

**Universidade de Évora**  
**Área Departamental de Ciências Exactas**  
**Departamento de Química**

**Especificação e Prototipação de  
Sistemas de Gestão e Controlo da  
Qualidade da Água de Albufeiras**

por

**Henrique Agostinho de Oliveira Moiteiro Vicente**

Tese submetida à Universidade de Évora para a obtenção do grau de Doutor em  
Química e elaborada sob a orientação do Doutor José Maria dos Santos Arteiro e do  
Professor Doutor José Carlos Ferreira Maia Neves

Évora

2004

# Especificação e Prototipação de Sistemas de Gestão e Controlo da Qualidade da Água de Albufeiras

Henrique Agostinho de Oliveira Moiteiro Vicente

Departamento de Química

Rua Romão Ramalho, 59

7000-671 Évora

Portugal

hvicente@uevora.pt



152226

Tese submetida à Universidade de Évora para a obtenção do grau de Doutor em  
Química e elaborada sob a orientação do Doutor José Maria dos Santos Arteiro  
e do Professor Doutor José Carlos Ferreira Maia Neves

## Agradecimentos

Deveria enumerar aqui uma lista infindável de pessoas que muito me ajudaram a atingir este objectivo. Seria impossível nomeá-las todas, como gostaria, pois correria sempre o risco de esquecer alguém. No entanto, não posso deixar de fazer referência a algumas delas, agradecendo em particular ao meu orientador, Doutor José Maria dos Santos Arteiro, com o qual tenho o privilégio de trabalhar desde 1993 e que acompanhou o meu trabalho de forma empenhada, ajudando-me, encorajando-me e inculcando-me o espírito de objectividade e clareza necessário a qualquer trabalho académico. Ao Professor Doutor José Carlos Ferreira Maia Neves, do Departamento de Informática da Universidade do Minho, meu co-orientador, a quem devo um agradecimento muito especial pelos ensinamentos que me proporcionou, pela forma inexcelável como me recebeu e me apoiou, contribuindo decisivamente para que esta dissertação fosse uma realidade. Não posso esquecer o Doutor Manuel Filipe Santos, o Doutor Paulo Cortez, o Mestre Hélder Quintela do Departamento de Sistemas de Informação da Universidade do Minho e o Mestre Rui Tavares do Departamento de Informática da Universidade de Évora pela colaboração, pelo interesse e pela disponibilidade sempre manifestada.

Expresso ainda ao Doutor António Manuel Neto Vaz, Presidente do Conselho do Departamento de Química, o meu agradecimento pela disponibilidade, amizade, apoio e estímulo, bem como pela ajuda na resolução de problemas de natureza burocrática. À Doutora Margarida Figueiredo do Departamento de Química, expresso também o meu agradecimento pelo apoio, encorajamento, interesse, estímulo e amizade. Aos meus colegas do Departamento de Química dirijo um abraço de amizade pelo interesse demonstrado.

Não posso esquecer, e agradecer também a todos os meus amigos e à minha família todo o carinho, apoio, compreensão e amizade.

Por último, quero prestar aqui, a título póstumo, uma sentida homenagem ao Professor Doutor César Augusto Nunes Viana de quem fui aluno e com quem tive o privilégio de trabalhar desde a minha vinda para a Universidade de Évora e que influenciou decisivamente todo o meu percurso académico.



## Resumo

É possível re-equacionar os problemas subjacentes à temática da modelação da qualidade da água em albufeiras através de metodologias de resolução de problemas que emanam da Área Científica da Inteligência Artificial, assim como de ferramentas usadas na procura de soluções como as Árvores de Decisão, a Indução de Regras, as Redes Neurais Artificiais e o Raciocínio Simbólico.

O presente trabalho descreve modelos para a previsão da qualidade da água em albufeiras. Um destes modelos fundamenta-se em técnicas de Descoberta de Conhecimento em Bases de Dados, enquanto o outro se baseia no Raciocínio Simbólico. Os sistemas sub-simbólicos apresentam-se como complementares dos sistemas simbólicos, permitindo a partir destes, elaborar sobre sistemas de previsão e/ou simulação do estado futuro da qualidade das águas em albufeiras.

Verificou-se que as fontes de informação na área do ambiente, particularmente no que concerne à qualidade da água em albufeiras, são variadas, heterogéneas e complexas, encontrando-se dispersas por diversas entidades, organismos e instituições. Esta dispersão não só torna difícil definir e caracterizar os problemas como impede a tomada das melhores decisões para os solucionar em tempo útil. Uma solução para este tipo de senãos passa pela integração, difusão e armazenamento desta informação diversificada num corpo único de forma a permitir a partilha de conhecimento com quem efectivamente dele necessita.

Neste trabalho é apresentada uma Agência para a Integração, Difusão e Arquivo de Informação sobre a **QUALIDADE da Água de Albufeiras - AQUA**. A agência **AQUA** disponibiliza trabalhadores electrónicos inteligentes para tratar uma multitude de tarefas, tais como a comunicação entre os diferentes sub-sistemas, o enviar e receber informação (e.g. resultados das análises laboratoriais, resultados fornecidos pelos sistemas de aquisição automática de dados, informação relevante sobre as bacias, tais como cartas, levantamentos topográficos, modelos digitais do terreno, fontes de poluição pontual, utilização do solo, dados meteorológicos e climáticos, caracterização geomorfológica), o gerir, guardar e respon-

der a pedidos de informação. A agência AQUA é, na sua essência, uma *Data Warehousing* associada a Sistemas MultiAgente. O objectivo principal passa por integrar, difundir, e arquivar grandes volumes de informação das mais diversas proveniências por forma a facilitar o rápido acesso à informação. A agência AQUA constitui um exemplo de aplicação de técnicas inovadoras da Área Científica da Inteligência Artificial a domínios do interesse público, neste caso à gestão e controlo dos recursos hídricos.

# Abstract

New alternative approaches for modeling Water Quality Control in reservoirs emerged from the Artificial Intelligence arena.

The present work reports models for the water quality forecasting. One of them uses Symbolic Approaches and the other uses Knowledge Discovery Techniques in Databases. The Symbolic Approaches so far developed appear as the most adequate for the resolution of this type of problems because they work with any and every information that outcomes from the reservoir. This model does not depend on a process of simplification and elimination of data like the Knowledge Discovery Techniques.

Information sources in water quality control of reservoirs are distributed, heterogeneous, large and complex. The solution is to integrate, diffuse and archive this information under a dynamic framework, in order to share its knowledge with every information system that needs it. In this thesis, it is presented **AQUA** - Agency for Integration, Diffusion and Archive of Information about Water **QUALITY** in Reservoirs, which is an agency that supplies intelligent electronic workers called reactive agents, in charge of some tasks, such as communication, sending and receiving information managing and saving the information and answering to information requests. **AQUA** is, in its essence, a data warehouse to which is associated a multitude of Multi-Agent Systems. The main goals are to integrate, diffuse and archive large sets of information from different sources (laboratories, services, units, technical equipments) and to provide tools in order to implement communication with humans based on web based services. It is a promising example of the application of innovative Artificial Intelligence techniques in areas of the public domain, like the area of Water Quality Control.





# Conteúdo

Agradecimentos .....	<i>iii</i>
Resumo.....	<i>v</i>
Abstract.....	<i>vi</i>
Lista de Figuras .....	<i>xiii</i>
Lista de Tabelas.....	<i>xvii</i>
Notação.....	<i>xix</i>
1 Introdução .....	1
1.1 Caracterização do Local de Trabalho .....	2
1.1.1 Caracterização Geral da Zona Envolvente .....	4
1.1.2 Caracterização das Albufeiras Estudadas .....	5
1.1.2.1 Albufeira do Divôr.....	6
1.1.2.2 Albufeira de Monte Novo.....	9
1.1.2.3 Albufeira da Vigia.....	12
1.2 O Processo de Tomada de Decisão.....	15
1.3 A Gestão do Conhecimento .....	19
1.4 A Gestão do Conhecimento e a Inteligência Artificial.....	20
1.5 Inovação e Valor Acrescentado do Trabalho.....	21
1.6 Estrutura do Trabalho.....	24
1.7 Conclusões.....	25
2 Modelos Dinâmicos de Ecossistemas Aquáticos de Uso Genérico .....	27
2.1 Características dos Ecossistemas .....	33
2.2 Modelos de Estrutura Dinâmica.....	40
2.2.1 Aplicação a Ecossistemas Aquáticos.....	43

2.3 Modelos Baseados na Teoria do Caos.....	45
2.4 Modelos Baseados na Teoria da Catástrofe.....	46
2.5 Outras Técnicas de Modelação.....	50
2.5.1 Modelos Recorrendo à Programação Orientada ao Objecto.....	50
2.5.2 Modelos de Base Individual.....	52
2.6 Caracterização do Estado Trófico das Albufeiras Estudadas.....	53
2.7 Caracterização da Qualidade da Água das Albufeiras Estudadas.....	54
2.8 Conclusões.....	60
<b>3 Agentes, Sistemas MultiAgente e Medida da Qualidade de Informação em Bases de</b>	
<b>Dados/Conhecimento.....</b>	<b>63</b>
3.1 Agentes.....	64
3.2 Programação em Lógica Estendida.....	65
3.3 Propriedades dos Agentes.....	71
3.4 Sistemas MultiAgente.....	76
3.5 A Agência AQUA.....	78
3.6 Conclusões.....	81
<b>4 Redes Neurais Artificiais e Descoberta de Conhecimento em Bases de</b>	
<b>Dados/Conhecimento.....</b>	<b>83</b>
4.1 Fundamentos Biológicos.....	85
4.2 Perspectiva Histórica.....	88
4.3 Benefícios das Redes Neurais Artificiais.....	89
4.4 A Inteligência Artificial e as Redes Neurais Artificiais.....	90
4.5 Arquitecturas de Rede.....	90
4.6 Aprendizagem.....	92
4.7 Neurónio Artificial ou Nodo.....	94
4.8 A Arquitectura Unidireccional.....	97
4.9 Capacidades e Limitações.....	99
4.10 Descoberta de Conhecimento e Bases de Dados.....	100
4.11 Data Mining.....	101
4.11.1 Definição de Data Mining.....	102
4.11.2 Áreas Relacionadas com o Data Mining.....	104
4.11.3 Abordagens do Data Mining.....	106

4.11.4 Objectivos do Data Mining.....	106
4.11.4.1 Classificação.....	106
4.11.4.2 Previsão.....	107
4.11.4.3 Regressão Linear.....	108
4.11.4.4 Segmentação.....	108
4.11.4.5 Associação ou Dependência.....	109
4.11.4.6 Sumariação.....	110
4.11.4.7 Visualização.....	110
4.11.4.8 Detecção de Desvios.....	110
4.12 Descoberta de Conhecimento em Base de Dados.....	110
4.12.1 Selecção.....	111
4.12.2 Pré-Processamento.....	112
4.12.3 Transformação.....	112
4.12.4 Data Mining.....	112
4.12.5 Interpretação e Avaliação.....	113
4.13 Problemas Inerentes ao Processo de DCBD.....	113
4.14 Modelos e Técnicas Utilizadas no Processo de Data Mining.....	115
4.15 Conclusões.....	115
5 Uma Agência para a Integração, Difusão e Arquivo de Informação sobre a Qualidade da Água de Albufeiras.....	117
5.1 Resolução de Problemas em Ambiente Distribuído.....	125
5.2 O Modelo AQUA.....	126
5.3 AQUA - Modelação Simbólica.....	126
5.4 Estrutura Global do Sistema.....	128
5.5 Data Warehousing e Melhores Decisões.....	128
5.6 AQUA - Modelação Sub-Simbólica.....	140
5.7 Análise de Séries Temporais.....	141
5.8 Redes Neurais Artificiais.....	144
5.9 Generalização.....	146
5.10 Previsão da Qualidade da Água em Albufeiras Utilizando RNAs.....	150
5.11 Árvores de Decisão e Indução de Regras.....	152
5.12 A Arquitectura do Sistema AQUA.....	160
5.13 Os Agentes do Sistema AQUA.....	162

5.14 A Estrutura Lógica que Suporta a Agência AQUA.....	165
5.14.1 O Agente Monitor .....	165
5.14.2 Agentes Proxy.....	166
5.14.3 Os Agentes de Decisão .....	167
5.14.4 Os Agentes de Computação.....	167
5.14.5 Os Agentes de Recursos .....	168
5.14.6 Os Agentes de Interação e Explicação .....	169
5.14.7 O Ambiente qenv .....	170
5.15 Conclusões.....	171
6 Conclusões Finais e Perspectivas Para o Futuro .....	173
Bibliografia.....	175
Apêndice A - Conjunto de Regras Gerado a Partir dos Dados de Qualidade da Água para as Albufeiras do Divôr, Monte Novo e Vigia - Modelo 1 .....	193
Apêndice B - Conjunto de Regras Gerado a Partir de um Grupo Restrito de Parâmetros de Qualidade da Água para as Albufeiras do Divôr, Monte Novo e Vigia - Modelo 2.....	199

# Lista de Figuras

Figura 1.1	Localização das albufeiras do Divôr, de Monte Novo e da Vigia.....	4
Figura 1.2	Vista parcial da albufeira do Divôr.....	6
Figura 1.3	Localização da bacia hidrográfica da albufeira do Divôr.....	7
Figura 1.4	Usos do solo na área da bacia hidrográfica da albufeira do Divôr.....	7
Figura 1.5	Fontes de poluição pontual na bacia hidrográfica da albufeira do Divôr.....	8
Figura 1.6	Vista parcial da barragem de Monte Novo.....	9
Figura 1.7	Localização da bacia hidrográfica da albufeira de Monte Novo.....	10
Figura 1.8	Usos do solo na área da bacia hidrográfica da albufeira de Monte Novo.....	10
Figura 1.9	Fontes de poluição pontual na bacia hidrográfica da albufeira de Monte Novo.....	11
Figura 1.10	Vista parcial da albufeira da Vigia.....	12
Figura 1.11	Localização da bacia hidrográfica da albufeira da Vigia.....	13
Figura 1.12	Usos do solo na área da bacia hidrográfica da albufeira da Vigia.....	13
Figura 1.13	Fontes de poluição pontual na bacia hidrográfica da albufeira da Vigia.....	14
Figura 2.1	Esquema do desenvolvimento da modelação ambiental.....	29
Figura 2.2	Influência dos factores externos sobre a composição do ecossistema.....	39
Figura 2.3	Esquema do procedimento utilizado no desenvolvimento de modelos de estrutura dinâmica.....	43
Figura 2.4	$-dx/dt$ em função de $x$ para seis valores de $a$ .....	47
Figura 2.5	Representação de $x$ em função de $a$ .....	48
Figura 2.6	Capacidade ecológica tampão em função do tempo.....	49
Figura 2.7	Ligações entre as classes num modelo de migração animal.....	52
Figura 3.1	Agente a interactuar com o meio.....	65
Figura 3.2	A Agência AQUA.....	80
Figura 3.3	Agentes da plataforma AQUA.....	80

Figura 4.1	Representação esquemática de um neurónio biológico.....	86
Figura 4.2	Representação esquemática de uma sinapse.....	86
Figura 4.3	Arquitectura de uma RNU.....	91
Figura 4.4	Arquitectura de uma RNR.....	92
Figura 4.5	Estrutura geral de um nodo.....	95
Figura 4.6	Alguns exemplos de funções de activação.....	96
Figura 4.7	A função sigmoid para três valores $k$ .....	97
Figura 4.8	Estrutura de uma RNU com a topologia 2-3-1-2.....	98
Figura 4.9	Exemplo de Classificação.....	107
Figura 4.10	Exemplo de Regressão Linear.....	108
Figura 4.11	Exemplo de Segmentação.....	109
Figura 4.12	Processo de DCBD.....	111
Figura 5.1	A arquitectura da agência AQUA.....	125
Figura 5.2	Os agentes da agência AQUA.....	126
Figura 5.3	Singularidade.....	133
Figura 5.4	Um Espaço Hiperdimensional em que pontifica a qualidade da informação que é intrínseca ao Programa 5.3.....	134
Figura 5.5	Um Espaço Hiperdimensional em que pontifica a qualidade da informação respeitante ao mês de Janeiro.....	135
Figura 5.6	Um Espaço Hiperdimensional em que pontifica a qualidade da informação respeitante ao mês de Fevereiro.....	136
Figura 5.7	Um Espaço Hiperdimensional em que pontifica a qualidade da informação respeitante ao mês de Março.....	137
Figura 5.8	Um Espaço Hiperdimensional em que pontifica a qualidade da informação respeitante ao mês de Abril.....	138
Figura 5.9	Um Espaço Hiperdimensional em que pontifica a qualidade da informação.....	139
Figura 5.10	Um Espaço Hiperdimensional que denota uma situação óptima no que concerne aos termos em que se deve operar no que respeita à gestão e controlo da qualidade da água.....	139
Figura 5.11	Um Espaço Hiperdimensional que denota uma situação de total descoor- denação no que respeita à gestão e controlo da qualidade da água.....	140
Figura 5.12	Série temporal não estacionária.....	142
Figura 5.13	Série temporal com um período de sazonalidade de quatro.....	143

---

Figura 5.14 Autocorrelações típicas de séries temporais.....	143
Figura 5.15 Uma RNU completamente interligada, com dois nodos de entrada, dois nodos intermédios, um nodo de saída, conexões de bias e atalhos.....	145
Figura 5.16 Deslocamentos de uma série temporal. ....	146
Figura 5.17 Generalização e sobre-ajustamento.....	147
Figura 5.18 Topologia para uma possível RNA a utilizar na previsão da qualidade da informação sobre a qualidade da água. ....	152
Figura 5.19 Árvore de decisão gerada a partir dos dados de qualidade da água para as albufeiras do Divôr, Monte Novo e Vigia.....	158
Figura 5.20 O comportamento dos agentes e/ou SMA na agência AQUA.....	164





# Lista de Tabelas

Tabela 2.1	Esforço de modelação para alguns ecossistemas.....	31
Tabela 2.2	Esforço de modelação para alguns problemas ambientais.....	32
Tabela 2.3	Hierarquia dos mecanismos de regulação.....	34
Tabela 2.4	Algumas das funções objectivo propostas. ....	41
Tabela 2.5	Comparação entre as previsões do modelo e os resultados experimentais referentes ao lago Søbygaard.....	44
Tabela 2.6	Critérios utilizados para obtenção do estado trófico de uma albufeira.....	54
Tabela 2.7	Estado trófico das três albufeiras em estudo nos últimos anos hidrológicos. ...	54
Tabela 2.8	Qualidade das águas superficiais para produção de água para consumo humano.....	55
Tabela 2.9	Classificação dos parâmetros de qualidade das águas superficiais em grupos...57	
Tabela 2.10	Frequência mínima de amostragem e de análise das águas superficiais. ....	58
Tabela 2.11	Usos permitidos por classe de qualidade da água.....	58
Tabela 2.12	Grelha proposta pelo INAG para a classificação qualitativa das águas superficiais.....	59
Tabela 2.13	Classificação qualitativa das águas do Divôr, Monte Novo e Vigia.....	60
Tabela 3.1	Distribuição do serviço docente no departamento. ....	67
Tabela 3.2	Disciplinas atribuídas a um dado professor, em que há um nulo do tipo desconhecido. ....	68
Tabela 3.3	Disciplinas atribuídas a um professor, em que pontua informação do tipo disjuntivo. ....	69
Tabela 3.4	Informação acerca dos funcionários de uma empresa em que há um nulo do tipo não permitido.....	70
Tabela 4.1	Algumas funções de activação.....	96
Tabela 4.2	Períodos que precedem o Data Mining. ....	102

---

Tabela 5.1	Programa e sua expressão em termos do operador qualidade para o mês de Janeiro .....	135
Tabela 5.2	Programa e sua expressão em termos do operador qualidade para o mês de Fevereiro.....	136
Tabela 5.3	Programa e sua expressão em termos do operador qualidade para o mês de Março. ....	137
Tabela 5.4	Programa e sua expressão em termos do operador qualidade para o mês de Abril.....	138
Tabela 5.5	Conjunto de casos de treino/teste para RNA.....	150
Tabela 5.6	Matriz de confusão do modelo 1. ....	157
Tabela 5.7	Matriz de confusão obtida para o modelo 2. ....	159
Tabela 5.8	Avaliação dos dois modelos obtidos. ....	159
Tabela 5.9	Descrição dos eventos do agente monitor. ....	166
Tabela 5.10	Descrição dos eventos dos agentes Proxy.....	166
Tabela 5.11	Descrição dos eventos dos agentes de decisão. ....	167
Tabela 5.12	Descrição dos eventos dos agentes de computação. ....	168
Tabela 5.13	Descrição dos eventos do agente de recursos.....	169
Tabela 5.14	Construção da operação para a definição do local de armazenamento. ....	169
Tabela 5.15	Descrição dos eventos dos agentes de interação e explicação.....	170

# Notação

## Acrónimos

**AA** - Aprendizagem Automática.

**AG** - Algoritmo Genético.

**AIC** - Critério de Informação de Akaike.

**AQUA** - Agência para a integração difusão e arquivo de informação sobre a **Q**ualidade da  
Água de albufeiras.

**BC** - Bases de Conhecimento.

**BD** - Bases de Dados.

**BIC** - Critério de Informação de Bayes.

**CBO** - Carência Bioquímica de Oxigénio.

**CQO** - Carência Química de Oxigénio.

**CS** - Cálculo Simbólico.

**DCBD** - Descoberta de Conhecimento em Bases de Dados.

**DW** - Data Warehousing.

**IA** - Inteligência Artificial.

**IAD** - Inteligência Artificial Distribuída.

**INAG** - INstituto da ÁGua.

**JTD** - Janela Temporal Deslizante.

**LAB** - Laboratórios de Análises Biológicas.

**LAFQ** - Laboratórios de Análises Físico-Químicas.

**MNQE** - Média Normalizada do Quadrado dos Erros.

**MQE** - Média do Quadrado dos Erros.

**NPA** - Nível de Pleno Armazenamento.

**PLC** - Programação em Lógica Contextual.

**PLE** - Programação em Lógica Estendida.

**PST** - Previsão de Séries Temporais.

**QN** - Quadro Negro

**RBF** - *Radial Basis-Functions*.

**RMQE** - Raiz da Média do Quadrado dos Erros.

**RNA** - Rede Neuronal Artificial.

**RNR** - Rede Neuronal Recorrente.

**RNU** - Rede Neuronal Unidireccional.

**RS** - Raciocínio Simbólico.

**SADAB** - Sistema de Arquivo e Difusão de Análises Biológicas.

**SAT** - Sistema de Apoio ao Técnico.

**SADAFQ** - Sistema de Arquivo e Difusão de Análises Físico-Químicas.

**SADDM** - Sistema de Arquivo e Difusão de Dados Meteorológicos.

**SMA** - Sistema MultiAgente.

**SPD** - Sistema de Processamento Distribuído.

**SQE** - Soma do Quadrado dos Erros.

**SRDM** - Serviço de Registo de Dados Meteorológicos.

---

## Símbolos Gerais e Abreviaturas

agc - agentes de computação.

agd - agentes de decisão.

agie - agentes de interacção e explicação.

agm - agente monitor.

agp - agentes *proxy*.

agr - agentes de recurso.

$c_x$  - concentração da espécie química  $x$ .

$C$  - conjunto de conexões pesadas unidireccionais.

$|C|$  - cardinalidade do conjunto  $C$ .

$D$  - domínio de  $f(x)$ .

$E$  - conjunto de nodos de entrada.

$eq$  - refere-se ao estado de equilíbrio químico.

$fa$  - função de activação.

$f(x)$  - função sobre a variável  $x$ .

$F$  - conjunto de funções de activação.

$I$  - conjunto de nodos intermédios.

$K_a$  - constante de rearejamento.

$\mathcal{L}$  - operador qualidade.

$n$  - número de entradas de um modelo.

$n_i$  - número de nodos intermédios.

$N$  - número de exemplos de treino.

$p$  - número de parâmetros livres de um modelo.

$R$  - constante dos gases ideais.

$s_i$  - valor de saída do nodo  $i$ .

$S$  - conjunto de nodos de saída.

$t$  - instante de tempo.

$T$  - temperatura.

$u_i$  - valor de integração do nodo  $i$ .

$w_{ij}$  - conexão entre o nodo  $j$  e o nodo  $i$ .

$x_t$  - série temporal.

$\hat{x}_t$  - modelo de série temporal.

# Capítulo 1

## Introdução

O problema da modelação da qualidade da água em albufeiras situa-se no âmbito das preocupações e necessidades de natureza ambiental que se têm vindo a afirmar como dos mais relevantes para um desenvolvimento sustentado e harmonioso das populações. Portugal não é excepção, constituindo as albufeiras um caso em que a resolução e prevenção de problemas relacionados com a qualidade da água se tem posto e continuará a pôr com grande acuidade [Rodrigues, 1992] [Neves et al., 1998].

A construção de barragens para armazenar água cria reservatórios que muitas vezes se constituem em receptores de efluentes domésticos, industriais e agrícolas, os quais não raras vezes provocam alterações nas características da água armazenada. Estes lagos artificiais têm como objectivo principal a produção de água potável para abastecimento de populações, rega e aproveitamento hidroeléctrico, pelo que é determinante conhecer e preservar a qualidade dessa água.

Entre os muitos senãos que aqui se podem referenciar, a eutrofização assume-se como um dos mais sérios problemas inerentes a muitas das albufeiras. Este processo é causado pelo aumento natural e/ou artificial da concentração de nutrientes que estas massas de água interiores sofrem. O crescimento anormal de algas, típico de águas eutrofizadas, implica a deterioração da qualidade da água e provoca a diminuição dos níveis de oxigénio dissolvido, com graves prejuízos para todo o ecossistema em geral, e para a fauna piscícola em particular. É por conseguinte necessário prevenir e actuar de uma forma pró-activa, pelo que o recurso a métodos de previsão, que têm subjacentes processos de tomada de decisão, oriundos da Área Científica da Inteligência Artificial, pode contribuir para solucionar ou amenizar o problema [Cortez, 2002] [Abelha, 2004].

Ora a previsão do comportamento das massas de água em resposta ao aumento dos nutrientes é difícil, dado que os processos físicos, químicos e biológicos envolvidos são altamente complexos. Como exemplo, tem-se que os modelos existentes, embora fornecendo boas estimativas da concentração de clorofila-*a*, não podem ser aplicados sem primeiramente passarem por um longo e moroso processo de calibração. Além disso, na maior parte dos casos, reduzem a solução de problemas extremamente complexos à simples resolução de sistemas equações lineares, ignorando factores que afectam decisivamente o processo de eutrofização (e.g. a natureza do ecossistema).

## 1.1 Caracterização do Local de Trabalho

A água tem sido um dos principais factores de desenvolvimento e fixação das comunidades humanas, extremamente dependentes da disponibilidade hídrica. Ao longo da História, as populações foram-se instalando, preferencialmente nas zonas costeiras e nas margens de rios e de lagos. Esta proximidade permitiu uma fácil utilização do recurso água, assim como permitiu a exploração dos recursos vivos suportados pelos ecossistemas aquáticos.

Ora, a explosão demográfica e o crescente nível de desenvolvimento tecnológico, ao apelarem a uma maior utilização da água, contribuíram para a sua crescente deterioração. De um modo geral, a principal causa de deterioração está, sem dúvida, relacionada com as entradas de nutrientes e de produtos químicos resultantes das práticas agrícolas. Contudo, em Portugal, a entrada directa na água de efluentes urbanos, industriais ou agrícolas sem tratamento ou com tratamento inadequado constituem outra importante causa de degradação da qualidade da água e, consequentemente, dos ecossistemas aquáticos.

A crescente intervenção humana nas linhas de água tem originado alterações morfológicas. Estas alterações têm efeitos sobre o funcionamento dos ecossistemas aquáticos, na medida em que diminuem a diversidade do *habitat* e alteram as características hidrológicas, repercutindo-se nos processos naturais de autodepuração, sobre a erosão e o transporte sólido.

No Sul de Portugal a grande maioria das alterações morfológicas resultam das actividades agro-pecuárias. Contudo, contrariamente ao que se passa na Europa Central e até mesmo na zona Norte de Portugal as alterações morfológicas são reduzidas devido à



baixa densidade populacional e ao regime extensivo das explorações agrícolas [Saraiva et al., 1994].

Em Portugal, os poucos lagos naturais de água doce existentes são de pequena dimensão, de natureza glaciária e situam-se na Serra da Estrela. Nos últimos trinta anos, contudo, passámos de uma situação onde praticamente não existiam lagos para cerca de duas centenas, de natureza artificial (i.e. albufeiras).

A maior parte dos lagos naturais de água doce situam-se em zonas próximas das cabeceiras, enquanto as albufeiras se situam, em geral, no curso médio ou final dos rios, apresentando áreas de drenagem alongadas que incluem, nalguns casos, os afluentes principais. Esta situação contribui decisivamente para que a diferença de idades entre um lago natural e uma albufeira seja substancial e significativa. Por exemplo, há que tomar-se nota de que os lagos marginais alpinos adquiriram a sua forma actual e perpetuaram-se ao longo de milhares de anos, enquanto apenas algumas albufeiras portuguesas poderão ultrapassar cinco décadas de existência, devido, sobretudo, à contaminação, à fertilização contínua e à acumulação de sedimentos no fundo [Margalef, 1994].

As alterações mais notórias nas características ecológicas das albufeiras são o aumento da produção de matéria orgânica no epilimnio, localizado acima da termoclina (i.e. a zona de gradiente térmico, onde a temperatura varia de forma brusca), e a depleção do oxigénio no hipolimnio, situado abaixo da termoclina, que corresponde à zona de decomposição e retorno dos elementos, na forma inorgânica, à massa de água. Quando uma grande quantidade de fitoplâncton está presente na albufeira pode conduzir a uma sobressaturação de oxigénio na camada superior devida à fotossíntese. No entanto esse aumento não tem correspondência nas camadas inferiores porque se escapa, por difusão, para a atmosfera. Assim, o aumento da matéria orgânica requer maior quantidade de oxigénio para a sua oxidação, conduzindo ao consumo do oxigénio do hipolimnio e à deposição da matéria orgânica não decomposta [Margalef, 1994].

Numa albufeira sujeita a entradas de matéria orgânica e de nutrientes este processo é agravado, levando a uma situação de anóxia que se estende por todo o hipolimnio. Este processo complementa o estado de eutrofia da massa de água e é acompanhado pelo desenvolvimento de organismos indesejáveis, produtores de toxinas que inibem, lesam ou matam invertebrados e vertebrados, acarretando problemas de saúde pública se a água se destinar ao abastecimento de populações [Margalef, 1994].

### 1.1.1 Caracterização Geral da Zona Envolvente

A albufeira é um sistema lântico, intimamente ligada à sua bacia de drenagem e ao sistema terrestre envolvente. Consequentemente, a localização geográfica, o tipo de clima e as características geológicas são determinantes nos processos físicos, químicos e biológicos que implicam trocas de energia e de matéria entre os meios aquático e terrestre.

O presente estudo recai sobre três albufeiras situadas no distrito de Évora, a distâncias compreendidas entre os 15 e os 30 km. Duas delas, Monte Novo e Vigia, fazem parte da bacia hidrográfica do rio Guadiana, enquanto a outra, Divôr, está incluída na bacia hidrográfica do rio Tejo. A albufeira de Monte Novo dista aproximadamente 10 km da albufeira da Vigia e cerca de 28 km da albufeira do Divôr, enquanto que as albufeiras do Divôr e da Vigia distam cerca de 30 km (Figura 1.1).

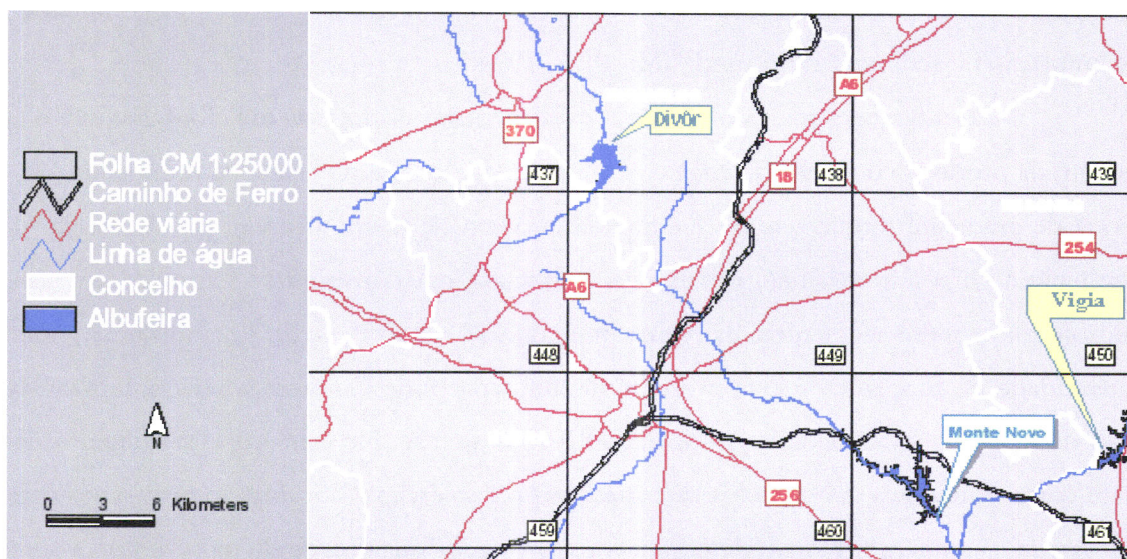


Figura 1.1: Localização das albufeiras do Divôr, de Monte Novo e da Vigia. Adaptada de Matoso [Matoso et al., 2004].

Estas albufeiras situam-se em zonas de peneplanícies com altitudes que não ultrapassam 515 m. O clima é mediterrânico, apresentando uma insolação média anual de 2800 a 3000 horas de Sol, com a temperatura média anual a situar-se entre os 15 °C e os 17,5 °C. O número de dias no ano com precipitação varia, em média, entre os 50 e os 100, registando-se nos meses com temperaturas menos elevadas (i.e. de Novembro a Abril). A precipitação média anual varia entre 500 e 1000 mm m<sup>-2</sup> em anos pluviosos, podendo não ultrapassar os 50 mm m<sup>-2</sup> em anos de seca [Cunha et al., 1980].

A unidade geológica fundamental da área em estudo é a zona de Ossa Morena, constituída maioritariamente por granitos de duas micas e rochas xisto-grauváquias. Do ponto de vista litológico predominam as rochas graníticas e xistosas recobertas por formações continentais bastante recentes (e.g. arenitos, areias e calcários).

### 1.1.2 Caracterização das Albufeiras Estudadas

De acordo com a classificação das albufeiras de águas públicas (decreto regulamentar n° 2/88 de 20 de Janeiro), as massas de água em estudo são consideradas como albufeiras protegidas. Albufeiras protegidas são aquelas cuja água é ou se prevê que venha a ser utilizada para abastecimento de populações e, aquelas cuja protecção é ditada por razões de defesa ecológica.

As zonas de protecção das albufeiras de águas públicas classificadas como protegidas terão a largura de 500 m, medida na horizontal, contada a partir da linha do Nível de Pleno Armazenamento (NPA).

De acordo com o decreto-lei n° 152/97 de 19 de Junho as albufeiras em estudo são consideradas zonas sensíveis. Uma determinada massa de água será identificada como zona sensível se pertencer a uma das seguintes categorias:

- (i) - Lagos naturais de água doce, outras extensões de água doce, estuários e águas costeiras que se revelem eutróficos ou susceptíveis de se tornarem eutróficos num futuro próximo, se não forem tomadas medidas de protecção;
- (ii) - Águas doces de superfície destinadas à captação de água potável, cujo teor em nitratos possa exceder a concentração estabelecida nas disposições pertinentes da Directiva n°75/440/CEE de 16 de Julho, relativa à qualidade das águas superficiais destinadas à produção de água potável, se não forem tomadas medidas de protecção; e
- (iii) - Zonas em que é necessário outro tratamento para além do previsto no artigo 5º, para cumprir o disposto nas directivas do Conselho.

### 1.1.2.1 Albufeira do Divôr

A albufeira do Divôr (Figura 1.2) localiza-se no concelho de Arraiolos, na ribeira com o mesmo nome, afluente do rio Sorraia, fazendo parte da bacia hidrográfica do rio Tejo. Iniciou o seu funcionamento em 1965 e destinou-se à produção de água para o abastecimento da cidade de Évora. Actualmente apenas uma pequena parcela é destinada à produção de água para consumo humano, servindo cerca de 1200 habitantes dos concelhos de Évora e Arraiolos, destinando-se fundamentalmente à irrigação. Quando o nível de pleno armazenamento é atingido (i.e. a cota 261,4 m), a área inundada ocupa 2,39 km<sup>2</sup>, apresenta a altura de 23 m e um comprimento de coroamento de 636 m. A sua capacidade máxima de armazenamento é 11,90 × 10<sup>6</sup> m<sup>3</sup>, sendo a capacidade útil de 11,89 × 10<sup>6</sup> m<sup>3</sup>. Esta albufeira situa-se numa zona bastante plana, apresentando uma profundidade máxima de 8 m e uma profundidade média relativamente baixa, cerca de 4,5 m [Matoso et al., 2004]. Nos anos de seca as necessidades de água provocam uma forte diminuição dos volumes armazenados e a descida da água põe a descoberto grande parte da área da albufeira.



Figura 1.2: Vista parcial da albufeira do Divôr.

A bacia hidrográfica apresenta declives suaves, entre 4% e 8%, ocupa 43 km<sup>2</sup> e situa-se nos concelhos de Évora e Arraiolos (Figura 1.3). As formações litológicas predominantes na bacia hidrográfica são os xistos e os grauvaques do Silúrico e do Ordovício.

Na área da bacia hidrográfica existem dois pequenos aglomerados urbanos. Os solos são, fundamentalmente, destinados às culturas arvenses de sequeiro e às explorações agro-silvo-pastoris e florestais, embora existam culturas arbóreas-arbustivas e culturas arvenses de regadio (Figura 1.4).

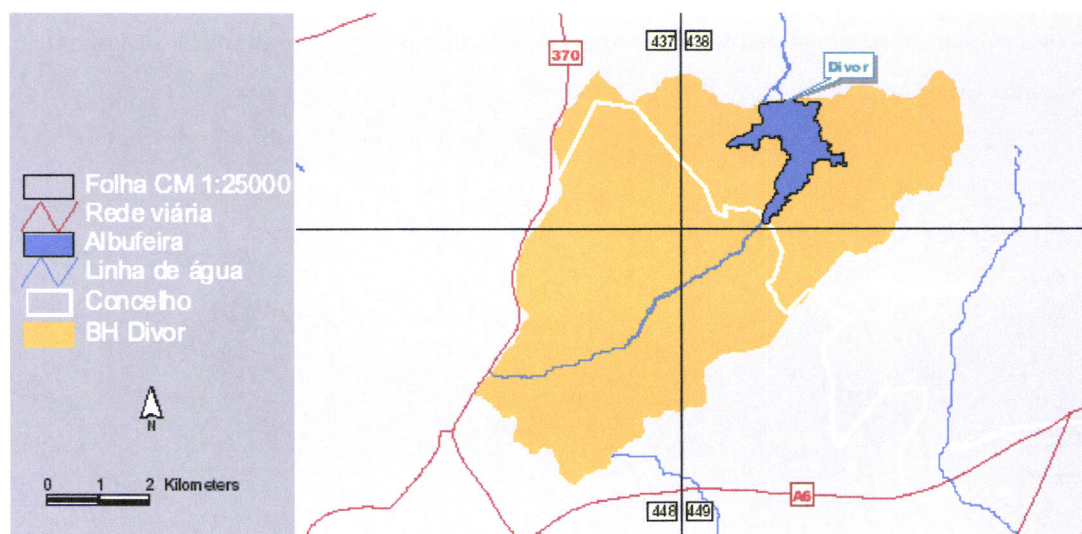


Figura 1.3: Localização da bacia hidrográfica da albufeira do Divôr. Adaptada de Matoso [Matoso et al., 2004].

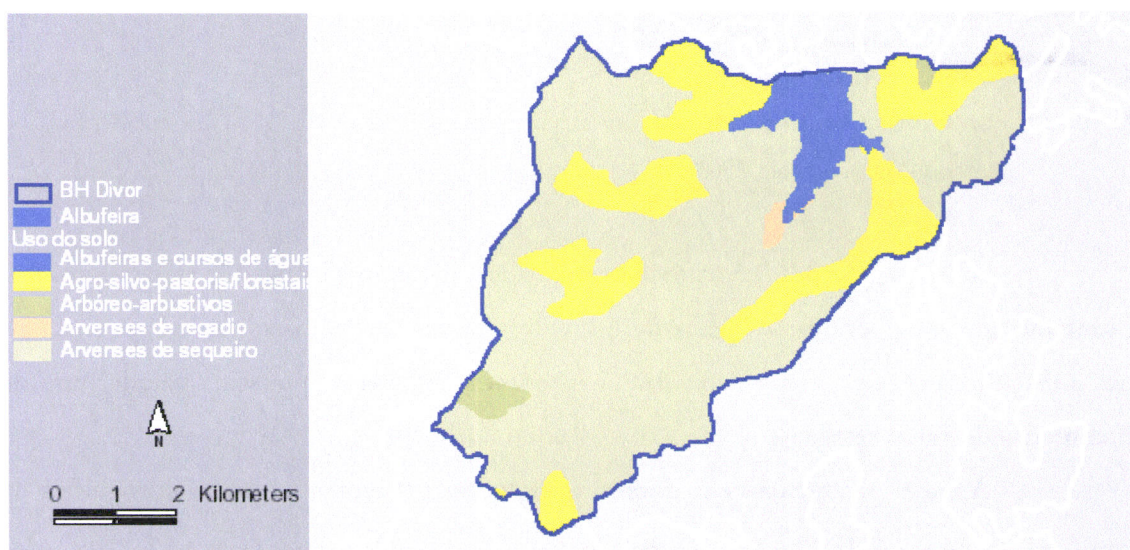


Figura 1.4: Usos do solo na área da bacia hidrográfica da albufeira do Divôr. Adaptada de Matoso [Matoso et al., 2004].

Na Figura 1.5 encontram-se localizadas as principais fontes de poluição pontual da bacia hidrográfica da albufeira. O levantamento apresentado diz respeito exclusivamente às fontes de poluição de carácter pontual, não incluindo as fontes de poluição difusa, importantes na região, atendendo aos usos do solo. As actividades pecuárias consideradas dizem respeito aos regimes intensivo e semi-intensivo. O regime extensivo é considerado uma for

ma de poluição difusa. As principais fontes de poluição pontual resultam da bovinicultura e das zonas urbanas [DRAOTA, 1999].

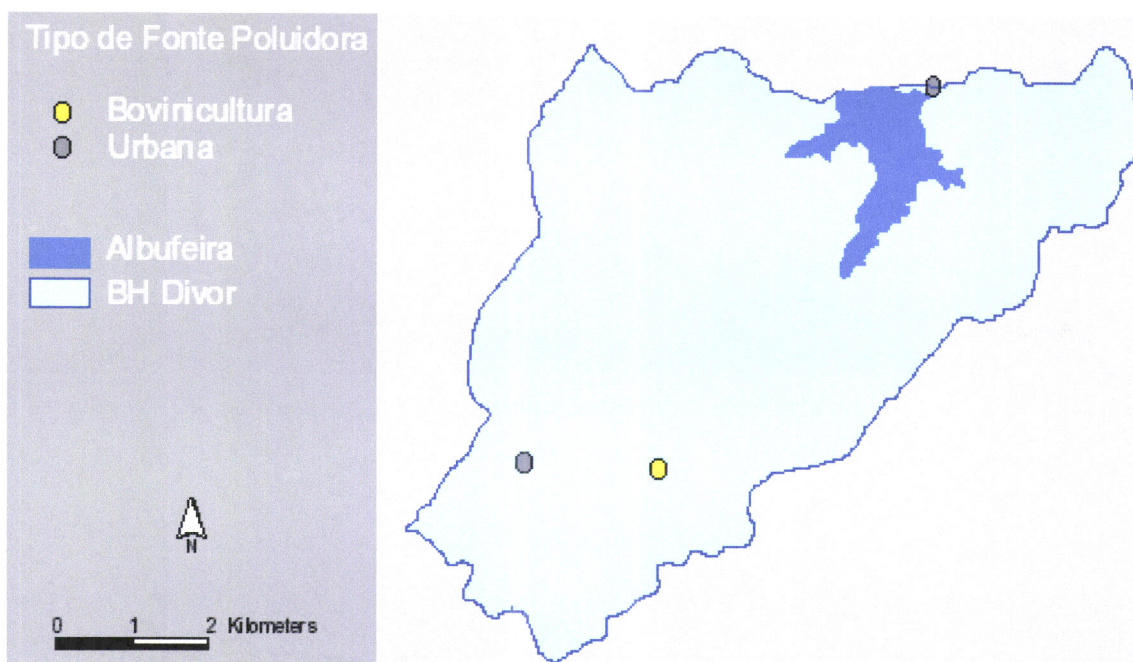


Figura 1.5: Fontes de poluição pontual na bacia hidrográfica da albufeira do Divôr. Adaptada de Matoso [Matoso et al., 2004].

As fontes de poluição urbana assinaladas possuem sistemas de tratamento, nomeadamente fossas sépticas e lagoas de estabilização que fazem baixar as cargas lançadas nas linhas de água (i.e. a carga rejeitada). A exploração bovina não possui qualquer tipo de tratamento de águas residuais.

A carga produzida pelas fontes de poluição existentes na bacia hidrográfica da albufeira do Divôr é de  $72,5 \text{ Kg}_{\text{CBO}_3} \text{ dia}^{-1}$ , o que corresponde ao equivalente a 1208 habitantes. A matéria orgânica produzida não é lançada, na sua totalidade, para as linhas de água, em virtude de existirem sistemas de tratamento. A carga rejeitada, que entra nas linhas de água, é estimada em  $53,2 \text{ Kg}_{\text{CBO}_3} \text{ dia}^{-1}$ , ou seja, o equivalente a 887 habitantes. A pequena diferença entre a carga produzida e a carga rejeitada prende-se com o facto da exploração bovina lançar directamente na linha de água e de as fossas sépticas não serem um sistema de tratamento muito eficiente [DRAOTA, 1999].

### 1.1.2.2 Albufeira de Monte Novo

A albufeira de Monte Novo localiza-se no concelho de Évora, sendo um aproveitamento do rio Degebe, afluente do rio Guadiana. Iniciou o seu funcionamento em 1982 e destina-se à produção de água para consumo humano, servindo cerca de 48000 habitantes dos concelhos de Évora e Arraiolos. Quando o nível de pleno armazenamento é atingido (i.e. a cota 196 m), a área inundada ocupa 2,77 km<sup>2</sup>, apresenta a altura de 30 m, um comprimento de coroamento de 160 m e uma profundidade média de 5,51 m sendo a profundidade máxima de 18 m. A capacidade máxima de armazenamento é  $15,28 \times 10^6$  m<sup>3</sup>, a que corresponde uma capacidade útil de  $14,78 \times 10^6$  m<sup>3</sup> [Matoso et al., 2004].



Figura 1.6: Vista parcial da barragem de Monte Novo.

A bacia hidrográfica chega a apresentar declives com inclinação compreendida entre 15 % e 25%. Ocupa 267 km<sup>2</sup>, estendendo-se aos concelhos de Évora e Arraiolos (Figura 1.7). As formações litológicas predominantes na bacia hidrográfica são arenitos, quartzodioritos, xistos e grauvaques.

Na área da bacia hidrográfica existem poucas zonas urbanas e praticam-se, sobretudo, as culturas arvenses de sequeiro, embora existindo também explorações agro-silvo-pastoris e florestais, culturas arbóreas-arbustivas, arvenses de regadio e pequenas manchas de eucaliptais (Figura 1.8).

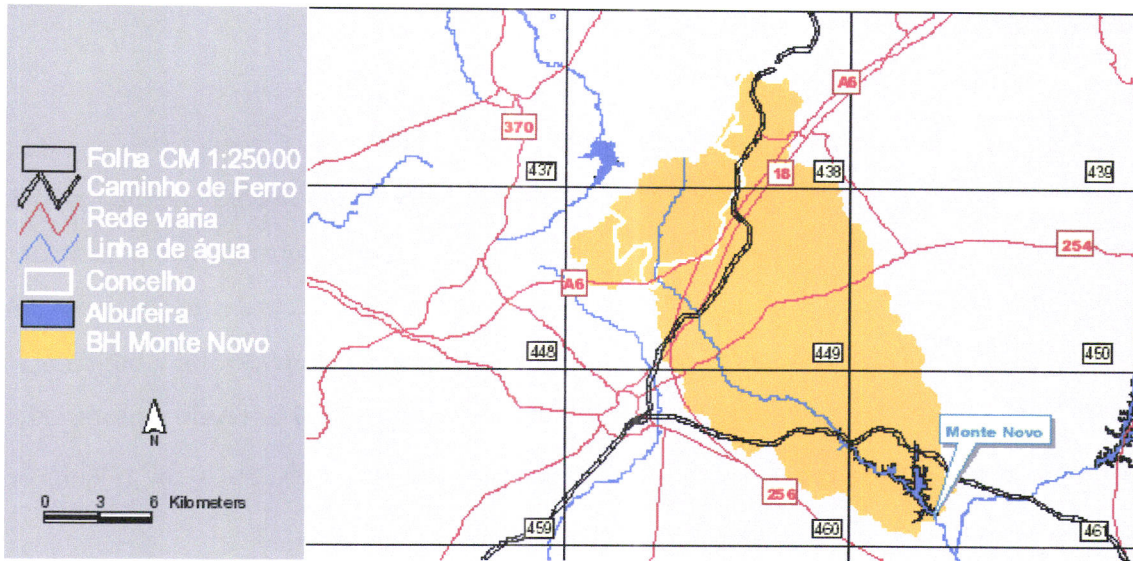


Figura 1.7: Localização da bacia hidrográfica da albufeira de Monte Novo. Adaptada de Matoso [Matoso et al., 2004].

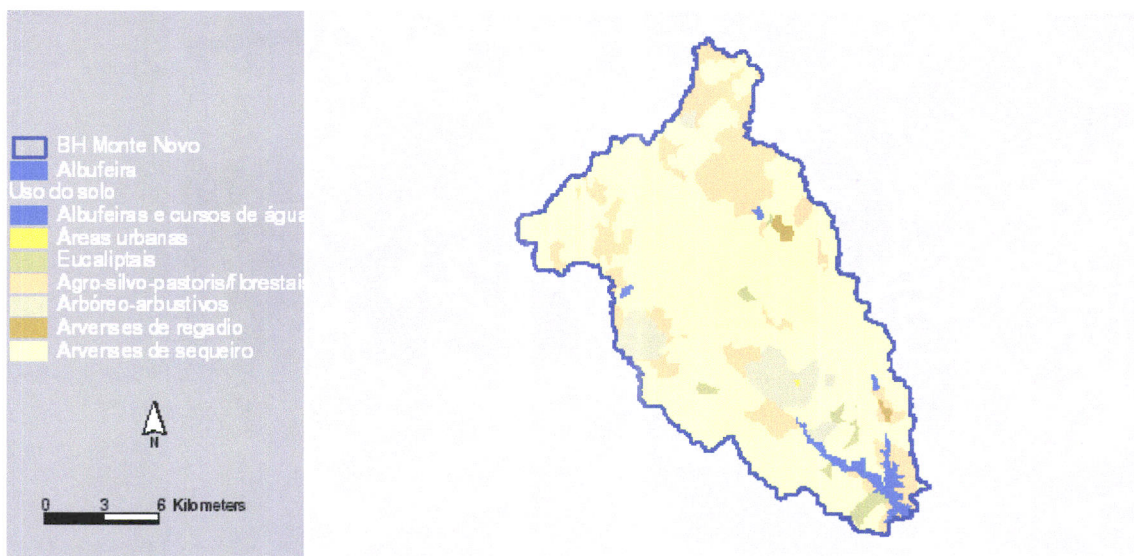


Figura 1.8: Usos do solo na área da bacia hidrográfica da albufeira de Monte Novo. Adaptada de Matoso [Matoso et al., 2004].

As principais fontes de poluição pontual resultam da bovinicultura, da suinicultura, das áreas urbanas, de um lagar de azeite e da indústria de curtumes. Na Figura 1.9 encontram-se localizadas as principais fontes de poluição pontual da bacia hidrográfica da albufeira. O levantamento apresentado, como se referiu anteriormente, diz respeito somente às fontes de poluição de carácter pontual. As actividades pecuárias assinaladas, tal como



no caso da albufeira anterior, apenas dizem respeito aos regimes de exploração intensivo e semi-intensivo.

As fontes de poluição inventariadas possuem sistemas de tratamento de águas residuais. As explorações pecuárias e o lagar de azeite possuem sistemas de retenção, comportando-se como fontes de poluição difusa, uma vez que a matéria orgânica recolhida é usada como fertilizante nos terrenos agrícolas. A indústria de curtumes procede a um tratamento prévio das águas residuais antes de lançá-las na rede pública.

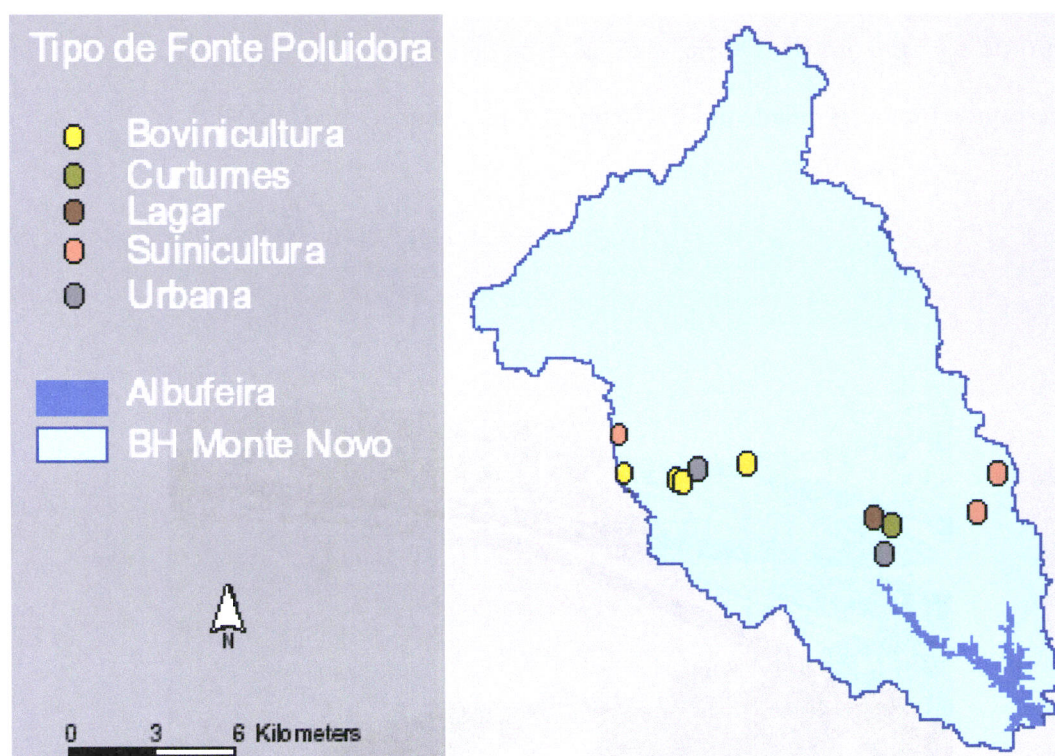


Figura 1.9: Fontes de poluição pontual na bacia hidrográfica da albufeira de Monte Novo. Adaptada de Matoso [Matoso et al., 2004].

