis (Ullman, 1989) .

De um modo geral, os compostos contendo um número de anéis maior que quatro têm maior relevância do ponto de vista ambiental, pois a maioria deles apresenta

propriedades carcinogênicas e são extremamente resistentes à degradação enzimática (Chen *et alli*, 1996).

A contaminação humana se dá geralmente através da contaminação ambiental, pela inalação ou absorção pela pele, ou ingestão de alimentos contaminados tanto pelo processo de cozimento, que pode gerar PAHs, quanto pela deposição atmosférica. Estudos utilizando o benzo[a]pireno como parâmetro de medida de absorção de PAHs relatam que a quantidade deste composto ingerido por via oral é maior do que através da inalação, porém, os efeitos da contaminação por inalação são maiores, devido à formação de carcinomas bronquiais. Os compostos aderidos às partículas são parcialmente dissolvidos e metabolizados nos pulmões. Os metabólitos podem atuar sobre as células, ou entrar na corrente sanguínea e chegar ao fígado, ou pode ser excretado sem passar por qualquer modificação, pela bÍlis. Apenas uma pequena parte dos compostos é eliminada com a urina na forma metabolizada ou na forma original (Pereira Netto *et alli*, 2000; Ullman, 1989).

Os PAHs maiores e mais hidrofóbicos tendem a ser mais carcinogênicos, enquanto os PAHs menores (dois e três ciclos) não são carcinogênicos. Apesar disto, estudos recentes demonstram que sob o efeito da luz alguns deles podem se tornar extremamente tóxicos (Ankley *et alli*, 1995).

Os efeitos carcinogênicos dos PAHs são muito conhecidos, principalmente na pele e pulmão, e têm sido observados em pessoas expostas às inúmeras fontes emissoras existentes: exaustão de motores a diesel ou gasolina, queima de carvão, fotocopiadoras, fumaça de cigarro, exaustão de incineração de plantas e rejeitos, a fuligem de chaminé de vários processos industriais, queimadas no campo e nas florestas etc. (Ullman, 1989; Pereira Netto *et alli*, 2000).

O mecanismo da ação carcinogênica dos PAHs é complexo, e envolve uma série de enzimas que catalisam reações de oxidação, redução e hidrólise, pois a substância em si não é ativa, e sim seus metabólitos finais. Este mecanismo para benzo[a]pireno tem sido muito estudado, e pode ser demonstrado como a seguir (Ullman, 1989; Lehner *et alli*, 1996).

Primeiramente o composto é convertido num *óxido de areno*, um epóxido em que a aromaticidade de um dos anéis foi destruída, reação que pode ser observada na figura 39 (Morrison & Boyd, 1993).

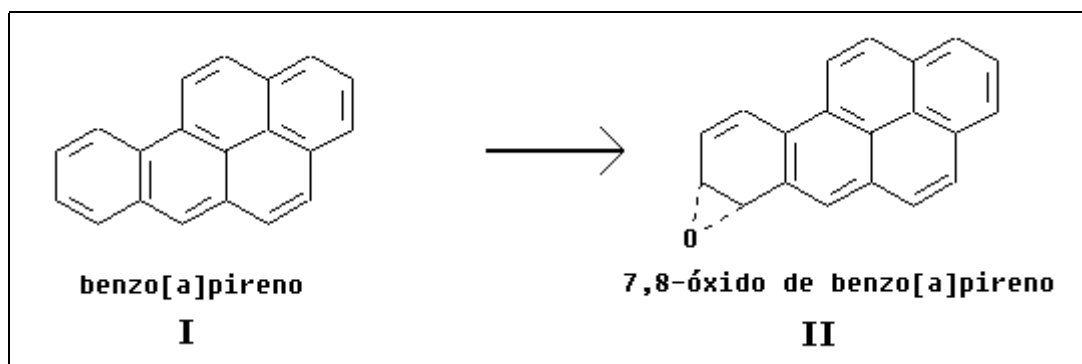


Figura 39: Conversão do benzo[a]pireno (I) em 7,8-óxido de benzo[a]pireno (II).

Em presença de água, ocorre a hidrólise (catalisada enzimaticamente) do anel do epóxido, com a formação de um diol-1,2, o benzo[a]pireno-dihidro-7,8-diol-7,8, reação que pode ser observada na figura 40 (Morrison & Boyd, 1993).

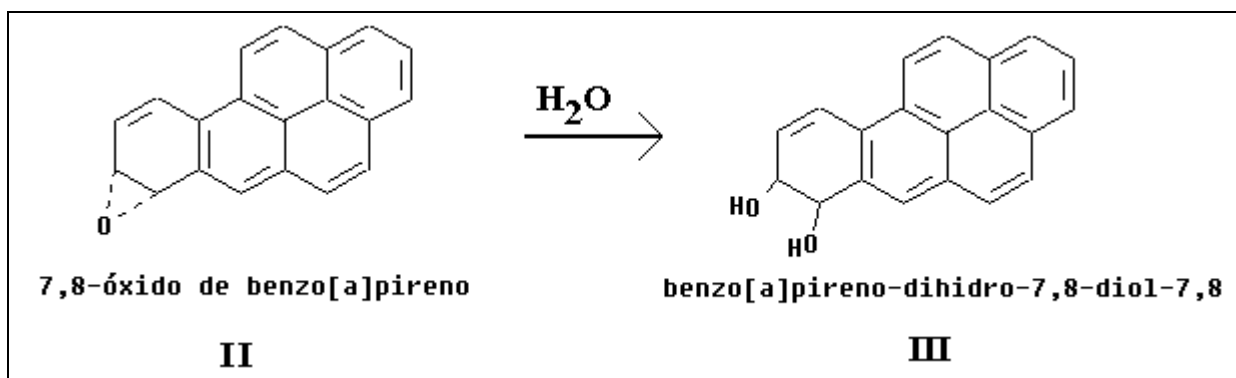


Figura 40: Conversão do benzo[a]pireno (II) em benzo[a]pireno-dihidro-7,8-diol-7,8 (III).

Alguns destes dióis sofrem novamente epoxidação com formação de dihidroxi-epóxidos (figura 41) que se acredita serem os verdadeiros compostos com ação carcinogênica (Morrison & Boyd, 1993).

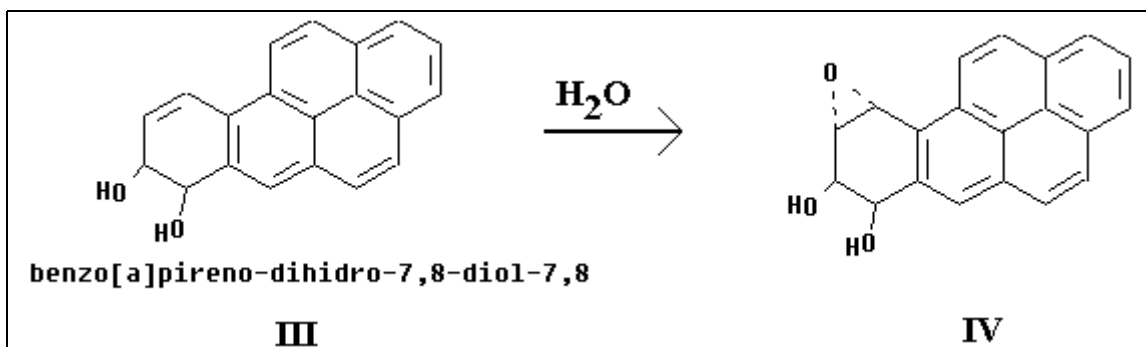


Figura 41: Formação do dihidroxi-epóxido (IV)

Em testes *in vitro*, ao se misturar o composto (IV) com o DNA, o grupo -NH_2 da base de guanina do nucleosídeo ataca o C-10 do epóxido e abre o anel com inversão química, formando o composto (V), observado na figura 42. Este grupo hidrocarboneto é muito volumoso, e sua ligação com a guanina impede-a de encaixar na hélice dupla do DNA e formar ligações de hidrogênio com uma citosina na extremidade oposta da hélice. Esta deterioração provoca mutações, e com isto, uma maior probabilidade de carcinogênese (Morrison & Boyd, 1993).

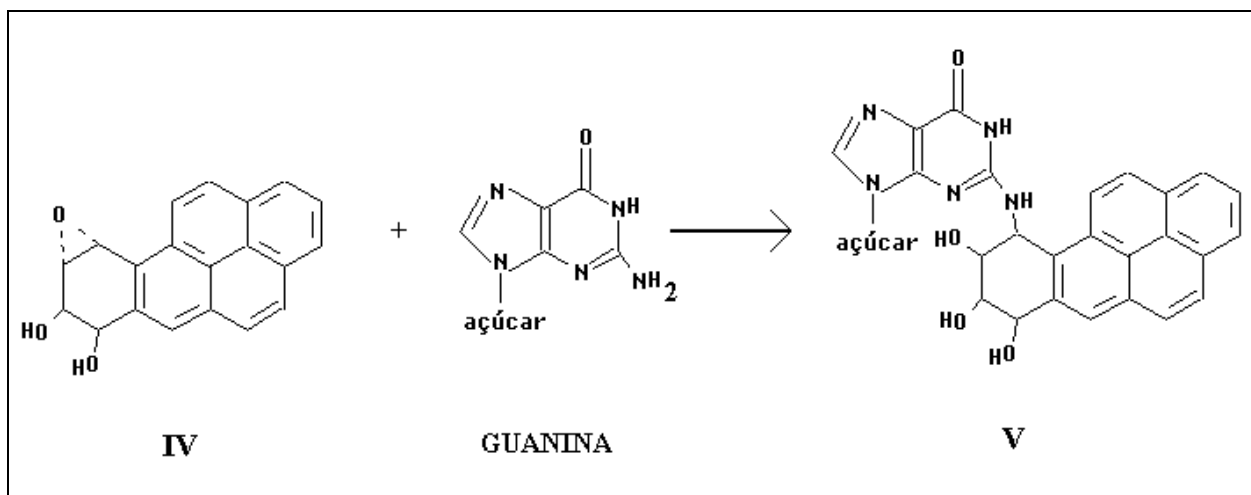


Figura 42: Conversão do composto (IV) no composto (V).

Neste trabalho foi estudado um conjunto de PAHs não substituídos contendo entre 2 e 7 anéis de 5 e 6 átomos de carbono, cujas estruturas, fórmulas moleculares e respectivos números de registro no *Chemical Abstracts* poderão ser observados mais adiante no quadro 16.

A Pesquisa dos Compostos

A pesquisa dos compostos foi feita basicamente através da utilização de 5 fontes:

- *Chemical Abstracts*
- *Illustrated Handbook of Physical-Chemical Properties and Environmental Fate for Organic Chemicals*, volumes 1, 2 e 3 (Mackay *et alli*, 1998).
- *Handbook of Data on Organic Compounds*, Vol. I e II (West & Astle, 1987).
- *ChemFinder.Com* – banco de dados acessível pela *Internet* para pesquisa de nome e estrutura dos compostos e suas propriedades físico-químicas, toxidez e cuidados no manuseio. A busca pode ser feita pela entrada dos nomes oficiais e comuns do composto, o número de registro no *CAS* ou a fórmula molecular.
- *Handbook of Environmental Data on Organic Chemistry* (Verschueren, 1996).

A primeira etapa do trabalho consistiu no levantamento e escolha das estruturas dos PAHs a serem utilizados nos cálculos. Inicialmente foi reunido um conjunto de PAHs não substituídos e algumas de suas propriedades foram procuradas no *Illustrated Handbook of Physical-Chemical Properties and Environmental Fate for Organic Chemicals* (Mackay *et alli*, 1998). Para alguns compostos, principalmente aqueles com número maior de anéis, não haviam valores determinados experimentalmente para algumas propriedades. Estes compostos foram utilizados então como o *conjunto de previsão*, isto é, aquelas amostras que teriam seus valores estimados através do modelo de regressão desenvolvido neste trabalho.

Das propriedades existentes na obra de Mackay *et alli* (1998) foi escolhida uma das que apresentaram o maior número de moléculas com valores determinados experimentalmente pelo mesmo método: o coeficiente de partição octanol-água. Outras três propriedades foram incluídas no conjunto de dados: O ponto de ebulição (Karcher *et alli*, 1988), o índice de retenção (Sander & Wise, 1986) e a fototoxidez (Mekenyan *et alli*, 1994), e novas moléculas, não constantes no *Handbook* de Mackay *et alli* (1998) foram acrescentadas.

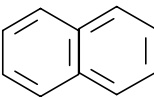
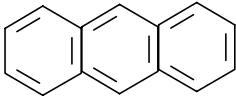
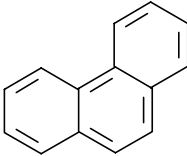
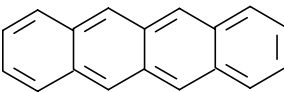
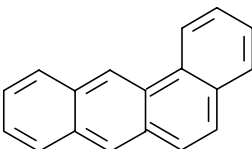
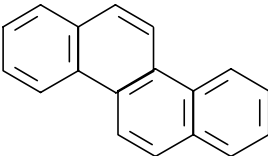
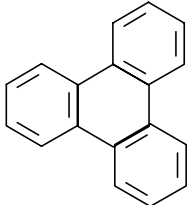
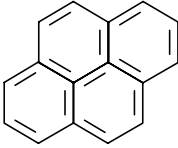
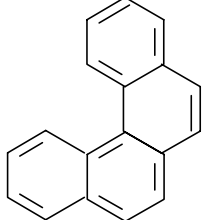
Estas duas últimas obras não apresentavam todas as informações sobre os compostos (estrutura, registro no CAS, e fórmula molecular), apenas seus nomes comuns e, em Mekenyan *et alli* (1994), as estruturas. Como citado anteriormente, pelo tipo de processo que dá origem a esta classe de compostos, existe uma imensa variedade de estruturas similares, ou com mesma fórmula molecular, e cada uma possui diferentes nomes e sinônimos muito parecidos, o que pode causar uma imensa confusão a respeito da identidade de cada composto.

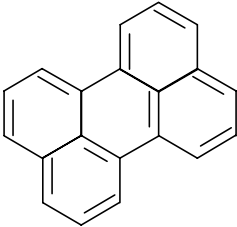
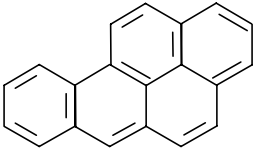
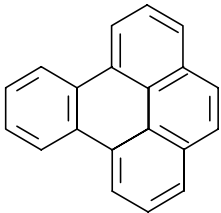
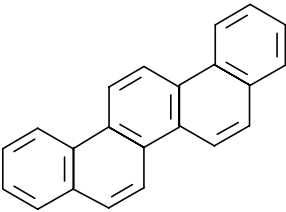
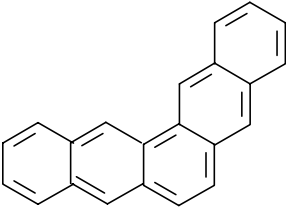
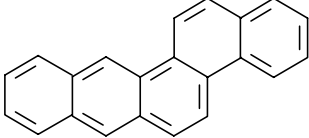
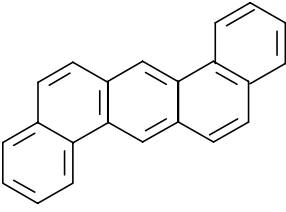
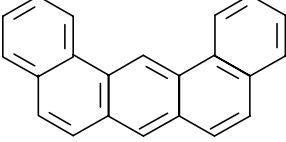
A segunda etapa da busca pelas estruturas consistiu então em pesquisar no *Chemical Abstracts* os nomes IUPAC de cada composto, seus sinônimos mais comuns, a estrutura molecular e o número de registro no *Chemical Abstracts* (número CAS), que funciona como o número de identidade do composto. Este não foi um trabalho trivial, pois muitas vezes o número de registro do CAS fornecido pelo handbook de Mackay *et alli* (1998) não correspondia no *Chemical Abstracts*, à mesma estrutura (caso do composto 57 e 67). Neste caso, outras fontes (fontes 3-6) foram consultadas para garantir a identidade de cada molécula colocada na lista, sempre considerando o número de registro no *Chemical Abstracts*.

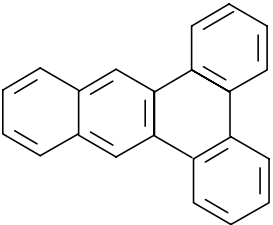
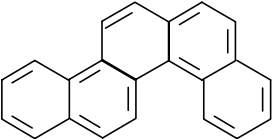
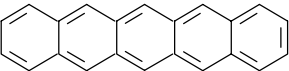
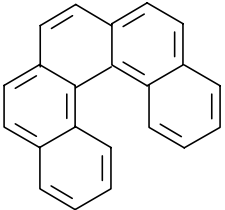
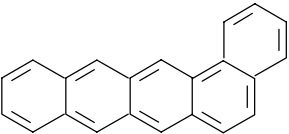
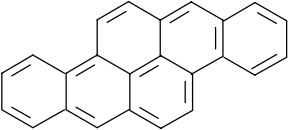
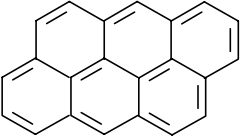
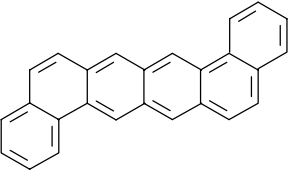
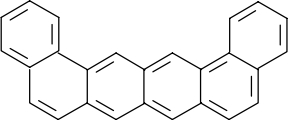
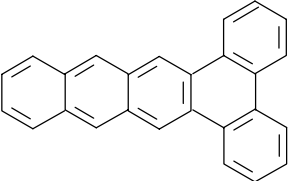
Os resultados na busca pelo site do *ChemFinder*, à partir do nome oficial destes compostos, foram encontrados dois números de registro no CAS para a mesma estrutura, como consta no quadro 16 mais adiante. Em alguns casos ainda, não foi encontrado o composto em nenhum dos bancos de dados utilizados (moléculas 27, 35, 38 e 51, do quadro 16. A molécula 51 foi retirada da publicação de Sander & Wise (1987), que contém os dados de índice de retenção e as moléculas 27, 35 e 38 foram retiradas do trabalho de Karcher *et alli* (1988) que contém os dados de ponto de ebulição. Em alguns casos, estas obras utilizaram o nome usual, e não o oficial, e por isto não foram encontradas em nenhum dos bancos de dados pesquisados. Mesmo assim, estas estruturas foram utilizadas dentro do *conjunto de previsão*.

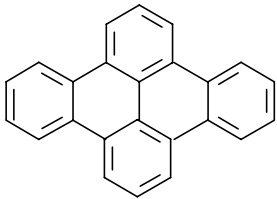
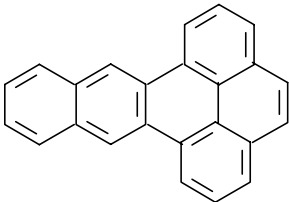
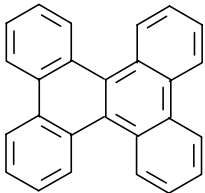
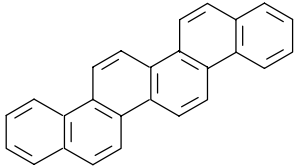
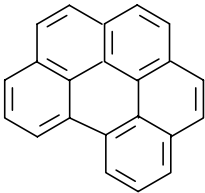
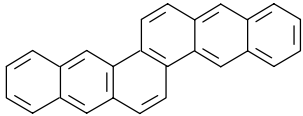
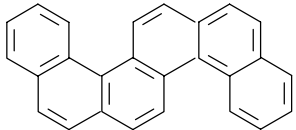
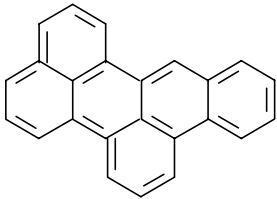
No quadro 16 pode ser observado o conjunto de moléculas escolhidas para o trabalho, bem como suas respectivas fórmulas moleculares, fórmulas estruturais, sinônimos mais freqüentes e número de registro no *Chemical Abstracts*.

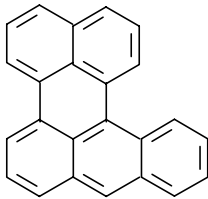
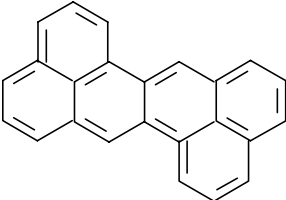
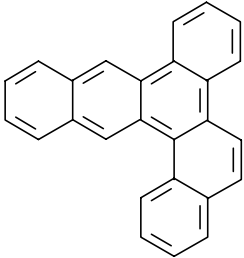
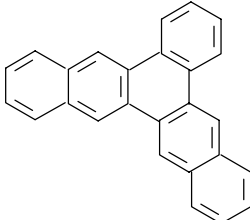
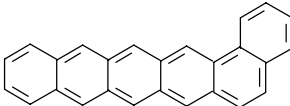
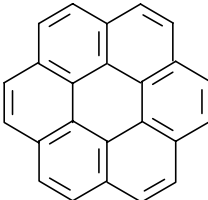
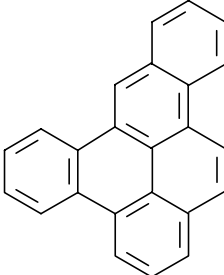
Quadro 16: Lista dos hidrocarbonetos aromáticos policíclicos utilizados neste trabalho

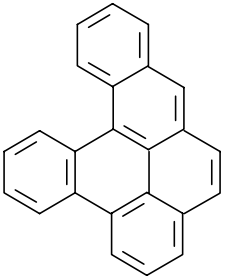
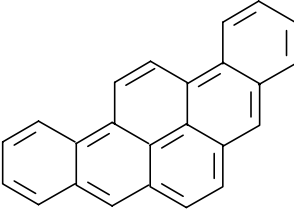
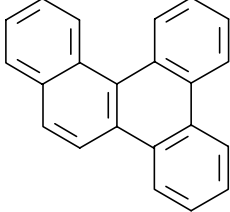
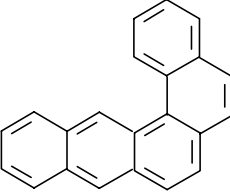
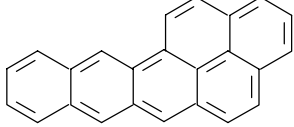
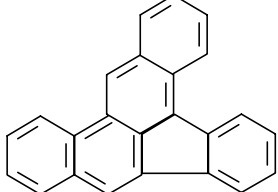
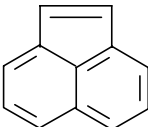
No	estrutura	FM	Nome no <i>Chemical Abstracts</i>	sinônimos mais comuns	No. CAS
1		C ₁₀ H ₈	naphthalene		91-20-3
2		C ₁₄ H ₁₀	anthracene		120-12-7
3		C ₁₄ H ₁₀	Phenanthrene		85-01-8
4		C ₁₈ H ₁₂	naphthacene	benzo[b]anthracene	92-24-0
5		C ₁₈ H ₁₂	benzo[a]anthracene	1.2-benzanthracene	56-55-3
6		C ₁₈ H ₁₂	chrysene		218-01-9
7		C ₁₈ H ₁₂	triphenylene		217-59-4
8		C ₁₆ H ₁₀	pyrene		129-00-0
9		C ₁₈ H ₁₂	benzo[c]phenanthrene		195-97-7

<u>No</u>	<u>estrutura</u>	<u>FM</u>	<u>Nome no <i>Chemical Abstracts</i></u>	<u>sinônimos mais comuns</u>	<u>No. CAS</u>
10		C ₂₀ H ₁₂	perylene		198-55-0
11		C ₂₀ H ₁₂	benzo[a]pyrene		50-32-8
12		C ₂₀ H ₁₂	benzo[c]pyrene	4.5-benzopyrene	192-97-2
13		C ₂₂ H ₁₄	picene	benzo[a]chrysene	213-46-7
14		C ₂₂ H ₁₄	pentaphene		222-93-5
15		C ₂₂ H ₁₄	benzo[b]chrysene		214-17-5
16		C ₂₂ H ₁₄	dibenzo[a, h]anthracene		53-70-3
17		C ₂₂ H ₁₄	dibenzo[a, j]anthracene		224-41-9

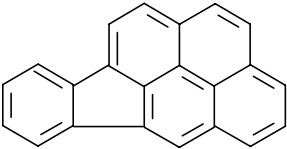
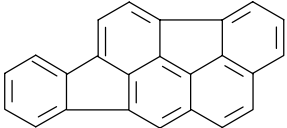
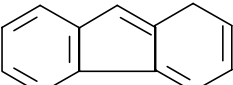
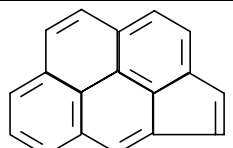
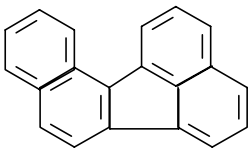
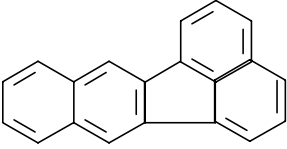
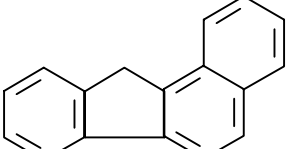
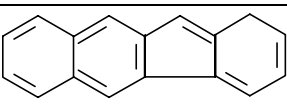
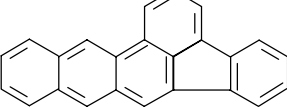
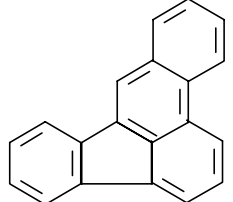
No	estrutura	FM	Nome no <i>Chemical Abstracts</i>	sinônimos mais comuns	No. CAS
18		C ₂₂ H ₁₄	benzo[b]triphenylene	dibenz[a.c]anthracene	215-58-7
19		C ₂₂ H ₁₄	benzo[c]chrysene	1.2.:5.6- dibenzophenanthrene	194-69-4
20		C ₂₂ H ₁₄	Pentacene		135-48-8
21		C ₂₂ H ₁₄	dibenzo[c. g]phenanthrene		188-52-3
22		C ₂₂ H ₁₄	benzo[a]naphthacene		226-88-0
23		C ₂₂ H ₁₄	dibenzo[b. def]chrysene		189-64-0
24		C ₂₂ H ₁₂	dibenzo[def. mno]chrysene	anthanthrene	191-26-4
25		C ₂₆ H ₁₆	dibenzo[a.j]naphthacene	dibenzo[a.j]naphthacene	227-04-3
26		C ₂₆ H ₁₆		dibenzo[a.l]naphthacene	
27		C ₂₆ H ₁₆	dibenzo[a.c]naphthacene		216-00-2

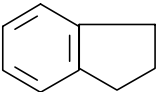
No	estrutura	FM	Nome no <i>Chemical Abstracts</i>	sinônimos mais comuns	No. CAS
28		C ₂₄ H ₁₄	dibenzo[el]naphthacene	Dibenzo[fg.op]naphthacene	192-51-8
29		C ₂₄ H ₁₄	dibenzo[de.gr]naphthacene		193-09-9
30		C ₂₆ H ₁₆	dibenzo[g.p]chrysene		191-68-4
31		C ₂₆ H ₁₆	benzo[c]picene		217-37-8
32		C ₂₂ H ₁₂	benzo[ghi]perylene	1.12-benzoperylene	191-24-2
33		C ₂₆ H ₁₆	dibenzo[b.k]chrysene		217-54-9
34		C ₂₆ H ₁₆		dibenzo[cl]chrysene	
35		C ₂₄ H ₁₄	benzo[b]perylene		197-70-6

No	estrutura	FM	Nome no <i>Chemical Abstracts</i>	sinônimos mais comuns	No. CAS
36		C ₂₄ H ₁₄	benzo[a]perylene		191-85-5
37		C ₂₄ H ₁₄		dibenzo[de,mn]naphthacene	
38		C ₂₆ H ₁₆	naphtho[2.3-g]chrysene	dibenzo[a,n]triphenylene	196-64-5
39		C ₂₆ H ₁₆	benzo[h]pentaphene		214-91-5
40		C ₂₆ H ₁₆	benzo[a]pentacene		239-98-5
41		C ₂₄ H ₁₄	coronene		191-07-1
42		C ₂₄ H ₁₄	naphtho[1.2.3.4-def]chrysene	dibenzo[a,e]pyrene	192-65-4

No	estrutura	FM	Nome no <i>Chemical Abstracts</i>	sinônimos mais comuns	No. CAS
43		C ₂₄ H ₁₄	dibenzo[def,p]chrysene	dibenzo[a,l]pyrene	191-30-0
44		C ₂₄ H ₁₄	benzo[rst]pentaphene	dibenzo[a,i]pyrene	189-55-9
45		C ₂₂ H ₁₄	benzo[g]chrysene	dibenzo[b,q]phenanthrene	196-78-1
46		C ₂₂ H ₁₄	2,3:5,6-Dibenzophenanthrene		195-06-2
47		C ₂₄ H ₁₄	naphtho[2.1.8-ql]naphthacene		196-42-9
48		C ₂₄ H ₁₄	dibenzo[a,e]aceantrylene	dibenzo[a,e]fluoranthene	5385-75-1
49		C ₁₂ H ₈	Acenaphthylene	Acenaphtelene	208-96-8

No	estrutura	FM	Nome no <i>Chemical Abstracts</i>	sinônimos mais comuns	No. CAS
50		C ₂₄ H ₁₄	dibenzo[a,k]fluoranthene		
51		C ₂₄ H ₁₄	naphtho[2,3-k]fluoranthene		207-18-1
52		C ₁₆ H ₁₀	fluoranthene		206-44-0
53		C ₂₂ H ₁₂		dibenzo[k,mno]fluoranthene	
54		C ₁₂ H ₁₀	1,2-dihydroacenaphthylene	acenaphthene	83-32-9
55		C ₁₃ H ₁₀	9H – fluorene	Fluorene	86-73-7
56		C ₁₇ H ₁₂	benzo[b]fluorene	11H-Benzo[b]fluorene	243-17-4
57		C ₁₇ H ₁₂	benzo[c]fluorene		205-12-9
58		C ₁₈ H ₁₀	benzo[ghi]fluoranthene		203-12-3
59		C ₂₀ H ₁₂	benzo[a]aceanthrylene	benzo[a]fluoranthene	203-33-8

No	estrutura	FM	Nome no <i>Chemical Abstracts</i>	sinônimos mais comuns	No. CAS
60		C ₂₂ H ₁₂	indeno[1.2.3.cd]pyrene		193-39-5
61		C ₂₂ H ₁₂	indeno[1.2.3-cd] fluoranthene		193-43-1
62		C ₁₃ H ₁₀	1H-fluorene		244-36-0
63		C ₁₈ H ₁₀	cyclopenta[cd]pyrene		27208-37-3
64		C ₂₀ H ₁₂	benzo[j]fluoranthene		205-82-3
65		C ₂₀ H ₁₂	benzo[k]fluoranthene		207-08-9
66		C ₁₇ H ₁₂	benzo[a]fluorene	11H-Benzo[a]fluorene	238-84-6
67		C ₁₇ H ₁₂	1H-benzo[b]fluorene		14458-76-5
68		C ₂₄ H ₁₄	dibenz[e.k]acephenanthrylene		206-06-4
69		C ₂₀ H ₁₂	benzo[b]fluoranthene		205-99-2

<u>No</u>	<u>estrutura</u>	<u>FM</u>	<u>Nome no <i>Chemical Abstracts</i></u>	<u>sinônimos mais comuns</u>	<u>No. CAS</u>
70		C ₉ H ₁₀	indan		496-11-7

Propriedades Moleculares

As propriedades moleculares não necessárias para estimar o comportamento dos contaminantes no meio ambiente, por exemplo, o tempo necessário para que um composto sofra degradação por intemperismo é diferente para cada espécie química. Algumas são mais estáveis, outras são rapidamente degradáveis. A dispersão de um composto no ar, por exemplo, está ligada à pressão de vapor (Burkard *et alli*, 1985), a tendência à aderir ao sedimento ou se dispersar no meio aquoso depende do coeficiente de partição octanol-água (Manahan, 1993). Isto se aplica também ao entendimento de como uma droga atua nos organismos, e por isto o estudo das propriedades moleculares é tão importante em química ambiental e medicinal.

Este trabalho aborda quatro propriedades moleculares de interesse: o ponto de ebulição, o coeficiente de partição octanol-água, o índice de retenção e a fototoxicidez.

Ponto de Ebulição (pe)

O ponto de ebulição é uma das principais propriedades físico-químicas usadas para caracterizar e identificar uma substância pura, e é descrito como sendo a temperatura em que a pressão de vapor de um líquido se iguala à pressão externa, permitindo assim a evaporação desta substância. Em termos termodinâmicos, o ponto de ebulição de um composto é a temperatura em que a energia térmica das partículas vence as forças de coesão que permitem mantê-las no estado de agregação que caracteriza o líquido (Atkins, 1994).

Estas forças de coesão, ou forças intermoleculares, estão intimamente relacionadas à natureza estrutural da molécula. Elas apresentam-se sob duas formas: as *forças de van der Waals*, observada em moléculas apolares (é o caso dos PAHs) e a

interação entre dipolos, que ocorre em moléculas polares (Morrison & Boyd, 1993). Estas forças variam em função do tamanho e da geometria molecular e estas diferenças também se refletem nas suas propriedades físico-químicas. Por exemplo: moléculas maiores tendem a apresentar pontos de ebulição mais elevados, pois apresentam uma área superficial de contato maior que conseqüentemente favorece a formação destas interações.

O ponto de ebulição é geralmente muito fácil de se determinar experimentalmente no entanto, quando um composto não está disponível, ou seu manuseio envolve um grande risco à saúde, um procedimento para prever o ponto de ebulição é muito útil e já tem sido aplicado com sucesso em muitos trabalhos. Um exemplo deste tipo de aplicação é o trabalho de Arupjyoti & Iragavarapu (1998), que construíram modelos de regressão utilizando redes neurais para prever o ponto de ebulição de alcanos e alcoóis, com desvios entre 5 e 9°C.

Coeficiente de Partição Octanol-Água (Log K_{ow} ou Log P)

O coeficiente de partição octanol-água é um parâmetro que mede a *lipofilicidade* (ou sob outro ponto de vista a *hidrofobicidade*) de um composto, é a estimativa da solubilidade em fase aquosa ou orgânica (em geral usa-se *n*-octanol). Este coeficiente é a razão entre a concentração do composto na fase orgânica e a concentração do composto na fase aquosa, como descreve a equação 30. Como o valor de K_{ow} é muito pequeno, costuma-se expressá-lo na sua forma logarítmica (Kubinyi, 1993). Os PAHs são muito hidrofóbicos, então a concentração na fase orgânica é muito grande, resultando em altos valores de K_{ow} , da ordem de 10^5 ou 10^6 .

$$K_{ow} = \frac{C_{org}}{C_{aq}} \quad (30)$$

Compostos que apresentam $K_{ow} > 1$ são *lipofílicos* (ou hidrofóbicos) e compostos que apresentam $K_{ow} < 1$ são *hidrofílicos* (van de Waterbeemd, 1995).

A determinação destas concentrações pode ser feita experimentalmente por meio de diversos métodos. Neste trabalho utilizamos Log K_{ow} medido pelo método *shake*

flask, em que o composto é dissolvido nas duas fases, por agitação, e as concentrações nas duas fases são determinadas por métodos gravimétricos ou instrumentais (espectroscopia de UV, análises cromatográficas, fluorescência e outros), após o equilíbrio. A razão C_{org}/C_{aq} é então calculada (Mackay *et alli*, 1998) .

Kubinyi (1993) cita uma série de vantagens da utilização do *n*-octanol para estimar a lipofilicidade dos compostos:

- Sua estrutura apresenta uma longa cadeia carbônica com um grupo polar na ponta, que favorece o transporte através das membranas biológicas;
- A presença do grupo hidroxila, capaz de formar ligações de hidrogênio e deste modo interagir com uma infinidade de grupos polares de diferentes solutos;
- Apesar de seu caráter lipofílico, o *n*-octanol dissolve muito mais compostos orgânicos que os solventes alcanos, cicloalcanos ou aromáticos;
- Enquanto a fase aquosa deste sistema não contém praticamente nenhum octanol dissolvido, a fase octanol dissolve uma significativa quantidade de água (2.3 mol.L^{-1} , correspondente à razão molar de *n*-octanol/água de aproximadamente 4/1). Desta forma, os grupos polares não precisam ser desidratados ao serem transferidos da fase aquosa para a fase orgânica;
- *n*-octanol tem uma baixa pressão de vapor, de forma que as medidas podem ser reproduzidas com facilidade, mas por outro lado sua pressão de vapor é alta o suficiente para permitir sua remoção sob condições moderadas;
- *n*-octanol não absorve numa larga faixa de radiação UV, tornando a determinação quantitativa de muitos compostos relativamente fácil;

O $\log K_{ow}$ é essencial para se entender os mecanismos de transporte e distribuição de um composto sobre o meio em que ele está inserido. Por exemplo, no caso de uma droga, este parâmetro será necessário para se entender seu transporte e distribuição através dos sistemas biológicos, como por exemplo as membranas biológicas, a absorção gastrointestinal, ou a transferência através da barreira cérebro-sangue. No caso de um poluente aquático, é este parâmetro que vai determinar se o composto irá se depositar no sedimento, nos tecidos dos organismos, ou se vai se dispersar na água.

O Log K_{ow} é utilizado ainda para estimar a degradação das substâncias no meio ambiente. Os hidrocarbonetos que são mais *hidrossolúveis* desaparecem primeiro, enquanto que aqueles mais *lipossolúveis* permanecem por mais tempo aderidos ao sedimento.

Verhaar *et alli* (1996) utilizaram o Log K_{ow} para classificar a toxidez aquática de alguns poluentes para as espécies *Pimephales promelas* e *Poecilia reticulata* utilizando métodos de QSAR; Croin & Schultz (1997) utilizaram o Log K_{ow} para estimar o potencial tóxico de alguns compostos orgânicos para a espécie *Vibrio fisheri*, também através de QSAR.