A carga produzida pelas fontes de poluição existentes na bacia hidrográfica da albufeira de Monte Novo é de  $502,9 \text{ Kg}_{\text{CBO}_5} \text{ dia}^{-1}$ , o que corresponde ao equivalente a 8382 habitantes. Contudo, em virtude de existirem sistemas de tratamento a carga rejeitada, que entra nas linhas de água, é estimada em  $23,4 \text{ Kg}_{\text{CBO}_5} \text{ dia}^{-1}$ , ou seja, ao equivalente a 390 habitantes [DRAOTA, 1999].

### 1.1.2.3 Albufeira da Vigia

A albufeira da Vigia (Figura 1.10) localiza-se no concelho de Redondo, na ribeira do Vale de Vasco, afluente do rio Degebe, estando integrada na bacia hidrográfica do rio Guadiana. Entrou em funcionamento em 1981 e destina-se, fundamentalmente, à produção de água para consumo humano, servindo cerca de 15500 habitantes dos concelhos de Redondo e Reguengos de Monsaraz. Quando o nível de pleno armazenamento é atingido (i.e. cota 224 m), a área inundada ocupa 2,62 km<sup>2</sup>, apresenta uma altura de 30 m e um comprimento de coroamento de 300 m. A capacidade de armazenamento é  $16,73 \times 10^6 \text{ m}^3$ , a que corresponde uma capacidade útil de  $15,58 \times 10^6 \text{ m}^3$  [Matoso et al., 2004].



Figura 1.10: Vista parcial da albufeira da Vigia.

A bacia hidrográfica ocupa 125 km<sup>2</sup>, apresentando, em certas zonas, declives com inclinações situadas entre 8% e 15%. Está localizada nos concelhos de Redondo e Alandroal (Figura 1.11). As formações litológicas predominantes na bacia hidrográfica são xistos, grauvaques, arenitos, calcários e quartzodioritos.

Na área da bacia hidrográfica existe um aglomerado urbano e praticam-se, sobretudo, culturas arvenses de sequeiro, culturas arbóreas-arbustivas e explorações agro-silvo-pastoris e florestais, existindo também culturas arvenses de regadio (Figura 1.12).

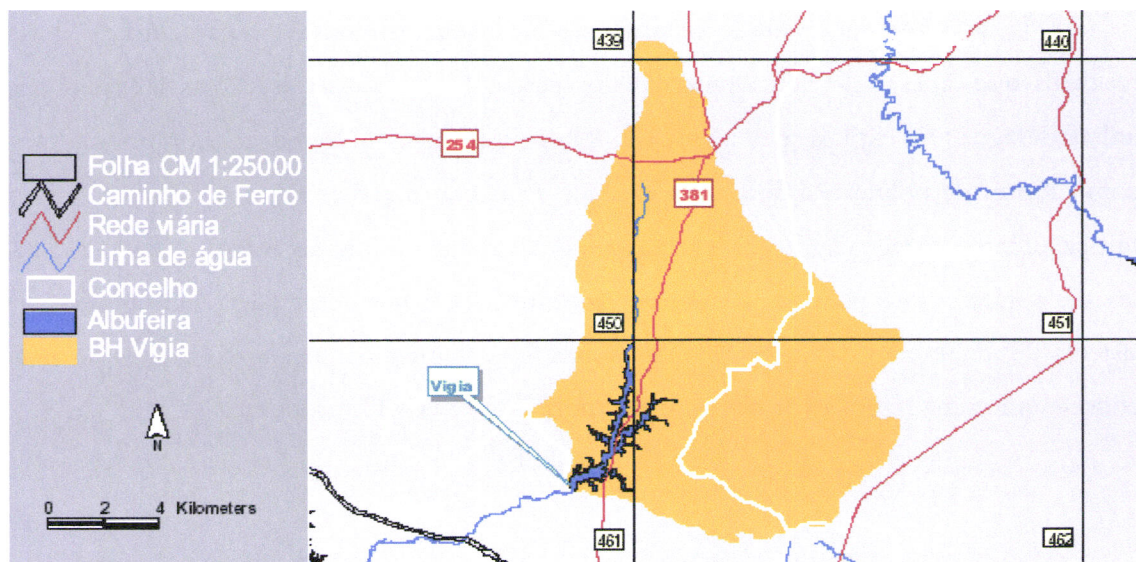


Figura 1.11: Localização da bacia hidrográfica da albufeira da Vigia. Adaptada de Matoso [Matoso et al., 2004].

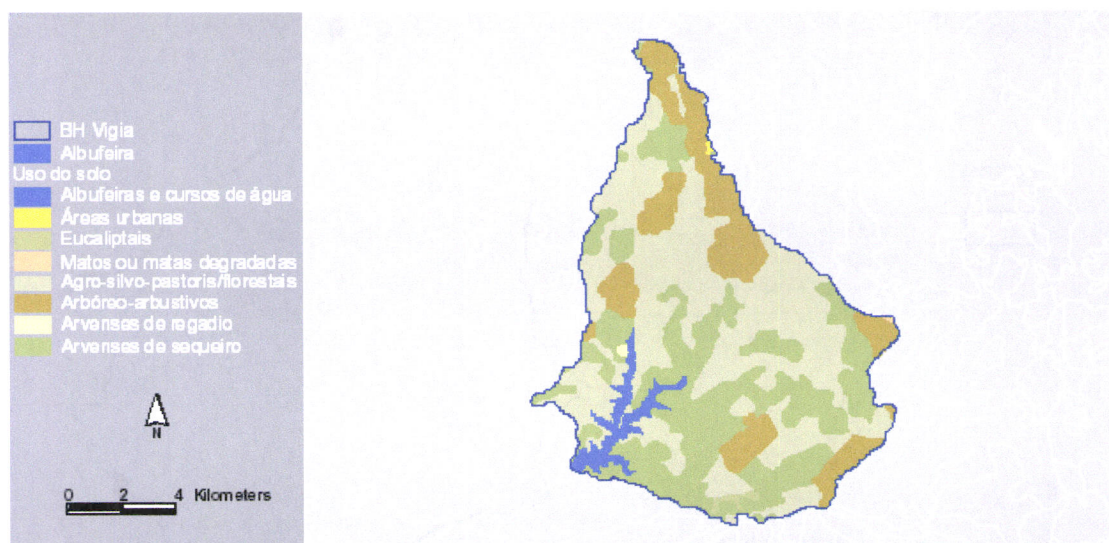


Figura 1.12: Usos do solo na área da bacia hidrográfica da albufeira da Vigia. Adaptada de Matoso [Matoso et al., 2004].

As principais fontes de poluição pontual resultam da suinicultura, dos lagares, da bovinicultura, e da área urbana. Na Figura 1.13 encontram-se localizadas as principais fontes de poluição pontual da bacia hidrográfica da albufeira. O levantamento apresentado, como foi sublinhado anteriormente, diz respeito somente às fontes de poluição de carácter pontual. As actividades pecuárias assinaladas apenas dizem respeito aos regimes intensivo e semi-intensivo, uma vez que o regime extensivo é considerado uma forma de poluição difusa.

As fontes de poluição identificadas possuem sistemas de tratamento de águas residuais que, apesar de deficientes nalguns casos, fazem baixar as cargas lançadas nas linhas de água. As explorações pecuárias possuem sistemas de retenção, comportando-se como fontes de poluição difusa, uma vez que a matéria orgânica recolhida é usada como fertilizante nos terrenos agrícolas. Os lagares utilizam tanques de decantação, sendo as fontes de poluição que mais carga lançam nas linhas de água desta bacia hidrográfica. O sistema de tratamento das águas urbanas utiliza fossa séptica conjugado com filtro de areia conseguindo uma remoção de 50% em relação à carga de  $\text{CBO}_5$  produzida.

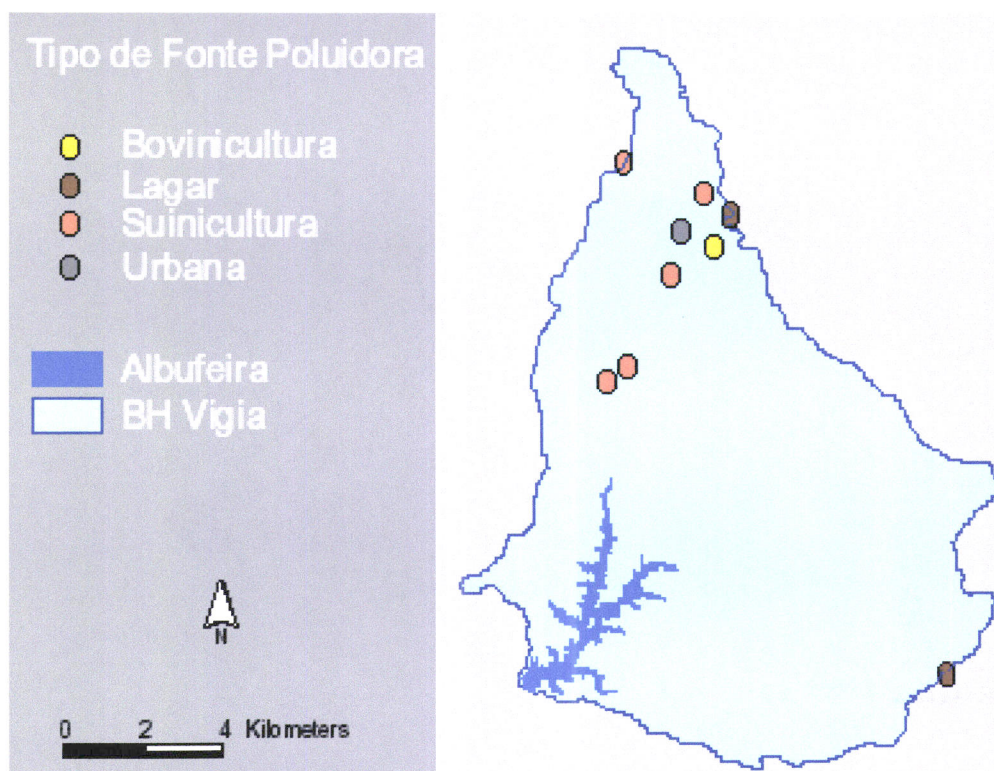


Figura 1.13: Fontes de poluição pontual na bacia hidrográfica da albufeira da Vigia. Adaptada de Matoso [Matoso et al., 2004].

A carga produzida pelas fontes de poluição existentes na bacia hidrográfica da albufeira da Vigia é de  $1679,7 \text{ Kg}_{\text{CBO}_5} \text{ dia}^{-1}$ , o que corresponde ao equivalente a 27995 habitantes. A carga rejeitada, que entra nas linhas de água, é  $723,9 \text{ Kg}_{\text{CBO}_5} \text{ dia}^{-1}$ , ou seja, o equivalente a 12065 habitantes [DRAOTA,1999].

## 1.2 O Processo de Tomada de Decisão

O processo de tomada de decisão, visto segundo uma vertente humanizante ou numa perspectiva não tecnológica, passou a ser um procedimento comum em algumas organizações, instituições e empresas na parte final do século XX, e espera-se que venha a generalizar-se ao longo do século XXI. A situação do indivíduo depende não raras vezes dessas decisões, quando é necessário pesar a oportunidade de uma intervenção do foro ambiental, que é a situação que emana do presente trabalho ou quando se pede a aprovação de um crédito bancário, se subscreve um seguro numa seguradora ou se se debate com a aceitação de uma encomenda de um cliente (i.e. considerando o estado do meio ambiente com base nos resultados obtidos em laboratório, o comportamento do cliente em termos dos seus compromissos comerciais e financeiros, ou a capacidade de produção e a carteira de encomendas, que porventura se lhe possa associar).

Face ao exposto, é natural que se levantem algumas questões, algumas das quais aqui se colocam, na forma:

- (i) - Será possível substituir o ser humano, com capacidade de decisão, por um autómato de decisão?;
- (ii) - Poderão esses sistemas ter um desempenho superior ao do ser humano?;
- (iii) - Os sistemas são desenvolvidos com base no conhecimento transmitido pelo ser humano com capacidade de decisão. Qual é a importância do conhecimento no processo de tomada de decisão?; e
- (iv) - Algumas decisões podem ser razoavelmente automatizadas, outras requerem a intervenção por parte de peritos ou especialistas. Poderão os peritos e os especialistas serem também parte integrante do sistema? Será possível substituí-los, igualmente, por autómatos?

O propósito deste trabalho passa, assim, também pela análise e desenvolvimento de sistemas baseados em conhecimento, com capacidade de raciocínio e inteligência, baseados em entidades que se pretendem autónomas, e que têm as suas raízes na Área Científica da Inteligência Artificial. Esses sistemas deverão ser interactivos, reactivos, distribuídos e servirão de suporte, entre outros, a sistemas de apoio à tomada de decisão. Sistemas deste

tipo pressupõem a necessidade de armazenamento de grandes volumes de dados, a estabilidade e um desempenho satisfatório à escala em que serão utilizados [Abelha, 2004].

O primeiro passo que aqui será dado passará pela procura de um enquadramento em termos da literatura disponível, de forma a definir e caracterizar cada uma das Áreas Científicas que estão na base deste trabalho.

**Definição 1.1** - Um sistema de apoio à decisão é um sistema interactivo, flexível e adaptável, baseado em entidades autónomas (i.e. os agentes), susceptíveis de socialização e tendo como objectivo a solução de problemas não estruturados de modo a melhorar a tomada de decisões. Usa dados, conhecimento, próprio ou de terceiros, dispondo de uma interface amigável com o utilizador [Turban e Aronson, 1998].

Os sistemas de apoio à decisão apresentam, não raras vezes, as seguintes limitações:

- (i) - A dificuldade em se adaptarem, aprenderem e evoluírem com o contacto com o utilizador;
- (ii) - A dificuldade em contemplarem novas situações;
- (iii) - O processo de tomada de decisão é feito, normalmente, em grupo. Contudo, os sistemas de apoio à decisão não contemplam, normalmente, esta valência;
- (iv) - A tomada de decisão requer informação externa à organização, a qual, não raras vezes, os sistemas de apoio à decisão não consideram; e
- (v) - Para uma grande parte dos problemas, os sistemas de apoio à decisão apresentam dificuldades em sustentar as decisões propostas, dificultando um diálogo sustentável com o utilizador.

**Definição 1.2** - A Inteligência Artificial é a área do conhecimento que estuda o modo de fazer com que os computadores realizem tarefas nas quais, de momento, os humanos levam vantagem [Rich e Knight, 1991].

**Definição 1.3** - A Inteligência Artificial é o ramo da Ciência dos Computadores que tem a ver com a automatização de comportamentos inteligentes, próprios dos seres humanos [Luger e Stubblefield, 1998].

Há aqui uma oposição clara entre o apoio à decisão com intervenção humana e a automatização da decisão. Isto não quer dizer que os sistemas de apoio à decisão não possam ser classificados como inteligentes, ou que na Inteligência Artificial não possa haver interacção com o ser humano. No entanto, em termos de definição, a separação deve ser transparente. Sente-se, também, uma certa dicotomia (i.e. os sistemas de apoio à decisão visam essencialmente ajudar o ser humano, enquanto que os sistemas inteligentes parecem querer substituí-lo).

**Definição 1.4** - Um sistema baseado em conhecimento é, como o nome sugere, um sistema que usa conhecimento e raciocínio para obter conclusões. Difere dos programas tradicionais de processamento de dados devido à sua capacidade de expressão e ao modo de operação [Winstansley, 1991].

**Definição 1.5** - Um sistema baseado em conhecimento é um sistema computorizado que usa conhecimento sobre um certo domínio ou universo de discurso, para obter soluções para problemas desse mesmo domínio. Assim, a solução para um dado problema é essencialmente a mesma que pode ser obtida por alguém familiarizado com esse mesmo domínio do conhecimento, quando confrontado com o mesmo problema [Gonzalez e Dankel, 1993].

O conhecimento foi reconhecido como o principal factor de sucesso dos sistemas inteligentes, nomeadamente a partir da década de 70. Mais recentemente, há novas leituras para a questão do conhecimento, como por exemplo a Gestão do Conhecimento e o conceito de Sociedade do Conhecimento. Pode-se dizer que o conhecimento beneficia o desempenho dos sistemas e favorece a tomada de decisão. Mais conhecimento permite tomar (automaticamente ou não) melhores decisões, mais correctas, menos arriscadas, mais baratas. O conhecimento pode ser obtido através de peritos ou especialistas ou automaticamente a partir de dados, e em certas situações usando técnicas de descoberta e extracção de conhecimento. Os sistemas baseados em conhecimento são um caso particular de sistemas inteligentes. O conhecimento estabelece uma ponte entre o sistema inteligente e o ser humano, que neste processo assume um papel central. Estes sistemas, porém, também apresentam algumas limitações, destacando-se entre estas:

- (i) - A dificuldade em encontrar e partilhar as fontes que disponibilizam o conhecimento a terceiros e o modo de obter tal conhecimento, assim como a forma a seguir para caracterizar os peritos e a sua competência;

- (ii) - O projecto de desenvolvimento de sistemas de apoio à decisão é demorado e, não raras vezes, é-se colocado perante dificuldades acrescidas nos procedimentos que visam a descoberta e a extracção do conhecimento;
- (iii) - O conhecimento não é absoluto e evolui ao longo do tempo (há um ciclo de vida do conhecimento); e
- (iv) - A principal causa de descontinuação dos sistemas baseados em conhecimento passa por se ter o utilizador, confrontado com custos excessivos em termos de manutenção.

**Definição 1.6** - Os agentes são entidades baseadas em sistemas de *hardware* ou *software* que apresentam, entre outras, valências como a autonomia e a capacidade de socialização, podendo ser reactivos e pró-activos [Wooldridge e Jennings, 1995].

O agente pode ser um elemento fundamental de um sistema inteligente, em particular no que concerne à automatização dos processos de raciocínio e tomada de decisão. Para sobreviver, os agentes são forçados a dotarem-se de atributos que lhes permitam o planeamento das suas acções e a definição de estratégias, de capacidade de previsão, e de coordenação, i.e. assumirem-se como a peça central numa nova metodologia de computação, a dos Sistemas MultiAgente, com vista à resolução de problemas [Coelho, 1995].

Os Sistemas MultiAgente assumem-se, assim, como um elo na evolução natural para os sistemas inteligentes, funcionando como elementos de substituição ou de delegação de tarefas, habitualmente executadas pelos humanos. No entanto, aos sistemas baseados em agentes são normalmente apontadas algumas limitações, e em particular:

- (i) - A ausência de um controlador global do sistema;
- (ii) - A ausência de uma perspectiva global do universo de discurso; e
- (iii) - A falta de confiança e o receio na delegação de competências, por parte de humanos.

Para que alguém possa delegar tarefas, numa dada entidade, em primeiro lugar há que estabelecer relações bilaterais de confiança [Jennings e Wooldridge, 1998]. As organizações necessitam de amadurecer a sua experiência relativamente a componentes de



*software* autónomos e, o processo de ganho de confiança em entidades, como os agentes, por parte dos utilizadores, tem o seu tempo de incubação.

### 1.3 A Gestão do Conhecimento

A Gestão do Conhecimento tem a ver com o aumento do desempenho das organizações, ao permitir-lhes capturar, partilhar e aplicar o conhecimento que emana de um ou mais colectivos, de forma a tomar decisões em tempo real. A Gestão do Conhecimento é a pedra angular para que as empresas, organizações e instituições sobrevivam numa sociedade e economia competitivas. Por outro lado, o fornecimento da informação, do conhecimento e da experiência, de forma a tomar as melhores decisões, deve ser concretizado em tempo real. Por tempo real entende-se o tempo disponível para tomar a decisão, i.e. desenvolver uma acção que afecte positivamente os resultados. A Gestão do Conhecimento está ligada a questões de eficiência, de produtividade, de qualidade e do recurso à experiência e a formas de meta-conhecimento. Outros pressupostos, com implicações a médio e longo prazo, são necessários para produzir resultados. Não é suficiente apostar na tecnologia, há que melhorar processos internos e externos em termos de gestão e organização, motivar e formar as pessoas, olhar para o seu comportamento dentro das organizações e, sustentar uma maior capacidade de inovação.

Uma das tarefas mais árduas dentro de uma organização tem naturalmente a ver com a partilha de conhecimento entre indivíduos, quer estes estejam próximos ou separados, de forma a que seja possível:

- (i) - Partilhar o que sabem;
- (ii) - Aplicar o conhecimento de terceiros;
- (iii) - Capturar e partilhar o conhecimento num formato passível de ser aplicado dentro da organização;
- (iv) - Melhorar os canais de comunicação, sejam estes locais ou remotos;
- (v) - Interagir com os colegas, em qualquer local, em qualquer altura e de diferentes formas; e

- (vi) - Utilizar novas aplicações, tecnicamente avançadas, assim como o uso de ferramentas de suporte à decisão, como a mineração de dados, a extração de conhecimento e a simulação.

A informação pode então ser definida como um conjunto de dados que são interpretados e o conhecimento como a informação em acção, i.e. a informação transformada em capacidade para dar corpo a acções efectivas. Uma acção pressupõe uma alteração ao estado do mundo em que se desenvolve. O conhecimento é visto como o recurso mais importante da sociedade contemporânea, muitas vezes referida por sociedade da informação ou do conhecimento, tal como aconteceu com a qualidade e os serviços de apoio ao cliente em décadas anteriores. A capacidade de aprender rapidamente e continuamente é uma chave para o sucesso das organizações.

## 1.4 A Gestão do Conhecimento e a Inteligência Artificial

A Gestão do Conhecimento e a Inteligência Artificial têm pelo menos um objetivo comum, i.e. melhorar o desempenho dos actores individuais ou de suas comunidades, usando para isso os computadores. A Inteligência Artificial assumiu sempre um papel central na temática subjacente à Gestão do Conhecimento, em termos do desenvolvimento de novas tecnologias, ferramentas ou metodologias de programação e computação, aplicadas à resolução de problemas. Ora, ao falar-se em conhecimento, há que ter em atenção a sua forma de armazenamento e os processos de inferência e/ou pesquisa que lhe estão associados, ou seja, há que atender:

- (i) - Às Bases de Conhecimento - terminologias diferentes podem ser usadas para caracterizá-las, tais como as de memória colectiva ou repositório de conhecimento. A Inteligência Artificial e a Lógica Matemática têm aqui uma palavra a dizer em especial no que respeita aos processos de aquisição e representação do conhecimento; e
- (ii) - Aos Processos de Pesquisa - os motores de pesquisa são elementos chave dos sistemas de Gestão do Conhecimento. Recorrem a técnicas bem conhecidas na Inteligência Artificial, que têm a ver com a interpretação

---

da língua natural, dos processos de indexação e recuperação de casos em sistemas baseados em casos, das tecnologias que recorrem a metodologias de resolução de problemas baseadas em agentes e de resolução de problemas em ambientes distribuídos.

O conhecimento é normalmente difícil de extrair, efêmero e apenas aproximadamente correcto. Este é um dos sinais quando se pretende desenvolver sistemas inteligentes, em que pontuam as incertezas, a incompletude e as situações imprevistas.

Há assim que recorrer a outros mecanismos computacionais, oriundos da Área Científica da Inteligência Artificial, como o Raciocínio Simbólico e as Redes Neurais Artificiais, as quais são constituídas por elementos computacionais relativamente simples, operando em paralelo, inspiradas no funcionamento do sistema nervoso central dos seres humanos, e que têm vindo a resolver problemas complicados em quase todos os domínios da ciência e da tecnologia [Zupan e Gasteiger, 1993] [Rumelhart et al., 1994] [Luger e Stubblefield, 1998] [Abelha, 2004]. Também na área da limnologia e do controlo e gestão da eutrofização foram aplicadas nos últimos anos do século XX, com resultados prometedores, e são um campo em franca expansão nestes primeiros anos do século XXI [Lek et al., 1996] [Maier e Dandy, 1996] [Maier et al., 1998] [Neves et al., 1998] [Wen e Lee, 1998] [Scardi e Harding Jr., 1999] [Schleiter et al., 1999] [Karul et al., 2000] [Recknagel, 2001] [Werner e Obach, 2001] [Recknagel, 2003].

## 1.5 Inovação e Valor Acrescentado do Trabalho

A relevância de qualquer tipo de informação para a resolução de um dado problema está na razão directa da frequência com que é considerada e do modo como é aplicada nessas situações. A fonte de informação ideal seria aquela que facultasse apenas o conhecimento e os dados relevantes para a resolução do problema, e o fizesse com um mínimo de esforço por parte do utilizador. As deficiências com que as actuais fontes externas de informação se apresentam tornam claro que algo mais é necessário, e este parece ser o mote para o aparecimento de alternativas.

Muito se tem dito acerca das ferramentas que transformarão a Gestão Ambiental, porém algumas das suas características estão finalmente a emergir. Tais ferramentas de-

vem dar resposta a questões deveras complexas e em tempo útil, para o que se terão de ligar a bases de dados e de conhecimento, terem como base a electrónica e a robótica, serem portáteis e fáceis de utilizar. As ferramentas não poderão ter uma atitude passiva, ou seja, limitarem-se a responder a questões colocadas pelos utilizadores, mas subscrever um comportamento pró-activo, o que passa por interrogar o gestor ambiental e explicar-lhe as suas decisões, procurando ajudar, encontrando uma solução técnica eficiente, desenvolvendo a alternativa mais amiga do ambiente e, jamais se assumindo como uma entidade indesejada, i.e. são ferramentas com que os gestores ambientais irão tratar a informação quer no presente quer no futuro, o que passa por que se estabeleçam ligações entre os dados ambientais e a sua difusão através da *Internet*, com o gestor a posicionar-se na sua intersecção, não tendo contudo que prestar atenção à tecnologia que está a ser utilizada. A resposta a este problema pode ser considerada por um mero exercício de quantificação dado na forma:

$$\text{utilidade - da - informação - ambiental} = \frac{\text{relevância} \times \text{validade}}{\text{trabalho - de - acesso}} \quad (1.1)$$

em que o factor relevância foi já definido no texto. Por validade entende-se uma medida da qualidade da informação que está a ser objecto de estudo. No que respeita ao trabalho-de-acesso, está-se perante uma medida do tempo e do esforço realizado para extrair a informação desejada. Tudo isto implicou a concepção e adaptação de novas metodologias de registo de dados ambientais.

Neste trabalho aplicam-se técnicas da Área Científica da Inteligência Artificial ao desenvolvimento de modelos para três albufeiras do distrito de Évora. Com este estudo pretende dar-se uma pequena contribuição para o esforço que há a fazer em Portugal e particularmente no Alentejo, na preservação e gestão adequada dos recursos hídricos. Para além da descrição do que normalmente está associado à caracterização da qualidade da água em albufeiras, é objectivo deste trabalho fazer uma revisão da literatura sobre modelos da qualidade da água. Tenta-se ainda fazer uma descrição tão completa quanto possível do estado actual dos conhecimentos e do estado das aplicações da modelação da qualidade da água. Há, no entanto, a noção que é um trabalho sempre inacabado.

Em resumo, os objectivos principais deste trabalho são os seguintes:

- (i) - Descrever os problemas e fenómenos associados à qualidade da água em albufeiras;

- (ii) - Fazer a revisão bibliográfica sobre a modelação da qualidade da água em albufeiras;
- (iii) - Utilizar as ferramentas da Área Científica da Inteligência Artificial no desenvolvimento de modelos de qualidade da água para três albufeiras situadas no Alentejo, em particular no distrito de Évora;
- (iv) - Realçar a importância de tais ferramentas e dos modelos na gestão e exploração sustentada deste tipo de aproveitamento hídrico;
- (v) - Realçar a capacidade dos modelos na descrição do estado da qualidade da água em tempo real e de fazerem previsões sobre cenários futuros;
- (vi) - Alargar a aplicabilidade do modelo a albufeiras de grandes dimensões, tendo em vista a sua futura utilização em Alqueva;
- (vii) - Alertar para a importância da obtenção de dados hidrológicos, meteorológicos, e de qualidade da água que possibilitem desenvolver e aplicar este tipo de modelos;
- (viii) - Conhecer quais os parâmetros fundamentais a determinar para efectuar a caracterização completa da qualidade da água nas albufeiras estudadas;  
e
- (ix) - Apresentação de modelos (recorrendo quer ao Cálculo Simbólico quer às Redes Neurais Artificiais) para a resolução de problemas que condicionavam o processo de quantificação da qualidade das águas numa albufeira.

## 1.6 Estrutura do Trabalho

A tese está estruturada e organizada em seis capítulos, não se contabilizando nestes a bibliografia. O primeiro capítulo corresponde a uma introdução, onde se faz o enquadramento do trabalho, apresentam-se os objectivos e se descreve a sua estrutura.

No segundo capítulo faz-se, em primeiro lugar, uma abordagem à evolução da modelação de ecossistemas. Situam-se no tempo as cinco gerações de modelos e caracteriza-se cada uma delas. Em segundo lugar referem-se, ainda que de um modo bastante geral, as características principais dos ecossistemas e o modo como os modeladores as tentaram incorporar nos modelos. Abordam-se novas vias para a resolução do problema, nomeadamente as que se baseiam na teoria do caos, na teoria da catástrofe e em técnicas de programação, como a programação orientada ao objecto, a programação orientada ao indivíduo (ou agente) e modelos baseados em conhecimento. Neste capítulo também se caracterizam as massas de água em estudo, nomeadamente no que concerne ao estado trófico e à qualidade da água.

Nos terceiro e quarto capítulos aborda-se a utilização de técnicas e metodologias de resolução de problemas oriundas da Área Científica da Inteligência Artificial na modelação de sistemas ambientais, respectivamente através do Cálculo Simbólico, das Redes Neurais Artificiais (RNAs) e da Descoberta de Conhecimento em Bases de Dados. No que concerne ao Cálculo Simbólico é tratada, no capítulo três, a problemática associada à definição das propriedades e das arquitecturas em que se fundamentam as entidades computacionais que se irão utilizar, denominadas de Agentes ou Sistemas MultiAgente. No que diz respeito ao capítulo quatro, abordam-se modelos inspirados na Natureza (i.e. as RNAs), descreve-se a sua constituição, apresentam-se os diferentes tipos e o modo como funcionam. É tratada ainda a problemática da Descoberta de Conhecimento em Bases de Dados, apresentando-se modelos e técnicas de *Data Mining*, destacando os objectivos e as abordagens.

No capítulo cinco é apresentada e dada uma especificação formal no que concerne à arquitectura de suporte do sistema **AQUA** (i.e. Agência para a Integração, Difusão e Arquivo de Informação sobre a **Q**ualidade da **Á**gua em Albufeiras). Apresentam-se e

discutem-se os resultados conseguidos através da Modelação Simbólica e da Modelação Sub-simbólica (i.e. RNAs e Árvores de Decisão).

Por fim, no sexto e último capítulo apresentam-se as conclusões finais e perspectiva-se o trabalho que se pretende desenvolver no futuro, direccionado para a barragem de Alqueva.

## 1.7 Conclusões

Neste capítulo apresenta-se a questão central desta tese (i.e. a previsão do comportamento das massas de água em resposta ao aumento de nutrientes). Referem-se as principais dificuldades do tipo de abordagens que têm sido tentadas (e.g. a elevada complexidade, a necessidade de processos morosos de calibração e a redução de processos complexos a equações lineares). Dá-se ênfase às potencialidades do tipo de abordagem proposta; i.e. o carácter pró-activo dos métodos de previsão e a característica distribuída da ferramenta integrada de suporte ao processo de tomada de decisão e de gestão do conhecimento. A referida abordagem está fundamentada em disciplinas da Área Científica da Inteligência Artificial, enquadrando-se nas temáticas do Raciocínio Simbólico, das Redes Neurais Artificiais e da Descoberta de Conhecimento em Base de Dados.

Apresentam-se e caracterizam-se as albufeiras onde se centrou o presente estudo, quer do ponto de vista geomorfológico ou climático, referindo-se ainda os principais usos do solo e as fontes de poluição difusa existentes na área da bacia hidrográfica de cada uma das albufeiras.





## Capítulo 2

### Modelos Dinâmicos de Ecossistemas

### Aquáticos de Uso Genérico

A espécie humana começou, desde tempos imemoriais, a utilizar modelos, uma visão simplificada da realidade, como ferramenta para ajudar a perceber e/ou resolver problemas. Obviamente que um modelo não contém todos os pormenores do sistema real, sob pena de, ele próprio, se transformar no sistema real. Um modelo deve ter em linha de conta os elementos do universo de discurso que são essenciais para a resolução do problema em causa.

Por exemplo, durante muitos anos utilizaram-se modelos, à escala natural, de navios para determinar qual o perfil que minimizava a resistência da água. Esses modelos contemplavam o perfil do navio, as suas dimensões, mas não incluíam pormenores como o equipamento de bordo ou o equipamento das cabinas. Tais detalhes são, obviamente, irrelevantes para o objectivo do modelo. Este exemplo é ilustrativo da filosofia por detrás da construção e utilização de um modelo. Um modelo desenvolvido para um ecossistema deve conter os aspectos essenciais do problema a resolver, o que nem sempre é fácil de seleccionar, uma vez que os ecossistemas são muito mais complexos do que o exemplo a que nos temos vindo a reportar.

Os modelos ambientais podem ser comparados a mapas. Existem diversos tipos de cartas, destinadas a servirem objectivos diferentes. O geólogo procura um certo tipo de informação, diferente da que procura um arqueólogo ou um automobilista.

Para além da escolha do tipo de mapa há ainda um outro aspecto importante, a selecção da escala. Por exemplo, um automobilista que pretenda fazer uma viagem apenas por auto-estrada, basta-lhe uma escala de banda larga. Quem pretenda explorar uma deter-

minada zona para conhecer as aldeias mais recônditas precisa de informação sobre estradas secundárias, municipais e muitas vezes sobre caminhos rurais, i.e. seguir informação pormenorizada e a escala a escolher tem de contemplar este pormenor.

Um mapa não contém toda a informação para a área a que se refere, já que por um lado, iria dificultar a sua consulta e por outro lado perdia rapidamente significado (e.g. podemos encontrar informações sobre a localização de parques de estacionamento mas já não encontramos informação sobre o número de lugares vagos).

Do mesmo modo podem coexistir vários modelos para o mesmo ecossistema de acordo com o fim em vista. Contudo, em todos é dada ênfase aos aspectos que têm interesse para o estudo. A inclusão de pormenores menos importantes pode dificultar e até mesmo impedir que se atinjam as metas definidas.

Os modelos podem ser físicos (e.g. o do navio na minimização da resistência da água) ou obtidos através de uma especificação, que recorre a métodos formais, como os oriundos da Lógica Matemática.

O movimento das populações, quer a nível tecnológico quer a nível urbano causa impactos ambientais. Submetem-se os ecossistemas aos mais diversos tipos de agressão como a poluição e, com isso, assiste-se ao crescimento cada vez mais rápido de algas, de bactérias e de espécies indesejadas, o que leva, por vezes, a alterações profundas desses ecossistemas. Torna-se pois fundamental e prioritário prever qual o efeito ambiental que uma determinada descarga de efluentes irá produzir.

A modelação de ecossistemas remonta aos anos vinte (Figura 2.1), aos trabalhos de Streeter e Phelps [Streeter e Phelps, 1925] sobre o balanço de oxigénio num rio e de Lotka [Lotka, 1924] e de Volterra [Volterra, 1926] sobre as relações predador-presa. Nas décadas de cinquenta e de sessenta assistiu-se ao desenvolvimento de modelos de dinâmica de populações e de modelos mais específicos para rios - modelos de segunda geração.

O aumento da utilização de modelos como ferramenta de gestão ambiental deu-se nos anos setenta, onde surgiu o primeiro modelo de eutrofização. Estes modelos, designados por terceira geração, caracterizam-se por serem demasiado complexos. Em certa medida, foi a revolução operada na tecnologia computacional que “criou” esta geração de modelos. Tornou-se, contudo, óbvio, em meados da década de setenta, que as limitações da modelação não eram de natureza computacional ou formal, mas sim devido à falta de dados e de conhecimentos sobre os sistemas e sobre os processos a modelar.

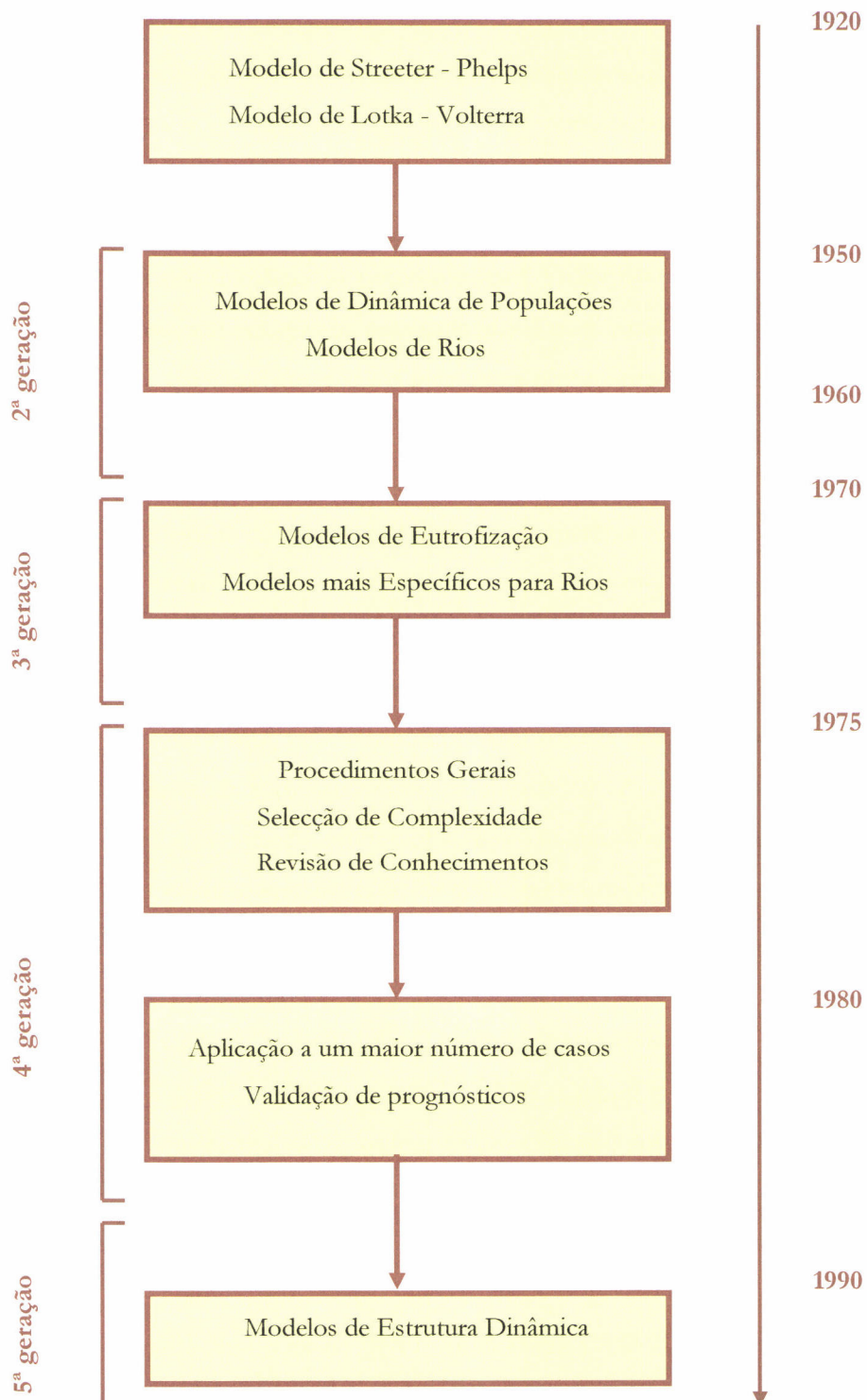


Figura 2.1: Esquema do desenvolvimento da modelação ambiental.

Os modeladores tornaram-se mais críticos na aceitação dos modelos. Fizeram notar que a base dos modelos ambientais estava no conhecimento aprofundado do ecossistema, do problema em causa e dos processos. A investigação ambiental ganhou um cariz quantitativo. Como resultado definiram-se procedimentos a seguir, entre os quais se destacam a conceptualização do problema, a selecção dos parâmetros, a verificação, a calibração, a análise de sensibilidade e a validação.

Em paralelo, começou a dar-se importância a estudos de complexidade dos modelos, fruto da ponderação de vários factores como os dados disponíveis, o problema a resolver, o ecossistema em causa e os conhecimentos adquiridos [Jørgensen, 1995a].

No período compreendido entre meados dos anos setenta até meados da década seguinte desenvolveram-se modelos designados pela quarta geração. São caracterizados por assentarem numa análise cuidadosa das bases de conhecimento científico e revelam simplicidade e realismo. Muitos modelos foram validados a partir de resultados comprovados no terreno e para alguns houve mesmo a possibilidade de se validarem os prognósticos [Jørgensen, 1995a].

Presentemente assiste-se a uma reavaliação destes modelos. É bastante claro que os modelos são demasiado rígidos quando comparados com a enorme flexibilidade patenteada pelos ecossistemas. As hierarquias e os mecanismos de regulação não têm sido considerados nos modelos, os quais não têm sido capazes de prever a adaptabilidade, as variações de dinâmica e as alterações na própria estrutura dos ecossistemas. Esta quinta geração de modelos (i.e. os modelos de estrutura dinâmica) está a dar os primeiros passos. Diversas vias têm sido tentadas como a utilização de funções objectivo [Mejer e Jørgensen, 1979] [Straškraba, 1979] [Salomonsen e Jensen, 1996] [Ray et al., 2001], aplicação da teoria do caos [Hasting et al., 1993] [Suárez, 1999] [Vayenas e Pavlou, 2001], da teoria da catástrofe [Dubois, 1979], de modelos baseados na programação orientada ao objecto [Sequeira et al., 1991] [Baveco e Lingeman, 1992] [Silvert, 1993] [Sequeira et al., 1997] [He et al., 1999], modelos de base individual [De Angelis e Gross, 1992] [Grimm, 1999], assim como o recurso à Inteligência Artificial (IA) e Sistemas Inteligentes (SI) [Lek et al., 1996] [Maier e Dandy, 1996] [Maier et al., 1998] [Neves et al., 1998] [Wen e Lee, 1998] [Scardi e Harding Jr., 1999] [Schleiter et al., 1999] [Karul et al., 2000] [Recknagel, 2001] [Werner e Obach, 2001] [Recknagel, 2003] e modelos assentes em conhecimento difuso [Salski, 1992].

Jørgensen [Jørgensen, 1995a] fez uma avaliação subjectiva e meramente qualitativa, quer dos ecossistemas quer dos problemas que se tem tentado modelar utilizando uma escala de 0 a 5, atribuindo o valor 5 a ecossistemas/problemas sobre os quais têm recaído esforços muito intensos (muitos modelos devidamente calibrados e validados); o valor 4 para os que têm recolhido esforços intensos; o valor 3 para os que têm atraído algum esforço de modelação; o valor 2 para aqueles que têm poucos modelos devidamente calibrados e validados; o valor 1 no caso de haver um só modelo calibrado e validado ou vários cuja calibração e validação não seja suficiente e o valor 0 para aqueles em que ainda não foi feito praticamente nenhum esforço de modelação. Os resultados desta análise são apresentados na Tabela 2.1 para alguns ecossistemas e na Tabela 2.2 para alguns dos problemas que estes evidenciam.

A observação da Tabela 2.1 e da Tabela 2.2 revela que os maiores esforços de modelação se concentram em ecossistemas aquáticos, nomeadamente rios, lagos, albufeiras, barragens e estuários, e os problemas modelados estão relacionados com o balanço de oxigénio, com a eutrofização, e com a contaminação de águas subterrâneas. Outro campo que tem merecido atenção diz respeito à poluição com metais pesados e à poluição atmosférica, com destaque para os problemas da distribuição de poluentes, efeito de estufa e chuvas ácidas.

Tabela 2.1: Esforço de modelação para alguns ecossistemas.

<b>Ecossistema</b>	<b>Esforço de modelação</b>
Rios	5
Lagos, albufeiras e barragens	5
Estuários	5
Zonas costeiras	4
Alto mar	3
Zonas pantanosas	4
Deserto	1
Florestas	4
Terrenos agrícolas	5
Savana	2
Ecossistemas do Ártico	0

Tabela 2.2: Esforço de modelação para alguns problemas ambientais.

Problema	Esforço de modelação
Balanço de oxigénio	5
Eutrofização	5
Poluição com metais pesados	4
Poluição de ecossistemas terrestres com pesticidas	4
Protecção de parques naturais	3
Dinâmica de populações em parques naturais	3
Poluição de águas subterrâneas e superficiais	5
Efeito de estufa	5
Chuvas ácidas	5
Distribuição de poluentes no ar	5
Alterações num microclima	3
Destruição da camada de ozono	3

O problema da modelação de substâncias tóxicas só começou a ser tratado na segunda metade dos anos setenta. Até aí raramente estavam associados à modelação ambiental. Admitia-se que os problemas podiam ser resolvidos, eliminando a fonte da substância tóxica. Só durante os anos setenta se reconheceu a complexidade do problema, e os vários acidentes que lançaram substâncias tóxicas tanto na atmosfera como nas águas como nos solos, contribuíram para reforçar a necessidade de modelos ecotoxicológicos.

Os modelos desenvolvidos para ecossistemas tentam captar as características destes. Contudo, os ecossistemas diferem da maioria dos outros sistemas por serem extremamente adaptáveis, terem capacidade de auto-organização e terem um número muito elevado de mecanismos de regulação [Straškraba, 1994] [Straškraba, 2001]. O verdadeiro desafio aos modeladores reside no facto de serem necessários modelos capazes de reflectir essas características e têm sido feitas, nos últimos anos, algumas tentativas nesse sentido.

Os modelos incluídos nas quatro primeiras gerações são demasiado inflexíveis, assentam em conjuntos de parâmetros rígidos, representam as propriedades das variáveis de estado tal como eram no período de observação. São incapazes de prever a adaptabilidade, as variações de dinâmica e as alterações dos próprios ecossistemas. A sua validade é, muitas vezes, limitada no espaço e no tempo.

Os modelos de estrutura dinâmica ou de quinta geração tentam considerar as alterações na composição das espécies bem como a capacidade destas se adaptarem às condições que lhes são impostas. São designados por modelos de estrutura dinâmica para indicar a sua capacidade de prever as mudanças estruturais. Também são chamados a quinta geração de modelos ecológicos para distinguir e sublinhar que são diferentes dos anteriores, que podem fazer mais, nomeadamente prever alterações na composição das espécies.

Poder-se-á argumentar que a capacidade de substituir as espécies presentes por outras melhor adaptadas pode ser introduzida nos modelos de terceira e quarta geração incorporando, para intervalo de tempo considerado, todas as espécies existentes. Contudo, este procedimento apresenta duas desvantagens principais. Em primeiro lugar, tornaria os modelos demasiado pesados, uma vez que passariam a ter um grande número de variáveis de estado para cada nível trófico, o que implicaria o recurso a muitos parâmetros que teriam que ser validados e calibrados e, como se referiu anteriormente, a sua utilização não só aumenta a incerteza dos modelos como também os torna muito específicos [Nielsen, 1992]. Em segundo lugar, os modelos não perderiam a característica de rigidez, uma vez que os parâmetros permaneceriam fixos enquanto que, na verdade, estão constantemente a variar, mesmo que não ocorram alterações na composição das espécies [Fontaine, 1981].

Antes de se discutirem os modelos de estrutura dinâmica apresentam-se algumas características relevantes dos ecossistemas, sublinhando-se aspectos como a flexibilidade, a adaptabilidade e o potencial de evolução.

## 2.1 Características dos Ecossistemas

A ecologia estuda sistemas irreduzíveis [Wolfram 1984a] [Wolfram, 1984b] [Jørgensen, 1990a] [Jørgensen, 1990b] [Jørgensen, 1992a] [Jørgensen, 1992b] [Jørgensen, 1995a] [Jørgensen et al. 1995]. Não se podem planejar experiências simples que revelem relações que possam ser transferidas, integralmente, de uma situação num ecossistema para outro ecossistema. Isso é possível, por exemplo, com a lei da gravidade porque a relação entre a força e a aceleração da gravidade se reduz a uma equação linear, enquanto que o crescimento de microorganismos depende de muitos factores relacionados entre si, dependentes do tempo e governados por mecanismos de regulação [Straškraba, 1979] [Straškraba, 1994] [Straškraba, 2001].

Jørgensen [Jørgensen, 1997] apresentou uma hierarquia dos mecanismos de regulação. A Tabela 2.3 mostra que a complexidade impede o recurso a equações simples, que possam ser utilizadas de forma repetitiva. Um ecossistema tem tantos componentes a interagir que é impossível examinar todas as relações e, mesmo se houvesse essa possibilidade, não seria possível determiná-las porque as relações são diferentes quando o processo ocorre na natureza ou no laboratório, longe dos restantes componentes do ecossistema.

Tabela 2.3: Hierarquia dos mecanismos de regulação.

Nível	Processo de regulação	Exemplificação com o crescimento do fitoplâncton	Escala
1	Regulação através da concentração do meio	Assimilação de fósforo de acordo com a concentração deste no meio	min - h
2	Regulação pela necessidade	Assimilação de fósforo de acordo com a concentração intracelular deste	min - h
3	Regulação por outros factores externos	Concentração de clorofila de acordo com a radiação solar	Dias
4	Adaptação de propriedades	Alteração da temperatura óptima de crescimento	dias - meses
5	Seleção de outras espécies	Prevalcimento de espécies melhor adaptadas	semanas - anos
6	Seleção de outra cadeia alimentar	Prevalcimento da cadeia alimentar melhor ajustada	meses - anos
7	Mutações e novas recombinações sexuais	Aparecimento de novas espécies ou alterações nas propriedades das espécies	10 - 10 <sup>5</sup> anos

Podem-se examinar as partes, por redução a relações simples. Contudo, quando se analisam em conjunto, formam um todo que se comporta de modo diferente, não se limitando a ser um simples somatório. Um exemplo prático desta realidade é o estudo dos órgãos nos organismos a que pertencem e fora deles.

A constatação que é impossível separar e examinar os processos dos ecossistemas reais está contida nas frases “tudo está relacionado com tudo” ou “o todo é mais que a soma das partes” [Allen, 1988].

Allen [Allen, 1988] afirma que a última frase está relacionada com o potencial evolutivo que está implícito nos sistemas vivos. O ecossistema contém, nele próprio, a possibilidade de se transformar em algo diferente, ou seja, evoluir e desenvolver-se.



O potencial evolutivo está ligado à existência de liberdade microscópica, resultante da diversidade, complexidade e variabilidade dos seus elementos. A diversidade microscópica junta-se à complexidade, de tal modo que é impossível cobrir todas as possibilidades e detalhes do fenómeno observado. Tenta-se apreender pelo menos uma parte da realidade, com recurso a modelos. Contudo, um só modelo é manifestamente pouco. São necessários muitos modelos, usados em simultâneo para captar a realidade. Este aspecto tem sido reconhecido pela Ecologia dos Sistemas, também designada por Ecologia Holística. A abordagem reducionista tenta compreender as reacções dos ecossistemas analisando um, ou quando muito, alguns processos relacionados com um ou dois componentes. Os resultados são depois extrapolados e utilizados para explicar as observações feitas no ecossistema real mas, muitas vezes, essas extrapolações não são válidas e conduzem a conclusões erradas. A abordagem reducionista não considera que:

- (i) - As condições básicas, determinadas por factores externos, estão sempre a variar o que dificulta ou impossibilita a validação dos resultados no contexto do ecossistema;
- (ii) - As interacções entre processos e componentes podem fazer variar os referidos processos e as propriedades de todos os componentes biológicos, fazendo com que os resultados não sejam válidos; e
- (iii) - Não é possível a supervisão directa de muitos processos simultâneos.

Torna-se, por isso, necessária uma ferramenta para fazer a síntese. A modelação pode ser essa ferramenta, o caminho para se conhecer e compreender os ecossistemas, apontando soluções para os problemas ambientais que ameaçam a sobrevivência.

O problema ambiental básico reside no facto de a espécie humana ter feito imensos progressos, únicos na nossa história, e não terem sido compreendidas, a todos os níveis, as suas consequências. Se pensarmos no desenvolvimento da medicina, sistemas de comunicação, transportes e os compararmos com o que se passava à cem anos atrás as diferenças são abissais. Contudo, na escala da evolução, cem anos é um intervalo de tempo muito curto. Pode dizer-se que este progresso foi favorável à espécie humana mas as consequências para a ecosfera, para as condições globais, para a vida, para o clima e para a entidade Natureza não foram de todo acauteladas. Hoje em dia pode ver-se as consequências mas ainda não estão compreendidas, o que cria dificuldades em direccionar correctamente o desenvolvimento.

O que impede o *Homo Sapiens* de tomar as medidas necessárias para prever as consequências catastróficas e irreversíveis do seu próprio progresso? Anda tão fascinado com as possibilidades de progresso tecnológico que se esquece da Natureza? Será apenas curto de vista? Não queremos que tais questões tenham resposta afirmativa. O *Homo Sapiens* desenvolve-se na Natureza, da Natureza e pela Natureza. Ele ainda não sabe lidar com sistemas de tão grande complexidade e a Natureza apresenta uma complexidade impossível de conceber.

Necessitamos de um esforço científico massivo para aprender a lutar com a complexidade dos sistemas em geral e dos ambientais em particular. Que ferramentas devemos utilizar?; Como as usar com maior eficiência?; Quais as leis gerais que são válidas para os sistemas complexos?; Estarão todos os sistemas complexos organizados hierarquicamente de acordo com as mesmas leis básicas?.

Ulanowicz [Ulanowicz, 1986] sugeriu a descrição holística dos ecossistemas. Holismo significa descrever as propriedades do sistema ao nível do conjunto em vez da descrição exhaustiva de todos os componentes. Adoptando esta metodologia pensa-se que certas propriedades se tornam claras e outras, que de outro modo não eram detectadas, tornam-se visíveis.

A complexidade dos ecossistemas é, como foi referido anteriormente, a principal causa que limita o conhecimento e consequentemente, condiciona as possibilidades de gestão sustentada. Não se é capaz de conhecer todos os pormenores, mas tenta-se conhecer e compreender o comportamento médio e as reacções importantes do sistema, i.e. o conhecimento das propriedades dos ecossistemas não pode ser atingido pela análise dos detalhes, mas sim pelo exame do ecossistema como um todo.

O número de mecanismos de regulação é extremamente elevado e possibilita que os organismos e as populações sobrevivam e se reproduzam, apesar das variações das condições externas (níveis 3 e 4 da Tabela 2.3). O fitoplâncton, por exemplo, é capaz de ajustar a concentração de clorofila-*a* de acordo com a radiação solar, além disso pode apresentar diversos tamanhos, em ambientes diferentes, em função do que for mais benéfico para o crescimento e para a sobrevivência [Jørgensen, 1999].

Os mecanismos de regulação estão, eles próprios, em constante mudança, ou seja, a adaptação é ela própria adaptável. Assim se os mecanismos de regulação actuais não forem suficientes ocorrerá outro processo de regulação, hierarquicamente superior. O exemplo dado anteriormente, referente à alteração de tamanho dentro de uma mesma espé-

cie, é um processo limitado e quando o fim é atingido, outras espécies surgirão. Deste modo, quando é necessário encontrar uma melhor utilização dos recursos disponíveis, não somente os processos e os componentes são substituídos como também os mecanismos de regulação [Jørgensen e Nielsen, 1994] [Jørgensen, 1997].

Têm sido utilizados três conceitos diferentes para explicar o funcionamento dos ecossistemas:

- (i) - O conceito individualista - assume que as populações respondem a um ambiente externo de forma independente;
- (ii) - O conceito de super-organismo - Vê o ecossistema como um organismo de ordem superior [Margalef, 1968]. O ecossistema e os organismos são, contudo, diferentes num aspecto muito importante. Os ecossistemas podem ser desmantelados, sem todavia serem destruídos, i.e. são substituídos por outros. Patten [Patten, 1991] sublinhou que nos ecossistemas tanto os efeitos directos como os indirectos são significativos enquanto nos organismos apenas os efeitos directos são relevantes; e
- (iii) - O conceito das hierarquias - É um compromisso entre os dois conceitos anteriormente apresentados. Os sistemas de ordem superior apresentam propriedades emergentes, que se destacam das propriedades dos seus componentes [Allen e Star, 1982]. De acordo com O'Neill [O'Neill et al., 1986] este conceito parece ser uma boa forma de descrever e compreender os ecossistemas.

O passado dos ecossistemas e dos seus componentes é importante para prever as suas reacções e o seu futuro desenvolvimento. A importância dos aspectos históricos sublinha a necessidade de abordagens dinâmicas que suportem a ideia de não se observar duas vezes a mesma situação [Patten, 1983] [Patten, 1997].

Os ecossistemas são dinâmicos. Todos os seus componentes no geral, e os biológicos em particular, estão em mudança, as suas propriedades são transitórias e impedem o ecossistema de retornar à situação anterior.

A heterogeneidade e as variações no espaço e no tempo dificultam a apreensão dos aspectos essenciais do ecossistema e conseqüentemente, a sua modelação. Contudo, a teoria das hierarquias aplica estas variações no desenvolvimento de uma hierarquia natural, facilitando o estudo e a modelação dos ecossistemas.

A teoria de Darwin descreve a competição entre as espécies e estabelece que as melhor adaptadas, às condições apresentadas pelo ecossistema, irão sobreviver. A referida teoria descreve as alterações na estrutura ecológica e na composição das espécies mas não pode ser aplicada quantitativamente [Jørgensen e Nielsen, 1994] [Brown, 1995] [Jørgensen, 1997].

Todas as espécies presentes no ecossistema são confrontadas com a possibilidade de sobreviver e/ou crescer nas condições vigentes. Qualquer variação numa propriedade de uma espécie influenciará a evolução das outras espécies, estando portanto, implícito o conceito de coevolução [Jørgensen e Nielsen, 1994] [Jørgensen, 1997].

As espécies são, de um modo geral, mais sensíveis ao *stress* do que as propriedades funcionais do ecossistema. Schindler [Schindler, 1988] verificou que as propriedades funcionais, como a produção primária e a respiração, eram praticamente insensíveis à acidificação, enquanto observou efeitos ao nível da composição e da morfologia das espécies.

Todos os factores naturais são dinâmicos, estão em permanente mudança, havendo sempre espécies prontas para tomarem a dianteira e tornarem-se predominantes, no caso de estarem melhor ajustadas do que as que prevaleciam até aí. As recombinações sexuais, as mutações e o resultado dos processos de auto-organização garantem que num ecossistema exista um espectro de espécies, representando diferentes combinações de propriedades [Kauffman, 1996]. A questão que se levanta prende-se em saber quais as espécies mais capazes de sobreviver e crescer nas condições existentes e quais serão as melhor adaptadas no futuro (Figura 2.2).

A evolução é o resultado da relação entre a dinâmica dos factores externos e a dinâmica das espécies disponíveis. Os factores externos fazem variar as condições e as espécies disponíveis fornecem as soluções para o problema da sobrevivência.

A teoria de Darwin assume que as populações consistem em indivíduos que apresentam as propriedades de reprodução elevada, de herança e de variabilidade. A primeira prende-se com o facto de as populações produzirem, em média, mais descendência do que a necessária para substituir os que morrem. A segunda relaciona-se com o facto de a descendência se assemelhar mais com os progenitores do que com indivíduos escolhidos ao acaso, entre a população. A terceira está relacionada com a existência, dentro da mesma espécie, de indivíduos com propriedades diferentes [Stenseth, 1986] [Brown, 1995] [Jørgensen, 1997].

As três propriedades anteriormente referidas estão integradas na Figura 2.2. A reprodução elevada é necessária para se conseguir variações na composição relativa das espécies. A variabilidade está representada nas variações, a curto e a médio prazo, das espécies disponíveis enquanto que a herança é necessária para se observarem efeitos, a longo prazo, na adaptação. Sem herança, as novas gerações partiam do mesmo ponto e não era possível manter os resultados da adaptação.

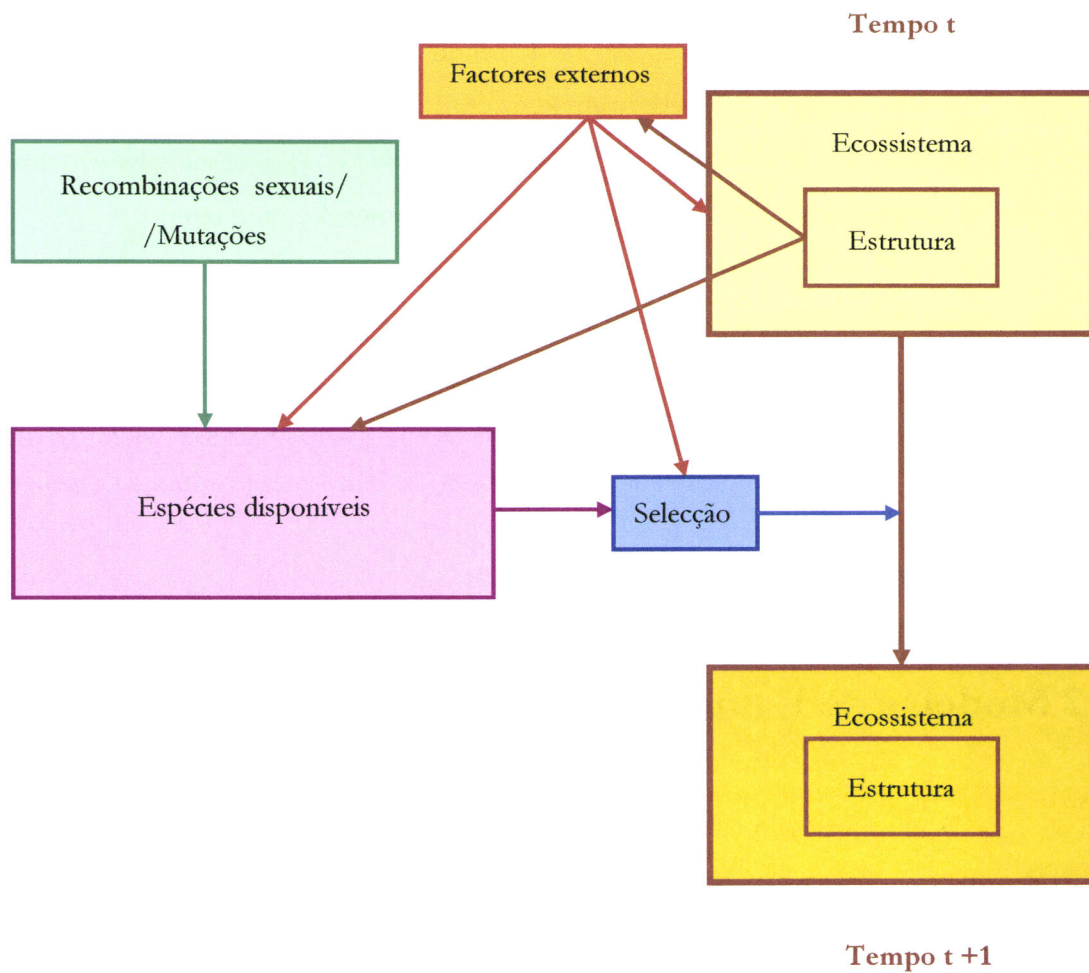


Figura 2.2: Influência dos factores externos sobre a composição do ecossistema.

As espécies são constantemente testadas em relação às condições que prevalecem em cada momento. A velocidade de crescimento de uma população pode ser utilizada para caracterizar a adaptação [Stenseth, 1986] [Brown, 1995]. Contudo, a adaptação tem que ser hereditária para ter efeitos a longo prazo na composição das espécies e na estrutura dos ecossistemas.

A teoria da selecção natural tem sido criticada por ser uma tautologia. A adaptação é medida pela sobrevivência e a sobrevivência dos mais aptos significa, por seu lado, a adaptação dos sobreviventes. Contudo, a teoria completa de Darwin, que inclui as três propriedades acima referidas, não deve ser rotulada desse modo. Existem diversas respostas para o problema da sobrevivência, tendo maior probabilidade de sobreviver e de crescer as espécies que possuam, nas condições existentes, as melhores combinações de propriedades [Ulanowicz, 1986] [Jørgensen, 1997].

Os registos fósseis mostram o aumento da diversidade das espécies e, por isso mesmo, testemunham que a evolução avança no sentido de a longo prazo, aumentar a complexidade. Existem contudo, forças destrutivas de curto prazo como, por exemplo, a poluição antropogénica e as catástrofes naturais, mas a probabilidade de se desenvolverem novas espécies e de serem utilizados novos nichos ecológicos aumenta com o tempo.

Importa salientar que a estrutura biológica é mais que um sistema não linear activo. No decurso da evolução, a estrutura biológica é constantemente alterada e deste modo, o mapa estrutural é ele próprio modificado [Schoffeniels, 1976]. A estrutura como um todo, torna-se assim a representação de toda a informação recebida. A componente biológica representa, através da sua complexidade, a síntese de toda a informação com que esteve em contacto.

## 2.2 Modelos de Estrutura Dinâmica

Os modelos incluídos nas quatro primeiras gerações apresentam estruturas rígidas, assentes num conjunto fixo de parâmetros. Incluem apenas os três primeiros níveis da hierarquia da regulação (Tabela 2.3), deixando de fora os outros, i.e. não consideram a capacidade dos ecossistemas de alterar e/ou substituir componentes.

Os níveis 4 e 5 da Tabela 2.3 (i.e. a adaptação de propriedades e selecção de outras espécies) têm sido introduzidos recentemente na modelação, através da inclusão de parâmetros que possam variar de acordo com uma função objectivo de natureza ecológica. O processo consiste em fazer variar os parâmetros mais importantes de forma a maximizar a referida função objectivo, sendo esse conjunto de valores escolhido para ser utilizado no modelo.

Várias funções objectivo têm sido propostas por diversos autores, como se pode constatar na Tabela 2.4. Contudo, apenas alguns modelos consideram as variações na composição das espécies e/ou a capacidade das espécies modificarem as suas propriedades, dentro de certos limites.

Tabela 2.4: Algumas das funções objectivo propostas.

<b>Tipo de sistema</b>	<b>Função objectivo</b>	<b>Referência</b>
Vários	Potência útil máxima ou fluxo de energia útil máximo	[Lotka, 1924] [Odum e Pinkerton, 1955]
Vários	Mínimo de entropia	[Glansdorff e Prigogine, 1971]
Vários	Máximo de exergia	[Mejer e Jørgensen, 1979] [Salomonsen e Jensen, 1996] [Ray et al., 2001]
Ambiental	Máximo de matéria orgânica persistente	[Whittaker et al., 1971] [O'Neil et al., 1975]
Ambiental	Máximo de biomassa	[Margalef, 1968] [Straškraba, 1979]

Bossel [Bossel, 1992] destacou os aspectos fundamentais para desenvolver modelos que possam descrever correctamente as capacidades dos ecossistemas:

- (i) - Existência - o ecossistema não pode apresentar quaisquer condições que possam mover as variáveis de estado para fora de zonas seguras;
- (ii) - Eficiência - o ganho de exergia a partir do exterior (ambiente) deve exceder sempre a quantidade dispendida;
- (iii) - Liberdade de acção - o sistema reage às variáveis externas com uma certa variabilidade;
- (iv) - Segurança - o sistema tem que fazer face a diferentes ameaças à sua segurança com estratégias diferentes e adequadas. Essas estratégias podem reflectir-se tanto ao nível das variáveis de estado como ao nível da estrutura interna;
- (v) - Adaptabilidade - se o sistema não puder evitar os factores que ameaçam a sua continuidade tem a possibilidade de se modificar para melhor se ajustar às novas condições; e

(vi) - Ter em conta os outros sistemas - o sistema responde ao comportamento dos outros sistemas.

Bossel [Bossel, 1992] aplicou o método da maximização do benefício, para seleccionar a estrutura de um modelo dinâmico, capaz de ter em conta as propriedades estruturais apresentadas na Tabela 2.3. O método, apesar de promissor ainda não foi explorado no que concerne a sistemas ambientais.

Straškraba [Straškraba, 1979] utilizou como princípio governante, a maximização da biomassa. O modelo possui uma rotina que calcula a biomassa para todas as combinações de parâmetros, dentro de um intervalo considerado realista, e selecciona o conjunto que origina, a cada instante, o valor máximo de biomassa.

A função objectivo mais utilizada nos modelos ambientais é contudo, a exergia. A teoria de Darwin pode ser reformulada, em termos de exergia, i.e. nas condições vigentes, um ecossistema em permanente evolução, selecciona as espécies que mais possam contribuir para a conservação (i.e. a sobrevivência) ou para o aumento da exergia (i.e. o crescimento) [Jørgensen e Nielsen, 1994] [Jørgensen, 1997].

Não é possível medir a exergia directamente mas, se a composição do sistema for conhecida, é possível calculá-la através da equação [Jørgensen e Mejer, 1979] [Jørgensen et al., 1995] [Jørgensen, 1988] [Jørgensen, 1997]:

$$Ex = RT \sum_{i=1}^{i=n} \left[ C_i \ln \left( \frac{C_i}{C_{i,eq}} \right) - (C_i - C_{i,eq}) \right] = \sum_{i=1}^{i=n} [\beta_i C_i] \quad (2.1)$$

onde  $R$  denota a constante dos gases ideais,  $T$  denota a temperatura,  $C_i$  denota a concentração do componente  $i$ , expressa em unidades adequadas,  $C_{i,eq}$  denota a concentração do componente  $i$  no equilíbrio e  $n$  denota o número de componentes.

A filosofia subjacente à quinta geração de modelos, reside em determinar permanentemente novos conjuntos de valores para os parâmetros mais importantes que traduzam as condições existentes no ecossistema (Figura 2.3).

A exergia já foi utilizada anteriormente como função objectivo em modelação [Jørgensen e Mejer, 1979] [Jørgensen, 1986] [Herendeen, 1989]. Não obstante, esses modelos não traduzem a flexibilidade dos ecossistemas uma vez que não utilizam parâmetros variáveis.



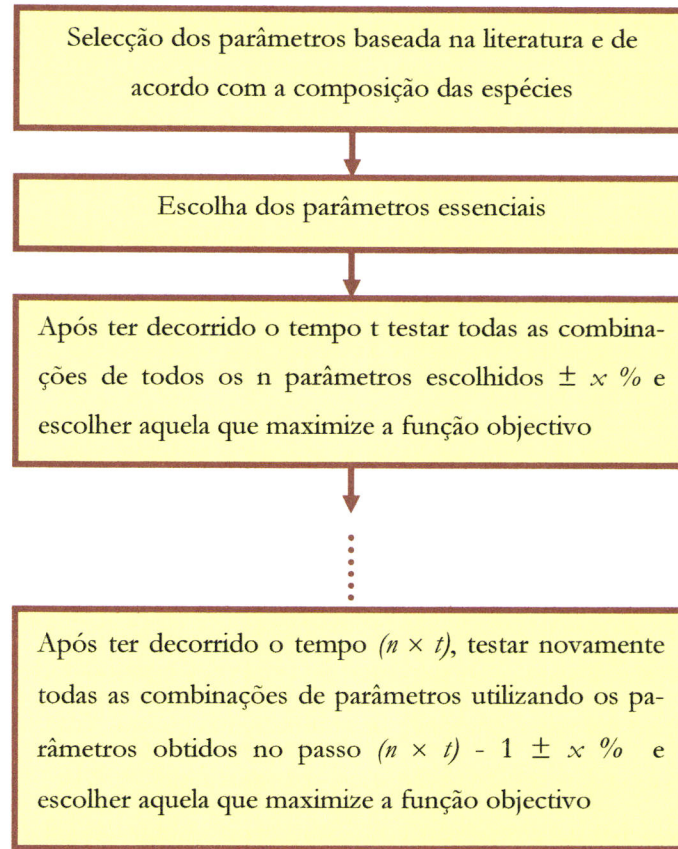


Figura 2.3: Esquema do procedimento utilizado no desenvolvimento de um modelo de estrutura dinâmica.

### 2.2.1 Aplicação a Ecossistemas Aquáticos

O cálculo do valor de exergia para fazer a selecção contínua dos parâmetros, foi aplicado a poucos casos respeitantes a ecossistemas aquáticos. Um dos casos em que o modelo já se encontra calibrado e validado diz respeito ao lago Søbygaard e apresenta-se em seguida.

O lago Søbygaard foi estudado entre 1982 e 1988. Durante esse período conheceu uma redução significativa nas entradas de nutrientes e o ecossistema sofreu alterações estruturais pronunciadas.

O lago Søbygaard é um lago pouco profundo, com um tempo de retenção curto. A entrada de nutrientes foi reduzida, de forma muito significativa, após 1982 (e.g. as entradas de fósforo passaram de  $30 \text{ g m}^{-2} \text{ ano}^{-1}$  para  $5 \text{ g m}^{-2} \text{ ano}^{-1}$ ) [Søndergaard, 1989] [Jeppesen et al., 1990].

Apesar da entrada de nutrientes ter sido reduzida, isso não se reflectiu quer na sua concentração quer na concentração de clorofila-*a*, durante o período de 1982 a 1985, devido à libertação de nutrientes acumulados nos sedimentos. Somente a partir de 1985 se observaram variações, sendo de destacar a diminuição da quantidade de peixes planctívoros e a diminuição da concentração média de fitoplâncton (passou de 700  $\mu\text{g dm}^{-3}$  em 1985 para 150  $\mu\text{g dm}^{-3}$  em 1988) acompanhada de um aumento de tamanho (passou de cerca de 100  $\mu\text{m}^3$  para valores situados entre 500 e 1000  $\mu\text{m}^3$ ) [Søndergaard, 1989] [Jeppesen et al., 1990].

O modelo desenvolvido considera apenas o ciclo do nutriente limitante, (i.e. o azoto), apresentando 6 variáveis de estado (i.e. a concentração de azoto solúvel e as concentrações de azoto presentes nos peixes, no zooplâncton, no fitoplâncton, nos detritos e nos sedimentos).

Foram escolhidos, como parâmetros variáveis, a velocidade máxima de crescimento do fitoplâncton e a sua velocidade de sedimentação, a que foram atribuídos, de acordo com a calibração inicial, respectivamente os valores 2,2  $\text{dia}^{-1}$  e 0,15  $\text{m dia}^{-1}$ .

O modelo foi desenvolvido com base no esquema apresentado na Figura 2.3, considerando a exergia como a função objectivo a ser maximizada. O intervalo de tempo escolhido foi de 5 dias e foi utilizado  $x\% = 10\%$ .

As previsões do modelo estão de acordo com o observado no que diz respeito à taxa de crescimento do fitoplâncton e à concentração de clorofila-*a*. Contudo, no que diz respeito à velocidade de sedimentação os valores previstos são numericamente mais baixos, embora tenha sido previsto um aumento da mesma magnitude, i.e. para o triplo (Tabela 2.5)[Søndergaard, 1989] [Jeppesen et al., 1990].

Tabela 2.5: Comparação entre as previsões do modelo e os resultados experimentais referentes ao lago Søbygaard.

Ano	Taxa máxima de crescimento do fitoplâncton/ $\text{dia}^{-1}$		Concentração de clorofila/ $\mu\text{gdm}^{-3}$		Velocidade de sedimentação do fitoplâncton/ $\text{m dia}^{-1}$	
	Previsto	Observado	Previsto	Observado	Previsto	Observado
1985	2,0	2,2	600	610	0,15	0,2
1988	1,2	1,1	200	195	0,45	0,6

Os autores, para ilustrarem a importância deste tipo de modelos, utilizaram os valores dos parâmetros referentes a 1988 nas condições de 1985, tendo o modelo previsto um comportamento caótico para o sistema. [Søndergaard, 1989] [Jeppesen et al., 1990].

A estimativa do valor dos parâmetros é, muitas vezes, o ponto mais fraco dos modelos quer por o número de observações ser insuficiente, quer por falta de informação na literatura, quer por serem conhecidos com pouca precisão. Contudo, estas dificuldades podem ser reduzidas, no caso da maioria dos valores dos parâmetros ser conhecida. Os valores em falta podem ser determinados calculando, no universo dos valores possíveis, a combinação que maximize a exergia. No caso de serem utilizados valores ligeiramente mais elevados os modelos podem prever comportamentos caóticos, tal como se verificou no exemplo anteriormente apresentado. O conjunto de valores que maximiza a função exergia pode ser tomado como a “fronteira do caos” [Jørgensen, 1995b] [Kauffman, 1996].

## 2.3 Modelos Baseados na Teoria do Caos

A teoria do caos está relacionada com o curso imprevisível dos acontecimentos. A noção de que uma borboleta que agite hoje o ar em Pequim pode influenciar tempestades no próximo mês em Nova Iorque, ilustra a referida teoria.

O efeito borboleta foi descoberto acidentalmente por Lorentz em 1961, quando trabalhava em modelos de previsão do tempo. Querendo examinar uma sequência longa com maior rapidez, abreviou o processo utilizando apenas três casas decimais em vez de seis. A nova série (que deveria reproduzir a precedente) divergia tão rapidamente que em poucos meses as semelhanças desapareciam. O modelo de Lorentz apresentava aquilo que se passou a designar por dependência sensível das condições iniciais [Gleick, 1989].

A natureza procura evitar rupturas e os ecossistemas têm muitas possibilidades de o fazer já que estão hierarquicamente organizados e possuem mecanismos de regulação. Não obstante, isso não significa que situações caóticas ou quase caóticas não se observem [Hasting et al., 1993] [Suárez, 1999] [Vayenas e Pavlou, 2001].

## 2.4 Modelos Baseados na Teoria da Catástrofe

A teoria da catástrofe, aplicada no seu sentido estrito é uma teoria de equilíbrio, que lida com pontos de atracção ou atractores. Um sistema dinâmico (governado por uma função potencial escalar dependente de, no máximo, 5 variáveis externas) pode ser modelado por uma de sete funções canónicas, que podem ser deduzidas analiticamente a partir da função potencial por transformações de coordenadas e outras técnicas matemáticas [Poston e Stewart, 1978].

Esta teoria tem sido aplicada em diversos campos do saber, tais como as Ciências Sociais, a Medicina, a Ecologia e a Economia [Zeeman, 1978] [Poston e Stewart, 1978] [Kempf, 1980] [Loehle, 1989].

Os sistemas vivos desenvolvem mecanismos para lidar com o *stress* provocado pelas modificações ambientais. Um desses mecanismos é a súbita variação de propriedades, a qual pode ser modelada recorrendo à teoria da catástrofe. Este tipo de comportamento não é, portanto, necessariamente negativo, pode reflectir uma adaptação rápida a uma nova situação. Além disso, muitos sistemas tiram partido das condições ambientais severas para testar a capacidade de sobrevivência dos seus componentes e/ou eliminar os mais fracos.

Dubois [Dubois, 1979] explicou as variações observadas na concentração de oxigénio de rios do sul da Bélgica, durante a Primavera e o Outono utilizando a teoria da catástrofe. Considerou que a variação da concentração de oxigénio ao longo do tempo era dada por:

$$\frac{dC(t)}{dt} = \underbrace{K_a(C_s - C(t))}_{\text{rearejamento}} + \underbrace{k_3 C(t)(1 - q \cdot C(t))}_{\text{fotossíntese}} - \underbrace{\frac{k_2 C(t)}{C(t) + k_1}}_{\text{respiração}} \quad (2.2)$$

onde  $C(t)$  e  $C_s$  denotam, respectivamente, a concentração de oxigénio no instante  $t$  e a concentração saturada de oxigénio à temperatura  $T$ ,  $K_a$  denota a constante de rearejamento e  $q$ ,  $k_1$ ,  $k_2$  e  $k_3$  denotam constantes.

Considerando  $x = \frac{C(t)}{k_1}$ ,  $a(T) = K_a C_s$ ,  $b = k_3 - K_a$ ,  $c = \frac{q \cdot k_3 \cdot k_1}{b}$  e  $d = \frac{k_2}{k_1}$  ob-

teve:

$$\frac{dx}{dt} = a(T) + bx(1 - cx) - \frac{dx}{1 + x} \quad (2.3)$$

A Figura 2.4 representa a variação de  $-\frac{dx}{dt}$  em função de  $x$  para seis valores de  $a(T)$  que ocorrem em seis alturas diferentes do ano, verificando-se que:

- (i) - Para  $a(T) = 0,5$  há apenas um atrator, correspondente a  $-\frac{dx}{dt} = 0$  e que ocorre quando  $x = S$ ;
- (ii) - Para  $a(T) = 1,0$  existem dois atratores,  $x = S$  e  $x = Q$ , permanecendo  $x$  em  $S$ , só saltando para o atrator,  $Q$  quando  $a(T) = 1,2$ ;
- (iii) - No caso de  $a(T) \geq 1,3$  o atrator  $Q$  é único; e
- (iv) - Quando se faz regressar  $a(T)$  ao valor 1,0 existem, tal como anteriormente, novamente dois atratores  $x = S$  e  $x = Q$ . Contudo,  $x$  permanece em  $Q$  e só para  $a(T) = 0,75$  é que retorna ao atrator  $S$ .

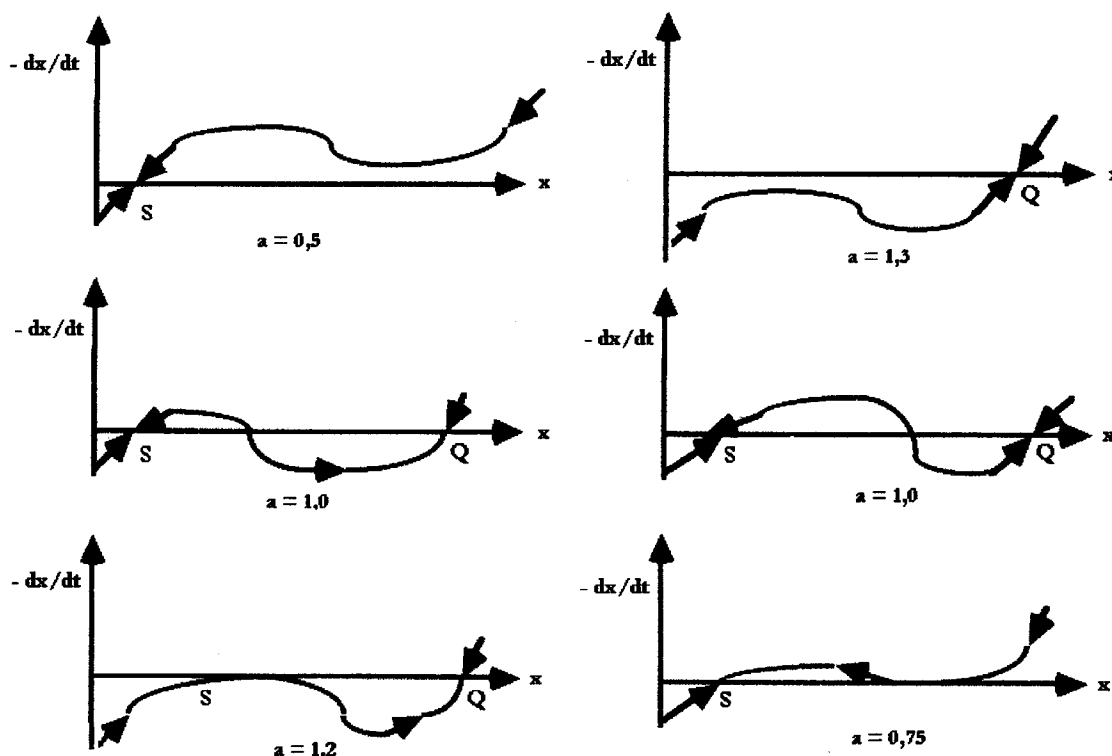


Figura 2.4:  $-\frac{dx}{dt}$  em função de  $x$  para seis valores de  $a$ .  $S$  e  $Q$  são atratores. As setas indicam a evolução de  $x$ . Adaptada de Dubois [Dubois, 1979].

Aumentando  $a(T)$  (i.e. durante os meses correspondentes à Primavera), verifica-se que a variação brusca da concentração de oxigénio ocorre quando  $a(T) = 1,2$ . Quando começa a ocorrer diminuição da temperatura (i.e. no Outono) o retorno à situação anterior só acontece quando  $a(T) = 0,75$ . Este efeito de histerese está representado na Figura 2.5.

A Figura 2.5 pode ser explicada considerando que perante concentrações elevadas de matéria orgânica, ocorre consumo de oxigénio tornando-se a sua presença dependente dos processos de rearejamento. No caso da água estar menos poluída o consumo de oxigénio será então menor e, conseqüentemente, a sua presença não estará tão dependente do rearejamento.

Dubois [Dubois, 1979] calculou a capacidade tampão,  $\beta$ , dada pela expressão  $\beta = \left( \frac{\partial C(t)}{\partial T} \right)^{-1}$  e representou  $\log(-\beta)$  em função do tempo (Figura 2.6). A Figura 2.6 mostra, tal como seria de esperar, que a capacidade tampão mais baixa coincide com as variações bruscas na concentração de oxigénio. Verifica-se ainda a existência de picos onde a capacidade tampão é muito elevada, o que significa que é muito difícil aumentar a concentração de oxigénio, o que corresponde ao período de Verão.

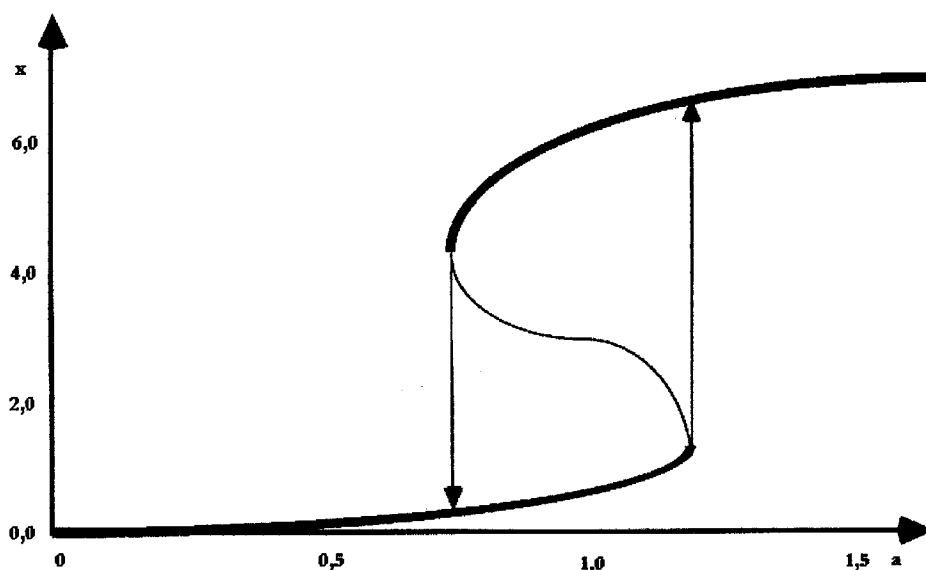


Figura 2.5: Representação de  $x$  em função de  $a$ . As setas destacam o efeito de histerese. Adaptada de Dubois [Dubois, 1979].

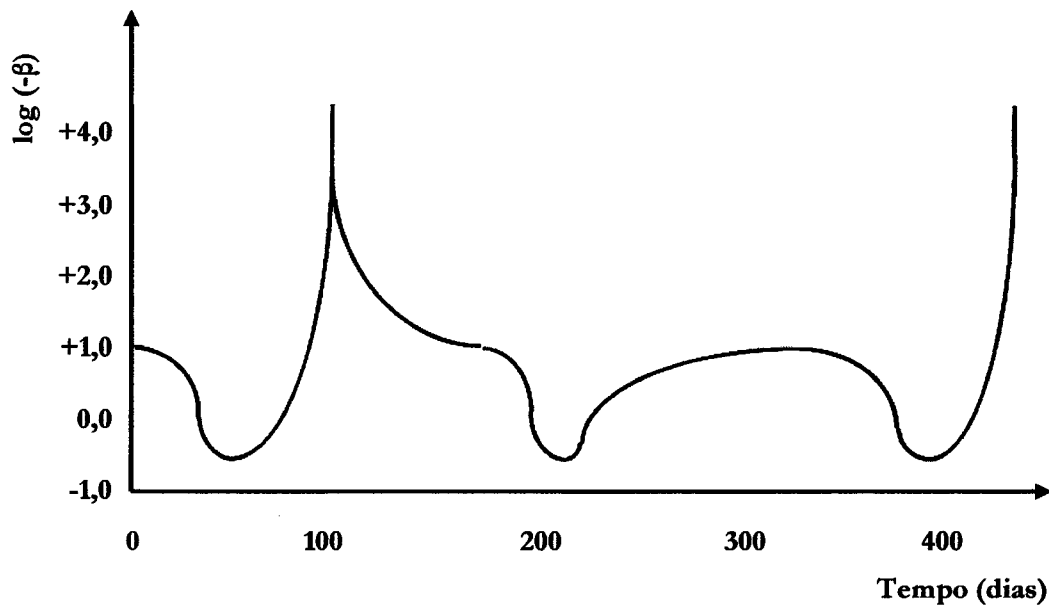


Figura 2.6: Capacidade ecológica tampão em função do tempo. Adaptada de Dubois [Dubois, 1979].

O caso apresentado é um exemplo ilustrativo da estabilidade descontínua exibida pelos ecossistemas e da forma como esta pode ser modelada. Contudo, não existem muitos casos de aplicação da teoria da catástrofe ao domínio das Ciências Ambientais. De acordo com Loehle [Loehle, 1989] existem dois factores capazes de explicar tal facto. Por um lado o facto de a base da teoria, a Topologia, ser uma área muito específica da Matemática e, por outro lado o facto de os procedimentos a seguir não estarem explicados, na literatura especializada, de forma a que especialistas de fora dessa área os possam utilizar.

Os ecossistemas tendem para um estado estacionário. Contudo, devido ao facto de os factores externos e das propriedades das espécies sofrerem variações permanentes o estado estacionário é variável. A Figura 2.5 ilustra o que se acabou de referir, i.e. a mesma ou praticamente a mesma combinação de factores externos origina estados estacionários diferentes. A escolha entre dois estados estacionários possíveis, depende da história recente do ecossistema.

A histerese pode também ser explicada pelo facto de o sistema manter a capacidade tampão tão elevada quanto é possível, de modo a que o retorno à situação anterior seja retardado o mais possível.

## 2.5 Outras Técnicas de Modelação

Neste ponto vamos referir, embora muito brevemente, duas técnicas de modelação que apresentam muitas potencialidades de aplicação a ecossistemas, nomeadamente os modelos recorrendo à programação orientada ao objecto e os modelos de base individual. Por detrás da aplicação destas técnicas de modelação, está subjacente, o facto de se reconhecer lacunas nos dados e a necessidade de modelos mais flexíveis.

Uma outra técnica de modelação que utiliza Sistemas Inteligentes e a Inteligência Artificial também poderia ser incluída neste ponto. Contudo, como a utilização desta técnica é um dos objectivos do trabalho será tratada noutro ponto.

### 2.5.1 Modelos Recorrendo à Programação Orientada ao Objecto

Os modelos que utilizam a programação orientada ao objecto, são baseados na ideia de que os programas podem descrever interacções entre representações abstractas de objectos reais, em vez de sequências de cálculos habitualmente associados à programação, referida como programação processual [Silvert, 1993].

O conceito central da programação orientada ao objecto é o conceito de classe, que descreve tanto a estrutura de um objecto como o conjunto de procedimentos para iniciá-lo e utilizá-lo no modelo. Um exemplo de classe é a definição de uma população, que constitui um bloco básico para a construção de muitos modelos em ecologia. Uma população é caracterizada por variáveis como, por exemplo, o tamanho médio, a idade, o número de indivíduos e apresenta processos dos quais se salientam a reprodução, o crescimento e a mortalidade. Cada população é única embora haja muitas semelhanças. Deste modo, pode tratar-se diferentes populações por adição dos aspectos particulares que necessitam de ser diferentes no contexto do modelo [Ferreira, 1995] [Sequeira et al., 1997].

A programação orientada ao objecto define procedimentos diferentes em módulos diferentes, que podem ser utilizados por várias classes. Permite ter várias versões de um mesmo processo (e.g. várias rotinas de crescimento) [Ferreira, 1995] [Sequeira et al., 1997].



Por vezes é difícil separar os processos relevantes daqueles que operam a um nível diferente e que não deveriam ser incluídos no modelo. Este tipo de programação oferece mecanismos que permitem esconder esta informação.

A hierarquia pode ser construída, por exemplo, descrevendo primeiro *populações*, depois *plantas*, a seguir *algas* e finalmente a *espécie*. Daqui resulta uma hierarquia que contém quatro classes, cada uma delas baseada na anterior. Em cada estágio é possível adicionar ou modificar informação adaptado-a ao nível da descrição. Por exemplo, a classe *plantas* deve incluir outros parâmetros, para além daqueles que são partilhados por todas as *populações* (e.g. a velocidade de crescimento e a capacidade de transporte). A classe *algas* partilha todas propriedades das classes anteriores mas deverá, além disso, reflectir a dependência em relação aos nutrientes, o que implica a redefinição, nesta classe, do crescimento. Por fim, a classe respeitante às *espécies* deve ter informação específica sobre a espécie em questão (e.g. a velocidade de sedimentação). Contudo todas as espécies herdam as propriedades comuns das classes *algas*, *plantas* e *populações*.

A programação orientada ao objecto oferece várias vantagens para a modelação de ecossistemas, entre as quais se destaca a ligação estreita entre objectos e classes. Permite desenvolver modelos relativamente simples de interpretar e que podem ser facilmente modificados. Na literatura encontram-se alguns exemplos da aplicação deste tipo de modelação a ecossistemas [Meyer e Pampagnin, 1979] [Muetzelfeldt, 1979] [Sequeira et al., 1991] [Baveco e Lingeman, 1992] [Silvert, 1993] [Ferreira, 1995] [Sequeira et al., 1997].

Downing e Reed [Downing e Reed, 1996] desenvolveram um modelo de migração animal utilizando esta técnica de modelação. A partir da classe *animal* desenvolveram a hierarquia apresentada na Figura 2.7. É utilizado o conceito de herança múltipla, ou seja a possibilidade de relação directa com mais do que uma classe de nível superior. Por exemplo, a classe *macho imaturo* herda as propriedades da classe *macho* e da classe *imaturo*. A classe *foca fêmea reprodutiva* herda as propriedades das classes *foca* e *fêmea reprodutiva*.

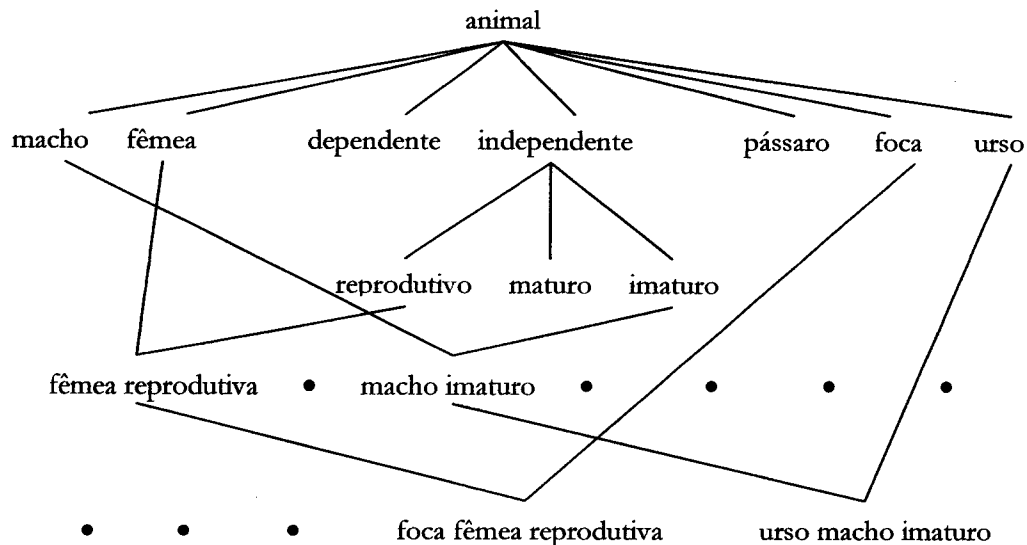


Figura 2.7: Ligações entre as classes num modelo de migração animal. Adaptada de Downing e Reed [Downing e Reed, 1996].

## 2.5.2 Modelos de Base Individual

Os modelos de base individual ou modelos orientados ao indivíduo ou agente, tentam traduzir a enorme variabilidade entre indivíduos, habitualmente representada por uma variável de estado. A modelação orientada ao indivíduo reconhece dois princípios básicos que são ignorados na maioria dos modelos ambientais, i.e. a individualidade de cada membro da população e o carácter localizado das interacções entre indivíduos. Sem desigualdades entre os membros de uma população, não seria possível nem a existência de competição nem o processamento da informação local [Mamedov e Udalov, 2002].

A defesa deste método tem sido feita com base no confronto dos pontos de vista reducionista e holístico. Contudo, o facto de os ecossistemas apresentarem as duas propriedades referidas, não invalida que também apresentem propriedades que não podem ser deduzidas a partir da soma dos componentes. Deste modo, torna-se necessário considerar quais as simplificações que se podem fazer para cada situação concreta a modelar, i.e. verificar em que situações é possível excluir a individualidade e/ou o carácter localizado das interacções [Lorek e Sonnenschein, 1998] [Grimm, 1999] [Lorek e Sonnenschein, 1999].

Não se vai desenvolver esta técnica uma vez que ela tem sido aplicada a situações que caem fora do âmbito deste trabalho. Contudo, De Angelis e Gross [De Angelis e

Gross, 1992] e Grimm [Grimm, 1999] apresentam vários exemplos de aplicação deste tipo de modelos.

## 2.6 Caracterização do Estado Trófico das Albufeiras Estudadas

A estabilização da qualidade da água de um reservatório estende-se por vários anos, após a sua construção, sendo a sua duração dependente de factores individuais tais como, a qualidade da água afluente, o volume de água armazenado, a sua utilização e as condições climáticas.

O envelhecimento de um reservatório vai depender das entradas de azoto e de fósforo que recebe, sendo tanto mais rápido quanto maiores forem essas cargas. O conceito de estado trófico de uma massa de água engloba, não só, as concentrações de nutrientes, mas também a sua produtividade.

Neste estudo a avaliação do estado trófico das albufeiras é feito segundo o critério adoptado pelo Instituto da Água (INAG), de acordo com as normas de 1982 da OCDE (Tabela 2.6). O referido critério tem por base três parâmetros, dois de natureza química (i.e. a concentração de fósforo total e o teor de oxigénio dissolvido) e um de natureza biológica (i.e. a concentração de clorofila-*a*). De acordo com o referido critério a massa de água pode ser classificada como ultra-oligotrófica, oligotrófica, mesotrófica, eutrófica e hiper-eutrófica. A atribuição faz-se parâmetro a parâmetro dando-se à massa de água a classificação do parâmetro mais desfavorável [URL 1] [DRAOTA, 2003] [Matoso et al., 2004].

Para classificar o estado trófico das albufeiras em estudo programou-se uma folha de cálculo do *Microsoft Excel* de forma a obter directamente a classificação da massa de água. De acordo com os resultados obtidos para os últimos anos hidrológicos, no que concerne às albufeiras tratadas neste estudo, estas apresentam-se eutrofizadas (Tabela 2.7).

Tabela 2.6: Critérios utilizados para obtenção do estado trófico de uma albufeira.

Parâmetros	Ultra- -oligotrófica	Oligotrófica	Mesotrófica	Eutrófica	Hiper- -eutrófica
Fósforo total/ $\mu\text{g}_\text{P}\text{dm}^{-3}$	$\leq 4$	$> 4$ $\leq 10$	$> 10$ $\leq 35$	$> 35$ $\leq 100$	$> 100$
Clorofila- <i>a</i> / $\mu\text{g dm}^{-3}$	$\leq 1$	$> 1$ $\leq 2,5$	$> 2,5$ $\leq 10$	$> 10$ $\leq 25$	$> 25$
Oxigénio dissolvido/ %					$< 40$

Tabela 2.7: Estado trófico das três albufeiras em estudo nos últimos anos hidrológicos.

Ano Hidrológico	Divôr	Monte Novo	Vigia
1998-1999	eutrófico	hiper-eutrófico	hiper-eutrófico
1999-2000	eutrófico	hiper-eutrófico	hiper-eutrófico
2000-2001	eutrófico	eutrófico	eutrófico
2001-2002	eutrófico	eutrófico	eutrófico

## 2.7 Caracterização da Qualidade da Água das Albufeiras Estudadas

A classificação das massas de água superficiais, no que concerne à sua qualidade, pode fazer-se segundo dois critérios. O primeiro, tem por objectivo definir o tipo de tratamento necessário, para a produção de água para consumo humano. O segundo critério, mais abrangente, classifica as massas de água de acordo com as suas características de qualidade para usos múltiplos.

### **Critério 1 - Classificação das águas brutas destinadas à produção de água para consumo humano**

Segundo este critério, a qualidade das águas doces superficiais, destinadas à produção de água para consumo humano, é obtida parâmetro a parâmetro, de acordo com os intervalos estipulados no anexo I do decreto-lei nº 236/98 de 1 de Agosto (Tabela 2.8). É atribuída à massa de água a classificação do parâmetro mais desfavorável.

Tabela 2.8: Qualidade das águas doces superficiais para produção de água para consumo humano.

Parâmetros	Expressão dos resultados	A 1		A 2		A 3	
		VMR	VMA	VMR	VMA	VMR	VMA
pH	Escala Sorensen	6,5-8,5	-	5,5-9,0	-	5,5-9,0	-
Cor	mg dm <sup>-3</sup> escala Pt-Co	10	20*	50	100*	50	200*
SST	mg dm <sup>-3</sup>	25	-	-	-	-	-
Temperatura	°C	22	25*	22	25*	22	25*
Condutividade	µS cm <sup>-1</sup> a 20°C	1000	-	1000	-	1000	-
Cheiro	Factor de diluição a 25°C	3	-	10	-	20	-
Nitratos**	mg dm <sup>-3</sup>	25	50*	-	50*	-	50*
Fluoretos	mg dm <sup>-3</sup>	0,7-1,0	1,5	0,7-1,7	-	0,7-1,7	-
Ferro dissolvido**	mg dm <sup>-3</sup>	0,1	0,3	0,1	2,0	1,0	-
Manganês**	mg dm <sup>-3</sup>	0,05	-	0,10	-	1,00	-
Cobre	mg dm <sup>-3</sup>	0,02	0,05*	0,05	-	1,00	-
Zinco	mg dm <sup>-3</sup>	0,5	3,0	1,0	5,0	1,0	5,0
Boro	mg dm <sup>-3</sup>	1,0	-	1,0	-	1,0	-
Arsénio	mg dm <sup>-3</sup>	0,01	0,05	-	0,05	0,05	0,10
Cádmio	mg dm <sup>-3</sup>	0,001	0,005	0,001	0,005	0,001	0,005
Crómio total	mg dm <sup>-3</sup>	-	0,05	-	0,05	-	0,05
Chumbo	mg dm <sup>-3</sup>	-	0,05	-	0,05	-	0,05
Selénio	mg dm <sup>-3</sup>	-	0,01	-	0,01	-	0,01
Mercúrio	mg dm <sup>-3</sup>	0,0005	0,001	0,0005	0,001	0,0005	0,001
Bário	mg dm <sup>-3</sup>	-	0,1	-	1,0	-	1,0
Cianetos	mg dm <sup>-3</sup>	-	0,05	-	0,05	-	0,05
Sulfatos	mg dm <sup>-3</sup>	150	250	150	250*	150	250*
Cloreto	mg dm <sup>-3</sup>	200	-	200	-	200	-

Cont.

Tabela 2.8 (cont.): Qualidade das águas doces superficiais para produção de água para consumo humano.

Parâmetros	Expressão dos resultados	A 1		A 2		A 3	
		VMR	VMA	VMR	VMA	VMR	VMA
Detergentes aniônicos	mg dm <sup>-3</sup>	0,2	-	0,2	-	0,5	-
Fosfatos**	mg P <sub>2</sub> O <sub>5</sub> dm <sup>-3</sup>	0,4	-	0,7	-	0,7	-
Fenóis	mg dm <sup>-3</sup>	-	0,001	0,001	0,005	0,010	0,100
Hidrocarbonetos dissolvidos ou emulsionados	mg dm <sup>-3</sup>	-	0,05	-	0,20	0,50	1,00
Hidrocarbonetos aromáticos polinucleares	mg dm <sup>-3</sup>	-	0,2	-	0,2	-	1,0
Pesticidas totais	µg dm <sup>-3</sup>	-	1,0	-	2,5	-	5,0
CQO**	mg O <sub>2</sub> dm <sup>-3</sup>	-	-	-	-	30	-
Oxigênio dissolvido**	% de saturação	70	-	50	-	30	-
CBO5 a 20°C**	mg O <sub>2</sub> dm <sup>-3</sup>	3	-	5	-	7	-
Azoto Kjeldahl	mg N dm <sup>-3</sup>	1	-	2	-	3	-
Azoto amoniacal	mg NH <sub>4</sub> <sup>+</sup> dm <sup>-3</sup>	0,05	-	1,00	1,50	2,00	4,00*
Substâncias extraíveis com clorofórmio	mg dm <sup>-3</sup>	0,1	-	0,2	-	0,5	-
Coliformes totais	n° / 100 cm <sup>-3</sup>	50	-	5000	-	50000	-
Coliformes fecais	n° / 100 cm <sup>-3</sup>	20	-	2000	-	20000	-
Estreptococos fecais	n° / 100 cm <sup>-3</sup>	20	-	1000	-	10000	-
Salmonelas		Ausência em 5 dm <sup>3</sup>	-	Ausência em 5 dm <sup>3</sup>	-	Ausência em 5 dm <sup>3</sup>	-

\* Estes limites podem ser excedidos em caso de condições geográficas ou meteorológicas excepcionais.

\*\* Os limites podem ser excedidos para estes parâmetros em lagos de pouca profundidade e baixa taxa de renovação.

De acordo com este critério, as massas de água são divididas em três classes: A1 (i.e. águas que necessitam de tratamento físico e desinfecção); A2 (i.e. águas que requerem tratamento físico-químico e desinfecção) e; A3 (i.e. águas que necessitam de tratamento físico, tratamento químico de afinação e desinfecção).

Considerando o parâmetro temperatura, as águas superficiais do Alentejo pertencem à classe A3. A exclusão deste parâmetro, contudo, não altera a situação no que concerne às três albufeiras em estudo uma vez que os coliformes totais (albufeira de Monte Novo), a carência bioquímica de oxigénio e o oxigénio dissolvido (albufeira da Vigia) e a carência bioquímica de oxigénio (albufeira do Divôr) são responsáveis pela referida classificação.

Os parâmetros de qualidade das águas doces superficiais destinadas à produção de água para consumo humano dividem-se em três grupos (i.e. G1, G2 e G3). Os parâmetros considerados em cada um dos grupos encontram-se na Tabela 2.9.

A frequência de amostragem e de análise das águas superficiais está regulamentada no decreto-lei nº 236/98 de 1 de Agosto, e depende da classe a que a água pertence e do parâmetro em causa (Tabela 2.10).

Tabela 2.9: Classificação dos parâmetros de qualidade das águas superficiais em grupos.

G1	G2	G3
pH	Ferro dissolvido	Fluoretos
Cor	Manganês	Boro
Sólidos suspensos totais	Cobre	Arsénio
Temperatura	Zinco	Cádmio
Condutividade	Sulfatos	Crómio total
Cheiro	Substâncias tensioactivas	Chumbo
Nitratos	Fenóis	Selénio
Cloretos	Azoto Kjeldahl	Mercúrio
Fosfatos	Estreptococos fecais	Bário
CQO		Cianetos
Oxigénio dissolvido		Hidrocarbonetos dissolvidos e emulsionados
CBO5		Hidrocarbonetos aromáticos polinucleares
Azoto amoniacal		Pesticidas totais
Coliformes totais		Substâncias extraíveis com clorofórmio
Coliformes fecais		Salmonelas

Tabela 2.10: Frequência mínima de amostragem e de análise das águas superficiais.

	Classe da água								
	A1			A2			A3		
	Grupo de parâmetros								
	G1	G2	G3	G1	G2	G3	G1	G2	G3
Frequência mínima de amostragem (número de vezes por ano)	4	2	1	8	4	3	12	6	3

**Critério 2 - Classificação das águas superficiais, de acordo com as suas características de qualidade para usos múltiplos**

Este critério, proposto pelo INAG, classifica as águas superficiais, de acordo com o seu nível de qualidade para usos múltiplos, dividindo-as em cinco classes. Na Tabela 2.11 apresentam-se as várias classes bem como os usos permitidos para cada uma delas.

Tabela 2.11: Usos permitidos por classe de qualidade da água.

Classe de qualidade	Nível de qualidade
A - Sem poluição	Águas consideradas como isentas de poluição, aptas a satisfazer potencialmente as utilizações mais exigentes a nível de qualidade.
B - Fracamente poluída	Águas com qualidade ligeiramente inferior à classe A, mas podendo também satisfazer potencialmente todas as utilizações.
C - Poluída	Águas com qualidade “aceitável”, suficiente para irrigação, para usos industriais e produção de água potável após tratamento rigoroso. Permite a existência de vida piscícola (espécies menos exigentes) mas com reprodução aleatória. Apta para recreio sem contacto directo.
D - Muito poluída	Águas com qualidade “mediocre”, apenas potencialmente aptas para irrigação, arrefecimento e navegação. A vida piscícola pode subsistir, mas de forma aleatória.
E - Extremamente poluída	Águas ultrapassando o valor máximo da classe D para um ou mais parâmetros. São consideradas como inadequadas para a maioria dos usos e podem ser uma ameaça para a saúde pública e ambiental.