Índice de Retenção (IR)

Existe uma imensa variedade de PAHs no meio ambiente. Grande parte deles possui características estruturais e propriedades muito similares, sendo comum a existência de estruturas isoméricas. Muitas vezes, certos isômeros são mais carcinogênicos ou mutagênicos do que outros, e a separação destes isômeros é necessária para se avaliar/estimar com precisão os efeitos da exposição às fontes poluentes.

Um dos métodos que têm sido utilizados para separação de isômeros e análise de PAHs é a cromatografia líquida (CL) de fase reversa sobre fases estacionárias quimicamente modificadas com octadecilsilano (C_{18}), que tem sido empregada com sucesso na resolução de misturas complexas de PAHs (Sander & Wise, 1986).

Como resultado da excelente seletividade da cromatografia líquida de fase reversa sobre fase de C_{18} , esta técnica tem sido especificada pela EPA⁴³ como o método de determinação de 16 tipos de PAHs em efluentes aquosos (Sander & Wise, 1986).

O mecanismo de retenção em cromatografia líquida é um complexo processo que envolve as interações entre a fase móvel e a fase estacionária, no caso, a amostra contendo os hidrocarbonetos e a coluna cromatográfica por onde a fase móvel tem que

⁴³ EPA – U. S. Environmental Protection Agency.

passar, respectivamente. Estas interações são muito difíceis de serem descritas em detalhes, mas estão intimamente relacionadas à estrutura molecular das duas fases.

Alguns modelos têm sido desenvolvidos para permitir o entendimento dos mecanismos que envolvem a separação das misturas cromatográficas, e alguns parâmetros são estimados para se quantificar esta separação. Estes parâmetros são os tempos (t_r) e índices (I_r) de retenção.

Sander & Wise (1987) utilizaram o índice de retenção para estudar a seletividade da cromatografia líquida (CL) de fase reversa sobre fases estacionárias quimicamente modificadas com octadecilsilano (C_{18}), para separação de misturas de hidrocarbonetos aromáticos policíclicos. Este índice de retenção foi determinado a partir da equação 31, em que o volume de eluição do soluto foi medido simultaneamente com o volume de eluição de alguns padrões representativos dos PAHs com número de anéis entre 1 e 5. Os padrões escolhidos foram: benzeno, naftaleno, fenantreno, benzo[a]antraceno e benzo[b]criseno e dibenzo[b.def]criseno.

$$\log I_x = \log I_n = \frac{\log R_x - \log R_n}{\log R_{(n+1)} - \log R_n} \quad (31)$$

Em que x representa o soluto e n e $n + 1$ representam respectivamente os menores e maiores padrões de eluição, e R representa o volume de eluição corrigido. Para o cálculo destes índices, considera-se os seguintes valores para os padrões: benzeno = 1; naftaleno = 2; fenantreno = 3; benz[a]antraceno = 4; benzo[b]criseno = 5; dibenzo[b.def]criseno = 6.

Os fatores que determinam a ordem de eluição dependem da interação de cada composto com a fase estacionária, e estas interações estão relacionadas à estrutura molecular das espécies envolvidas. Por exemplo, compostos maiores tendem a ter maiores índices de retenção, pois possuem uma maior área superficial, que permite interações mais fortes com a fase estacionária e conseqüentemente uma demora no tempo de eluição em relação aos compostos menores.

Sander & Wise compararam as seletividades características de PAHs não substituídos para diferentes colunas comerciais do tipo C_{18} de vários fornecedores,

através da técnica de cromatografia líquida de fase reversa. Os índices de retenção observados para cada coluna foram diferentes, devido às variações nos substratos de sílica utilizados como suporte e à variedade de procedimentos de produção de cada coluna. Os melhores resultados para a separação de misturas de PAHs foram obtidos utilizando colunas com fase estacionária polimérica (por exemplo, alguns isômeros não eram separados através da utilização de substratos monoméricos e oligoméricos). Os índices de retenção obtidos com esta coluna foram utilizados para a construção de um modelo de previsão deste índice para outros hidrocarbonetos aromáticos policíclicos.

Fototoxidez

Fototoxidez é a toxidez apresentada por alguns compostos ao serem expostos à radiação visível ou UV (Schaeffer, FAIC & Larson, 1999). Se considerarmos que tudo que está na superfície da terra está exposto à radiação do sol, podemos ter uma noção da importância deste efeito na avaliação do impacto ambiental de alguns compostos, dependendo é claro, das condições do ambiente a que estamos nos referindo. Por exemplo, se um composto fototóxico está contaminando o sedimento à 100 m de profundidade, onde luz solar não chega, este composto não apresenta risco às populações que habitam este local. E para que este composto possa permanecer no fundo deste corpo aquático, e não da superfície, é preciso que suas propriedades físico-químicas favoreçam este comportamento, e isto está relacionado às estruturas moleculares.

Mas voltando ao composto fototóxico, suponha que devido às propriedades deste composto, ele contamine um peixe, acumulando-se nos tecidos biológicos. Este peixe tem por hábito subir periodicamente à superfície da água, na região atingida pela luz solar. É fácil concluir que existe uma probabilidade muito grande deste peixe sofrer os efeitos da toxidez fotoinduzida do referido composto, dependendo é claro, da quantidade do composto que tenha contaminado o peixe.

A *fototoxicologia* é o estudo de como a luz afeta a toxidez dos compostos, e é de particular interesse em química medicinal e química ambiental. É também uma área de

estudo interdisciplinar, pois agrega conceitos de diversos campos de estudo, como física, química e biologia (Schaeffer, FAIC & , 1999).

As radiações eletromagnéticas, na forma de radiação ultravioleta – UV (290-400 nm)⁴⁴ e visível (400-780 nm), fornecem energia capaz de modificar a estrutura química dos compostos, e até mesmo iniciar algumas reações químicas. Para que uma reação ocorra, é necessário que a molécula absorva uma certa quantidade de energia capaz de quebrar alguma ligação, e nos casos em que a fototoxidez ocorre, modificar sua estrutura gerando espécies biologicamente ativas (Schaeffer, FAIC & Larson, 1999).

Um exemplo muito conhecido do efeito tóxico induzido pela luz é a fotodermatose provocada na pele por vegetais cítricos como o limão, a lima e o salsão, quando o indivíduo se expõe ao sol. Estas plantas contêm furocumarinas, substâncias que quando expostas à luz causam manchas e erupções na pele (Weber *et alli*, 1999; Schaeffer, FAIC & Larson, 1999). Outros compostos, como xantenos e porfirinas tornam-se tóxicos ao absorverem radiação UV/visível, e são utilizados como inseticidas (Amor & Jori, 2000).

Mas existem inúmeros outros processos que dependem da luz, e que não possuem este caráter tóxico, como é o caso da fotossíntese realizada pelas plantas e algas, e sem a qual o ciclo do oxigênio no planeta não seria possível; ou como a produção da vitamina D pelos animais vertebrados, necessária ao seu metabolismo.

A exposição dos PAHs à luz do sol resulta numa rápida modificação estrutural que ocorre geralmente através de reações de oxidação. Em geral, os produtos desta modificações são mais tóxicos do que os compostos de origem (Mallakin *et alli*, 2000).

Alguns trabalhos sobre a fototoxidez de PAHs têm sido publicados mas os efeitos e mecanismos ainda não são muito conhecidos. Sabe-se apenas que a toxidez ocorre devido ao sistema de orbitais- π dos PAHs, que absorvem fortemente na região do UV e do visível (Hatch & Burton, 1999; Peachey & Crosby, 1996; Chen *et alli*, 1996).

Mallakin *et alli* (2000) utilizou o antraceno, um conhecido PAH, para estudar uma possível rota de modificação sob os efeitos da radiação solar simulada. Sob os efeitos desta radiação, os orbitais- π do anel central do antraceno são menos estáveis,

tornando-se suscetível ao ataque do oxigênio molecular, sofrendo uma série de reações de foto-oxidação e gerando mais de vinte subprodutos, dos quais alguns apresentaram grande potencial tóxico. Os produtos desta modificação foram analisados utilizando-se cromatografia líquida de alta eficiência.

Pode-se dizer que a fototoxidez destes compostos ocorre em função das propriedades moleculares que afetam sua bioacumulação nos organismos, sua habilidade de absorver a luz solar, e as características do seu estado excitado triplete. Estes estados triplete são de particular interesse em química ambiental, pois são estas espécies que interagem com o oxigênio molecular e geram espécies com oxigênio reativo, que causam a citotoxicidade (Chen *et alli*, 1996).

No meio aquático, os PAHs hidrofóbicos podem acumular-se nos tecidos biológicos, e se tornarem tóxicos quando expostos à radiação UV e visível (Hatch & Burton Jr., 1999). Alguns estudos sobre a toxidez do fluoranteno em organismos aquáticos demonstram que organismos expostos à concentrações da ordem de 5µg/L apresentam fototoxidez somente quando expostos a radiação UV muito intensa (75,2 µW/cm²), independente do tempo de exposição (Ankley *et alli*, 1995) .

Esta propriedade depende diretamente da dose do composto e da intensidade da radiação absorvida, como observou Ankley *et alli* (1995) através de um estudo sobre os efeitos da intensidade da luz sobre a fototoxidez do fluoranteno em macroinvertebrados bênticos. Seus resultados concordam com vários outros estudos sobre o tema (Hatch & Burton, 1999; Schaeffer, FAIC & Larson, 1999; Mekenyan *et alli*, 1994; Peachy & Crosby, 1996).

A fototoxidez de compostos tem sido levada em consideração pela EPA⁴⁵ para elaborar os critérios de proteção da qualidade da água (Ankley *et alli*, 1995) .

Os dados de fototoxidez utilizados neste estudo de QSPR foram retirados do trabalho de Mekenyan *et alli* (1994), em que os autores avaliaram a relação entre os orbitais de fronteira de um conjunto de PAHs e fatores como a persistência, absorção de luz e a toxidez fotoinduzida destes compostos. Os autores propuseram um mecanismo de explicação para estimar a correlação entre a toxidez fotoinduzida dos PAHs para a

⁴⁴ A radiação UV é subdividida em UV-A (320-400 nm) e UV-B (290-320 nm).

⁴⁵ EPA – U. S. Environmental Protection Agency

espécie *Daphnia magna* com vários parâmetros como energias dos estados triplete e singlete, energia dos orbitais de fronteira (HOMO e LUMO) e o intervalo (GAP) entre estes orbitais, calculados pelos métodos PM3 e AM1. O tamanho do GAP ($E_{LUMO} - E_{HOMO}$) é proporcional à energia do fóton absorvido, como pode ser observado na relação dada pela equação 32. Desta forma, espécies com um grande intervalo entre os orbitais de fronteira absorvem em comprimentos de onda menores, que implicam em uma quantidade de energia maior e por isto poderão apresentar um aumento na fototoxicidez (Mekenyan *et alli*, 1994).

$$E_{LUMO} - E_{HOMO} = h\nu \quad (32)$$

Em que

E_{HOMO} = energia do HOMO

E_{LUMO} = energia do LUMO

h = constante de Planck (6.62608×10^{-34} J s)

ν = (frequência de absorção $\cong 1/\lambda$, em que λ é o comprimento de onda)

Descritores Moleculares

A estrutura molecular de um composto pode ser descrita por uma série de parâmetros ou descritores moleculares, que fornecem informações a respeito do formato, tamanho, simetria dos compostos, distribuição dos átomos, que podem ser físico-químicos (descrevendo parâmetros hidrofóbicos, estéricos ou eletrônicos), estruturais, topológicos, eletrônicos e geométricos, dependendo das características que se deseja descrever (van de Waterbeemd, 1995).

Descritores moleculares representam uma importante ferramenta para prever as propriedades das substâncias, para classificar estruturas químicas ou procurar similaridades entre elas. Para este estudo foram calculados alguns descritores topológicos, eletrônicos e geométricos: energia de HOMO (orbital ocupado de maior energia), energia de LUMO (orbital desocupado de menor energia), a diferença entre as energias do HOMO e do LUMO (GAP), dureza molecular, momento de dipolo,

eletronegatividade molecular, polarizabilidade, cargas, ovalidade, propriedades termodinâmicas, volume, área, além de 151 índices topológicos.

Será feito um breve comentário sobre cada tipo de descritor utilizado neste trabalho, que tem o intuito de ser apenas ilustrativo e não extensivo, no sentido de situar o leitor que não tenha familiaridade com o assunto.

Todos os descritores moleculares investigados neste trabalho encontram-se listados no apêndice I.

Os descritores eletrônicos

Estes descritores são estimados à partir dos cálculos de orbitais moleculares, e podem ser descritos por uma ou várias quantidades para cada molécula. Este tipo de descritor é capaz de descrever cada sítio da molécula, por exemplo, a distribuição de cargas para cada átomo na estrutura, e para tanto irá render vários índices. Os descritores eletrônicos utilizados neste trabalho foram: cargas, momento de dipolo molecular, orbitais de fronteira (HOMO e LUMO), eletronegatividade, dureza/moleza molecular e polarizabilidade. A seguir serão feitos alguns comentários a respeito de alguns destes descritores.

Cargas (q^+/q^-)

A densidade de cargas de uma estrutura molecular descreve a distribuição dos elétrons, que é responsável pelo comportamento químico de cada espécie. Estimar a magnitude desta carga não é tarefa trivial, pois uma molécula não é um sistema estático, não é um mero arranjo de prótons com cargas positivas e elétrons com carga negativa, que pode ser explicado em toda a sua complexidade através da simples utilização de equações da mecânica clássica. Cada molécula é um sistema dinâmico, em constante mudança, e para estimar parâmetros como as cargas são necessárias algumas aproximações, como a aproximação de *Born-Oppenheimer* (Hehre *et alli*, 1993). No entanto, o tratamento dos elétrons é um pouco mais complicado, eles podem ser vistos como uma distribuição de cargas negativas. A região abrangida por esta nuvem

eletrônica corresponde à superfície de van der Waals, e contém certa de 90 a 95% dos elétrons de toda a distribuição em um sistema molecular (Hehre *et alli*, 1993).

Com a utilização desta aproximação, a contribuição do núcleo para a carga total sobre um átomo em uma molécula é simples, basta levar em consideração o número atômico. No entanto, a contribuição dos elétrons é mais complicada e requer o entendimento de como os elétrons estão associados a estes átomos individualmente. A carga atômica em uma molécula pode ser calculada em termos de *distribuição* de carga atômica ao longo de todos os centros atômicos, e para tanto existe uma variedade de métodos para realizar estes cálculos. Neste trabalho foram utilizados os dois mais populares: o método de Mülliken (Mülliken, 1955), para cargas particionadas, e as cargas calculadas à partir do potencial eletrostático (Chirlian & France, 1987; Breneman & Wilberg, 1990).

Momento de dipolo (μ)

O momento de dipolo é um parâmetro relacionado à distribuição de cargas numa molécula, e tem sido largamente empregado na explicação de propriedades físico-químicas e reatividade molecular. O momento de dipolo resultante (μ) é soma dos dipolos locais (representados pelo vetor μ) e deve ser zero para moléculas neutras. Este momento de dipolo local é definido pela magnitude das cargas positiva e negativa multiplicadas pela distância entre elas (Hehre *et alli*, 1993).

Quando se trata de compostos que apresentam ligações covalentes, cada ligação química é feita através do compartilhamento dos elétrons por um par de átomos. Se estes átomos são diferentes (ou se são iguais mas o ambiente em que eles estão são diferentes), este compartilhamento permite uma deformação da nuvem eletrônica (caráter negativo) para um dos lados, favorecendo a formação de um caráter positivo na extremidade oposta. Costuma-se dizer que a ligação torna-se *polarizada*, ou seja, ocorre a formação de dois *pólos*: um positivo e um negativo, que podem ser representados na forma de um vetor ao longo do eixo da ligação (Hehre *et alli*, 1993).

Este dipolo é estimado em função da geometria molecular, e pode oferecer indicações a respeito da reatividade ou das forças intermoleculares que determinam o

estado de agregação das mesmas. O cálculo do momento dipolar total de uma molécula é o vetor soma dos momentos de dipolo das ligações individuais e costuma ser expresso em Debye (D).

Neste trabalho foram calculados dois tipos diferentes de dipolos moleculares: o momento de dipolo através do valor esperado e o dipolo calculado a partir do potencial eletrostático.

Energia do HOMO, Energia do LUMO e GAP

Estes parâmetros dizem respeito às energias dos orbitais de fronteira de uma molécula: o orbital ocupado de maior energia (HOMO⁴⁶), o orbital vazio de menor energia (LUMO⁴⁷) e o intervalo entre estes dois orbitais (equação 33), que estão relacionados ao potencial de ionização (PI) e à afinidade eletrônica (AE) da molécula, através das relações dadas nas equações 34 e 35 (Atkins, 1993; Shriver *et alli*, 1994).

$$GAP = E_{LUMO} - E_{HOMO} \quad (33)$$

$$PI = -E_{HOMO} \quad (34)$$

$$AE = -E_{LUMO} \quad (35)$$

Dureza e Moleza Molecular (η)

A média da separação entre os orbitais de fronteira, ou seja, o intervalo entre as energias do HOMO e do LUMO, é a medida da dureza de um átomo. Quando este intervalo é pequeno, uma pequena quantidade de energia é capaz de elevar um elétron do HOMO para o LUMO. Isto significa que a distribuição de elétrons é facilmente

⁴⁶ Highest Occupied Molecular Orbital

⁴⁷ Lowest Unoccupied Molecular Orbital

rearranjada por uma perturbação, então costuma-se dizer que a molécula é *mole* (pequeno intervalo entre HOMO e LUMO).

Se o contrário ocorre, ou seja, se o intervalo entre os orbitais de fronteira é grande, significa que será preciso uma energia maior para perturbar este sistema. Costuma-se dizer neste caso que a molécula é dura. A dureza (η) de uma molécula é calculada pela equação 36, em que PI representa o potencial de ionização e AE representa a eletroafinidade (Shriver *et alli*, 1994).

$$\eta = \frac{1}{2}(PI - AE) \quad (36)$$

Da mesma forma, é possível estimar a dureza (ou a moleza) de uma molécula. Ela será expressa em termos de energia de ionização entre o átomo neutro e o seu ânion. Neste trabalho não foram feitos os cálculos para o ânion, então foi utilizada uma aproximação em que a eletroafinidade (AE) foi estimada como sendo a energia do LUMO.

Polarizabilidade (α)

A polarizabilidade é a medida da susceptibilidade da molécula ser polarizada quando sofre uma perturbação externa, e é fornecida pela constante de proporcionalidade α , ilustrada na equação 37. A presença de um campo elétrico pode induzir a formação de dipolos numa molécula, que são proporcionais ao campo de força aplicado, segundo a equação aproximada 37 (Atkins, 1993).

$$\mu = \alpha E \quad (37)$$

A *polarizabilidade* é justamente a medida de quanto a molécula permite a formação dos dipolos, isto é, quanto ela pode ser polarizada, e é fornecida pela constante de proporcionalidade α . Quanto maior o valor de α , maior o momento de dipolo formado sob efeito de um campo elétrico. Se este campo é muito forte, esta

relação não é linear, e poderá ser descrita segundo a equação 38, onde β é a *hiperbolarizabilidade* (Atkins, 1993).

Quando o campo de força é expresso em Volts por metro (Vm^{-1}), a polarizabilidade é expressa em unidades de $(\text{Coulomb-metro})^2$ por joule ($\text{C}^2\text{m}^2\text{J}^{-1}$) (Atkins, 1993).

$$\mu = \alpha E + \frac{1}{2} \beta E^2 + \dots \quad (38)$$

A polarizabilidade está intimamente relacionada com a dureza de uma espécie (e por consequência com seus orbitais de fronteira). Uma espécie *dura* é dificilmente modificada por uma perturbação, então dificilmente será polarizada. Já uma espécie *mole* é facilmente polarizada (Shriver *et alli*, 1994).

Os descritores topológicos

Como foi enfatizado por diversas vezes neste texto, para se entender as diferenças no comportamento das moléculas, é fundamental conhecer as peculiaridades de cada estrutura química. Não basta definir o arranjo estrutural, número e tipo dos átomos para descrever exatamente as diferenças e peculiaridades entre uma molécula e outra. Dentro de um grupo de isômeros, por exemplo, o número e o tipo de átomos é o mesmo, e em alguns tipos de isomeria, até mesmo a estrutura, mas estas diferenças precisam ser identificadas no estudo das suas propriedades. Uma das dificuldades encontradas é justamente encontrar uma forma de representar esta estrutura em termos numéricos, de forma que cada detalhe na estrutura molecular seja descrito.

Uma maneira encontrada pelos químicos para contornar este tipo de dificuldade foi representar as estruturas em gráficos. Destes gráficos, é possível então derivar uma quantidade numérica que descreva a molécula. Com esta idéia em mente, alguns descritores moleculares foram desenvolvidos a partir da década de 40, e têm sido utilizados com sucesso em estudos de QSAR/QSPR.

Este tipo de descritor molecular é chamado de *índice topológico*, e descreve exatamente a disposição dos átomos na molécula, representando as informações estruturais na forma de uma tabela de conexão, que descreve o tipo de átomos e de ligações existentes, além das conexões formadas (van de Waterbeemd, 1995). Alguns exemplos de descritores topológicos utilizados neste trabalho são o índice de Wiener (*WIT*), o índice de conectividade Randic (*R*) e os índices WHIM-3D.

O primeiro índice topológico citado acima foi proposto em 1947 por Harry Wiener, e descreve as ramificações moleculares com base na matriz de distância molecular, cujos elementos são os números de ligação entre os pares apropriados na molécula. No cálculo do índice de Wiener são levados em consideração a distância entre um par de átomos, e o número de átomos de carbono que existe entre eles. O índice de Wiener é inversamente proporcional à compactação da molécula (Wiener, 1947; van de Waterbeemd, 1995).

O índice de conectividade Randic, proposto em 1975 por Milan Randic, descreve o tamanho molecular e o grau de ramificações com base no número de ligações que cada carbono realiza. À medida que aumenta o número de átomos, aumenta também o índice; por outro lado, quanto maior o grau de ramificação da molécula, menor o valor do índice (Randic, 1975; Randic, 1993; van de Waterbeemd, 1995).

Os índices WHIM-3D⁴⁸ foram desenvolvidos por Todeschini *et alli* (1994, 1995, 1996a, 1996b e 1997) e contém informações a respeito do tamanho, da forma, da simetria e da distribuição dos átomos nas moléculas. Eles são calculados a partir da geometria molecular, em termos de coordenadas *x*, *y* e *z*, dentro de diferentes esquemas de atribuição de pesos, resultando numa simples aproximação para representar as moléculas e seus fragmentos e conformações.

Os descritores geométricos

Numa série homóloga de compostos, o tamanho dos substituintes interfere nas complexas interações intermoleculares, refletindo desta forma nas variações dos valores

⁴⁸ *Weighted Holistic Invariant Molecular Descriptors*

das propriedades ou atividades biológicas. Os descritores geométricos dizem respeito ao tamanho e à forma da molécula e seus substituintes, e são importantes no entendimento das diversas propriedades físico-químicas e atividades biológicas dos compostos. É através da comparação entre o tamanho molecular de dois compostos similares que poderemos estimar qual terá o maior índice de retenção numa análise cromatográfica, por exemplo, ou qual terá o maior ponto de ebulição. Estes descritores são fundamentais no estudo das relações quantitativas entre estrutura e propriedades/atividades. Neste trabalho foram utilizados o volume molecular, a área molecular superficial e a ovalidade molecular.

Volume Molecular (VM)

O volume molecular pode ser estimado por meio de cálculos à partir do raio de van der Waals dos átomos individuais. A técnica mais utilizada nos últimos anos foi desenvolvida por Pearlman (van de Waterbeemd, 1995), cujo algoritmo utiliza uma integração numérica em que a molécula é vista como um conjunto de esferas atômicas parcialmente sobrepostas. O volume total molecular é a soma das contribuições de cada átomo (cada esfera) individualmente, subtraindo-se a região caracterizada pela intersecção entre duas esferas (van de Waterbeemd, 1995).

Área Superficial Molecular (VSA)

A área superficial de um átomo em uma molécula é a quantidade de área superficial daquele átomo que não apresenta intersecção com a área superficial de nenhum outro átomo. Se tomarmos cada átomo em uma molécula como sendo uma esfera com o raio igual ao raio de van der Waals, a área superficial (VSA^{49}) de cada átomo poderá ser calculada. A soma das áreas de cada átomo para uma determinada molécula irá fornecer a área superficial molecular de van der Waals (VSA molecular).

Dos vários algoritmos existentes para a estimativa da área molecular, os mais utilizados são aqueles desenvolvidos por Lee and Richards, Hermann, and Pearlman

(van de Waterbeemd, 1995), em que cada uma das várias esferas de uma molécula são mapeadas em vários planos. As intersecções entre estas esferas são calculadas, somando-se as áreas não ocupadas, obtendo-se assim a área superficial associada àquela molécula (van de Waterbeemd, 1995; Labute, P., 2000).

Ovalidade (O)

A ovalidade molecular é a razão entre a área superficial e a superfície esférica mínima e indica o grau de desvio esférico da estrutura molecular. É calculada a partir da equação 39, em que S é a área superficial molecular e V é o volume molecular (Huang, 1995; Bodor & Huang, 1997).

$$O = S / 4\pi \left(\frac{3V}{4\pi} \right)^{2/3} \quad (39)$$

O Cálculo dos Descritores Moleculares

Inicialmente a geometria de todas as moléculas foi otimizada e os descritores físico-químicos (termodinâmicos e eletrônicos) e estéricos calculados utilizando-se o método semi-empírico AM1 (Dewar *et alli*, 1985) do software Spartan para ambiente UNIX. Para o cálculo dos descritores topológicos as moléculas foram previamente construídas utilizando-se o software *HyperChem* para ambiente *Windows* (*HiperChem*, 1993). Os descritores foram então calculados usando o programa WHIM-3D que tem sido empregado com sucesso em estudos de QSPR (Todeschini *et alli*, 1994/1995/1996a/1996b/1997).

⁴⁹ *van der Waals surface area*

Apresentação dos Dados

Foram pesquisadas na literatura o ponto de ebulição (Karcher *et alli*, 1988), o coeficiente de partição octanol-água (Mackay *et alli*, 1998), o índice de retenção (Sander & Wise, 1986) e a fototoxidez (Mekenyan *et alli*, 1994) para estes compostos, sendo que para algumas moléculas ainda não existem os valores para estas propriedades determinados experimentalmente. As propriedades comuns a todas foram utilizadas na construção de modelos para prever as propriedades faltantes. Para a construção dos modelos de regressão foram utilizados apenas valores determinados experimentalmente pelo mesmo método, descartando-se os valores estimados através de cálculos. No quadro 17 pode ser observada a lista de propriedades que estão sendo estudadas.

Quadro 17: Lista das Propriedades em Estudo. Os dados experimentais utilizados para a construção dos modelos foram retirados da literatura

	Composto	pe / o C	Log Kow	IR	fototoxidez		Composto	pe / o C	Log Kow	IR	fototoxidez
1	naphthalene	218	3.37	2.00		36	benzo[a]perylene				
2	anthracene	340	4.54	3.2	-2.46	37	dibenzo[de, mn]naphthacene				
3	phenanthrene	338	4.57	3	-3.20	38	naphtho[2.3-g]chrysene				
4	naphthacene	440	5.76	4.51	-2.89	39	benzo[h]pentaphene				
5	benz[a]anthracene	435	5.91	4	-2.52	40	benzo[a]pentacene				
6	chrysene	431	5.86	4.10	-2.54	41	coronene	590	6.75		
7	triphenylene	429	5.49	3.7	-3.20	42	naphtho[1.2.3.4-def]chrysene	592		4.97	
8	pyrene	393	5.18	3.58	-2.31	43	dibenzo[def,p]chrysene	595		4.89	
9	benzo[c]phenanthrene			3.64		44	benzo[rs,t]pentaphene	594		5.73	
10	perylene	497	6.25	4.33	-2.67	45	benzo[g]chrysene			4.27	
11	benzo[a]pyrene	496	6.04	4.53	-2.43	46	2,3:5,6-Dibenzophenanthrene			4.33	
12	benzo[e]pyrene	493		4.28	-2.95	47	naphtho[2.1.8-gra]naphthacene			5.87	
13	picene	519		5.18		48	dibenzo[a,e]aceanthrylene			4.90	
14	pentaphene			4.67		49	acenaphthylene	270	4		
15	benzo[b]chrysene			5		50	dibenzo[a,k]fluoranthene			4.90	
16	dibenz[a, h]anthracene	535	6.75	4.73	-2.44	51	naphtho[2.3-k]fluoranthene			5.92	
17	dibenz[a, j]anthracene	531		4.56		52	fluoranthene	383	5.22	3.37	-2.81
18	benzo[b]triphenylene	535		4.40		53	dibenzo[k, mno]fluoranthene				
19	benzo[c]chrysene			4.45		54	acenaphthene	279	3.92		
20	pentacene					55	9H - fluorene	294	4.18	2.70	-3.20
21	dibenzo[c, g]phenanthrene			4.07		56	benzo[b]fluorene	398	5.75	3.84	-3.16
22	benzo[a]naphthacene			4.99		57	benzo[c]fluorene	406		3.49	
23	dibenzo[b, def]chrysene	596		6		58	benzo[ghi]fluoranthene	422			
24	dibenzo[def, mno]chrysene	547		5.08		59	benzo[a]fluoranthene			4.22	
25	dibenzo[a, j]naphthacene					60	indeno[1.2.3.cd]pyrene	534		4.84	
26	dibenzo[a, l]naphthacene					61	indeno[1.2.3-cd] fluoranthene	531		4.93	
27	dibenzo[a, c]naphthacene					62	1H-fluorene				
28	dibenzo[el]naphthacene					63	cyclopenta[cd]pyrene	439			
29	dibenzo[de, gr]naphthacene			4.92		64	benzo[j]fluoranthene	480		4.24	
30	dibenzo[g, p]chrysene					65	benzo[k]fluoranthene	481	6	4.42	-2.89
31	benzo[c]picene					66	benzo[a]fluorene	403	5.40	3.72	-3.17
32	benzo[ghi]perylene	542	6.50	4.76	-2.36	67	1H-benzo[b]fluorene				
33	dibenzo[b, k]chrysene					68	dibenz[e, k]acephenanthrylene			5.27	
34	dibenzo[c]chrysene					69	benzo[b]fluoranthene	481	5.80	4.29	
35	benzo[b]perylene			5.04		70	indan		3.33		

Análise dos Dados

Alguns estudos têm sido feitos na tentativa de encontrar relações entre as propriedades moleculares das substâncias e a estrutura de suas moléculas e para isto têm sido largamente utilizados uma variedade de descritores moleculares dos mais diversos tipos; hidrofóbicos, estéricos, eletrônicos, estruturais, topológicos e geométricos. A utilização de apenas um destes parâmetros, como por exemplo o peso molecular para explicar o ponto de ebulição, não fornece a informação suficiente para se estabelecer uma relação flexível o bastante para explicar os mais sutis desvios desta correlação. É necessário a utilização de outros parâmetros, que ajudem a explicar estes desvios, por exemplo, a diferença no ponto de ebulição de dois isômeros estruturais (que possuem o mesmo peso molecular) poderá ser explicada por algum parâmetro que indique a presença de uma ramificação.

No intuito de construir o modelo matemático que forneça a melhor correlação entre as propriedades estudadas neste trabalho procurou-se trabalhar com uma diversidade de tipos de descritores moleculares, na tentativa de descrever ao máximo as peculiaridades moleculares em todos os seus aspectos.

Isto não significa necessariamente que foram escolhidos os melhores descritores de *todos* os existentes atualmente para a construção do modelo, significa apenas que neste trabalho foram utilizados *alguns* dos descritores que forneceram a *melhor* regressão, dentre um grupo extenso de descritores moleculares pesquisados, mas que não contém todos os descritores existentes atualmente (e nem obtidos por todos os diferentes métodos conhecidos) em trabalhos de QSAR.

Seleção de Variáveis

A primeira dificuldade a ser resolvida é a escolha dos descritores que fornecerão o melhor modelo matemático que correlacione as variáveis independentes com as variáveis dependentes. Se o método for eficiente, será possível gerar um modelo capaz de prever as propriedades estudadas com um erro mínimo.

Dos 168 parâmetros estudados (cargas, momento de dipolo, energias do HOMO e LUMO, área, volume, ovalidade, eletronegatividade, dureza, polarizabilidade, peso molecular e mais 151 índices topológicos), foram utilizados na elaboração dos modelos apenas aqueles com maior correlação com a variável dependente, no caso, *ponto de ebulição, coeficiente de partição octanol-água, índice de retenção e fototoxidez*. Isto foi feito através da construção de gráficos de variável independente *vs.* variável dependente, que podem ser conferidos no apêndice II, e observando-se as correlações entre elas. Aquelas variáveis que não apresentaram correlação com as variáveis dependentes foram descartadas.

Ainda assim, algumas das variáveis apresentaram alta correlação entre si. Foi feita então uma outra seleção de variáveis para utilizar o mínimo necessário de variáveis referentes à mesma característica molecular (geométrica, eletrônica ou topológica), levando-se em consideração o coeficiente de cada variável no vetor de regressão e nos pesos, minimizando o risco de repetir informações e reduzindo assim a dimensão da matriz de dados original.

Isto foi feito gerando modelos de regressão com estas variáveis, e escolhendo aquelas com maior peso no vetor de regressão. Dentre estas, foram testadas as combinações possíveis entre as variáveis restantes e escolhidas aquelas que apresentaram melhores resultados.

Na escolha do modelo mais adequado para a previsão das propriedades foram levados em consideração a somatória do quadrado dos erros de previsão (PRESS), o coeficiente de correlação e o erro de validação apresentado no gráfico de valor medido *vs.* valor previsto. A tabela de dados contendo todas as variáveis estudadas para a construção de cada modelo encontra-se no apêndice I.

QSPR para ponto de ebulição (p.e.)

Apresentação dos Resultados para Ponto de Ebulição

Para a construção do modelo de regressão para ponto de ebulição, foi utilizado um conjunto de 36 moléculas com valores determinados experimentalmente segundo o método descrito por Karcher *et alli* (1988). Uma vez validado, o modelo foi utilizado para prever os valores de outras 34 moléculas.

Dentre as variáveis com maior peso no vetor de regressão e nos pesos, foram selecionadas volume, peso molecular e índice de conectividade Randic (Randic, 1975; Randic, 1993). No quadro 18 podem ser observados os coeficientes que cada uma das variáveis selecionadas têm no vetor de regressão.

Quadro 18: Coeficientes das variáveis no vetor de regressão para ponto de ebulição

variáveis	β
Volume	0.3318
PM	0.3339
R	0.3319

Para a propriedade estudada, os resultados da análise de regressão pelo método dos mínimos quadrados parciais são mostrados no quadro 19. O número ótimo de variáveis latentes encontradas para o modelo desenvolvido foi 1, que descreve 99.5% da variância total dos dados, e o modelo apresentou boa qualidade, com coeficiente de correlação 0.995.

Quadro 19: Resultados dos modelos de QSPR utilizando PLS com validação cruzada para ponto de ebulição

Variáveis Latentes	SEV	PRESS	% de Variância	r
1	10	3601.49	99.53	0.9952
2	8.33	2500.68	99.62	0.9975
3	7.67	2118.11	100	0.9975

Na figura 43 é apresentado o gráfico de valor medido vs. valor previsto do melhor modelo obtido, cujos valores com os erros de validação podem ser encontrados no quadro 20.

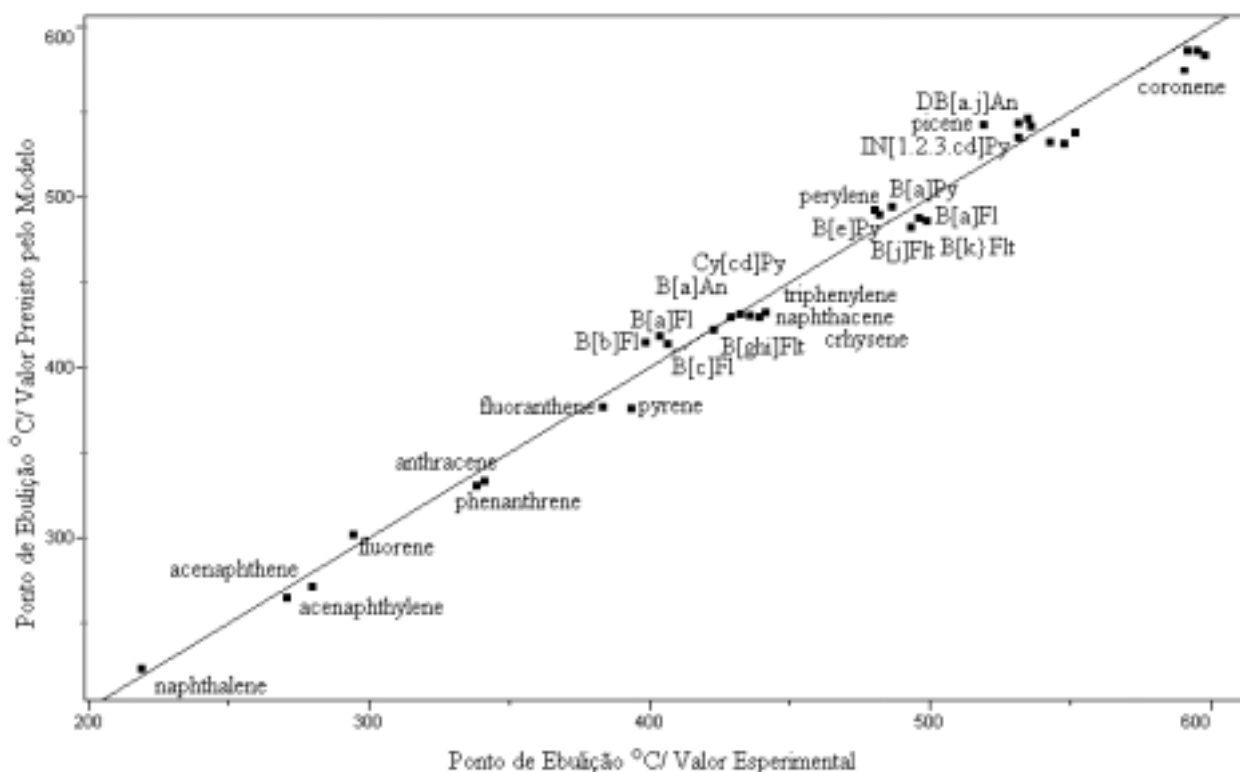


Figura 43: Gráfico de valor medido vs. valor previsto para ponto de ebulição.