A classificação da massa de água é, de acordo com este critério, feita parâmetro a parâmetro tendo como base a grelha de catorze parâmetros apresentada na Tabela 2.12. A



inclusão da massa de água numa das classes (i.e. A, B, C, D ou E) é determinada pelo segundo parâmetro mais desfavorável [URL 1] [DRAOTA, 2003] [Matoso et al., 2004].

Tabela 2.12: Grelha proposta pelo INAG para a classificação qualitativa das águas superficiais.

Parâmetro	Classe				
	A	B	C	D	E
pH (escala Sorensen)	6,5 - 8,5		6,0 - 9,0		5,5 - 9,5
Temperatura / °C	≤ 20	> 20 ≤ 25	> 25 ≤ 28	> 28 ≤ 30	> 30
Condutividade / $\mu\text{S cm}^{-2}$	≤ 750	> 750 ≤ 1000	> 1000 ≤ 1500	> 1500 ≤ 3000	> 3000
SST / $\text{mg dm}^{-3}$	≤ 25,0	> 25,0 ≤ 30,0	> 30,0 ≤ 40,0	> 40,0 ≤ 80,0	> 80,0
Oxigénio dissolvido (% sat)	≥ 90	< 90 ≥ 70	< 70 ≥ 50	< 50 ≥ 30	< 30
Oxidabilidade / $\text{mg}_{\text{O}_2} \text{dm}^{-3}$	≤ 3,0	> 3,0 ≤ 5,0	> 5,0 ≤ 10,0	> 10,0 ≤ 25,0	> 25,0
CBO5 a 20°C / $\text{mg}_{\text{O}_2} \text{dm}^{-3}$	≤ 3,0	> 3,0 ≤ 5,0	> 5,0 ≤ 8,0	> 8,0 ≤ 20,0	> 20,0
CQO / $\text{mg}_{\text{O}_2} \text{dm}^{-3}$	≤ 10,0	> 10,0 ≤ 20,0	> 20,0 ≤ 40,0	> 40,0 ≤ 80,0	> 80,0
Azoto amoniacal / $\text{mg}_{\text{NH}_4^+} \text{dm}^{-3}$	≤ 0,10	> 0,10 ≤ 1,00	> 1,00 ≤ 2,00	> 2,00 ≤ 5,00	> 5,00
Nitratos / $\text{mg}_{\text{NO}_3} \text{dm}^{-3}$	≤ 5,0	> 5,0 ≤ 25,0	> 25,0 ≤ 50,0	> 50,0 ≤ 80,0	> 80,0
Azoto Kjeldahl / $\text{mg}_{\text{N}} \text{dm}^{-3}$	≤ 0,50	> 0,50 ≤ 1,00	> 1,00 ≤ 2,00	> 2,00 ≤ 3,00	> 3,00
Fosfatos / $\text{mg}_{\text{P}_2\text{O}_5} \text{dm}^{-3}$	≤ 0,54		> 0,54 ≤ 0,94		> 0,94
Coliformes totais / n° / 100 $\text{cm}^{-3}$	≤ 50	> 50 ≤ 5000	> 5000 ≤ 50000	> 50000	
Coliformes fecais / n° / 100 $\text{cm}^{-3}$	≤ 20	> 20 ≤ 2000	> 2000 ≤ 20000	> 20000	

Para determinar a qualidade da água das albufeiras em estudo programou-se uma folha de cálculo do *Microsoft Excel* de forma a obter directamente a classe a que pertence a massa de água. Nos últimos anos as albufeiras em estudo têm, na maior parte das vezes, pertencido à classe C, ou seja, são águas com qualidade “aceitável”, suficiente para irrigação, para usos industriais e para produção de água potável após ajustamento (Tabela 2.13).

Os resultados obtidos após a aplicação dos critérios 1 e 2 às três albufeiras estudadas (i.e. Divôr, Monte Novo e Vigia), mostra que as classificações obtidas quer por um quer por outro são concordantes. Qualquer deles aponta para a necessidade de se proceder a ajustamentos para que a água das albufeiras possa ser considerada apta para o abastecimento das populações.

Tabela 2.13: Classificação qualitativa das águas do Divôr, Monte Novo e Vigia.

	Ano	Divôr	Monte Novo	Vigia
<b>Civil</b>	1996	D	C	C
	1997	C	D	C
	1998	C	D	C
<b>Hidrológico</b>	1998-1999	C	C	C
	1999-2000	D	C	C
	2000-2001	D	C	D
	2001-2002	C	C	C

## 2.8 Conclusões

Os ecossistemas são caracterizados por possuírem propriedades como adaptabilidade, capacidade de auto-organização, flexibilidade, elevado número de mecanismos de regulação e potencial evolutivo.

A modelação de um ecossistema requer, forçosamente, que se tenha em consideração tais propriedades. Caso isso não se verifique, corre-se o risco dos modelos apenas reflectirem o comportamento do ecossistema no período das observações, originando previsões desajustadas no caso de ocorrerem variações do ecossistema, quer ao nível da sua dinâmica, quer ao nível da sua composição.

A necessidade de considerar as referidas propriedades na fase de especificação e posterior modelação levou a que se reconsiderasse toda a envolvente directamente relacionada com a resolução deste tipo de problemas. Foram tentadas diversas vias e abordagens (e.g. maximização de uma função objectivo, modelos baseados na teoria do caos, na teoria da catástrofe, recurso à programação orientada ao objecto, à programação orientada ao indivíduo ou agente e, como no caso do presente trabalho, o recurso a metodologias de resolução de problemas oriundas da Área Científica da Inteligência Artificial).

As albufeiras que se pretendem modelar classificam-se, do ponto de vista do estado trófico como eutróficas ou hiper-eutróficas. A qualidade da água, qualquer que seja o critério para a classificar, é deficitária, necessitando de ajustamentos para que possa ser dada como apta para abastecimento de populações.



## Capítulo 3

# Agentes, Sistemas MultiAgente e Medida da Qualidade de Informação em Bases de Dados/Conhecimento

Os Sistemas MultiAgente (SMA) assumem-se como uma nova metodologia para a resolução de problemas em ambientes distribuídos, que privilegia o facto de a solução ou soluções para um dado problema poder ser entendida (e em particular em sistemas que assentam as suas bases formais na Lógica Matemática) como uma consequência natural da demonstração de problemas, i.e. a computação baseada em agentes contribui para a emergência de um novo enquadramento no que toca à problemática associada à resolução de problemas e uma revolução na análise e no desenvolvimento de *software*. Não será pois de admirar que a problemática associada aos agentes e Sistemas MultiAgente esteja a interferir com o cerne das diferentes áreas do conhecimento [Gruber, 1991], i.e. os agentes não configuram apenas uma muito promissora tecnologia, mas antes se assumem como uma nova forma de pensar, um paradigma conceptual para analisar problemas e conceber sistemas, para lidar com questões de complexidade, distribuição e interactividade. Uma nova forma de computação e inteligência emerge.

Contudo, não existe uma definição universalmente aceite para caracterizar o que se entende por agente. Um agente assume-se como uma figura metafórica utilizada em inúmeras áreas do conhecimento humano, que vão desde a Psicologia, à Economia, à Sociologia, às Ciências da Computação até à Área do Ambiente. No que concerne às Ciências da Computação, o conceito de agente foi introduzido pela mão da comunidade científica da Inteligência Artificial, que para isso se serviu do sistema *Actor* [Hewitt, 1977]. Aí foram apresentadas, pela primeira vez, novas entidades que foram designadas como objectos autocon-

tidos, interactivos e concorrentes, que comunicavam entre si por troca de mensagens e que demonstravam possuir um certo tipo de comportamento. Designaram-se essas entidades por *actores*.

Por outro lado, o avanço tecnológico, em matéria de computação e de comunicação, permitiu o aparecimento da Inteligência Artificial Distribuída, passível de se subdividir, em termos da resolução de problemas, em Sistemas de Processamento Distribuído e em Sistemas MultiAgente [Bond, 1990]. As aplicações do tipo distribuído são apelativas quando aplicadas na obtenção de soluções para problemas que contemplam múltiplos intervenientes e encapsulam a complexidade e as especificidades próprias dos sistemas físicos que lhes deram origem. Os intervenientes têm uma base de conhecimento individual ou partilhada, com capacidade de processamento e de aquisição de conhecimento, assim como de formas de raciocínio, sendo por isso catalogados como *inteligentes*, i.e. partindo de uma definição muito genérica de agente, pode afirmar-se que estes são entidades em que os humanos delegam tarefas a serem executadas, e que têm a ver, por exemplo, com a sua capacidade de cálculo, de memorização ou de perseverança [Machado, 2002].

### 3.1 Agentes

Definir o conceito de agente tem sido, como já foi mencionado, uma tarefa de difícil concretização, devido ao facto de um agente ser usado para caracterizar muitas e diferentes espécies de entidades. Daí resulta que as múltiplas definições existentes na literatura se prendam com inúmeros e diferentes contextos de utilização e/ou de desenvolvimento de aplicações. No dicionário da Língua Portuguesa [Costa e Melo, 1997], é apresentada uma definição de agente, e que pode ser dada na forma:

**Definição 3.1** - Agente: s.m. algo que age, capaz de produzir um efeito.

Uma outra definição passa por considerar um agente como algo que obtém informação e/ou conhecimento do seu universo de discurso através de sensores e actua nesse ambiente através de actuadores [Russel e Norvig, 1995] (Figura 3.1). Pode dizer-se que os agentes são componentes persistentes e activos que percebem, raciocinam, actuam e comunicam [Huhns e Singh, 1998]. Pode perceber-se que os agentes são entidades que habi-

tam em ambientes complexos, sentem esse ambiente e actuam de modo autónomo, procurando executar um conjunto de tarefas para as quais receberam procuração [Maes, 1990].

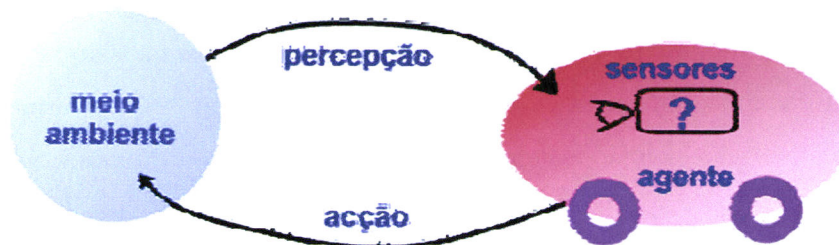


Figura 3.1: Agente a interactuar com o meio. Adaptada de Abelha [Abelha, 2004].

Na comunidade científica existe, no entanto, um certo acordo sobre uma das facetas dessas entidades, a que se prende com o conceito de autonomia. Por exemplo, Wooldridge [Wooldridge, 1999] afirma que um agente corporiza um sistema computacional capaz de uma acção flexível e autónoma, desenvolvido num determinado meio ou sobre um dado universo de discurso, o que se apresenta como uma definição algo abrangente (entende-se por flexível um agente que é reactivo, pró-activo e com capacidade de socialização).

## 3.2 Programação em Lógica Estendida

Para a caracterização formal, e implementação de tais entidades, tornou-se necessário o recurso a ferramentas de cunho formal e passíveis de serem utilizadas em prototipação rápida, como é o caso da Programação em Lógica Estendida (PLE), introduzida por [Neves, 1984] [Gelfond e Lifschitz, 1990] e que resultou da aplicação da negação forte ou explícita à Programação em Lógica (PL). O objectivo principal da PLE passou por resolver o problema do tratamento da informação incompleta. Um programa lógico é um conjunto finito de cláusulas na forma:

$$A_0 \leftarrow A_1 \wedge \dots \wedge A_m \wedge \text{não } A_{m+1} \wedge \dots \wedge \text{não } A_n$$

tal que, para cada número inteiro  $i \in [0 \dots n]$ ,  $A_i$  é um átomo. Os termos da forma  $A$  ou  $\text{não } A$  são chamados literais. A negação fraca, aqui denotada pelo termo *não*, usada na PL convencional é interpretada como negação por falsificação ou negação por falha na prova:  $\text{não } A$  é verdadeiro sempre que não é possível provar  $A$ , e  $\text{não } A$  é falso sempre que é possível provar  $A$ . No entanto, em alguns casos, é muito útil ter uma negação explícita ou forte ( $\neg$ ) considerando  $\neg A$  verdadeiro quando é possível provar  $\neg A$ . Os termos  $A$  e  $\text{não } A$ , sendo  $A$  um literal, são chamados literais estendidos, enquanto os literais simples são  $A$  ou  $\neg A$ , ao contrário do que sucede na PL em geral, onde  $A$  e  $\text{não } A$  são literais.

Os valores nulos são elementos fundamentais para a representação de informação desconhecida ou não aplicável, como normalmente acontece nas bases de dados e/ou bases de conhecimento. Para representar, de forma coerente e rigorosa os valores nulos, é necessário uma linguagem para especificar e interrogar bases de dados e/ou de conhecimento, e que lhe seja reconhecido esse potencial. Recorreu-se, numa primeira aproximação, à lógica de primeira ordem. No entanto, apesar dos resultados alcançados e que permitiram perceber melhor o que se denomina por informação incompleta, alguns senãos subsistiram:

- (i) - Os sistemas baseados na lógica de primeira ordem não parecem ter poder expressivo suficiente para permitir uma representação clara e modular de definições por omissão e declarações de normas. Isto impede o sistema de possuir algum conhecimento de senso comum. As dificuldades que existem em se exprimir formas simples do Pressuposto do Mundo Fechado em lógica de primeira ordem é uma manifestação típica de um dos senãos já mencionados;
- (ii) - A lógica de primeira ordem não é suficientemente vocacionada para a expressão de diferentes tipos de fechos transitivos que há que levar em linha de conta quando se interpreta uma base de dados; e
- (iii) - Existe um vazio entre os métodos de resposta a questões usados nas bases de dados e os métodos usados na lógica de primeira ordem e nas suas variações. Isto tem inibido a utilização da lógica de primeira ordem para a caracterização das propriedades computacionais das bases de dados.



No entanto, a utilização de linguagens baseadas na PLE pode não só preservar o poder de expressão da lógica de primeira ordem, mas também ajudar a descrever formas de incompletude que ocorrem nas bases de dados e, ultrapassar os inconvenientes referidos em epígrafe. Primeiro há que verificar que a PLE é capaz de fazer pelo menos aquilo que a lógica de primeira ordem permite, além de permitir representar em bases de dados e/ou base de conhecimento valores nulos. Por exemplo, atenda-se a um departamento universitário onde leccionam dois professores: o *Miguel* e o *João*. No departamento são leccionadas, entre outras, quatro disciplinas, nomeadamente o *Prolog*, o *Lisp*, o *Informix* e o *Basic* (A Tabela 3.1 apresenta a distribuição do serviço docente de tal departamento).

Tabela 3.1: Distribuição do serviço docente no departamento.

Professor	Disciplina
miguel	prolog
joão	lisp

Assume-se que a Tabela 3.1 tem informação completa. Para caracterizar formalmente este universo, considere-se que *miguel* e *joão* são objectos do tipo *Professor*, as disciplinas *prolog* e *lisp* objectos do tipo *Disciplina*, e um predicado *ensina* com dois argumentos, o primeiro do tipo *Professor* e o segundo do tipo *Disciplina*. A Tabela 3.1 pode facilmente ser representada pelo Programa 3.1, em PLE, onde a primeira regra representa o Pressuposto do Mundo Fechado para o predicado *ensina*( ) e permite que se prove, por exemplo, que  $\neg \text{ensina}(\text{miguel}, \text{lisp})$ .

- (1)  $\neg \text{ensina}(X,Y) \leftarrow \text{não ensina}(X,Y)$
- (2)  $\text{ensina}(\text{miguel}, \text{prolog})$
- (3)  $\text{ensina}(\text{joão}, \text{lisp})$

Programa 3.1: Respeitante à informação constante da Tabela 3.1.

A situação altera-se quando a Tabela 3.1 é substituída pela Tabela 3.2. Aqui *professor* é um valor nulo do tipo *Professor*, que representa um professor desconhecido e possivelmente diferente do Miguel ou do João. É claro que, dada essa informação, o sistema deverá responder às questões:  $\text{ensina}(X, \text{lisp})?$  e  $\text{ensina}(X, \text{basic})?$  na forma:  $X = \text{joão}$  e  $X = \text{desco}$

*nhecido*, respectivamente. A resposta à questão *ensina(miguel, lisp)?* será *não* e, a resposta à questão: *ensina(miguel, basic)?* será *não sei*.

Tabela 3.2: Disciplinas atribuídas a um dado professor, em que há um nulo do tipo desconhecido.

Professor	Disciplina
miguel	prolog
joão	lisp
professor	basic

À primeira vista, parece suficiente acrescentar ao Programa 3.1, a regra:

(4) *ensina*(professor, basic).

No entanto, a regra não é suficiente para obter o efeito desejado, e.g. a resposta a *ensina(miguel, basic)?* dada pelo programa será *não* em vez de *não sei*. Para representar correctamente esta informação é necessário encontrar uma formalização própria do conceito de Pressuposto do Mundo Fechado numa linguagem com valores nulos. Isto é conseguido substituindo a regra (1) pelas novas regras (1a) e (1b) (Programa 3.2).

*ensina*(miguel, prolog).

*ensina*(joão, lisp).

*ensina*(professor, basic).

(1a)  $\neg \textit{ensina}(X, Y) \leftarrow$

$\textit{não ensina}(X, Y) \wedge$

$\textit{não excepção}_{\textit{ensina}}(X, Y).$

(1b)  $\textit{excepção}_{\textit{ensina}}(X, \textit{basic}) \leftarrow$

$\textit{ensina}(\textit{professor}, \textit{basic}).$

Programa 3.2: Representação de um nulo do tipo desconhecido.

O Programa 3.2 representa correctamente o conhecimento presente na Tabela 3.2. Note-se que os custos computacionais e conceptuais da solução são modestos. O tipo de nulo considerado é um valor nulo do tipo desconhecido e não necessariamente pertencente a um dado conjunto de valores. Considere-se, agora, a Tabela 3.3, onde  $\{\textit{miguel}, \textit{joão}\}$

denotam um novo tipo de valor nulo, um valor nulo do tipo desconhecido mas tomado de um conjunto finito de valores.

Tabela 3.3: Disciplinas atribuídas a um professor, em que pontua informação do tipo disjuntivo.

Professor	Disciplina
miguel	Prolog
joão	lisp
{miguel, joão}	informix
professor	basic

A Tabela 3.3 denota que o *miguel* ou o *joão* ensinam *informix*, embora não se saiba ao certo quem. Para representar essa informação é necessário expandir o Programa 3.2 com a regra (5), o que é feito na forma:

(5)  $\text{ensina}(\text{miguel}, \text{informix})$  ou  $\text{ensina}(\text{joão}, \text{informix})$ .

Em termos de significado, o Programa 3.3 resultante denota correctamente a informação constante da Tabela 3.3. No entanto a introdução de informação do tipo disjuntivo tem custos computacionais elevados. Para tornar este senão, a regra (5) é substituída pelas produções (5a) e (5b):

$\text{ensina}(\text{miguel}, \text{prolog})$ .

$\text{ensina}(\text{joão}, \text{lisp})$ .

$\text{ensina}(\text{professor}, \text{basic})$ .

$\neg \text{ensina}(X, Y) \leftarrow$

$\text{não ensina}(X, Y) \wedge$

$\text{não excepção}_{\text{ensina}}(X, Y)$ .

$\text{excepção}_{\text{ensina}}(X, \text{basic}) \leftarrow$

$\text{ensina}(\text{professor}, \text{basic})$ .

(5a)  $\text{excepção}_{\text{ensina}}(\text{miguel}, \text{informix})$ .

(5b)  $\text{excepção}_{\text{ensina}}(\text{joão}, \text{informix})$ .

Programa 3.3: Representação de um nulo do tipo desconhecido tomado de um conjunto finito de valores.

O próximo exemplo apresenta outro tipo de valor nulo, um valor nulo do tipo não permitido. Considere-se uma base de dados com informação acerca dos funcionários de uma determinada empresa, dada na Tabela 3.4, onde  $\omega$  é um valor nulo do tipo não permitido, i.e. assume-se que o salário dos executivos da empresa não devem estar presentes na base de dados (e qualquer tentativa de inserir essa informação deve ser rejeitada ou provocar uma situação de inconsistência). A interpretação de  $\omega$  como um valor nulo do tipo desconhecido, leva à formulação do Programa 3.4 que representa correctamente a informação definida na Tabela 3.4. A resposta à questão *salário(miguel, 1000)?* é *não* e, a resposta à questão *salário(joão, 1000)?* é *não sei*. No entanto, a representação dada é insuficiente para capturar o estado da base de dados à medida que o tempo passa, e que está relacionada com o facto de  $\omega$  ser um nulo do tipo não permitido, i.e. nenhum alerta de inconsistência ocorre pela inserção na base de dados da cláusula *salário(joão,1000)*.

Tabela 3.4: Informação acerca dos funcionários de uma empresa em que há um nulo do tipo não permitido.

Executivo	Funcionário	Salário
joão	joão	$\omega$
	miguel	500

- (1) *executivo(joão)*.
- (2)  $\neg \text{executivo}(X) \leftarrow \text{não executivo}(X)$ .
- (3) *salário(joão,  $\omega$ )*.
- (4) *salário(miguel, 500)*.
- (5)  $\neg \text{salário}(E, S) \leftarrow \text{não salário}(E, S) \wedge \text{não exceção}_{\text{salário}}(E)$ .
- (6)  $\text{exceção}_{\text{salário}}(E) \leftarrow \text{executivo}(E)$ .

Programa 3.4: Respeitante à informação constante da Tabela 3.4.

Ora esta situação de inconsistência pode ser ultrapassada, se acrescentarmos ao Programa 3.4, as regras:

executivo(joão).  
 $\neg$ executivo(X)  $\leftarrow$   
     não executivo(X).  
 salário(joão,  $\omega$ ).  
 salário(miguel, 500).  
 $\neg$  salário (E, S)  $\leftarrow$   
     não salário (E, S)  $\wedge$   
     não exceção<sub>salário</sub>(E).  
 exceção<sub>salário</sub>(E)  $\leftarrow$   
     executivo(E).  
 nulo( $\omega$ ).  
 $\neg$ (executivo(U)  $\wedge$  salário(U,X)  $\wedge$  não nulo(X)).

Programa 3.5: Representação de um valor nulo do tipo não permitido.

É fácil de verificar que agora a inserção da cláusula *salário(joão, 1000)* na base de dados é rejeitada pelo facto de dar corpo a uma inconsistência, i.e. não é possível ser-se (*o ordenado do joão é 1000*) e não ser-se (*o ordenado do joão é desconhecido*) em simultâneo.

### 3.3 Propriedades dos Agentes

Wooldridge e Jennings [Wooldridge e Jennings, 1995] apresentam aquilo a que nos dias de hoje se pode definir como o fio condutor que levou a uma das primeiras formulações e/ou quantificações, sob um ponto de vista formal, da noção de agente ou Sistema MultiAgente, em termos de propriedades do foro antropopático (i.e. próprias dos seres humanos) e adjectivadas como sendo de *fracas* ou *fortes*. Estas entidades (agentes e Sistemas MultiAgente) são seriadas a partir do cálculo de uma média pesada dos *valores de verdade* dos seus atributos. Todo este processo se desenvolve na óptica de uma lógica multivalor, com os *valores de verdade* a pertencerem ao intervalo [0, 1] [Analide e Neves, 2002] [Analide, 2004]. Estes *valores de verdade* são dados pelo inverso da cardinalidade do conjunto das excep

ções às extensões dos predicados constantes da Base de Conhecimento (BC) dos agentes, e dadas por produções do tipo:

**Definição 3.2** - Dado o predicado,  $\text{nome-do-predicado}(A_1, \dots, A_p, \dots, A_n)$ , o conjunto de clausulas na forma  $\text{excepção}_{\text{nome-do-predicado}}(A_1, \dots, A_p, \dots, A_n)$  denota as exceções à extensão do predicado,  $\text{nome-do-predicado}(A_1, \dots, A_p, \dots, A_n)$ , em que  $n \geq 0$  e  $i = 1, \dots, n$ .

Ora, será a partir da análise da(s) extensão(ões) de predicados na forma  $\text{excepção}_{\text{nome-do-predicado}}(\dots)$ , que se elaborará sobre a personificação de entidades computacionais, aqui denominadas de agentes e/ou Sistemas MultiAgente.

Em particular, um agente ou Sistema MultiAgente é considerado *fraco* quando não apresenta um conjunto mínimo de atributos, a partir dos quais se possam definir e quantificar um certo conjunto de atitudes como, por exemplo:

(i) - Autonomia - os agentes operam sem a intervenção de outros agentes, e têm controlo sobre as suas acções e o seu estado de conhecimento interno. Autonomia é valorada em termos de valores de verdade pertencentes ao intervalo [0, 1], i.e. é aferida a partir da análise das exceções à extensão do predicado  $\text{autonomia}(A_1, \dots, A_p, \dots, A_n)$  ou seja, a partir de produções na forma:

...  
 $\text{excepção}_{\text{autonomia}}(\dots, a_j^i, \dots)$ .  
 ...

em que  $i$  denota o agente e  $a_j^i$  o atributo.

A análise de uma atitude como a autonomia é obtida fazendo uso de expressões do tipo [Wooldridge e Jennings, 1995] [Novais, 2003]:

$$V^i(\text{autonomia}) = \frac{\sum_{j=1}^n w_j^i v_j^i(a_j)}{n} \quad (3.1)$$

em que  $w_j^i$  denota o inverso da cardinalidade do conjunto das exceções ao predicado  $\text{autonomia}()$  para o atributo  $a_j^i$ , i.e. dá uma medida de relevância do atributo  $a_j^i$  na qualificação de  $\text{autonomia}$  para o agente  $i$ .  $v_j^i(a_j)$  denota, por sua vez, numa escala de valores 0...1, a con

tribuição do atributo  $a_j^i$  para a qualificação da autonomia do agente  $i$ , i.e. estabelece uma relação de ordem sobre os atributos do predicado  $autonomia()$ , em termos de uma valoração que mede quão preponderante é esse atributo para a caracterização de  $autonomia$ . Por exemplo, tem-se que  $autonomia$ , do grego  $autonomia$ , ou seja, a liberdade que cada um tem de viver segundo os seus próprios princípios e leis, pode ser aqui entendida como sendo função das extensões ou das exceções às extensões dos predicados  $princípios()$  e  $direito()$ . Se a lei a seguir for a lei geral, pode-se escrever que (em PLE):

$$\begin{aligned} & \text{direito}(\text{lei-geral}). \\ & \neg \text{direito}(X) \leftarrow \\ & \quad \text{não direito}(X). \end{aligned}$$

Neste caso a função de pontuação  $v_j^i$  toma a forma:

$$V_1^i(\text{direito}) = 1 \tag{3.2}$$

em que o número 1, em  $v_j^i()$ , denota o primeiro atributo do predicado  $direito()$ , e neste caso o único. O valor de  $v_j^i(\text{direito})$  é tomado do intervalo de valores  $[0...1]$ , e sendo este o valor máximo do intervalo, pode afirmar-se que não há qualquer ambiguidade quanto ao enquadramento legal a seguir.

No que respeita aos  $princípios$ , e partindo do pressuposto que estes são enumeráveis (e.g. Os Dez Mandamentos), e que se encontram quantificados, numa escala de 1...10, é possível escrever que:

$$\begin{aligned} & \text{excepção}_{\text{princípio}}(7). \\ & \text{excepção}_{\text{princípio}}(3). \\ & \text{excepção}_{\text{princípio}}(9). \\ & \neg \text{princípio}(X) \leftarrow \\ & \quad \text{não princípio}(X), \\ & \quad \text{não excepção}_{\text{princípio}}(X). \end{aligned}$$

o que denota que não se está seguro quanto à contribuição dos *princípios* para a caracterização da *autonomia* de cada um. Como os valores de *princípios* foram tomados do intervalo [1...10], é possível escrever que:

$$V_2^i(\textit{princípio}) = \left( \frac{10 - \frac{7+3+9}{3}}{10-1} \right) = 0,41 \quad (3.3)$$

Sabe-se também que  $w_2^i = 0,33(3)$  valor este que é dado pelo inverso da cardinalidade do conjunto de exceções à extensão do predicado *princípio*( ).  $w_2^i$  dá uma medida da relevância dos princípios que cada um subscreve ao definir o seu conceito de autonomia. Tem-se então que  $v^i$  (*autonomia*), que no caso presente se irá denominar simplesmente por *autonomia*, é dada por:

$$\begin{aligned} \frac{\sum_{j=1}^2 w_j^i v_j^i(a_j)}{2} &= \\ &= \frac{w_{\textit{direito}}^i v_{\textit{direito}}^i(\textit{direito}) + w_{\textit{princípio}}^i v_{\textit{princípio}}^i(\textit{princípio})}{2} = \\ &= \frac{1 \times 1 + 0,33 \times 0,41}{2} = 0,57 \quad (3.4) \end{aligned}$$

Em termos de representação de conhecimento, o recurso à cardinalidade das exceções à extensão de um dado predicado, permite aferir o processo de esvaziamento-de-informação/aprendizagem sobre um dado tema ou assunto (e.g. à medida que a cardinalidade aumenta, o tema ou assunto em mãos torna-se menos relevante para a resolução do problema).

(ii) - Reactividade - os agentes têm percepção do que ocorre no seu universo de discurso e respondem adequada e atempadamente a mudanças ocorridas nesse ambiente. A análise e implementação deste tipo de atitude é feita em termos da extensão do predicado *reactividade*( $R_1, \dots, R_j, \dots, R_n$ ) e do conjunto de exceções à referida extensão, ou seja, em termos das produções:



...  
 excepção<sub>reactividade</sub>(...,  $r_j^i$ , ...).

...

em que  $i$  denota o agente e  $r_j^i$  o atributo. Aqui, reactividade podia ser dada, por exemplo, em termos dos predicados *cor()*, *desejo()* e *acessibilidade()* que, no seu conjunto, são passíveis de quantificar o sentido de posse de um dado indivíduo por um certo objecto, quando este é posto perante a possibilidade de o vir a adquirir ou a perder.

(iii) - Pró-actividade - os agentes são capazes de tomar a iniciativa, conduzindo as suas próprias acções segundo um comportamento que é dirigido por objectivos. Esta atitude é analisada em moldes semelhantes aos apresentados para autonomia e reactividade; e

(iv) - Sociabilidade - os agentes interactuam com os seus pares, comunicando com estes, competindo ou cooperando na resolução de problemas. Esta atitude é analisada em moldes semelhantes aos apresentados para autonomia, reactividade e pró-actividade.

A noção de agente fraco tem a sua génese em áreas como a da Computação Distribuída e da Inteligência Artificial (IA), sendo possível afirmar que quer os agentes quer os Sistemas MultiAgente dão corpo a uma nova metodologia de computação ou forma de inteligência.

Por outro lado, um agente ou Sistema MultiAgente é considerado forte, quando a(s) entidade(s) com que se depara é (são) eminentemente cognitiva(s), passível(eis) de desenvolver a sua própria consciência, de se apresentar como tendo um conjunto de mais variáveis como a perceptibilidade, a sentimentalidade e o emocionismo [Wooldridge e Jennings, 1995]. Por conseguinte, um conjunto mínimo de atributos tem de ser considerado, e em que se destacam [Ferber, 1999] [Nwana, 1996] [Russel e Norvig, 1995]:

(i) - A Mobilidade - um agente diz-se móvel quando se movimenta através da rede executando as tarefas de que foi incumbido ou que lhe foram delegadas. Esta atitude é analisada em moldes semelhantes aos apresentados para a qualificação de autonomia, reactividade e pró-actividade, no texto precedente;

(ii) - A Intencionalidade - a intencionalidade é a capacidade que o agente apresenta para a definição de objectivos assim como das estratégias para os atingir. Esta atitude é analisada em moldes semelhantes aos apresentados para a qualificação de mobilidade;

(iii) - A Aprendizagem - a aprendizagem é a capacidade que o agente apresenta de adquirir conhecimento. A actualização da base de conhecimento é feita através da assimilação de padrões de comportamento ou de preferências manifestadas pelo próprio agente;

(iv) - A Competência - um agente é competente quando conduz com sucesso e eficiência as tarefas de que é incumbido. A competência está normalmente relacionada com a confiança depositada no agente, por terceiros;

(v) - A Veracidade - um agente não fornece, propositadamente, informação falsa;

(vi) - A Racionalidade - um agente racional não aceita realizar tarefas que lhe pareçam impossíveis de executar, contraditórias com os seus princípios (e.g. éticos), ou quando não são compensados em termos do risco, custo e esforço envolvido;

(vii) - A Benevolência - um agente benevolente adopta como seus, os objectivos de terceiros, desde que estes não entrem em conflito com os que perfilha. Os agentes benevolentes são levados a realizar todas as tarefas que lhe sejam imputadas; e

(viii) - A Emotividade - certas características próprias do ser humano têm vindo a migrar e a constituir-se como parte constituinte de agentes e Sistemas MultiAgente [Anallide e Neves, 2002]. Os primeiros passos nesta área foram dados, entre outros, por Rao e Georgeff [Rao e Georgeff, 1995], que introduziram as arquitecturas do tipo CDI (i.e. *Crenças, Desejos e Intenções*) ao nível do agente e Sistemas MultiAgente, quer como forma de representação de conhecimento, quer como estrutura computacional.

### 3.4 Sistemas MultiAgente

O estudo dos Sistemas MultiAgente (SMA) teve a sua génese na área da Inteligência Artificial Distribuída (IAD) volvidos que são 25 (vinte e cinco) anos sobre o advento da área científica da Inteligência Artificial. Os SMA têm-se vindo a afirmar como uma das áreas do conhecimento em que se realiza investigação de qualidade, em particular no que respeita à resolução de problemas em ambientes distribuídos, surgindo a todo o momento novos produtos, quer seja nos sectores comercial, industrial ou de serviços, ou mesmo em termos de investigação e desenvolvimento de protótipos. Por outro lado, o

crescimento que se afirma como exponencial no que concerne ao uso da Internet e a sua propensão para ligar entre si pessoas e máquinas, está a fazer com que se olhe ao surgimento de uma nova era, a dos SMA, como consequência directa de tal grau de interconectividade. Contudo, construir SMA requer muito mais do que a compreensão de protocolos de comunicação e paradigmas computacionais associados à programação distribuída. É preciso dominar tecnologias que têm a ver com as linguagens de comunicação entre agentes, arquitecturas de agentes, resolução de problemas em ambientes distribuídos, concepção de protocolos, assim como ser perito em teoria das organizações.

Neste trabalho, o *focus* tem a ver com problemas que respeitam à integração de conhecimento oriundo de diferentes fontes e domínios, num mundo cartesiano em termos de espaço e tempo, procurando passar-se a novos estádios (integração) de conhecimento, fluxos de informação, e interactividade entre agentes e SMA, por forma a que se possa aprofundar o conhecimento existente acerca dos problemas inerentes à gestão dos recursos hídricos e à avaliação da qualidade da água de albufeiras em tempo útil.

Minsk, em *The Society of Mind* [Minsk, 1986], lançou os SMA como a base para a concretização da inteligência em sistemas computacionais, assumindo, contudo, que cada entidade constituinte pode ser o mais simples possível e que, da sua interacção, podem emergir novas formas de inteligência. É então possível avançar para uma definição do SMA, na forma [Durfée e Rosenchien, 1994] [O'Hare e Jennings, 1996]:

**Definição 3.3** - Um Sistema MultiAgente compreende um conjunto de entidades que cooperam por forma a solucionar um dado problema o que, em princípio, estará para além das suas capacidades individuais.

A importância dos SMA assenta na especificidade de cada agente e nas suas possibilidades de relacionamento. A interacção entre agentes num SMA obriga, assim, à análise, definição, especificação e implementação de um certo conjunto base de funcionalidades, aqui concretizadas em termos de camadas, definidas na forma:

- (i) - A plataforma de comunicação;
- (ii) - A linguagem de comunicação;
- (iii) - A ontologia; e
- (iv) - A arquitectura.

Em resumo, será através da activação de protocolos de comunicação que se efectuará a troca e a descodificação de mensagens. Por outro lado, são também necessários protocolos de interacção que tornem possível o diálogo, i.e. a troca estruturada de informação através de mensagens.

A problemática do processo da interacção exige que se atenda à forma de coordenação entre agentes. Os dois modos de coordenação que se irão equacionar dão pelo nome de cooperação e competição. Na cooperação, os agentes integram o seu conhecimento e outras mais valias para alcançar um objectivo comum, o que individualmente não conseguiriam. Quando se passa para a competição, constata-se que os agentes têm interesses próprios, por vezes antagónicos; neste caso a negociação irá fornecer os mecanismos e as formas de coordenação que irão permitir a ultrapassagem destes senãos.

### 3.5 A Agência AQUA

Como se pode constatar não existe uma definição de agente que seja universalmente aceite. Na agência **AQUA**, estas entidades são compreendidas como utensílios computacionais, podendo ser entidades de *software* e/ou *hardware*, onde se manifestam propriedades como as que se enunciam no que se segue:

- (i) - Autonomia, i.e. quando existem entidades que tem habilidade para actuar sem a intervenção directa dos seus patronos (normalmente humanos);
- (ii) - Reactividade, i.e. quando existem entidades que embebidas num dado ambiente, percebem através de sensores e agem por reacção a estímulos (e.g. revêm as suas crenças concordando ou em reacção a novos dados);
- (iii) - Pró-actividade, i.e. quando existem entidades que exibem capacidade e inteligência para resolver problemas (e.g. planeamento e resolução de problemas); e
- (iv) - Sociabilidade, i.e. a inteligência de um sistema desenvolve-se pela interacção entre todas as entidades que o constituem) [Neves et al., 1997].

A agência **AQUA** é todavia um conjunto de SMA puramente comunicativo, i.e. não existe qualquer influencia externa e os agentes comunicam apenas através de mensagens. A Figura 3.2 ilustra a arquitectura geral dos SMA da agência **AQUA**. Na Figura 3.3 apresenta-se uma visão simplificada dos diferentes tipos de agentes envolvidos, e que podem ser caracterizados nos seguintes termos [Santos, 1999] [Alves et al., 2002]:

- (i) - Os Agentes *Proxy* fazem a ponte entre os utilizadores e o sistema nos termos e tipos de questões que podem ser formuladas, das explicações pedidas, das decisões que têm de ser tomadas e/ou da visualização de resultados finais. Os interfaces do sistema funcionam todos em ambiente *Web* utilizando páginas *Hypermedia*, podendo ser acedidas a partir de um qualquer *Web browser*;
- (ii) - Os Agentes de Decisão têm capacidades de mediação, actuam aceitando tarefas dos Agentes *Proxy*. Estas tarefas podem ser decompostas em tarefas elementares e então enviadas para serem processadas pelos agentes de computação, os quais, por sua vez, dão de volta resultados;
- (iii) - Os Agentes de Computação aceitam pedidos para execução de tarefas específicas dos Agentes de Decisão, retornando os resultados;
- (iv) - Os Agentes de Recursos possuem todo o conhecimento necessário para aceder a dados; e
- (v) - Os Agentes de Interação e Explicação, baseiam a sua actuação em procedimentos argumentativos, trabalhando com dados e conhecimento fornecidos pelos Agentes *Proxy* e/ou Agentes de Decisão. Tendo em atenção que o plano de execução recebido dos Agentes de Decisão pode ser parcial, e apenas quando uma tarefa é completamente executada, a informação compilada e a sua explicação podem ser encaminhadas para os Agentes *Proxy* e/ou Agentes de Decisão.

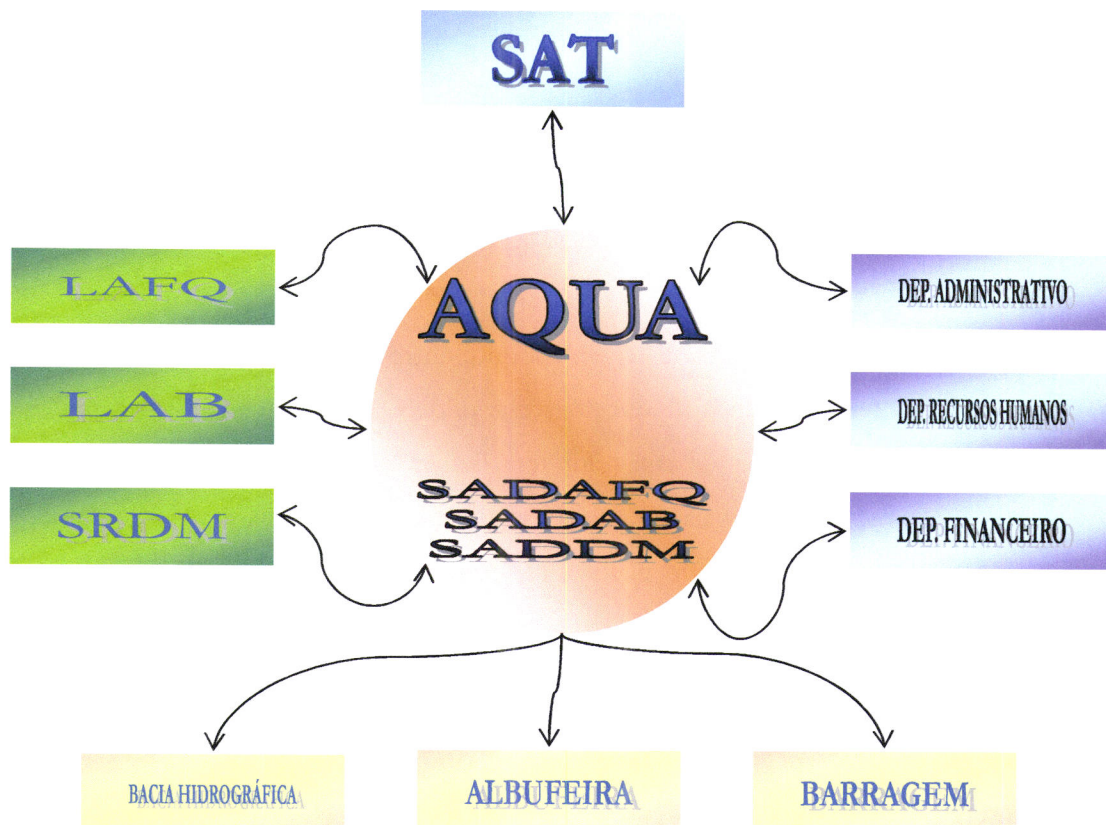


Figura 3.2: A Agência AQUA.

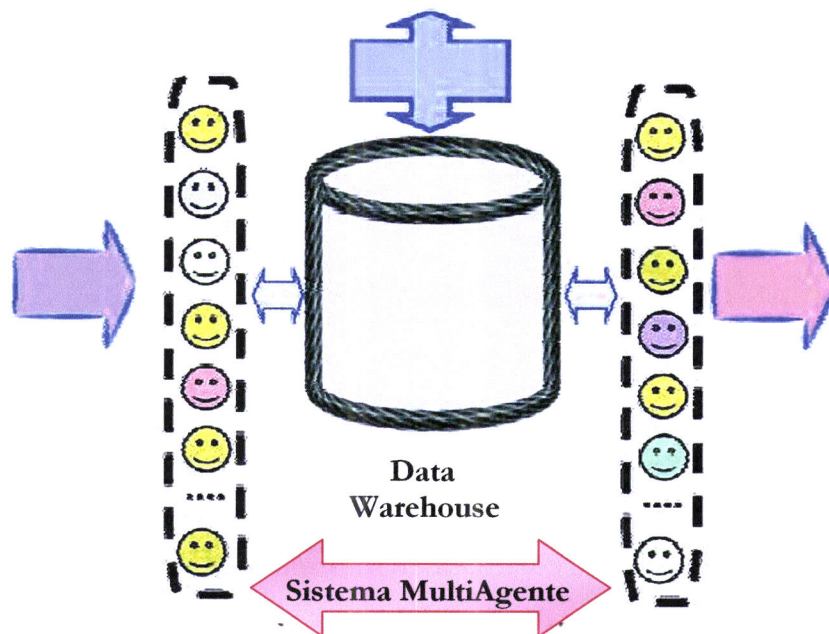


Figura 3.3: Agentes da plataforma AQUA. Adaptada de Abelha [Abelha, 2004].

## 3.6 Conclusões

Neste capítulo, é apresentada uma panorâmica sobre os SMA, áreas de aplicação e diferentes implementações, i.e. pretende justificar-se as opções tomadas na definição da arquitectura da agência **AQUA**. No final é feita uma primeira apresentação da agência **AQUA**, na perspectiva dos SMA que a integram.





## Capítulo 4

# Redes Neurais Artificiais e Descoberta de Conhecimento em Bases de Dados/Conhecimento

Na sociedade actual a informação tem assumido um papel cada vez mais importante. Contudo, dada a grande quantidade de informação disponível há o perigo de esta nos submergir se não conseguirmos seleccionar a informação certa para um problema específico. Torna-se pois vital analisar os dados e extrair deles conhecimento, generalizar, sublinhar princípios e estruturar a informação. Numa frase, devemos ser capazes de aprender a partir das observações.

A análise de dados não é um tema novo, tem sido feita desde à vários anos principalmente recorrendo a métodos estatísticos. Contudo, desde muito cedo, que ficou claro que o cérebro humano analisa dados e trata a informação de um modo diferente. Esta constatação abriu caminho para o desenvolvimento de Sistemas Inteligentes, os quais emanam da Área Científica da Inteligência Artificial (IA).

O objectivo da IA passa pelo desenvolvimento de paradigmas e algoritmos para tentar realizar tarefas cognitivas as quais são executadas eficientemente pelos seres humanos. Apesar de ser recente, tem proporcionado desenvolvimentos relevantes em diversas áreas do conhecimento como, por exemplo, na compreensão da linguagem natural, no raciocínio lógico matemático, na teoria dos jogos ou nos sistemas de apoio à decisão.

De acordo com Haykin [Haykin, 1999] um sistema de IA deve preencher três requisitos: armazenar conhecimentos, aplicar o conhecimento adquirido para resolver problemas e adquirir novo conhecimento através da experiência.

O primeiro requisito está relacionado com a representação do conhecimento que provém, fundamentalmente, de dois tipos de informação, o estado conhecido (representado por factos sobre o que é e o que se conhece) e observações ou medições sobre o ambiente (este tipo de informação pode conter ruído e está sujeito a erro).

O segundo requisito, a aplicação do conhecimento adquirido, diz respeito ao raciocínio, definido como a capacidade para a resolução de problemas, que pode ser encarada como um problema de procura.

O último requisito envolve a aprendizagem que permite aumentar o conhecimento através da experiência. A aprendizagem pode envolver duas fases distintas de processamento da informação - a forma indutiva e a forma dedutiva. Na primeira têm-se a aprendizagem baseada em casos, ou seja, determinam-se padrões e regras a partir dos dados e das experiências. Na segunda forma as regras são utilizadas para criar novos factos.

No âmbito da aprendizagem um aspecto importante a realçar prende-se com a relação entre o modelo de aprendizagem e o seu ambiente. Neste contexto existem, fundamentalmente, três paradigmas em que se baseia a aprendizagem [Patterson, 1996]:

- (i) - Aprendizagem supervisionada - é uma técnica bastante divulgada, na qual são fornecidas as respostas correctas ao sistema. A aprendizagem é levada a cabo a partir de um conjunto de casos de treino ou exemplos onde cada um deles é formado por um vector de entrada e um vector de saída ou de resposta;
- (ii) - Aprendizagem de reforço - nesta técnica não se apresenta a resposta correcta ao sistema. Apenas se dá uma indicação sobre se a resposta apresentada pelo sistema é correcta ou não, sendo esta informação utilizada no melhoramento da sua eficácia; e
- (iii) - Aprendizagem não supervisionada - neste caso não chega ao sistema qualquer indicação externa sobre a resposta correcta. A aprendizagem é feita a partir de descobertas de características nos dados de entrada, adaptando-se a regularidades estatísticas ou agrupamentos de padrões dos exemplos de treino.

As técnicas para a resolução de problemas na área da IA passam essencialmente por duas abordagens. A primeira concepção, designada por simbolista, foi baseada, fundamentalmente, em estudos em Psicologia e na Lógica Matemática. É uma abordagem essen-

cialmente matemática, onde o conhecimento é representado pelo uso de uma linguagem simbólica, através de uma colecção de factos e regras. No segundo tipo de abordagem, designado por movimento conexionista, o desafio consiste em desenvolver e implementar modelos cognitivos artificiais que reproduzam as capacidades naturais do cérebro humano - as Redes Neuronais Artificiais [Weigend e Gershenfeld, 1994].

## 4.1 Fundamentos Biológicos

A maior parte da investigação em Redes Neuronais Artificiais (RNAs) foi inspirada e influenciada pelo sistema nervoso dos seres humanos. Este recebe milhares de informações dos diferentes receptores sensoriais e a combinação dessa informação com a informação armazenada (obtida por hereditariedade ou pelos mecanismos de aprendizagem) produz a resposta a ser dada. Esta quantidade enorme de informação é processada e apenas uma pequena parte dela é relevante para o funcionamento do corpo humano, (e.g. se alguém fica exposto continuamente a um cheiro, passado algum tempo esse cheiro deixa de ser percebido).

O cérebro humano é uma estrutura extremamente complexa e paralela. Em termos de velocidade de processamento, um neurónio é cerca de cinco a seis vezes mais lento do que uma porta lógica de silício. Contudo, esta lentidão é compensada pela utilização de uma estrutura paralela [Haykin, 1999]. O neurónio é uma célula complexa, que responde a sinais electroquímicos, sendo composto por um núcleo, por um corpo celular, por um numeroso conjunto de dendrites (entidades que recebem sinais de outros neurónios) e por um axónio que transmite um estímulo a outros neurónios [Bose e Liang, 1996] (Figura 4.1).

As conexões entre o terminal de um neurónio e as dendrites de outro denominam-se por sinapses. Estima-se que o córtex humano seja constituído por cerca de dez biliões de neurónios e sessenta triliões de sinapses [Haykin, 1999].

As sinapses são junções especializadas, em que uma das células influencia a outra através da passagem de mensagens químicas. O neurónio que envia a mensagem denomina-se por pré-sináptico e o que a recebe por pós-sináptico. A separá-los existe a fenda sináptica, com cerca de 20 a 40 nm de largura. Na zona terminal do axónio existe uma zona em forma de botão, contendo vesículas com determinadas espécies químicas designadas por neurotransmissores (Figura 4.2).

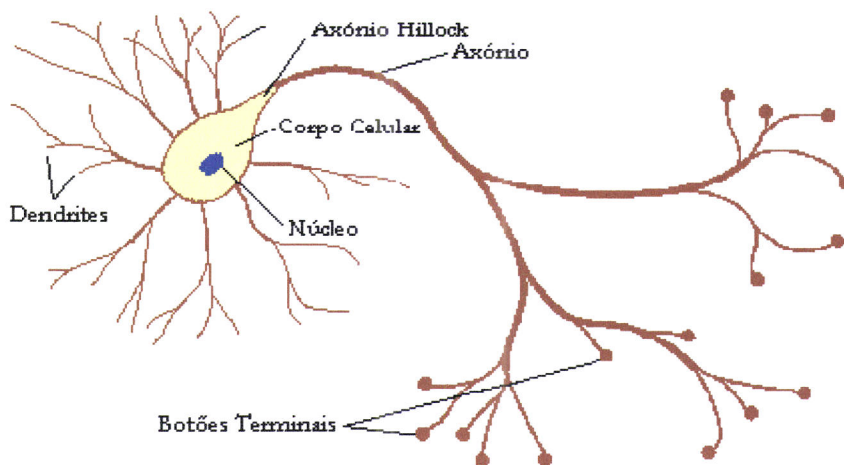


Figura 4.1: Representação esquemática de um neurônio biológico.

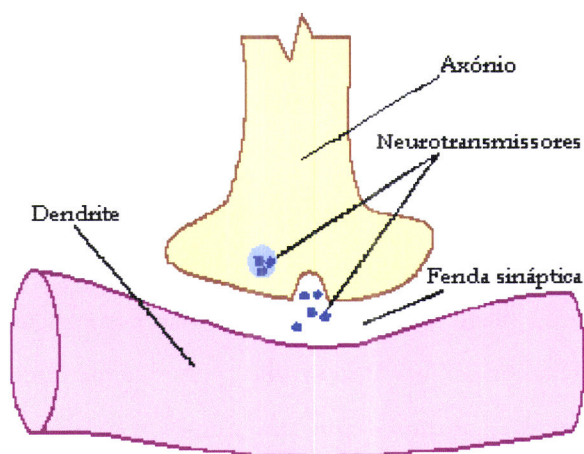


Figura 4.2: Representação esquemática de uma sinapse.

A liberação do neurotransmissor ocorre por excitação, após a chegada do potencial de ação à membrana pré-sináptica. Este é o sinal para as vesículas contendo os neurotransmissores libertarem o seu conteúdo para a fenda sináptica.

A membrana pós-sináptica contém receptores específicos para as moléculas de neurotransmissores libertadas, cuja ligação vai permitir a formação do potencial de ação, que se propaga ao longo do neurônio pós-sináptico.

A actividade da sinapse depende do equilíbrio entre a liberação de neurotransmissor e a sua destruição. Este último passo é fundamental, pois o excesso de neurotrans-

missores na fenda sináptica impossibilitaria a membrana pós-sináptica de receber novo estímulo, durante um espaço de tempo relativamente longo. Dado que a eficiência do sistema nervoso depende da sua capacidade de transmitir o máximo de mensagens, num mínimo de tempo, o neurotransmissor tem que ser removido rapidamente.

Os neurónios respondem, como se referiu anteriormente, a sinais electroquímicos que fazem o axónio libertar substâncias químicas, os neurotransmissores, existentes na membrana pré-sináptica que atravessam a membrana pós-sináptica possibilitando o desenvolvimento de um potencial de membrana.

Cada neurónio “decide” se irá ou não gerar um potencial de acção após somar todos os estímulos que recebe (excitatórios e inibitórios). Esta “soma” de estímulos pode ser espacial ou temporal. A primeira resulta da adição de estímulos simultâneos de sinapses localizadas em diversos locais da célula pós-sináptica, enquanto a segunda resulta da adição de estímulos de uma mesma sinapse em rápida sequência.

Cada neurónio pode receber até mil estímulos sinápticos mas apenas produz uma resposta, um potencial de acção formado no seu axónio. O potencial de acção é uma alteração brusca e intensa do potencial de membrana, durando apenas cerca de um ou dois milisegundos.

Os neurónios tendem a agruparem-se em camadas, considerando-se três tipos principais de conexões:

- (i) - Divergentes, onde um neurónio pode estar ligado a outros neurónios via uma arborização do axónio;
- (ii) - Convergentes, onde vários neurónios podem estar conectados a um único neurónio; e
- (iii) - Encadeadas ou cíclicas, as quais podem envolver vários neurónios e formarem ciclos.

As RNAs são, pois, modelos simplificados do sistema nervoso central dos seres humanos. São processadores paralelos compostos por unidades de processamento designadas por neurónios artificiais ou nodos. Comportam-se de forma semelhante ao cérebro humano no modo como adquirem o conhecimento (por processos de aprendizagem) e no modo como o conhecimento é armazenado (nas ligações entre nodos designadas apenas por ligações).

## 4.2 Perspectiva Histórica

Do ponto de vista histórico, as RNAs tal como hoje as entendemos, surgiram em 1943 com o trabalho de McCulloch e Pitts [McCulloch e Pitts, 1943], publicado no “*Bulletin of Mathematical Biophysics*” intitulado “*A logical calculus of the ideas immanent in nervous activity*”. Neste trabalho os autores propuseram um modelo simplificado de neurónios biológicos. O modelo baseia-se no facto do neurónio, num dado instante, ou estar inactivo ou disparando o que corresponde ao *verdadeiro* ou *falso* da lógica proposicional ou ao *zero* e ao *um* da álgebra booleana.

O trabalho de McCulloch e Pitts fez surgir o interesse pelas RNAs e atraiu para a área muitos investigadores. Contudo, alguns resultados menos positivos e a falta de algoritmos de aprendizagem levou a que a abordagem conexionista ficasse em segundo plano na década de setenta, ressurgindo no início dos anos oitenta com diversas contribuições, das quais se destacam:

- (i) - RNAs baseadas em redes neuronais recorrentes com conexões simétricas [Hopfield, 1982];
- (ii) - Redes não supervisionadas do tipo *Kohonen* [Kohonen, 1982];
- (iii) - O aparecimento do algoritmo da retro-propagação que se tornou o algoritmo de treino mais utilizado em redes neuronais unidireccionais [Rumelhart, Hinton e Williams, 1986];
- (iv) - O desenvolvimento de redes denominadas *Radial Basis-Functions* (RBF) como alternativa às redes neuronais unidireccionais [Broomhead e Lowe, 1988]; e
- (v) - O surgimento de uma classe de redes supervisionadas denominadas *support vector machines* para a regressão e reconhecimento de padrões [Boser et al., 1992].

Actualmente tenta-se obter não só redes mais eficientes como também melhores algoritmos de treino [Sharda e Rampal, 1996].

## 4.3 Benefícios das Redes Neurais Artificiais

O poder computacional de uma RNA alicerça-se em dois aspectos fundamentais: numa topologia que premeia o paralelismo e, por outro lado, na sua capacidade de aprendizagem e generalização, i.e. conseguir responder de modo adequado a novas situações, com base em experiências passadas. São estas duas características que tornam possível a resolução de problemas que, de outra forma, seriam intratáveis.

Isto não quer dizer que as RNAs sejam caixas mágicas, que consigam por si dar resposta a qualquer problema. Em vez disso, precisam não raras vezes de serem integradas com outros sistemas ou novos paradigmas para a resolução de problemas. Convém, ainda, reconhecer que se está ainda muito longe de atingir uma arquitectura que simule o cérebro humano [Haykin, 1999].

No entanto, as RNAs apresentam características únicas, que não se encontram em outros mecanismos ou técnicas para a resolução de problemas [Azoff, 1994] [Patterson, 1996] [Haykin, 1999] [Cortez, 2002]:

- (i) - Aprendizagem e generalização, i.e. conseguindo descrever o todo a partir de algumas partes, constituindo-se como formas eficientes de aprendizagem e armazenamento de conhecimento;
- (ii) - Processamento maciçamente paralelo, i.e. permitindo que tarefas complexas sejam realizadas num curto espaço de tempo;
- (iii) - Não linearidade, i.e. atendendo a que a maioria dos problemas reais e a resolver são de natureza não linear;
- (iv) - Adaptabilidade, i.e. podendo adaptar a sua topologia de acordo com mudanças do ambiente;
- (v) - Robustez e degradação suave, i.e. permitindo processar o ruído ou informação incompleta de forma eficiente, assim como sendo capazes de manter o seu desempenho quando há desactivação de algumas das suas conexões e/ou nodos; e
- (vi) - Flexibilidade, i.e. apresentam-se com um grande domínio de aplicabilidade.

## 4.4 A Inteligência Artificial e as Redes Neurais Artificiais

Os sistemas clássicos da IA podem ser comparados às RNAs, a três níveis de complexidade [Memmi, 1989]:

- (i) - Explicação - nos primórdios da IA, a ênfase estava na construção de representações simbólicas, admitindo-se a existência de uma representação mental para a cognição. Por outro lado, a ênfase nas RNAs está no desenvolvimento de modelos de Processamento Paralelo Distribuído, procurando-se uma explicação neurobiológica para os fenómenos cognitivos;
- (ii) - Estilo de Processamento - na IA dita de primeira geração, o processamento é, não raras vezes, sequencial e, mesmo quando não existe uma ordem pré-definida, as operações são executadas *passo a passo*. Este estilo é influenciado pela natureza sequencial da linguagem natural e lógica, assim como da estrutura da máquina de *von Neumann*. Em contraste, o paralelismo é essencial às RNAs, dando-lhe flexibilidade, robustez e imunidade ao ruído; e
- (iii) - Estrutura da Representação - a IA utiliza uma linguagem simbólica, com uma estrutura quase linguística, onde surgem expressões complexas a partir de símbolos simples. Ao contrário, nas RNAs o conhecimento é armazenado na sua estrutura interna, no valor dos pesos das suas conexões, dados na forma de valores pertencentes ao conjunto dos números reais.

## 4.5 Arquitecturas de Rede

A forma como os nodos se interligam numa estrutura de rede é denominada de arquitectura ou topologia. Existem inúmeros tipos de arquitecturas de RNAs ou topologias,



cada um com as suas próprias potencialidades. Em geral, caem dentro de duas categorias principais [Haykin, 1999]:

- (i) - Redes Neurais Unidireccionais (RNUs) - são organizadas por camadas, pois não existem ciclos, dado que as conexões se propagam sempre numa só direcção (podem ser convergentes ou divergentes). Na sua forma mais simples uma rede é composta por uma camada de entrada, cujos valores de saída são fixados externamente, e por uma camada de saída. No entanto, pode possuir uma ou mais camadas internas, cujos nodos são designados por nodos intermédios (Figura 4.3); e

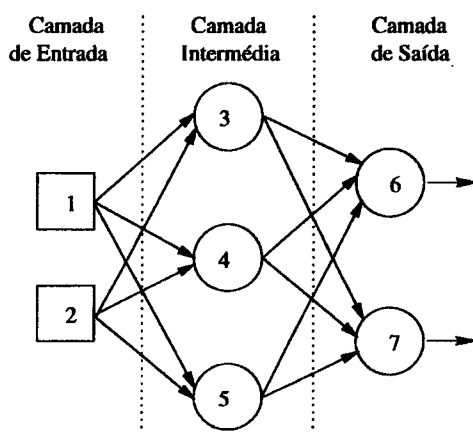


Figura 4.3: Arquitectura de uma RNU. Adaptada de Cortez [Cortez, 2002].

- (ii) - Redes Neurais Recorrentes (RNRs) - a recorrência existe em sistemas dinâmicos quando uma saída de um elemento influencia, de algum modo, a entrada para esse mesmo elemento, criando-se assim um ou mais circuitos fechados (Figura 4.4). Ao conter ciclos, as saídas não são função exclusivamente das conexões entre nodos, mas também de uma dimensão temporal, i.e. está-se na presença de um procedimento recursivo, que obedecerá naturalmente a uma certa condição de paragem, com a última iteração a ser dada como a saída para o nodo [Rojas, 1996].

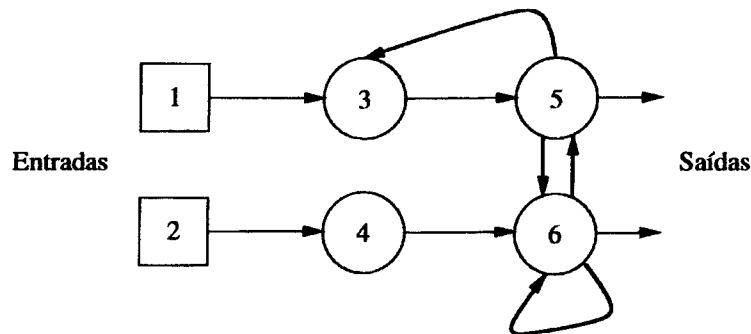


Figura 4.4: Arquitectura de uma RNR.. Adaptada de Rojas [Rojas, 1996].

## 4.6 Aprendizagem

Como já foi referido, uma propriedade importante das RNAs está na sua capacidade para aprender a partir do seu ambiente. A aprendizagem de uma RNA envolve a sequência de eventos:

- (i) - A RNA é estimulada por eventos que emanam de um dado ambiente;
- (ii) - Certos parâmetros livres, normalmente os pesos das conexões, são alterados em resultado deste estímulo; e
- (iii) - A RNA responde de uma forma diferente a quaisquer estímulos que venham desse ambiente, em virtude das alterações que a sua estrutura interna está sofrendo.

A aprendizagem é executada a partir de um conjunto bem conhecido de regras, chamado de algoritmo de aprendizagem ou de treino.

A escolha da arquitectura e do método de aprendizagem é influenciada pela tarefa a ser desempenhada pela RNA, sendo de referenciar, entre outras, as categorias de [Patterson, 1996] [ Haykin, 1999]:

(i) - Diagnóstico - esta é uma tarefa deveras comum em áreas distintas como a Medicina, a Engenharia ou a Ecologia. Trata-se essencialmente de uma tarefa de classificação, i.e. exige uma correcta associação entre entradas, que denotam os dados indicadores de um estado (e.g. parâmetros físico-químicos), com o correspondente diagnóstico (e.g. qualidade da água). De uma forma geral, as RNAs são integradas em sistemas periciais porven-

tura mais elaborados ou complexos, em que se recorre a diferentes técnicas de aprendizagem (e.g. aprendizagem por regras);

(ii) - Reconhecimento de padrões - formalmente, esta tarefa define-se como o processo pelo qual um sinal/padrão recebido é atribuído a uma de diversas categorias possíveis. Em primeiro lugar, é necessário treinar uma rede, onde os padrões, associados à respectiva categoria, são alimentados à rede de uma forma repetitiva. Mais tarde, um padrão novo é fornecido à rede, que terá de ser capaz de identificar a categoria correcta, de acordo com a informação assimilada;

(iii) - Regressão/Previsão - a ideia é conceber uma RNA capaz de modelar uma função desconhecida  $f(\bullet)$ , que se aproxime da função  $F(\bullet)$  dada por um conjunto de vectores etiquetados, i.e. compostos por um par entrada-saída, de forma que a distância euclidiana seja ínfima para todas as entradas, ou seja:

$$\forall x, \|F(x) - f(x)\| < \rho \quad (4.1)$$

onde  $\rho$  denota um valor real próximo de zero. A regressão é uma tarefa perfeita para a aprendizagem supervisionada. Na previsão, um caso particular de regressão, pretende-se “adivinhar” o comportamento futuro de um dado sistema com base no seu comportamento passado (e.g. a previsão de séries temporais);

(iv) - Controlo - esta tarefa envolve um processo ou uma parte crítica de um sistema que tem de ser mantido sob controlo. O principal objectivo do controlador é fornecer os sinais apropriados para um dado sistema de controlo, de forma a que a saída deste acompanhe uma entrada de referência. Como exemplos temos o controlo de veículos autónomos, de robôs ou de processos de fabrico, onde as RNAs tem sido utilizadas com bastante sucesso; e

(v) - Filtragem/Compressão de dados - o termo filtro usualmente designa um dispositivo ou algoritmo que extrai informação de interesse, a partir de um conjunto de dados em bruto, que apresentam um certo ruído. Por sua vez, a compressão envolve uma redução de um espaço  $n$ -dimensional para um espaço  $m$ -dimensional, sendo  $n > m$ . Ambas as tarefas se tornam particularmente importantes quando existem quantidades enormes de dados a serem processados (e.g. o processamento de imagens).

## 4.7 Neurónio Artificial ou Nodo

Um nodo, termo usado para distinguir um neurónio artificial de um neurónio natural, é a unidade de processamento chave para a operação de uma RNA. Embora existam diversos tipos de nodos, em princípio estes comportam-se como um comparador, que produz uma saída quando o efeito cumulativo das entradas excede um dado valor limite. Um nodo pode ser dissecado de acordo com o exposto na Figura 4.5 [Rojas, 1996]:

- (i) - Um conjunto de conexões ( $w_j$ ), cada uma etiquetada por um peso, i.e. um número real ou binário (embora seja mais comum na forma real) que tem um efeito excitatório para valores positivos e inibitório para valores negativos. Assim o sinal ou estímulo ( $x_j$ ), como entrada da conexão, é multiplicado pelo correspondente peso  $w_j$ , onde  $i$  denota o nodo objecto de estudo e  $j$  o nodo de onde partiu o sinal. Pode ainda existir uma conexão extra, denominada de *bias*, cuja entrada toma o valor +1, que estabelece uma certa tendência ou inclinação no processo computacional, i.e. adiciona uma constante ( $w_{i0}$ ) para que se estabeleçam as condições operacionais correctas para o nodo;
- (ii) - Um integrador, que reduz os  $n$  argumentos de entrada (estímulos) a um único valor. Nas RNUs é utilizada a função adição  $\Sigma$ , pesando todas as entradas numa combinação linear; e
- (iii) - Uma função de activação ( $f_a$ ), que pode condicionar o sinal de saída, introduzindo uma componente de não linearidade no processo computacional.

Em termos formais tem-se que este nodo é descrito pelas seguintes equações:

$$u_i = \sum (1 \times w_{i0}, x_1 \times w_{i1}, x_2 \times w_{i2}, \dots, x_n \times w_{in}) \quad (4.2)$$

$$s_i = f_a(u_i) \quad (4.3)$$

para o nodo  $i$  com  $n$  entradas e uma saída, onde o  $u_i$  representa o ganho do nodo  $i$  e  $s_i$  a saída do nodo.

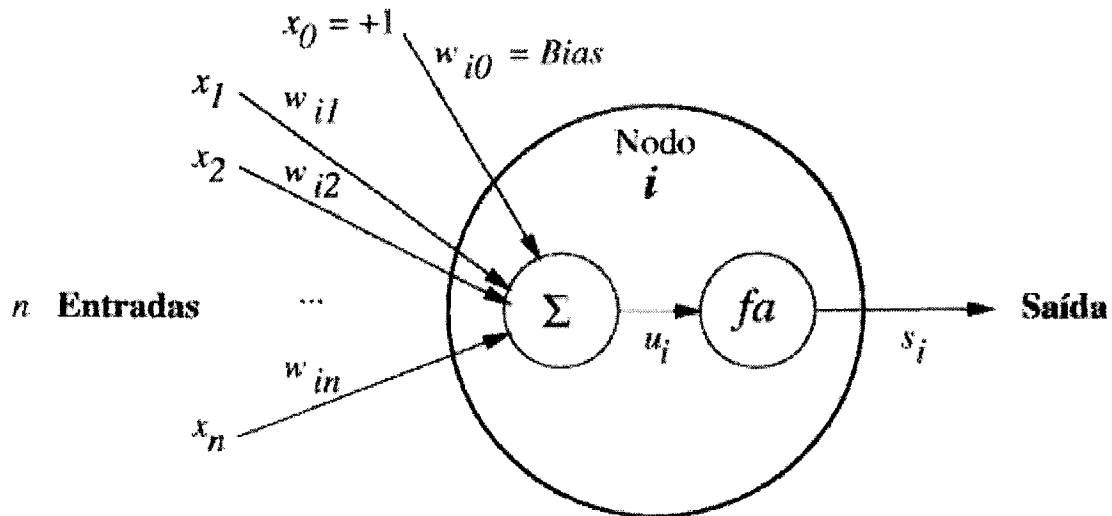


Figura 4.5: Estrutura geral de um nó. Adaptada de Rojas [Rojas, 1996].

A Tabela 4.1 mostra algumas das funções de activação mais utilizadas, onde  $k$  denota a inclinação da função,  $mod$  o resto de uma divisão inteira e  $sign(x) = \frac{x}{|x|}$ , se  $x \neq 0$  e  $sign(0) = 0$  (Figura 4.6) [Azoff, 1994] [Cortez, 2002].

A função limiar, também designada por função *Heaviside*, é normalmente utilizada em nodos do tipo McCulloch-Pitts [McCulloch e Pitts, 1943], em que a saída toma o valor +1 apenas se o ganho for não-negativo, de acordo com uma filosofia do tudo ou nada. Em seguida, aparecem duas outras funções lineares, com a última a ter como contradomínio o intervalo  $[0, 1]$ .

A não linearidade é definida em termos das restantes funções (i.e. logística, tangente hiperbólica, seno, coseno, gaussiana e quadrada), onde uma especial atenção deve ser dada à função logística, também conhecida por função *sigmoid*. Esta função, cuja forma é aproximada por um “S”, é de longe a função de activação mais utilizada em RNAs. É uma função crescente que exhibe um balanceamento gracioso entre um comportamento linear e não linear (Figura 4.7) [Jordan, 1995]. Ao variar a inclinação ( $k$ ) obtêm-se funções com diferentes declives. No limite, quando  $k$  se aproxima do infinito, a função tende para a limiar.

Tabela 4.1: Algumas funções de activação.

Nome	Função	Contradomínio
limiar	$\begin{cases} 1 & , u_i \geq 0 \\ 0 & , u_i < 0 \end{cases}$	{0,1}
linear	$u_i$	$] -\infty, +\infty [$
por troços	$\begin{cases} 1 & , u_i \geq 0,5 \\ k u_i & , -0,5 < u_i < 0,5 \\ 0 & , u_i \leq -0,5 \end{cases}$	[0,1]
logística	$\frac{1}{1 + e^{-k u_i}}$	[0,1]
tangente hiperbólica	$\tanh(k u_i)$	[-1,1]
seno	$\sin(u_i \bmod 2\pi)$	[-1,1]
coseno	$\cos(u_i \bmod 2\pi)$	[-1,1]
gaussiana	$\frac{-u_i^2}{e^{2k^2}}$	[-1,1]
quadrada	$-\text{sign}(u_i)u_i^2$	$] -\infty, +\infty [$

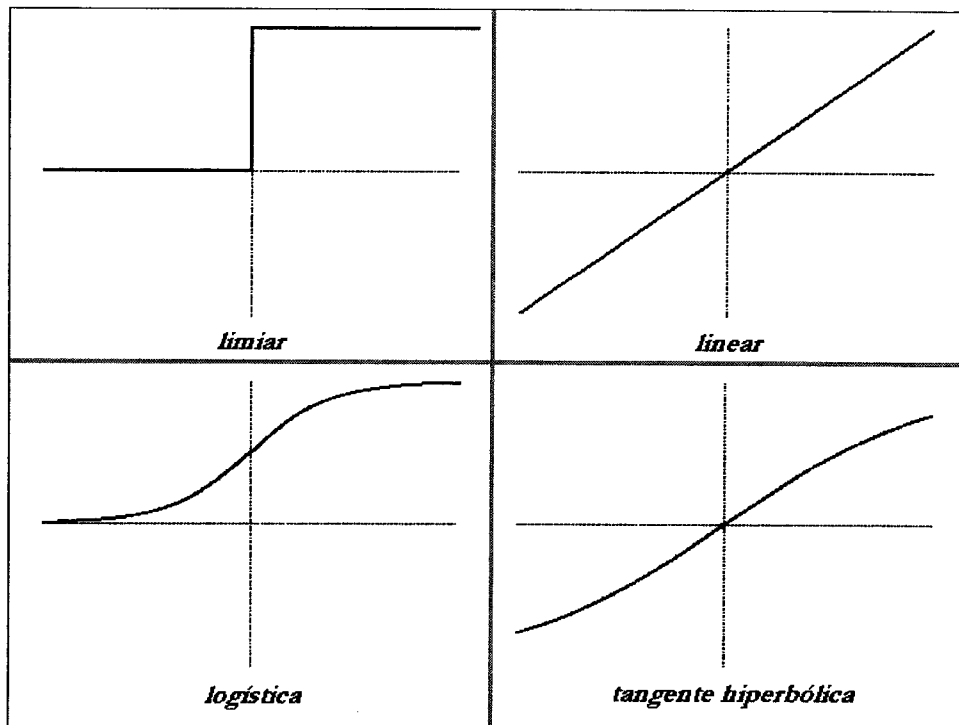


Figura 4.6: Alguns exemplos de funções de activação.

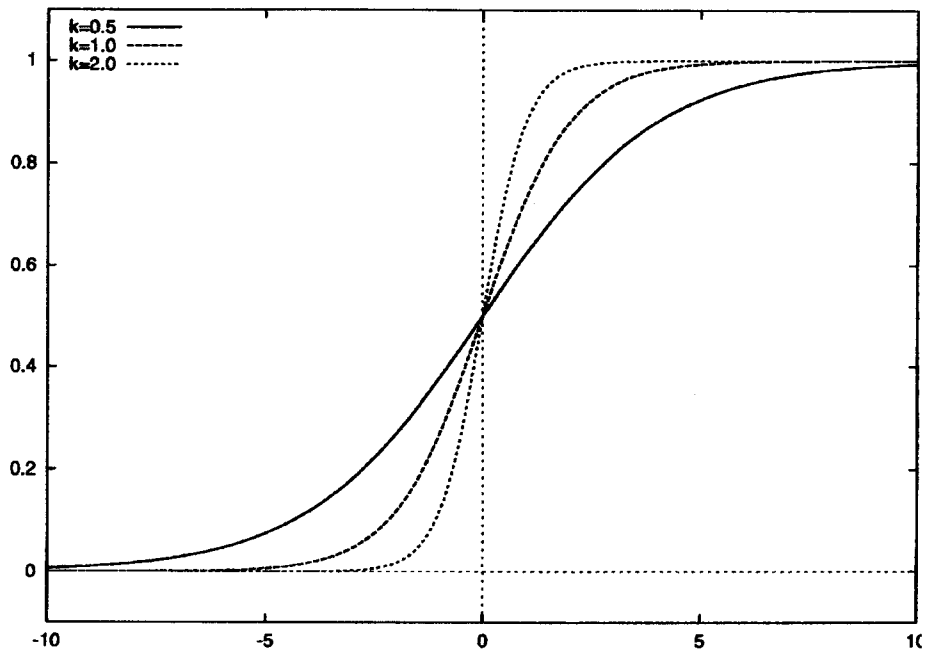


Figura 4.7: A função sigmoide para três valores  $k$ . Adaptada de Cortez [Cortez, 2002]

## 4.8 A Arquitectura Unidireccional

As RNUs constituem uma das mais importantes e populares classes de RNAs, com um vasto leque de aplicabilidade, utilizáveis, por exemplo, em problemas de memória associativa, classificação, reconhecimento de padrões e regressão [Patterson, 1996].

Em termos formais, estas são definidas pelo tuplo  $(E, I, S, C, F)$  [Gallant, 1993] [Rojas, 1996], contendo (Figura 4.8):

- (i) - um conjunto de nodos de entrada ( $E$ ), onde pontificam os estímulos enviados a partir do ambiente;
- (ii) - um conjunto de nodos intermédios ( $I$ ), unidades internas de processamento que aumentam a capacidade de aprendizagem de tarefas complexas;
- (iii) - um conjunto de nodos de saída ( $S$ ), que devolvem a resposta da rede;

- (iv) - Um conjunto de conexões pesadas unidireccionais ( $C$ ), definidas pelo tuplo  $(i, j, w)$  ou abreviadamente  $w_{ij}$ , em que  $i \in I \cup S$ ,  $j \in E \cup I$ ,  $j < i$  e  $w \in \mathfrak{R}$ ; e
- (v) - Um conjunto de funções de activação ( $F$ ), normalmente do tipo não linear e diferenciável, sendo a função logística uma das mais utilizadas (Tabela 4.1).

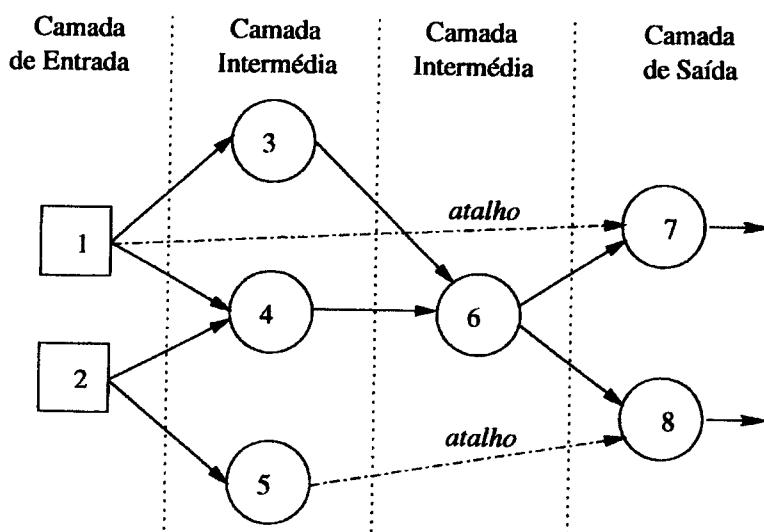


Figura 4.8: Estrutura de uma RNU com a topologia 2-3-1-2. Adaptada de Cortez [Cortez, 2002]

O sinal de entrada propaga-se para a frente, através da rede, camada por camada, não existindo ciclos. Nas redes multicamada as unidades Intermédias ( $I$ ) dividem-se em  $c$  subconjuntos, ou camadas  $(I_1, I_2, \dots, I_c)$ . É comum representar as camadas pela forma  $|E| - |I_1| - \dots - |I_c| - |S|$ , onde  $| \cdot |$  denota a cardinalidade dos conjuntos.

Uma RNU designa-se por completamente interligada quando contém todas as ligações possíveis entre os nodos de duas camadas adjacentes (Figura 4.3). Caso contrário, designa-se por parcialmente interligada. Por vezes existem ligações directas entre as entradas e os nodos de saída ou que “saltam” camadas (Figura 4.8). Estas conexões designam-se por atalhos [Prechelt, 1994].

A não linearidade, a existência de nodos intermédios e o seu alto grau de conectividade tornam a arquitectura RNU muito poderosa como uma máquina de aprendizagem. Por outro lado, são estas mesmas características que dificultam um estudo de complexidade ao processo de aprendizagem [Haykin, 1999].



A topologia ideal (em termos de camadas, nodos e conexões) de uma RNU depende fundamentalmente do problema a ser resolvido. Em geral, tenta-se reduzir o número de escolhas (ou espaço de procura) adoptando certos pressupostos, como o uso de redes completamente interligadas. Também é comum adoptar-se apenas uma camada intermédia, ficando a escolha dos nodos intermédios a ser decidida por regras empíricas. Trata-se de uma estratégia que leva a que se esteja perante diferentes soluções para o problema, sem um critério de decisão contratuado, dado que não se atribui grande importância ao modelo em termos de desempenho, em parte devido ao enorme poder das RNUs.

## 4.9 Capacidades e Limitações

Uma rede RNU pode ser vista como uma forma prática para efectuar uma qualquer correspondência não linear. A questão que surge é: dada uma função  $g$  desconhecida que faz uma correspondência entre padrões de um espaço  $n$ -dimensional para um espaço  $m$ -dimensional,  $g : \mathcal{R}^n \rightarrow \mathcal{R}^m$ , será possível encontrar uma RNU que realize a correspondência exigida ou pelo menos se aproxime dela com alguma eficácia? A resposta vem do Teorema Universal de Aproximação de Funções que estipula que uma camada intermédia é suficiente para uma RNU computar uma aproximação de uma qualquer função contínua, embora o teorema não afirme que esta estrutura seja óptima em termos de tempo de aprendizagem e generalização [Haykin, 1999].

Outro aspecto importante diz respeito ao tempo de aprendizagem. A aprendizagem implica a procura dos elementos desconhecidos de uma RNA, normalmente pelo ajuste dos pesos. Ora, a aprendizagem numa rede com 100 (cem) pesos é bastante mais pesada, em termos computacionais, do que uma rede com 10 (dez) pesos, sendo uma relação bem maior do que o factor 1:10 poderia sugerir. Seria muito útil que o tempo de aprendizagem fosse limitado por uma função polinomial sobre o número de variáveis. No entanto, tal não se verifica.

Está igualmente provado que o problema geral de aprendizagem em RNAs é intratável, i.e. não pode ser resolvido de forma eficiente para todas as instâncias. Não é conhecido um algoritmo que consiga realizar a aprendizagem num tempo polinomial, sendo improvável que tal algoritmo venha a existir [Rojas, 1996].

Contudo, devido às capacidades de correspondência não linear entre padrões, as RNUs têm sido utilizadas com sucesso em diversos domínios, incluindo, entre outros, Processos de Controlo Industrial, Processos de Controlo Ambiental, Robótica e Automação, Estatística, Gestão e Controlo de Produção, Medicina, Ecologia, Economia, Química [Zupan e Gasteiger, 1993] [Patterson, 1996] [Sharda e Rampal, 1996] [Neves et al., 1998] [Recknagel e Wilson, 2000] [Cortez, 2002]. Normalmente, para um bom conjunto de dados de treino, uma rede com apenas uma (ou quanto muito duas) camada(s) intermédia(s) serve para aprender uma dada tarefa com mestria.

## 4.10 Bases de Dados e Descoberta de Conhecimento

O avanço da tecnologia tem proporcionado novas formas de criar e armazenar dados. Diariamente, as organizações acumulam informação sobre diversos processos (e.g. contabilidade, facturação, transacções comerciais, transacções governamentais, análises na área do ambiente e investigação científica), tendo como base o pressuposto de que grandes volumes de dados podem ser fonte de conhecimento, e que este pode ser extraído com ajuda de técnicas apropriadas. A existência desta mais valia pode contribuir para melhorar o comportamento das organizações, descobrindo tendências e especificidades, tornando mais célere e eficaz a acção dos gestores. Como consequência deste aumento efectivo de informação, o seu processamento através dos métodos tradicionais (e.g. com o recurso à estatística) tornou-se cada vez mais difícil e complexo [Goebel e Gruenwald, 1999].

As ferramentas convencionais de análise de dados (e.g. folhas de cálculo e consultas *ad-hoc*) sofrem de um sem número de senãos, que em nada potenciam a detecção de singularidades nesse emaranhado de dados [Michalski et al., 1998]. Esta limitação, combinada com o aumento significativo de dados armazenados em Bases de Dados (BD), tanto em número de registos como em complexidade (e.g. texto, *DBase*, *Informix*, *Oracle*, *Sybase*, *Microsoft -Structured Query Language* (MS-SQL), *MS-Access* e *MS-Excel*), é ainda condicionada pelos múltiplos domínios em que se aplica (e.g. Comercial, *Business Intelligence*, *Customer Relationship Management* (CRM), Saúde, Finanças, Ambiente, Investigação Científica e Internet).

Tendo em vista responder a um certo número de senãos (e.g. os que resultam de se considerarem grandes volumes de dados, múltiplas fontes de dados, múltiplos domínios de aplicação) instituiu-se a área da Descoberta de Conhecimentos em Base de Dados

(DCBD), recorrendo a modelos e técnicas de *Data Mining* para extrair informação, padrões de comportamento e/ou tendências, previamente desconhecidas, e isto de forma autónoma e porventura automática [Thuraisingham 1999] [Apte et al., 2002].

A investigação em *Data Mining* e DCBD tem evoluído e atraído a atenção dos investigadores, com base na ideia de que grandes volumes de dados são fonte de conhecimento, implicitamente representado e que pode ser extraído com ajuda de técnicas oriundas da área científica da Inteligência Artificial (IA) e da Aprendizagem Automática (AA) (e.g. Algoritmos Genéticos (AG), Redes Neurais Artificiais (RNAs), Árvores de Decisão e Sistemas de Classificação). Os objectivos em vista são os de estudar, de investigar e desenvolver processos que permitam extrair conhecimento que se revele inovador, útil e válido, implicitamente incluído nos grandes volumes de dados e representá-lo de forma acessível e legível para o utilizador.

## 4.11 Data Mining

Com o surgir do fenómeno de crescente acumulação de dados, emergiram novos problemas, que por sua vez deram lugar a novas oportunidades. Saiu reforçada a importância da informação, pois na directa proporção da sua quantidade existe uma porção de conhecimento que pode ser utilizada para otimizar as tomadas de decisões. Surgem também os Sistemas de Gestão de Bases de Dados (SGBD) que permitem o acesso avançado aos dados armazenados, mas estes são ainda uma pequena parte do que se pode alcançar.

As consultas tradicionais em BD (e.g. *Query reports*, MS-SQL) contrastam com a abordagem proporcionada pelo *Data Mining* simplesmente porque estão limitados a questões simples, tais como “Qual foi o meu rendimento total nos últimos 5 anos?”. A análise multidimensional, *On-Line Analytical Processing* (OLAP), possibilita fazer consultas mais complexas, tais como a comparação das vendas programadas e das efectuadas numa determinada região nos anos de interesse. O *Data Mining*, por seu lado, através do uso de algoritmos específicos ou de mecanismos de pesquisa, tenta descobrir padrões discerníveis e tendências nos dados e inferir regras para os mesmos [Fayyad et al., 1996] [Han e Kamber, 2001]. A análise de dados pode fornecer um conhecimento adicional acerca do problema em estudo, ao permitir ir além dos dados guardados explicitamente. É a partir dessa possibilidade que a utilização do *Data Mining* evidencia visíveis benefícios.



A principal diferença entre *Data Mining* e outras ferramentas de análise de dados reside na forma como estas exploram as relações entre os dados. As diversas ferramentas de análise disponíveis utilizam um método baseado na verificação, i.e. o utilizador constrói hipóteses sobre relações específicas e então corrobora-as ou refuta-as através da ferramenta. Esse modelo torna-se dependente da intuição e habilidade do utilizador em propor hipóteses interessantes, em manipular a complexidade dos atributos e em refinar a análise, baseada nos resultados de consultas à BD. Já o processo de *Data Mining* fica responsável pela geração de hipóteses, garantindo maior rapidez, aperfeiçoamento, autonomia e fiabilidade aos resultados.

No período de tempo que antecedeu o aparecimento do *Data Mining* como que se viveu uma revolução, em que afloraram novas abordagens de exploração dos dados armazenados em BD (Tabela 4.2) [Sousa, 2004].

Tabela 4.2: Períodos que precedem o Data Mining.

Época	Pergunta	Tecnologias	Características
Colecção de dados (1960's)	“Qual foi o meu rendimento total nos últimos 5 anos?”	Computadores, <i>tapes</i> , discos	Resultado retrospectivo de dados estatísticos
Acesso aos dados (1980's)	“Qual o número de vendas em Lisboa em Março?”	Bases de dados relacionais, <i>Structured Query Language</i> , <i>On-line DataBase Connectivity</i>	Resultado retrospectivo e dinâmico dos dados ao nível dos registos
<i>Data Warehousing</i> e Suporte à Decisão (1990's)	“Qual o número de vendas em Lisboa em Março? Induzir para o Porto”	<i>On-Line Analytic Processing</i> , Bases de Dados Multi-dimensionais, <i>Data Warehousings</i>	Resultado retrospectivo e dinâmico dos dados em níveis múltiplos
<i>Data Mining</i> (actualidade)	“O que é provável acontecer ao número de vendas no próximo mês? Porquê?”	Algoritmos avançados, multi-processamento, Bases de Dados de grandes dimensões	Resultado em perspectiva da informação

#### 4.11.1 Definição de Data Mining

O termo *Data Mining* pode ser referido através de outras formas como extração de conhecimento, arqueologia de dados (*Data Archeology*), colheita de informações (*infor-*

*mation harvesting*) e *Data Dredging* [Fayyad et al., 1996]. Consoante o autor, a abordagem e a área de especialização, são várias as definições encontradas para o termo *Data Mining*, sendo entre muitas, as mais correntes:

“*Data Mining* passa pela aplicação de algoritmos para extracção de padrões de comportamento a partir de dados, sem os passos adicionais do processo DCBD (tais como a inclusão de conhecimento anterior e a interpretação correcta dos resultados)” [...]

“*Data Mining* denota a procura de padrões de interesse numa determinada forma de representação de dados, ou conjunto das representações: classificação, árvores de decisão, regras de indução, regressão, segmentação, e assim por diante” [...]

“*Data Mining* é o processo de encontrar padrões e relações em BD de grande dimensão, previamente desconhecidos e potencialmente interessantes” [Fayyad et al., 1996].

“O termo *Data Mining* é vulgarmente utilizado como um sinónimo do processo de extracção de informação útil das BD [...]. O componente de *Data Mining* de DCBD está relacionado com os meios algorítmicos através dos quais os padrões são extraídos e os dados são enumerados” [Fayyad, 1997].

“*Data Mining* refere-se à aplicação de métodos de Aprendizagem Automática (AA), bem como de outros métodos, para a enumeração de padrões sobre os dados” [Michalski et al., 1998].

“*Data Mining* é o processo de extrair informação ou conhecimento de conjuntos de dados para os propósitos da tomada de decisão” [Chang et al., 2001].

“*Data Mining* refere-se à utilização de uma variedade de técnicas para identificar *nuggets*<sup>1</sup> de informação ou conhecimento decisivo em conjuntos de dados, extraíndo-os de tal forma que eles possam ser utilizados nas áreas de suporte à decisão, previsão e estimativa. Esses dados são por norma volumosos, mas de pouco valor na sua forma bruta, e de pouco serve acedendo-lhes de forma directa, é a informação escondida nos dados que é valiosa” [Clementine, 1998].

Uma análise a estas definições evidencia características comuns a todas elas, o que se pode sintetizar na seguinte definição: aplicação de métodos e técnicas em grandes BD, para encontrar tendências ou padrões com o intuito de descobrir conhecimento. Contudo, o *Data Mining* é apenas uma etapa de um processo mais abrangente, denominado de Descoberta de Conhecimento em Bases de Dados (DCBD), sendo sustentado por três

---

<sup>1</sup> nuggets - pedaços de conhecimento que podem ser utilizados no processo de decisão.

pilares fundamentais, dos quais depende o sucesso do projecto de *Data Mining* [Berry e Linoff, 2000]:

- (i) - Modelos e técnicas;
- (ii) - Os dados; e
- (iii) - A modelação de dados.

Os conceitos *Data Mining* e AA tendem a ser confundidos, no entanto existem dois critérios que os diferenciam. O primeiro, diz respeito ao número e à qualidade dos dados e o segundo está relacionado com os algoritmos utilizados.

Quanto ao primeiro critério, é de referir que as BD utilizadas em *Data Mining* são constituídas, na maioria dos casos, por grandes quantidades de dados com alguns problemas de falta de qualidade (e.g. ruído, dados omissos, dados incorrectos). Pelo contrário, na AA os dados são escolhidos de modo a evidenciarem apenas determinadas características, que o algoritmo deve tratar [Holsheimer et al., 1995].

No que concerne ao segundo critério, é de salientar que os algoritmos utilizados em *Data Mining* deverão apresentar a propriedade de escalabilidade, i.e. conhecidos os recursos do sistema (e.g. memória, velocidade de processamento) o tempo de execução do algoritmo deve crescer linearmente com o tamanho do conjunto de dados. Alguns algoritmos usados no domínio do *Data Mining* tiveram a sua origem na AA, tendo sofrido alterações no sentido de se tornarem escaláveis [Ramakrishnan, 1998].

### 4.11.2 Áreas Relacionadas com o Data Mining

*Data Mining* é uma área multidisciplinar e tem captado interesse de investigadores de áreas como a AA e a IA. Utiliza modelos cognitivos ou estratégias de aprendizagem, bem como recorre a novos paradigmas de aprendizagem para a aquisição de conhecimento automático, nomeadamente [Fayyad et al., 1996] [Han e Kamber 2001]:

- (i) - Reconhecimento de Padrões - nesta área concentra-se o estudo sobre as teorias e os algoritmos para extracção de padrões e modelos, principalmente a identificação de padrões especiais. Os modelos trabalham com

todo o conhecimento inferido devendo, no entanto, ser avaliados para determinar se reflectem ou não um conhecimento útil e interessante;

- (ii) - BD e *Data Warehouse* (DW) - na área de BD existem tecnologias específicas, bem como uma série de pesquisas que objectivam melhorar a exploração das características dos dados a serem trabalhados;
- (iii) - Estatística e Matemática - é frequente que modelos matemáticos ou estatísticos sejam construídos para a geração de regras, padrões e regularidades. No caso específico da Estatística, esta disponibiliza procedimentos e testes para as tarefas de *Data Mining* (e.g. verificar se estimativas e procedimentos de pesquisa estão consistentes sob determinados critérios de avaliação e identificar o grau de incerteza);
- (iv) - Sistemas Periciais - são programas complexos de IA criados com vista a resolver problemas do mundo real. Inicialmente, estes sistemas ofereciam apenas mecanismos para a representação do conhecimento, raciocínio e explicação. Posteriormente foram incorporadas ferramentas para a aquisição do conhecimento. Esta designação está em desuso, sendo substituída por termos como Sistemas de Apoio à Decisão, Sistemas Inteligentes, Sistemas de Conhecimento; e
- (v) - Visualização de Dados - a visualização de dados desempenha um papel importante já que em vários momentos existe a necessidade da verificação do processo, podendo-se recorrer a gráficos, ícones e figuras.

O mecanismo de aquisição ou extracção de conhecimento consiste num algoritmo ou combinação de algoritmos que acede aos dados registados na BD e os compacta ou os traduz em informação. Nas últimas décadas foram desenvolvidos vários tipos de algoritmos de aprendizagem. Distinguem-se, sobretudo, pela forma como traduzem a informação descoberta e pelo processo como é feita essa mesma descoberta, sendo alguns mais adequados a determinados tipos de problemas.

### 4.11.3 Abordagens do Data Mining

Em *Data Mining* existem dois tipos de abordagens ou estratégias que podem ser usadas para obter informação relevante [Berry e Linoff, 2000]:

- (i) - Modelo de verificação, abordagem directa ou *top-down* (usada quando se sabe o que pesquisar); e
- (ii) - Modelo de descoberta, abordagem indirecta ou *bottom-up* (quando a pesquisa incide sobre os dados sem indicação do objectivo de pesquisa).

A abordagem directa utiliza hipóteses e testa a sua validade de acordo com os dados. O responsável (utilizador) formula hipóteses e realiza consultas para as corroborar ou refutar. Neste tipo de abordagem recorre-se a uma variedade de técnicas, tais como consultas (*queries*), análises multidimensionais e visualização.

Na abordagem indirecta o sistema descobre automaticamente a informação importante que está implícita nos dados (e.g. descobrir numa BD de um banco grupos de clientes para uma campanha de *mailing*). Os dados são pesquisados de forma a encontrar padrões frequentes, tendências e generalizações, sem a intervenção ou orientação humana.

### 4.11.4 Objectivos do Data Mining

Os modelos e as técnicas de *Data Mining* podem ser aplicados a vários tipos de problemas e, conseqüentemente, terem objectivos distintos consoante o contexto em que estão a ser utilizados. Seguidamente são apresentados alguns desses objectivos.

#### 4.11.4.1 Classificação

A classificação consiste em encontrar uma função que associa um caso a uma classe, dentro de diversas classes (discretas) possíveis, de forma a referenciar um novo objecto de acordo com um dado modelo. As técnicas utilizadas em classificação recorrem a conjuntos de treino com exemplos, para construir modelos adequados à descrição das classes, que posteriormente são aplicados a dados não classificados [Thuraisingham, 1999].



Na Figura 4.9 é apresentado um exemplo de classificação para determinar se os empréstimos bancários podem ou não ser concedidos. O exemplo baseia-se em 23 casos de pedidos de empréstimo, e utiliza, como atributos, o valor do empréstimo e os rendimentos. Os dados foram catalogados em 2 classes onde o  $x$  representa os maus pagadores e o  $0$  denota os bons pagadores. Através de uma simples representação num espaço bi-dimensional, obtém-se duas classes de dados que permitem ao banco decidir a atribuição dos futuros empréstimos [Fayyad et al., 1996].

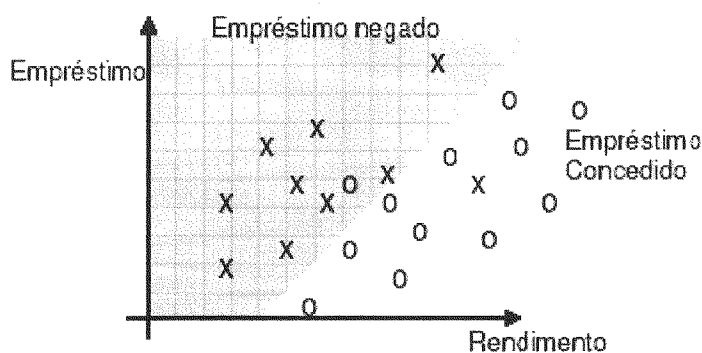


Figura 4.9: Exemplo de Classificação. Adaptada de Fayyad [Fayyad et al., 1996].

A classificação é o primeiro dos objectivos a atingir em *Data Mining*, sendo as Árvores de Decisão, Indução de Regras, AG e as RNAs as técnicas mais aplicadas [Chapman et al., 2000].

#### 4.11.4.2 Previsão

A previsão permite extrapolar e obter não só o valor de variáveis desconhecidas, com base em valores conhecidos para algumas das variáveis referenciadas (e.g. num dado problema) mas também a descoberta de padrões. Por exemplo, com base nas habilitações académicas de um certo número de funcionários de uma organização, no actual emprego e nos padrões da indústria, pode prever-se o seu salário nos próximos anos [Thuraisingham, 1999].

As técnicas de *Data Mining* mais aplicadas na previsão são as Árvores de Decisão, AG e as RNAs [Cortez et al., 1995] [Cortez et al., 1996] [Cortez et al., 1999] [Chapman et al., 2000] [Cortez, 2002].

### 4.11.4.3 Regressão Linear

Trata-se de encontrar uma função para uma previsão de uma variável, i.e. consiste na procura de uma função que represente de uma forma aproximada o comportamento de variáveis. Os métodos de regressão linear permitem a discriminação dos dados através da combinação dos atributos de entrada, o que equivale a determinar rectas de separação dos dados. A Figura 4.10 representa o resultado de uma regressão linear simples, onde o valor do empréstimo é considerado como uma função linear do rendimento. A técnica de *Data Mining* mais utilizada em regressão linear são as RNAs e podem ser usadas para prever, por exemplo, a quantidade de biomassa presente numa floresta, a probabilidade de um doente sobreviver através de resultados de um conjunto de diagnósticos, o consumo de novos produtos em função dos gastos com publicidade [Fayyad et al., 1996].

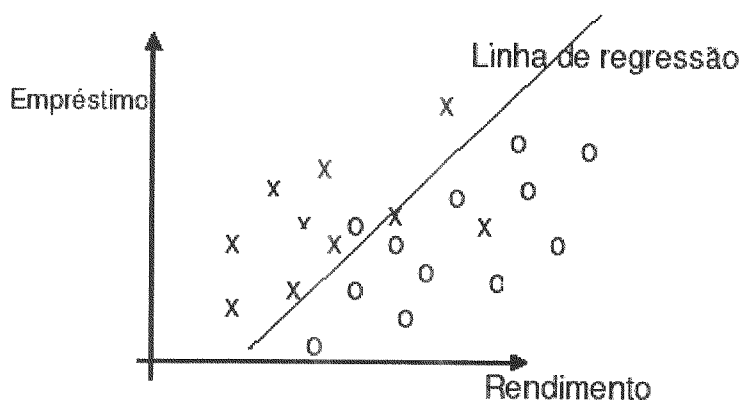


Figura 4.10: Exemplo de Regressão Linear. Adaptada de Fayyad [Fayyad et al., 1996].

### 4.11.4.4 Segmentação

A Segmentação ou *Clustering* permite identificar um conjunto finito de categorias ou segmentos para descrever os dados, e.g. identificar grupos homogéneos de objectos em que cada grupo é uma classe. Dentro da mesma classe os objectos são semelhantes e entre classes são discrepantes, i.e. a segmentação assegura que a similaridade inter-segmentos é baixa e que a similaridade intra-segmentos é alta [Thuraisingham, 1999].

No contexto da segmentação as ferramentas de *Data Mining* descobrem diferentes grupos, podendo ser aplicados a diversos problemas (e.g. detectar defeitos de fabrico,

encontrar grupos com afinidades para determinados produtos bancários, encontrar consumidores tipo em BD de *Marketing*, agrupar moradias de acordo com determinadas características). As técnicas de *Data Mining* mais aplicadas em segmentação são as RNAs. A Figura 4.11 mostra um exemplo de segmentação de empréstimos bancários, agrupando os dados em três agrupamentos distintos. Contudo, alguns dados podem pertencer a mais do que um grupo devido à intersecção destes.

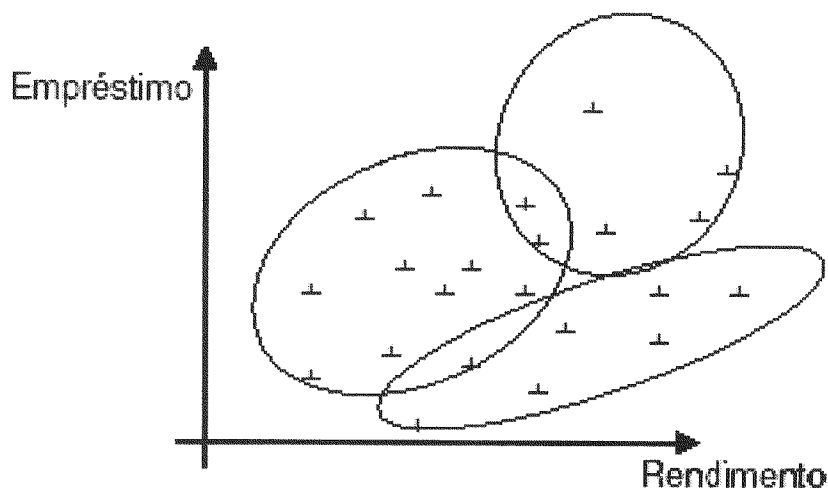


Figura 4.11: Exemplo de Segmentação. Adaptada de Sousa [Sousa, 2004].

#### 4.11.4.5 Associação ou Dependência

A associação pretende encontrar um modelo que descreva dependências significativas entre variáveis, i.e. identificar grupos de dados associados, identificar factos que possam ser directa ou indirectamente associados. De um modo geral, a associação é usada quando se procura identificar dados que possam ser colocados juntos, num mesmo pacote de negociação (e.g. produtos que podem ser comprados/vendidos conjuntamente).

As associações surgem quando várias ocorrências estão ligadas num único evento, podendo ser apresentadas ao nível estrutural ou ao nível quantitativo. Ao nível estrutural, o modelo é representado de uma forma gráfica, com variáveis localmente dependentes de outras. Ao nível quantitativo, o modelo especifica a força (ou peso) das dependências de acordo com uma escala numérica [Fayyad et al., 1996].

#### 4.11.4.6 Sumariação

A sumariação utiliza métodos para encontrar uma descrição compacta para um subconjunto de dados. Os métodos de sumariação mais sofisticados derivam de regras de resumo e da descobertas de relações funcionais entre variáveis. As técnicas de sumariação são aplicadas à análise exploratória de dados e à geração automática de relatórios.

#### 4.11.4.7 Visualização

A visualização trata da apresentação dos resultados (finais ou intermédios) obtidos com as ferramentas de *Data Mining*. Pretende dar-se a conhecer informações complexas através de um formato acessível ao utilizador (e.g. utilizando diagramas ou gráficos para a representação/visualização de padrões e tendências).

#### 4.11.4.8 Detecção de Desvios

A detecção de desvios visa a descoberta de alterações significativas nos dados, a partir de valores medidos previamente ou de valores normativos. A detecção de desvios pode ser aplicada na detecção de fraudes e na detecção de doenças [Thuraisingham 1999].

### 4.12 Descoberta de Conhecimento em Base de Dados

O termo DCBD foi formalizado em 1989 e refere-se a um processo que envolve a identificação e o reconhecimento de padrões numa BD de uma forma automática, i.e. a obtenção, sem nenhuma formulação prévia de hipóteses, informações relevantes e desconhecidas que podem ser úteis no processo de tomada de decisão [Fayyad et al., 1996].

O processo de DCBD depende de uma nova geração de ferramentas e técnicas de análise de dados, que envolve diversas etapas. A principal, que forma o núcleo do processo e que muitas vezes se confunde com ele, chama-se *Data Mining*. A Figura 4.12 esquema-

tiza o processo de DCBD e de *Data Mining*, onde a expressão DCBD se refere a todo o procedimento de descoberta do conhecimento e a designação *Data Mining* compreende a etapa referente à aplicação de algoritmos para extracção de padrões a partir dos dados [Fayyad et al., 1996].

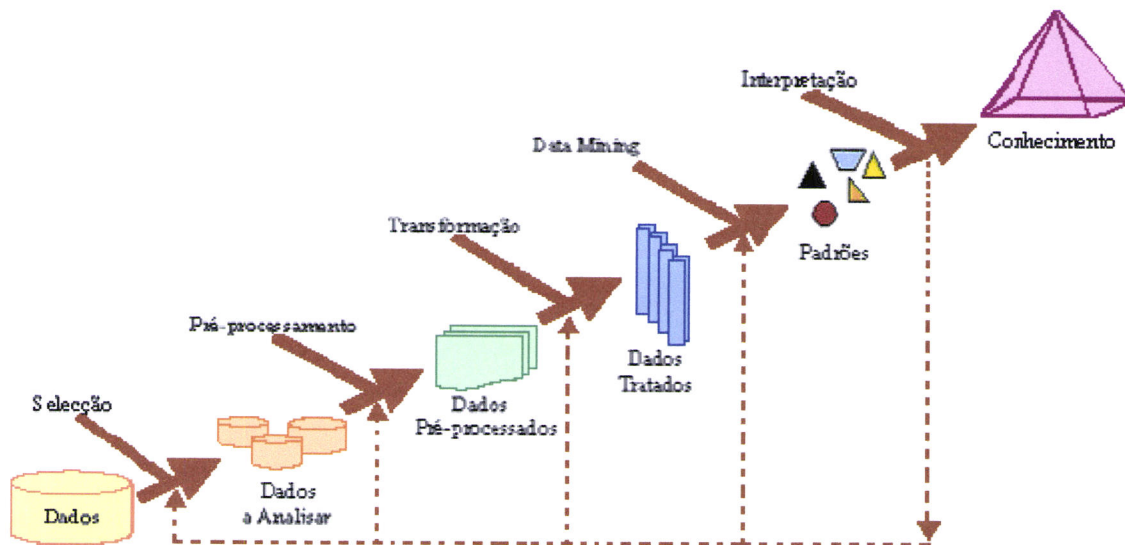


Figura 4.12: Processo de DCBD. Adaptada de Sousa [Sousa, 2004].

A Figura 4.12 mostra que o processo da aquisição de conhecimento é composto por cinco etapas que se passam a apresentar em seguida.

#### 4.12.1 Selecção

Uma vez definido o domínio sobre o qual se pretende executar o processo de descoberta de conhecimento, é feita a selecção e recolha do conjunto de dados ou variáveis necessárias ao estudo. Em seguida, é feita uma selecção ou segmentação dos dados de acordo com critérios que determinam subconjuntos (e.g. classes de qualidade da água).