Todos os erros de validação encontram-se abaixo de 3%, com exceção das moléculas de piceno (4.89%), benzo[b]fluoreno (4.46%) e benzo[a]fluoreno (3.96%), o que é considerado satisfatório em estudos de QSAR. O quadro 20 apresenta os valores medidos experimentalmente e os valores previstos pelo modelo.

Quadro 20: Erros de validação para o ponto de ebulição (° C)

	Composto	Valor Medido	Valor Previsto	Resíduo	Erro de Validação (%)
1	naphthalene	218	222.93	-4.93	-2.26
2	anthracene	340	332.91	7.09	2.08
3	phenanthrene	338	332.72	5.28	1.56
4	naphthacene	440	431.91	8.09	1.84
5	benz[a]anthracene	435	431.56	3.44	0.79
6	chrysene	431	431.24	-0.24	-0.06
7	triphenylene	429	430.51	-1.51	-0.35
8	pyrene	393	376.73	16.27	4.14
10	perylene	497	488.67	8.33	1.68
11	benzo[a]pyrene	496	488.85	7.15	1.44
12	benzo[e]pyrene	493	488.23	4.77	0.97
13	picene	519	544.40	-25.40	-4.89
16	dibenz[a, h]anthracene	535	544.31	-9.31	-1.74
17	dibenz[a, j]anthracene	531	543.95	-12.95	-2.44
18	benzo[b]triphenylene	535	543.80	-8.80	-1.65
23	dibenzo[b, def]chrysene	596	587.22	8.78	1.47
24	dibenzo[def, mno]chrysene	547	532.59	14.41	2.63
32	benzo[ghi]perylene	542	532.56	9.44	1.74
41	coronene	590	575.70	14.30	2.42
42	naphtho[1.2.3.4-def]chrysene	592	586.58	5.42	0.92
43	dibenzo[def,p]chrysene	595	586.37	8.63	1.45
44	benzo[rst]pentaphene	594	586.69	7.31	1.23
49	acenaphthylene	270	265.40	4.60	1.70
52	fluoranthene	383	377.93	5.07	1.32
54	acenaphthene	279	271.24	7.76	2.78
55	9H - fluorene	294	302.92	-8.92	-3.03
56	benzo[b]fluorene	398	415.74	-17.74	-4.46
57	benzo[c]fluorene	406	414.99	-8.99	-2.21
58	benzo[ghi]fluoranthene	422	421.98	0.02	0.00
60	indeno[1.2.3.cd]pyrene	534	535.30	-1.30	-0.24
61	indeno[1.2.3-cd] fluoranthene	531	535.84	-4.84	-0.91
63	cyclopenta[cd]pyrene	439	431.43	7.57	1.72
64	benzo[j]fluoranthene	480	492.21	-12.21	-2.54
65	benzo[k]fluoranthene	481	491.85	-10.85	-2.25
66	benzo[a]fluorene	403	418.98	-15.98	-3.96
69	benzo[b]fluoranthene	481	491.58	-10.58	-2.20

Quadro 21: Valores de Ponto de Ebulição medidos experimentalmente e previstos pelo modelo (os valores previstos encontram-se em **negrito)**

Composto	pe/ o C)	Composto	pe/ o C)
1 naphthalene	218	36 benzo[a]perylene	587.13
2 anthracene	340	37 dibenzo[de. mn]naphthacene	587.49
3 phenanthrene	338	38 naphtho[2.3-g]chrysene	653.89
4 naphthacene	440	39 benzo[h]pentaphene	655.09
5 benz[a]anthracene	435	40 benzo[a]pentacene	656.13
6 chrysene	431	41 coronene	590
7 triphenylene	429	42 naphtho[1.2.3.4-def]chrysene	592
8 pyrene	393	43 dibenzo[def.p]chrysene	595
9 benzo[c]phenanthrene	431	44 benzo[rst]pentaphene	594
10 perylene	497	45 benzo[g]chrysene	542.74
11 benzo[a]pyrene	496	46 2.3:5.6-Dibenzophenanthrene	543.56
12 benzo[e]pyrene	493	47 naphtho[2.1.8-qua]naphthacene	587.76
13 picene	519	48 dibenzo[a.e]aceantrylene	588.59
14 pentaphene	543.99	49 acenaphthylene	270
15 benzo[b]chrysene	543.54	50 dibenzo[a.k]fluoranthene	588.81
16 dibenzo[a. h]anthracene	535	51 naphtho[2.3-k]fluoranthene	589.18
17 dibenzo[a. j]anthracene	531	52 fluoranthene	383
18 benzo[b]triphenylene	535	53 dibenzo[k. mno]fluoranthene	534.13
19 benzo[c]chrysene	543.14	54 acenaphthene	279
20 pentacene	544.45	55 9H - fluorene	294
21 dibenzo[c. g]phenanthrene	542.93	56 benzo[b]fluorene	398
22 benzo[a]naphthacene	544.01	57 benzo[c]fluorene	406
23 dibenzo[b. def]chrysene	596	58 benzo[ghi]fluoranthene	422
24 dibenzo[def. mno]chrysene	547	59 benzo[a]fluoranthene	489.93
25 dibenzo[a.j]naphthacene	655.76	60 indeno[1.2.3.cd]pyrene	534
26 dibenzo[a. l]naphthacene	655.75	61 indeno[1.2.3-cd] fluoranthene	531
27 dibenzo[a.c]naphthacene	655.09	62 1H-fluorene	309.59
28 dibenzo[e]naphthacene	586.25	63 cyclopenta[cd]pyrene	439
29 dibenzo[de.gr]naphthacene	587.19	64 benzo[j]fluoranthene	480
30 dibenzo[g.p]chrysene	654.29	65 benzo[k]fluoranthene	481
31 benzo[c]picene	654.84	66 benzo[a]fluorene	403
32 benzo[ghi]perylene	542	67 1H-benzo[b]fluorene	418.75
33 dibenzo[b.k]chrysene	655.65	68 dibenz[e.k]acephenanthrylene	610.90
34 dibenzo[cl]chrysene	654.89	69 benzo[b]fluoranthene	481
35 benzo[b]perylene	586.87	70 indan	212.03

Os valores de ponto de ebulição utilizados para a construção do modelo de regressão encontram-se numa faixa entre 218 °C (naftaceno) e 596 °C (dibenzo[b. def]criseno). Alguns dos valores previstos foram extrapolados e encontram-se acima desta faixa, como pode ser observado no quadro acima.

Discussão dos Resultados para Ponto de Ebulição

O ponto de ebulição de um composto pode ser explicado em termos termodinâmicos como sendo a temperatura em que a energia térmica das partículas é capaz de vencer as forças de coesão que mantêm a substância no estado de agregação que caracteriza a fase líquida (Atkins, 1993). Desta forma, o ponto de ebulição permite estimar a dispersão do composto na atmosfera.

Para o caso das moléculas estudadas, as forças de van der Waals que ocorrem em moléculas apolares são menos intensas do que as forças do tipo interação entre dipolos que caracterizam as moléculas polares (Morrison & Boyd, 1993). A intensidade relativa deste tipo de interação está intimamente relacionada com o tamanho e o modo de ramificação do esqueleto molecular e para se estimar este tamanho são utilizados parâmetros como peso molecular, volume e área superficial, dos quais foram utilizados apenas os dois primeiros para a construção do modelo.

Apesar de apresentar alta correlação com o ponto de ebulição, como pode ser observado no apêndice II, a área superficial molecular não foi utilizada na construção do modelo matemático, pois é altamente correlacionada com o volume molecular. A utilização dos dois parâmetros simultaneamente implica na repetição de informações que podem muito bem serem descritas por uma variável apenas, no caso, o volume molecular.

Para ilustrar a influência do tamanho molecular no ponto de ebulição, tomemos como exemplo duas moléculas de diferentes pesos moleculares (e portanto diferentes volumes): antraceno e o naftaceno, com pontos de ebulição 340 e 440 respectivamente, cujas estruturas podem ser visualizadas nas figuras 44(a) e 44(b). O naftaceno apresenta um volume maior, e portanto uma maior área superficial, capaz de interagir com outra molécula de forma mais intensa do que o antraceno, sendo necessária uma quantidade de energia maior para romper as forças de coesão que as mantêm no estado líquido.

A importância da ramificação ou, em se tratando de PAHs, da disposição dos anéis na estimativa do ponto de ebulição pode ser constatada pela observação de outros dois compostos isômeros: o naftaceno (p.e.= 440) e o trifenileno (p.e.= 429), nas figuras 44(b) e 44(c). Os dois compostos apresentam mesmo peso molecular, mas apresentam

um formato diferente. O primeiro é mais linear, enquanto o segundo apresenta uma forma mais arredondada e compacta, em consequência da disposição dos anéis. Sendo assim, a energia necessária para romper as interações entre duas moléculas de naftaceno são mais intensas do que para o caso do trifenileno, resultando num menor ponto de ebulição para este último. A indicação deste tipo de variação na estrutura molecular é fornecida pelo índice de conectividade Randic, que foi criado inicialmente para compostos acíclicos, e mais tarde foi desenvolvido para compostos cíclicos e aromáticos (Randic, 1975; Randic, 1993).

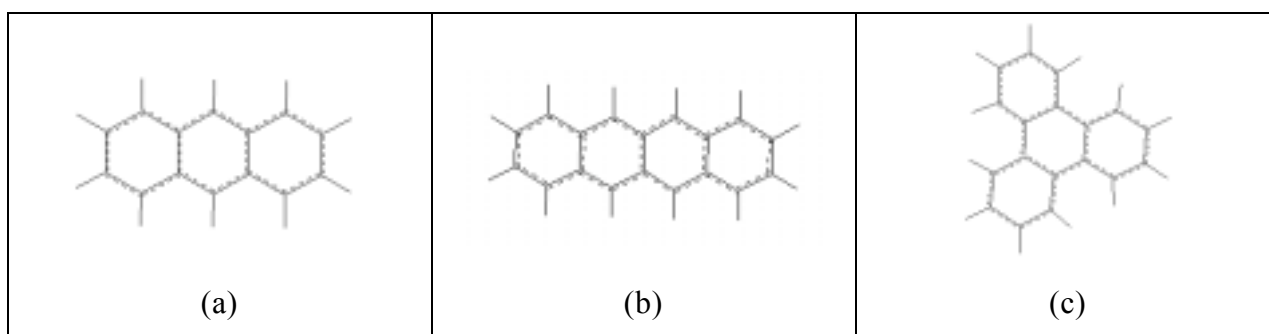


Figura 44: Estruturas do antraceno (a), naftaceno (b) e trifenileno (c).

Por meio do índice de Randic é possível diferenciar estruturas isoméricas, estimar o tamanho da molécula e o número de átomos (quanto maior o número de átomos, maior o índice) e verificar o grau de ramificação ou ciclização da estrutura (quanto mais ramificações/ciclos, menor o valor do índice) (Randic, 1975; Randic, 1993).

Os erros de validação encontram-se em torno de 3%, com poucas exceções que ultrapassam este valor: piceno (4.89%), benzo[b]fluoreno (4.46%) e benzo[a]fluoreno (3.96%).

Exceto a molécula de piceno, as outras duas moléculas fazem parte de um grupo relativamente pequeno, se comparado ao restante do conjunto, das espécies que apresentam em sua estrutura um anel de cinco átomos de carbono que não possui aromaticidade. Juntamente com estas duas, outras 3 moléculas utilizadas na construção do modelo também apresentam um anel de cinco átomos de carbono não aromático, porém estas apresentaram erros menores. São elas: acenafteno (2.78%), o fluoreno

(3.03%) e o benzo[c]fluoreno (2.21%), cujas estruturas podem ser observadas na figura 45. Os erros de validação encontram-se no quadro 20.

Estas moléculas apresentam uma estrutura planar (exceto o benzo[c]fluoreno), com apenas os dois átomos de hidrogênio que estão ligados ao átomo de carbono com hibridação sp^3 fora do plano da molécula, como pode ser observado na figura 45.

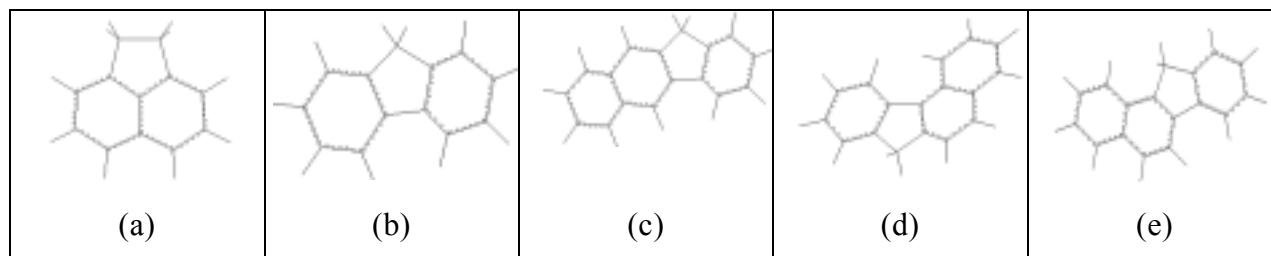


Figura 45: Estruturas do acenafteño (a), fluoreno (b), benzo[b]fluoreno (c), benzo[c]fluoreno (d) e benzo[a]fluoreno (e).

O número de moléculas no conjunto de modelagem que representa esta classe de compostos (aqueles que apresentam um anel não aromático) é pequeno em relação ao conjunto total, e pode não ser muito representativo. Isto explicaria o erro maior em alguns casos. Mas não se pode afirmar isto com certeza, seria apenas uma possível explicação. Basta observar o caso do piceno, que apresenta o maior erro dentre todas as moléculas utilizadas no conjunto de modelagem, e no entanto possui várias similaridades com outras moléculas: é planar como a maioria delas, possui em sua estrutura 5 anéis, como mais de 10 outras moléculas etc. Uma outra explicação possível para a dimensão do erro de validação pode estar relacionada ao erro embutido nas medidas experimentais, mas que não poderão ser checados, uma vez que os dados foram retirados da literatura. Seria interessante realizar uma nova determinação experimental do ponto de ebulição para estas moléculas, para checar estes valores, mas neste caso as medidas deveriam ser feitas para todo o conjunto, uma vez que o erro experimental está atrelado aos procedimentos analíticos e, mesmo utilizando um mesmo método, pode variar de uma analista para outro.

Sobre a ausência de descritores eletrônicos

O ponto de ebulição é uma propriedade que depende basicamente da forma e tamanho da molécula, e portanto um modelo de regressão para a previsão desta propriedade necessita somente de variáveis relacionadas à parâmetros geométricos e topológicos, descartando-se a utilização de parâmetros relacionados às características eletrônicas da molécula. Apesar de algumas delas apresentarem uma certa correlação com a variável dependente, como é o caso da energia do HOMO, energia do LUMO e GAP, como pode ser observado no apêndice II, estas variáveis apresentam baixo valor de pesos.

Para avaliar a influência da utilização de uma variável eletrônica para a construção do modelo de previsão para ponto de ebulição, a variável energia do LUMO foi incluída no conjunto de modelagem e um novo modelo foi testado, tendo sido observada a ocorrência de uma distorção dos resultados obtidos. O número de variáveis latentes necessários subiu de 1 para 2, para a mesma % de variância descrita, como pode ser comparado entre os quadros 19⁵⁰ e 22.

Quadro 22: Resultados dos modelos de QSPR utilizando PLS com validação cruzada para ponto de ebulição com as variáveis energia do LUMO, volume, PM e R

Variáveis Latentes	SEV	PRESS	% de Variância	r
1	13.63	6737.22	87.43	0.9915
2	8.97	2897.99	99.67	0.9965
3	8.27	2465.1	99.83	0.9973

O modelo anterior apresentou erros menores que 3%, com algumas exceções acima de 4%. O modelo construído com a inclusão da variável energia do LUMO apresentou erros acima de 3%, com algumas exceções acima de 5%. Decidiu-se então excluir as variáveis eletrônicas da construção do modelo de previsão para ponto de ebulição.

⁵⁰ Cf. página 114.

QSPR para coeficiente de partição octanol-água (Log de K_{ow})

Apresentação dos Resultados para Log K_{ow}

Na construção do modelo de regressão para coeficiente de partição octanol-água, foi utilizado um conjunto de 22 moléculas com valores determinados experimentalmente, segundo o método descrito por Mackay *et alli* (1998). Uma vez validado, o modelo foi utilizado para prever os valores de outras 48 moléculas.

Foram selecionadas as variáveis volume, peso molecular e índice de conectividade Randic (Randic, 1975; Randic 1993). No quadro 23 podem ser observados os coeficientes de cada variável no vetor de regressão.

Quadro 23: Coeficientes das variáveis no vetor de regressão para Log K_{ow}

variáveis	β
Volume	0.3309
PM	0.3294
R	0.3262

Quadro 24: Resultados dos modelos de QSPR utilizando PLS com validação cruzada para Log K_{ow} utilizando as variáveis volume molecular, PM e R

Variáveis Latentes	SEV	PRESS	% de Variância	r
1	0.1958	0.8437	99.54	0.9842
2	0.1763	0.684	99.88	0.988
3	0.1769	0.6887	100	0.9886

Os resultados da análise de regressão pelo método dos mínimos quadrados parciais para o coeficiente de partição octanol-água, são mostrados no quadro 24. O número ótimo de variáveis latentes encontradas para o modelo desenvolvido foi 1, que descreve 99.54% da variância total dos dados, e o modelo apresentou boa qualidade, com coeficientes de correlação 0.984.

Na figura 46 é apresentado o gráfico de valor medido vs. valor previsto do melhor modelo obtido, cujos valores com os erros de validação podem ser encontrados no quadro 25.

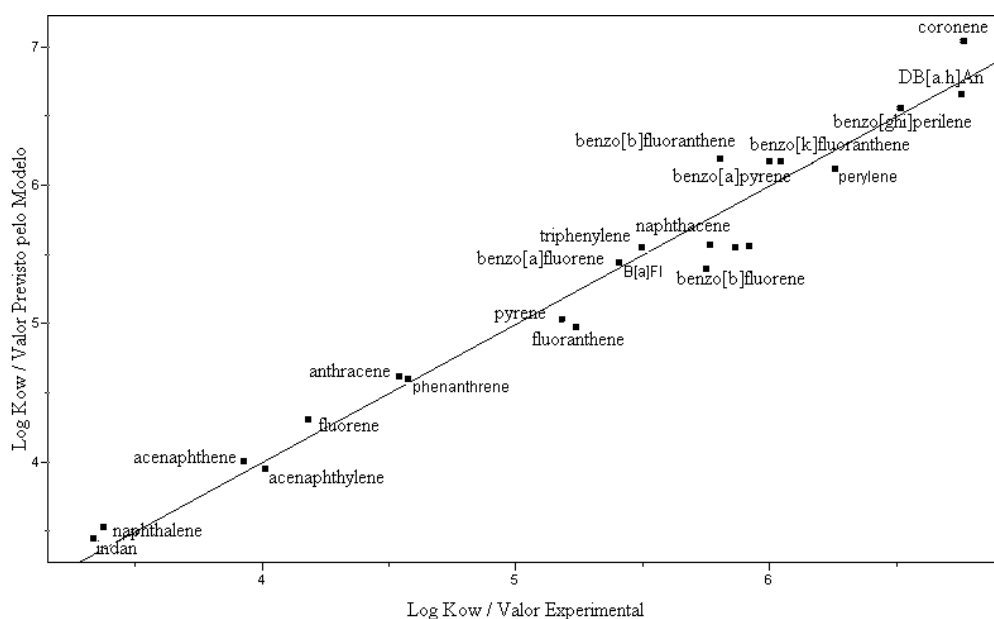


Figura 46: Gráfico de valor medido vs. valor previsto para $\text{Log } K_{ow}$.

Quadro 25: Erros de validação para $\text{Log } K_{ow}$

	Compostos	Valor Medido	Valor Previsto	Resíduo	Erro de Validação %
1	naphthalene	3.37	3.54	-0.17	-5.13
2	anthracene	4.54	4.62	-0.08	-1.66
3	phenanthrene	4.57	4.61	-0.04	-0.82
4	naphthacene	5.76	5.57	0.19	3.33
5	benz[a]anthracene	5.91	5.57	0.34	5.81
6	chrysene	5.86	5.57	0.29	5.01
7	triphenylene	5.49	5.56	-0.07	-1.24
8	pyrene	5.18	5.04	0.14	2.71
10	perylene	6.25	6.13	0.12	1.96
11	benzo[a]pyrene	6.04	6.18	-0.14	-2.34
16	dibenz[a, h]anthracene	6.75	6.66	0.09	1.39
32	benzo[ghi]perylene	6.5	6.56	-0.06	-0.99
41	coronene	6.75	7.06	-0.31	-4.54
49	acenaphthylene	4	3.95	0.05	1.20
52	fluoranthene	5.22	5.04	0.18	3.53
54	acenaphthene	3.92	4.02	-0.10	-2.46
55	9H - fluorene	4.18	4.32	-0.14	-3.35
56	benzo[b]fluorene	5.75	5.41	0.34	6.00
65	benzo[k]fluoranthene	6	6.18	-0.18	-2.95
66	benzo[a]fluorene	5.4	5.45	-0.05	-0.86
69	benzo[b]fluoranthene	5.8	6.21	-0.41	-7.05
70	indan	3.33	3.45	-0.12	-3.72

Todos os erros de validação encontram-se abaixo de 5%, com exceção das molécula de naftaleno, benzo[a]antraceno, criseno e benzo[b]fluoreno, que

apresentaram erros de validação de 5.13, 5.81, 5.01, 6.00 e 7.05% respectivamente. O quadro 26 apresenta os valores medidos experimentalmente e os valores previstos pelo modelo.

Quadro 26: Valores para Log Kow determinados experimentalmente e previstos pelo modelo (em negrito)

	Composto	log Kow		Composto	log Kow
1	naphthalene	3.37	36	benzo[a]perylene	7.10
2	anthracene	4.54	37	dibenzo[de. mn]naphthacene	7.11
3	phenanthrene	4.57	38	naphtho[2.3-g]chrysene	7.76
4	naphthacene	5.76	39	benzo[h]pentaphene	7.77
5	benz[a]anthracene	5.91	40	benzo[a]pentacene	7.78
6	chrysene	5.86	41	coronene	6.75
7	triphenylene	5.49	42	naphtho[1.2.3.4-def]chrysene	7.10
8	pyrene	5.18	43	dibenzo[def.p]chrysene	7.10
9	benzo[c]phenanthrene	5.57	44	benzo[rst]pentaphene	7.10
10	perylene	6.25	45	benzo[g]chrysene	7.67
11	benzo[a]pyrene	6.04	46	2.3:5.6-Dibenzophenanthrene	6.67
12	benzo[e]pyrene	6.13	47	naphtho[2.1.8-qla]naphthacene	7.11
13	picene	6.67	48	dibenzo[a.e]aceantrylene	7.12
14	pentaphene	6.68	49	acenaphthylene	4
15	benzo[b]chrysene	6.67	50	dibenzo[a.k]fluoranthene	7.12
16	dibenz[a. h]anthracene	6.75	51	naphtho[2.3-k]fluoranthene	7.12
17	dibenz[a. j]anthracene	6.67	52	fluoranthene	5.22
18	benzo[b]triphenylene	6.67	53	dibenzo[k. mno]fluoranthene	6.58
19	benzo[c]chrysene	6.67	54	acenaphthene	3.92
20	pentacene	6.68	55	9H - fluorene	4.18
21	dibenzo[c. g]phenanthrene	6.67	56	benzo[b]fluorene	5.75
22	benzo[a]naphthacene	6.68	57	benzo[c]fluorene	5.41
23	dibenzo[b. def]chrysene	7.10	58	benzo[ghi]fluoranthene	5.48
24	dibenzo[def. mno]chrysene	6.57	59	benzo[a]fluoranthene	6.14
25	dibenzo[a.j]naphthacene	7.78	60	indeno[1.2.3.cd]pyrene	6.58
26	dibenzo[a. l]naphthacene	7.78	61	indeno[1.2.3-cd] fluoranthene	6.59
27	dibenzo[a.c]naphthacene	7.77	62	1H-fluorene	4.29
28	dibenzo[el]naphthacene	7.09	63	cyclopenta[cd]pyrene	5.57
29	dibenzo[de.gr]naphthacene	7.10	64	benzo[j]fluoranthene	6.16
30	dibenzo[g.p]chrysene	7.76	65	benzo[k]fluoranthene	6
31	benzo[c]picene	7.77	66	benzo[a]fluorene	5.40
32	benzo[ghi]perylene	6.50	67	1H-benzo[b]fluorene	5.45
33	dibenzo[b.k]chrysene	7.77	68	dibenz[e.k]acephenanthrylene	7.33
34	dibenzo[cl]chrysene	7.77	69	benzo[b]fluoranthene	5.80
35	benzo[b]perylene	7.10	70	indan	3.33

Os valores de coeficiente de partição octanol-água utilizados para a construção do modelo de regressão encontram-se numa faixa entre 3.37 (naftaleno) e 6.75 (dibenz[a,h]antraceno e coroneno). Alguns dos valores previstos foram extrapolados e encontram-se acima desta faixa, como pode ser observado no quadro acima.

Discussão dos resultados para Log K_{ow}

Todos os compostos utilizados neste estudo apresentam Log K_{ow} acima de 1, e são lipofílicos, sendo que a maioria é extremamente hidrofóbica. Mas as pequenas variações nas suas estruturas moleculares resultam em variações nesta lipofilicidade.

Assim como o ponto de ebulição, o coeficiente de partição octanol-água está intimamente relacionado com as interações intermoleculares, que dependem neste caso do formato e do tamanho molecular. Por este motivo o modelo de regressão para o coeficiente de partição octanol-água de PAHs que apresentou melhores resultados utiliza as mesmas variáveis que o modelo para o ponto de ebulição: volume, peso molecular e o índice de conectividade Randic. O peso que cada uma destas variáveis apresenta no vetor de regressão são bem próximos daqueles apresentados no caso do ponto de ebulição, como pode ser observado nos quadros 18 (para ponto de ebulição) e 23 (para Log K_{ow}), em que as variáveis apresentam coeficientes em torno de 0.3.⁵¹

Para ilustrar a influência do tamanho molecular no valor desta propriedade, tomemos como exemplo as mesmas duas moléculas de diferentes pesos moleculares (e portanto diferentes volumes) citadas no caso anterior: o antraceno e o naftaceno, com Log K_{ow} igual a 4.54 e 5.76, respectivamente. Para estes dois casos, a lipofilicidade cresce à medida que aumenta o número de átomos de carbono na estrutura, o que resulta em maiores volumes e peso moleculares.

A influência da ramificação ou da disposição dos anéis pode ser discutida a partir da observação dos dois isômeros estruturais utilizados no caso anterior: o naftaceno e o trifenileno, que possuem respectivamente Log K_{ow} igual a 5.76 e 5.49. Os dois compostos apresentam o mesmo peso molecular, mas diferem no formato. Esta

⁵¹ Cf. páginas 114 e 122.

diferença estrutural é descrita pelo índice de conectividade Randic (Randic, 1975; Randic, 1993).

A estrutura do naftaceno é mais linear, resultando num formato maior, mais alongado e portanto mais volumoso, favorecendo desta forma a solubilização na fase orgânica. O trifenileno apresenta um formato mais compacto, devido à disposição dos anéis, resultando numa estrutura com menor volume.

Neste caso também a área superficial não foi utilizada na construção do modelo de regressão para coeficiente de partição octanol-água por apresentar alta correlação com o volume molecular, como pode ser observado no apêndice II.

O modelo apresentou erro de validação menores que 5%, que é considerado muito satisfatório para estudos de QSAR, exceto para algumas moléculas. Benzo[b]fluoreno e benzo[b]fluoranteno foram as únicas que apresentaram erros de validação muito maiores do que 5%, como pode ser observado no quadro 28.

Como citado anteriormente, o benzo[b]fluoreno faz parte de um grupo relativamente pequeno, se comparado ao restante do conjunto, das espécies que apresentam em sua estrutura um anel não aromático de cinco átomos de carbono. Na construção do modelo para previsão de ponto de ebulição, foram utilizadas 5 destas moléculas, cujas estruturas podem ser observadas na figura 45⁵². No caso do coeficiente de partição octanol-água, uma destas moléculas não foi utilizada na construção do modelo, o benzo[c]fluoreno, que por suas características estruturais se assemelha à molécula de benzo[b]fluoreno. É muito possível que esta molécula que não foi utilizada na construção do modelo influencia na descrição da molécula de benzo[b]fluoranteno, e sua retirada do conjunto de modelagem contribuiu para o aumento do erro de validação deste composto. Neste caso, todas as moléculas apresentam estrutura planar, com apenas os átomos de hidrogênio ligados aos carbonos com hibridação sp^3 fora do plano da molécula, como pode ser observado na figura 45.

A molécula de benzo[b]fluoranteno pertence à outro grupo muito pequeno de estruturas que possuem um anel de cinco átomos de carbono, mas que apresentam aromaticidade, cujas estruturas podem ser observadas na figura 47.

⁵² Cf. página 120.

Para a construção do modelo de previsão para o ponto de ebulição foram utilizadas seis destas estruturas (figura 47), enquanto para o modelo de previsão do coeficiente de partição octanol-água foram utilizadas apenas três: as moléculas de fluoranteno, benzo[k]fluoranteno e benzo[b]fluoranteno, que podem não ser suficientes para representar o conjunto de estruturas contendo um anel de cinco átomos de carbono, levando à um aumento no erro de validação.

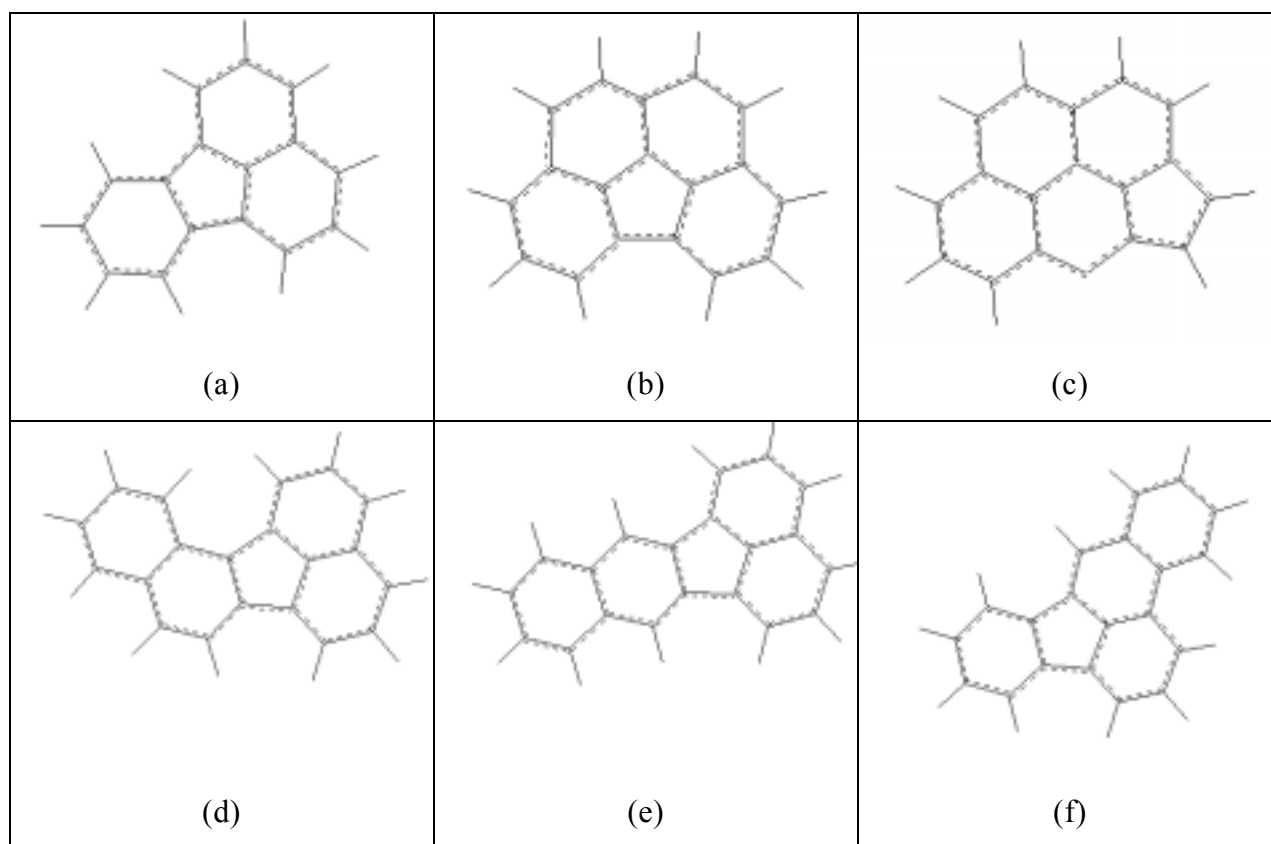


Figura 47: Estruturas do fluoranteno (a), benzo[ghi]fluoranteno (b), ciclopenta[cd]pireno (c), benzo[j]fluoranteno (d), benzo[k]fluoranteno (e) e benzo[b]fluoranteno (f).

Mas a origem deste erro pode não estar apenas na escolha do conjunto de modelagem, como explicado anteriormente. Este erro pode ter origem também nas medidas experimentais, ou até mesmo para algum outro fator desconhecido. Uma outra molécula que apresenta maior grau de similaridade com o conjunto também apresentou um erro de validação um pouco acima dos 5% encontrados (benz[a]antraceno – 5.81).

Não se pode afirmar com certeza a causa deste erro de validação obtido, apenas se pode concluir que o modelo não é adequado para prever estas 3 moléculas.

Assim como para o ponto de ebulição, neste caso também não foi necessária a utilização de variáveis eletrônicas, apesar de algumas delas apresentarem correlação com o coeficiente de partição octanol-água, como pode ser observado no apêndice II.

A inclusão da variável energia do LUMO, a pretexto de se testar as variações no modelo, acarretou numa distorção dos resultados. Como pode ser observado no quadro 37 houve um aumento no número de variáveis latentes de 1 para 2, e o erro de validação aumentou um pouco (acima de 5%). Decidiu-se então não utilizar nenhuma variável eletrônica na construção do modelo.

Quadro 27: Resultados dos modelos de QSPR utilizando PLS com validação cruzada para $\text{Log } K_{ow}$ utilizando as variáveis energia do LUMO, volume molecular, PM e R

Variáveis Latentes	SEV	PRESS	% de Variância	r
1	0.2334	0.1199	88.87	0.979
2	0.1947	0.8342	99.52	0.985
3	0.1777	0.6889	99.92	0.989

QSPR para Índice de Retenção (IR)

Apresentação dos Resultados para Índice de Retenção

Para a construção do modelo de regressão para índice de retenção, foi utilizado um conjunto de 44 moléculas com valores determinados experimentalmente segundo o método descrito por Sander & Wise (1987). Uma vez validado, o modelo foi utilizado para prever os valores de outras 25 moléculas.

Foram selecionadas as variáveis energia do LUMO, dureza, polarizabilidade, peso molecular e índice de Wiener (Wiener, 1947). No quadro 28 podem ser observados os coeficientes que cada variável tem no vetor de regressão.

Quadro 28: Coeficiente das variáveis no vetor de regressão para índice de retenção

variáveis	β
Elumo	-0.1321
η	-0.1246
α	0.2387
PM	0.2595
WIT	0.2729

Os resultados da análise de regressão pelo método dos mínimos quadrados parciais para o índice de retenção são mostrados no quadro 29. O número ótimo de variáveis latentes encontradas para o modelo desenvolvido foi 2 e o modelo apresentou boa qualidade, com coeficiente de correlação 0.948.

Quadro 29: Resultados dos modelos de QSPR utilizando PLS com validação cruzada para Índice de Retenção

Variáveis Latentes	SEV	PRESS	% de Variância	r
1	0.2493	2.7353	81.73	0.9436
2	0.2432	2.6017	97.52	0.9477
3	0.2596	2.9565	97.95	0.9489

Na figura 48 é apresentado o gráfico de valor medido vs. valor previsto do melhor modelo obtido, cujos valores com os erros de validação podem ser encontrados no quadro 30.

Apesar do gráfico apresentar uma faixa de pontos mais larga que o desejável, os erros de validação são satisfatórios, com valores abaixo de 10%, e o coeficiente de correlação obtido foi muito bom (0.948), como pode ser observado no quadro 29. O quadro 31 apresenta os valores medidos experimentalmente e os valores previstos pelo modelo.