A maioria das empresas e dos organismos já possuem BD, contudo, nem sempre todos os dados necessários estão disponíveis, o que exige um trabalho de compatibilização.

### 4.12.2 Pré-Processamento

Esta etapa comporta a fase de limpeza dos dados (*data cleaning*) que visa adequá-los aos algoritmos que posteriormente serão utilizados. Inclui também a integração de dados heterogêneos e a eliminação de dados incompletos e/ou inconsistentes.

Na fase de pré-processamento de dados são estabelecidas as estratégias para eliminar ou minimizar os problemas referentes à omissão de dados e à presença de ruídos e erros. Nesta etapa os dados são reconfigurados para assegurar a construção de uma BD uniforme e consistente (e.g. a variável sexo pode ser guardada como *m* e *f* ou *0* e *1*) [Fayyad et al., 1996].

### 4.12.3 Transformação

Os dados pré-processados devem ainda passar por uma transformação que os armazena no formato adequado, facilitando o uso das técnicas de Data Mining. Nesta fase, o uso de *Data Warehouse* (DW) tem-se expandido consideravelmente, pois nessas estruturas as informações estão armazenadas de uma forma mais eficiente [Inmon 1997].

### 4.12.4 Data Mining

A fase de *Data Mining* envolve a selecção dos métodos e técnicas a serem aplicados. Essa escolha depende fundamentalmente do objectivo do processo de *Data Mining* (i.e. segmentação, previsão, classificação, associação, sumariação, visualização). Nesta fase podem ser aplicadas diversas técnicas distintas ao problema (e.g. RNAs, Regras de Indução, Árvores de Decisão).

### 4.12.5 Interpretação e Avaliação

Os resultados do processo de descoberta de conhecimento podem ser visualizados de diversas formas. Estas, porém, devem possibilitar uma análise criteriosa para identificar a necessidade de retornar a uma qualquer das fases anteriores do processo de DCBD.

Os padrões identificados pelo sistema são interpretados como conhecimento, o qual pode ser usado para o suporte à tomada de decisão. Este conhecimento traduz-se, por exemplo, em entendimento dos conteúdos da BD ou na explicação de um fenómeno observado. O objectivo é, portanto, construir uma representação do conhecimento que seja acessível e clara para os utilizadores finais.

Antes de se poder dar o processo de DCBD como concluído deve fazer-se a verificação da qualidade da informação obtida. Esta verificação pode ser encarada como uma fase de Avaliação do Desempenho do mecanismo de aprendizagem. Inclui-se nesta etapa porque a determinação dos valores de Confiança e Suporte são dados extremamente importantes para a validação dos modelos e devem ser fornecidos ao utilizador final.

A Figura 4.12, que esquematiza a DCBD, pode sugerir a existência de um percurso linear, no entanto, isso geralmente não se verifica, uma vez que em cada etapa pode ser identificada a necessidade de retorno para cada uma das fases anteriores.

## 4.13 Problemas Inerentes ao Processo de DCBD

No processo de DCBD existem alguns problemas como a representação do conhecimento extraído, a complexidade da pesquisa, o controlo da operação de descoberta, a selecção do objectivo de *Data Mining* mais apropriado e a escolha dos métodos e técnicas adequados. Essas decisões dependem, essencialmente, dos objectivos do estudo, da BD utilizada, do domínio e da aplicação.

Além destes problemas, existem alguns desafios referentes às BD e ao próprio sistema a ser implementado [Fayyad et al., 1996]:

- (i) - Volume da BD - as BD com muitas tabelas possuem grande número de registos, podendo originar uma enorme variedade de padrões, combinações e hipóteses. A solução passa quer pela utilização de algoritmos que

enumerem todas as regras de associação, quer por métodos que incluam amostragem, métodos de aproximação e processamento paralelo.

- (ii) - Alta dimensionalidade da BD - a alta dimensionalidade (medida pelo número de campos de uma BD) aumenta de forma exponencial o espaço de procura e também a probabilidade do algoritmo encontrar padrões falsos. Uma solução possível é a utilização de métodos para identificar variáveis irrelevantes.
- (iii) - Dados inconsistentes - atributos com valores nulos e atributos importantes para o processo podem não estar presentes na BD. A solução passa pela utilização da validação cruzada ou de métodos estatísticos.
- (iv) - Ruído na BD - este é um tipo de problema muito comum que pode ser minorado recorrendo a métodos estatísticos para identificar variáveis ocultas e as suas dependências ou utilizar amostras com muitos dados.
- (v) - Dados irregulares - os dados podem estar arquivados em BD diferentes (e.g. *Access*, *Oracle*, *Informix*, *SQL*, *DB2*), e podem apresentar domínios diferentes para definir uma mesma informação e variar em termos de qualidade. A solução para este problema passa por construir um DW.
- (vi) - Interacção com o utilizador - os sistemas devem ser autónomos e extrair preferencialmente hipóteses úteis. Além disso devem ser configurados para a aplicação e para as BD de cada utilizador, de acordo com as suas necessidades e os seus conhecimentos.
- (vii) - Representação da informação - a informação descoberta deve ser clara, acessível para o utilizador, caso contrário pode interpretar-se o conhecimento erradamente. Uma solução possível será a inclusão de representações gráficas, linguagem natural e técnicas de visualização de dados.

Todos estes aspectos são importantes e devem ser considerados durante o processo de desenvolvimento de um sistema de DCBD. Nos projectos de DCBD, normalmente 80% do tempo refere-se às etapas de preparação dos dados, enquanto que os restantes 20% referem-se à etapa de *Data Mining* propriamente dita [Adriaans e Zantinge 1996].



## 4.14 Modelos e Técnicas Utilizadas no Processo de Data Mining

Os modelos e as técnicas mais utilizadas no processo de *Data Mining* são as Árvores de Decisão, as RNAs, os AG, a Indução de Regras, os Conjuntos Aproximados, as Redes de Bayes e os Sistemas de Classificação.

Por modelo entende-se a representação matemática de um sistema (ou processo) com o objectivo de o estudar, i.e. uma estruturação da representação do conhecimento com vista a atingir certo resultado [Rud, 2001]. Os modelos podem ser estáticos, quando representam o sistema num determinado instante ou podem ser dinâmicos quando o representam ao longo do tempo, acompanhando as alterações inerentes ao seu funcionamento.

Como técnica compreende-se o conjunto de processos baseados em conhecimentos científicos, de cálculo ou experimentação, utilizados para obter certo resultado.

Na construção de um modelo definem-se as características principais do sistema, recolhem-se os dados necessários para a sua construção (i.e. conjunto de treino) e para a sua posterior validação (i.e. conjunto de teste). Aos dados são aplicados algoritmos que deverão identificar padrões e relações. O modelo é a generalização de um padrão, i.e. a instânciação das variáveis.

## 4.15 Conclusões

Neste capítulo foi feita a caracterização e a descrição do funcionamento das Redes Neurais Unidireccionais (RNUs).

As RNUs revelam-se adequadas para uma aprendizagem supervisionada, sendo capazes de aprender correspondências entre conjuntos de entradas e saídas. As RNUs encerram todas as vantagens das RNAs, como seja a aprendizagem não linear, a tolerância ao ruído, a transparência e a flexibilidade, sendo recomendáveis para lidar com problemas de regressão e de classificação.

A arquitectura de uma RNU é definida por um conjunto de nodos, agrupados por camadas, conexões unidireccionais pesadas, definidas entre os nodos, e um conjunto de funções de activação.

Neste capítulo abordou-se ainda o processo de Descoberta de Conhecimento em Bases de Dados. Este processo recorre a modelos e técnicas de *Data Mining* para extrair informação útil (que, de outro modo, dificilmente seria revelada) e representá-la de forma acessível e legível para o utilizador.

O processo de DCBD é essencialmente composto por três etapas: o pré-processamento, o *Data Mining* e o pós-processamento, sendo cada uma destas etapas constituídas por várias sub-tarefas. Na etapa de pré-processamento, os dados são reconfigurados com vista à obtenção de uma BD consistente. A etapa de *Data Mining* visa a descoberta de conhecimento, enquanto na etapa de pós-processamento se converte o conhecimento obtido numa linguagem perceptível para o utilizador final e se procede à análise do desempenho, recorrendo-se a métricas de avaliação (e.g. a confiança e o suporte).

## Capítulo 5

# Uma Agência para a Integração, Difusão e Arquivo de Informação sobre a Qualidade da Água de Albufeiras

Nos dias de hoje tem-se acesso a uma grande quantidade de dados físico-químicos, biológicos e meteorológicos para muitas das albufeiras portuguesas, principalmente aquelas que se destinam à produção de água para o abastecimento das populações. Contudo, estes dados são obtidos pelas mais diversas entidades e encontram-se dispersos.

Por outro lado, os técnicos ambientais têm dificuldade em aceder de uma forma expedita a toda esta informação para que possam avaliar a qualidade da informação disponível, e com esta fazer previsões e tomar as melhores decisões em tempo útil.

Tendo em atenção o exposto, passar-se-á a apresentar o trabalho desenvolvido com o objectivo de concretizar uma arquitectura integrada e distribuída para o Sub-sistema de Arquivo e Difusão de Análises Físico-Químicas (SADAFQ), o Sub-sistema de Arquivo e Difusão de Análises Biológicas (SADAB), o Sub-sistema de Arquivo e Difusão de Dados Meteorológicos (SADDM) e o Sub-sistema de Apoio ao Técnico (SAT), em que a autonomia, a aprendizagem e a cooperação são mais valias a que se recorre como forma de melhorar, não apenas o seu desempenho, mas também a sua esfera de aplicabilidade. O sistema, uma Agência para a Integração, Difusão e Arquivo de Informação sobre a Qualidade da Água de Albufeiras (AQUA), foi elaborado com o recurso a agentes inteligentes, que vão responsabilizar-se por múltiplas tarefas, como a comunicação com os diferentes sub-sistemas a operar nos laboratórios, enviar e receber informação (e.g. análises físico-químicas, análises biológicas, dados meteorológicos), gerir, guardar e responder a pedidos de informação, usando como plataforma de comunicação o paradigma do Quadro Negro (QN)

[Santos, 1999]. Neste são definidas diferentes formas de comunicação, uma para cada um dos sub-sistemas **AQUA**, ou seja:

- (i) - O Sub-sistema de Informação Físico-Química (SIFQ), que se assume como o identificador da estrutura de comunicação do SADAFQ;
- (ii) - O Sub-sistema de Informação Biológica (SIB), que denota o identificador da estrutura de comunicação do SADAB; e
- (iii) - O Sub-sistema de Informação Meteorológica (SIM), que responde pelo identificador da estrutura de comunicação do SADDM.

Estas estruturas foram adoptadas para potenciar a resolução de problemas através de um recurso à Lógica Matemática [Neves, 1984][Engelmore et al., 1988] [Jagannathan et al., 1989] [Neves et al., 1994] [Kowalski, 1994]. Por outro lado, e apesar dos muitos adjectivos que abonam a favor da Programação em Lógica (PL), esta não disponibiliza as abstracções necessárias para o desenvolvimento de *software* no contexto dos Sistemas MultiAgente (SMA). Estas limitações estão identificadas e têm motivado o aparecimento de várias extensões à Programação em Lógica, e em especial no contexto da Inteligência Artificial Distribuída (IAD), i.e. dos Sistemas de Processamento Distribuído (SPD) e Sistemas MultiAgente (SMA). Uma das extensões criadas, que é aplicada na especificação do sistema **AQUA**, tem os seus alicerces na Programação em Lógica Contextual (PLC) [Cavedon e Tilhar, 1995] [Denti et al., 1995] [Giunchiglia et al., 1993] [Cimatti e Serafini, 1995] [Santos e Neves, 1998]. Para ilustrar o conceito de base que está na origem da PLC, suponha-se que *Teo* é uma teoria lógica, i.e. um conjunto identificado de cláusulas universalmente quantificadas. Colocar uma expressão na forma de uma implicação:

$$Teo \rightarrow BC_{agente}$$

em relação a um programa *Prog*, significa que a prova de  $BC_{agente}$  seja realizada a partir do programa *Prog* a que se associam os axiomas ou hipóteses contidos na teoria *Teo*.

A semântica operacional de todo este processo pode ser formalizada através da regra de inferência:

$$\frac{Prog \cup Teo \vdash BC_{agente}}{Prog \vdash Teo \rightarrow BC_{agente}}$$

onde o símbolo  $\vdash$  denota a relação de derivabilidade e  $\rightarrow$  denota a implicação lógica. Esta regra é facilmente implementada se o programa *Prog* for enriquecido com as cláusulas de *Teo* antes de tentar provar  $BC_{agente}$  e, a seguir, quer seja ou não seja  $BC_{agente}$  objecto de prova, atendendo a que ocorreu uma transição de estado, há que substituir *Prog* por  $Prog \cup Teo$ . Esta é a ideia chave subjacente à PLC e faz com que cada teoria lógica seja concebida numa perspectiva de alguma incompletude relativamente ao conhecimento e numa óptica de um mundo aberto. Os dois programas passam a ser construídos pela composição dinâmica de teorias, como forma de criar as bases de conhecimento que levam à designação de contextos.

O programa *Prog* é aqui dado em termos da extensão do meta-predicado  $demo(T, V)$ , em que *demo*, *T* e *V* denotam, respectivamente, o demonstrador de teoremas, o teorema ou conjunção de termos lógicos que será objecto de uma tentativa de demonstração, e um valor de verdade tomado do conjunto  $\{verdadeiro, falso, desconhecido\}$ . O valor de verdade *desconhecido*, que mede a qualidade da informação em termos das excepções às extensões dos predicados que definem o contexto em que o teorema *T* é colocado, toma valores do intervalo  $]0, 1[$ .  $demo()$  é dado através das produções:

$$demo(T, verdadeiro) \leftarrow T.$$

$$demo(T, falso) \leftarrow \neg T.$$

$$demo(T, desconhecido) \leftarrow \text{não } \neg T \wedge \text{não } T.$$

em que  $\neg$  denota a negação forte ou clássica, e *não* denota a negação por falha na prova ou negação fraca.

Esta linha de pensamento segue o trabalho desenvolvido por Neves [Neves, 1984] e Kowalski [Kowalski, 1994], no que respeita aos processos de assimilação de conhecimento em bases de conhecimento. Torna possível definir contexto como a semântica de um mundo possível, que emprega na sua caracterização um vocabulário extremamente rico, sob o ponto de vista sintáctico, de termos para representar o tempo, acontecimentos, situações e teorias, i.e. ao representarem-se situações, sintacticamente, como teorias e, agregando linguagem objecto e linguagem meta, tem-se uma alternativa à semântica de situações, comum na teoria dos modelos [Alchourrón et al., 1985], a qual se serve da relação entre linguagem e experiência para especificar a noção de consequência lógica.

O contexto pode ser interpretado (a um determinado nível de abstracção) como um ambiente, uma hierarquia de componentes de *software* com herança, uma instância de

um objecto, uma linha de raciocínio, um subsistema, um agente ou uma agência (comunidade de agentes), i.e. o termo contexto designa aqui uma situação em que um dado problema pode ser resolvido, e que é definida na forma:

- (i) - O problema;
- (ii) - A informação constante do universo de discurso, ou sejam as proposições dadas em contexto, e que caracterizam o problema em análise;
- (iii) - O conhecimento relevante, ou seja o conhecimento a partir do qual a solução do problema pode ser obtida;
- (iv) - As indicações metodológicas, ou seja quando e como proceder para se atingir uma solução; e
- (v) - As entidades que comungam do processo de resolução do problema.

O termo *problema* é para ser tomado numa perspectiva minimalista e refere-se ao objectivo a ser atingido. É para ser interpretado em termos de uma conjugação de termos lógicos ou teorema dado em contexto, ou seja:

- (i) - Que definem o espaço de soluções, o conhecimento relevante;
- (ii) - Que definem os meios a utilizar na solução do problema e as indicações metodológicas; e
- (iii) - Que indicam o caminho a seguir para se obter uma solução para o problema.

A resolução de um problema, se complexo, poderá suscitar uma grande variedade de questões (e.g. divisão em sub-problemas, problemas relacionados e/ou derivados) e, desse modo, fazer com que o caminho a seguir contemple uma pluralidade de contextos. Também se deve distinguir o contexto para o grupo (e.g. as comunicações entre as entidades que comungam do processo de resolução de problemas), do contexto (individual) em que se move cada um desses participantes, sendo todos estes contextos possivelmente diferentes, mesmo nos casos em que existe uma cooperação estreita entre as partes.

Para a especificação do sistema AQUA foi utilizado um formalismo orientado para a especificação, modelação e raciocínio sobre um SMA que utiliza a noção de contexto e incorpora algumas facilidades da programação orientada ao objecto, nomeadamente a

abstracção, o encapsulamento, a modularidade e a hierarquia [Cavedon e Tilhar, 1995] [Neves e Machado, 1997]. Através deste formalismo é possível de uma forma flexível e modular, especificar:

- (i) - Os componentes do sistema (i.e. os agentes);
- (ii) - O processo de socialização, quer de agentes, quer de Sistemas MultiAgente, quer porventura de outras formas de agregação e cooperação que se possam estabelecer entre outros tipos de entidades;
- (iii) - O procedimento de coordenação dos agentes; e
- (iv) - O comportamento global do sistema.

Este formalismo permite a especificação de cada uma das entidades que compõem o sistema, uma vez que cada entidade apenas reporta ao seu próprio universo, assim como às interacções entre si. Nesta óptica, cada agente é modelado num contexto em separado, com a sua própria lógica, o que apresenta algumas vantagens competitivas. Há que atender ainda a que:

- (i) - Um agente é uma entidade tão simples quanto possível, de tal modo que apenas precisa de ser relacionado com as proposições no mundo ou universo de discurso com o qual está directamente relacionado;
- (ii) - Agentes com diferentes atributos podem ser modelados como tendo interesses distintos ou, pelo menos, perspectivas diferentes do mesmo mundo ou universo de discurso; e
- (iii) - O desempenho de um agente pode ser maximizado, desde que lhe seja limitado o acesso a conhecimento, na medida em que se está em presença de conhecimento que se aplica à resolução de um certo tipo de problema.

Por outro lado, quaisquer afirmações dadas em contexto não devem ser questionadas dentro desse contexto, i.e. são consideradas como sendo contextualmente e necessariamente verdadeiras. Nesse sentido limitam as possíveis soluções para o problema. Além do mais, determinam parcialmente o conjunto de operações justificáveis e, por isso, a lógica subjacente ao contexto. A função do conhecimento relevante é diferente. Este não determi-

na as soluções possíveis para o problema, mas impõe condições para uma possível solução correcta, ou pelo menos para a eliminação de possíveis soluções. As indicações metodológicas especificam as operações que devem ou não devem ser realizadas de forma a atingir a solução ou aproximar-se desta. Não raras vezes, formam uma receita explícita para a resolução do problema; porém, se são fracas, apenas dirigem o processo de resolução em si. Em geral, as indicações metodológicas ajudam a determinar os passos que o processo de resolução do problema deve contemplar dentro do contexto.

É também importante perceber que os elementos de um contexto podem variar de contexto para contexto. Isto não significa apenas que os invariantes de um contexto podem estar ausentes em um outro, e o mesmo conhecimento preencher funções diferentes em contextos diferentes, podendo mesmo levar a gerar contradições ou rejeições. Se num dado contexto não for possível encontrar uma solução para o problema, o processo de resolução de problemas deve abordar um problema derivado num outro contexto, e colocar, possivelmente entre outras, as questões:

- (i) - Quais são os factores do anterior contexto que são responsáveis pela inexistência de uma resolução para o problema?; e
- (ii) - Quais desses elementos possuem uma justificação para uma alteração do seu valor (na forma de uma extensão ou de uma substituição) de forma a tornar o problema ou parte deste resolúvel?

Por exemplo, atenda-se ao predicado *sair-com*( $X, Y$ ), o qual denota que o *indivíduo*  $X$  sai com a *rapariga*  $Y$ . Considere-se agora a situação em que o João sai com *alguém*, mas não se sabe com quem (i.e. o João tem um conjunto de amigas com quem sai, mas não as enumera). Em PLC, isto pode ser dado pelos axiomas ou produções:

$$\begin{aligned} & \text{sair-com}(\text{joão}, \text{alguém}). \\ & \neg \text{sair-com}(X; Y) \leftarrow \text{não sair-com}(X, Y), \\ & \qquad \qquad \qquad \text{não excepção}_{\text{sair-com}}(X, Y). \\ & \text{excepção}_{\text{sair-com}}(X, Y) \leftarrow \text{sair-com}(X, \text{alguém}). \end{aligned}$$

Programa 5.1: João sai com *alguém*, não se sabe é com quem.



Se agora for colocada ao sistema a questão, com quem é que sai o João?, porventura dada na forma:

$\text{sair-com}(\text{joão}, X)?$

obtém-se como solução o termo *alguém* (i.e.  $X = \text{alguém}$ ). Ora esta solução para o problema pode estar muito aquém do pretendido. Se se pudessem enumerar as amigas do João ter-se-ia como solução  $X \in \{\text{Joana}, \text{Filipa}, \text{Aldina}\}$  (Programa 5.2). Neste caso o termo *alguém* foi substituído pelos elementos do conjunto  $\{\text{Joana}, \text{Filipa}, \text{Aldina}\}$ .

$\neg \text{sair-com}(X, Y) \leftarrow \text{não sair-com}(X, Y),$   
 $\text{não excepção}_{\text{sair-com}}(X, Y).$

$\text{excepção}_{\text{sair-com}}(\text{joão}, \text{joana}).$

$\text{excepção}_{\text{sair-com}}(\text{joão}, \text{filipa}).$

$\text{excepção}_{\text{sair-com}}(\text{joão}, \text{aldina}).$

Programa 5.2: João sai com *alguém*, sendo que *alguém* é tomado do conjunto finito de valores  $\{\text{Joana}, \text{Filipa}, \text{Aldina}\}$ .

A partir de uma posição estritamente formal, é assim possível integrar diferentes tipos de situações, quer em termos de argumentação, do raciocínio, da modificação da interpretação de constantes lógicas e não lógicas ou mesmo de linguagens de programação com características particulares. A Lógica Contextual aqui estendida nos mesmos termos em que decorreu o processo para a Programação em Lógica, captura os aspectos dinâmicos típicos do raciocínio humano. A dinâmica da prova em Programação em Lógica Contextual não invalida os mecanismos de inferência utilizados em PLE, mas restringe a sua aplicação a premissas que verificam certas condições, i.e. tudo passa por equacionar os processos de raciocínio a partir de um conjunto de hipóteses consideradas consistentes até ser provada a sua inconsistência. Este processo é dinâmico pelo facto de a derivação de uma inconsistência (ou a disjunção de inconsistências) poder obrigar a rever as derivações calculadas a jusante. Por exemplo, a derivação de  $q$  a partir de  $\text{não } p$  e  $p \vee q$  pressupõe um comportamento consistente de  $p$ : se simultaneamente  $p$  e  $\text{não } p$  são deriváveis das premissas,  $p \vee q$  é uma consequência lógica de  $p$  mas as premissas não garantem  $q$ . Assim, se  $q$  é derivável de  $\text{não } p$  e  $p \vee q$  em algum estágio da prova e  $p$  é derivável num estágio posterior, então  $q$  não pode

ser considerado derivável nesse contexto a partir desse estágio. Tem-se então que um Sistema MultiAgente (SMA) passará a ser especificado como um multi-sistema composto por agentes, sendo caracterizado por:

- (i) - Um contexto; e
- (ii) - Um conjunto de regras ponte que funcionam como o interface entre os agentes, sistemas de agentes e/ou suas comunidades, i.e. um conjunto não vazio de subsistemas, correspondendo cada um destes a um multi-sistema com o seu próprio contexto.

As regras ponte, que não são mais que regras de reflexão, podem agora ser descritas nos termos:

- (i) - Regras subordinantes na forma  $\frac{ocorre(e, i)}{C_k : ocorre(e', i')}$ , denotando que se no sistema ocorre o evento  $e$  no instante  $i$ , então no contexto  $C_k$  ocorrerá o evento  $e'$  no instante  $i'$ ;
- (ii) - Regras subordinadas na forma  $\frac{C_k : ocorre(e, i)}{ocorre(e', i')}$ , denotando que se num determinado contexto  $C_k$  ocorre o evento  $e$  no instante  $i$ , então no sistema ocorrerá o evento  $e'$  no instante  $i'$ ; e
- (iii) - regras ordinárias na forma  $\frac{C_k : ocorre(e, i)}{C_l : ocorre(e', i')}$ , denotando que se num determinado contexto  $C_k$  ocorre o evento  $e$  no instante  $i$ , então no contexto  $C_l$  ocorrerá o evento  $e'$  no instante  $i'$ .

As regras ponte dão corpo, de uma forma explícita, à interação e coordenação entre dois sub-sistemas. As restrições impostas por estas regras reflectem a esperança de que qualquer procedimento seja efectuado (com privacidade) dentro dos componentes do sistema (racionalidade emergente). Os componentes de um sistema não são visíveis fora desse sistema. A um dado sistema, agente ou SMA, estão associadas:

- (i) - Um conjunto de fluentes, que designam as propriedades do agente, e que se podem alterar ao longo do tempo;

- (ii) - Um conjunto dos valores que cada fluente pode tomar num determinado instante e contexto;
- (iii) - Um conjunto de eventos que o agente é capaz de executar; e
- (iv) - Um conjunto de instantes de tempo.

A cada tipo de evento estão ainda associados dois conjuntos de pares do tipo fluente-valor, as pré-condições e as pós-condições de cada evento.

## 5.1 Resolução de Problemas em Ambiente Distribuído

A arquitectura aqui apresentada foi concebida para ser de alto desempenho com regras de distribuição de cargas computacionais geradas por um processo de aprendizagem e cooperação [Alves, 2002]. Está implementada num conjunto de computadores constituídos por um conjunto de nodos ligados numa *intranet*, tendo sido implementado em *Windows* e *Linux*, utilizando a linguagem XML para a comunicação entre entidades [Abelha et al., 2003] [Abelha, 2004].

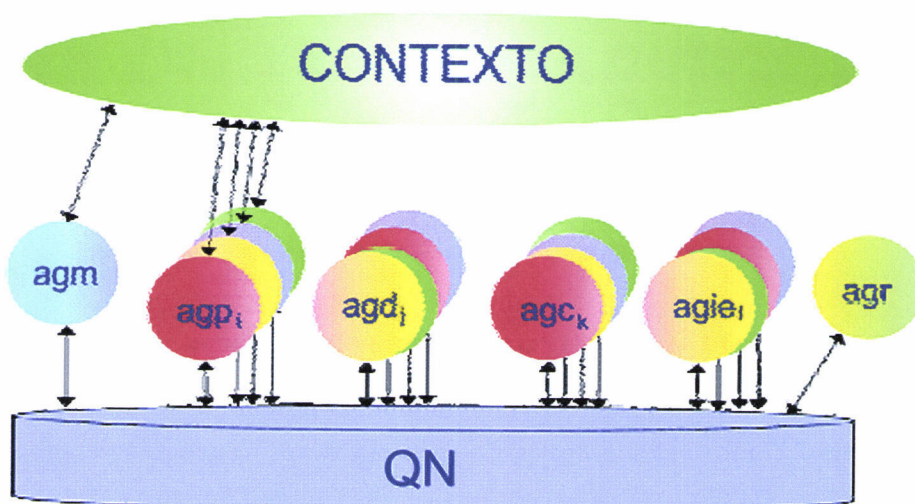


Figura 5.1: A arquitectura da agência AQUA. Adaptada de Abelha [Abelha, 2004].

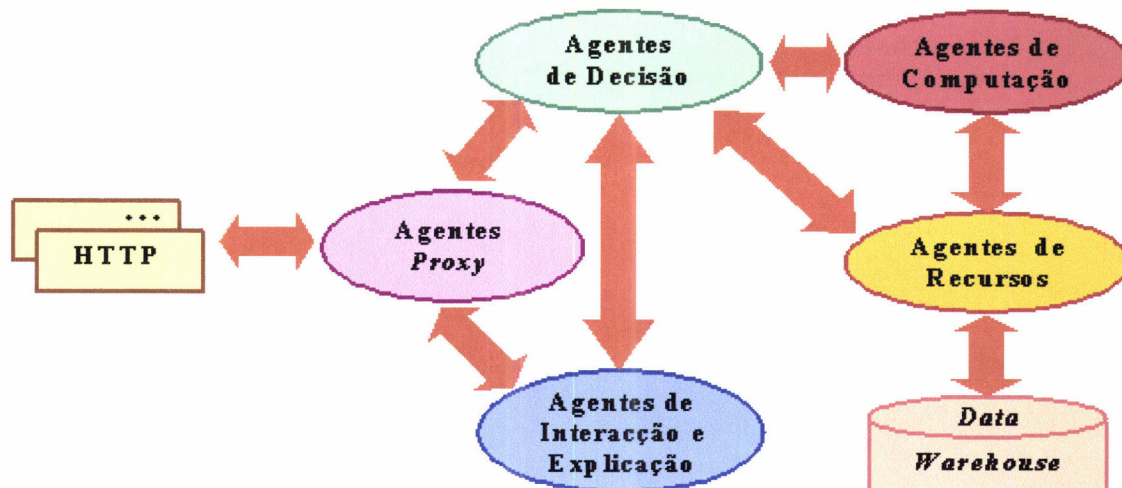


Figura 5.2: Os agentes da agência AQUA. Adaptada de Abelha [Abelha, 2004].

## 5.2 O Modelo AQUA

O modelo **AQUA** é uma agência (ou um conjunto de SMA) em que estas entidades são entendidas como utensílios computacionais, podendo ser entidades de *software* e/ou *hardware*, que se assumem como autónomas, reactivas, pro-activas e sociais.

A zona de memória partilhada da agência, designada por Quadro Negro (QN), que para além de suportar a comunicação entre agentes e SMA, facilita também a partilha de conhecimento entre estes.

## 5.3 AQUA - Modelação Simbólica

A necessidade de evitar falhas ao nível global do sistema, mesmo na situação de alguns dos seus componentes entrarem em ruptura, levou à introdução de um mecanismo de meta-execução como unidade básica de controlo de procedimentos e protecção dos mesmos.

Por forma a tornar este processo simples, o desenvolvimento de sistemas computacionais em que há demonstração de teoremas, socorre-se de um conjunto de facilidades como lemas, linguagens de cariz tático e de procedimentos de demonstração de teoremas, de tal modo que a construção de uma prova possa ser automatizada tanto quanto pos-

sível. Isto é, uma derivação formal de um teorema pode apenas ser concretizada através do recurso a tais facilidades, uma vez que esta (i.e. a derivação formal de um teorema) é de uma envergadura tal que não se tornaria prático para o utilizador de um tal sistema introduzir manualmente cada um dos termos que corporiza o teorema a demonstrar. Uma linguagem táctica (e.g. XML) ou outra facilidade similar, é apenas o meio de que o utilizador se *so*corre para automatizar o seu trabalho.

Uma outra forma para se olhar para o que acaba de ser exposto passa por se ser especialista na demonstração de teoremas colocados num certo contexto, contexto este elaborado com o propósito de derivar, de uma forma dinâmica, novos contextos. A ideia central que está por detrás deste juízo, está em que, dada uma descrição formal de uma *Teoria* que utiliza a mesma linguagem com que se elaborou o contexto, é possível, associando à *Teoria* o programa *Prog*, já referido anteriormente, derivar novos contextos. De facto, a descrição de uma teoria num dado contexto é uma meta-teoria, e os teoremas derivados a partir desta teoria são meta-teoremas na meta-teoria.

As vantagens e utilidade da meta-interpretação como técnica utilizada na programação em lógica é amplamente reconhecida tanto na teoria como na prática. Meta-interpretadores têm sido utilizados com evidentes vantagens quer no desenvolvimento de ferramentas algorítmicas de correcção de erros, controlo de expressões em programas lógicos ou para flexibilizar sistemas baseados em conhecimento. Os meta-interpretadores em lógica mais comuns estão ao nível da redução de cláusulas ou axiomas a utilizar na demonstração de teoremas. O meta-interpretador que se lista de seguida é a forma mais comum de uma tal situação:

demo(verdade).

demo((P,Q)) : -

demo(P),

demo(Q).

demo(P) : -

clause((P : - Corpo)),

demo(Corpo).

demo(P) : -

sys(P),

P.

Aqui, a extensão do predicado *demo()* é acrescida de predicados do sistema, permitindo a execução ao nível meta de *demo(objectivo)*. Esta meta-chamada executa simplesmente o objectivo *objectivo* dentro de *demo*, sendo o resultado computacional equivalente ao obtido pela execução directa de *objectivo*.

## 5.4 Estrutura Global do Sistema

Para implementar este sistema, distribuído por natureza, todos os interfaces com o utilizador final usam tecnologias *web*, o que torna o sistema extremamente amigável. É executado via *browser*, com ferramentas *freeware* e sistemas de bases de dados com licenças disponibilizadas directamente pelo Ministério do Ambiente e da Agricultura. A *intranet* foi implementada usando um PC com o sistema operativo *LINUX* e o motor de base de dados *ORACLE*; computadores com o sistema operativo *Windows* e *Internet Information Server* (*IIS*); uma rede de computadores com o sistema operativo *LINUX* para repositório de dados, vídeos e/ou relatórios de qualidade da água de albufeiras, rios e de laboratórios. As ligações com o exterior são potenciadas via *router* ISDN com RAS (*Remote Access Service*) [Alves et al., 2001] [Alves et al., 2002].

## 5.5 Data Warehousing e Melhores Decisões

A chave para o sucesso para quem trabalha em sistemas que têm em vista estabelecer uma medida da qualidade da água, quer ao nível da direcção, quer ao nível da intervenção pública, passa por serem tomadas as melhores decisões em tempo útil. Ora uma *data warehouse* pode prestar em cima do acontecimento informação correcta (ou de grande utilidade), quer do tipo ambiental quer do tipo financeiro, bem como disponibilizar ferramentas de análise dos dados, que se transformarão num factor decisivo e numa mais valia no momento de uma tomada de decisão. Uma integração efectiva dos dados deverá dar aos interventores de uma qualquer organização ambiental ou do Ministério da Agricultura, não só um melhor conhecimento do custo da água, bem como dos sistemas de tratamento de água mais adequados, mas também uma medida da qualidade dessa água [Abelha et al.,

2004]. Mesmo a localização de focos de poluição e a resolução dos problemas por estes levantados pode passar por uma *data warehouse*.

Há uma imensidade de dados a tratar quando se olha para a prática ambiental, mas será que essa informação tem significado? Aqui o maior problema passa por transformar os dados em informação com um mínimo de conteúdo, a ser utilizada atempadamente e de uma forma em que seja possível explorar certos tipos de relacionamentos. *Data warehousing* e o uso de sistemas inteligentes tornam-se componentes centrais para qualquer organização, e a sua inclusão em sistemas de apoio à decisão tornou-se mandatária. Para a prática da agricultura e para o controlo ambiental chegou o momento de equacionar o mesmo tipo de questões. No dia a dia, o técnico pode aceder a informação sobre a qual correm as aplicações de natureza administrativa, e combiná-la com dados laboratoriais ou mesmo informação de cariz financeira oriunda de outros sistemas, e carregá-la numa *data warehouse*. O desafio passa por transformar e carregar estes dados numa estrutura ou modelo  $n$ -dimensional, atendendo ao tipo de questões a que o sistema terá de responder. Ao olhar o problema segundo esta perspectiva, de cunho operacional, há que mencionar as Aplicações de Processamento Analítico Online (OLAP). Aqui os dados são organizados sob a forma de cubos de dados, segundo diferentes categorias ou dimensões, a que se associam medidas ou valores de natureza quantitativa. Com este modelo organizativo tem-se como objectivo facultar ao utilizador a possibilidade de recorrer a diferentes formas de análise e interpretação de dados, socorrendo-se de dados que foram agregados atendendo aos desígnios do utilizador.

As medidas ou valores de natureza quantitativa já referidas no texto, irão contemplar os parâmetros da qualidade da água de que o pH, o teor em oxigénio dissolvido, as concentrações em nitratos, fosfatos e em clorofila-*a* são exemplos.

Considere-se, por exemplo, o caso em que as dimensões do cubo contemplam o pH, o teor em oxigénio dissolvido (oD) e as concentrações das espécies nitrato, fosfato e clorofila-*a*, medidas durante um certo período de tempo (em meses). Considere-se, ainda, que estes elementos são dados, respectivamente, através das extensões dos predicados  $\text{pH} : X, Y \rightarrow \text{booleano}$ ,  $\text{oD} : X, Y \rightarrow \text{booleano}$ ,  $\text{nitratos} : X, Y \rightarrow \text{booleano}$ ,  $\text{fosfatos} : X, Y \rightarrow \text{booleano}$ , e  $\text{clorofila} : X; Y \rightarrow \text{booleano}$ , e respectivas excepções às suas extensões. Os meses que irão ser objecto de estudo são os meses de Janeiro, Fevereiro, Março e Abril. Se se tiver agora em linha de conta que para o mês de Janeiro não se conhece o valor do teor em oxigénio dissolvido nem de clorofila-*a*, embora esta possa ser tomada do conjunto de valores  $\{0.26, 0.28, 0.32\}$ . Quanto ao mês de Fevereiro não se conhece o valor do pH nem de

fosfatos, embora esta possa ser tomado do conjunto de valores  $\{0.04, 0.05, 0.07\}$ . Quanto ao mês de Março, sabe-se o teor em oxigénio dissolvido e a concentração em clorofila-*a*, embora não se conheça o pH e as concentrações de nitratos e de fosfatos embora estas possam ser tomadas, respectivamente, dos conjuntos de valores  $\{0.24, 0.32\}$ ,  $\{0.07, 0.08\}$  e  $\{0.16, 0.18, 0.21\}$ . No que respeita ao mês de Abril, nada se sabe sobre a concentração de clorofila-*a*, embora se saiba que os valores de pH, do oxigénio dissolvido e dos nitratos possam ser tomados, respectivamente, dos conjuntos de valores  $\{0.24, 0.26, 0.42\}$ ,  $\{0.47, 0.48\}$  e  $\{0.13, 0.15, 0.17, 0.18\}$ . Por outro lado sabe-se que o teor em fosfatos é 0.12. Ora o que acaba de ser exposto, em termos da PLE, toma a forma:

(i) - pH

pH(janeiro, 0.32).

pH(fevereiro, pH).

$\neg$ pH(X,Y)  $\leftarrow$  não pH(X,Y),  
 não excepção<sub>pH</sub>(X,Y).

excepção<sub>pH</sub>(X,Y)  $\leftarrow$  pH (X, pH).

excepção<sub>pH</sub>(março, 0.24).

excepção<sub>pH</sub>(março, 0.32).

excepção<sub>pH</sub>(abril, 0.24).

excepção<sub>pH</sub>(abril, 0.26).

excepção<sub>pH</sub>(abril, 0.42).

(ii) - Oxigénio dissolvido (oD)

oD(janeiro, oD).

oD(fevereiro, 0.46).

oD(março, 0.45).

$\neg$ oD(X,Y)  $\leftarrow$  não oD(X,Y),  
 não excepção<sub>oD</sub>(X,Y).

excepção<sub>oD</sub>(X,Y)  $\leftarrow$  OD(X, oD).

excepção<sub>oD</sub>(abril, 0.47).

excepção<sub>oD</sub>(abril, 0.48).



(iii) - Nitratos

nitratos(janeiro, 0.24).

nitratos(fevereiro, 0.28).

$\neg$ nitratos (X,Y)  $\leftarrow$  não nitratos(X,Y),  
 não exceção<sub>nitratos</sub>(X,Y).

exceção<sub>nitratos</sub>(X,Y)  $\leftarrow$  nitratos(X, nitratos).

exceção<sub>nitratos</sub>(março, 0.07).

exceção<sub>nitratos</sub>(março, 0.08).

exceção<sub>nitratos</sub>(abril, 0.13).

exceção<sub>nitratos</sub>(abril, 0.15).

exceção<sub>nitratos</sub>(abril, 0.17).

exceção<sub>nitratos</sub>(abril, 0.18).

(iv) - Fosfatos

fosfatos(janeiro, 0.21).

fosfatos(abril, 0.12).

$\neg$ fosfatos(X,Y)  $\leftarrow$  não fosfatos (X,Y),  
 não exceção<sub>fosfatos</sub>(X,Y).

exceção<sub>fosfatos</sub>(X,Y)  $\leftarrow$  fosfatos(X, fosfatos).

exceção<sub>fosfatos</sub>(fevereiro, 0.04).

exceção<sub>fosfatos</sub>(fevereiro, 0.05).

exceção<sub>fosfatos</sub>(fevereiro, 0.07).

exceção<sub>fosfatos</sub>(março, 0.16).

exceção<sub>fosfatos</sub>(março, 0.18).

exceção<sub>fosfatos</sub>(março, 0.21).

(v) - Clorofila-a

clorofila(fevereiro, 0.24).

clorofila(março, 0.20).

clorofila(abril, clorofila).

$$\neg \text{clorofila}(X,Y) \leftarrow \text{não clorofila}(X,Y),$$

$$\text{não exceção}_{\text{clorofila}}(X,Y).$$

$$\text{exceção}_{\text{clorofila}}(X,Y) \leftarrow \text{clorofila}(X, \text{clorofila}).$$

$$\text{exceção}_{\text{clorofila}}(\text{janeiro}, 0.26).$$

$$\text{exceção}_{\text{clorofila}}(\text{janeiro}, 0.28).$$

$$\text{exceção}_{\text{clorofila}}(\text{janeiro}, 0.32).$$

Programa 5.3: Aplicação da PLE a alguns parâmetros de qualidade da água.

Por outro lado, tem-se que no espaço  $n$ -dimensional que está a ser objecto de consideração (e que é descrito pelo Programa 5.3), cada uma das suas dimensões não só reflete o nome de um predicado tomado do universo de discurso ou contexto em que se desenvolve ou se procura uma solução para um dado problema, mas também consubstancia uma medida da qualidade da informação em jogo em termos do operador de qualidade  $\mathcal{L}$  (i.e. tomando em consideração a extensão de um predicado ou as exceções a esta, é possível marcar cada eixo de uma estrutura  $n$ -dimensional com os valores tomados do intervalo  $[0, 1]$ , em que à intersecção destes se atribui o valor zero e às suas extremidades o valor um). Tem-se então que:

$$(i) - \text{Em termos do predicado pH}() - \mathcal{L}_{\text{janeiro}} = 1, \mathcal{L}_{\text{fevereiro}} = \lim_{N \rightarrow \infty} \frac{1}{N} = 0 \quad (N \gg 0),$$

$$\mathcal{L}_{\text{março}} = \frac{1}{2} = 0,5 \text{ e } \mathcal{L}_{\text{abril}} = \frac{1}{3} = 0,33(3);$$

$$(ii) - \text{Em termos do predicado oD}() - \mathcal{L}_{\text{janeiro}} = \lim_{N \rightarrow \infty} \frac{1}{N} = 0 \quad (N \gg 0), \mathcal{L}_{\text{fevereiro}} = 1,$$

$$\mathcal{L}_{\text{março}} = 1 \text{ e } \mathcal{L}_{\text{abril}} = \frac{1}{2} = 0,5;$$

$$(iii) - \text{Em termos do predicado nitratos}() - \mathcal{L}_{\text{janeiro}} = 1, \mathcal{L}_{\text{fevereiro}} = 1, \mathcal{L}_{\text{março}} = \frac{1}{2} =$$

$$= 0,5 \text{ e } \mathcal{L}_{\text{abril}} = \frac{1}{4} = 0,25;$$

(iv) - Em termos do predicado fosfatos() -  $\mathcal{L}_{\text{janeiro}} = 1$ ,  $\mathcal{L}_{\text{fevereiro}} = \frac{1}{3} = 0,33(3)$ ,

$$\mathcal{L}_{\text{março}} = \frac{1}{3} = 0,33(3) \text{ e } \mathcal{L}_{\text{abril}} = 1; \text{ e}$$

(v) - Em termos do predicado clorofila() -  $\mathcal{L}_{\text{janeiro}} = \frac{1}{3} = 0,33(3)$ ,  $\mathcal{L}_{\text{fevereiro}} = 1$ ,

$$\mathcal{L}_{\text{março}} = 1 \text{ e } \mathcal{L}_{\text{abril}} = \lim_{N \rightarrow \infty} \frac{1}{N} = 0 \text{ (} \mathbb{N} \gg 0 \text{)}.$$

em que  $N$  denota a cardinalidade do conjunto das extensões para um dado predicado e,  $\mathcal{L}_{\text{mês}}$  nos dá uma medida da qualidade dessa informação.

É de realçar que para o caso de  $N = 0$ , se tem que  $\mathcal{L} = \infty$  (infinito). Dado, por exemplo, o predicado *singularidade*, a sua extensão seria dada pela cláusula:

$\neg$ singularidade  $\leftarrow$   
 não singularidade,  
 não exceção<sub>singularidade</sub>.

i.e. está-se mediante uma singularidade (Figura 5.3). Nada se sabe sobre a extensão ou exceções à extensão de *singularidade*, pelo que todas as hipóteses sobre o desenvolvimento de um qualquer sistema estão em aberto ou são plausíveis.

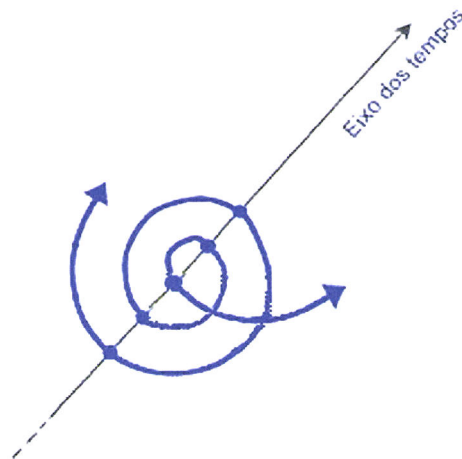


Figura 5.3: Singularidade. Adaptada de Abelha, [Abelha, 2004].

Face ao exposto, torna-se evidente que é possível caracterizar e quantificar a qualidade da informação que vai dentro de uma *data warehouse*, que inclui os níveis de verda-

de 1 (*verdadeiro*), 0 (*falso*) e os *valores de verdade* pertencentes ao intervalo ]0, 1[. De notar que por dimensão se passará a entender um atributo estrutural (aqui dado na forma da extensão à exceção de um predicado) de uma figura geométrica  $n$ -dimensional (no caso presente  $n = 5$ ), que corporiza uma sucessão de elementos do mesmo tipo, no que respeita à percepção que o utilizador passa a ter quanto à qualidade dos dados que tem vindo a manipular (Figura 5.4).

As Figuras 5.4, 5.10 e 5.11 dão, pelo seu lado, uma medida do potencial que esta representação transporta. Na Figura 5.4, a área a tracejado dá uma medida da qualidade da informação sobre a qualidade da água, quando considerada no seu todo. Nas Figuras 5.10 e 5.11 têm-se, respectivamente, documentadas as situações em que se deve procurar operar (o seu óptimo), e aquela que é a sua antítese.

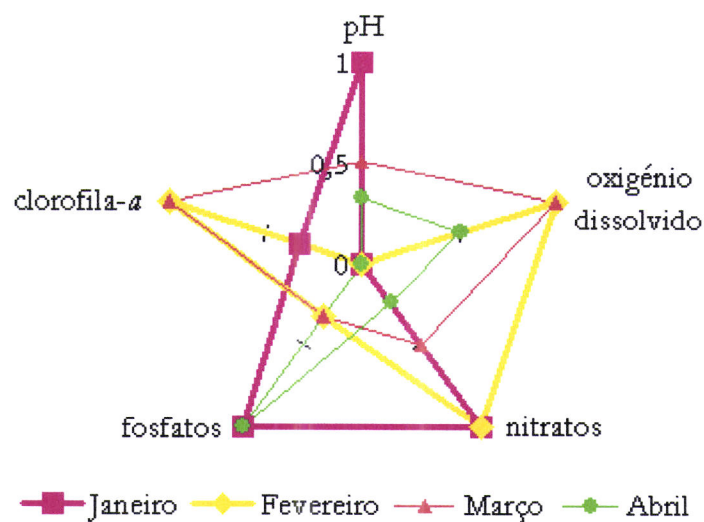


Figura 5.4: Um Espaço Hiperdimensional em que pontifica a qualidade da informação que é intrínseca ao Programa 5.3.

Esta figura (Figura 5.4) resulta de um processo de composição aplicado às Figuras 5.5, 5.6, 5.7 e 5.8.

Tabela 5.1: Programa e sua expressão em termos do operador qualidade para o mês de Janeiro.

Programa	Operador Qualidade
<p>pH(janeiro, 0.32).  <math>\neg \text{pH}(X,Y) \leftarrow \text{não pH}(X,Y),</math>                      não exceção<sub>pH</sub>(X,Y).                      exceção<sub>pH</sub>(X,Y) <math>\leftarrow \text{pH}(X, \text{pH}).</math></p>	$\mathcal{L}_{\text{pH}} = 1$
<p>oD(janeiro, oD).  <math>\neg \text{oD}(X,Y) \leftarrow \text{não oD}(X,Y),</math>                      não exceção<sub>oD</sub>(X,Y).                      exceção<sub>oD</sub>(X,Y) <math>\leftarrow \text{OD}(X, \text{oD}).</math></p>	$\mathcal{L}_{\text{oxigênio dissolvido}} = \lim_{N \rightarrow \infty} \frac{1}{N} = 0 \quad (N \gg 0)$
<p>nitratos(janeiro, 0.24).  <math>\neg \text{nitratos}(X,Y) \leftarrow \text{não nitratos}(X,Y),</math>                      não exceção<sub>nitratos</sub>(X,Y).                      exceção<sub>nitratos</sub>(X,Y) <math>\leftarrow \text{nitratos}(X, \text{nitratos}).</math></p>	$\mathcal{L}_{\text{nitratos}} = 1$
<p>fosfatos(janeiro, 0.21).  <math>\neg \text{fosfatos}(X,Y) \leftarrow \text{não fosfatos}(X,Y),</math>                      não exceção<sub>fosfatos</sub>(X,Y).                      exceção<sub>fosfatos</sub>(X,Y) <math>\leftarrow \text{fosfatos}(X, \text{fosfatos}).</math></p>	$\mathcal{L}_{\text{fosfatos}} = 1$
<p><math>\neg \text{clorofila}(X,Y) \leftarrow \text{não clorofila}(X,Y),</math>                      não exceção<sub>clorofila</sub>(X,Y).                      exceção<sub>clorofila</sub>(X,Y) <math>\leftarrow \text{clorofila}(X, \text{clorofila}).</math>                      exceção<sub>clorofila</sub>(janeiro, 0.26).                      exceção<sub>clorofila</sub>(janeiro, 0.28).                      exceção<sub>clorofila</sub>(janeiro, 0.32).</p>	$\mathcal{L}_{\text{clorofila}} = \frac{1}{3} = 0,33(3)$

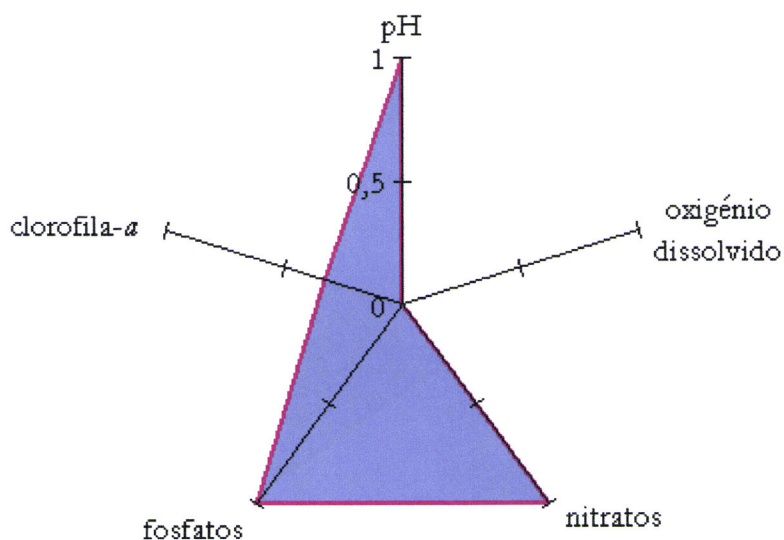


Figura 5.5: Um Espaço Hiperdimensional em que pontifica a qualidade da informação respeitante ao mês de Janeiro.

Tabela 5.2: Programa e sua expressão em termos do operador qualidade para o mês de Fevereiro.

Programa	Operador Qualidade
<p>pH(fevereiro, pH).  <math>\neg</math>pH(X,Y) <math>\leftarrow</math> não pH(X,Y),                      não exceção<sub>pH</sub>(X,Y).                      exceção<sub>pH</sub>(X,Y) <math>\leftarrow</math> pH (X, pH).                      oD(fevereiro, 0.46).  <math>\neg</math>oD(X,Y) <math>\leftarrow</math> não oD(X,Y),                      não exceção<sub>oD</sub>(X,Y).                      exceção<sub>oD</sub>(X,Y) <math>\leftarrow</math> OD(X, oD).                      nitratos(fevereiro, 0.28).  <math>\neg</math>nitratos (X,Y) <math>\leftarrow</math> não nitratos(X,Y),                      não exceção<sub>nitratos</sub>(X,Y).                      exceção<sub>nitratos</sub>(X,Y) <math>\leftarrow</math> nitratos(X, nitratos).  <math>\neg</math>fosfatos(X,Y) <math>\leftarrow</math> não fosfatos (X,Y),                      não exceção<sub>fosfatos</sub>(X,Y).                      exceção<sub>fosfatos</sub>(X,Y) <math>\leftarrow</math> fosfatos(X, fosfatos).                      exceção<sub>fosfatos</sub>(fevereiro, 0.04).                      exceção<sub>fosfatos</sub>(fevereiro, 0.05).                      exceção<sub>fosfatos</sub>(fevereiro, 0.07).                      clorofila(fevereiro, 0.24).  <math>\neg</math>clorofila (X,Y) <math>\leftarrow</math> não clorofila (X,Y),                      não exceção<sub>clorofila</sub>(X,Y).                      exceção<sub>clorofila</sub>(X,Y) <math>\leftarrow</math> clorofila(X, clorofila).</p>	<p><math>\mathcal{L}_{\text{pH}} = \lim_{N \rightarrow \infty} \frac{1}{N} = 0 \quad (N \gg 0)</math></p> <hr/> <p><math>\mathcal{L}_{\text{oxigênio dissolvido}} = 1</math></p> <hr/> <p><math>\mathcal{L}_{\text{nitratos}} = 1</math></p> <hr/> <p><math>\mathcal{L}_{\text{fosfatos}} = \frac{1}{3} = 0,33(3)</math></p> <hr/> <p><math>\mathcal{L}_{\text{clorofila}} = 1</math></p>

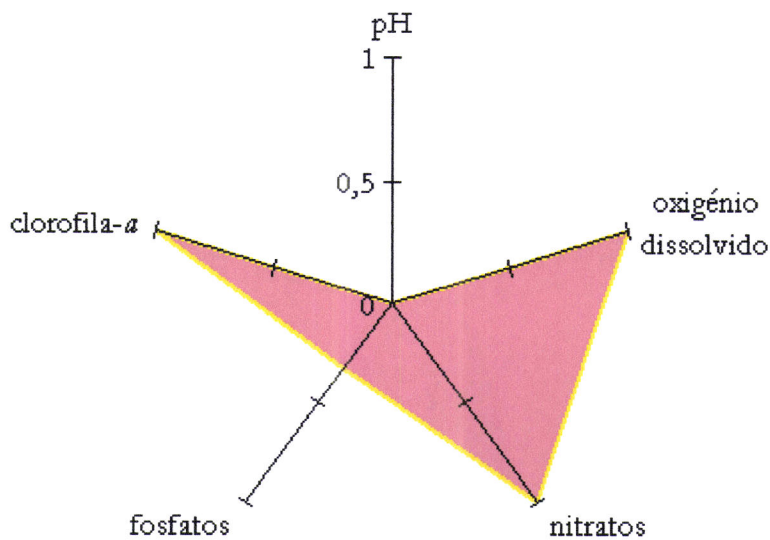


Figura 5.6: Um Espaço Hiperdimensional em que pontifica a qualidade da informação respeitante ao mês de Fevereiro.