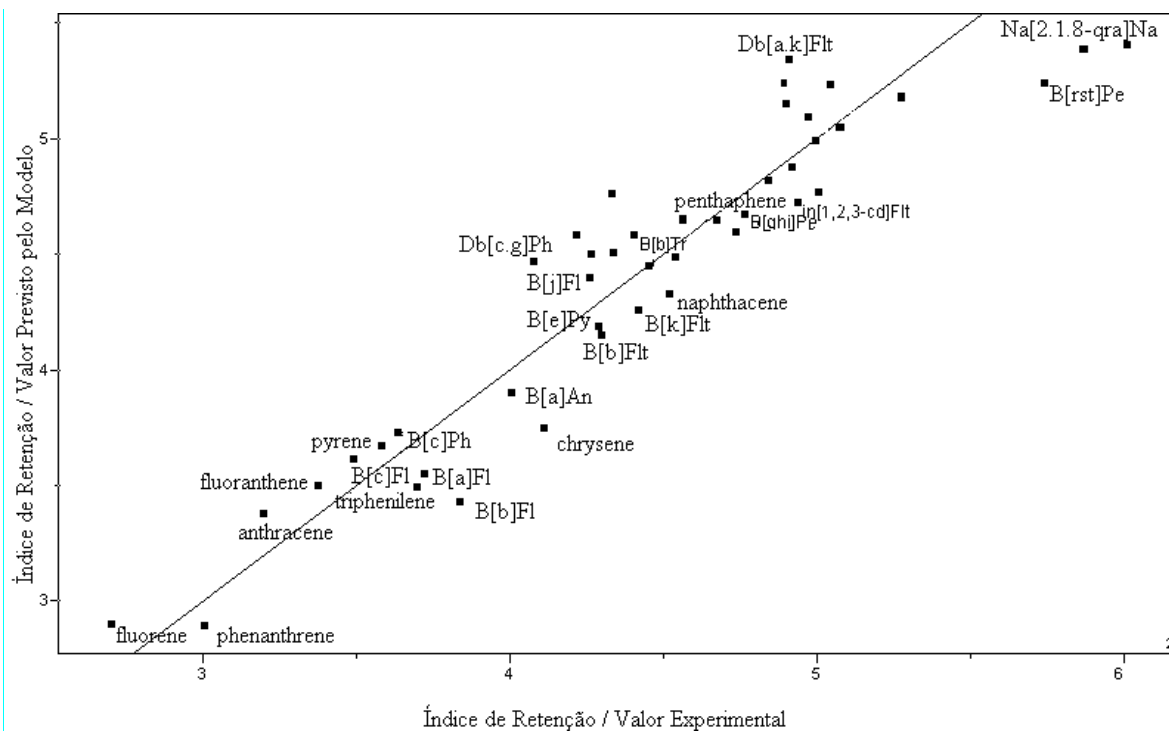


Figura 48: Gráfico de valor medido vs. valor previsto para índice de retenção.

Quadro 30: Erros de validação para Índice de Retenção

	composto	Valor Medido	Valor Previsto	Resíduos	Erros de Validação
2	anthracene	3.2	3.39	-0.19	-5.79
3	phenanthrene	3	2.91	0.09	3.04
4	naphthacene	4.51	4.33	0.18	4.031
5	benz[a]anthracene	4	3.91	0.09	2.19
6	chrysene	4.10	3.75	0.35	8.43
7	triphenylene	3.7	3.49	0.21	5.58
8	pyrene	3.58	3.67	-0.09	-2.62
9	benzo[c]phenanthrene	3.64	3.72	-0.08	-2.20
10	perylene	4.33	4.51	-0.18	-4.18
11	benzo[a]pyrene	4.53	4.47	0.06	1.29
12	benzo[e]pyrene	4.28	4.19	0.09	2.17
14	pentaphene	4.67	4.65	0.02	0.46
15	benzo[b]chrysene	5	4.78	0.22	4.33
16	dibenz[a. h]anthracene	4.73	4.59	0.14	2.88
17	dibenz[a. j]anthracene	4.56	4.66	-0.10	-2.18
18	benzo[b]triphenylene	4.40	4.59	-0.19	-4.31
19	benzo[c]chrysene	4.45	4.46	-0.01	-0.22
21	dibenzo[c,g]phenanthrene	4.07	4.47	-0.40	-9.80
22	benzo[a]naphthacene	4.99	5.00	-0.01	-0.19
23	dibenzo[b. def]chrysene	6	5.40	0.60	9.96
24	dibenzo[def. mno]chrysene	5.08	5.06	0.02	0.49
29	dibenzo[de.gr]naphthacene	4.92	4.89	0.03	0.66
32	benzo[ghi]perylene	4.76	4.68	0.08	1.68
35	benzo[b]perylene	5.04	5.24	-0.20	-4.05
42	naphtho[1.2.3.4-def]chrysene	4.97	5.10	-0.13	-2.68
43	dibenzo[def.p]chrysene	4.89	5.15	-0.26	-5.30
44	benzo[rst]pentaphene	5.73	5.25	0.48	8.45
45	benzo[g]chrysene	4.27	4.52	-0.25	-5.88
46	2.3:5.6-Dibenzophenanthrene	4.33	4.76	-0.43	-10.03
47	naphtho[2.1.8-qr]naphthacene	5.87	5.39	0.48	8.21
48	dibenz[a.e]aceantrylene	4.90	5.24	-0.34	-6.85
50	dibenzo[a.k]fluoranthene	4.90	5.35	-0.45	-9.22
52	fluoranthene	3.37	3.51	-0.14	-4.30
55	9H - fluorene	2.70	2.90	-0.20	-7.41
56	benzo[b]fluorene	3.84	3.44	0.40	10.44
57	benzo[c]fluorene	3.49	3.62	-0.13	-3.80
59	benzo[a]fluoranthene	4.22	4.59	-0.37	-8.66
60	indeno[1.2.3.cd]pyrene	4.84	4.82	0.02	0.52
61	indeno[1.2.3-cd] fluoranthene	4.93	4.73	0.20	4.06
64	benzo[j]fluoranthene	4.24	4.40	-0.16	-3.76
65	benzo[k]fluoranthene	4.42	4.26	0.16	3.56
66	benzo[a]fluorene	3.72	3.54	0.18	4.94
68	dibenz[e.k]acephenanthrylene	5.27	5.18	0.09	1.67
69	benzo[b]fluoranthene	4.29	4.17	0.12	2.80

Quadro 31: Valores para Índice de Retenção determinados experimentalmente e previstos pelo modelo (em negrito)

	Composto	IR		Composto	IR
1	naphthalene	2.00/ 2.28	36	benzo[a]perylene	5.54
2	anthracene	3.2	37	dibenzo[de. mn]naphthacene	5.65
3	phenanthrene	3	38	naphtho[2.3-g]chrysene	5.48
4	naphthacene	4.51	39	benzo[h]pentaphene	5.33
5	benz[a]anthracene	4	40	benzo[a]pentacene	6.13
6	chrysene	4.10	41	coronene	4.91
7	triphenylene	3.7	42	naphtho[1.2.3.4-def]chrysene	4.97
8	pyrene	3.58	43	dibenzo[def.p]chrysene	4.89
9	benzo[c]phenanthrene	3.64	44	benzo[rst]pentaphene	5.73
10	perylene	4.33	45	benzo[g]chrysene	4.27
11	benzo[a]pyrene	4.53	46	2.3:5.6-Dibenzophenanthrene	4.33
12	benzo[e]pyrene	4.28	47	naphtho[2.1.8-qla]naphthacene	5.87
13	picene	5.18/ 4.51	48	dibenz[a.e]aceantrylene	4.90
14	pentaphene	4.67	49	acenaphthylene	2.85
15	benzo[b]chrysene	5	50	dibenzo[a.k]fluoranthene	4.90
16	dibenz[a. h]anthracene	4.73	51	naphtho[2.3-k]fluoranthene	5.92/ 5.11
17	dibenz[a. j]anthracene	4.56	52	fluoranthene	3.37
18	benzo[b]triphenylene	4.40	53	dibenzo[k. mno]fluoranthene	4.53
19	benzo[c]chrysene	4.45	54	acenaphthene	2.57
20	pentacene	5.40	55	9H - fluorene	2.70
21	dibenzo[c. g]phenanthrene	4.07	56	benzo[b]fluorene	3.84
22	benzo[a]naphthacene	4.99	57	benzo[c]fluorene	3.49
23	dibenzo[b. def]chrysene	6	58	benzo[ghi]fluoranthene	3.78
24	dibenzo[def. mno]chrysene	5.08	59	benzo[a]fluoranthene	4.22
25	dibenzo[a.j]naphthacene	5.77	60	indeno[1.2.3.cd]pyrene	4.84
26	dibenzo[a. l]naphthacene	5.75	61	indeno[1.2.3-cd] fluoranthene	4.93
27	dibenzo[a.c]naphthacene	5.71	62	1H-fluorene	-
28	dibenzo[el]naphthacene	4.84	63	cyclopenta[cd]pyrene	4.12
29	dibenzo[de.gr]naphthacene	4.92	64	benzo[j]fluoranthene	4.24
30	dibenzo[g.p]chrysene	5.36	65	benzo[k]fluoranthene	4.42
31	benzo[c]picene	5.44	66	benzo[a]fluorene	3.72
32	benzo[ghi]perylene	4.76	67	1H-benzo[b]fluorene	3.47
33	dibenzo[b.k]chrysene	5.80	68	dibenz[e.k]acephenanthrylene	5.27
34	dibenzo[cl]chrysene	5.34	69	benzo[b]fluoranthene	4.29
35	benzo[b]perylene	5.04	70	indan	1.59

Os valores de índice de retenção utilizados para a construção do modelo de regressão encontram-se numa faixa entre 2 (naftaleno) e 6.00 (dibenzo[b.def]criseno). Apenas um dos valores previstos foi extrapolado e encontra-se acima desta faixa, como pode ser observado no quadro acima.

Discussão dos Resultados para Índice de Retenção

O índice de retenção é um parâmetro utilizado em métodos cromatográficos para separar misturas de hidrocarbonetos, e indica a ordem de eluição de cada composto. Os fatores que determinam esta ordem de eluição são complexos, e envolvem as interações entre a fase móvel, constituída pela amostra contendo os hidrocarbonetos, e a fase estacionária, constituída pela coluna cromatográfica por onde a fase móvel tem que passar. Estas interações são muito difíceis de serem descritas em detalhes, mas estão intimamente relacionadas à estrutura molecular das duas fases, sendo que as variações no valor do índice de retenção se devem às variações estruturais dos compostos analisados. Por exemplo: compostos maiores tendem a ter maiores índices de retenção, pois possuem uma maior área superficial, que permite um contato maior, favorecendo interações mais fortes com a fase estacionária e conseqüentemente uma demora no tempo de eluição em relação aos compostos menores. Para ilustrar este tipo de comportamento pode-se comparar os valores de índices de retenção do antraceno ($I_r = 3.2$) e do naftaceno ($I_r = 4.51$), cujas estruturas podem ser observadas nas figura 49(a) e 49(b).

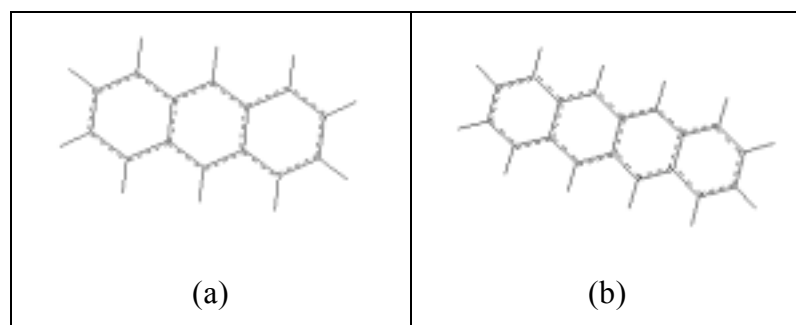


Figura 49: Estruturas do antraceno (a) e do naftaceno (b).

Para estimar o tamanho do composto são utilizados parâmetros como peso molecular, volume e área superficial, dos quais apenas o peso molecular foi utilizado para a construção do modelo.

Todavia, a presença de ramificações no esqueleto molecular, a disposição dos ciclos e anéis, ou a planaridade da estrutura podem alterar esta ordem de eluição. Sander & Wise (1987) citam a influência da razão entre o comprimento e a largura dos compostos na ordem de eluição, e conseqüentemente, no valor do índice de retenção.

Num grupo de isômeros de mesmo peso molecular, os compostos com maior razão entre comprimento e largura, isto é, aqueles compostos que apresentam uma estrutura mais linear, demoram mais para eluir e apresentam um índice de retenção maior.

Para ilustrar este tipo de comportamento podemos retornar ao exemplo do naftaceno ($I_r = 4.51$), e compará-lo com os seus isômeros de mesmo peso molecular benz[a]antraceno ($I_r = 4.00$), criseno ($I_r = 4.10$) e trifenileno ($I_r = 3.7$). O índice de retenção é maior para aquelas moléculas mais lineares. As estruturas podem ser observadas na figura 50.

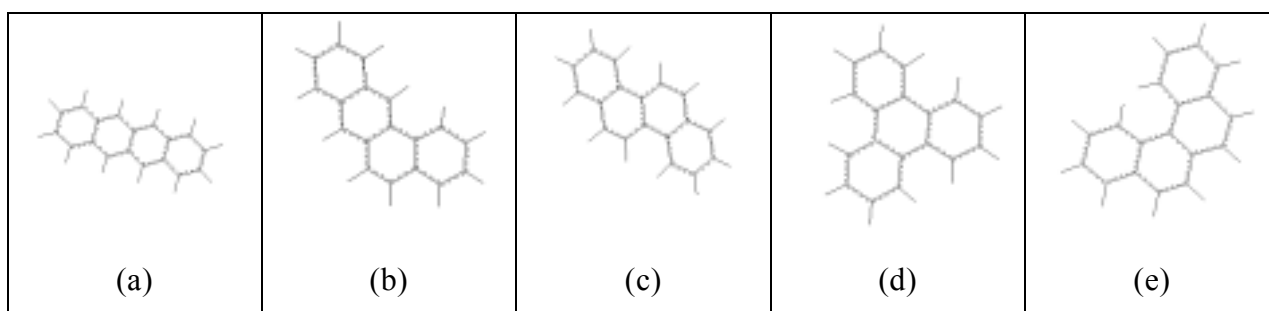


Figura 50: Estruturas do naftaceno (a), benz[a]antraceno (b), criseno (c), trifenileno (d) e benzo[c]fenantreno (e).

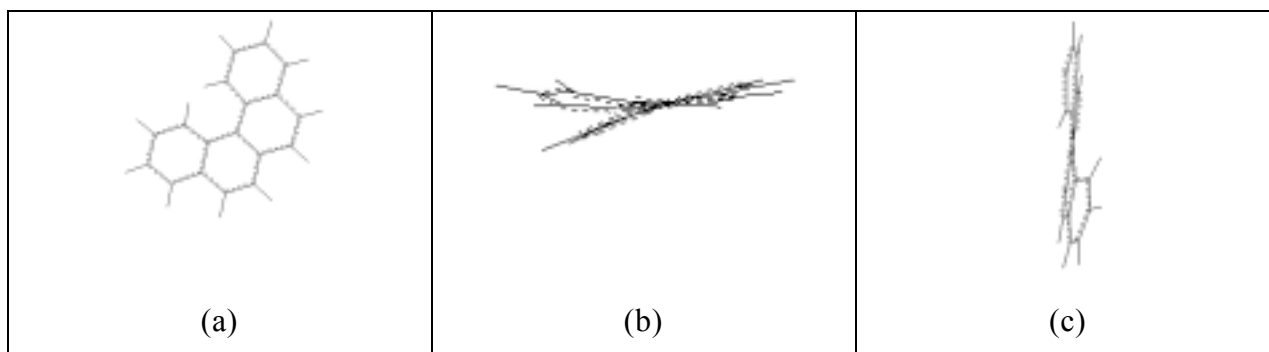


Figura 51: Estrutura do benzo[c]fenantreno mostrando a não planaridade da molécula: (a) visão inicial; (b) rotação vertical; (c) rotação horizontal.

No entanto, o benzo[c]fenantreno ($I_r = 3.64$) constitui uma exceção a esta regra. Isto acontece porque agora entra em jogo a interferência de um outro fator: a planaridade do composto. Dentre todos os isômeros utilizados no exemplo, o benzo[c]fenantreno é o único que apresenta uma estrutura ligeiramente não planar e helicoidal (os outros são planares), devido à repulsão estérica entre os átomos de

hidrogênio dos anéis que se encontram muito próximos (Figura 51), e esta peculiaridade se reflete na sua ordem de eluição (Sander & Wise, 1988).

Existem vários descritores topológicos que informam estes detalhes estruturais dos compostos: número de ciclos, quantidade de átomos de carbono no anel (5 ou 6), índice de conectividade Randic etc. Neste trabalho, o índice que apresentou melhores resultados para a construção do modelo de regressão foi o índice de *Wiener* (WIT), que de uma forma geral descreve o grau de compactação da molécula levando em consideração as distâncias entre os pares de átomos, e o número de átomos de carbono existentes entre eles (Wiener, 1947).

Outro parâmetro que interfere na força das interações entre a fase móvel (a amostra contendo os hidrocarbonetos) e a fase estacionária é a *polarizabilidade* do composto, que indica a susceptibilidade à formação de dipolos instantâneos, ou dipolos temporários ao longo da estrutura química. A polarizabilidade é inversamente proporcional à *dureza* de uma espécie, e conseqüentemente está correlacionada com os orbitais de fronteira (Shriver *et alli*, 1994).

Quanto mais a estrutura permitir a formação de dipolos (quanto mais *polarizável* e portanto menos *dura* ela for), mais intensa será a interação entre a fase móvel e a fase estacionária, e portanto, maior o índice de retenção. Se observarmos os gráficos de índice de retenção *vs.* polarizabilidade no apêndice II veremos que existe uma curva ascendente.

A estrutura será mais suscetível à formação de dipolos quanto menos dura (ou mais mole) ela for. Observando o gráfico de índice de retenção *vs.* dureza no apêndice II pode-se observar que ocorre a formação de uma curva descendente, por isto possui valor negativo no vetor de regressão ($\eta = -0.1246$).

Apesar da alta correlação entre as variáveis energia do LUMO, polarizabilidade e dureza, que poderia implicar na repetição de informações, o modelo de regressão que apresentou os melhores resultados na previsão do índice de retenção foi aquele que continha estas três variáveis simultaneamente. A retirada de qualquer uma delas resulta numa piora do modelo.

O modelo apresentou erro de validação em torno de 10%, que é considerado satisfatório para estudos de QSAR, com coeficiente de correlação 0.9477.

Apresentação dos Resultados para Fototoxidez

Para a construção do modelo de regressão para fototoxidez, foi utilizado um conjunto de 14 moléculas com valores determinados experimentalmente retirados do trabalho de Mekenyan *et alli* (1994) para prever os valores de outras 56 moléculas. A fototoxidez foi expressa em termos de tempo letal ajustado médio, ALT⁵³ (*lethal time*, LT50, normalizado à uma constante de concentração).

Para tanto, foram selecionadas as variáveis eletrônicas energia do HOMO (E_{HOMO}), energia do LUMO (E_{LUMO}) e o GAP (intervalo) entre estes dois parâmetros.

A correlação existente entre estes parâmetros e a variável dependente não é linear, como pode ser observado nas figuras 52, 53 e 54. Foi feito então um ajuste para linearizar esta relação, através da utilização de Gaussianas, ajustadas de acordo com as equações 40, 41 e 42.

$$E_{HOMO}' = \exp[-(E_{HOMO} - 8.06)^2] \quad (40)$$

$$E_{LUMO}' = \exp[-(E_{LUMO} - 0.91)^2] \quad (41)$$

$$GAP' = \exp[-(GAP - 7.2)^2] \quad (42)$$

Os gráficos destas variáveis já linearizadas em relação à fototoxidez podem ser observados nas figuras 55, 56 e 57.

⁵³ ALT = log(1/LT50), em que ALT = Adjusted Lethal Time.

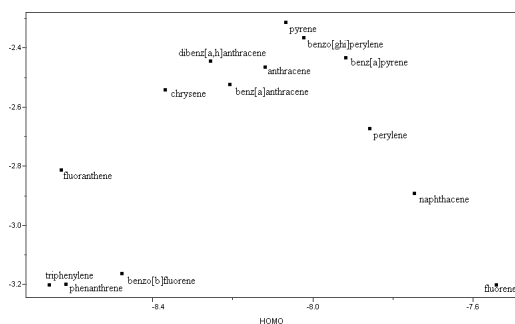


Figura 52: Gráfico de Fototoxidez vs. Energia do HOMO.

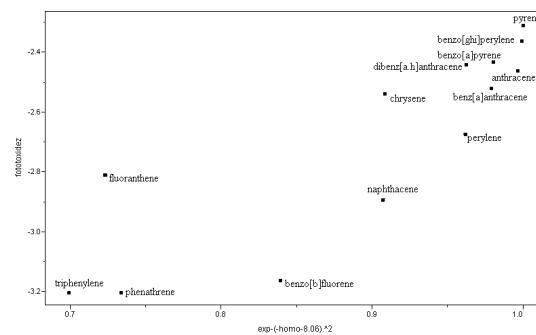


Figura 55: Gráfico de Fototoxidez vs. Energia do HOMO, após a linearização.

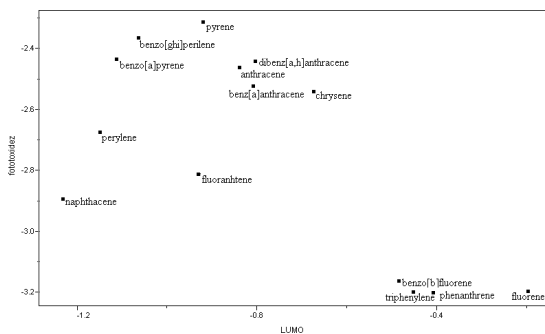


Figura 53: Gráfico de Fototoxidez vs. Energia do LUMO.

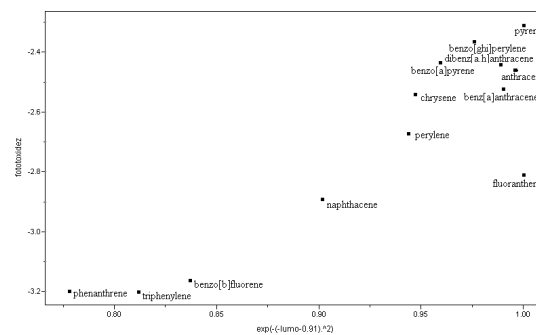


Figura 56: Gráfico de Fototoxidez vs. Energia do LUMO, após a linearização.

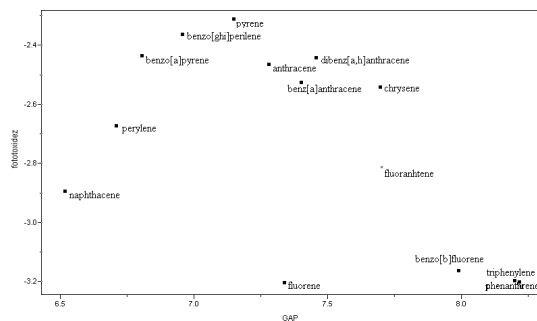


Figura 54: Gráfico de Fototoxidez vs. GAP.

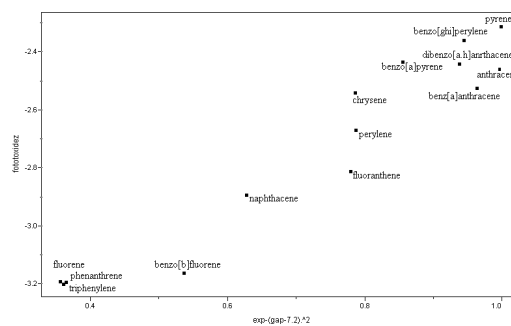


Figura 57: Gráfico de Fototoxidez vs. GAP, após a linearização.

No quadro 32 podem ser observados os coeficientes que cada variável tem no vetor de regressão.

Quadro 32: Coeficientes de cada variável no vetor de regressão para Fototoxidez.

variáveis	β
$\exp(-(\text{Elumo}-0.91)^2)$	0.47439
$\exp(-(\text{Ehomo}-8.06)^2)$	0.36969
$\exp(-(\text{GAP}-7.2)^2)$	0.28819

Os resultados da análise de regressão pelo método dos mínimos quadrados parciais para a fototoxidez são mostrados no quadro 33. O número ótimo de variáveis latentes utilizadas para o modelo desenvolvido foi 2 e o modelo apresentou boa qualidade, com coeficientes de correlação 0.967.

Quadro 33: Resultados dos modelos de QSPR utilizando PLS com validação cruzada para Fototoxidez

Variáveis Latentes	SEV	PRESS	% de Variância	r
1	0.0989	0.137	72.51	0.9637
2	0.1013	0.1438	91.1	0.9671
3	0.1868	0.4886	100	0.9674

Na figura 58 é apresentado o gráfico de valor medido vs. valor previsto do melhor modelo obtido, cujos valores com os erros de validação podem ser encontrados no quadro 34.

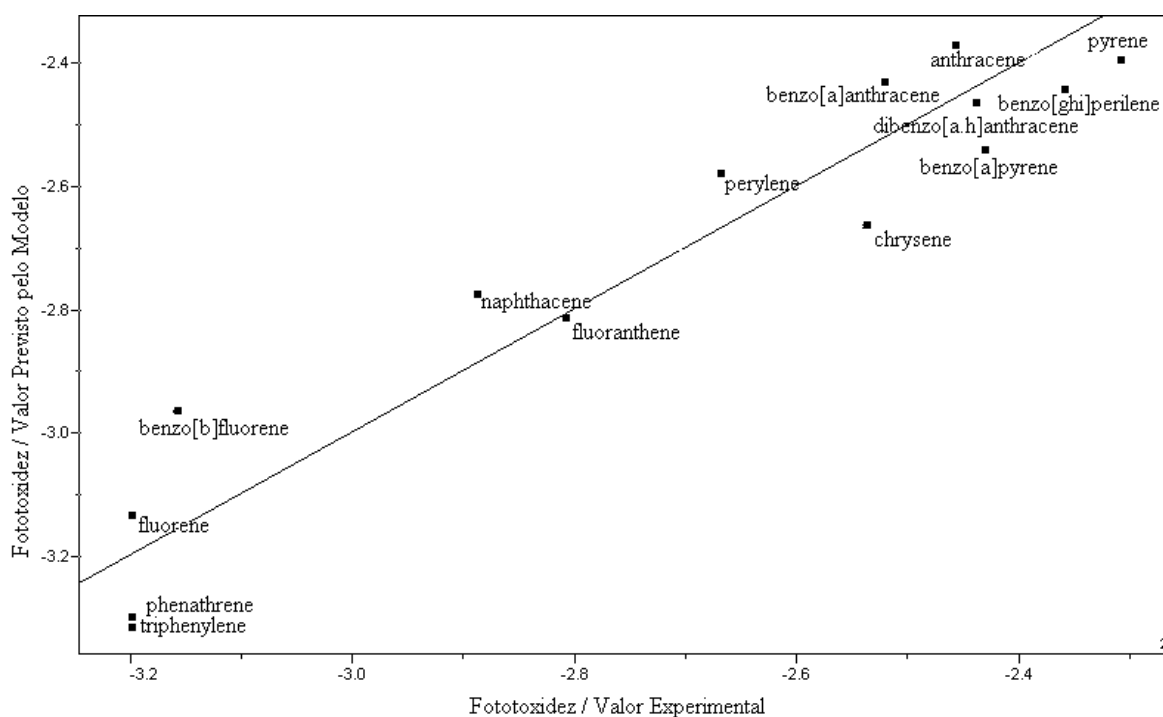


Figura 58: Gráfico de valor medido vs. valor previsto para Fototoxidez.

Quadro 34: Erros de Validação para Fototoxidez.

	Composto	Valor Medido	Valor Previsto	Resíduos	Erros de Validação (%)
2	anthracene	-2.46	-2.37	-0.09	3.68
3	phenanthrene	-3.2	-3.31	0.11	-3.48
4	naphthacene	-2.89	-2.77	-0.12	4.04
5	benz[a]anthracene	-2.52	-2.43	-0.09	3.49
6	chrysene	-2.54	-2.66	0.12	-4.79
7	triphenylene	-3.2	-3.30	0.10	-3.05
8	pyrene	-2.31	-2.39	0.08	-3.67
10	perylene	-2.67	-2.58	-0.09	3.44
11	benzo[a]pyrene	-2.43	-2.54	0.11	-4.40
16	dibenz[a. h]anthracene	-2.44	-2.46	0.02	-0.88
32	benzo[ghi]perylene	-2.36	-2.44	0.08	-3.36
52	fluoranthene	-2.81	-2.81	0.00	-0.13
55	fluorene	-3.2	-3.13	-0.07	2.09
56	benzo[b]fluorene	-3.16	-2.96	-0.20	6.27

Todos os erros de validação encontram-se abaixo de 5%, exceto para a molécula de benzo[b]fluoreno que apresentou um erro de 6.27%. Um novo modelo foi testado, sem esta amostra, mas não houve uma melhora significativa no resultado obtido. Decidiu-se manter este composto no conjunto de modelagem.

O quadro 35 apresenta os valores medidos experimentalmente e os valores previstos pelo modelo.

Quadro 35: Valores para Fototoxidez determinados experimentalmente e previstos pelo modelo (em negrito)

	Composto	fototoxidez		Composto	fototoxidez
1	naphthalene	-3.59	36	benzo[a]perylene	-3.38
2	anthracene	-2.46	37	dibenzo[de. mn]naphthacene	-3.58
3	phenanthrene	-3.20	38	naphtho[2.3-g]chrysene	-2.39
4	naphthacene	-2.89	39	benzo[h]pentaphene	-2.49
5	benz[a]anthracene	-2.52	40	benzo[a]pentacene	-3.30
6	chrysene	-2.54	41	coronene	-2.40
7	triphenylene	-3.20	42	naphtho[1.2.3.4-def]chrysene	-2.43
8	pyrene	-2.31	43	dibenzo[def.p]chrysene	-2.49
9	benzo[c]phenanthrene	-2.76	44	benzo[rst]pentaphene	-2.63
10	perylene	-2.67	45	benzo[g]chrysene	-2.50
11	benzo[a]pyrene	-2.43	46	2.3:5.6-Dibenzophenanthrene	-2.39
12	benzo[e]pyrene	-2.95/-2.42	47	naphtho[2.1.8-qr]naphthacene	-2.91
13	picene	-2.59	48	dibenz[a.e]aceantrylene	-2.62
14	pentaphene	-2.42	49	acenaphthylene	-3.19
15	benzo[b]chrysene	-2.40	50	dibenzo[a.k]fluoranthene	-2.71
16	dibenz[a. h]anthracene	-2.44	51	naphtho[2.3-k]fluoranthene	-2.39
17	dibenz[a. j]anthracene	-2.41	52	fluoranthene	-2.81
18	benzo[b]triphenylene	-2.43	53	dibenzo[k. mno]fluoranthene	-2.53
19	benzo[c]chrysene	-2.65	54	acenaphthene	-3.42
20	pentacene	-3.57	55	9H - fluorene	-3.20
21	dibenzo[c. g]phenanthrene	-2.63	56	benzo[b]fluorene	-3.16
22	benzo[a]naphthacene	-2.65	57	benzo[c]fluorene	-2.61
23	dibenzo[b. def]chrysene	-3.03	58	benzo[ghi]fluoranthene	-2.86
24	dibenzo[def. mno]chrysene	-3.15	59	benzo[a]fluoranthene	-2.68
25	dibenzo[a.j]naphthacene	-2.53	60	indeno[1.2.3.cd]pyrene	-2.62
26	dibenzo[a. l]naphthacene	-2.53	61	indeno[1.2.3-cd] fluoranthene	-2.85
27	dibenzo[a.c]naphthacene	-2.56	62	1H-fluorene	<3.20
28	dibenzo[el]naphthacene	-2.48	63	cyclopenta[cd]pyrene	-2.67
29	dibenzo[de.gr]naphthacene	-2.46	64	benzo[j]fluoranthene	-2.55
30	dibenzo[g.p]chrysene	-2.38	65	benzo[k]fluoranthene	-2.89/-2.45
31	benzo[c]picene	-2.45	66	benzo[a]fluorene	-3.17/-2.78
32	benzo[ghi]perylene	-2.36	67	1H-benzo[b]fluorene	-2.99
33	dibenzo[b.k]chrysene	-2.61	68	dibenz[e.k]acephenanthrylene	-2.44
34	dibenzo[cl]chrysene	-2.48	69	benzo[b]fluoranthene	-2.76
35	benzo[b]perylene	-2.62	70	indan	<3.20

Os valores de fototoxidez utilizados para a construção do modelo de regressão encontram-se numa faixa entre -3.20 (fenantreno, trifenileno e fluoreno) e -2.36

(benzo[ghi]perileno). Alguns dos valores previstos foram extrapolados e encontram-se acima desta faixa, como pode ser observado no quadro acima.

Discussão dos Resultados para Fototoxidez

A toxidez de um composto é o resultado da influência de uma série de parâmetros conjugados, que podem estar tanto relacionados à fatores internos, como a estrutura dos compostos, quanto à fatores externos, como as condições ambientais que permitem ao composto atuar como poluente: exposição à luz, intensidade da luz irradiada, tempo de exposição etc.; e nos casos em que esta toxidez é estimada por meio de testes laboratoriais para se fornecer um valor, o método utilizado também é considerado um fator importante.

Como mostrado anteriormente (Mekenyan *et alli*, 1994) os descritores eletrônicos HOMO, LUMO e GAP, utilizadas neste trabalho, apresentam uma correlação não linear com a fototoxidez dos compostos (Figuras 53, 54 e 55). Este comportamento é resultante da competição entre fatores como a estabilidade química dos compostos, a absorção de energia necessária para induzir à fototoxidez e a intensidade da radiação à qual o composto foi exposto.

Os orbitais de fronteira, HOMO e LUMO, podem ser usados respectivamente para estimar a capacidade molecular de aceitar ou doar um par eletrônico, e o GAP expressa a energia necessária para excitar um elétron do HOMO para o LUMO. Sendo assim, a dimensão destas variáveis deveria ser proporcional à energia necessária para induzir à fototoxidez. Por exemplo, moléculas com um GAP grande absorvem em menores comprimentos de onda, em regiões de maior energia ($GAP = h\nu$), e deveriam estar relacionados com grandes valores de fototoxidez.

No entanto, como pode-se observar nas Figuras 53, 54 e 55, esta correlação não é linear. Num primeiro momento, a curva de fototoxidez aumenta até um valor máximo, e começa a decrescer. Este fenômeno está relacionado à competição entre fatores externos e fatores internos, como descrito a seguir:

Absorção de luz vs. estabilidade química:

O GAP é utilizado para estimar a dureza (η) do composto, ($\eta = \text{GAP}/2$), e é um parâmetro relacionado à estabilidade molecular (Atkins, 1994; Zhou & Parr, 1989). A dureza absoluta é utilizada para estimar a ennergia de estabilização em sistemas aromáticos. Compostos aromáticos que apresentam um pequeno valor de dureza (GAP pequeno) são considerados menos estáveis do que aqueles com um grande intervalo entre os seus orbitais de fronteira. Se para um determinado composto a toxidez é característica do composto em si (e não dos seus derivados), o entendimento da estabilidade deste composto será necessário para estimar o impacto ambiental deste no meio ambiente. Se o composto se degrada facilmente (é *mole*, *i. e.* pequeno GAP), sua persistência no meio ambiente será menor, resultando num menor potencial tóxico.

Quanto menor o GAP de um composto, menor a energia necessária para induzir à fototoxidez e portanto, maior deveria ser seu potencial tóxico. No entanto, estas moléculas apresentam menor estabilidade química que aquelas com um grande GAP, o que resulta num menor potencial fototóxico. Estas moléculas encontram-se do lado esquerdo das curvas nas Figuras 52, 53 e 54.

Absorção de luz vs. intensidade da radiação:

Moléculas com um grande GAP (lado direito das curvas na Figuras 52, 53 e 54) absorvem radiação em comprimentos de onda maiores (menor enegria) e apresentam maior estabilidade por causa dos altos valores de dureza, e deveriam portanto exibir maior fototoxidez. No entanto, fatores externos como a energia e a intensidade da radiação também afetam o potencial fototóxico (Ankley et alli, 1995; Mekenyan et alli, 1994; Schaeffer & Larson, 1999). No meio ambiente, sob a luz natural do sol, os organismos estão expostos às radiações UV-A, UV-B e visível, com intensidades que variam de acordo com a ordem $I_{\text{UV-B}} < I_{\text{UV-A}} < I_{\text{visível}}$. Desta maneira, moléculas com um GAP pequeno absorvem na região do visível, moléculas com um GAP intermediário absorvem na região do UV-A e moléculas com um GAP maiores absorvem na região do UV-B (Mekenyan *et alli*, 1994). Compostos com grandes intervalos de energia entre os

orbitais de fronteira, em que seria esperada uma maior toxidez devido à alta absorção de radiação e à grande estabilidade dos compostos, absorvem fortemente numa região de menores comprimentos de onda, que não estão presentes na radiação solar que atinge a superfície do planeta, e portanto não encontram energia suficiente no meio ambiente, para ativar a sua fototoxidez (Mekenyan *et alli*, 1994).

O resultado da competição entre estes fatores é uma curva que lembra uma gaussiana, entre as variáveis eletrônicas e a propriedade estudada. Como estimado por Mekenyan *et alli* (1994), a máxima fototoxidez ocorre para um valor de GAP de $7,2 \pm 0,4$ eV, que caracteriza um tempo letal ajustado ($\log (1/ALT)) \leq -2.95$).

Com base nestas informações e levando-se em consideração os ajustes descritos pelas equações 40-42, propõe-se neste trabalho duas novas correlações para se estimar a fototoxidez de PAHs, baseadas nas energias dos orbitais de fronteira. É possível observar nas curvas das figuras 52, 53 e 54 que dentre as moléculas estudadas, aquelas que apresentaram um potencial fototóxico ($\log (1/ALT)) \leq -2.95$) encontram-se na região central, numa faixa de E_{HOMO} de -8.06 ± 0.20 eV, e numa faixa de E_{LUMO} ao redor de -0.91 ± 0.20 eV. Nas figuras 59, 60 e 61 podem ser observadas as curvas de fototoxidez vs. orbitais de fronteira para todo o conjunto de PAHs estudados, inclusive aqueles cujos valores foram previstos pelo modelo.