Tabela 5.4: Programa e sua expressão em termos do operador qualidade para o mês de Abril.

Programa	Operador Qualidade
$\neg \text{pH}(X,Y) \leftarrow$ não pH(X,Y), não exceção <sub>pH</sub> (X,Y). $\text{exceção}_{\text{pH}}(X,Y) \leftarrow$ pH (X, pH). $\text{exceção}_{\text{pH}}(\text{abril}, 0.24)$ . $\text{exceção}_{\text{pH}}(\text{abril}, 0.26)$ . $\text{exceção}_{\text{pH}}(\text{abril}, 0.42)$ .	$\mathcal{L}_{\text{pH}} = \frac{1}{3} = 0,33(3)$
$\neg \text{oD}(X,Y) \leftarrow$ não oD(X,Y), não exceção <sub>oD</sub> (X,Y). $\text{exceção}_{\text{oD}}(X,Y) \leftarrow$ OD(X, oD). $\text{exceção}_{\text{oD}}(\text{abril}, 0.47)$ . $\text{exceção}_{\text{oD}}(\text{abril}, 0.48)$ .	$\mathcal{L}_{\text{oxigênio dissolvido}} = \frac{1}{2} = 0,5$
$\neg \text{nitratos}(X,Y) \leftarrow$ não nitratos(X,Y), não exceção <sub>nitratos</sub> (X,Y). $\text{exceção}_{\text{nitratos}}(X,Y) \leftarrow$ nitratos(X, nitratos). $\text{exceção}_{\text{nitratos}}(\text{abril}, 0.13)$ . $\text{exceção}_{\text{nitratos}}(\text{abril}, 0.15)$ . $\text{exceção}_{\text{nitratos}}(\text{abril}, 0.17)$ . $\text{exceção}_{\text{nitratos}}(\text{abril}, 0.18)$ . $\text{fosfatos}(\text{abril}, 0.12)$ .	$\mathcal{L}_{\text{nitratos}} = \frac{1}{4} = 0,25$
$\neg \text{fosfatos}(X,Y) \leftarrow$ não fosfatos (X,Y), não exceção <sub>fosfatos</sub> (X,Y). $\text{exceção}_{\text{fosfatos}}(X,Y) \leftarrow$ fosfatos(X, fosfatos). $\text{clorofila}(\text{abril}, \text{clorofila})$ .	$\mathcal{L}_{\text{fosfatos}} = 1$
$\neg \text{clorofila}(X,Y) \leftarrow$ não clorofila (X,Y), não exceção <sub>clorofila</sub> (X,Y). $\text{exceção}_{\text{clorofila}}(X,Y) \leftarrow$ clorofila(X, clorofila).	$\mathcal{L}_{\text{clorofila}} = \lim_{N \rightarrow \infty} \frac{1}{N} = 0 \quad (N \gg 0)$

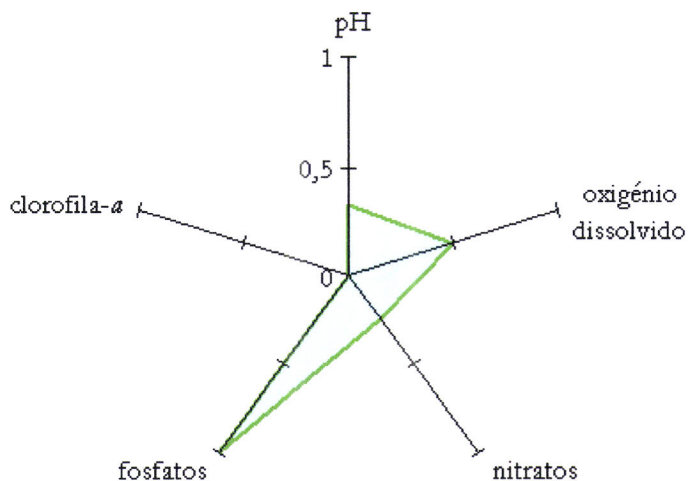


Figura 5.8: Um Espaço Hiperdimensional em que pontifica a qualidade da informação respeitante ao mês de Abril.



As áreas coloridas de Figura 5.9 dão-nos uma medida da evolução da qualidade da informação sobre a qualidade da água para o mês em consideração, com valores tomados do intervalo[0, 1].

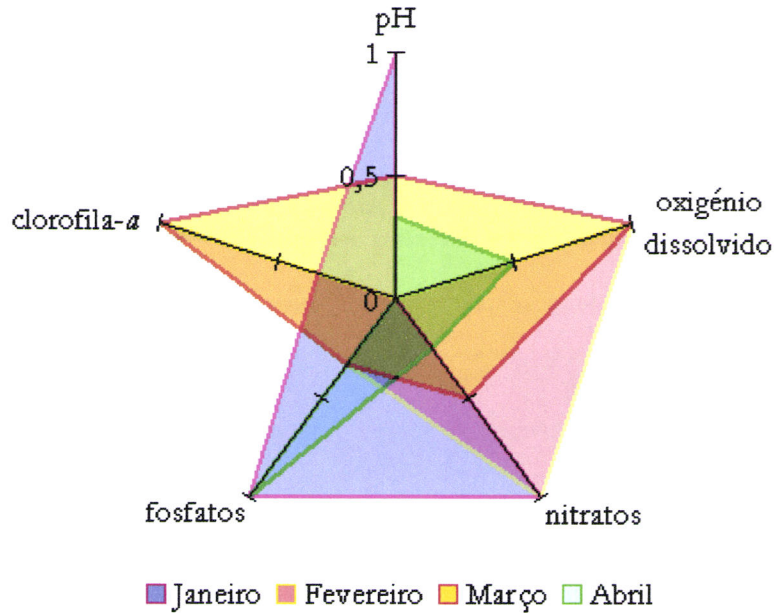


Figura 5.9: Um Espaço Hiperdimensional em que pontifica a qualidade da informação, resultante da sobreposição das Figuras 5.5, 5.6, 5.7 e 5.8.

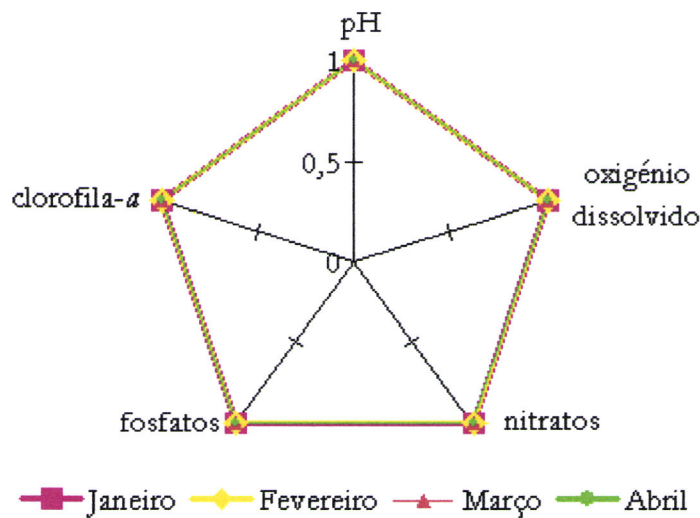


Figura 5.10: Um Espaço Hiperdimensional que denota uma situação ótima no que concerne aos termos em que se deve operar no que respeita à gestão e controlo da qualidade da água.

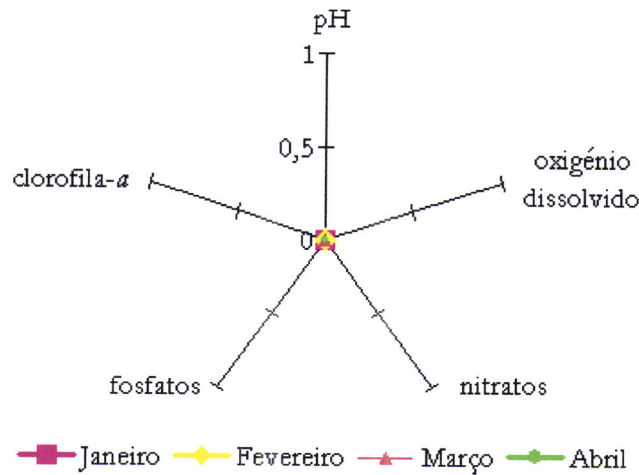


Figura 5.11: Um Espaço Hiperdimensional que denota uma situação de total descoordenação no que respeita à gestão e controlo da qualidade da água.

É assim possível, através do cálculo de uma simples área, quantificar não só a **qualidade** da informação sobre a qualidade da água na albufeira (tomada de valores de verdade pertencentes ao intervalo  $[0, 1]$ ), para um dado instante ou período de tempo, como também gerar possíveis cenários para a sua evolução, e isto sem escamotear a possibilidade de no presente serem tomadas medidas que venham a condicionar esse já mencionado futuro.

## 5.6 AQUA - Modelação Sub-Simbólica

Hoje em dia assiste-se a uma globalização da economia, e por conseguinte a um **acréscimo** de concorrência entre as organizações, cujas estruturas, processos e tecnologias se optimizados, lhe darão vantagens competitivas. No entanto, inovar num ambiente de incerteza pode ser desastroso, pelo que é natural pressupor que as organizações estão interessadas em obter previsões fundamentadas sobre o seu futuro.

Uma das técnicas em uso socorre-se da Previsão de Séries Temporais (PST), que se baseia em observações passadas de uma dada variável [Cortez et al., 1995] [Cortez et al., 1996] [Cortez et al., 1999][ Cortez, 2002]. Os métodos convencionais de previsão, desenvolvidos a partir de disciplinas como a Investigação Operacional ou a Estatística, providenciam previsões precisas, quando estão envolvidos dados lineares. Contudo, quando um elevado grau de não linearidade é introduzido, essas técnicas não são as mais adequadas, pelo

que uma alternativa surge com os modelos inspirados na Natureza, que passar-se-ão a descrever.

## 5.7 Análise de Séries Temporais

Uma série temporal é uma colecção de observações  $x_1, x_2, \dots, x_n$  registadas de um modo sequencial, no que respeita ao comportamento apresentado por um dado sistema. Exemplos de séries temporais surgem nos mais diversos domínios, desde a Economia, Demografia, Ecologia, Agricultura ou Engenharia, sendo a Análise de Séries Temporais uma área importante da Estatística [Chatfield, 1989].

Um modelo de uma série temporal ( $\hat{x}_t$ ) assume que as observações são dependentes, i.e. padrões passados irão recorrer no futuro, pelo que a série é previsível. Uma série diz-se determinística quando é previsível a 100%. Todavia, a maior parte das séries contém um elemento estocástico, sendo o futuro apenas parcialmente determinado por valores passados, pelo que as previsões exactas são impossíveis de se obter.

A análise de séries temporais contém vários objectivos, que não são necessariamente idênticos, nomeadamente os indicados a seguir [Weigend e Gershenfeld, 1994]:

- (i) - Descrição, onde se pretende descobrir, com nenhum ou pouco conhecimento *à priori*, no que respeita ao sistema em estudo, algumas das suas propriedades fundamentais, como o número de graus de liberdade ou quantidade de ruído;
- (ii) - Explicação, cujo o objectivo é encontrar um modelo que capture características sobre o comportamento de um sistema a longo prazo;
- (iii) - Previsão, que almeja prever com precisão a evolução de um sistema a longo prazo; e
- (iv) - Controlo, onde se tenta controlar o comportamento de um dado processo, atendendo a certos parâmetros (e.g. qualidade).

De entre estes, a ênfase neste trabalho será dada à previsão, uma importante tarefa em Ciências Ambientais. Os modelos de previsão tratam o sistema como se de uma caixa

negra se tratasse, não tendo como objectivo determinar os factores que afectam o seu comportamento. Existem, porventura, algumas razões para esta postura. A primeira deve-se ao facto do sistema não ser perfeitamente determinado, sendo difícil identificar as relações de causa-efeito em jogo. A segunda reside no facto de se desejar prever o que acontece e não como acontece, i.e. existem séries temporais reais impossíveis de axiomatizar, tangíveis à previsão (e.g. a previsão de ocorrência de manchas solares ou a previsão do comportamento do mercado de acções [McCluskey, 1993]).

Em geral, as técnicas tradicionais de análise de séries temporais baseiam-se no processo de decomposição, onde se identificam os factores que influenciam os valores de uma série temporal, sendo de referir entre estes a tendência e a sazonalidade [Makridakis e Wheelwright, 1989].

A tendência dá-nos uma medida do crescimento ou declínio dos valores de uma série. Uma série que possua este tipo de comportamento é chamada de não estacionária. Na Figura 5.12 representa-se uma série não estacionária típica. Várias forças se conjugam para que esta componente seja comum em situações que têm a ver com o estudo da inflação, das mudanças tecnológicas, do crescimento da população ou dos aumentos de produtividade.

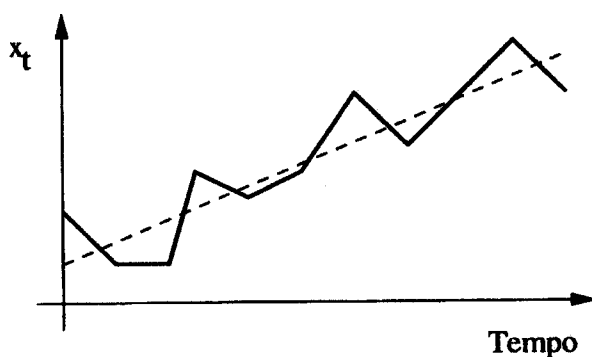


Figura 5.12: Série temporal não estacionária. Adaptada de Cortez [Cortez, 2002].

A sazonalidade corporiza uma flutuação periódica, com uma duração, por exemplo, semanal, mensal ou trimestral. O padrão repete-se, ao longo do tempo, após cada  $K$  períodos, sendo  $K$  o factor sazonal. Esta situação é comum, sendo de referir, por exemplo, a comercialização de produtos que dependem do clima ou que se vendem em determinados período de ano. A Figura 5.13 representa uma série sazonal com um período de sazonalidade de quatro, correspondendo às quatro estações do ano.

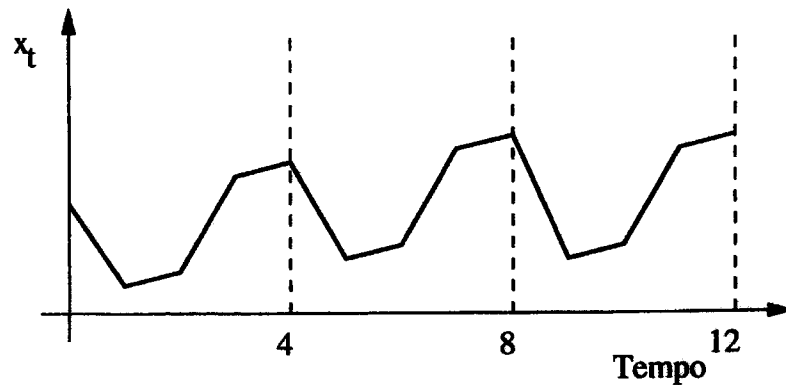


Figura 5.13: Série temporal com um período de sazonalidade de quatro. Adaptada de Cortez [Cortez, 2002].

Uma ferramenta estatística importante para a análise de séries temporais é o coeficiente de autocorrelação, definido como a correlação entre a série e ela própria (daí o termo auto), deslocada de  $k$  períodos de tempo [Box e Jenkins, 1976]:

$$r_k = \frac{\sum_{t=1}^s (x_t - \bar{x}_t)(x_{t+k} - \bar{x}_t)}{\sum_{t=1}^{s-k} (x_t - \bar{x}_t)^2} \quad (5.1)$$

onde  $s$  representa o tamanho da série e  $\bar{x}_t$  o valor médio da série temporal. As autocorrelações são úteis para testar se uma dada série é previsível e para a decomposição das principais componentes de uma série, como a tendência e a sazonalidade (Figura 5.14).

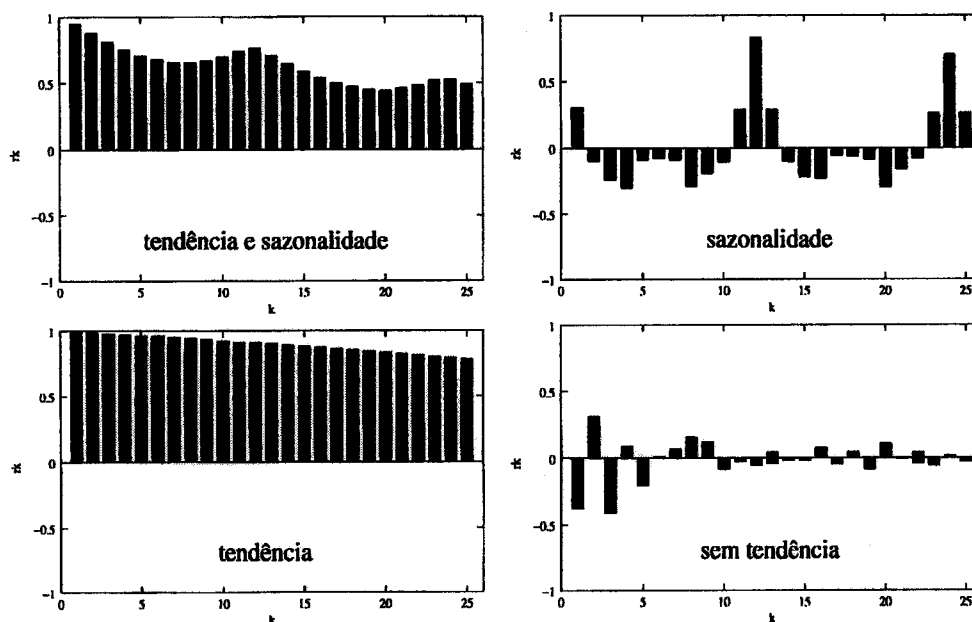


Figura 5.14: Autocorrelações típicas de séries temporais. Adaptada de Cortez [Cortez, 2002].

Para avaliar o desempenho global de um modelo de previsão, torna-se necessário utilizar uma medida do erro, sendo que uma das formas mais utilizadas recorre ao quadrado de erro. Em particular, tem-se a Soma do Quadrado dos Erros (SQE), a Média do Quadrado dos Erros (MQE), a Raiz da Média do Quadrado dos Erros (RMQE) e a Média Normalizada do Quadrado dos Erros (MNQE):

$$SQE = \sum_{i=1}^L e^2 \quad (5.2)$$

$$MQE = \frac{SQE}{L} \quad (5.3)$$

$$RMQE = \sqrt{MQE} \quad (5.4)$$

$$MNQE = \frac{SQE}{\sum_{i=1}^L (x_t - \bar{x}_t)^2} \quad (5.5)$$

em que  $L$  denota o número de previsões a efectuar e  $e$  denota o erro de previsão, dado na forma  $e = x - \hat{x}$  (i.e. a diferença entre o valor actual da série e o que foi previsto).

## 5.8 Redes Neurais Artificiais

O uso de técnicas inspiradas na Natureza para a Previsão de Séries Temporais (PST) começou nos fins dos anos oitenta, com o trabalho de Lapedes e Farber [Lapedes e Farber, 1987], RNUs foram utilizadas para prever séries caóticas determinísticas. De início, as tentativas de aplicação de RNAs ao mercado bolsista revelaram-se algo desapontadoras, situação que se foi progressivamente esbatendo. Por exemplo, já no início dos anos noventa, Shoneberg [Shoneberg, 1990] refere a inaplicabilidade dos métodos estatísticos convencionais às previsões de curto prazo dos valores em bolsa. Era, então, já óbvio que os métodos de alisamento eram incapazes de detectar as pequenas variações de preço dos produtos financeiros no dia a dia, justificando-se ainda o uso de novas ferramentas neste domínio.

A aplicação de RNAs para a previsão envolve todo um conjunto de etapas como a análise da série temporal, o pré-processamento dos dados, a escolha do modelo neuronal, o treino da rede e a validação do desempenho desta. Relativamente à escolha do modelo, existem diversos tipos de RNAs passíveis de serem aplicadas à PST, como sejam as

RBFs [Shi et al., 1999] ou as RNRs [Ulbricht, 1994], embora a maioria dos estudos opte pelas RNUs [Tang e Fishwick, 1993] [Cortez et al., 1995] [Faraday e Chatfield, 1998] [Cortez, 2002].

Estudos há que parecem indicar que as RNUs são as mais indicadas para a análise do comportamento dos mercados financeiros. Papadourakis [Papadourakis et al., 1993] utilizaram RNUs para a previsão de preços de produtos financeiros, com resultados encorajadores. Um estudo dos índices bolsistas do mercado Kuala Lumpur, revelou elementos de trabalho interessantes sem um uso excessivo de dados do mercado ou do conhecimento deste [Yao e Poh, 1995]. Outros casos de interesse são referenciados por Freisleben e Ripper [Freisleben e Ripper, 1995], em que os resultados obtidos se revelaram superiores aos obtidos por regressão linear.

Dentro deste contexto, a topologia base é definida por uma rede completamente interligada, com apenas uma camada intermédia, com conexões de *bias* e atalhos, dado que estes acrescentam uma componente linear ao modelo (Figura 5.15). A função logística é aplicada aos nodos intermédios, para permitir uma aprendizagem não linear. No nodo de saída, é utilizada a função linear para escalonar os dados de saída, visto que a função logística tem um contradomínio limitado ( $[0, 1]$ ). Esta estratégia tem a vantagem de evitar o uso de técnicas de filtragem, que podem levar a uma perda de informação (e.g. escalonamento).

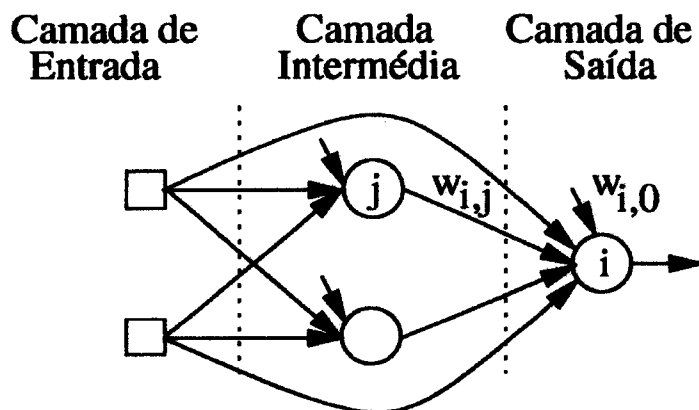


Figura 5.15: Uma RNU completamente interligada, com dois nodos de entrada, dois nodos intermédios, um nodo de saída, conexões de *bias* e atalhos. Adaptada de Cortez [Cortez, 2002].

Uma Janela Temporal Deslizante (JTD) define um conjunto de deslocamentos temporais passados ( $k_i$ ), definidos a partir do tempo  $t$  (Figura 5.16), sendo representada

pela sequência  $JTD = \langle k_1, k_2, \dots, k_n \rangle$ , para uma JTD de comprimento  $n$ . Por exemplo, a aplicação da  $JTD = \langle 1, 3 \rangle$  à série 1, 2, 3, 4, 5 produzirá as sequências  $\langle 1, 3 \rangle$  e  $\langle 2, 4 \rangle$ , para os instantes  $t = 4$  e  $t = 5$ .

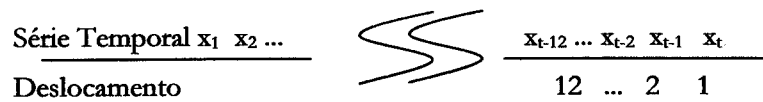


Figura 5.16: Deslocamentos de uma série temporal. Adaptada de Cortez [Cortez, 2002].

No caso da previsão de curto prazo, o nodo de saída da RNU é treinado para efectuar correspondências entre os valores passados, gerados a partir da JTD, com o valor presente. Assim, o modelo geral fornecido pela RNU é dado por:

$$\hat{x}_t = w_{S,0} + \sum_{i=1}^n x_{t-k_i} w_{S,i} + \sum_{j=n+1}^{S-1} fa \left( \sum_{i=1}^n x_{t-k_i} w_{j,i} + w_{j,0} \right) w_{S,j} \quad (5.6)$$

em que  $S$  denota o nodo de saída,  $fa$  a função logística  $\frac{1}{1+e^{-x}}$  e  $n$  o número de nodos de entrada ( $n = |E|$ ).

## 5.9 Generalização

Diz-se que um modelo de aprendizagem possui uma boa generalização quando a correspondência entre as entradas e as saídas é correcta (ou próximo disso) para dados de teste, retirados da mesma população, nunca antes utilizados na criação ou treino da rede. O processo de aprendizagem pode ser visto como um ajustamento de curvas ou de aproximação de funções, onde a rede tenta efectuar uma boa interpolação não linear dos dados [Rojas, 1996].

A Figura 5.17 mostra como podem ocorrer duas generalizações distintas para o mesmo conjunto de dados de treino. Aqui, uma boa generalização ocorre com a curva  $A$ , com um erro mínimo para os dados de teste. O mesmo já não sucede com a curva  $B$ , que origina um erro maior para os casos de teste, isto apesar de apresentar um menor erro para



os dados de treino. Tal fenómeno, designado de sobre-ajustamento<sup>1</sup>, ocorre quando um modelo memoriza em demasia os exemplos de treino, tratando-se de um dos problemas mais sérios relacionados com o uso da RNAs [Sarle, 1995].

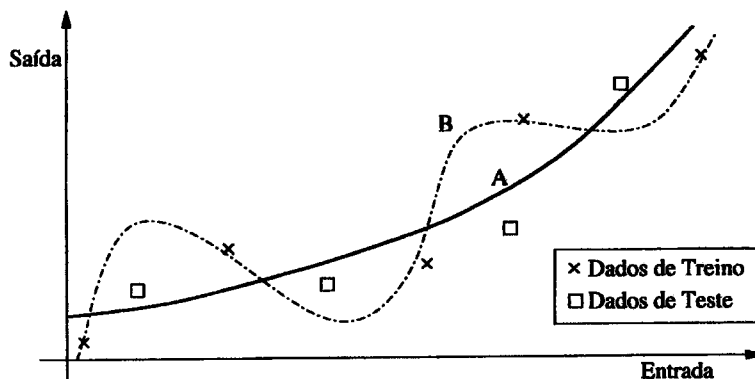


Figura 5.17: Generalização e sobre-ajustamento. Adaptada de Cortez [Cortez, 2002].

Durante o processo de aprendizagem, o modelo pode captar certas características, como o ruído, que estão presentes nos dados de treino mas não na função implícita a ser aprendida. Este exemplo ilustra os dois objectivos contraditórios da aproximação funcional. Por um lado tem-se a minimização do erro de treino, pelo outro há a minimização do erro para as entradas desconhecidas.

A generalização é influenciada por três factores [Haykin, 1999] [Sarle, 1999]:

- (i) - A complexidade do problema a ser aprendido - trata-se de um factor de difícil controlo. As entradas devem conter informação suficiente para permitir a obtenção das saídas desejadas, i.e. tem de existir uma função matemática com algum grau de precisão que relacione as entradas com as saídas. Por outro lado, convém que esta função seja suave, i.e. pequenas alterações nas entradas devem provocar pequenas alterações nas saídas, para a maior parte dos casos;
- (ii) - Os casos de treino - a sua cardinalidade deve ser elevada, com exemplos (ou amostras na terminologia estatística) que sejam representativos acerca do ambiente. A generalização é sempre efectuada a partir de dois tipos de situações: interpolação e extrapolação. No primeiro caso, um va-

<sup>1</sup> Conhecido em inglês pelo termo *overfitting*.

lor é determinado a partir da informação dos valores vizinhos. A segunda situação engloba tudo o resto, ou seja, casos fora do domínio dos dados de treino. Enquanto que a interpolação pode ser efectuada com relativa acuidade, o mesmo já não se passa com a extrapolação, notoriamente menos fiável; e

- (iii) - Complexidade do modelo de aprendizagem - medida pelo número de parâmetros livres ( $p$ ) que contém e pela sua capacidade de aprendizagem. Um modelo que não seja suficientemente representativo do problema em equação irá falhar na aproximação da função a aprender. Por outro lado, um modelo demasiado ajustado aos dados do problema em equação irá fixar o ruído, provocando sobre-ajustamento.

A melhor forma de evitar o fenómeno de sobre-ajustamento é utilizar uma quantidade elevada de casos de treino. Quando este número for trinta vezes superior ao número de parâmetros, então é muito improvável que ocorra uma perda de generalização [Sarle, 1999]. O problema é que nem sempre existem suficientes casos de treino disponíveis e não se deve reduzir o número de parâmetros de um modo arbitrário, devida a problemas de falta de representatividade. Assim, dado um número fixo de casos de treino existem duas alternativas para controlar a complexidade ou a representatividade de um modelo (i.e. a regularização e a selecção de modelos) [Sarle, 1995].

A regularização baseia-se num controlo dos valores dos parâmetros de um modelo para obter uma boa generalização, envolvendo o uso de restrições ou penalidades sobre estes, de modo a que a rede aprenda funções mais suaves. Como exemplo tem-se a paragem antecipada, um dos mais populares métodos de regularização, onde os dados de treino são divididos em dois tipos de casos: de treino e de validação. Os primeiros são utilizados na aprendizagem do modelo, enquanto os últimos são utilizados para aferir da qualidade deste, i.e. para estimar o erro da generalização. Durante a fase de aprendizagem, calcula-se o erro de validação de forma periódica, parando-se quando este ameaça aumentar.

A regularização diminui o efeito de perda de generalização pelo estímulo na aprendizagem de funções suaves. No entanto, utiliza uma estrutura fixa, que deve ser especificada em avanço pelo utilizador. Embora se possa utilizar uma grande estrutura, com um grande número de nodos intermédios, na prática, a optimização dos parâmetros torna-se de difícil ajuste, exigindo um elevado esforço computacional. Mas ainda, em geral, são méto-

dos que exigem um delicado balanço, controlado por um (ou mais) parâmetro(s) de regularização. Uma alternativa distinta baseia-se na procura de um modelo adequado, em termos de representatividade (i.e. complexidade). Assim, um problema que seja de difícil aprendizagem para um modelo poderá ser facilmente aprendido por outro.

A abordagem estatística para a selecção de modelos passa pela estimativa do erro de generalização para cada um dos modelos, ou topologias da rede, escolhendo-se o modelo que minimiza essa estimativa. Existem diversos métodos para estimar a capacidade de generalização de uma RNA, como seja a Validação Cruzada<sup>2</sup> ou o *Bootstrapping* [Efron e Tibshirani, 1993] [Kohavi, 1995], que são deveras exigentes em termos computacionais. Uma alternativa razoável é dada pela teoria da informação, onde o erro de generalização é estimado a partir de estatísticas simples, que acrescentam uma penalidade que depende da complexidade do modelo, nomeadamente:

- (i) - Critério de Informação de Akaike, conhecido por AIC<sup>3</sup>, segundo o clausulado [Akaike, 1973]:

$$AIC = N \ln(SQE/N) + 2p \quad (5.7)$$

- (ii) - Critério de Informação de Bayes, designado de BIC<sup>4</sup> ou SBC, e dado na forma [Schwarz, 1978]:

$$BIC = N \ln(SQE/N) + p \ln(N) + p \quad (5.8)$$

onde  $SQE$  denota o somatório do quadrado dos erros para todos os exemplos de treino e  $N$  e  $p$  denotam o número de exemplos de treino e o número de parâmetros livres, respectivamente.

---

<sup>2</sup> Técnica conhecida em inglês por K-fold Validation.

<sup>3</sup> Do inglês Akaike Information Criterion.

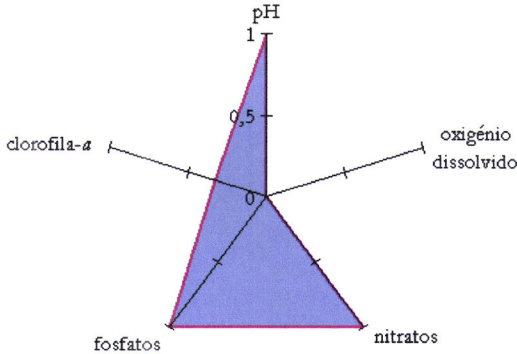
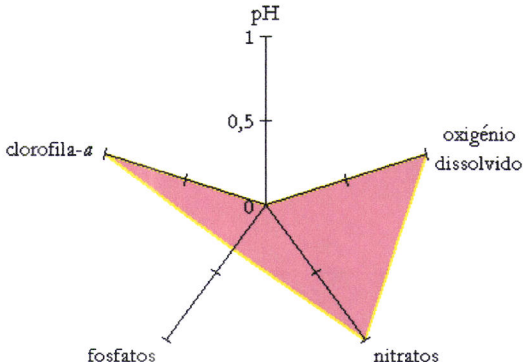
<sup>4</sup> Acrónimo tirado do inglês Bayesian Information Criterion

## 5.10 Previsão da Qualidade da Água em Albufeiras Utilizando RNAs

Todos os métodos de previsão partem do princípio de que as experiências do passado serão úteis no futuro, i.e. assume-se, assim, e agora reportando-nos ao caso em estudo, que as condições que condicionaram o estado da qualidade da água nas albufeiras no passado, se repetirão no futuro.

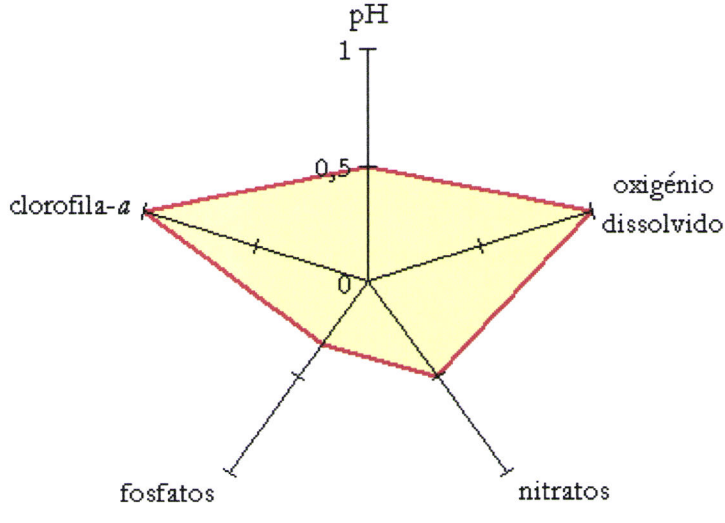
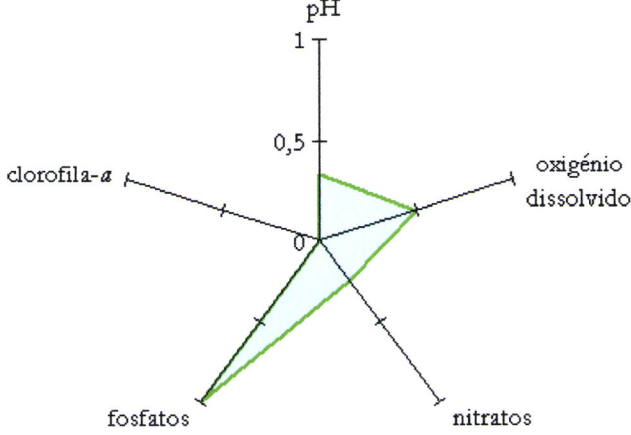
É então possível, entrando em linha de conta com os valores associados ao operador qualidade  $\mathcal{L}$  (para cada uma das leituras a que nos reportamos já neste texto (i.e. os dados sobre a qualidade da água para os meses de Janeiro, Fevereiro, Março e Abril, referidos anteriormente)), e à medida da qualidade da informação sobre a qualidade da água que lhe pode ser associada por uma leitura no hiperespaço, criar um conjunto de casos de treino/teste e apresentar uma RNA a ser utilizada em previsão, ou seja:

Tabela 5.5: Conjunto de casos de treino/teste para RNA.

Operador Qualidade $\mathcal{L}$ (entrada para a RNA)	Medida da qualidade da informação sobre a qualidade da água dada pela área a tracejado das figuras que se seguem, fruto de uma representação no hiperespaço do Operador Qualidade $\mathcal{L}$ (saída da RNA)
<p>Janeiro</p> <p><math>\mathcal{L}_{\text{pH}} = 1</math></p> <p><math>\mathcal{L}_{\text{oxigénio dissolvido}} = 0</math></p> <p><math>\mathcal{L}_{\text{nitratos}} = 1</math></p> <p><math>\mathcal{L}_{\text{fosfatos}} = 1</math></p> <p><math>\mathcal{L}_{\text{clorofila}} = 0,33(3)</math></p>	
<p>Fevereiro</p> <p><math>\mathcal{L}_{\text{pH}} = 0</math></p> <p><math>\mathcal{L}_{\text{oxigénio dissolvido}} = 1</math></p> <p><math>\mathcal{L}_{\text{nitratos}} = 1</math></p> <p><math>\mathcal{L}_{\text{fosfatos}} = 0,33(3)</math></p> <p><math>\mathcal{L}_{\text{clorofila}} = 1</math></p>	

Cont.

Tabela 5.5 (cont.): Conjunto de casos de treino/teste para RNA.

<b>Operador Qualidade <math>\mathcal{L}</math></b> <b>(entrada para a RNA)</b>	<b>Medida da informação sobre a qualidade da água dada pela área a tracejado das figuras que se seguem, fruto de uma representação no hiperespaço do Operador Qualidade <math>\mathcal{L}</math> (saída da RNA)</b>
<p>Março</p> <p><math>\mathcal{L}_{\text{pH}} = 0,5</math></p> <p><math>\mathcal{L}_{\text{oxigénio dissolvido}} = 1</math></p> <p><math>\mathcal{L}_{\text{nitratos}} = 0,5</math></p> <p><math>\mathcal{L}_{\text{fosfatos}} = 0,33(3)</math></p> <p><math>\mathcal{L}_{\text{clorofila}} = 1</math></p>	 <p>A radar chart with five axes: pH (top), oxigénio dissolvido (right), nitratos (bottom-right), fosfatos (bottom-left), and clorofila-a (left). The axes are scaled from 0 to 1. The data points are: pH=0.5, oxigénio dissolvido=1, nitratos=0.5, fosfatos=0.33, and clorofila-a=1. The area between these points is shaded yellow.</p>
<p>Abril</p> <p><math>\mathcal{L}_{\text{pH}} = 0,33(3)</math></p> <p><math>\mathcal{L}_{\text{oxigénio dissolvido}} = 0,5</math></p> <p><math>\mathcal{L}_{\text{nitratos}} = 0,25</math></p> <p><math>\mathcal{L}_{\text{fosfatos}} = 1</math></p> <p><math>\mathcal{L}_{\text{clorofila}} = 0</math></p>	 <p>A radar chart with five axes: pH (top), oxigénio dissolvido (right), nitratos (bottom-right), fosfatos (bottom-left), and clorofila-a (left). The axes are scaled from 0 to 1. The data points are: pH=0.33, oxigénio dissolvido=0.5, nitratos=0.25, fosfatos=1, and clorofila-a=0. The area between these points is shaded light blue.</p>

Atendendo, porém, a que a cardinalidade do conjunto dos casos de treino/teste obtidos é de 4 (quatro), não se pode elaborar mais sobre este modelo, ficando a constituir mais um tema para trabalho futuro. É contudo possível, apresentar já uma topologia para uma possível RNA a utilizar na previsão, e que tomaria a forma:

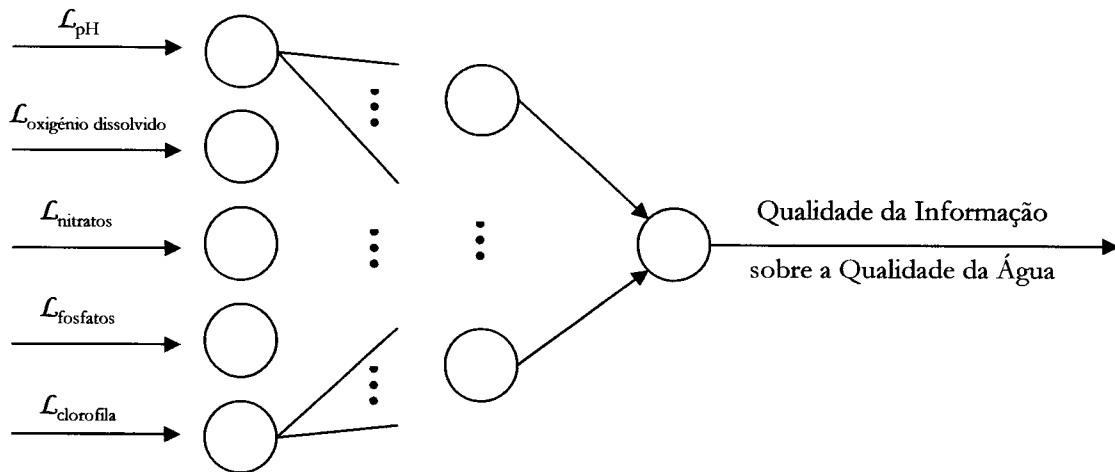


Figura 5.18: Topologia para uma possível RNA a utilizar na previsão da qualidade da água.

## 5.11 Árvores de Decisão e Indução de Regras

A definição mais simples que se pode dar de Árvore de Decisão é a de uma forma de representação de um conjunto de regras que seguem uma hierarquia de classes ou valores. Expressam uma lógica simples condicional e graficamente são semelhantes a uma árvore (Figura 5.19).

Uma Árvore de Decisão é uma representação de um conjunto de regras, que classificam as instâncias, desde o nó raiz até aos nós terminais (folhas). Cada nó da árvore, especifica um teste para os atributos da instância (variáveis), e cada ramo descendente desse nó corresponde a um dos valores possíveis para esse atributo. Uma instância é classificada começando por testar o atributo especificado pelo nó raiz, e depois seguindo o ramo correspondente ao valor do atributo.

Os algoritmos de indução de Árvores de Decisão permitem construir árvores a partir dos dados de treino, de uma forma recursiva subdividindo o referido conjunto até à obtenção de nós puros, que representam apenas uma única classe, ou a satisfação de um critério [Rich et al., 1991]. As árvores geradas têm a seguinte estrutura [Michalski et al., 1998]:

- (i) - Folhas ou nós puros - corresponde às classes/objectos;
- (ii) - Nós internos - corresponde aos atributos; especifica algum teste efectuado num único atributo, com duas ou mais sub-árvores que representam saídas possíveis; e

(iii) - Ramos - corresponde aos valores dos atributos.

Esta técnica funciona através da criação e treino de subconjuntos de informação para os quais é inferida uma ou mais regras. O primeiro nó da Árvore de Decisão mostra o factor mais correlacionado com os objectos de saída. Os restantes factores são subsequentemente classificados como nós e relacionados com os nós anteriores, possibilitando uma visualização fácil e rápida dos factores que mais se relacionam com os objectos de saída. As ferramentas baseadas em Árvores de Decisão permitem explorar a estrutura da árvore, encontrar grupos, associar dados, bem como seleccionar dados fundamentais em qualquer nó da árvore [Kufrin, 1997] [Quinlan,1998] [Han e Kamber, 2001].

Após a fase de construção das árvores, encontra-se uma estrutura especializada que está ajustada aos dados. A poda passa a desempenhar um papel importante, produzindo árvores menores com melhor potencial e precisão quando novos casos são considerados [Quinlan, 1998] [Han e Kamber, 2001].

O objectivo principal da poda é remover partes da árvore que não contribuem para a precisão da classificação, produzindo estruturas menos complexas e de fácil compreensão. Existem dois tipos de poda [Cios et al., 1998]:

- (i) - Durante a aprendizagem; e
- (ii) - Após a aprendizagem.

A poda durante a aprendizagem é feita em simultâneo com a construção da árvore, i.e. são testados os atributos enquanto esta se constrói. A vantagem consiste em evitar a construção de sub-árvores que depois são podadas. As desvantagens prendem-se com o facto de se terem que tomar decisões em tempo real, decisões estas que podem ser algo complexas, implicando maiores custos do algoritmo de indução [Han e Kamber, 2001].

A maioria dos algoritmos utiliza a técnica de poda após a aprendizagem (Algoritmo 5.1) [Quinlan, 1998]. A vantagem reside no facto de muitas vezes só ser possível determinar se uma sub-árvore deve ser podada depois da construção da árvore, i.e. depois de uma análise do contexto. Mesmo após a poda uma árvore pode ainda representar uma estrutura complexa e de difícil compreensão. Neste caso, a indução de regras é utilizada para extrair regras menos complexas e com uma maior precisão [Kufrin, 1997] [Cios et al., 1998].

**Algoritmo 5.1**

*Seja T uma árvore de decisão completa não podada;*  
*calcular o erro estático para cada folha da árvore;*  
*para todo o nó  $\neq$  folha no sentido das folhas para a raiz*  
*calcular o erro estático e o erro propagado*  
*se: erro estático < erro propagado*  
*então podar a árvore passando esse nó a ser uma folha*

o erro estático é o erro de classificação de um nó, dado por  $E_{Est}(nó) = \frac{N - n + k - 1}{N + k}$  onde

$N$  denota o número exemplos classificados no nó,  $n$  denota o número de exemplos pertencentes à classe maioritária (com mais exemplos) e  $k$  o número de classes total do problema. O erro propagado é o erro de classificação de um nó devido à propagação de erros das sub-árvores, dado por  $E_{Prop}(nó) = \sum_i P_i E_{Est}(nó_i)$  onde  $i$ ,  $P_i$  e  $nó_i$  denotam, respectivamente, o ramo do nó, a frequência relativa dos exemplos no ramo  $i$  e o sub-nó do nó.

Na fase da poda surge o problema de como prever as taxas de erro [Han e Kamber, 2001]. De acordo com Quinlan [Quinlan, 1998], existem dois tipos de técnicas para prognóstico de taxas de erro:

- (i) - Uma das técnicas prevê uma taxa de erro com base num conjunto separado de dados (conjunto de teste), que é distinto do conjunto de treino. Na construção da árvore os dados de teste não são utilizados, permitindo obter estimativas da taxa de erro confiáveis caso existam dados suficientes [Quinlan, 1997]. A desvantagem apresentada por esta técnica está associada à necessidade de um grande volume de dados, por forma que se possa separar um conjunto para construção da árvore e um outro para a poda. Quando os dados são escassos, pode-se utilizar a técnica da validação cruzada. Nesta técnica, os dados de treino são divididos em  $n$  conjuntos de igual tamanho, de preferência com a mesma distribuição de classes. A árvore é construída utilizando casos de todos os blocos e testada com um bloco reservado para esse efeito [Quinlan, 1998]; e
- (ii) - A outra técnica de prognóstico de taxas de erro utiliza apenas os dados de treino a partir do qual a árvore foi construída [Quinlan, 1998].



A principal desvantagem das Árvores de Decisão prende-se com o facto de necessitarem de uma considerável quantidade de dados para descobrir as estruturas complexas. Por outro lado, as suas vantagens baseiam-se no facto de levarem em consideração as regras mais relevantes para a compreensão do sistema. Para além da vantagem exposta anteriormente, as Árvores de Decisão são, em geral, obtidas mais rapidamente e produzem resultados com precisão similar a outros métodos de classificação [Quinlan, 1998].

As Árvores de Decisão constituem-se como um dos métodos mais simples e rápidos de tratar problemas com muitas dimensões. Uma outra vantagem prende-se com a expressividade associada ao modo como é feita a representação do conhecimento.

A Indução de Regras refere-se à detecção, em bases de dados, de tendências e padrões, apresentados sob forma de regras [Quinlan, 1998] [Berson et al., 2000].

O objectivo a atingir com a Indução de Regras passa por encontrar dependências entre os atributos e seus valores, através da análise de um conjunto de probabilidades do foro condicional. Em geral, os resultados são apresentados na forma de regras  $X \rightarrow Y$ , que significa que se  $X$  está presente, então  $Y$  também tem probabilidade de estar presente. O elemento  $X$  pode ser uma combinação de atributos e valores, formando assim regras mais complexas. As regras têm dois graus associados, a Confiança e o Suporte. O Suporte é o número de casos onde a regra é encontrada, i.e. o número de vezes em que  $X$  e  $Y$  aparecem associados. A Confiança é a probabilidade condicional da regra, ou seja, a probabilidade do segundo elemento estar presente. É calculada dividindo do número de casos onde  $X$  e  $Y$  aparecem juntos (i.e. o Suporte Absoluto), pelo número de casos onde somente  $X$  aparece [Quinlan, 1998].

A Indução de Regras corresponde ao processo de descoberta de padrões numa série de dados, i.e. criar hipóteses que conduzem a padrões. Esta técnica consiste na descoberta de regras de previsão, do tipo **Se ... Então**, onde o **Se**, a condição da regra, especifica alguns atributos e o **Então** da regra fornece um valor para um determinado atributo cujo conhecimento é desejável [Fayyad et al., 1996].

As regras devem, preferencialmente, satisfazer três propriedades [Kuftrin, 1997] [Quinlan, 1998]:

- (i) - Previsões correctas - se o **Se** da regra é verdadeiro, o **Então** da regra também é verdadeiro;

- (ii) - Compreensíveis - as regras representam conhecimento a um alto nível de abstracção. Devem ser de fácil compreensão para o utilizador; e
- (iii) - Úteis para a tomada de decisão - está relacionado com o facto da regra expressar conhecimento que é ou não esperado por parte do utilizador.

As vantagens da técnica de Indução de Regras está no modo directo de lidar com os dados, o desempenho, as facilidades de explicação e compreensão das regras. Em contrapartida, sendo as regras altamente heurísticas, a sua falta de profundidade, a dificuldade de manuseamento da informação incompleta ou de valores fora do contexto, ou a explicação baseada na prova e não nos fundamentos teóricos constituem as suas principais desvantagens [Langley e Simon, 1995].

Normalmente, a Indução de Regras e as Árvores de Decisão surgem associadas, utilizando-se a Indução de Regras para representar o conhecimento representado nas Árvores de Decisão.

No presente estudo obteve-se um conjunto de regras (Apêndice A) e gerou-se uma árvore de decisão (Figura 5.19) para a base de dados contendo os parâmetros de qualidade da água para as albufeiras estudadas (i.e. Divôr, Monte Novo e Vigia). A divisão da água em classes de qualidade foi feita com base no critério de classificação das águas superficiais, de acordo com as suas características de qualidade para usos múltiplos (Tabela 2.11).

Na base de dados inicial havia dados omissos, tendo sido necessário, na fase de pré-processamento dos dados, estabelecer estratégias para resolver este problema. Os dados em falta resultam da divisão dos parâmetros em três grupos distintos, com frequências de amostragem e de análise diferentes, nos termos expostos no capítulo 2. Deste modo, a base de dados apresentava registos mensais para os parâmetros do grupo G1, registos bimensais para os do grupo G2 e apenas três registos anuais para os parâmetros do grupo G3, tendo-se atribuído às lacunas o valor conhecido da leitura imediatamente anterior.

Conseguiram-se obter regras cujo suporte é elevado (ver apêndice A). A acuidade do modelo, medida em termos da percentagem de acertos relativamente ao número de casos apresentados é de 96,4% (134 acertos/139 casos) para o conjunto de treino e de 76,8% (53 acertos/69 casos) para o conjunto de teste. Na Tabela 5.6 apresenta-se a matriz de confusão obtida para o modelo, designado por modelo 1. Um aspecto muito importante a realçar é o facto da acuidade do modelo 1 ser 100% para a classe B, i.e. quando o modelo prevê que a água é fracamente poluída a probabilidade de ela o ser é muito elevada. Nas

restantes classes, onde a acuidade não é tão significativa a tendência é para dar uma medida da qualidade da água inferior à real, o que é aceitável, atendendo a que as albufeiras se destinam à produção de água para abastecimento de populações.

Tabela 5.6: Matriz de confusão do modelo 1.

Classe	Conjunto de treino			Conjunto de teste		
	D	C	B	D	C	B
E	0	0	0	0	1	0
D	66	3	0	25	6	0
C	2	58	0	8	25	1
B	0	0	10	0	0	3

A análise da árvore de decisão (Figura 5.19) mostra que o oxigénio dissolvido é o parâmetro que mais influencia a classificação da qualidade da água, secundado pelos parâmetros que reflectem a quantidade de matéria orgânica presente (i.e. a carência química e bioquímica de oxigénio). Para os níveis mais baixos de oxigénio dissolvido (i.e.  $\leq 61,8\%$ ), verifica-se que o parâmetro representativo do oxigénio consumido pelos microorganismos na oxidação aeróbia da matéria orgânica (a carência bioquímica de oxigénio), se torna preponderante para a atribuição da classe à massa de água. Para os teores mais elevados de oxigénio (i.e.  $> 61,8$ ) a classificação da massa de água depende da quantidade de oxigénio necessária para oxidar quimicamente a matéria orgânica (a carência química de oxigénio).

O modelo 1 faz a classificação da massa de água com base no critério de análise das águas superficiais, de acordo com as suas características de qualidade para usos múltiplos, utilizando cerca de uma dezena de parâmetros. No entanto, a maioria dos casos, é classificada com base em seis parâmetros (oxigénio dissolvido, CBO<sub>5</sub>, CQO, fósforo total, azoto Kjeldahl e transparência).

A determinação dos parâmetros referidos anteriormente é demorada e impede que se afira no imediato a qualidade da água da albufeira. Para tentar contornar este senão procurou-se desenvolver um novo modelo, denominado modelo 2, que privilegiasse parâmetros para os quais se pudesse fazer aquisição automática de dados, com recurso a sensores instalados na albufeira. No apêndice B, apresenta-se o conjunto de regras obtido para o modelo 2 onde, tal como no modelo 1, o oxigénio dissolvido se assume como o parâmetro que mais influencia o processo.