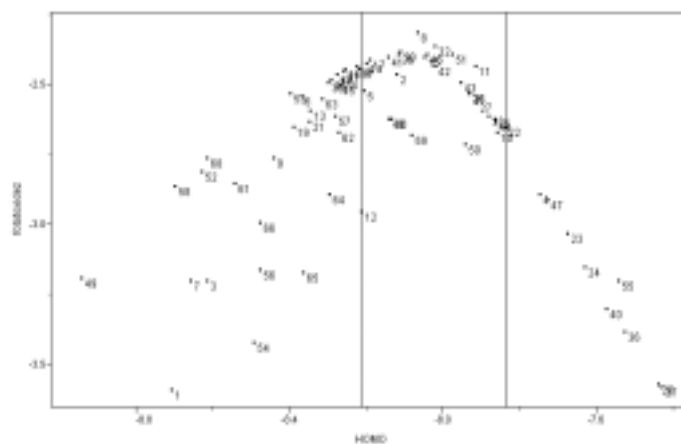


Figura 59: Gráfico de Fototoxidez vs. Energia do HOMO para todos os PAHs.

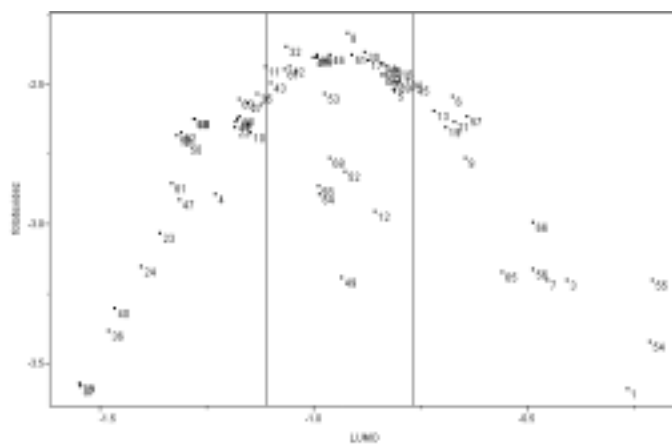


Figura 60: Gráfico de Fototoxidez vs. Energia do LUMO para todos os PAHs.

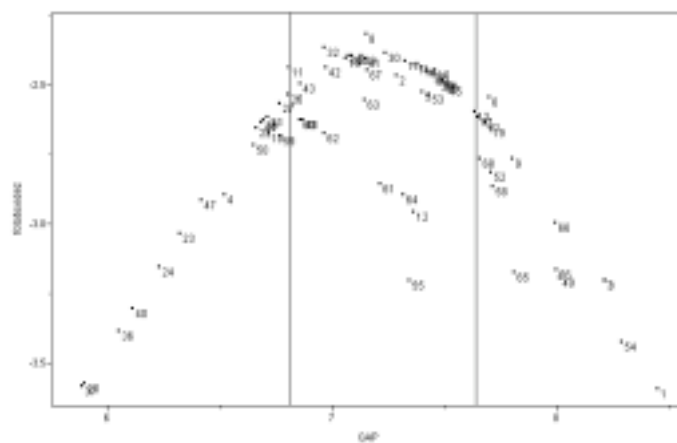


Figura 61: Gráfico de Fototoxidez vs. GAP para todos os PAHs.

É possível observar que aqueles compostos que apresentam fototoxicidade encontram-se ao redor do ponto de máximo da gaussiana, dentro dos intervalos descritos anteriormente. Pela utilização do modelo de previsão construído à partir dos orbitais de fronteira foi possível estimar que os seguintes compostos são potencialmente fototóxicos:

Piceno	dibenzo[g.p]criseno	nafto[2.1.8-qla]naftaceno
Pentafeno	benzo[c]piceno	dibenzo[a.e]aceantrileno
Benzo[b]criseno	dibenzo[b.k]criseno	dibenzo[a.k]fluoranteno
Dibenz[a.j]antraceno	dibenzo[cl]criseno	nafto[2.3-k]fluoranteno
Benzo[b]trifenileno	benzo[b]perileno	dibenzo[k.mno]fluoranteno
Benzo[c]criseno,	nafto[2.3-g]criseno	benzo[c]fluoreno
Dibenzo[c.g]fenantreno	benzo[h]pentafeno	benzo[ghi]fluoranteno
Benzo[a]naftaceno	coroneno	benzo[a]fluoranteno
Dibenzo[a.j]naftaceno	nafto[1.2.3.4-def]criseno	indeno[1.2.3.cd]pireno
Dibenzo[a.l]naftaceno	dibenzo[def.p]criseno	ciclopenta[cd]pireno
Dibenzo[a.c]naftaceno	benzo[rst]pentafeno	benzo[j]fluoranteno
Dibenzo[el]naftaceno	benzo[g]criseno	1H-benzo[b]fluoreno
Dibenzo[de.gr]naftaceno	1.2:5.6-dibenzofenantreno	dibenz[e.k]acefenantrileno

Conclusão

Este trabalho apresentou duas diferentes aplicações de métodos quimiométricos em áreas bem distintas: uma em química ambiental e outra em química médica, com a intenção de divulgar as possibilidades da utilização de estatística multivariada nos mais diversos campos da química.

No primeiro capítulo foi apresentada uma breve introdução aos métodos de análise multivariada, enfatizando as duas técnicas empregadas neste trabalho: *a análise de componentes principais* (PCA) e *a regressão por mínimos quadrados parciais* (PLS).

No segundo capítulo a PCA foi aplicada num estudo do perfil de distribuição de hidrocarbonetos policíclicos aromáticos (PAH) em amostras de sedimento coletadas ao longo de duas baías em *Prince William Sound*, no Alasca, onde em 1989 ocorreu o derramamento de óleo do petroleiro *Exxon Valdez*.

Num trabalho como este, os teores de cada grupo de hidrocarbonetos têm que ser analisados um a um, e é muito difícil detectar as suas correlações, como por exemplo se o aumento de um determinado PAH está relacionado ao aumento de um outro PAH; ou se com o aumento da profundidade ocorre uma maior incidência deste PAH e não daquele outro. A procura por estas informações leva muito tempo, e como não é possível analisar todos os dados de uma única vez, é bem possível que alguma informação, possa não ser detectada, como por exemplo uma correlação entre compostos, ou a influência de uma determinada variável na descrição de uma das duas baías.

Pela utilização de análise de componentes principais é possível ter uma visão geral do conjunto de dados e de todas as correlações possíveis entre as variáveis, sempre de uma forma global, inteira, e portanto, muito mais completa.

Isto pode ser visto na identificação das variáveis que contribuem para a caracterização de cada uma das duas baías, por exemplo. Através da observação do gráfico dos escores e dos pesos, foi possível visualizar rapidamente que, apesar de haver a presença de todos estes hidrocarbonetos ao longo das duas baías, o perfil de distribuição dos teores destes compostos é muito diferenciado para cada região, e em

cada profundidade. A utilização da PCA permitiu identificar que algumas variáveis, como D2/P2, C13D e C14C estão relacionadas ao aporte de PAHs provocado pelo derramamento, e portanto contribuem para a caracterização da baía que foi mais atingida pelo acidente com o petroleiro; assim como relacionou à variável Pyro com a proximidade de pontos de maior aporte de hidrocarbonetos de origem pirogênica, na baía utilizada como referência por não ter sido contaminada tão diretamente.

Foi possível ainda observar a distribuição destes compostos ao longo das profundidades em cada uma das duas baías, e perceber que o óleo derramado foi aos poucos se depositando nas áreas costeiras, devido ao movimento das correntes marítimas.

Outro ponto interessante é a detecção de amostras com um comportamento diferente do restante do conjunto, os *outliers*. Estas amostras apresentaram teores de alguns PAHs muito altos, ou muito baixos, e geralmente foram coletadas em regiões onde o aporte deste PAHs era diferenciado. Por exemplo as amostras R081, R091, F041 e R032, que apresentam alto teor da variável Pyro, e que estão localizadas perto de regiões com maior atividade humana e das ruínas de uma antiga mina de carvão.

O método permitiu detectar estas amostras e rapidamente identificar quais as variáveis que contribuem para este comportamento peculiar, como por exemplo um teor de um certo grupo de PAH mais alto que o usual. Sem a aplicação do método, isto somente seria possível após um longo e cuidadoso exame dos teores de cada tipo de hidrocarboneto e a comparação com as amostras da mesma região.

Com esta pequena aplicação de um método multivariado no tratamento de dados químicos é possível avaliar as significativas contribuições que a quimiometria pode oferecer à química ambiental. Este tipo de ferramenta permite otimizar a extração de informação química de uma grande quantidade de dados de forma rápida e objetiva.

No terceiro capítulo foi mostrado um estudo das relações entre estrutura-propriedades e estrutura-atividade (QSPR/QSAR) de um conjunto de 70 hidrocarbonetos policíclicos aromáticos (PAHs).

Foram escolhidos alguns descritores moleculares calculados que apresentassem características geométricas, eletrônicas e topológicas para a construção de modelos de

regressão para 3 propriedades moleculares e 1 atividade: ponto de ebulição, coeficiente de partição octanol-água, índice de retenção e fototoxidez.

Estes modelos foram construídos por meio da utilização do *método dos mínimos quadrados parciais* com *preprocessamento autoescalado* e *validação cruzada*. Uma vez validados, estes modelos foram utilizados para prever estas propriedades para moléculas que não possuem estes valores estimados experimentalmente.

O estudo destas relações constitui uma importante ferramenta na pesquisa e conhecimento de compostos químicos, e tem sido largamente empregado em química farmacêutica e medicinal, em modelagem molecular e na investigação de novas drogas.

Muitas vezes, a determinação experimental das propriedades de um composto não é uma tarefa simples. O seu manuseio pode envolver algum risco, ou ser difícil, no caso dos compostos que se decompõem rapidamente. Neste caso, os estudos de QSPR permitem estimar estas propriedades com um excelente grau de confiança e assim, evitar os riscos que este tipo de determinação pode apresentar.

Este foi o caso da determinação do ponto de ebulição, do coeficiente de partição octanol-água e do índice de retenção, para o conjunto de PAHs escolhidos. Em geral, alguns destes compostos são muito tóxicos, outros são de difícil manuseio, pois sublimam facilmente, o que torna a determinação destas propriedades um tanto difícil.

No caso do coeficiente de partição octanol-água ainda, estes compostos são extremamente hidrofóbicos, e praticamente quase não se dissolvem na fase aquosa, o que dificulta sua determinação experimental.

Esta técnica é muito investigada também nos estudos da atividade biológica de compostos homólogos. Muitas vezes, o analista se vê diante de uma batelada de compostos com possível potencial tóxico, e os testes experimentais envolvem uma grande quantidade de delicadas operações que são muito suscetíveis a erros. A utilização de métodos de QSAR permite estimar por meio de cálculos e análise estrutural, quais seriam as moléculas que possivelmente apresentariam uma determinada atividade biológica. Este é o caso da fototoxidez dos PAHs.

Em geral, o custo da determinação experimental destes parâmetros é muito elevado. Os estudos de QSAR/QSPR permitem reduzir este custo. Desta forma, é possível descobrir quais as moléculas que poderiam apresentar atividade biológica e

somente então realizar as investigações para apenas estas moléculas, minimizando muito o custo de um projeto.

Apêndice I

Apresenta a tabela de dados dos estudos de QSAR e QSPR apresentados no capítulo 3.

	nome no cas	Q-	Q+	dipolo	cccpe	E lumo	E homo	GAP	area	volume
1	naphtalene	-0.176574	0.123663	0.001145	0.0081	-0.2650	-8.7099	8.4449	159.0527	131.9503
2	anthracene	-0.242131	0.116172	0.003088	0.0084	-0.8417	-8.1212	7.2795	206.4759	177.2617
3	phenanthrene	-0.1648	0.122816	0.022734	0.0233	-0.4085	-8.6171	8.2086	203.1945	176.7936
4	naphthacene	-0.223292	0.137431	0.00691	0.0218	-1.2321	-7.7488	6.5167	253.8928	222.5818
5	benzo[a]anthracene	-0.219814	0.135023	0.050532	0.03	-0.8116	-8.2079	7.3963	250.6318	222.1094
6	chrysene	-0.175844	0.123409	0.002475	0.0106	-0.6762	-8.3697	7.6935	246.9786	221.5272
7	triphenylene	-0.135509	0.111853	0.001518	0.0129	-0.4532	-8.6584	8.2052	243.0391	220.8467
8	pyrene	-0.171165	0.124824	-0.002136	0.0192	-0.9225	-8.0692	7.1467	217.8428	193.5795
9	benzo[c]phenanthrene	-0.143929	0.127338	0.047076	0.0625	-0.6456	-8.4438	7.7982	245.6475	221.1480
10	perylene	-0.16159	0.120907	0.036483	0.0312	-1.1508	-7.8598	6.7090	258.7597	238.2124
11	benzo[a]pyrene	-0.242926	0.140925	0.033749	0.0294	-1.1142	-7.9173	6.8031	262.2384	238.6522
12	benzo[e]pyrene	-0.155323	0.125033	0.032207	0.022	-0.8580	-8.2149	7.3569	258.5826	238.0639
13	picene	-0.176308	0.124354	0.030102	0.0305	-0.7209	-8.3487	7.6278	290.8768	266.2941
14	pentaphene	-0.220436	0.137339	0.082173	0.0474	-0.8400	-8.2022	7.3622	298.0454	267.3996
15	benzo[b]chrysene	-0.224813	0.143632	0.029982	0.0368	-0.9948	-8.0511	7.0563	294.3653	266.8390
16	dibenzo[a, h]anthracene	-0.215794	0.126384	0.007559	0.0121	-0.8041	-8.2570	7.4529	294.6701	266.9324
17	dibenzo[a, j]anthracene	-0.241648	0.138265	0.120375	0.095	-0.8736	-8.1916	7.3180	294.3490	266.7377
18	benzo[b]triphenylene	-0.214559	0.143071	0.078096	0.0729	-0.8319	-8.2255	7.3936	290.0115	266.0922
19	benzo[c]chrysene	-0.198909	0.127443	0.042745	0.0247	-0.6931	-8.3898	7.6967	291.2930	266.3563
20	pentacene	-0.202783	0.134356	0.109838	0.1243	-1.5500	-7.4414	5.8914	301.2299	267.9507
21	dibenzo[c, g]phenanthrene	-0.149569	0.120154	0.030797	0.0257	-0.6732	-8.3498	7.6766	290.0853	266.0904
22	benzo[a]naphthacene	-0.233668	0.141217	0.062465	0.052	-1.1857	-7.8407	6.6550	298.0321	267.4128
23	dibenzo[b, def]chrysene	-0.255897	0.144586	0.047277	0.038	-1.3630	-7.6784	6.3154	306.1395	283.5631
24	dibenzo[def, mno]chrysene	-0.229818	0.136143	0.021319	0.0192	-1.4067	-7.6315	6.2248	277.1642	255.5712
25	dibenzo[a, j]naphthacene	-0.220606	0.141016	0.005792	0.0179	-1.1349	-7.9321	6.7972	342.2179	312.2648
26	dibenzo[al]naphthacene	-0.220085	0.137888	0.102594	0.0645	-1.1352	-7.9345	6.7993	342.1814	312.2526
27	dibenzo[a, c]naphthacene	-0.237341	0.14395	0.064774	0.0671	-1.1564	-7.9148	6.7584	337.8583	311.4491
28	dibenzo[e]naphthacene	-0.137078	0.1159	0.005309	0.0156	-0.8276	-8.2948	7.4672	298.4228	282.1115
29	dibenzo[de, gr]naphthacene	-0.21349	0.142989	0.081288	0.0819	-0.8336	-8.2774	7.4438	306.0306	283.3872
30	dibenzo[g, p]chrysene	-0.128664	0.112429	0.004240	0.0087	-0.8832	-8.1128	7.2296	329.5913	310.4633
31	benzo[c]picene	-0.170425	0.125814	0.046107	0.0466	-0.8279	-8.2597	7.4318	334.5104	311.1329
32	benzo[ghi]perylene	-0.168448	0.127496	0.068652	0.0497	-1.0662	-8.0235	6.9573	273.9605	255.1830
33	dibenzo[b, k]chrysene	-0.230923	0.141216	0.002804	0.0248	-1.1775	-7.8832	6.7057	341.7884	312.1273
34	dibenzo[cl]chrysene	-0.216487	0.128877	0.059064	0.0163	-0.7827	-8.2649	7.4822	335.5407	311.1978
35	benzo[b]perylene	-0.16358	0.135591	0.053555	0.062	-1.1806	-7.8666	6.6860	302.4596	282.8705
36	benzo[a]perylene	-0.225428	0.131065	0.124478	0.1234	-1.4836	-7.5284	6.0448	302.5047	283.1900
37	dibenzo[de, mn]naphthacene	-0.185147	0.144613	0.029093	0.0314	-1.5482	-7.4305	5.8823	306.0786	283.6270
38	naphtho[2,3-g]chrysene	-0.234884	0.145414	0.104992	0.108	-0.9944	-8.1177	7.1233	332.3547	309.9723
39	benzo[h]pentaphene	-0.197664	0.135009	0.001197	0.0548	-0.8089	-8.3009	7.4920	337.8917	311.4461
40	benzo[a]pentacene	-0.207036	0.1369	0.080938	0.0743	-1.4685	-7.5763	6.1078	345.4223	312.7085
41	coronene	-0.140384	0.119999	0.035652	0.0401	-1.0021	-8.1438	7.1417	289.2667	272.4353
42	naphtho[1,2,3,4-def]chrysene	-0.20003	0.142007	0.051303	0.0605	-1.0583	-8.0233	6.9650	302.3366	282.8178
43	dibenzo[def, p]chrysene	-0.207092	0.134685	0.070978	0.0421	-1.1022	-7.9553	6.8531	302.9476	282.8979
44	benzo[rs]pentaphene	-0.217185	0.141164	0.06767	0.0483	-1.1838	-7.8650	6.6812	306.3398	283.4943
45	benzo[g]chrysene	-0.148251	0.116766	0.035819	0.0175	-0.7660	-8.2705	7.5045	287.9205	265.8624
46	2,3:5,6-Dibenzophenanthrene	-0.216687	0.134671	0.054215	0.0393	-0.9636	-8.0436	7.0800	294.6767	266.8634
47	naphtho[2,1,8-gra]naphthacene	-0.215576	0.137964	0.073243	0.0670	-1.3169	-7.7299	6.4130	309.6593	283.9536
48	dibenzo[a, e]aceanthrylene	-0.213057	0.131865	0.492226	0.5277	-1.2806	-8.1423	6.8617	310.6715	284.9750
49	acenaphthylene	-0.1204	0.1227	0.34265	0.3542	-0.9359	-8.9428	8.0069	175.2256	149.9876
50	dibenzo[a, k]fluoranthene	-0.2303	0.1467	0.223597	0.2322	-1.2992	-7.9434	6.6442	313.2985	285.2393
51	naphtho[2,3-k]fluoranthene	-0.2106	0.1334	0.02765	0.0372	-0.9127	-7.9780	7.0653	317.5809	285.6889
52	fluoranthene	-0.1312	0.1184	0.239307	0.2378	-0.9294	-8.6301	7.7007	222.8304	195.1020
53	dibenzo[k, mno]fluoranthene	-0.1595	0.1231	0.23636	0.2482	-0.9755	-8.4000	7.4245	280.3476	257.0611
54	1,2-dihydroacenaphthylene	-0.164	0.1142	0.557096	0.5597	-0.2132	-8.4944	8.2812	183.2279	155.6393
55	9H - fluorene	-0.7398	0.3944	0.7975364	0.8227	-0.2088	-7.5446	7.3358	196.5286	166.8248
56	benzo[b]fluorene	-0.1896	0.1253	0.338153	0.3427	-0.4880	-8.4783	7.9903	244.4588	212.2972
57	benzo[c]fluorene	-0.1687	0.1274	0.442474	0.4577	-0.6415	-8.2835	7.6420	239.5300	211.6880
58	benzo[ghi]fluoranthene	-0.14	0.1126	0.067289	0.0769	-0.9911	-8.6995	7.7084	234.9511	211.8821
59	benzo[a]aceanthrylene	-0.2428	0.1332	0.381872	0.3967	-1.3219	-8.0850	6.7631	266.0202	239.9710
60	indeno[1,2,3-cd]pyrene	-0.1841	0.1232	0.466314	0.477	-1.2835	-8.1363	6.8528	282.0055	256.9774
61	indeno[1,2,3-cd] fluoranthene	-0.1247	0.1181	0.00013	0.0001	-1.3350	-8.5435	7.2085	286.4592	258.3600
62	1H-fluorene	-0.2334	0.1604	0.0980	0.0926	0.26005	-8.5955	8.8555	193.52	164.58
63	cyclopenta[cd]pyrene	-0.1467	0.1259	0.6940	0.6816	-1.3123	-8.2727	6.9604	234.42	211.87
64	benzo[j]fluoranthene	-0.1871	-0.1342	0.3665	0.3479	-1.1767	-8.3165	7.1398	265.90	239.95
65	benzo[k]fluoranthene	-0.1838	0.1380	0.1135	-0.1200	-0.9892	-8.2994	7.3102	270.1861	240.39
66	benzo[a]fluorene	-0.1759	0.1181	0.3720	-0.1038	-0.5607	-8.3656	7.8049	243.65	212.28
67	1H-benzo[b]fluorene	-0.1896	0.1253	0.3431	0.3159	-0.4885	-8.4779	7.9894	244.4619	212.291
68	dibenz[e, k]acephenanthrylene	-0.2197	0.1410	0.3944	0.3982	-1.0702	-8.2215	7.1513	315.12	285.45
69	benzo[b]fluoranthene	-0.1561	0.1219	0.3083	0.295561	-0.9654	-8.6166	7.6512	267.71	240.02
70	indan	-0.1462	0.1090	0.4634	-0.4467	0.5418	-9.1127	9.6545	157.45	127.54

	eletronegatividade	dureza	ovalidade	polarizabi	q_mull	q_mull_pos	HF	MW	NAT	SKC	NBO	NH
1	4.4875	4.2225	1.2690	14.5141	-0.1279	0.1322	40.582	128.18	18	10	11	8
2	4.4814	3.6397	1.3530	21.2966	-0.1268	0.1343	62.927	178.24	24	14	16	10
3	4.5128	4.1043	1.3339	18.7282	-0.1283	0.1331	57.447	178.24	24	14	16	10
4	4.4904	3.2583	1.4295	27.3665	-0.1264	0.1345	86.954	228.3	30	18	21	12
5	4.5098	3.6982	1.4131	24.5932	-0.1279	0.1344	78.254	228.3	30	18	21	12
6	4.5229	3.8468	1.3949	23.7076	-0.1281	0.1336	76.218	228.3	30	18	21	12
7	4.5558	4.1026	1.3755	22.3106	-0.1278	0.1333	214.096	228.3	30	18	21	12
8	4.4958	3.5733	1.3461	23.0164	-0.1283	0.1369	70.028	202.26	26	16	19	10
9	4.5447	3.8991	1.3890	23.3918	-0.1308	0.1369	88.352	228.3	30	18	21	12
10	4.5053	3.3545	1.3924	28.0041	-0.1297	0.1355	90.645	252.32	32	20	24	12
11	4.5157	3.4015	1.4094	27.7382	-0.1269	0.1347	87.543	252.32	32	20	24	12
12	4.5364	3.6784	1.3920	26.0020	-0.1280	0.1347	84.014	252.32	32	20	24	12
13	4.5348	3.8139	1.4532	27.5208	-0.1274	0.1343	94.352	278.36	36	22	26	14
14	4.5211	3.6811	1.4849	28.3664	-0.1261	0.1359	98.710	278.36	36	22	26	14
15	4.5230	3.5281	1.4686	29.2342	-0.1274	0.1349	97.710	252.32	32	20	24	12
16	4.5306	3.7264	1.4698	28.0668	-0.1272	0.1354	94.219	278.36	36	22	26	14
17	4.5326	3.6590	1.4689	28.4418	-0.1331	0.1373	104.246	278.36	36	22	26	14
18	4.5287	3.6968	1.4496	28.1692	-0.1300	0.1377	98.566	278.36	36	22	26	14
19	4.5415	3.8483	1.4550	27.3357	-0.1301	0.1435	99.232	278.36	36	22	26	14
20	4.4957	2.9457	1.4987	33.2215	-0.1277	0.1376	113.869	278.36	36	22	26	14
21	4.5115	3.8383	1.4500	27.3692	-0.1328	0.1484	103.498	278.36	36	22	26	14
22	4.5132	3.3275	1.4848	30.5484	-0.1275	0.1352	101.502	278.36	36	22	26	14
23	4.5207	3.1577	1.4667	32.9926	-0.1281	0.1372	110.429	302.38	38	24	29	14
24	4.5191	3.1124	1.4231	31.0343	-0.1270	0.1363	100.295	276.34	34	22	27	12
25	4.5335	3.3986	1.5375	33.7286	-0.1271	0.1347	116.435	328.42	42	26	31	16
26	4.5348	3.3997	1.5373	33.7210	-0.1272	0.1351	116.429	328.42	42	26	31	16
27	4.5356	3.3792	1.5205	33.7866	-0.1277	0.1349	118.474	328.42	42	26	31	16
28	4.5612	3.7336	1.4346	29.2572	-0.1288	0.1344	101.769	302.38	38	24	29	14
29	4.5555	3.7219	1.4668	29.4276	-0.1276	0.1349	103.768	302.38	38	24	29	14
30	4.4980	3.6148	1.4865	32.2504	-0.1289	0.1433	121.124	352.44	44	28	34	16
31	4.5438	3.7159	1.5065	31.7127	-0.1291	0.1369	114.012	328.42	42	26	31	16
32	4.5449	3.4786	1.4081	28.5940	-0.1277	0.1349	91.328	276.34	34	22	27	12
33	4.5304	3.3529	1.5360	34.0111	-0.1260	0.1353	119.459	328.42	42	26	31	16
34	4.5238	3.7411	1.5109	31.5738	-0.1299	0.1447	258.531	328.42	42	26	31	16
35	4.5236	3.3430	1.4514	31.7015	-0.1293	0.1344	106.949	302.38	38	24	29	14
36	4.5060	3.0224	1.4505	33.9068	-0.1350	0.1424	130.220	302.38	38	24	29	14
37	4.4893	2.9411	1.4662	34.5268	-0.1296	0.1385	120.959	302.38	38	24	29	14
38	4.5560	3.5617	1.5005	32.5293	-0.1354	0.1446	136.243	328.42	42	26	31	16
39	4.5549	3.7460	1.5207	31.5661	-0.1271	0.1345	115.264	328.42	42	26	31	16
40	4.5224	3.0539	1.5504	36.0786	-0.1273	0.1351	125.991	328.42	42	26	31	16
41	4.5729	3.5708	1.4233	29.4287	-0.1164	0.1385	100.349	300.36	36	24	30	12
42	4.5408	3.4825	1.4510	30.8117	-0.1272	0.1345	103.829	302.38	38	24	29	14
43	4.5287	3.4265	1.4537	31.1690	-0.1299	0.1433	110.397	302.38	38	24	29	14
44	4.5244	3.3406	1.4679	31.7678	-0.1267	0.1351	106.453	302.38	38	24	29	14
45	4.5183	3.7522	1.4400	27.8327	-0.1303	0.1436	99.533	278.36	36	22	26	14
46	4.5036	3.5400	1.4701	29.1636	-0.1299	0.1441	102.492	278.36	36	22	26	14
47	4.5234	3.2065	1.4822	32.6929	-0.1266	0.1349	109.641	302.38	38	24	29	14
48	4.7115	3.4309	1.4835	31.3102	-0.1301	0.1379	125.502	302.38	38	24	29	14
49	4.9394	4.0035	1.2835	17.0668	-0.1443	0.1455	80.737	152.2	20	12	14	8
50	4.6213	3.3221	1.4951	32.0296	-0.1331	0.1373	128.601	302.38	38	24	29	14
51	4.4453	3.5327	1.5139	30.7358	-0.1325	0.1374	126.743	302.38	38	24	29	14
52	4.7797	3.8503	1.3698	21.5443	-0.1366	0.1362	87.921	202.26	26	16	19	10
53	4.6877	3.7122	1.4339	27.3474	-0.1374	0.1376	131.599	276.34	34	22	27	12
54	4.3538	4.1406	1.3095	16.8321	-0.1225	0.1336	42.592	154.22	22	12	14	10
55	3.6679	3.8767	1.3410	19.1064	-0.4892	0.3226	152.632	166.23	23	13	15	10
56	4.4832	3.9952	1.4204	22.1648	-0.1332	0.1367	72.010	216.29	29	17	20	12
57	4.4625	3.8210	1.3945	23.0518	-0.1340	0.1356	75.912	216.29	29	17	20	12
58	4.8453	3.8542	1.3670	22.8845	-0.1392	0.1361	116.554	226.28	28	18	22	10
59	4.7035	3.3815	1.4245	27.9729	-0.1370	0.1368	111.373	252.32	32	20	24	12
60	4.7099	3.4264	1.4427	29.0669	-0.1337	0.1370	116.340	276.34	34	22	27	12
61	4.9393	3.6043	1.4603	28.0860	-0.1276	0.1369	146.448	313.38	38	25	31	13
62	4.1677	4.4278	1.332	999999	-0.1317	0.14985	76.098	165.22	22	13	15	9
63	4.7925	3.4802	1.3640	25.0704	-0.1382	0.1459	109.42	226.28	28	18	22	10
64	4.75	3.57	1.4239	26.7991	-0.1404	0.1366	109.07	252.32	32	20	24	12
65	4.60	3.69	1.4451	26.0929	-0.1337	0.1374	105.07	252.32	32	20	24	12
66	4.46	3.90	1.4158	22.6540	-0.1319	0.1364	74.17	216.29	29	17	20	12
67	4.4832	3.9947	1.4205	22.1666	-0.1332	0.1367	72.01	216.29	29	17	20	12
68	4.65	3.58	1.5031	30.4554	-0.1338	0.1373	123.82	289.36	36	23	28	13
69	4.77	3.80	1.4328	25.4749	-0.1349	0.1375	103.24	252.32	32	20	24	12
70	4.28	4.83	1.2850	11.6359	-0.1576	0.1320	11.50	118.19	19	9	10	10

	NC	NCIC	NCIT	NR05	NR06	NR09	NR10	ISIZ	IAC	AAC	ZM1	ZM2
1	10	2	3	0	2	0	1	33.219	17.839	0.991	50	57
2	14	3	5	0	3	0	2	53.303	23.517	0.98	76	90
3	14	3	5	0	3	0	2	53.303	23.517	0.98	76	91
4	18	4	6	0	4	0	2	75.059	29.129	0.971	102	123
5	18	4	7	0	4	0	3	75.059	29.129	0.971	102	124
6	18	4	6	0	4	0	2	75.059	29.129	0.971	102	125
7	18	4	7	0	4	0	3	75.059	29.129	0.971	102	126
8	16	4	8	0	4	0	4	64	24.992	0.961	94	117
9	18	4	6	0	4	0	2	75.059	29.129	0.971	102	125
10	20	5	10	0	5	0	5	86.439	30.542	0.954	120	152
11	20	5	10	0	5	0	5	86.439	30.542	0.954	120	151
12	20	5	11	0	5	0	6	86.439	30.542	0.954	120	152
13	22	5	8	0	5	0	3	98.107	34.707	0.964	128	159
14	22	5	8	0	5	0	3	98.107	34.707	0.964	128	157
15	20	5	10	0	5	0	5	86.439	30.542	0.954	120	151
16	22	5	8	0	5	0	3	98.107	34.707	0.964	128	158
17	22	5	9	0	5	0	4	98.107	34.707	0.964	128	158
18	22	5	8	0	5	0	3	98.107	34.707	0.964	128	159
19	22	5	9	0	5	0	4	98.107	34.707	0.964	128	159
20	22	5	8	0	5	0	3	98.107	34.707	0.964	128	156
21	22	5	8	0	5	0	3	98.107	34.707	0.964	128	159
22	22	5	8	0	5	0	3	98.107	34.707	0.964	128	157
23	24	6	12	0	6	0	6	110.039	36.079	0.949	146	185
24	22	6	14	0	6	0	8	98.107	31.847	0.937	138	177
25	26	6	10	0	6	0	4	122.211	40.266	0.959	154	191
26	26	6	10	0	6	0	4	122.211	40.266	0.959	154	191
27	26	6	10	0	6	0	4	122.211	40.266	0.959	154	192
28	24	6	12	0	6	0	6	110.039	36.079	0.949	146	187
29	24	6	12	0	6	0	6	110.039	36.079	0.949	146	185
30	28	7	14	0	7	0	7	134.606	41.609	0.946	172	221
31	26	6	10	0	6	0	4	122.211	40.266	0.959	154	193
32	22	6	14	0	6	0	8	98.107	31.847	0.937	138	178
33	26	6	10	0	6	0	4	122.211	40.266	0.959	154	191
34	26	6	10	0	6	0	4	122.211	40.266	0.959	154	193
35	24	6	12	0	6	0	6	110.039	36.079	0.949	146	186
36	24	6	12	0	6	0	6	110.039	36.079	0.949	146	186
37	24	6	12	0	6	0	6	110.039	36.079	0.949	146	185
38	26	6	10	0	6	0	4	122.211	40.266	0.959	154	193
39	26	6	10	0	6	0	4	122.211	40.266	0.959	154	192
40	26	6	10	0	6	0	4	122.211	40.266	0.959	154	190
41	24	7	18	0	7	0	11	110.039	33.059	0.918	156	204
42	24	6	12	0	6	0	6	110.039	36.079	0.949	146	186
43	24	6	12	0	6	0	6	110.039	36.079	0.949	146	186
44	24	6	12	0	6	0	6	110.039	36.079	0.949	146	185
45	22	5	8	0	5	0	3	98.107	34.707	0.964	128	160
46	22	5	9	0	5	0	4	98.107	34.707	0.964	128	158
47	24	6	12	0	6	0	6	110.039	36.079	0.949	146	184
48	24	6	12	1	5	3	3	110.039	36.079	0.949	146	186
49	12	3	6	1	2	2	1	43.02	19.419	0.971	68	83
50	24	6	11	1	5	2	3	110.039	36.079	0.949	146	185
51	24	6	11	1	5	2	3	110.039	36.079	0.949	146	184
52	16	4	7	1	3	2	1	64	24.992	0.961	94	118
53	22	6	13	1	5	3	4	98.107	31.847	0.937	138	178
54	12	3	6	1	2	2	1	43.02	21.869	0.994	68	83
55	13	3	5	1	2	2	0	48.106	22.717	0.988	72	87
56	17	4	7	1	3	2	1	69.487	28.375	0.978	98	120
57	17	4	7	1	3	2	1	69.487	28.375	0.978	98	121
58	18	5	12	1	4	4	3	75.059	26.328	0.94	112	144
59	20	5	9	1	4	2	2	86.439	30.542	0.954	120	152
60	22	6	14	1	5	3	5	98.107	31.847	0.937	138	178
61	25	7	15	2	5	5	3	116.096	35.219	0.927	160	208
62	13	3	5	1	2	2	0	48.106	21.472	0.976	72	87
63	18	5	12	1	4	2	5	75.059	26.328	0.94	112	143
64	20	5	10	1	4	3	2	86.439	30.542	0.954	120	152
65	20	5	10	1	4	3	2	86.439	30.542	0.954	120	151
66	17	4	7	1	3	2	1	69.487	28.375	0.978	98	121
67	17	4	7	1	3	2	1	69.487	28.375	0.978	98	120
68	23	6	13	1	5	3	4	104.042	33.97	0.944	142	182
69	20	5	10	1	4	3	2	86.439	30.542	0.954	120	152
70	9	2	3	1	1	1	0	28.529	18.962	0.998	46	53

	CHI0	CHI1	CHI2	CHI0A	CHI1A	CHI2A	ROUV	WIT	WIA	FP	GSI	BAL
1	6.812	4.966	4.089	0.681	0.451	0.292	218	109	2.422	28	15	1.925
2	9.38	6.933	6.081	0.67	0.433	0.276	558	279	3.066	44	24	1.682
3	9.38	6.949	5.994	0.67	0.434	0.272	542	271	2.978	44	24	1.74
4	11.949	8.899	8.072	0.664	0.424	0.269	1138	569	3.719	60	33	1.465
5	11.949	8.916	7.986	0.664	0.425	0.266	1106	553	3.614	60	33	1.512
6	11.949	8.933	7.899	0.664	0.425	0.263	1090	545	3.562	60	33	1.538
7	11.949	8.949	7.836	0.664	0.426	0.261	1026	513	3.353	60	33	1.642
8	10.535	7.933	7.212	0.658	0.418	0.258	724	362	3.017	56	31	1.671
9	11.949	8.933	7.909	0.664	0.425	0.264	1058	529	3.458	60	33	1.587
10	13.104	9.933	9.054	0.655	0.414	0.251	1308	654	3.442	72	40	1.545
11	13.104	9.916	9.128	0.655	0.413	0.254	1360	680	3.579	72	40	1.487
12	13.104	9.933	9.054	0.655	0.414	0.251	1304	652	3.432	72	40	1.553
13	14.518	10.916	9.804	0.66	0.42	0.258	1926	963	4.169	76	42	1.361
14	14.518	10.882	9.977	0.66	0.419	0.263	1958	979	4.238	76	42	1.335
15	13.104	9.916	9.128	0.655	0.413	0.254	1360	680	3.579	72	40	1.487
16	14.518	10.899	9.891	0.66	0.419	0.26	1942	971	4.203	76	42	1.346
17	14.518	10.899	9.891	0.66	0.419	0.26	1910	955	4.134	76	42	1.367
18	14.518	10.916	9.828	0.66	0.42	0.259	1814	907	3.926	76	42	1.449
19	14.518	10.916	9.814	0.66	0.42	0.258	1862	931	4.03	76	42	1.408
20	14.518	10.865	10.064	0.66	0.418	0.265	2022	1011	4.377	76	42	1.288
21	14.518	10.916	9.824	0.66	0.42	0.259	1798	899	3.892	76	42	1.461
22	14.518	10.882	9.977	0.66	0.419	0.263	1974	987	4.273	76	42	1.321
23	15.673	11.899	11.045	0.653	0.41	0.251	2284	1142	4.138	88	49	1.336
24	14.259	10.899	10.358	0.648	0.404	0.247	1678	839	3.632	84	47	1.432
25	17.087	12.865	11.882	0.657	0.415	0.258	3162	1581	4.865	92	51	1.188
26	17.087	12.865	11.882	0.657	0.415	0.258	3130	1565	4.815	92	51	1.199
27	17.087	12.882	11.82	0.657	0.416	0.257	2970	1485	4.569	92	51	1.271
28	15.673	11.933	10.896	0.653	0.411	0.248	2124	1062	3.848	88	49	1.437
29	15.673	11.899	11.045	0.653	0.41	0.251	2220	1110	4.022	88	49	1.372
30	18.242	13.916	12.821	0.651	0.409	0.247	3184	1592	4.212	104	58	1.346
31	17.087	12.899	11.709	0.657	0.416	0.255	3114	1557	4.791	92	51	1.212
32	14.259	10.916	10.271	0.648	0.404	0.245	1630	815	3.528	84	47	1.475
33	17.087	12.865	11.882	0.657	0.415	0.258	3146	1573	4.84	92	51	1.199
34	17.087	12.899	11.729	0.657	0.416	0.255	2922	1461	4.495	92	51	1.292
35	15.673	11.916	10.971	0.653	0.411	0.249	2176	1088	3.942	88	49	1.399
36	15.673	11.916	10.981	0.653	0.411	0.25	2136	1068	3.87	88	49	1.428
37	15.673	11.899	11.045	0.653	0.41	0.251	2244	1122	4.065	88	49	1.351
38	17.087	12.899	11.743	0.657	0.416	0.255	2810	1405	4.323	92	51	1.351
39	17.087	12.882	11.82	0.657	0.416	0.257	2890	1445	4.446	92	51	1.312
40	17.087	12.848	11.969	0.657	0.414	0.26	3210	1605	4.938	92	51	1.169
41	15.413	11.899	11.489	0.642	0.397	0.239	2004	1002	3.63	96	54	1.409
42	15.673	11.916	10.971	0.653	0.411	0.249	2164	1082	3.92	88	49	1.411
43	15.673	11.916	10.981	0.653	0.411	0.25	2132	1066	3.862	88	49	1.433
44	15.673	11.899	11.045	0.653	0.41	0.251	2284	1142	4.138	88	49	1.336
45	14.518	10.933	9.751	0.66	0.42	0.257	1766	883	3.823	76	42	1.492
46	14.518	10.899	9.9	0.66	0.419	0.261	1878	939	4.065	76	42	1.395
47	15.673	11.882	11.12	0.653	0.41	0.253	2332	1166	4.225	88	49	1.303
48	15.673	11.916	10.983	0.653	0.411	0.25	2180	1090	3.949	88	49	1.407
49	7.966	5.949	5.297	0.664	0.425	0.265	332	166	2.515	40	22	1.845
50	15.673	11.899	11.057	0.653	0.41	0.251	2256	1128	4.087	88	49	1.358
51	15.673	11.882	11.122	0.653	0.41	0.253	2400	1200	4.348	88	49	1.265
52	10.535	7.949	7.139	0.658	0.418	0.255	728	364	3.033	56	31	1.677
53	14.259	10.916	10.273	0.648	0.404	0.245	1696	848	3.671	84	47	1.425
54	7.966	5.949	5.297	0.664	0.425	0.265	332	166	2.515	40	22	1.845
55	8.673	6.449	5.652	0.667	0.43	0.269	438	219	2.808	42	23	1.762
56	11.242	8.416	7.644	0.661	0.421	0.264	942	471	3.463	58	32	1.522
57	11.242	8.433	7.567	0.661	0.422	0.261	906	453	3.331	58	32	1.589
58	11.69	8.933	8.357	0.649	0.406	0.246	956	478	3.124	68	38	1.593
59	13.104	9.933	9.056	0.655	0.414	0.252	1352	676	3.558	72	40	1.504
60	14.259	10.916	10.273	0.648	0.404	0.245	1690	845	3.658	84	47	1.428
61	16.121	12.399	11.847	0.645	0.4	0.242	2452	1226	4.087	98	55	1.294
62	8.673	6.449	5.652	0.667	0.43	0.269	438	219	2.808	42	23	1.762
63	11.69	8.916	8.431	0.649	0.405	0.248	966	483	3.157	68	38	1.569
64	13.104	9.933	9.054	0.655	0.414	0.251	1356	678	3.568	72	40	1.5
65	13.104	9.916	9.131	0.655	0.413	0.254	1396	698	3.674	72	40	1.453
66	11.242	8.433	7.558	0.661	0.422	0.261	922	461	3.39	58	32	1.56
67	11.242	8.416	7.644	0.661	0.421	0.264	942	471	3.463	58	32	1.522
68	14.966	11.416	10.617	0.651	0.408	0.247	2002	1001	3.957	86	48	1.362
69	13.104	9.933	9.056	0.655	0.414	0.252	1352	676	3.558	72	40	1.504
70	6.104	4.466	3.736	0.678	0.447	0.287	158	79	2.194	26	14	1.97

	ICEN	IDE	IDM	IDDE	IDDM	IDMT	IDET	ITOP	IC	SIC	CIC	BIC
1	1.522	2.117	5.334	1.522	3.308	581.454	95.247	0.722	1.522	0.458	1.8	0.38
2	1.95	2.57	6.317	1.557	3.787	1762.319	233.831	0.863	1.95	0.512	1.857	0.425
3	1.842	2.492	6.332	2.522	3.788	1715.868	226.775	0.863	1.842	0.484	1.965	0.402
4	2.281	2.918	7.039	2.281	4.147	4005.32	446.53	0.918	1.975	0.474	2.195	0.399
5	2.414	2.855	7.052	3.197	4.147	3899.944	436.773	0.918	2.197	0.527	1.973	0.439
6	2.503	2.83	7.056	2.948	4.146	3845.513	432.993	0.918	1.837	0.44	2.333	0.371
7	1.585	2.611	7.091	1.585	4.148	3637.83	399.45	0.918	1.585	0.38	2.585	0.323
8	1.75	2.458	6.746	2.25	3.985	2442.072	294.962	0.954	1.75	0.437	2.25	0.364
9	2.071	2.717	7.076	3.059	4.149	3743.314	415.633	0.918	2.017	0.484	2.153	0.407
10	1.571	2.645	7.406	1.846	4.305	4843.263	502.644	0.971	2.122	0.491	2.2	0.41
11	2.661	2.802	7.383	3.646	4.301	5020.3	532.37	0.971	2.284	0.529	2.038	0.442
12	2.046	2.633	7.408	3.322	4.304	4829.717	500.219	0.971	2.122	0.491	2.2	0.417
13	2.732	3.115	7.626	3.278	4.433	7344.111	719.463	0.946	1.79	0.401	2.67	0.341
14	2.482	3.136	7.626	2.845	4.433	7466.134	724.521	0.946	2.231	0.5	2.228	0.422
15	2.661	2.802	7.383	3.646	4.301	5020.3	532.37	0.971	2.284	0.529	2.038	0.442
16	2.55	3.127	7.626	2.845	4.434	7405.075	722.268	0.946	2.187	0.49	2.273	0.411
17	2.379	3.054	7.639	2.897	4.436	7295.716	705.589	0.946	2.187	0.49	2.273	0.417
18	2.231	2.924	7.663	3.278	4.435	6950.416	675.551	0.946	2.163	0.485	2.297	0.412
19	2.422	3.014	7.648	3.664	4.435	7119.845	696.122	0.946	2.006	0.45	2.453	0.382
20	2.55	3.201	7.613	2.3	4.434	7696.415	739.374	0.946	1.936	0.434	2.523	0.366
21	2.118	2.888	7.669	3.027	4.437	6894.226	667.095	0.946	2.04	0.458	2.419	0.389
22	2.659	3.151	7.622	3.482	4.434	7523.304	727.864	0.946	2.231	0.5	2.228	0.425
23	2.689	3.076	7.899	3.189	4.559	9020.633	848.869	0.98	2.418	0.527	2.167	0.446
24	2.413	2.793	7.675	3.278	4.441	6438.957	645.135	0.994	2.118	0.475	2.341	0.398
25	2.931	3.377	8.099	3.239	4.674	12804.52	1097.413	0.961	2.258	0.48	2.443	0.406
26	2.632	3.325	8.108	3.671	4.676	12688.78	1080.733	0.961	2.258	0.48	2.443	0.411
27	2.507	3.215	8.129	3.547	4.675	12071.97	1044.914	0.961	2.295	0.488	2.405	0.416
28	2.117	2.85	7.937	2.522	4.564	8429.405	786.607	0.98	1.855	0.405	2.73	0.344
29	2.23	2.961	7.92	3.418	4.563	8790.92	817.238	0.98	2.418	0.527	2.167	0.448
30	2.182	2.957	8.393	3.807	4.786	13361.15	1117.644	0.985	2.156	0.448	2.651	0.386
31	2.931	3.357	8.098	3.547	4.671	12608.86	1090.873	0.961	1.738	0.37	2.962	0.315
32	1.972	2.662	7.692	3.278	4.444	6268.612	614.833	0.994	2.04	0.458	2.419	0.389
33	2.931	3.37	8.1	3.181	4.672	12741.9	1095.143	0.961	2.258	0.48	2.443	0.409
34	2.449	3.178	8.135	3.239	4.677	11885.41	1032.825	0.961	2.046	0.435	2.655	0.37
35	2.201	2.916	7.927	3.689	4.564	8624.796	804.763	0.98	2.288	0.499	2.297	0.424
36	2.262	2.845	7.938	4.168	4.565	8477.812	785.103	0.98	2.372	0.517	2.213	0.434
37	2.23	2.986	7.914	3.022	4.565	8879.738	824.269	0.98	2.418	0.527	2.167	0.448
38	2.332	3.051	8.157	3.642	4.675	11460.6	991.601	0.961	2.324	0.495	2.376	0.421
39	2.449	3.139	8.145	3.393	4.673	11769.17	1020.105	0.961	2.295	0.488	2.405	0.416
40	2.869	3.398	8.096	3.748	4.674	12993.79	1104.293	0.961	2.2	0.468	2.501	0.398
41	1.5	2.689	7.95	1.5	4.571	7966.031	742.07	1	1.5	0.327	3.085	0.278
42	2.272	2.894	7.931	4.335	4.563	8581.049	798.822	0.98	2.288	0.499	2.297	0.424
43	2.345	2.84	7.939	3.918	4.564	8462.849	783.724	0.98	2.372	0.517	2.213	0.443
44	2.689	3.076	7.899	3.418	4.559	9020.633	848.869	0.98	2.418	0.527	2.167	0.446
45	2.329	2.854	7.674	3.391	4.436	6776.355	659.259	0.946	1.959	0.439	2.501	0.373
46	2.392	3.024	7.647	3.447	4.435	7180.421	698.538	0.946	2.295	0.515	2.164	0.437
47	2.717	3.113	7.893	3.97	4.56	9203.495	859.298	0.98	2.376	0.518	2.209	0.438
48	2.218	2.889	7.932	4.502	4.56	8645.617	797.377	0.98	2.396	0.523	2.189	0.444
49	1.626	2.108	5.9	2.418	3.573	979.449	139.11	0.918	1.73	0.482	1.855	0.4
50	2.382	3.001	7.915	4.335	4.56	8928.421	828.219	0.98	2.507	0.547	2.078	0.462
51	2.471	3.149	7.885	3.387	4.561	9462.34	869.007	0.98	2.455	0.535	2.13	0.45
52	2.055	2.448	6.745	3.125	3.981	2455.352	293.777	0.954	2.031	0.508	1.969	0.422
53	2.333	2.805	7.671	4.187	4.438	6505.255	647.863	0.994	2.356	0.528	2.103	0.446
54	1.626	2.108	5.9	2.418	3.573	979.449	139.11	1.459	2.418	0.675	1.167	0.569
55	2.162	2.387	6.112	2.777	3.68	1338.539	186.197	1.239	2.162	0.584	1.538	0.492
56	2.463	2.782	6.883	3.572	4.064	3241.715	378.304	1.221	2.676	0.655	1.412	0.557
57	2.116	2.68	6.903	3.735	4.065	3126.907	364.542	1.221	2.587	0.633	1.501	0.533
58	1.969	2.466	7.101	3.281	4.154	3394.512	377.309	0.991	2.086	0.5	2.084	0.417
59	2.359	2.759	7.39	4.022	4.299	4995.706	524.265	0.971	2.19	0.507	2.131	0.427
60	2.243	2.793	7.674	4.096	4.439	6484.142	645.298	0.994	2.538	0.569	1.921	0.484
61	2.596	3.028	8.028	4.564	4.62	9842.648	908.262	0.999	2.614	0.563	2.03	0.479
62	2.162	2.387	6.112	2.777	3.68	1338.539	186.197	1.42	3.027	0.818	0.673	0.689
63	1.875	2.514	7.097	4.059	4.156	3427.762	384.634	0.991	1.927	0.462	2.242	0.385
64	2.166	2.768	7.388	3.784	4.3	5009.355	525.906	0.971	2.122	0.491	2.2	0.41
65	2.185	2.837	7.375	3.422	4.299	5147.753	539.068	0.971	2.409	0.557	1.913	0.466
66	2.184	2.736	6.892	3.97	4.064	3177.053	372.088	1.221	2.587	0.633	1.501	0.538
67	2.463	2.782	6.883	3.572	4.064	3241.715	378.304	1.379	3.013	0.737	1.074	0.627
68	2.442	2.993	7.779	3.762	4.498	7787.12	757.121	0.988	2.284	0.505	2.24	0.426
69	2.359	2.759	7.39	4.022	4.299	4995.706	524.265	0.971	2.19	0.507	2.131	0.427
70	1.436	1.896	5.028	1.975	3.158	397.205	68.263	1.53	2.281	0.72	0.889	0.616

	1K	2K	PHI	Sv	Se	Sp	Ss	MAXDN	MAXDP	DELS	UI	HYF
1	5.483	2.144	1.176	12.391	17.551	13.032	19.333	0.356	0.12	1.426	2.585	-0.96
2	7.573	2.846	1.54	16.989	23.439	17.79	26.667	0.355	0.241	2.843	3	-0.975
3	7.573	2.846	1.54	16.989	23.439	17.79	26.667	0.355	0.178	2.734	3	-0.975
4	9.668	3.577	1.921	21.586	29.327	22.548	34	0.361	0.275	4.304	3.322	-0.982
5	9.668	3.577	1.921	21.586	29.327	22.548	34	0.361	0.299	4.191	3.322	-0.982
6	9.668	3.577	1.921	21.586	29.327	22.548	34	0.361	0.236	4.082	3.322	-0.982
7	9.668	3.577	1.921	21.586	29.327	22.548	34	0.327	0.212	3.924	3.322	-0.982
8	8.118	2.733	1.387	18.989	25.439	19.79	30	0.332	0.212	3.716	3.17	-0.979
9	9.668	3.577	1.921	21.586	29.327	22.548	34	0.36	0.212	4.056	3.322	-0.982
10	10.221	3.488	1.782	23.586	31.327	24.548	37.333	0.341	0.247	4.898	3.459	-0.984
11	10.221	3.488	1.782	23.586	31.327	24.548	37.333	0.349	0.309	5.025	3.459	-0.984
12	10.221	3.488	1.782	23.586	31.327	24.548	37.333	0.336	0.247	4.878	3.459	-0.984
13	11.765	4.317	2.309	26.184	35.215	27.306	41.333	0.366	0.27	5.47	3.585	-0.986
14	11.765	4.317	2.309	26.184	35.215	27.306	41.333	0.366	0.321	5.688	3.585	-0.986
15	10.221	3.488	1.782	23.586	31.327	24.548	37.333	0.349	0.309	5.025	3.459	-0.984
16	11.765	4.317	2.309	26.184	35.215	27.306	41.333	0.366	0.333	5.579	3.585	-0.986
17	11.765	4.317	2.309	26.184	35.215	27.306	41.333	0.366	0.356	5.582	3.585	-0.986
18	11.765	4.317	2.309	26.184	35.215	27.306	41.333	0.365	0.333	5.421	3.585	-0.986
19	11.765	4.317	2.309	26.184	35.215	27.306	41.333	0.366	0.27	5.447	3.585	-0.986
20	11.765	4.317	2.309	26.184	35.215	27.306	41.333	0.366	0.309	5.803	3.585	-0.986
21	11.765	4.317	2.309	26.184	35.215	27.306	41.333	0.365	0.235	5.417	3.585	-0.986
22	11.765	4.317	2.309	26.184	35.215	27.306	41.333	0.366	0.333	5.692	3.585	-0.986
23	12.323	4.242	2.178	28.184	37.215	29.306	44.667	0.353	0.325	6.375	3.7	-0.987
24	10.81	3.48	1.71	25.586	33.327	26.548	40.667	0.331	0.343	5.896	3.585	-0.986
25	13.863	5.062	2.699	30.782	41.103	32.064	48.667	0.369	0.349	7.112	3.807	-0.989
26	13.863	5.062	2.699	30.782	41.103	32.064	48.667	0.37	0.355	7.116	3.807	-0.989
27	13.863	5.062	2.699	30.782	41.103	32.064	48.667	0.371	0.367	6.962	3.807	-0.989
28	12.323	4.242	2.178	28.184	37.215	29.306	44.667	0.321	0.269	6.084	3.7	-0.987
29	12.323	4.242	2.178	28.184	37.215	29.306	44.667	0.361	0.355	6.38	3.7	-0.987
30	14.425	4.996	2.574	32.782	43.103	34.064	52	0.347	0.303	7.484	3.907	-0.99
31	13.863	5.062	2.699	30.782	41.103	32.064	48.667	0.369	0.293	6.89	3.807	-0.989
32	10.81	3.48	1.71	25.586	33.327	26.548	40.667	0.33	0.269	5.759	3.585	-0.986
33	13.863	5.062	2.699	30.782	41.103	32.064	48.667	0.369	0.333	7.108	3.807	-0.989
34	13.863	5.062	2.699	30.782	41.103	32.064	48.667	0.37	0.293	6.852	3.807	-0.989
35	12.323	4.242	2.178	28.184	37.215	29.306	44.667	0.353	0.367	6.252	3.7	-0.987
36	12.323	4.242	2.178	28.184	37.215	29.306	44.667	0.353	0.325	6.221	3.7	-0.987
37	12.323	4.242	2.178	28.184	37.215	29.306	44.667	0.349	0.367	6.429	3.7	-0.987
38	13.863	5.062	2.699	30.782	41.103	32.064	48.667	0.371	0.367	6.826	3.807	-0.989
39	13.863	5.062	2.699	30.782	41.103	32.064	48.667	0.371	0.355	6.958	3.807	-0.989
40	13.863	5.062	2.699	30.782	41.103	32.064	48.667	0.37	0.349	7.223	3.807	-0.989
41	11.424	3.519	1.675	27.586	35.327	28.548	44	0.318	0.274	6.568	3.7	-0.987
42	12.323	4.242	2.178	28.184	37.215	29.306	44.667	0.353	0.367	6.227	3.7	-0.987
43	12.323	4.242	2.178	28.184	37.215	29.306	44.667	0.353	0.332	6.201	3.7	-0.987
44	12.323	4.242	2.178	28.184	37.215	29.306	44.667	0.353	0.325	6.375	3.7	-0.987
45	11.765	4.317	2.309	26.184	35.215	27.306	41.333	0.365	0.27	5.286	3.585	-0.986
46	11.765	4.317	2.309	26.184	35.215	27.306	41.333	0.366	0.333	5.556	3.585	-0.986
47	12.323	4.242	2.178	28.184	37.215	29.306	44.667	0.363	0.344	6.53	3.7	-0.987
48	12.323	4.242	2.178	28.184	37.215	29.306	44.667	0.349	0.367	6.102	3.7	-0.987
49	6.013	1.978	0.991	14.391	19.551	15.032	22.667	0.324	0.178	2.442	2.807	-0.969
50	12.323	4.242	2.178	28.184	37.215	29.306	44.667	0.356	0.367	6.252	3.7	-0.987
51	12.323	4.242	2.178	28.184	37.215	29.306	44.667	0.362	0.367	6.458	3.7	-0.987
52	8.118	2.733	1.387	18.989	25.439	19.79	30	0.328	0.22	3.512	3.17	-0.979
53	10.81	3.48	1.71	25.586	33.327	26.548	40.667	0.339	0.367	5.663	3.585	-0.986
54	6.243	2.1	1.093	14.989	21.439	15.79	21.667	0.267	0.249	2.441	2.585	-0.969
55	6.904	2.463	1.308	15.989	22.439	16.79	24.167	0.396	0.218	2.62	2.807	-0.972
56	9.005	3.207	1.699	20.586	28.327	21.548	31.5	0.42	0.338	3.985	3.17	-0.98
57	9.005	3.207	1.699	20.586	28.327	21.548	31.5	0.42	0.275	3.841	3.17	-0.98
58	8.709	2.73	1.321	20.989	27.439	21.79	33.333	0.316	0.247	4.371	3.322	-0.982
59	10.221	3.488	1.782	23.586	31.327	24.548	37.333	0.344	0.34	4.804	3.459	-0.984
60	10.81	3.48	1.71	25.586	33.327	26.548	40.667	0.329	0.374	5.69	3.585	-0.986
61	12.164	3.868	1.882	28.885	37.271	29.927	46	0.33	0.409	6.716	3.755	-0.988
62	6.823	2.417	1.268	15.69	21.495	16.411	24.667	0.472	0.934	3.728	2.907	-0.972
63	8.709	2.73	1.321	20.989	27.439	21.79	33.333	0.324	0.317	4.584	3.322	-0.982
64	10.221	3.488	1.782	23.586	31.327	24.548	37.333	0.351	0.27	4.805	3.459	-0.984
65	10.221	3.488	1.782	23.586	31.327	24.548	37.333	0.351	0.333	4.951	3.459	-0.984
66	9.005	3.207	1.699	20.586	28.327	21.548	31.5	0.424	0.267	3.799	3.17	-0.98
67	9.005	3.207	1.699	20.586	28.327	21.548	31.5	0.434	0.33	4.294	3.17	-0.98
68	11.563	3.853	1.937	26.885	35.271	27.927	42.667	0.349	0.328	5.941	3.644	-0.987
69	10.221	3.488	1.782	23.586	31.327	24.548	37.333	0.344	0.34	4.804	3.459	-0.984
70	5.041	1.871	1.048	11.989	18.439	12.79	15.833	0.198	0.238	1.47	2	-0.954

	L1u	L2u	L3u	P1u	P2u	G1u	G2u	G3u	E1u	E2u	E3u	L1m
1	4.516	2.325	0	0.66	0.34	1	1	1	0.54	0.536	0	3.169
2	7.972	2.462	0	0.764	0.236	1	1	1	0.547	0.546	0	6.188
3	6.743	2.98	0	0.693	0.307	1	0.179	1	0.53	0.522	0	5.082
4	12.415	2.545	0	0.83	0.17	1	1	1	0.55	0.553	0	10.196
5	10.548	3.174	0	0.769	0.231	0.169	0.169	1	0.542	0.518	0	8.517
6	10.376	3.049	0	0.773	0.227	1	1	1	0.529	0.536	0	8.321
7	5.954	5.954	0	0.5	0.5	0.224	0.245	1	0.497	0.497	0	4.248
8	6.175	3.824	0	0.618	0.382	1	1	1	0.499	0.498	0	4.43
9	7.734	4.016	0.146	0.65	0.338	0.188	0.169	0.206	0.538	0.529	0.36	6.036
10	7.729	4.822	0	0.616	0.384	0.2	1	1	0.566	0.568	0	5.897
11	9.866	3.747	0	0.725	0.275	0.167	0.167	1	0.513	0.509	0	7.693
12	6.948	5.57	0	0.555	0.445	0.167	1	1	0.486	0.476	0	5.084
13	14.429	3.225	0	0.817	0.183	0.162	0.162	0.889	0.539	0.554	0.516	12.092
14	13.846	3.6	0	0.794	0.206	0.162	0.162	0.856	0.54	0.484	0.435	11.645
15	9.866	3.747	0	0.725	0.275	0.167	0.167	1	0.513	0.509	0	7.693
16	14.551	3.396	0	0.811	0.189	1	1	1	0.545	0.527	0	12.22
17	12.67	4.238	0	0.749	0.251	1	0.162	1	0.566	0.533	0	10.592
18	10.044	5.141	0	0.661	0.339	0.162	0.162	0.924	0.515	0.484	0.357	8.019
19	11.491	3.722	0.177	0.747	0.242	0.162	0.162	0.177	0.53	0.518	0.364	9.478
20	17.846	2.6	0	0.873	0.127	1	1	0.713	0.552	0.558	0.41	15.192
21	7.418	5.166	0.548	0.565	0.393	0.177	0.162	0.162	0.535	0.519	0.374	5.855
22	15.247	3.096	0	0.831	0.169	0.162	0.162	0.653	0.557	0.523	0.472	12.938
23	13.841	3.577	0	0.795	0.205	0.16	0.16	1	0.512	0.511	0	11.303
24	9.481	4.401	0	0.683	0.317	0.195	0.195	0.644	0.517	0.516	0.398	7.289
25	19.639	3.358	0	0.854	0.146	0.156	0.192	0.397	0.567	0.525	0.517	17.131
26	18.167	3.794	0	0.827	0.173	1	0.156	0.82	0.589	0.522	0.377	15.87
27	15.028	4.776	0	0.759	0.241	0.156	0.156	0.504	0.548	0.471	0.341	12.777
28	9.875	4.605	0.059	0.679	0.317	0.16	0.16	0.16	0.476	0.473	0.351	7.718
29	11.18	4.918	0.001	0.694	0.306	0.16	0.16	0.308	0.504	0.469	0.342	9.026
30	9.612	6.291	0.187	0.597	0.391	0.155	0.155	0.155	0.557	0.564	0.372	7.715
31	19.517	3.252	0	0.857	0.143	0.169	0.169	0.914	0.539	0.559	0	16.814
32	7.367	5.819	0	0.559	0.441	0.18	0.164	1	0.542	0.538	0	5.454
33	19.993	3.068	0	0.867	0.133	0.156	0.169	1	0.528	0.528	0	17.249
34	14.969	4.12	0.151	0.778	0.214	0.169	0.18	0.156	0.538	0.53	0.368	12.678
35	11.107	4.805	0	0.698	0.302	0.16	0.16	0.766	0.518	0.495	0.507	8.885
36	8.801	5.764	0.182	0.597	0.391	0.16	0.174	0.187	0.499	0.474	0.344	6.867
37	12.204	4.395	0	0.735	0.265	0.16	0.16	0.832	0.579	0.554	0.507	10.09
38	8.896	7.405	0.4	0.533	0.443	0.156	0.156	0.192	0.465	0.449	0.312	7.108
39	12.1	6.592	0.02	0.647	0.352	0.169	0.156	0.217	0.484	0.44	0.25	9.889
40	20.807	3.069	0	0.871	0.129	0.156	0.156	0.562	0.57	0.525	0.467	18.27
41	6.879	6.859	0	0.501	0.499	0.191	0.255	1	0.517	0.519	0	4.949
42	9.782	5.943	0	0.622	0.378	0.16	0.16	1	0.514	0.506	0.404	7.599
43	8.604	6.064	0.168	0.58	0.409	0.16	0.174	0.16	0.47	0.46	0.333	6.609
44	13.27	3.905	0	0.773	0.227	0.16	0.187	1	0.515	0.497	0	10.818
45	8.988	5.211	0.214	0.624	0.362	0.162	0.177	0.162	0.492	0.471	0.341	7.054
46	10.297	4.477	0.173	0.689	0.3	0.162	0.162	0.177	0.539	0.491	0.398	8.498
47	14.275	3.62	0	0.798	0.202	0.16	0.174	1	0.534	0.51	0	11.876
48	9.233	6.253	0	0.596	0.404	0.16	0.16	1	0.497	0.499	0	7.089
49	4.04	3.167	0	0.561	0.439	0.188	0.188	1	0.512	0.504	0	2.69
50	11.3	5.363	0	0.678	0.322	0.16	0.16	1	0.517	0.476	0	9.105
51	16.398	3.313	0	0.832	0.168	0.174	0.174	0.545	0.567	0.521	0.499	13.883
52	6.643	3.827	0	0.634	0.366	0.175	0.797	1	0.53	0.504	0	4.86
53	10.052	4.117	0	0.709	0.291	0.164	0.164	0.959	0.524	0.505	0.374	7.819
54	3.822	3.696	0.151	0.498	0.482	1	0.183	1	0.486	0.57	0.182	2.686
55	6	2.392	0.069	0.709	0.283	0.181	0.181	1	0.515	0.563	0.087	4.646
56	9.41	2.496	0.055	0.787	0.209	0.171	0.19	0.807	0.538	0.564	0.069	7.784
57	7.577	3.438	0.074	0.683	0.31	0.171	0.171	0.209	0.52	0.546	0.111	5.999
58	6.252	4.461	0	0.584	0.416	0.172	0.172	1	0.532	0.534	0	4.494
59	9.199	4.145	0	0.689	0.311	0.167	0.167	1	0.506	0.466	0	7.154
60	9.82	4.299	0	0.695	0.305	0.164	0.164	1	0.519	0.505	0	7.608
61	13.27	3.98	0	0.769	0.231	0.16	0.16	0.862	0.534	0.512	0.308	10.763
62	6.842	2.121	0.072	0.757	0.235	0.198	0.183	0.635	0.579	0.565	0.091	4.945
63	6.723	4.386	0	0.605	0.395	0.172	0.172	0.892	0.528	0.529	0	4.931
64	9.749	4.134	0	0.702	0.298	0.167	0.167	0.746	0.542	0.52	0.486	7.676
65	11.428	3.692	0	0.756	0.244	0.167	1	1	0.559	0.509	0	9.167
66	9.606	2.887	0.059	0.765	0.23	0.171	0.171	1	0.522	0.535	0.069	7.806
67	10.695	2.507	0.055	0.807	0.189	0.171	0.19	0.62	0.562	0.54	0.069	8.511
68	12.59	3.49	0	0.783	0.217	0.162	0.162	0.477	0.51	0.496	0.427	10.117
69	9.945	4.269	0	0.7	0.3	0.167	0.167	0.687	0.509	0.47	0.331	7.729
70	3.919	2.02	0.291	0.629	0.324	0.191	1	0.191	0.551	0.52	0.203	2.808

	L2m	L3m	P1m	P2m	G1m	G2m	G3m	E1m	E2m	E3m	L1v	L2v
1	1.247	0	0.718	0.282	1	1	1	0.266	0.154	0	3.628	1.614
2	1.293	0	0.827	0.173	1	1	1	0.33	0.151	0	6.779	1.68
3	1.719	0	0.747	0.253	1	0.179	1	0.301	0.174	0	5.632	2.137
4	1.318	0	0.885	0.115	1	1	1	0.371	0.148	0	10.919	1.718
5	1.845	0	0.822	0.178	0.169	0.188	1	0.353	0.175	0	9.178	2.278
6	1.713	0	0.829	0.171	1	1	1	0.341	0.17	0	8.991	2.149
7	4.247	0	0.5	0.5	0.224	0.224	1	0.253	0.253	0	4.804	4.803
8	2.405	0	0.648	0.352	1	1	1	0.257	0.197	0	4.99	2.861
9	2.639	0.077	0.69	0.302	0.169	0.169	0.169	0.328	0.229	0.101	6.589	3.088
10	3.172	0	0.65	0.35	0.2	1	1	0.329	0.246	0	6.48	3.697
11	2.284	0	0.771	0.229	0.167	0.167	1	0.312	0.189	0	8.385	2.75
12	3.825	0	0.571	0.429	0.167	1	1	0.26	0.224	0	5.677	4.38
13	1.853	0	0.867	0.133	0.162	0.162	0.889	0.379	0.183	0.198	12.846	2.295
14	2.238	0	0.839	0.161	0.162	0.162	1	0.382	0.187	0.118	12.354	2.677
15	2.284	0	0.771	0.229	0.167	0.167	1	0.312	0.189	0	8.385	2.75
16	1.995	0	0.86	0.14	1	1	1	0.385	0.183	0	12.971	2.447
17	2.74	0	0.794	0.206	1	0.162	1	0.396	0.223	0	11.262	3.223
18	3.554	0	0.693	0.307	0.162	0.162	0.924	0.329	0.231	0.08	8.671	4.066
19	2.368	0.097	0.794	0.198	0.162	0.162	0.162	0.361	0.209	0.111	10.127	2.804
20	1.335	0	0.919	0.081	1	1	0.609	0.4	0.147	0.139	16.048	1.743
21	3.693	0.333	0.593	0.374	0.177	0.162	0.162	0.335	0.265	0.144	6.358	4.168
22	1.771	0	0.88	0.12	0.162	0.162	0.653	0.401	0.171	0.234	13.682	2.198
23	2.119	0	0.842	0.158	0.16	0.16	1	0.341	0.179	0	12.106	2.58
24	2.828	0	0.72	0.28	0.195	0.195	0.6	0.306	0.213	0	7.972	3.318
25	1.963	0	0.897	0.103	0.156	0.169	0.291	0.432	0.18	0.293	17.933	2.41
26	2.367	0	0.87	0.13	1	0.156	0.846	0.449	0.203	0.115	16.605	2.823
27	3.207	0	0.799	0.201	0.156	0.169	0.435	0.396	0.212	0.09	13.497	3.709
28	3.078	0.033	0.713	0.284	0.167	0.174	0.174	0.29	0.211	0.109	8.4	3.561
29	3.299	0	0.732	0.268	0.16	0.174	0.348	0.328	0.211	0.088	9.707	3.811
30	4.49	0.097	0.627	0.365	0.155	0.155	0.155	0.359	0.287	0.1	8.312	5.057
31	1.848	0	0.901	0.099	0.169	0.18	0.841	0.401	0.181	0	17.678	2.298
32	4.024	0	0.575	0.425	0.164	0.164	0.822	0.297	0.257	0	6.05	4.583
33	1.723	0	0.909	0.091	0.156	0.156	1	0.393	0.167	0	18.127	2.153
34	2.662	0.076	0.822	0.173	0.156	0.18	0.156	0.388	0.224	0.096	13.41	3.129
35	3.19	0	0.736	0.264	0.16	0.16	0.766	0.332	0.219	0.203	9.587	3.701
36	4.05	0.106	0.623	0.367	0.16	0.16	0.16	0.306	0.236	0.119	7.478	4.592
37	2.81	0	0.782	0.218	0.16	0.16	0.894	0.396	0.227	0.197	10.759	3.311
38	5.578	0.231	0.55	0.432	0.156	0.156	0.169	0.298	0.255	0.104	7.679	6.163
39	4.812	0.01	0.672	0.327	0.156	0.156	0.186	0.323	0.234	0.057	10.596	5.381
40	1.739	0	0.913	0.087	0.156	0.156	0.351	0.44	0.168	0.232	19.082	2.164
41	4.949	0	0.5	0.5	0.191	0.162	1	0.27	0.268	0	5.539	5.534
42	4.219	0	0.643	0.357	0.174	0.16	1	0.31	0.255	0.169	8.289	4.764
43	4.371	0.095	0.597	0.395	0.16	0.16	0.174	0.277	0.239	0.112	7.24	4.906
44	2.405	0	0.818	0.182	0.16	0.16	0.963	0.342	0.189	0	11.593	2.879
45	3.639	0.122	0.652	0.336	0.162	0.162	0.162	0.304	0.23	0.106	7.677	4.145
46	3.046	0.095	0.73	0.262	0.162	0.162	0.177	0.367	0.227	0.124	9.078	3.507
47	2.171	0	0.845	0.155	0.16	0.16	1	0.37	0.183	0	12.634	2.629
48	4.566	0	0.608	0.392	0.16	0.16	1	0.293	0.266	0	7.766	5.1
49	1.998	0	0.574	0.426	0.188	0.188	1	0.227	0.201	0	3.13	2.379
50	3.742	0	0.709	0.291	0.16	0.16	1	0.336	0.231	0	9.799	4.255
51	1.95	0	0.877	0.123	0.16	0.16	0.514	0.406	0.18	0.263	14.678	2.381
52	2.398	0	0.67	0.33	0.175	0.797	1	0.284	0.198	0	5.432	2.857
53	2.652	0	0.747	0.253	0.164	0.164	0.883	0.317	0.21	0	8.514	3.109
54	2.302	0.022	0.536	0.46	1	0.183	1	0.24	0.221	0.004	3.077	2.782
55	1.296	0.01	0.781	0.218	0.181	0.207	1	0.309	0.165	0.002	5.103	1.666
56	1.346	0.007	0.852	0.147	0.171	0.171	0.807	0.368	0.164	0.001	8.321	1.726
57	2.129	0.017	0.737	0.261	0.171	0.171	0.19	0.326	0.21	0.005	6.52	2.562
58	2.943	0	0.604	0.396	0.172	0.172	1	0.275	0.232	0	5.044	3.418
59	2.656	0	0.729	0.271	0.167	0.167	1	0.306	0.191	0	7.805	3.13
60	2.797	0	0.731	0.269	0.164	0.18	1	0.312	0.214	0	8.297	3.265
61	2.541	0	0.809	0.191	0.16	0.16	0.661	0.352	0.209	0.072	11.536	2.985
62	1.294	0.01	0.791	0.207	0.183	0.183	0.781	0.302	0.211	0.002	5.568	1.567
63	2.862	0	0.633	0.367	0.172	0.172	1	0.284	0.225	0	5.492	3.339
64	2.605	0	0.747	0.253	0.167	0.167	0.635	0.337	0.207	0.172	8.335	3.092
65	2.238	0	0.804	0.196	0.167	1	1	0.359	0.187	0	9.886	2.701
66	1.626	0.008	0.827	0.172	0.171	0.171	1	0.345	0.17	0.001	8.4	2.043
67	1.4	0.007	0.858	0.141	0.171	0.171	0.575	0.356	0.169	0.001	9.232	1.766
68	2.101	0	0.828	0.172	0.162	0.162	0.37	0.329	0.18	0.187	10.893	2.537
69	2.744	0	0.738	0.262	0.167	0.167	0.899	0.308	0.194	0.095	8.434	3.229
70	1.167	0.06	0.696	0.289	0.191	1	0.191	0.283	0.174	0.009	3.221	1.484

	L3v	P1v	P2v	G1v	G2v	G3v	E1v	E2v	E3v	L1e	L2e	L3e
1	0	0.692	0.308	1	1	1	0.349	0.258	0	4.466	2.285	0
2	0	0.801	0.199	1	1	1	0.396	0.254	0	7.903	2.417	0
3	0	0.725	0.275	1	0.179	1	0.369	0.268	0	6.678	2.931	0
4	0	0.864	0.136	1	1	1	0.426	0.252	0	12.327	2.496	0
5	0	0.801	0.199	0.169	0.169	1	0.41	0.267	0	10.467	3.122	0
6	0	0.807	0.193	1	1	1	0.397	0.267	0	10.295	2.996	0
7	0	0.5	0.5	0.224	0.224	1	0.323	0.323	0	5.887	5.886	0
8	0	0.636	0.364	1	1	1	0.326	0.279	0	6.105	3.767	0
9	0.099	0.674	0.316	0.169	0.169	0.169	0.391	0.313	0.169	7.667	3.962	0.143
10	0	0.637	0.363	0.2	1	1	0.398	0.334	0	7.654	4.755	0
11	0	0.753	0.247	0.167	0.167	1	0.37	0.274	0	9.777	3.687	0
12	0	0.565	0.435	0.167	1	1	0.325	0.294	0	6.872	5.498	0
13	0	0.848	0.152	0.162	0.162	0.889	0.427	0.28	0.286	14.335	3.17	0
14	0	0.822	0.178	0.162	0.162	1	0.43	0.267	0.2	13.757	3.546	0
15	0	0.753	0.247	0.167	0.167	1	0.37	0.274	0	9.777	3.687	0
16	0	0.841	0.159	1	1	1	0.433	0.275	0	14.457	3.34	0
17	0	0.777	0.223	1	0.162	1	0.447	0.308	0	12.586	4.178	0
18	0	0.681	0.319	0.162	0.162	0.824	0.384	0.303	0.147	9.962	5.077	0
19	0.123	0.776	0.215	0.162	0.162	0.177	0.412	0.294	0.178	11.41	3.667	0.174
20	0	0.902	0.098	1	1	0.653	0.446	0.251	0.211	17.739	2.549	0
21	0.403	0.582	0.381	0.177	0.162	0.177	0.394	0.338	0.208	7.355	5.106	0.54
22	0	0.862	0.138	0.162	0.162	0.609	0.448	0.263	0.308	15.154	3.043	0
23	0	0.824	0.176	0.16	0.16	1	0.392	0.266	0	13.736	3.516	0
24	0	0.706	0.294	0.195	0.195	0.644	0.366	0.294	0	9.388	4.335	0
25	0	0.882	0.118	0.156	0.169	0.291	0.474	0.271	0.358	19.537	3.301	0
26	0	0.855	0.145	1	0.156	0.846	0.492	0.289	0.183	18.073	3.735	0
27	0	0.784	0.216	0.156	0.156	0.415	0.442	0.284	0.154	14.936	4.712	0
28	0.041	0.7	0.297	0.16	0.16	0.187	0.344	0.283	0.171	9.786	4.542	0.058
29	0	0.718	0.282	0.16	0.174	0.363	0.38	0.282	0.151	11.091	4.851	0.001
30	0.125	0.616	0.375	0.155	0.155	0.155	0.417	0.364	0.167	9.533	6.216	0.184
31	0	0.885	0.115	0.169	0.18	0.841	0.443	0.28	0	19.407	3.195	0
32	0	0.569	0.431	0.164	0.164	0.822	0.366	0.334	0	7.286	5.743	0
33	0	0.894	0.106	0.156	0.156	1	0.434	0.261	0	19.881	3.013	0
34	0.1	0.806	0.188	0.156	0.192	0.156	0.433	0.309	0.164	14.875	4.061	0.148
35	0	0.721	0.279	0.16	0.16	0.766	0.387	0.294	0.285	11.015	4.738	0
36	0.13	0.613	0.376	0.16	0.16	0.16	0.362	0.302	0.178	8.72	5.693	0.179
37	0	0.765	0.235	0.16	0.16	0.894	0.45	0.315	0.28	12.117	4.329	0
38	0.286	0.544	0.436	0.156	0.156	0.156	0.347	0.311	0.159	8.823	7.331	0.394
39	0.013	0.663	0.337	0.156	0.156	0.217	0.371	0.293	0.103	12.01	6.519	0.019
40	0	0.898	0.102	0.156	0.169	0.351	0.479	0.261	0.306	20.703	3.015	0
41	0	0.5	0.5	0.162	0.237	1	0.336	0.338	0	6.795	6.776	0
42	0	0.635	0.365	0.16	0.16	1	0.369	0.325	0.233	9.692	5.872	0
43	0.118	0.59	0.4	0.16	0.16	0.167	0.333	0.301	0.17	8.521	5.994	0.165
44	0	0.801	0.199	0.16	0.174	0.963	0.393	0.27	0	13.168	3.843	0
45	0.152	0.641	0.346	0.162	0.162	0.177	0.36	0.299	0.168	8.91	5.148	0.211
46	0.12	0.714	0.276	0.162	0.177	0.162	0.419	0.301	0.197	10.225	4.419	0.17
47	0	0.828	0.172	0.16	0.16	1	0.418	0.269	0	14.176	3.56	0
48	0	0.604	0.396	0.16	0.16	1	0.352	0.332	0	9.144	6.183	0
49	0	0.568	0.432	0.188	0.188	1	0.308	0.284	0	3.986	3.12	0
50	0	0.697	0.303	0.16	0.16	1	0.389	0.299	0	11.209	5.296	0
51	0	0.86	0.14	0.16	0.16	0.545	0.454	0.269	0.331	16.294	3.257	0
52	0	0.655	0.345	0.175	0.797	1	0.354	0.281	0	6.57	3.769	0
53	0	0.733	0.267	0.164	0.164	0.883	0.376	0.288	0.213	9.957	4.055	0
54	0.066	0.519	0.47	1	0.183	1	0.315	0.323	0.035	3.78	3.645	0.146
55	0.03	0.751	0.245	0.181	0.181	1	0.372	0.273	0.016	5.949	2.351	0.067
56	0.023	0.826	0.171	0.171	0.171	0.807	0.421	0.27	0.012	9.347	2.451	0.053
57	0.036	0.715	0.281	0.19	0.171	0.209	0.385	0.303	0.025	7.515	3.387	0.072
58	0	0.596	0.404	0.172	0.172	1	0.346	0.314	0	6.179	4.397	0
59	0	0.714	0.286	0.167	0.167	1	0.364	0.266	0	9.115	4.084	0
60	0	0.718	0.282	0.164	0.164	1	0.371	0.292	0	9.726	4.236	0
61	0	0.794	0.206	0.16	0.16	0.661	0.404	0.288	0.127	13.162	3.918	0
62	0.03	0.777	0.219	0.183	0.183	0.781	0.383	0.309	0.016	6.768	2.089	0.07
63	0	0.622	0.378	0.172	0.172	1	0.352	0.307	0	6.648	4.322	0
64	0	0.729	0.271	0.167	0.167	0.635	0.397	0.292	0.254	9.664	4.071	0
65	0	0.785	0.215	0.167	1	1	0.418	0.273	0	11.336	3.632	0
66	0.025	0.802	0.195	0.171	0.171	1	0.4	0.268	0.012	9.536	2.838	0.057
67	0.023	0.838	0.16	0.171	0.171	0.62	0.419	0.268	0.012	10.61	2.464	0.053
68	0	0.811	0.189	0.162	0.162	0.37	0.382	0.262	0.252	12.486	3.432	0
69	0	0.723	0.277	0.167	0.167	0.899	0.366	0.269	0.155	9.854	4.206	0
70	0.146	0.664	0.306	0.191	1	0.191	0.372	0.281	0.051	3.882	1.992	0.283

	P1e	P2e	G1e	G2e	G3e	E1e	E2e	E3e	L1p	L2p	L3p	P1p
1	0.662	0.338	1	1	1	0.528	0.518	0	3.768	1.726	0	0.686
2	0.766	0.234	1	1	1	0.538	0.526	0	6.963	1.801	0	0.795
3	0.695	0.305	1	0.179	1	0.519	0.505	0	5.803	2.267	0	0.719
4	0.832	0.168	1	1	1	0.543	0.532	0	11.146	1.844	0	0.858
5	0.77	0.23	0.169	0.169	1	0.533	0.501	0	9.387	2.414	0	0.795
6	0.775	0.225	1	1	1	0.521	0.517	0	9.201	2.286	0	0.801
7	0.5	0.5	0.224	0.224	1	0.485	0.485	0	4.979	4.978	0	0.5
8	0.618	0.382	1	1	1	0.488	0.483	0	5.168	3.005	0	0.632
9	0.651	0.337	0.188	0.169	0.169	0.529	0.515	0.346	6.763	3.229	0.106	0.67
10	0.617	0.383	0.2	1	1	0.555	0.552	0	6.666	3.865	0	0.633
11	0.726	0.274	0.167	0.167	1	0.504	0.493	0	8.605	2.898	0	0.748
12	0.556	0.444	0.167	1	1	0.476	0.464	0	5.867	4.557	0	0.563
13	0.819	0.181	0.162	0.162	0.889	0.532	0.535	0.501	13.084	2.435	0	0.843
14	0.795	0.205	0.162	0.162	0.827	0.533	0.469	0.419	12.579	2.816	0	0.817
15	0.726	0.274	0.167	0.167	1	0.504	0.493	0	8.605	2.898	0	0.748
16	0.812	0.188	1	1	1	0.538	0.51	0	13.209	2.59	0	0.836
17	0.751	0.249	1	0.162	1	0.559	0.518	0	11.474	3.376	0	0.773
18	0.662	0.338	0.162	0.162	0.924	0.507	0.472	0.342	8.878	4.228	0	0.677
19	0.748	0.24	0.162	0.162	0.177	0.522	0.503	0.351	10.332	2.942	0.131	0.771
20	0.874	0.126	1	1	0.653	0.546	0.536	0.397	16.319	1.872	0	0.897
21	0.566	0.393	0.177	0.162	0.162	0.526	0.507	0.363	6.518	4.318	0.425	0.579
22	0.833	0.167	0.162	0.162	0.57	0.55	0.505	0.461	13.918	2.333	0	0.856
23	0.796	0.204	0.16	0.16	1	0.504	0.494	0	12.363	2.728	0	0.819
24	0.684	0.316	0.195	0.195	0.644	0.507	0.5	0.388	8.193	3.477	0	0.702
25	0.855	0.145	0.156	0.169	0.291	0.561	0.507	0.506	18.188	2.552	0	0.877
26	0.829	0.171	1	0.156	0.874	0.583	0.506	0.364	16.839	2.969	0	0.85
27	0.76	0.24	0.156	0.156	0.456	0.541	0.458	0.328	13.726	3.869	0	0.78
28	0.68	0.316	0.16	0.16	0.16	0.467	0.46	0.338	8.618	3.716	0.044	0.696
29	0.696	0.304	0.16	0.16	0.308	0.496	0.457	0.328	9.925	3.975	0	0.714
30	0.598	0.39	0.155	0.155	0.155	0.548	0.551	0.358	8.504	5.239	0.134	0.613
31	0.859	0.141	0.169	0.169	0.778	0.533	0.54	0	17.954	2.44	0	0.88
32	0.559	0.441	0.195	0.164	0.822	0.531	0.524	0	6.243	4.764	0	0.567
33	0.868	0.132	0.156	0.156	1	0.522	0.51	0	18.406	2.29	0	0.889
34	0.779	0.213	0.156	0.18	0.156	0.532	0.515	0.354	13.643	3.278	0.108	0.801
35	0.699	0.301	0.16	0.16	0.766	0.51	0.481	0.492	9.812	3.864	0	0.717
36	0.598	0.39	0.16	0.16	0.174	0.49	0.463	0.333	7.674	4.766	0.138	0.61
37	0.737	0.263	0.16	0.16	0.71	0.571	0.538	0.492	10.973	3.472	0	0.76
38	0.533	0.443	0.156	0.156	0.169	0.457	0.44	0.302	7.861	6.349	0.303	0.542
39	0.647	0.351	0.169	0.156	0.186	0.477	0.43	0.24	10.822	5.563	0.014	0.66
40	0.873	0.127	0.156	0.156	0.415	0.564	0.506	0.457	19.34	2.3	0	0.894
41	0.501	0.499	0.191	0.255	1	0.505	0.507	0	5.732	5.725	0	0.5
42	0.623	0.377	0.16	0.16	1	0.504	0.494	0.392	8.51	4.939	0	0.633
43	0.58	0.408	0.16	0.174	0.16	0.461	0.45	0.322	7.442	5.078	0.126	0.589
44	0.774	0.226	0.16	0.16	1	0.507	0.482	0	11.842	3.031	0	0.796
45	0.624	0.361	0.162	0.162	0.162	0.483	0.46	0.329	7.875	4.306	0.162	0.638
46	0.69	0.298	0.162	0.162	0.162	0.532	0.478	0.384	9.262	3.654	0.128	0.71
47	0.799	0.201	0.16	0.16	1	0.527	0.493	0	12.877	2.776	0	0.823
48	0.597	0.403	0.16	0.16	1	0.488	0.488	0	7.983	5.271	0	0.602
49	0.561	0.439	0.188	0.188	1	0.499	0.489	0	3.268	2.499	0	0.567
50	0.679	0.321	0.16	0.16	1	0.509	0.464	0	10.021	4.419	0	0.694
51	0.833	0.167	0.16	0.16	0.545	0.56	0.503	0.488	14.933	2.519	0	0.856
52	0.635	0.365	0.175	0.797	1	0.518	0.489	0	5.614	3.002	0	0.652
53	0.711	0.289	0.164	0.164	0.815	0.514	0.49	0.363	8.739	3.256	0	0.729
54	0.499	0.481	1	0.183	1	0.475	0.554	0.171	3.196	2.928	0.08	0.515
55	0.711	0.281	0.181	0.181	1	0.506	0.544	0.081	5.243	1.779	0.036	0.743
56	0.789	0.207	0.171	0.171	0.807	0.531	0.544	0.064	8.488	1.845	0.028	0.819
57	0.685	0.309	0.171	0.171	0.209	0.512	0.53	0.104	6.682	2.697	0.042	0.709
58	0.584	0.416	0.172	0.193	1	0.519	0.519	0	5.222	3.571	0	0.594
59	0.691	0.309	0.167	0.184	1	0.497	0.452	0	8.012	3.281	0	0.709
60	0.697	0.303	0.164	0.164	1	0.51	0.491	0	8.52	3.417	0	0.714
61	0.771	0.229	0.16	0.16	0.71	0.525	0.496	0.295	11.788	3.129	0	0.79
62	0.758	0.234	0.183	0.183	0.781	0.566	0.548	0.085	5.763	1.652	0.037	0.773
63	0.606	0.394	0.172	0.172	0.892	0.516	0.513	0	5.672	3.493	0	0.619
64	0.704	0.296	0.167	0.167	0.635	0.533	0.505	0.47	8.546	3.247	0	0.725
65	0.757	0.243	0.167	1	1	0.549	0.493	0	10.116	2.848	0	0.78
66	0.767	0.228	0.171	0.171	1	0.515	0.517	0.064	8.586	2.173	0.03	0.796
67	0.808	0.188	0.171	0.19	0.62	0.553	0.522	0.064	9.457	1.88	0.028	0.832
68	0.784	0.216	0.162	0.162	0.388	0.502	0.479	0.415	11.143	2.677	0	0.806
69	0.701	0.299	0.167	0.167	0.687	0.5	0.456	0.319	8.659	3.384	0	0.719
70	0.63	0.324	0.191	1	0.191	0.541	0.506	0.193	3.339	1.575	0.17	0.657

	P2p	G1p	G2p	G3p	E1p	E2p	E3p	L1s	L2s	L3s	P1s	P2s
1	0.314	1	1	1	0.376	0.296	0	2.992	1.081	0	0.735	0.265
2	0.205	1	1	1	0.417	0.292	0	6.001	1.128	0	0.842	0.158
3	0.281	1	0.179	1	0.392	0.302	0	4.901	1.552	0	0.76	0.24
4	0.142	1	1	1	0.444	0.29	0	9.998	1.155	0	0.896	0.104
5	0.205	0.169	0.169	1	0.429	0.3	0	8.327	1.679	0	0.832	0.168
6	0.199	1	1	1	0.416	0.302	0	8.132	1.547	0	0.84	0.16
7	0.5	0.224	0.224	1	0.347	0.347	0	4.074	4.073	0	0.5	0.5
8	0.368	1	1	1	0.35	0.307	0	4.253	2.24	0	0.655	0.345
9	0.32	0.169	0.169	0.169	0.412	0.342	0.193	5.865	2.483	0.068	0.697	0.295
10	0.367	0.2	1	1	0.421	0.365	0	5.73	3.013	0	0.655	0.345
11	0.252	0.167	0.167	1	0.39	0.305	0	7.509	2.121	0	0.78	0.22
12	0.437	0.167	1	1	0.347	0.319	0	4.912	3.656	0	0.573	0.427
13	0.157	0.162	0.162	0.889	0.443	0.316	0.317	11.901	1.695	0	0.875	0.125
14	0.183	0.162	0.162	0.827	0.446	0.296	0.229	11.458	2.078	0	0.846	0.154
15	0.252	0.167	0.167	1	0.39	0.305	0	7.509	2.121	0	0.78	0.22
16	0.164	1	1	1	0.449	0.308	0	12.026	1.832	0	0.868	0.132
17	0.227	1	0.162	1	0.464	0.338	0	10.406	2.574	0	0.802	0.198
18	0.323	0.162	0.162	0.824	0.403	0.327	0.173	7.843	3.395	0	0.698	0.302
19	0.219	0.162	0.162	0.162	0.428	0.324	0.202	9.301	2.219	0.088	0.801	0.191
20	0.103	1	1	0.653	0.462	0.289	0.237	14.982	1.172	0	0.927	0.073
21	0.383	0.162	0.162	0.162	0.414	0.362	0.231	5.698	3.545	0.309	0.596	0.371
22	0.144	0.162	0.162	0.703	0.464	0.297	0.331	12.745	1.61	0	0.888	0.112
23	0.181	0.16	0.16	1	0.408	0.297	0	11.116	1.961	0	0.85	0.15
24	0.298	0.195	0.195	0.644	0.386	0.322	0	7.115	2.67	0	0.727	0.273
25	0.123	0.156	0.169	0.291	0.487	0.304	0.38	16.938	1.804	0	0.904	0.096
26	0.15	1	0.156	0.846	0.506	0.32	0.208	15.681	2.209	0	0.877	0.123
27	0.22	0.156	0.156	0.456	0.457	0.309	0.177	12.601	3.051	0	0.805	0.195
28	0.3	0.16	0.16	0.187	0.362	0.308	0.193	7.557	2.932	0.03	0.718	0.279
29	0.286	0.16	0.174	0.398	0.397	0.307	0.174	8.852	3.144	0	0.738	0.262
30	0.378	0.155	0.155	0.155	0.436	0.391	0.192	7.564	4.354	0.086	0.63	0.363
31	0.12	0.169	0.18	0.841	0.457	0.316	0	16.612	1.693	0	0.908	0.092
32	0.433	0.164	0.164	0.822	0.39	0.361	0	5.29	3.865	0	0.578	0.422
33	0.111	0.156	0.156	1	0.447	0.295	0	17.046	1.565	0	0.916	0.084
34	0.192	0.156	0.18	0.169	0.448	0.338	0.19	12.497	2.511	0.067	0.829	0.167
35	0.283	0.16	0.16	0.766	0.405	0.321	0.314	8.714	3.04	0	0.741	0.259
36	0.379	0.16	0.16	0.16	0.381	0.325	0.199	6.705	3.9	0.098	0.626	0.364
37	0.24	0.16	0.16	0.894	0.468	0.346	0.309	9.919	2.656	0	0.789	0.211
38	0.437	0.156	0.156	0.156	0.363	0.33	0.179	6.952	5.426	0.212	0.552	0.431
39	0.339	0.156	0.156	0.204	0.387	0.313	0.121	9.713	4.65	0.008	0.676	0.324
40	0.106	0.156	0.156	0.381	0.492	0.295	0.33	18.075	1.581	0	0.92	0.08
41	0.5	0.205	0.191	1	0.359	0.362	0	4.793	4.79	0	0.5	0.5
42	0.367	0.16	0.16	1	0.389	0.349	0.255	7.432	4.064	0	0.646	0.354
43	0.402	0.16	0.16	0.167	0.352	0.323	0.191	6.452	4.221	0.087	0.6	0.392
44	0.204	0.16	0.174	1	0.41	0.3	0	10.632	2.247	0	0.826	0.174
45	0.349	0.162	0.162	0.162	0.378	0.322	0.191	6.887	3.486	0.111	0.657	0.333
46	0.28	0.162	0.162	0.177	0.436	0.327	0.224	8.332	2.893	0.085	0.737	0.256
47	0.177	0.16	0.16	1	0.435	0.3	0	11.691	2.012	0	0.853	0.147
48	0.398	0.16	0.16	1	0.372	0.354	0	6.927	4.415	0	0.611	0.389
49	0.433	0.188	0.188	1	0.335	0.314	0	2.535	1.849	0	0.578	0.422
50	0.306	0.16	0.16	1	0.407	0.323	0	8.934	3.59	0	0.713	0.287
51	0.144	0.16	0.16	0.545	0.47	0.301	0.354	13.696	1.802	0	0.884	0.116
52	0.348	0.175	0.797	1	0.378	0.31	0	4.689	2.24	0	0.677	0.323
53	0.271	0.164	0.164	0.883	0.396	0.316	0.234	7.648	2.507	0	0.753	0.247
54	0.472	1	0.183	1	0.34	0.358	0.051	2.58	2.048	0	0.557	0.443
55	0.252	0.181	0.181	1	0.393	0.312	0.023	4.547	1.121	0	0.802	0.198
56	0.178	0.171	0.171	0.807	0.438	0.308	0.018	7.68	1.176	0	0.867	0.133
57	0.286	0.171	0.171	0.228	0.405	0.336	0.035	5.876	1.954	0.008	0.75	0.249
58	0.406	0.172	0.172	1	0.371	0.342	0	4.335	2.791	0	0.608	0.392
59	0.291	0.167	0.167	1	0.384	0.292	0	6.987	2.501	0	0.736	0.264
60	0.286	0.164	0.164	1	0.391	0.319	0	7.438	2.651	0	0.737	0.263
61	0.21	0.16	0.16	0.661	0.422	0.317	0.148	10.592	2.404	0	0.815	0.185
62	0.222	0.183	0.183	0.781	0.411	0.344	0.023	4.687	1.2	0	0.796	0.204
63	0.381	0.172	0.172	0.892	0.376	0.336	0	4.768	2.704	0	0.638	0.362
64	0.275	0.167	0.167	0.635	0.417	0.322	0.284	7.5	2.447	0	0.754	0.246
65	0.22	0.167	1	1	0.438	0.303	0	8.982	2.08	0	0.812	0.188
66	0.201	0.171	0.171	1	0.417	0.303	0.018	7.703	1.455	0	0.841	0.159
67	0.165	0.171	0.19	0.62	0.44	0.304	0.018	8.259	1.253	0	0.868	0.132
68	0.194	0.162	0.162	0.37	0.4	0.292	0.275	9.938	1.954	0	0.836	0.164
69	0.281	0.167	0.167	0.746	0.386	0.295	0.177	7.552	2.587	0	0.745	0.255
70	0.31	0.191	1	0.191	0.4	0.316	0.07	2.621	1.01	0.014	0.719	0.277

	G1s	G2s	G3s	E1s	E2s	E3s	Tu	Tm	Tv	Te	Tp	Ts
1	1	1	1	0.237	0.116	0	6.841	4.416	5.242	6.751	5.494	4.072
2	1	1	1	0.31	0.115	0	10.434	7.481	8.459	10.32	8.764	7.129
3	1	0.208	1	0.28	0.141	0	9.723	6.801	7.769	9.61	8.07	6.452
4	1	1	1	0.357	0.114	0	14.96	11.514	12.637	14.823	12.99	11.153
5	0.193	0.193	1	0.338	0.145	0	13.722	10.361	11.456	13.589	11.801	10.006
6	1	1	1	0.325	0.138	0	13.425	10.035	11.14	13.291	11.487	9.679
7	0.308	0.308	1	0.232	0.232	0	11.908	8.495	9.607	11.772	9.957	8.147
8	1	1	1	0.237	0.171	0	9.999	6.835	7.851	9.871	8.174	6.493
9	0.193	0.193	0.193	0.31	0.202	0.08	11.897	8.751	9.776	11.772	10.099	8.416
10	0.25	1	1	0.311	0.222	0	12.552	9.069	10.177	12.409	10.531	8.743
11	0.188	0.188	1	0.297	0.163	0	13.613	9.978	11.134	13.464	11.504	9.63
12	0.188	1	1	0.243	0.205	0	12.518	8.909	10.057	12.37	10.424	8.568
13	0.183	0.183	0.781	0.367	0.153	0.168	17.654	13.945	15.141	17.505	15.519	13.596
14	0.183	0.183	0.392	0.37	0.161	0.091	17.447	13.883	15.032	17.303	15.396	13.536
15	0.188	0.188	1	0.297	0.163	0	13.613	9.978	11.134	13.464	11.504	9.63
16	1	1	1	0.373	0.154	0	17.947	14.215	15.418	17.797	15.799	13.858
17	1	0.183	1	0.382	0.196	0	16.908	13.332	14.485	16.764	14.85	12.98
18	0.183	0.183	0.531	0.314	0.211	0.058	15.185	11.572	12.737	15.039	13.106	11.238
19	0.183	0.183	0.183	0.347	0.184	0.09	15.389	11.943	13.054	15.251	13.406	11.608
20	1	1	0.366	0.389	0.113	0.114	20.446	16.527	17.79	20.288	18.19	16.154
21	0.183	0.183	0.183	0.317	0.244	0.125	13.132	9.881	10.929	13.001	11.261	9.552
22	0.183	0.183	0.302	0.389	0.141	0.205	18.344	14.709	15.881	18.197	16.252	14.355
23	0.179	0.179	1	0.33	0.154	0	17.418	13.422	14.685	17.252	15.09	13.077
24	0.238	0.211	0.392	0.292	0.19	0	13.882	10.117	11.29	13.723	11.67	9.784
25	0.175	0.198	0.208	0.423	0.152	0.273	22.997	19.094	20.343	22.838	20.74	18.742
26	1	0.175	0.797	0.439	0.177	0.094	21.96	18.237	19.428	21.809	19.807	17.89
27	0.175	0.175	0.268	0.385	0.192	0.07	19.804	15.984	17.206	19.649	17.595	15.652
28	0.179	0.204	0.204	0.279	0.192	0.091	14.539	10.829	12.002	14.385	12.378	10.519
29	0.179	0.179	0.24	0.316	0.192	0.068	16.099	12.325	13.518	15.943	13.901	11.997
30	0.172	0.172	0.172	0.345	0.27	0.08	16.09	12.302	13.494	15.932	13.877	12.004
31	0.198	0.198	0.797	0.391	0.152	0	22.769	18.662	19.976	22.602	20.394	18.305
32	0.211	0.183	0.878	0.28	0.237	0	13.185	9.478	10.633	13.029	11.007	9.155
33	0.175	0.175	1	0.384	0.138	0	23.061	18.972	20.28	22.895	20.697	18.612
34	0.175	0.219	0.187	0.377	0.199	0.075	19.239	15.416	16.64	19.084	17.029	15.075
35	0.179	0.179	0.592	0.32	0.199	0.179	15.912	12.075	13.288	15.753	13.677	11.753
36	0.179	0.179	0.179	0.292	0.219	0.102	14.747	11.023	12.2	14.593	12.578	10.704
37	0.179	0.179	0.436	0.383	0.203	0.173	16.599	12.9	14.069	16.446	14.444	12.576
38	0.175	0.175	0.198	0.285	0.241	0.087	16.702	12.917	14.128	16.548	14.513	12.59
39	0.175	0.175	0.23	0.312	0.219	0.042	18.712	14.711	15.991	18.549	16.398	14.372
40	0.175	0.175	0.268	0.43	0.139	0.202	23.875	20.009	21.246	23.718	21.64	19.657
41	0.228	0.179	1	0.254	0.251	0	13.738	9.898	11.073	13.571	11.457	9.583
42	0.179	0.179	0.407	0.296	0.236	0.145	15.725	11.817	13.053	15.563	13.449	11.496
43	0.179	0.179	0.179	0.264	0.223	0.095	14.836	11.075	12.264	14.68	12.645	10.76
44	0.179	0.179	0.883	0.331	0.165	0	17.175	13.223	14.473	17.011	14.873	12.879
45	0.183	0.183	0.183	0.29	0.211	0.087	14.414	10.814	11.975	14.269	12.342	10.485
46	0.183	0.183	0.183	0.353	0.205	0.101	14.947	11.639	12.706	14.814	13.043	11.31
47	0.179	0.179	1	0.358	0.158	0	17.895	14.047	15.263	17.736	15.653	13.703
48	0.179	0.179	1	0.28	0.249	0	15.486	11.655	12.866	15.327	13.254	11.342
49	0.218	0.218	1	0.202	0.172	0	7.206	4.688	5.509	7.106	5.767	4.384
50	0.179	0.179	1	0.323	0.213	0	16.663	12.847	14.053	16.505	14.44	12.524
51	0.179	0.179	0.283	0.395	0.154	0.238	19.712	15.833	17.059	19.551	17.452	15.498
52	0.2	1	1	0.264	0.173	0	10.47	7.258	8.289	10.339	8.616	6.929
53	0.183	0.183	0.531	0.303	0.187	0	14.169	10.471	11.623	14.013	11.996	10.155
54	1	0.218	1	0.221	0.175	0	7.67	5.01	5.925	7.572	6.203	4.628
55	0.213	0.213	1	0.296	0.124	0	8.462	5.952	6.798	8.367	7.059	5.668
56	0.197	0.197	1	0.358	0.125	0	11.961	9.137	10.07	11.852	10.361	8.856
57	0.197	0.197	0.236	0.313	0.176	0.001	11.089	8.145	9.117	10.974	9.42	7.837
58	0.193	0.193	1	0.256	0.209	0	10.714	7.438	8.462	10.576	8.793	7.126
59	0.188	0.188	1	0.292	0.17	0	13.345	9.81	10.934	13.199	11.293	9.488
60	0.183	0.211	1	0.298	0.192	0	14.119	10.405	11.562	13.962	11.937	10.09
61	0.177	0.177	0.653	0.341	0.187	0.057	17.249	13.304	14.521	17.08	14.917	12.996
62	0.213	0.213	0.64	0.272	0.18	0	9.035	6.249	7.165	8.926	7.452	5.887
63	0.193	0.193	1	0.266	0.201	0	11.109	7.793	8.83	10.97	9.165	7.472
64	0.188	0.188	0.622	0.322	0.183	0.147	13.883	10.281	11.427	13.735	11.793	9.947
65	0.188	1	1	0.345	0.162	0	15.12	11.405	12.587	14.968	12.964	11.062
66	0.197	0.197	1	0.336	0.136	0	12.553	9.44	10.468	12.432	10.789	9.158
67	0.197	0.197	0.491	0.335	0.135	0	13.257	9.918	11.021	13.127	11.365	9.513
68	0.181	0.181	0.246	0.318	0.155	0.165	16.08	12.217	13.43	15.919	13.821	11.892
69	0.188	0.188	0.516	0.294	0.173	0.075	14.213	10.474	11.663	14.06	12.043	10.138
70	0.24	1	0.24	0.247	0.13	0	6.23	4.034	4.85	6.158	5.084	3.645

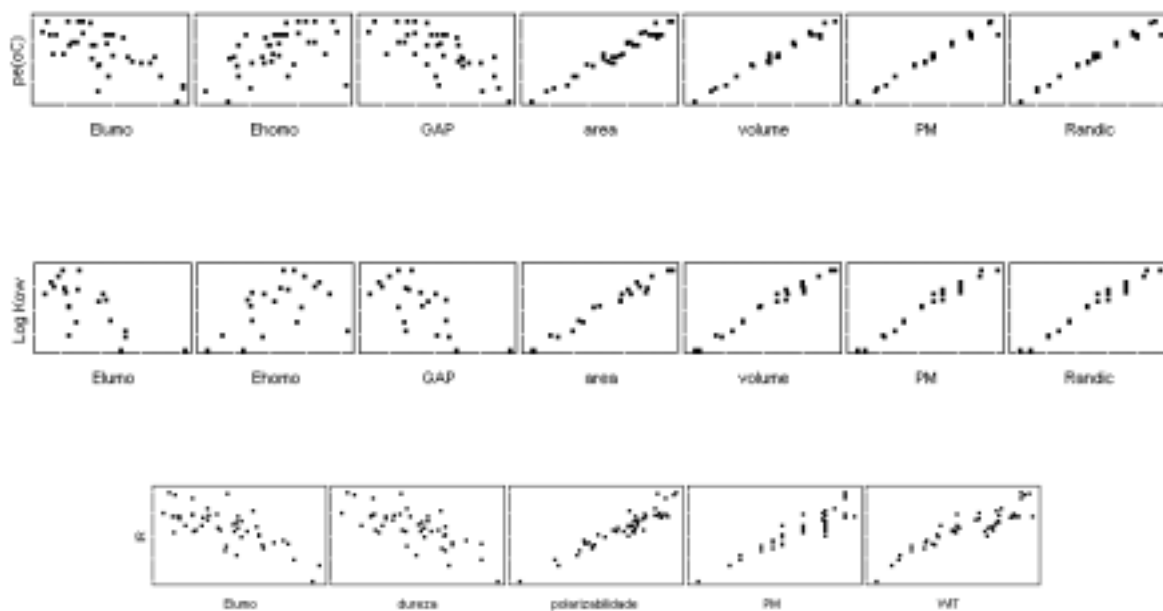
	Au	Am	Av	Ae	Ap	As	Gu	Gm	Gs	Ku	Km	Kv
1	10.5	3.951	5.856	10.204	6.505	3.233	1	1	1	0.5	0.576	0.538
2	19.63	7.999	11.39	19.101	12.538	6.769	1	1	1	0.646	0.741	0.702
3	20.095	8.737	12.035	19.576	13.156	7.604	0.564	0.564	0.593	0.54	0.621	0.587
4	31.594	13.442	18.76	30.77	20.551	11.546	1	1	1	0.745	0.828	0.796
5	33.483	15.711	20.909	32.674	22.663	13.977	0.306	0.317	0.334	0.653	0.733	0.702
6	31.637	14.258	19.319	30.843	21.032	12.584	1	1	1	0.659	0.744	0.711
7	35.45	18.043	23.076	34.648	24.786	16.593	0.38	0.369	0.456	0.5	0.5	0.5
8	23.615	10.654	14.276	22.994	15.532	9.526	1	1	1	0.5	0.5	0.5
9	32.779	16.594	21.308	32.039	22.902	15.129	0.187	0.169	0.193	0.482	0.534	0.511
10	37.273	18.707	23.957	36.391	25.763	17.264	0.585	0.585	0.63	0.5	0.5	0.5
11	36.97	17.575	23.056	36.049	24.941	15.928	0.303	0.303	0.328	0.587	0.657	0.63
12	38.7	19.446	24.865	37.781	26.735	17.957	0.55	0.55	0.573	0.5	0.5	0.5
13	46.535	22.402	29.481	45.439	31.862	20.169	0.286	0.286	0.297	0.726	0.801	0.773
14	49.853	26.063	33.078	48.778	35.429	23.807	0.282	0.297	0.236	0.69	0.758	0.733
15	36.97	17.575	23.056	36.049	24.941	15.928	0.303	0.303	0.328	0.587	0.657	0.63
16	49.417	24.38	31.742	48.282	34.214	22.029	1	1	1	0.716	0.789	0.762
17	53.698	29.026	36.3	52.584	38.738	26.788	0.545	0.545	0.568	0.624	0.692	0.666
18	51.634	28.497	35.254	50.578	37.533	26.626	0.29	0.29	0.261	0.5	0.539	0.521
19	45.456	23.593	29.989	44.46	32.144	21.648	0.167	0.162	0.183	0.62	0.69	0.664
20	46.397	20.28	27.966	45.215	30.545	17.562	0.893	0.848	0.716	0.809	0.879	0.853
21	45.219	24.805	30.745	44.283	32.753	23.054	0.167	0.167	0.183	0.437	0.449	0.445
22	47.213	22.909	30.075	46.116	32.478	20.52	0.258	0.258	0.216	0.747	0.819	0.792
23	49.504	23.952	31.232	48.299	33.721	21.795	0.295	0.295	0.318	0.692	0.763	0.736
24	41.729	20.613	26.453	40.696	28.484	18.993	0.291	0.284	0.27	0.524	0.581	0.559
25	65.949	33.63	43.214	64.497	46.412	30.56	0.228	0.197	0.193	0.781	0.846	0.822
26	68.917	37.565	46.883	67.512	49.988	34.639	0.504	0.51	0.519	0.741	0.805	0.782
27	71.779	40.977	50.063	70.386	53.106	38.444	0.231	0.226	0.202	0.638	0.699	0.677
28	46.327	24.113	30.403	45.272	32.562	22.474	0.16	0.172	0.195	0.519	0.569	0.55
29	54.997	29.781	36.999	53.813	39.459	27.84	0.199	0.213	0.198	0.542	0.598	0.577
30	63.447	35.823	43.708	62.145	46.4	33.963	0.155	0.155	0.172	0.483	0.488	0.486
31	63.469	31.078	40.618	62.002	43.816	28.118	0.296	0.294	0.315	0.786	0.851	0.827
32	42.864	21.948	27.729	41.841	29.741	20.447	0.309	0.281	0.324	0.5	0.5	0.5
33	61.343	29.713	39.032	59.91	42.155	26.681	0.298	0.29	0.313	0.8	0.864	0.841
34	64.547	34.918	43.62	63.2	46.542	32.39	0.168	0.164	0.193	0.667	0.734	0.709
35	53.366	28.346	35.48	52.187	37.919	26.486	0.27	0.27	0.267	0.547	0.604	0.582
36	53.382	28.968	35.911	52.228	38.288	27.194	0.173	0.16	0.179	0.481	0.486	0.484
37	53.638	28.35	35.62	52.457	38.092	26.351	0.277	0.284	0.241	0.603	0.673	0.647
38	72.407	42.582	51.284	71.041	54.218	40.345	0.167	0.16	0.183	0.464	0.473	0.47
39	80.131	47.727	57.229	78.655	60.423	45.289	0.179	0.166	0.192	0.498	0.508	0.499
40	63.852	31.772	41.302	62.414	44.479	28.584	0.24	0.205	0.202	0.807	0.87	0.847
41	47.181	24.494	30.653	46.045	32.816	22.957	0.365	0.314	0.344	0.5	0.5	0.5
42	58.137	32.057	39.488	56.907	42.029	30.204	0.295	0.303	0.235	0.5	0.5	0.5
43	54.637	29.932	36.955	53.469	39.36	28.166	0.165	0.165	0.179	0.483	0.487	0.486
44	51.82	26.02	33.382	50.605	35.897	23.886	0.31	0.291	0.305	0.659	0.727	0.702
45	49.88	26.971	33.623	48.827	35.875	25.166	0.167	0.162	0.183	0.478	0.483	0.481
46	48.659	26.982	33.355	47.676	35.495	25.057	0.167	0.167	0.183	0.533	0.595	0.572
47	51.679	25.783	33.218	50.468	35.748	23.523	0.303	0.295	0.318	0.697	0.768	0.742
48	57.731	32.368	39.605	56.537	42.077	30.586	0.295	0.295	0.318	0.5	0.5	0.5
49	12.792	5.375	7.445	12.438	8.166	4.687	0.328	0.328	0.362	0.5	0.5	0.5
50	60.606	34.069	41.689	59.366	44.282	32.075	0.295	0.295	0.318	0.517	0.563	0.546
51	54.333	27.073	34.95	53.067	37.619	24.683	0.255	0.236	0.209	0.748	0.815	0.791
52	25.421	11.654	15.519	24.764	16.856	10.505	0.519	0.519	0.585	0.5	0.504	0.5
53	41.385	20.738	26.47	40.38	28.459	19.175	0.296	0.288	0.261	0.564	0.62	0.599
54	15.265	6.293	8.949	14.868	9.845	5.284	0.568	0.568	0.602	0.47	0.493	0.483
55	14.935	6.079	8.702	14.542	9.583	5.098	0.32	0.335	0.356	0.564	0.671	0.626
56	24.145	10.547	14.596	23.541	15.948	9.029	0.297	0.287	0.338	0.68	0.778	0.739
57	26.863	12.909	17.026	26.236	18.409	11.541	0.183	0.177	0.209	0.525	0.605	0.573
58	27.893	13.228	17.242	27.169	18.648	12.101	0.31	0.31	0.334	0.5	0.5	0.5
59	38.133	18.998	24.425	37.228	26.288	17.476	0.303	0.303	0.328	0.534	0.594	0.571
60	42.217	21.282	27.093	41.198	29.11	19.721	0.3	0.309	0.338	0.543	0.597	0.576
61	52.811	27.347	34.435	51.572	36.888	25.468	0.281	0.257	0.274	0.654	0.714	0.692
62	15.159	6.461	8.941	14.753	9.793	5.624	0.284	0.297	0.307	0.636	0.687	0.666
63	29.489	14.111	18.335	28.732	19.811	12.893	0.298	0.31	0.334	0.5	0.5	0.5
64	40.302	19.997	25.77	39.344	27.748	18.351	0.275	0.26	0.28	0.553	0.62	0.594
65	42.192	20.518	26.7	41.172	28.814	18.683	0.55	0.55	0.573	0.634	0.706	0.678
66	28.475	12.771	17.421	27.773	18.98	11.204	0.308	0.308	0.338	0.648	0.74	0.704
67	27.537	11.989	16.557	26.835	18.096	10.351	0.272	0.256	0.267	0.71	0.787	0.757
68	43.945	21.251	27.636	42.858	29.835	19.423	0.232	0.213	0.201	0.674	0.742	0.717
69	42.45	21.212	27.236	41.445	29.303	19.532	0.267	0.292	0.263	0.55	0.607	0.585
70	9.645	3.513	5.464	9.4	6.096	2.698	0.331	0.331	0.386	0.443	0.544	0.496

	Ke	Kp	Ks	Du	Dm	Dv	De	Dp	Ds	Vu	Vm	Vv
1	0.5	0.529	0.602	0.359	0.14	0.202	0.349	0.224	0.118	17.341	8.367	11.098
2	0.649	0.692	0.763	0.364	0.16	0.217	0.355	0.237	0.142	30.064	15.48	19.849
3	0.542	0.579	0.639	0.35	0.158	0.213	0.341	0.231	0.14	29.818	15.538	19.804
4	0.747	0.787	0.845	0.368	0.173	0.226	0.358	0.245	0.157	46.554	24.957	31.397
5	0.655	0.693	0.748	0.353	0.176	0.226	0.345	0.243	0.161	47.204	26.072	32.366
6	0.662	0.702	0.76	0.355	0.17	0.221	0.346	0.239	0.155	45.063	24.292	30.459
7	0.5	0.5	0.5	0.331	0.168	0.215	0.324	0.231	0.155	47.358	26.538	32.683
8	0.5	0.5	0.5	0.332	0.151	0.202	0.324	0.219	0.136	33.615	17.489	22.127
9	0.482	0.505	0.545	0.476	0.219	0.291	0.464	0.316	0.197	49.205	26.565	33.102
10	0.5	0.5	0.5	0.378	0.192	0.244	0.369	0.262	0.178	49.824	27.776	34.134
11	0.589	0.622	0.67	0.341	0.167	0.215	0.332	0.232	0.153	50.584	27.553	34.19
12	0.5	0.5	0.5	0.321	0.162	0.206	0.313	0.222	0.149	51.218	28.355	34.923
13	0.728	0.765	0.813	0.536	0.253	0.331	0.522	0.358	0.229	64.19	36.347	44.622
14	0.693	0.726	0.77	0.486	0.229	0.299	0.474	0.324	0.207	67.3	39.946	48.11
15	0.589	0.622	0.67	0.341	0.167	0.215	0.332	0.232	0.153	50.583	27.553	34.19
16	0.719	0.754	0.802	0.357	0.189	0.236	0.349	0.252	0.176	67.364	38.595	47.16
17	0.626	0.659	0.703	0.366	0.206	0.252	0.359	0.267	0.193	70.606	42.358	50.785
18	0.5	0.516	0.547	0.452	0.213	0.278	0.44	0.301	0.195	66.821	40.07	47.992
19	0.622	0.656	0.702	0.47	0.227	0.295	0.459	0.318	0.207	68.413	37.72	46.54
20	0.812	0.846	0.891	0.507	0.229	0.303	0.493	0.329	0.205	66.843	36.807	45.756
21	0.438	0.443	0.451	0.476	0.248	0.313	0.465	0.336	0.229	79.363	41.891	52.362
22	0.749	0.785	0.832	0.517	0.269	0.34	0.505	0.364	0.245	65.558	37.618	45.956
23	0.694	0.729	0.775	0.341	0.174	0.219	0.333	0.235	0.161	66.922	37.375	45.917
24	0.526	0.553	0.591	0.477	0.173	0.22	0.465	0.236	0.161	55.611	30.73	37.743
25	0.783	0.815	0.856	0.536	0.302	0.367	0.525	0.39	0.283	88.952	52.727	63.56
26	0.743	0.775	0.815	0.496	0.256	0.321	0.484	0.344	0.237	90.88	55.803	66.312
27	0.64	0.67	0.708	0.453	0.233	0.293	0.443	0.314	0.216	91.586	56.962	67.271
28	0.52	0.544	0.578	0.433	0.204	0.266	0.422	0.288	0.187	63.54	35.726	43.637
29	0.543	0.571	0.607	0.438	0.209	0.271	0.427	0.293	0.192	71.132	42.116	50.533
30	0.483	0.485	0.489	0.498	0.249	0.316	0.485	0.34	0.232	90.87	51.473	62.466
31	0.788	0.821	0.861	0.366	0.194	0.241	0.358	0.257	0.181	86.239	49.74	60.594
32	0.5	0.5	0.5	0.36	0.185	0.233	0.352	0.25	0.172	56.05	31.427	38.362
33	0.803	0.834	0.874	0.352	0.187	0.232	0.344	0.247	0.174	84.404	48.685	59.312
34	0.669	0.702	0.743	0.479	0.236	0.302	0.467	0.326	0.217	93.073	52.903	64.457
35	0.549	0.576	0.612	0.507	0.251	0.322	0.494	0.346	0.232	69.278	40.421	48.768
36	0.482	0.484	0.486	0.439	0.22	0.281	0.429	0.302	0.204	77.38	42.94	52.583
37	0.605	0.639	0.683	0.547	0.273	0.348	0.533	0.375	0.253	70.237	41.25	49.69
38	0.464	0.469	0.475	0.408	0.219	0.272	0.399	0.291	0.204	115.482	64.674	78.942
39	0.498	0.499	0.514	0.391	0.205	0.256	0.382	0.274	0.191	100.4	62.897	73.952
40	0.809	0.841	0.879	0.52	0.28	0.349	0.509	0.372	0.257	87.734	51.783	62.551
41	0.5	0.5	0.5	0.346	0.179	0.224	0.337	0.24	0.168	60.919	34.392	41.726
42	0.5	0.5	0.5	0.474	0.244	0.309	0.463	0.331	0.226	73.862	43.874	52.541
43	0.483	0.485	0.488	0.421	0.209	0.268	0.411	0.288	0.194	78.22	43.76	53.419
44	0.661	0.694	0.738	0.338	0.177	0.221	0.33	0.237	0.165	68.995	39.243	47.854
45	0.478	0.48	0.485	0.435	0.213	0.275	0.424	0.297	0.196	74.327	40.92	50.438
46	0.535	0.565	0.605	0.476	0.24	0.306	0.465	0.329	0.22	71.596	41.075	49.889
47	0.699	0.734	0.78	0.348	0.184	0.229	0.34	0.245	0.172	69.574	39.829	48.481
48	0.5	0.5	0.5	0.332	0.186	0.228	0.325	0.242	0.176	73.217	44.022	52.47
49	0.5	0.5	0.5	0.339	0.143	0.197	0.329	0.216	0.125	19.999	10.063	12.954
50	0.519	0.541	0.57	0.331	0.189	0.229	0.324	0.243	0.179	77.269	46.916	55.743
51	0.75	0.783	0.826	0.529	0.283	0.351	0.517	0.375	0.263	74.046	42.906	52.009
52	0.5	0.5	0.515	0.345	0.16	0.212	0.336	0.23	0.146	35.891	18.912	23.808
53	0.566	0.593	0.63	0.468	0.176	0.293	0.456	0.316	0.164	55.554	31.209	38.093
54	0.471	0.481	0.5	0.413	0.155	0.224	0.4	0.249	0.132	25.07	11.437	15.442
55	0.567	0.614	0.703	0.388	0.159	0.221	0.377	0.243	0.14	24.394	12.089	15.754
56	0.683	0.729	0.801	0.391	0.178	0.234	0.38	0.255	0.161	37.401	19.762	24.999
57	0.527	0.564	0.625	0.392	0.18	0.238	0.382	0.258	0.164	39.882	21.266	26.738
58	0.5	0.5	0.5	0.355	0.169	0.22	0.346	0.238	0.155	38.607	20.666	25.704
59	0.536	0.564	0.605	0.324	0.166	0.21	0.316	0.225	0.154	51.478	28.808	35.359
60	0.545	0.571	0.606	0.342	0.175	0.221	0.333	0.237	0.163	56.336	31.688	38.655
61	0.656	0.685	0.722	0.451	0.211	0.273	0.439	0.296	0.195	70.06	40.651	48.956
62	0.637	0.66	0.694	0.411	0.172	0.236	0.4	0.259	0.151	25.241	12.772	16.37
63	0.5	0.5	0.5	0.352	0.17	0.22	0.343	0.237	0.156	40.598	21.904	27.165
64	0.555	0.587	0.631	0.516	0.239	0.314	0.502	0.341	0.217	54.186	30.278	37.197
65	0.636	0.67	0.718	0.356	0.182	0.23	0.347	0.247	0.169	57.312	31.923	39.287
66	0.651	0.694	0.762	0.375	0.172	0.227	0.365	0.246	0.157	42.663	22.313	28.316
67	0.712	0.748	0.802	0.391	0.175	0.233	0.38	0.254	0.157	42.263	21.996	27.954
68	0.677	0.709	0.753	0.478	0.232	0.299	0.465	0.322	0.213	60.026	33.469	41.067
69	0.551	0.578	0.617	0.437	0.199	0.263	0.425	0.286	0.18	56.663	31.686	38.899
70	0.446	0.485	0.579	0.425	0.155	0.235	0.413	0.262	0.126	18.178	7.742	11.01

	Ve	Vp	Vs	pe(oC)	log Kow	ir	phototox
1	16.955	11.999	7.306	218	3.37	2	-3.6
2	29.421	21.302	13.898	340	4.54	3.2	-2.46
3	29.186	21.226	14.056	338	4.57	3	-3.20
4	45.593	33.542	22.699	440	5.76	4.51	-2.89
5	46.263	34.463	23.983	435	5.91	4	-2.52
6	44.134	32.519	22.263	431	1649	4.10	-2.54
7	46.42	34.744	24.74	429	5.49	3.7	-3.20
8	32.865	23.706	16.019	393	5.18	3.58	-2.31
9	48.157	35.322	24.529	-	-	3.64	-
10	48.8	36.293	26.007	497	6.25	4.33	-2.67
11	49.513	36.445	25.558	496	6.04	4.53	-2.43
12	50.151	37.158	26.524	493	-	4.28	-2.95
13	62.944	47.382	33.765	519	-	5.18	-
14	66.082	50.825	37.343	-	-	4.67	-
15	49.513	36.445	25.558	-	-	5	-
16	66.079	50.013	35.886	535	6.75	4.73	-2.44
17	69.348	53.588	39.768	531	-	4.56	-
18	65.618	50.64	37.864	535	-	4.40	-
19	66.981	49.541	35.067	-	-	4.45	-
20	65.503	48.736	33.716	-	-	-	-
21	77.557	55.986	38.847	-	-	4.07	-
22	64.314	48.73	34.876	-	-	4.99	-
23	65.551	48.811	34.872	596	-	6	-
24	54.42	40.154	28.777	547	-	5.08	-
25	87.34	67.156	49.304	-	-	-	-
26	89.323	69.796	52.53	-	-	-	-
27	90.038	70.703	54.096	-	-	-	-
28	62.224	46.343	33.663	-	-	-	-
29	69.791	53.379	39.845	-	-	4.92	-
30	88.959	66.266	48.814	-	-	-	-
31	84.604	64.21	46.423	-	-	-	-
32	54.87	40.748	29.602	542	6.50	4.76	-2.36
33	82.805	62.852	45.292	-	-	-	-
34	91.198	68.383	49.577	-	-	-	-
35	67.94	51.596	38.24	-	-	5.04	-
36	75.718	55.911	40.464	-	-	-	-
37	68.903	52.537	38.927	-	-	-	-
38	113.043	83.861	60.939	-	-	-	-
39	98.702	77.654	60.04	-	-	-	-
40	86.138	66.122	48.243	-	-	-	-
41	59.616	44.273	32.539	590	6.75	-	-
42	72.47	55.478	41.7	592	-	4.97	-
43	76.56	56.75	41.308	595	-	4.89	-
44	67.617	50.77	36.764	594	-	5.73	-
45	72.752	53.694	38.323	-	-	4.27	-
46	70.18	52.879	38.416	-	-	4.33	-
47	68.204	51.401	37.226	-	-	5.87	-
48	71.864	55.331	41.928	-	-	4.90	-
49	19.544	13.933	9.071	270	4	-	-
50	75.871	58.722	44.599	-	-	4.90	-
51	72.619	55.072	40.181	-	-	5.92	-
52	35.104	25.472	17.434	383	5.22	3.37	-2.81
53	54.393	40.455	29.33	-	-	-	-
54	24.458	16.795	9.913	279	3.92	-	-
55	23.848	16.978	10.765	294	4.18	2.70	-3.20
56	36.612	26.749	17.885	398	5.75	3.84	-3.16
57	39.04	28.578	19.467	406	-	3.49	-
58	37.745	27.441	19.227	422	-	-	-
59	50.427	37.581	26.964	-	-	4.22	-
60	55.159	41.046	29.811	534	-	4.84	-
61	68.652	51.805	38.464	531	-	4.93	-
62	24.665	17.594	11.511	-	-	-	-
63	39.702	28.976	20.365	439	-	-	-
64	53.079	39.541	28.298	480	-	4.24	-
65	56.14	41.778	29.746	481	6	4.42	-2.89
66	41.746	30.33	20.362	403	540	3.72	-3.17
67	41.346	29.957	19.864	-	-	-	-
68	58.777	43.656	31.315	-	-	5.27	-
69	55.504	41.347	29.671	481	5.80	4.29	-
70	17.75	12.075	6.38	-	3.33	-	-

Apêndice II

Apresenta os gráficos das variáveis independentes *versus* variável dependente para os estudos de QSAR e QSPR apresentados no capítulo 3.



Referências Bibliográficas

- AMOR, T. B. & JORI, G., "Sunlight-Activated Insecticides: Historical Background and Mechanisms of Phototoxic Activity", *Insect Biochemistry and Molecular Biology*, 30, 2000, 915-925.
- ANKLEY, G. T.; ERICKSON, R. J.; PHIPPS, G. L.; MATTSON, V. R.; KOSIAN, P. A.; SHEEDY, B. R.; COX, J. S., "Effects of Light Intensity on the Phototoxicity of Fluoranthene to a Benthic Macroinvertebrate", *Environ. Sci. Technol.*, 29, 1995, 2828-2833.
- ARUPJYOTI, S. & IRAGAVARAPU, S., "New Electrotopological Descriptors for Prediction of Boiling Points of Alkanes and Aliphatic Alcohols through Artificial Neural Network and Multiple Linear Regression Analysis", *Computers Chem.*, v. 22, n. 6, 1998, 515-522.
- ATKINS, P. W., *Physical Chemistry*, 5th. Ed., Oxford University Press, 1994.
- BARRATT, E. S.; STANFORD, M. S.; DOWDY, L.; LIEBMAN, M. J.; KENT, T. A., "Impulsive and Premeditated Aggression: a Factorial Analysis of Self-reported Acts", *Psychiatry Research*, n.86, 1999, 163-173.
- BEEBE, K. R.; PELL, R. J.; SEASHOLTZ, M. B., *Chemometrics: A Practical Guide*, Wiley-Interscience Publication, 1998.
- BODOR, N. S. & HUANG, M., "Theoretical Study of the Stereoisomers of Tetrahydrocannabinols", *International Journal of Quantum Chemistry: Quantum Biology Symposium*, vol. 61, 1997, 127-135.
- BOEHM, P. D.; PAGE, D. S.; GILFILLAN, E. S.; BENICE, A. E.; BURNS, W. A.; MANKIEWICZ, P. J., "Study of the facts and Effects of the Exxon Valdez Oil Spill on Benthic Sediments in Two Bays in Prince William Sound, Alaska. 1. Study Design, Chemistry, and Source Fingerprinting", *Environ. Sci. Technol.*, v. 32, 1998, 567-576.
- BRENEMAN, C. M. & WILBERG, K. B., "Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials. The need for High Sampling Density in

- Formamide Conformational Analysis”, *J. Computational Chem*, vol. 11, no. 3, 1990, 361-373.
- BURKHARD, L. P.; ANDREN, A. W.; ARMSTRONG, D. E., “Estimation of Vapor Pressure for Polyhalogenated Biphenyls: A Comparison of Eleven Predictive Methods”, *Environ. Sci. Technol.*, 19, 1985, 500-507.
- BURNS, W. A.; MANKIEWICZ, P. J.; BENICE, A. E.; PAGE, D. S.; PARKERS, K. R., “A Principal-Component and Least-Square Method for Allocating Polycyclic Aromatic Hydrocarbons in Sediment to Multiple Sources”, *Environmental Toxicology and Chemistry*, v. 16, n. 6, 1997, 1119-1131.
- CHEN, J., W.; KONG, L. R.; ZHU, C. M.; HUANG, Q. G.; WANG, L. S., “Correlation Between Photolysis Rate Constants for Polycyclic Aromatic Hydrocarbon and Frontier Molecular Orbital Energy”, *Chemosphere*, v. 33, n. 6, 1996, 1143-1150.
- CHIRLIAN, L. E. & FRANCI, M., “Atomic Charges Derived from Electrostatic Potentials: A Detailed Study” *J. Computational Chem.*, vol. 8, 1987, 894-905.
- COFFIN, R. B.; CIFUENTES, L. A.; PRITCHARD, P. H., “Assimilation of Oil-Derived Carbon and Remedial Nitrogen Application by Intertidal Food Chains on a Contaminated Beach in Prince William Sound, Alaska”, *Marine Environmental Research*, v. 44, n.1, 1997, 27-39.
- CRONIN, M. T. D. & SCHULTZ, T. W., “Validation of *Vibrio fischeri* Acute Toxicity Data: Mechanism of Action-Based QSARs for Non-polar Narcotics and Polar Narcotic Phenols”, *The Science of the Total Environment*, 204, 1997, 75-88.
- DEWAR, M. J. S.; ZOEBISCH, E. G.; HEALY, E. F.; STEWART, J. J. P., “AM1: a New General purpose Quantum Mechanical Molecular Model”, *J. Am. Chem. Soc.*, 107, 1985, 3902-3909.
- FERREIRA, M. M. C; ANTUNES, A. M.; MELGO, M. S.; VOLPE, P. L. O., “Quimiometria I: Calibração Multivariada, um Tutorial”, *Química Nova*, v. 22, n. 5, 1999, 724-731.
- FERREIRA, M. M. C; FARIA, C. G.; PAES, E. T., “Oceanographic Characterization of Northern São Paulo Coast: a chemometric study”, *Chemom. Intell. Lab. Syst.*, 47, 1999, 289-297.

- GALT, J. A.; LEHR, W. J.; PAYTON, D. L., "Fate and Transport of the Exxon Valdez Oil Spill, Part 4", *Env. Scie. Technol.*, v. 25, n. 2, 1991, 202-208.
- GELADI, P., "Notes os the History and Nature of Partial Least Squares (PLS) Modelling", *Journal of Chemometrics*, vol. 2, 1988, 231-246.
- GELADI, P.& KOWALSKI, B. R., "Partial Least Square Regression: a Tutorial", *Anal. Chim. Acta*, 185, 1986, 1-17.
- HASWELL, S. J. & WALMSTEY, A. D., "Multivariate Data Visualization Methods Based on Multi-elemental Analysis of Wines and Coffes using Total Reflection X-Ray Fluorescence Analysis", *Journal of Analytical Atomic Spectrometry*, v.13, n.2, 1998, 131-134.
- HATCH, A. C. & BURTON Jr., G. A. "Photo-induced toxicity of PAHs to Hyalella Azteca and Chironomus tentans: effects of mixtures and behavior", *Environmental Pollution*, 106, 1999, 157-167.
- HEHRE, W. J.; BURKE, L. D.; SHUSTERMAN, A. J.; PIETRO, W. J., *Experiments in Computational Organic Chemistry*, Wavefunction, Inc., 1993.
- HOSTETTLER, F. D. & KVENVOLDEN, K. A., "Geochemical Changes in Crude Oil Spilled from the Exxon Valdez Supertanker into Prince William Sound, Alaska", *Org. Geochem.*, v. 21, n. 8/9, 1994, 927-936.
- HUANG, M.; DOERGE, D.; BODOR, N.; POP, E.; BREWSTER, M. E., "Nitrogen Radical Cations as Intermediates in enzymatically Mediated Oxidative Deaminations-Applications of Molecular Parametric Models", *International Journal of Quantum Chemistry: Quantum Biology Symposium*, vol. 22, 1995, 171-179.
- KARCHER, W.; FORDHAM, R. J.; DUBOIS, J. J.; GLAUDE, P. G. J. M.; LIGHART, J. A. M., *Spectral Atlas of Polycyclic Aromatic Compounds*, 1988.
- KARELSON, M.; PERKSON, A., "QSPR Prediction of Densities of Organic Liquids", *Computer & Chemistry*, 23, 1999, 49-59.
- KELSO, D. D. & KENDZIOREK, M., "Alaska's Response to the Exxon Valdez Oil Spill, Part 1", *Env. Scie. Technol.*, v. 25, n. 1, 1991, 16-23.
- KOWALSKI, B. R., "Chemometrics, Mathematics and Statistics in Chemistry", NATO ASI Series, Cosenza, Italy, September, 12-23, 1983.

- KROP, H. B.; VELZEN, J. M. VAN; PARSONS, J. R. GOVES, H. A. J., “n-Octanol-Water PARTition Coefficients, Aqueous Solubilities and henry’s Law Constants of Fatty Acid Esteres”, *Chemosphere*, vol. 34, n. 1, 1997, 107-119.
- KUBINYI, H., “*QSAR: Hansch Analysis and Related Approachs*”, VCH, 1993.
- LABUTE, P., “A Widely Applicable Set of Descriptors”, *Journal of Molecular Graphics and Modelling*, vol. 18, 2000, 464-477.
- LEHNER, A. F.; HORN, J.; FLESHER, J. W.,”Benzylic Carbonic Ions as Ultimate Carcinogenns of Polynuclear Aromatic Hydrocarbons” *J. Molecular Structure*, 366, 1996, 203-217.
- LEITÃO, A.; MONTANARI, C. A.; DONNICI, C. L., “Sobre o Uso de Métodos Quimiométricos em Química Combinatória”, *Química Nova*, v. 23, n. 2, 2000, 178-184.
- MACKAY, D.; SHIU, W.Y.; MA, K.C., *Illustrated Handbook of Physical-Chemical Properties and Environmental Fate for Organic Chemicals*, Lewis, 1998.
- MAKI, A. W., “The Exxon Valdez Oil Spill: Initial Environmental Impact Assessment”, *Env. Scie. Technol.*, v. 25, n. 1, 1991, 24-29.
- MALLAKIN, A.; DIXON, D. G.; GRRENBERG, B. M., “Pathway of Anthracene Modification Under Simulated Solar Radiation”, *Chemosphere*, 40, 2000, 1435-1441.
- MALINOWSKI, E. R., *Factor Analysis in Chemistry*, 2th. Ed., John Wiley & Sons, Inc., 1991.
- MANAHAN, S. E., *Environmental Chemistry*, 6th ed., Lewis Publishers, Boca Ranton/Ann Arbor/London/Tokyo, 1994.
- MATLAB[®] for Windows, versão 5.0, *MathWorks Inc.*, 1995
- MARTÍN, M. J.; PABLOS, F.; GONZÁLEZ, A. G., “Application of Pattern Recognition to the Discrimination of Roasted Coffees”, *Analytica Chimica Acta*, n. 320, 1996, 191-197.
- MARTÍN, M. J.; PABLOS, F.; GONZÁLEZ, A. G., “Characterization of Green Coffee Varieties According to their Metal Content”, *Analytica Chimica Acta*, n. 358, 1998, 177-183.

- MARTÍN, M. J.; PABLOS, F.; GONZÁLEZ, A. G., “Characterization of *arabica* and *robusta* Roasted Coffee Varieties and Mixture Resolution According to their Metal Content”, *Food Chemistry*, v.66, 1999, 365-370.
- MEKENYAN, O. G.; ANKLEY, G. T.; VEITH, G. D.; CALL, D. J., “QSAR for Photoinduced Toxicity: I. Acute Lethality of Polycyclic Aromatic Hydrocarbons to *Daphnia Magna*”, *Chemosphere*, v. 28, n. 3, 1994, 567-582.
- MOITA NETO, J. M. & MOITA, G. C., “Uma Introdução à Análise Exploratória de Dados Multivariados”, *Química Nova*, v. 21, n. 4, 1998, 467-469.
- MORRISON, R. T.; BOYD, R. N., *Química Orgânica*, trad. Manoel Alves da Silva, 10^a. Ed., Fundação Calouste Gulbenkian, Lisboa, 1993, 1492-1494.
- MULLIKEN, R. S., “Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I-IV”, *J. Chem. Phys.*, vol. 23, no. 10, 1995, 1833-1846/ 2338-2346.
- PAGE, D. S.; BOEHM, P. D.; DOUGLAS, G. S.; BENICE, A. E.; BURNS W. A.; MANKIEWICZ, P. J., “The Natural Petroleum Hydrocarbon Background in Subtidal Sediments of Prince William Sound, Alaska, USA”, *Env. Toxicol. and Chem.*, v. 15, n. 8, 1996, 1266-1281.
- PEACHEY, R. L. & CROSBY, D. G., “Phototoxicity in Tropical Reef Animals”, *Marine Environmental Research*, v. 42, n. 1-4, 359-362.
- PEREIRA NETTO, A. D.; MOREIRA, J. C.; DIAS, A. E. X. O.; ARBILLA, G.; FERREIRA, L. F. V.; OLIVEIRA, A. S.; BAREK, J., “Avaliação da contaminação Humana por Hidrocarbonetos Policíclicos Aromáticos (HPAS) e seus Derivados Nitratos (NHPAS): uma Revisão Metodológica”, *Química Nova*, vol. 23, n. 6, 2000, 765-773.
- PERRY, R. H., *Perry's Chemicals Engineer's Handbook*, 5th. Ed., McGraw-Hill Inc., 1984.
- PIROUETTE Multivariate Data analysis form IBM PC Systems, Versão 2.02, Infometrix, Seattle, WA, 1990-1996.
- PRITCHARD, P. H. & COSTA, C. F., “EPA's Alaska Oil Spill Bioremediation Project, Part 5”, *Env. Scie. Technol.*, v. 25, n. 3, 1991, 372-379.

- RAMOS, E. U.; VAES, W. H. J.; VERHAAR, H. J. M.; M.HERMES, J. L.,
 “Quantitative Structure-Activity Relationships for the Aquatic Toxicity of Polar and NonPolar Narcotic Pollutants”, *J. Chem. Inf. Comp. Sci.*, 38, 1998, 845-852.
- RANDIC, M., “Novel Molecular Descriptors for Structure-Property Studies”, *Chemical Physics Letters*, v. 211 n. 4,5, 1993, 478-482.
- RANDIC, M., “On Characterization of Molecular Branching”, *J. Am. Chem. Soc.*, v. 97, n. 23, 1975, 6609-6615.
- ROHRBACK, B. G., “Computer-assisted Rating of Gasoline Octane”, *Trends in Anal. Chem.*, vol. 10, no. 9, 1991.
- SALAU, J. S. I.; TAULER, R.; BAYONA, J. M.; TOLOSA, I., “Input Characterization of Sediments Organic Contaminants and Molecular Markers in the Northwestern Mediterranean Sea by Exploratory Data Analysis”, *Environ. Scie. Technol.*, v. 31, 1997, 3482-3490.
- SANDER, L. C. & WISE, S. A., *Advances in Chromatography*, vol. 25, 1986, 139-219.
- SCHAEFFER, FAIC, D. J. & LARSON, R. A., “Phototoxicology”, *The Chemist*, July/August, 1999, 18-24
- SCHNOOR, J., “The Alaska oil Spill: Its Effects and Lessons”, *Environ. Scie. Technol.*, v. 25, n. 1, 1991, 14.
- SHARAF, M. A.; ILLMAN, D. L.; KOWALSKI, B. R., *Chemometrics*, John Wiley & Sons, 1986.
- SHORT, J. W.; KVENVOLDEN, K. A.; CARLSON, P. R.; HOSTETTLER, F. D.; ROSENBAUER, R. J.; WRIGHT, B. A., “Natural hydrocarbon Background in Benthic Sediments of Prince William Sounds, Alaska: Oil vs. Coal”, *Environ. Scie. Technol.* v.33, 1999, 34-42.
- SHORT, J. W.; HEINTZ, R. A., “Identification of Exxon Valdez Oil in Sediments and Tissues from Prince William sounds and the Northwestern Gulf of Alaska Based on a PAH Weathering Model”, *Environ. Scie. Technol.* 31, 1997, 2375-2384.
- SHRIVER, D. F.; ATKINS, P. W.; LANGFORD, C. H., *Inorganic Chemistry*, Oxford University Press, 2nd. Ed., 1994.
- SITE da ChemFinder.Com – Database and Internet Searching, consultado pela última vez em 04/012001: <http://www.chemfinder.com>

- SITE oficial sobre o derramamento de de óleo, organizado pelo *Exxon Valdez Oil Spill Trustee Council*: <http://www.oilspill.state.ak.us>, consultado em 24/10/2000.
- SITE da Petrobrás, consultado em 13/11/2000: <http://www.petrobras.com.br/a-crono.htm>
- SPILL: The Wreck of Exxon Valdez Final Report, Report of the Alaska Oil Spill Commission: February 1990, 5-14, retirado do site oficial do derramamento (<http://www.oilspill.state.ak.us>, consultado em 24/10/2000)
- TAYLOR, S. & COX, B. J., Anxiety Sensitivity: Multiple Dimensions and Hierarchical Structure, *Behavior Research and Therapy*, n. 36, 1998, 37-51.
- THOMAS, E. V., “A Primer on Multivariate Calibration”, *Anal. Chemistry*, v. 66, n.15, 1994, 795A– 803A.
- TODESCHINI, R; LASAGNI, M.; MARENGO, E, “New Molecular Descriptors for 2D and 3D Structures. Theory”, *Journal of Chemometrics*, vol. 8, 1994, 263-272.
- TODESCHINI, R; GRAMATICA, P.; PROVENZANI, R.; MARENGO, E, “Weighted Holistic Invariant Molecular Descriptors. Part 2”, *Chemometrics and Intelligent Laboratory Systems*, vol. 27, 1995, 221-229.
- TODESCHINI, R; VIGHI, M.; PROVENZANI, R.; FINIZIO, A.; GRAMATICA, P., “Modelling and Prediction by Using WHIM Descriptors in QSAR Studies: Toxicity of Heterogeneous Chemicals on *Daphnia magna*”, *Chemosphere*, vol. 32, no. 8, 1996, 1527-1545.
- TODESCHINI, R; BETTION, C.; GIURIN, G.; GRAMATICA, P.; MIANA, P.; ARGEES, E., “Modelling and Prediction by Using WHIM Descriptors in QSAR Studies: submitochondrial particles (SMP) as toxicity biosensors of chlorophenols.”, *Chemosphere*, vol. 33, no. 1, 1996, 71-79.
- TODESCHINI, & GRAMATICA, P., “3D-Modelling and Prediction by Using WHIM Descriptors. Part 5.”, *Quant. Struct.-Act. Relat.*, vol. 16, 1997, 113-119.
- TOLOSA, I.; BAYONA, J.; ALBAIGÉS, J., “Spatial and Temporal Distribution, Fluxes, and Bugets of Organochlorinated Compounds in Northwest Mediterranean Sediments”, *Environ. Sci. Technol.*, 29, 2519-2527.
- ULLMAN’S Encyclopedia of Industrial Chemistry, volume B7, 5th. Ed, 1989.
- VERHAAR, H. J. M.; RAMOS, E. U.; HERMENS, J. L., “Classifying Environmental Pollutants. 2: Separation of Class 1(Baseline Toxicity) and Class 2 (‘Polar

- Narcosis') Type Compounds Based on Chemical Descriptors", *Journal of Chemometrics*, v. 10, 1996, 149-162.
- VERSCHUEREN, K., *Handbook of Environmental Data on Organic Chemicals*, 3rd. Ed., Van Nostrand Reinhold, 1996.
- WANG, Z.; FINGAS, M.; PAGE, D. S., "Oil Spill Identification", *Journal of Chromatography A*, 843, 369-411, 1999.
- WATERBEEMD, H. van de, *Chemometric Methods in Molecular Design*, vol. 2, VCH, 1995.
- WEBER, I. C.; DAVIS, C. P.; GREESON, D. M., "Phytophotodermatitis: The Other "Lime" Disease", *The Journal of Emergency Medicine*, vol. 17, no. 2, 1999, 235-237.
- WEHENS, R.; GELDER, R.; KEMPERMAN, G. J.; ZWANENBURG, B.; BUYDENS, L. M. C., "Molecular Challenges in Modern Chemometrics", *Anal. Chim. Acta*, 400, 1999, 413-424.
- WELDSCHOLTE, C. M.; KROONENBERG, P. M.; ANTONIDES, G., Three-mode Analysis of Perceptions of Economic Activities in Eastern and Western Europe, *Journal of Economic Psychology*, n.19, 1998, 321-351.
- WEST, R. C. & ASTLE, M. J., *Handbook of Data on Organic Compounds*, vol. 1 e 2, CRC Press, Inc., Flórida, 4th. Ed., 1987.
- WESTERMEYER, W. E., "Oil Spill Response Capabilities in the United States, Part 3", *Env. Scie. Technol.*, v. 25, n. 2, 1991, 196-200.
- WIENER, H., "Structure Determination of Paraffin Boiling Point", *Journal of the American Chemical Society*, vol. 69, 1947, 17-20.
- WOLD, S.; ESBENSEN, K.; GELADI, P., "Principal Component Analysis", *Chemometrics and Intelligent Systems*, 2, 1987, 37-52.
- YAP, C. T., "Multi-variate Analysis of Trace Elements from XRF Studies for Classification According to Origin", *Applied Spectroscopy*, v.46, n.5, 1992, 843-847.
- ZHANG, C.; SELINUS, O., "Statistics and GIS in Environmental Geochemistry – Some Problems and Solutions", *Journal of Geochemical Exploration*, n. 64, 1998, 339-354.

- ZHAO, Y. H.; JI, G. D.; CRONIN, M. T. D.; DEARDEN, J. C., "QSAR Study of the Toxicity of Benzoic Acids to *Vibrio Fischeri*, *Daphnia magna* and Carp", *The Science of the Total environment*, 216, 1998, 205-215.
- ZHOU, Z. & PARR, R. G., "New Measures of Aromaticity: Absolute Hardness and Relative Hardness", *J. Am. Chem. Soc.*, 111, 1989, 7371-7379.