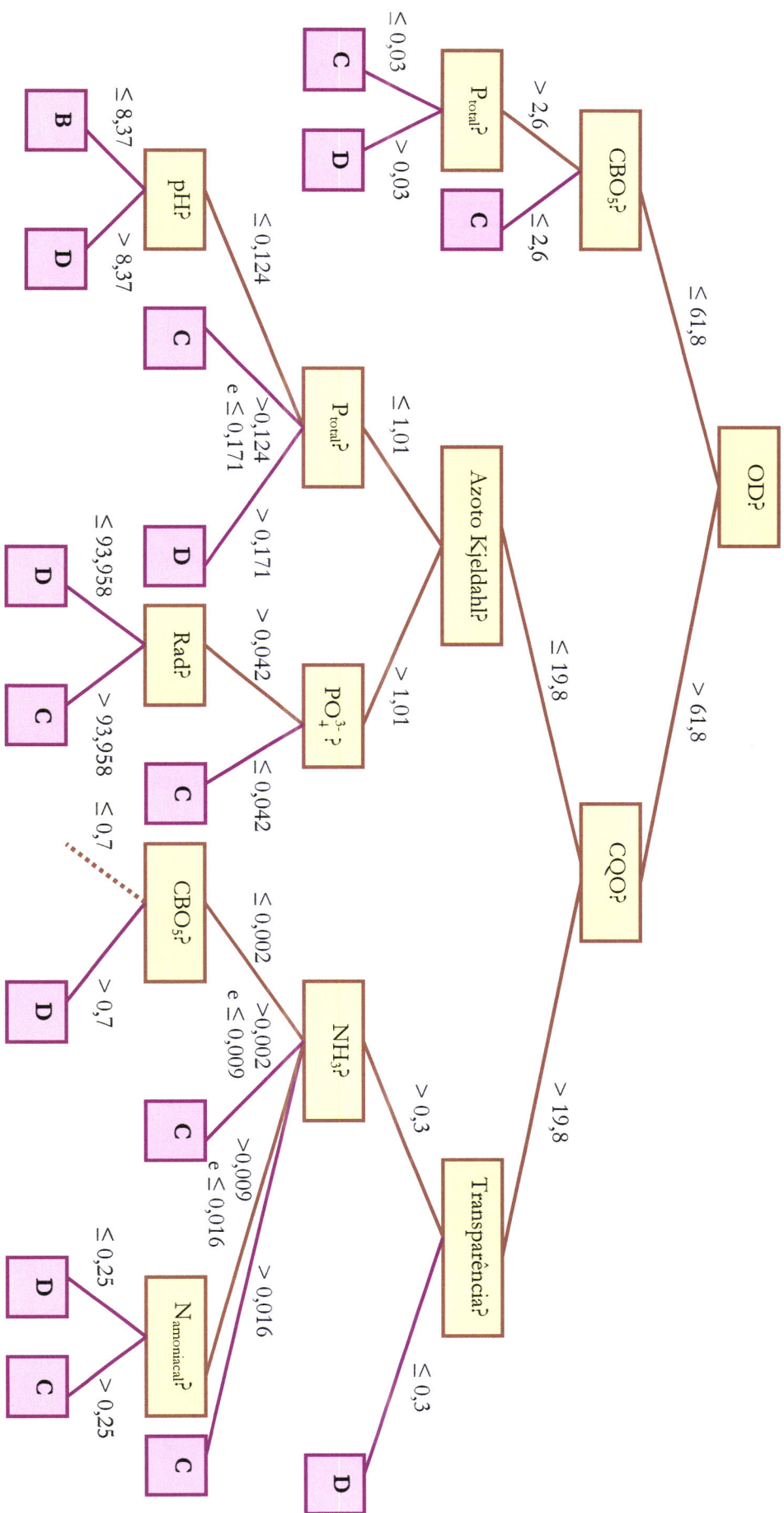


Figura 5.19: Árvore de decisão gerada a partir dos dados de qualidade da água para as albufeiras do Divôr, Monte Novo e Vigia.

As regras obtidas tem suporte elevado embora haja algumas para as quais a confiança é relativamente baixa. A acuidade do modelo é 87,0% (121 acertos/139 casos) para o conjunto de treino e 73,9% (51 acertos/69 casos) para o conjunto de teste. Na Tabela 5.7 apresenta-se a matriz de confusão obtida para o modelo 2, verificando-se que esta apresenta uma medida da qualidade da água inferior à real, para a classe B (o que não acontece no modelo 1, onde a acuidade foi de 100%). Nas restantes classes mantém-se o mesmo tipo de tendências, (i.e. o modelo prevê uma qualidade da água inferior à que na verdade se verifica).

Tabela 5.7: Matriz de confusão obtida para o modelo 2.

Classe	Conjunto de treino			Conjunto de teste		
	D	C	B	D	C	B
E	0	0	0	1	0	0
D	68	1	0	30	1	0
C	11	47	2	10	21	3
B	1	3	6	2	1	0

Na Tabela 5.8 faz-se uma avaliação comparativa dos dois modelos com base em alguns critérios que permitem mostrar que, apesar do modelo 2 apresentar uma acuidade para a qualidade da água ligeiramente inferior, tem a vantagem de permitir a aferição do estado da qualidade da água em tempo real.

Tabela 5.8: Avaliação dos dois modelos obtidos.

Critério de avaliação	Modelo 1	Modelo 2
Obtenção dos resultados analíticos	Demorada	Rápida
Métodos analíticos testados e aceites como válidos	Sim	Sim
Custo das análises	Dispendioso	Razoável
Possibilidade de medição na própria albufeira	Não	Sim
Possibilidade de medição remota	Não	Sim
Possibilidade de medição em contínuo	Não	Sim
Parâmetros mais influentes	OD, CBO <sub>5</sub> , CQO, P <sub>tot</sub> , N <sub>Kj</sub> , Transparência	OD, V <sub>vento</sub> , T <sub>água</sub> , Transparência, pH
Confiança na previsão	77 %	74 %
Aferição do estado da qualidade da água em tempo real	Não	Sim

## 5.12 A Arquitectura do Sistema AQUA

A arquitectura do sistema AQUA é dada na Figura 5.1, com base nas suas entidades constituintes, ou seja:

### **O Quadro Negro**

O Quadro Negro (QN) é uma zona de memória partilhada pelos agentes do sistema, que a utilizam como plataforma de comunicação e para partilha de estruturas de dados e informações de controlo e coordenação, de uma forma assíncrona e estruturada. Está dividida em áreas específicas de trabalho, com conjuntos de agentes especializados para cada uma das referidas áreas. Denota o ambiente em que operam os agentes do sistema.

### **O Agente Monitor**

O Agente Monitor (agm) permite a observação e o registo do funcionamento do sistema. Faz a interface com o administrador e pode, em algumas circunstâncias, tomar parte activa no controlo do sistema.

### **Os Agentes *Proxy***

Os Agentes *Proxy* (agps) fazem a ponte entre os utilizadores e o sistema nos termos e tipos de questões que podem ser formuladas, explicações pedidas, decisões que têm de ser tomadas e/ou visualização de resultados finais. Os interfaces do sistema funcionam todos em ambiente *Web* utilizando páginas *Hipermédia*, podendo ser acedidas a partir de um qualquer *Web browser*.

### **Os Agentes de Decisão**

Os Agentes de Decisão (agds) apresentam capacidades de mediação, actuam aceitando tarefas dos agps. Estas tarefas podem ser decompostas, sendo estas sub-tarefas enviadas para serem processadas nos agcs, integrando no final os resultados (devolvidos pelos agcs).

### **Os Agentes de Computação**

Os Agentes de Computação (agcs) aceitam pedidos para execução de tarefas específicas dos agds, devolvendo os resultados.

## Os Agentes de Recursos

Os Agentes de Recursos (agrs) possuem todo o conhecimento necessário para aceder a um determinado recurso e/ou dados.

## Os Agentes de Interação e Explicação

Os Agentes de Interação e Explicação (agies) baseiam a sua actuação em procedimentos argumentativos, trabalhando com dados e conhecimento fornecidos pelos agps e/ou agds. Tendo em atenção que o plano de execução recebido dos agds pode ser parcial, e apenas quando uma tarefa é completamente executada, a informação compilada e a sua explicação podem ser encaminhadas para os agps e/ou agds.

Para descrever o sistema AQUA seguiu-se a via sugerida por Cavedon e Tilhar com recurso quer à Programação em Lógica Contextual (PLC) [Cavedon e Tilhar, 1995], quer à Programação em Lógica Estendida (PLE) [Neves, 1984] [Traylor e Gelfond, 1993], e a considerações de natureza geométrica ou social, intrínsecas ao universo do discurso [Denti et al., 1995].

**Definição 5.1** - (O Sistema AQUA). Formalmente, o sistema AQUA dá forma a um multi-sistema  $\Xi = \langle C_{AQUA}, \Delta_{AQUA}, agm, agp_1, \dots, agp_i, agd_1, \dots, agd_j, agc_1, \dots, agc_k, agie_1, \dots, \dots, agie_l, agr, qenv \rangle$ , com  $i, j, k, l \geq 1$ , em que (Figura 5.1):

- (i) -  $C_{AQUA}$  denota o contexto em que o sistema se insere e opera, definido em termos de uma teoria ou programa em lógica dada pelo triplete  $\langle Lg, Ax, \Delta \rangle$ , onde  $Lg$  é a linguagem de programação em lógica, neste caso uma extensão à programação em lógica,  $Ax$  o conjunto de axiomas de  $Lg$  e  $\Delta$  o conjunto de regras de inferência;
- (ii) -  $\Delta_{AQUA}$  denota o conjunto das regras-ponte, que definem o processo de interação entre os componentes do sistema. Define a geometria do sistema, dando uma medida do significado e da interação possível entre as suas partes constituintes. Estabelece o modo como os eventos que ocorrem ao nível dos subsistemas ou agentes são coordenados, e expressa o grau de paralelismo presente no sistema;
- (iii) -  $agm$ , denota o agente monitor;

- (iv) -  $agp_o$ , denota um agente *proxy* ( $i = 1, \dots, o$ ), em que  $o$  é um parâmetro do sistema;
- (v) -  $agd_p$ , denota um agente de decisão ( $j = 1, \dots, p$ ), em que  $p$  é um parâmetro do sistema;
- (vi) -  $agc_k$ , denota um agente de conhecimento ( $k = 1, \dots, q$ ), em que  $q$  é um parâmetro do sistema;
- (vii) -  $agie_l$ , denota um agente de interacção e explicação ( $l = 1, \dots, r$ ), em que  $r$  é um parâmetro do sistema;
- (viii) -  $agr$ , denota o agente de recursos; e
- (ix) -  $qenv$ , denota o contexto em que se está a operar.

### 5.13 Os Agentes do Sistema AQUA

O sistema **AQUA** é uma agência, i.e. um sistema computacional no qual um conjunto de entidades semi-autónomas e heterogéneas interagem, através de uma insuspeita e terceira entidade, o QN, e que num processo de consolidação, a um nível meta, criam um Sistema Computacional (SC). Como se pode observar, o QN denota um espaço de memória partilhado, um meio de comunicação entre os demais componentes do sistema. Todas as interacções entre os agentes do sistema efectuam-se através do QN, colocando ou retirando informação com significado no contexto  $C_{AQUA}$ .

Os agentes de decisão, agentes computacionais, agentes *proxy* e o agente monitor, que compõem o sistema **AQUA**, são similares ao agente reactivo com estado interno apresentado em [Russel e Norvig, 1995] [Weiss, 1999], e possuem uma arquitectura do tipo não deliberativo [Coelho, 1995]. O comportamento destes agentes pode ser equacionado através de um processo de transição de estados. Deste ponto de vista assume-se que o estado do ambiente ou contexto em que um agente ou SMA se movimenta pode ser caracterizado como um conjunto (possivelmente infinito)  $EAA = eaa_1, eaa_2, \dots$  de estados ecológicos. Num determinado instante, o ambiente encontra-se num destes estados. A capacidade do agente actuar no ambiente pode, por outro lado, ser representada pelo conjunto  $Ev = ev_1,$



$ev_2, \dots$  de eventos (acções). Considere-se  $Per$  como o conjunto dos estímulos que o agente  $p$  pode receber do ambiente, e  $Int$  o conjunto de todos os seus estados internos. Num determinado instante o agente é caracterizado pelo estado interno em que se encontra. Com base nestes postulados é possível então descrever de forma abstracta o comportamento do agente através do conjunto de funções que são apresentadas a seguir:

- (i) - Observa :  $EAA \rightarrow Per$ , que faz o mapeamento de um estado ecológico numa percepção do agente ou SMA;
- (ii) - Pensa :  $Int \times Per \rightarrow Int$ , que possibilita o mapeamento de um estado interno do agente ou SMA e de um estímulo recebido do ambiente num novo estado interno do agente ou SMA ( $\times$  denota o produto cartesiano); e
- (iii) - Actua :  $mt \rightarrow Ev$ , que se encarrega de mapear um estado interno do agente ou SMA numa acção a executar sobre o ambiente.

#### **Algoritmo - (Ciclo de Vida dos Agentes e/ou SMA)**

O comportamento de um agente ou SMA (Figura 5.20) pode ser dado pelo procedimento [Santos, 1999]:

#### **Algoritmo 5.2 - *proc* Agente ou SMA $\equiv$**

*seja*  $Ev = ev_1, ev_2, \dots$  o conjunto de acções que representam a aptidão do agente ou SMA  
*actuar sobre o ambiente*  
*seja*  $Per$  o conjunto (não vazio) dos estímulos que o agente ou SMA pode receber do ambiente  
*seja*  $int_0$  o estado de conhecimento inicial do agente ou SMA  
 $Estado \leftarrow int_0$   
*inicia*(Estado)  
*Repetir*  
 $Estímulo \leftarrow observa\ Per$   
 $Estado \leftarrow pensa\ (Estado, Estímulo)$   
 $Atitude \leftarrow actua\ (Estado, Ev)$   
*até*  $Atitude = fim-de-sessão$   
*fim*

O *modus vivendi* de um agente ou SMA é descrito por uma sequência de transições de estado, na forma de *observa*, *pensa* e *actua*. O agente começa por um determinado estado inicial  $int_0$ . Então *observa* que o seu ambiente se encontra num estado  $aaa$ , e gera um estímulo  $observa(aaa)$ . O estado interno do agente ou SMA é actualizado e passa a igualar o resultado da função  $pensa(int_0, observa(aaa))$ . Esta acção é então executada sobre o ambiente e o agente ou SMA entra noutra ciclo, perscrutando o mundo com a função *observa*, actualizando o seu estado interno com a função *pensa* e seleccionando e perpetrando uma acção a executar com a função *actua*. O critério de paragem do ciclo de vida do agente ou SMA passa por se desligar do sistema (i.e. o processo computacional que está associado ao agente ou SMA é cancelado ao nível do sistema operativo da(s) máquina(s) em que o agente ou SMA opera).

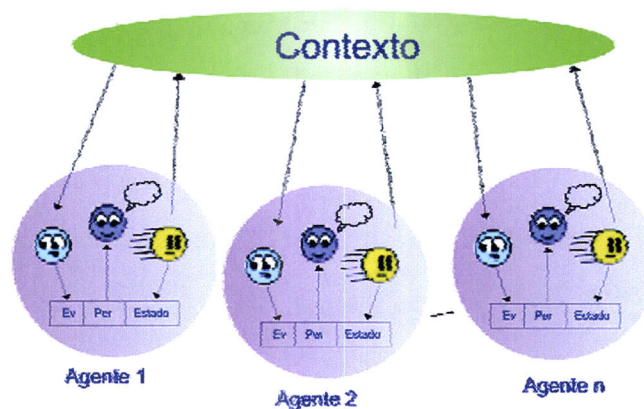


Figura 5.20: O comportamento dos agentes e/ou SMA na agência AQUA. Adaptada de Abelha [Abelha, 2004].

A coordenação a estabelecer entre os diferentes agentes que corporizam o sistema é obtida através da definição de um conjunto de relacionamentos entre estes, os quais se corporizam através de um conjunto de regras dado na forma:

- (i) - Quando o agente monitor é activado, os outros agentes do sistema devem ser igualmente activados, ou seja  $\frac{C_{agm} : ocorre(activar)}{ocorre(activar)}$ ;
- (ii) - Quando o agente monitor é desactivado, os restantes agentes do sistema devem ser também desactivados, ou seja  $\frac{C_{agm} : ocorre(desactivar)}{ocorre(desactivar)}$ ; e

- (iii) - Quando é efectuado, por exemplo, uma nova análise ambiental num dado laboratório, o agente *proxy* correspondente procede à sua classificação e ao envio do ficheiro para uma área específica do QN.

## 5.14 A Estrutura Lógica que Suporta a Agência AQUA

É agora altura de apresentar o conjunto de funcionalidades que corporizam a estrutura lógica da agência AQUA. Para isso vai ser necessário caracterizar formalmente os agentes do sistema (i.e. agente monitor, agentes *proxy*, agentes de decisão, agentes de computação, agentes de recurso, agentes de interacção e exploração) e os do meio envolvente.

### 5.14.1 O Agente Monitor

O agente monitor (*agm*) é o agente responsável pela monitorização do sistema. Recebe mensagens dos agentes *proxy* e através de uma interface amigável com o utilizador apresenta a informação sobre as actividades que se vão desenvolvendo no sistema. Os eventos do Agente Monitor estão descritos na Tabela 5.9. A operação de recepção de uma mensagem *M* pode então ser descrita por uma regra ponte ou de reflexão, dada na forma:

$$\frac{\text{recolhe - mensagem } (M) \wedge \text{grava - mensagem } (M)}{\text{ocorre (recepção)}}$$

A operação de apresentação de uma mensagem *M* numa janela *J* pode ser descrita de forma idêntica, sendo dada na forma:

$$\frac{\text{processa - mensagem } (M) \wedge \text{apresenta - mensagem } (M, J)}{\text{ocorre (apresentação)}}$$

Tabela 5.9: Descrição dos eventos do agente monitor.

Evento	Descrição
Iniciação	Inicia o funcionamento do agente.
Recepção	Recebe uma mensagem.
Apresentação	Envia uma mensagem para a janela e ficheiro respectivo.
Finalização	Termina.

### 5.14.2 Agentes Proxy

Os acontecimentos associados aos agentes *proxy* (agps) estão descritos na Tabela 5.10. Os agps são os agentes encarregados da gestão do ambiente da agência AQUA, implementando entre outras as seguintes funcionalidades:

- (i) - Problema - recebe as mensagens para processamento;
- (ii) - Soluções - propõe ao sistema os agentes de decisão que entende serem os mais indicados para resolver o problema; e
- (iii) - Reforço - recebe do ambiente a informação que mede o impacto que a solução do problema teve sobre esse mesmo ambiente e reporta essa mensagem aos agentes de interacção e explicação.

Tabela 5.10: Descrição dos eventos dos agentes *Proxy*.

Evento	Descrição
Iniciação	Activa o agente e reinicializa o relógio (tempo = 0).
Deteção	Verifica as mensagens enviadas pelo ambiente.
Colocação	Coloca as mensagens enviadas pelo ambiente na lista de mensagens a processar.
Resolução	Selecciona uma mensagem.
Repercussão	Escolhe o destinatário da mensagem.
Ciclo	Incrementa o tempo.
Finalização	Desliga o sistema.

### 5.14.3 Os Agentes de Decisão

Os eventos dos agentes decisão (agds) estão descritos na Tabela 5.11. Os agds tem como função principal receber uma tarefa, proceder à sua validação e decompô-la em sub-tarefas a distribuir pelos agentes de computação (agcs). A distribuição tem como principal objectivo não só garantir um certo equilíbrio na repartição de tarefas pelos agcs, assim como tirar partido das suas qualificações específicas.

Tabela 5.11: Descrição dos eventos dos agentes de decisão.

Evento	Descrição
Iniciação	Inicia o funcionamento do agente.
Recepção	Recebe uma mensagem.
Verificação	Verifica a consistência da mensagem.
Apresentação	Envia uma mensagem para a janela e ficheiro respectivo.
Decomposição	A tarefa associada à mensagem é decomposta em sub-tarefas.
Distribuição	Efectua a distribuição aos agentes de conhecimento.
Arquivo	Envia mensagem ao agente de recursos para arquivamento da informação.
Integração	Envia mensagem para integração com entidades externas.
Finalização	Termina.

### 5.14.4 Os Agentes de Computação

Os eventos associados aos agentes de computação (agcs) estão descritos na Tabela 5.12, tendo entre outras as tarefas:

- (i) - Facturação - inclui o pré-processamento da informação necessária para a sua integração e manipulação pela aplicação administrativa;
- (ii) - Arquivo - inclui a preparação da informação para armazenamento no *data warehouse*;

- (iii) - Relatórios Técnicos - inclui a preparação da informação necessária à compilação do relatório técnico e conseqüente geração de documentos, por exemplo em formato PDF e/ou XML; e
- (iv) - Integração de Imagem - inclui a integração de gráficos, mapas e outras imagens relevantes com o respectivo relatório e dados específicos do sistema (e.g. ano da construção da barragem, características hidráulicas, modelo digital do terreno, usos do solo e fontes de poluição na bacia hidrográfica).

Tabela 5.12: Descrição dos eventos dos agentes de computação.

Evento	Descrição
Iniciação	Inicia o funcionamento do agente.
Recepção	Recebe uma mensagem.
Verificação	Verifica se está apto a processar a mensagem.
Apresentação	Envia uma mensagem para a janela e ficheiro respectivo.
Execução	A sub-tarefa associada à mensagem é executadas.
Solução	Envia mensagem de sub-tarefa terminada e espera recompensa.
Finalização	Termina.

O número de agcs activos no sistema num dado instante varia com a carga de trabalho gerada pelos agds. A tarefa de chamar mais assistentes em caso de quebra no desempenho ou de desactivá-los em situações em que o agente está à algum tempo inactivo, é uma das responsabilidades do agm.

### 5.14.5 Os Agentes de Recursos

Os agentes de recursos (agrs) encarregam-se de monitorizar e gerir os recursos do sistema, além de proceder às cópias de segurança. Os eventos associados ao agrs estão descritos na Tabela 5.13. Quando surge uma mensagem de um agp a solicitar um local para armazenamento de informação, seja esta dada em forma de um tuplo, um ficheiro em for-

mato PDF ou um ficheiro em formato XML, o agr localiza esse local recorrendo à sua base de conhecimento, e envia uma mensagem ao agente servidor com essa informação. Este procedimento pode ser descrito pela regra:

$$\frac{\text{local - armazenamento}(R, T) \wedge \text{determina - local}(T, L) \wedge \text{local}(R, L)}{\text{local - de - armazenamento}(T)}$$

Tabela 5.13: Descrição dos eventos do agente de recursos.

Evento	Descrição
Iniciação	Inicia o funcionamento do agente.
Recepção	Recebe uma mensagem.
Verificação	Determina o melhor local para armazenar a informação.
Solução	Envia mensagem com informação sobre o local de armazenamento.
Gestão	Periodicamente activa a regra que analisa os recursos disponíveis e graus de utilização.
Segurança	Activa a regra de implementação dos procedimentos de segurança.
Finalização	Termina.

Tabela 5.14: Construção da operação para a definição do local de armazenamento.

Termo	Significado
<i>local-armazenamento</i> (R,T)	Recepção do pedido de armazenamento R e tamanho T.
<i>Determina-local</i> (T,L)	Determina o local onde se poderá efectuar o armazenamento.
<i>Local</i> (R, L)	Envio das coordenados do local ao Agente <i>Proxy</i> .
<i>local-armazenamento</i> (T)	Local de armazenamento de tamanho T.

### 5.14.6 Os Agentes de Interação e Explicação

Os agentes de interação e explicação (agies) encarregam-se das ligações a entidades externas bem como todos os pedidos de consulta, uma vez que detêm o conhecimen

to de como deve ser apresentada essa informação. Os eventos associados aos agies estão descritos na Tabela 5.15.

Tabela 5.15: Descrição dos eventos dos agentes de interacção e explicação.

Evento	Descrição
Iniciação	Inicia o funcionamento do agente.
Recepção	Recebe uma mensagem.
Verificação	Verifica se está apto a executar a tarefa.
Recolha	Recolhe a informação necessária à solução.
Solução	Envia a mensagem de sub-tarefa terminada e espera recompensa.
Finalização	Termina.

### 5.14.7 O Ambiente qenv

A interacção da agência AQUA com o seu meio ambiente (i.e. os equipamentos, os diferentes sistemas de informação da unidade ambiental, os utilizadores) é liderada pelo agm, com a ajuda dos agps. Estes agentes implementam três tipos de procedimentos, típicos de um sistema de aprendizagem reforçada, e que são referidos no que se segue:

- (i) - Receber do ambiente do sistema as mensagens para serem processadas (e.g. ficheiros em formato XML e/ou PDF);
- (ii) - Enviar para o ambiente do sistema as acções a executar, as quais são subscritas pelos agds (i.e. uma solução); e
- (iii) - Receber desse mesmo ambiente informação que meça o impacto que a solução proposta teve sobre esse mesmo ambiente (i.e. o reforço).

O sistema e a sua envolvente deverão fazer uso de uma linguagem passível de ser interpretada de um modo correcto por ambos os interlocutores. O ambiente pode, no caso presente, tomar uma das seguintes formas:

- (i) - Um agente (utilizador ou agente de *software*);
- (ii) - Um periférico (e.g. um computador, um autómato); e



(iii) - Uma base de dados e/ou de conhecimento.

Cada uma destas situações condiciona o comportamento do sistema [Russel e Norvig, 1995].

## 5.15 Conclusões

As arquitecturas de agentes assumem um papel fundamental nos sistemas computacionais actuais pela explosão das redes e pela necessidade de modelos adaptados a infra-estruturas complexas, heterogéneas, dinâmicas e não determinísticas. A definição de metodologias para a resolução de problemas, consensuais e adequadas aos mecanismos de coordenação, negociação e socialização, é uma necessidade absoluta, sem esquecer os critérios de qualidade, eficiência e competitividade que devem subscrever. Os sistemas simbólicos a que nos referimos no texto e aqui desenvolvidos apresentam-se como os mais adequados para a resolução deste tipo de problemas, na medida em que elaboram a partir de toda e qualquer informação que emana da albufeira. Os sistemas sub-simbólicos apresentam-se como complementares dos sistemas simbólicos, permitindo a partir destes elaborar sobre sistemas de previsão e/ou simulação do estado futuro da qualidade das águas em albufeiras.

Outro aspecto a não esquecer e que cada vez mais se afirma como fundamental, tem a ver com a questão da privacidade e da segurança da informação, valências estas que contudo extravazam os objectivos definidos de início para este trabalho.

A agência AQUA já ultrapassou a fase de testes beta em termos de produto final, estando a ser utilizada na determinação da qualidade da água da barragem de Alqueva, sendo os resultados obtidos bastante encorajadores.



## Capítulo 6

# Conclusões Finais e Perspectivas Para o Futuro

Após a conclusão do trabalho desenvolvido, chega o momento de olhar para trás e analisar o percurso feito e perspectivar o trabalho futuro.

Os modelos convencionais desenvolvidos para prever a qualidade da água de albufeiras são demasiado inflexíveis, assentam em conjuntos de parâmetros rígidos, representam as propriedades das variáveis de estado tal como eram no período de observação, sendo incapazes de prever a adaptabilidade, as variações de dinâmica e as alterações do ecossistema, o que os torna limitados no espaço e no tempo.

Os modelos convencionais demoram a ser obtidos, passam por etapas muito longas (e.g. verificação, análise de sensibilidade, calibração e validação) e são, na maior parte dos casos, específicos para a albufeira onde foram desenvolvidos.

Além das desvantagens já referidas há ainda a salientar o facto de os parâmetros a medir necessitarem da deslocação de equipas de amostragem à albufeira, preservação das amostras, transporte para os vários laboratórios e recurso a técnicas analíticas dispendiosas (quer ao nível de reagentes, quer ao nível do equipamento necessário, quer ao nível de pessoal qualificado, quer ao nível do tempo consumido até à obtenção do resultado), não permitindo a aferição da qualidade da água em tempo real.

O facto de estarem envolvidas diversas entidades nos processos referidos, promove a dispersão e dificulta o acesso rápido a esta informação por parte dos técnicos e gestores ambientais por forma a avaliarem a qualidade da informação disponível, e com ela fazer previsões e tomar, em tempo útil, as melhores decisões.

O presente trabalho pretende dar uma contribuição para a resolução de problemas nestas duas vertentes, recorrendo a técnicas e metodologias da Área Científica da Inteligência Artificial, do Cálculo Simbólico e da Descoberta de Conhecimento em Bases de Dados. Por um lado, desenvolveram-se modelos paralelos com recurso à Modelação Simbólica e à Modelação Sub-Simbólica que permitem aferir a qualidade da água em tempo real e potenciam a previsão. Os sistemas sub-simbólicos apresentam-se como complementares dos sistemas simbólicos, permitindo a partir destes elaborar sobre sistemas de previsão e/ou simulação do estado futuro da qualidade da água em albufeiras.

Apresentou-se ainda a agência AQUA, uma agência para a integração, a difusão e o arquivo de informação sobre a qualidade da água de albufeiras, que permite resolver o problema da dispersão de dados e facultar aos técnicos e aos gestores ambientais o apoio necessário na tomada das melhores decisões em tempo útil.

O desafio no futuro consiste na implementação de sistemas de aquisição automática de dados que possam alimentar os modelos desenvolvidos. Essa implementação passa pelo desenvolvimento de novas arquitecturas quer ao nível do *hardware*, quer do *software*, que permitam o controlo dos sensores instalados na albufeira, o registo dos dados obtidos e a cedência dessa informação de modo a que se tenha no momento a aferição da qualidade da água.

Um outro desafio que se coloca, prende-se com o facto da generalização desta tecnologia a outras albufeiras, nomeadamente à recente albufeira de Alqueva, onde a integração, a difusão e o arquivo de informação desde o início da sua entrada em funcionamento se pode revelar fundamental em todos os aspectos relacionados com a optimização de uma gestão sustentada, que por um lado responda às expectativas das populações e, por outro, preserve e valorize o ecossistema e toda a área envolvente.

# Bibliografia

- [Abelha, 2004] Abelha, A., *Sistemas Multiagente como Suporte a Trabalho Cooperativo em Unidades Hospitalares*, Tese de Doutoramento, Departamento de Informática, Universidade do Minho, Portugal, 2004.
- [Abelha et al., 2003] Abelha, A., Machado, J., Santos, M., Allegro, S., Rua, F., Paiva, M. e Neves, J., Agency for Integration, Diffusion and Archive of medical information, em *Proceedings of the Second LASTED International Conference on Artificial Intelligence and Applications*, Malaga, Espanha, 2003.
- [Abelha et al., 2004] Abelha A., Machado J., Alves V. e Neves, J., Data Warehousing through Multi-Agent systems in the Medical area, em *Proceedings of the first International Conference on Knowledge Engineering and Decision Support*, Porto, Portugal, 2004.
- [Alchourrón et al., 1985] Alchourrón, C.E., Gärdenfors, P., e Makinson, D., On the logic of theory change: partial meet functions for contractions and revision. *Journal of Symbolic Logic*, 50: 510-530, 1985.
- [Adriaans e Zantinge, 1996] Adriaans P. e Zantinge, D., *Data Mining*. Addison-Wesley, 1996.
- [Akaike, 1973] Akaike, H., Information theory and an extension of the maximum likelihood principle, em B. Petrov e F. Csaki eds. *2<sup>nd</sup> International Symposium on Information Theory*, 267-281. Budapest. 1973.
- [Allen, 1988] Allen, P.M., Evolution: Why the whole is greater than the sum of the parts, em W. Wolff, C.-J. Soeder e F.R. Drepper eds., *Ecodynamics: Contribution to Theoretical Ecology*, Part 1: Evolution. *Proc. Int. Workshop*, Springer-Verlag, Berlin, 2-30, 1988.
- [Allen e Starr, 1982] Allen, T.F.H. e Starr, T.B., *Prediction of photochemical transformation of pollutants in aquatic environment*. University of Chicago Press, 1982.

- [Alves, 2002] Alves, V., Resolução de Problemas em Ambientes Distribuídos Uma Contribuição nas Áreas da Inteligência Artificial e da Saúde, Tese de Doutoramento, Departamento de Informática, Universidade do Minho, Portugal, 2002.
- [Alves et al., 2002] Alves, V., Abelha, A., Ribeiro, V. e Machado, J., Interchanging Agents and Humans in problem solving in the area of the medicine, em *Proceedings of the 2nd LASTED International Conference on Artificial Intelligence and Applications*, Benalmadena, Espanha, 2002.
- [Analide, 2004] Analide C., Antropopatia em Entidades Virtuais, Tese de Doutoramento, Departamento de Informática, Universidade do Minho, 2004.
- [Analide e Neves, 2002] Analide, C. e Neves, J., Antropopatia em Entidades Virtuais. *Em Actas do Workshop de Teses e Dissertações em Inteligência Artificial (WTDLA02)*, Recife, Brasil, 2002.
- [Apte et al., 2002] Apte, C., Liu, B., Pednault, E., Smith, P., Business applications of Data Mining. *Communications of the ACM*, 45 (8): 49-53, 2002.
- [Azoff, 1994] Azoff, E., Neural Networks Time Series Forecasting of Financial Markets. John Wiley & Sons, Inc., USA. 1994.
- [Baveco e Lingeman, 1992] Baveco, J.M. e Lingeman, R., An object-oriented tool for individual-oriented simulation: host-parasitoid system application. *Ecol. Modelling*, 61: 267-286, 1992.
- [Berry e Linoff, 2000] Berry, M. e Linoff, G., Mastering Data Mining: The Art and the Science of Customer Relationships Management. John Wiley & Sons, Inc., USA, 2000.
- [Bond, 1990] Bond, A., A computational model for organizations of cooperating intelligent agents. *SIGOIS Bulletin*, 11, 1990.
- [Berson et al., 2000] Berson, A., Smith, S. e Thearling, K., Building Data Mining Applications. McGraw Hill, USA, 2000
- [Bose e Liang, 1996] Bose, N. e Liang, P., Neural Networks Fundamentals with Graphs, Algorithms and Applications. McGraw-Hill. USA. 1996.
- [Boser et al., 1992] Boser, B., Guyon, I. e Vapnik, V., A training algorithm for optimal margin classifiers. *Proceedings of the 5<sup>th</sup> Annual Workshop on Computational Learning Theory*, 144-152. San Mateo. 1992.
- [Bossel, 1992] Bossel, H., Real structure process description as the basis of understanding ecosystems. *Ecol. Modelling*, 63: 261-276 1992.

- 
- [Box e Jenkins, 1976] Box, G. e Jenkins, G., *Time Series Analysis: Forecasting and Control*. Holden Day. San Francisco. USA. 1976.
- [Broomhead e Lowe, 1988] Broomhead, D. e Lowe, D., Multivariable functional interpolation and adaptative networks. *Complex Systems*, 2: 321-355, 1988.
- [Brown, 1995] Brown, J.H., *Macroecology*. The University of Chicago Press, Chicago, 1995.
- [Cavedon e Tilhar, 1995] Cavelon, L. e Tilhar, G., A logical framework for multi-agent systems and joint attitudes. *Proceedings of First Australian Workshop on Distributed Artificial Intelligence*, Canberra, Australia, 1995.
- [Chang et al., 2001] Chang, G., Healey, M., McHugh, J. e Wang, J., *Mining the Word Wide Web - An Information Search Approach*. Kluwer Academic Publishers, USA, 2001.
- [Chapman et al., 2000] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. e Wirth, R., *CRISP-DM 1.0 - Step-by-step Data Mining Guide*, CRISP-DM Consortium, 2000.
- [Chatfield, 1989] Chatfield, C., *The Analysis of Time Series - An Introduction*. Chapman and Hall. UK. 1989.
- [Cios et al., 1998] Cios, K., Pedrycz, W. e Swiniarski, R., *Data Mining Methods for Knowledge Discovery*. USA, 1998.
- [Clementine, 1998] *Clementine User Guide version 5*, Integral Solutions Limited, 1998.
- [Cortez, 2002] Cortez, P., *Modelos Inspirados na Natureza para a Previsão de Séries Temporais*. Tese de Doutoramento, Departamento de Informática, Universidade do Minho, 2002.
- [Cortez et al., 1996] Cortez, P., Machado, J. e Neves, J., An Evolutionary Artificial Neural Network Time Series Forecasting System, em M. Hanza ed. *LASTED International Conference on Artificial Intelligence, Expert Systems and Neural Networks*, 278-281. Honolulu, Havai. IASTED ACTA Press. 1996.
- [Cortez et al., 1995] Cortez, P., Rocha, M., Machado, J. e Neves, J., A Neural Network Based Forecasting System, em *IEEE International Conference on Neural Networks Proceedings*. 5: 2689-2693, Perth, Austrália. 1995.
- [Cortez et al., 1999] Cortez, P., Rocha, e Neves, J., An Evolutionary and Connectionist Approach for Time Series Forecasting, em *Proceedings of International Conference on Systems Engineering - ICSE '99*. 19-24, Las Vegas, USA. 1999.

- [Cimatti e Serafini, 1995] Cimatti, A. e Serafini, L., Multiagent reasoning with belief contexts II: elaboration tolerance. *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS-95)*, 57-64, 1995.
- [Coelho, 1995] Coelho, H., *Inteligência Artificial em 25 Lições*. Fundação Calouste Gulbenkian, Lisboa, 1995.
- [Costa e Melo, 1997] Costa, J. e Melo, A., *Dicionário da Língua Portuguesa*. 7ª Edição, Porto Editora, Portugal, 1997.
- [De Angelis e Gross, 1992] De Angelis, D.L. e Gross, L.J., Individual-Based Models and Approaches, em D.L. De Angelis e L.J. Gross ed., *Ecology: Populations, Communities and Ecosystems*. Chapman and Hall, USA, 1992.
- [Denti et al., 1995] Denti, E., Natali, A. e Omicini, A., Robot Control Systems as Contextual Logic Programs, in *Logic Programming: Formal Methods and Practical Applications*, Studies in Computer Science and Artificial Intelligence 11, Elsevier, 1995.
- [Downing e Reed, 1996.] Downing, K. e Reed, M., Object-oriented migration modelling for biological impact assessment. *Ecol. Modelling*, 93: 203-219, 1996.
- [DRAOTA, 1999] Identificação e caracterização das fontes poluidoras de carácter pontual localizadas em bacias hidrográficas das albufeiras destinadas à produção de água e em bacias drenantes para zonas balneares. Relatório da Direcção Regional do Ambiente e Ordenamento do Território - Alentejo. 1999.
- [DRAOTA, 2003] Anuário de recursos hídricos do Alentejo - Ano Hidrológico 2001/2002. Relatório da Direcção Regional do Ambiente e Ordenamento do Território - Alentejo. 2003.
- [Dubois, 1979] Dubois, D.M., Catastrophe theory applied to water quality regulations of rivers, em S.E. Jørgensen ed., *State of Art of Ecological Modelling. Environmental Sciences and Applications. Proc. 7th Conf. Ecological Modelling*, Copenhaga, 751-758, 1979.
- [Durfee e Rosenchien, 1994] Durfee, E. e Rosenchien, J., Distributed problem solving and multiagent systems: comparisons and examples. *Proceedings of the International Workshop on Distributed Artificial Intelligence*, Seattle, USA, 1994.
- [Efron e Tibbshirani, 1993] Efron, B. e Tibbshirani, R., *An Introduction to the Bootstrap*. Chapman e Hall. USA. 1993.
- [Engelmore et al., 1988] Engelmore, R., Morgan, A. e Nii, H., HERSAY-II, em Engelmore, R., Morgan, T., eds, *Blackboard Systems*. Addison-Wesley Publishing Study Inc., USA, 1988.



- 
- [Faraday e Chatfield, 1998] Faraday, J. e Chatfield, C., Time series forecasting with neural networks: a case study. *Applied Statistics*, 47: 231-250, 1998.
- [Fayyad, 1997] Fayyad, U., Mining Databases: Towards Algorithms for Knowledge Discovery. *IEEE Computer Society Technical Committee on Data Engineering*, 21: 39-48, 1997.
- [Fayyad et al., 1996] Fayyad, U., Piatetsky-Shapiro, G., Smith, P. e Uthurusamy, R., Advances in Knowledge Discovery and Data Mining. AAAI Press/MIT Press, USA, 1996.
- [Ferber, 1999] Ferber, J., Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence. Addison-Wesley, 1999.
- [Ferreira, 1995] Ferreira, J., ECOWIN - an object-oriented ecological model for aquatic ecosystems. *Ecological Modelling*, 79: 21-34, 1995.
- [Freisleben e K. Ripper, 1995] Freisleben, B. e K. Ripper, K., Economic Forecasting using neural networks, em ICNN'95 - IEEE International Conference on Neural Networks Proceedings, Perth, Austrália. 1995.
- [Fontaine, 1981] Fontaine, T.D., A self-designing model for testing hypotheses of ecosystem development, em D. Dubois ed., Progress in Ecological Engineering and Management by Mathematical Modelling. *Proc. 2nd Int. Conf. State of Art of Ecological Modelling*, Liege, Bélgica, 281-291, 1981.
- [Gallant, 1993] Gallant, S., Neural Network Learning and Expert Systems. MIT Press. Cambridge. USA. 1993.
- [Gelfond e Lifschitz, 1990] Gelfond, M. e Lifschitz, V., Logic programs with classical negation. *Proceedings of the International Conference on Logic Programming*, 1990.
- [Giunchiglia et al., 1993] Giunchiglia, E., Traverso, P. e Giunchiglia, F., Multi-Context Systems as a Specification Framework for Complex Reasoning Systems, em J. Teur and T. Wetter eds., Formal Specification Methods for Complex Reasoning Systems, Ellis Horwood Publishers, 1993.
- [Glansdorff, e Prigogine, 1971] Glansdorff, P. e Prigogine, I., Thermodynamic Theory of Structure, Stability and Fluctuations. Wiley-Interscience, 1971.
- [Gleick, 1989] Gleick, J., Caos a Construção de uma Nova Ciência. Gradiva, Lisboa. 1989.
- [Goebel e Gruenwald, 1999] Goebel, M. e Gruenwald, L., A survey of Data Mining and knowledge discovery software tools. *ACM SIGKDD Explorations*, 1: 20-33, 1999.

- [Gonzalez e Dankel, 1993] Gonzalez, A. e Dankel, D., *The Engineering of Knowledge-based Systems*, Prentice Hall Inc., 1993.
- [Grimm, 1999] Grimm, V., Ten year of individual-based modelling in ecology: what have we learned and what could we learn in the future? *Ecological Modelling*, 115: 129-148, 1999.
- [Gruber, 1991] Gruber, T., The role of common ontology in achieving sharable, reusable knowledge bases, em J. Allen, R. Fikes, e E. Sandewall eds., *Principles of Knowledge Representation and Reasoning, KR'91. Proceedings of the Second International Conference*, Cambridge, Massachusetts, U.S.A., 1991.
- [Han e Kamber, 2001] Han, J. e Kamber, M., *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, USA, 2001.
- [Hasting et al., 1993] Hasting, A., Hom, C.L., Ellner, S., Turchin, P. e Godfray, H., Chaos in ecology: is mother nature a strange attractor?. *Ann. Rev. Ecol. Syst.* 34: 1-33, 1993.
- [Haykin, 1999] Haykin, S., *Neural Networks - A Comprehensive Foundation*. Prentice-Hall. New Jersey. 1999.
- [He et al., 1999] He, H., Mladenoff, D. e Boeder, J., An object-oriented forest landscape model and its representation of tree species. *Ecological Modelling*, 119: 1-19, 1999.
- [Herendeen, 1989] Herendeen, R., Energy intensity, residence time, exergy and ascendancy in dynamic ecosystems. *Ecol. Modelling*, 48: 19-44, 1989.
- [Hewitt, 1977] Hewitt, C., Viewing Control Structures as Patterns of Passing Messages. *Artificial Intelligence*, 8, 1977.
- [Holsheimer et al., 1995] Holsheimer, M., Kersten, M., Mannila, H. e Toivinen, H., A perspective on databases and data mining. *Proceedings of 1<sup>st</sup> International Conference of Knowledge Discovery and Data Mining*, 150-155, 1995.
- [Hopfield, 1982] Hopfield, J., Neural networks and physical systems with emergent collective computational abilities. *The National Academy of Science*. 79: 2554-2558, 1982.
- [Huhns e Singh, 1998] Huhns, M. e Singh, M., *Readings in Agents*, chapter Agents and Multiagent Systems: Themes, Approaches and Challenges. M. Huhns and M. Singh eds., Morgan Kaufmann Publishers, 1998.
- [Inmon et al., 1997] Inmon, W., Welch, H. e Glassey, K., *Managing the Data Warehouse Toolkit*. John Willey & Sons, 1997.

- 
- [Jagannathan et al., 1989] Jagannathan, V., Dodhiawala, R. e Baum, L., Blackboard Architectures and Applications. Academic Press, Inc., 1989.
- [Jennings e Wooldridge, 1998] Jennings, N. e Wooldridge, M., Agent Technology - Foundations, Applications and Markets. Springer-Verlag, 1998.
- [Jeppesen et al., 1990] Jeppesen, E.J., Mortensen, E., Sortkjaer, O., Kristensen, P., Bidstrup, J., Timmermann, M., Jensen, J.P., Hansen, A.M., Søndegaard, M., Müller, J.P., Jensen, J., Riemann, B., Lindegaard-Petersen, C., Bosselmann, S., Christoffersen, K., Dall, E. e Andersen, J.M., Fish manipulation as a lake restoration tool in a shallow, eutrophic temperate lakes. Cross-analysis of three Danish case studies. *Hydrobiologia*, 200/201: 205-218, 1990.
- [Jordan, 1995] Jordan, M., Why the logistic function? a tutorial discussion on probabilities and neural networks. MIT. USA. 1995.
- [Jørgensen, 1986] Jørgensen, S.E., Structural dynamic model. *Ecol. Modelläng*, 31: 1-9, 1986.
- [Jørgensen, 1988] Jørgensen, S.E., Fundamentals of Ecological Modelling. Elsevier, Amesterdão, 1988.
- [Jørgensen, 1990a] Jørgensen, S.E., Application of models in limnological research. *Verh. Internat. Verein. Limnol.*, 24: 61-67, 1990.
- [Jørgensen, 1990b] Jørgensen, S.E., Ecosystem theory, ecological buffer capacity, uncertainly and complexity. *Ecol. Modelläng*, 52: 125-133, 1990.
- [Jørgensen, 1992a] Jørgensen, S.E., Development models able to account for changes in species composition. *Ecol. Modelläng*, 62: 195-208, 1992.
- [Jørgensen, 1992b] Jørgensen, S.E., Parameters, ecological constraints and exergy. *Ecol. Modelläng*, 62: 163-170, 1992.
- [Jørgensen, 1994] Jørgensen, S.E., Models as instruments for combination of ecological theory and environmental practice. *Ecol. Modelläng*, 75/76: 5-20, 1994.
- [Jørgensen, 1995a] Jørgensen, S.E., State of art of ecological modelling in limnology. *Ecol. Modelläng*, 78: 101-115, 1995.
- [Jørgensen, 1995b] Jørgensen, S.E., The growth rate of zooplankton at the edge of chaos. *J. Theor. Biol.*, 175: 13-21, 1995.
- [Jørgensen, 1997] Jørgensen, S.E., Integration of Ecosystems Theories: A Pattern. 2<sup>a</sup> ed., Kluwer, Dordrecht, 1997.

- [Jørgensen, 1999] Jørgensen, S.E., State-of-art of ecological modelling with emphasis on development of structural dynamic models. *Ecol. Modelling*, 120: 75-96, 1999.
- [Jørgensen, e Mejer, 1979] Jørgensen, S.E. e Mejer, H.F., A holistic approach to ecological modelling. *Ecol. Modelling*, 7: 169-189, 1979.
- [Jørgensen e Nielsen, 1994] Jørgensen, S.E. e Nielsen, S.N., Models of the structural dynamics in lakes and reservoirs. *Ecol. Modelling*, 74: 39-46, 1994.
- [Jørgensen et al., 1995] Jørgensen, S.E., Nielsen, S.N. e Mejer, H., Emergy, environ, exergy and ecological modelling. *Ecol. Modelling*, 77: 99-109, 1995.
- [Karul et al., 2000] Karul, C. Soyupak, S., Çilesiz, A., Akbay, N. e Germen, E., Case studies on the use of neural networks in eutrophication modelling, *Ecological Modelling*, 134: 145-152, 2000.
- [Kauffman, 1996] Kauffman, S.A., *At Home in the Universe. The Search for Laws of Complexity.* Penguin Books, Oxford University Press, Oxford. 1996.
- [Kempf, 1980] Kempf, J., Multiple steady states and catastrophes in ecological models. *ISEM-Journal*, 2: 55-80, 1980.
- [Kohavi, 1995] Kohavi, R., A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*. Montreal, Canadá. 1995
- [Kohonen, 1982] Kohonen, T., Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43: 59-69, 1982.
- [Kowalski, 1994] Kowalski, R., Logic Without Model Theory, in *What is a Logical System? Studies in Logic and Computation*, Oxford Science Publications, 35-71, 1994.
- [Kuftrin, 1997] Kuftrin, R., Generating C4.5 Productions Rules in Parallel; *4<sup>th</sup> National Conference on Artificial Intelligence (AAAI-97)*, 1997.
- [Langley e Simon, 1995] Langley, P. e Simon, H., Applications of Machine Learning and Rule Induction, *Communications of the ACM*, 38, 1995.
- [Lapedes e Farber, 1987] Lapedes, A. e Farber, R., Non-linear signal processing using neural networks: prediction and system modelling. Technical Report LA-UR-87-2662, Los Alamos National Laboratory, USA, 1987.

- 
- [Lek et al., 1996] Lek S., Delacoste, M., Baran, P., Dimopoulos., Lauga, J. e Aulagnier, S., Application of neural networks to modelling nonlinear relationships in ecology. *Ecol. Modelling*, 90: 39-52, 1996.
- [Loehle, 1989] Loehle, C., Catastrophe theory in ecology: a critical review and an example of butterfly catastrophe. *Ecol. Modelling*, 49: 125-144, 1989.
- [Lorek e Sonnenschein, 1998] Lorek, H. e Sonnenschein, M., Object-oriented support for modelling and simulation of individual-oriented ecological models. *Ecol. Modelling*, 108: 77-96, 1998.
- [Lorek e Sonnenschein, 1998] Lorek, H. e Sonnenschein, M., Modelling and simulation software to support individual-based ecological modelling. *Ecol. Modelling*, 115: 199-216, 1999.
- [Lotka, 1924] Lotka, A.J., Contribution to the energetics of evolution. *Proc. Natl. Acad. Sci.*, 8: 147-150, 1924.
- [Luger e Stubblefield, 1998] Luger, G. e Stubblefield, W., Artificial Intelligence Structures and Strategies for Complex Problem Solving, Addison-Wesley, 1998.
- [Machado, 2002] Machado J., Agentes Inteligentes como Objectos dum Sistema Distribuído de Realidade Virtual, Tese de Doutoramento, Departamento de Informática, Universidade do Minho, 2002.
- [Maes, 1990] Maes, P., Situated Agents Can Have Goals, em Maes, P., ed., Designing Autonomous Agents. MIT Press, 1990.
- [Maier e Dandy, 1996] Maier, H. e Dandy, G., The use of artificial neural networks for the prediction of water quality parameters. *Water Resources Research*, 32: 1013-1022, 1996.
- [Maier et al., 1998] Maier, H., Dandy, G. e Burch, M., Use of artificial neural networks for modelling cyanobacteria *Anabaena* spp. in the River Murray, South Australia. *Ecological Modelling*, 105: 257-272, 1998.
- [Makridakis e Wheelwright, 1989] Makridakis, S. e Wheelwright S., Forecasting Methods for Management. John Wiley & Sons, NY, 5ª Ed., 1989.
- [Mamedov e Udalov, 2002] Mamedov, A. e Udalov, S., A computer tool to develop individual-based models for simulation of population interactions. *Ecological Modelling*, 147: 53-68, 2002.
- [Margalef, 1968] Margalef, R., Perspectives in Ecological Theory. Chicago University Press, 1968.

- [Margalef, 1994] Margalef, R., The place of continental waters in global ecology. *Limnology now, a paradigm of planetary problems*, em Margalef ed., 1-8. Elsevier NY. 1994.
- [Matoso et al., 2004] Matoso, A., Rasga, M., Santana, M. e Murteira, M., Principais albufeiras do Alentejo monitorizadas. Comissão de Coordenação da Região Alentejo, Direcção de Serviços de Monitorização Ambiental. 2004.
- [McCluskey, 1993] McCluskey, P., Feedforward and recurrent neural networks and genetic programs for stock market and time series forecasting. Thesis for master science degree, Brown University, USA. 1993.
- [McCulloch e Pitts, 1943] McCulloch, W. e Pitts, W., A logical calculus of the ideas immanent in nervous activity, *Bulletin of Mathematical Biophysics*, 5: 115-133. 1943.
- [Mejer, e Jørgensen, 1979] Mejer, H.F. e Jørgensen, S.E., Energy and ecological buffer capacity, em S.E. Jørgensen ed., *State of Art of Ecological Modelling. Environmental Sciences and Applications. Proc. 7th Conf. Ecological Modelling*, Copenhaga, 829-846, 1979.
- [Memmi, 1989] Memmi, D., Connectionism and artificial intelligence. *Neuro-Nimes'89 International workshop on neural networks and their applications*. 17-34. Nimes, França. 1989.
- [Meyer e Pampagnin, 1979] Meyer, J.A. e Pampagnin, N., The utility of the Simscript II language for the simulation of complex Predator-prey relationships, em S.E. Jørgensen ed., *State of Art of Ecological Modelling, Environmental Sciences and Applications. Proc. 7th Conf. Ecological Modelling*, Copenhaga, 801, 1979.
- [Michalski et al., 1998] Michalski, R., Bratko, I. e Miroslav, K., *Machine Learning and Data Mining Methods and Applications*. John Willey and Sons, Inc., Londres, 1998.
- [Minsk, 1986] Minsk, M., *The Society of Mind*. Simon and Schuster, 1986.
- [Muetzelfeldt, 1979] Muetzelfeldt, R.I., Towards an ecological-orientated simulation language, em S.E. Jørgensen ed., *State of Art of Ecological Modelling. Environmental Sciences and Applications. Proc. 7th Conf. Ecological Modelling*, Copenhaga, 771. 1979.
- [Neves, 1984] Neves, J., A Logic Interpreter to Handle Time and Negation in Logic DataBases. *Proceedings of the ACM'84 Annual Conference - The Fifth Generation Challenge*, San Francisco, California, USA, 1984.
- [Neves et al., 1988] Neves, J. Machado, J. Cortez, P. Arteiro, J. e Vicente, H., A distributed Web-based water quality diagnosis and simulation in lakes and reservoirs. *Proceedings of the 10th European Simulation Symposium and Exhibition*, Nottingham, 1998.

- 
- [Neves et al., 1994] Neves, J. Santos M. e Alves V., An Adaptable and Dynamic Architecture for Distributed Problem Solving Based on the Blackboard Paradigm. *Proceedings of the 7th Australian Joint Conference on Artificial Intelligence*, Armidale, Austrália, 1994.
- [Neves et al., 1997] Neves, J., Machado, J., Analide, C., Novais, P., and Abelha, A., Extended Logic Programming Applied to the Specification of Multi-Agent Systems and Their Computing Environments. *Proceedings of the 1997 IEEE - International Conference on Intelligent Processing Systems*, China, 1997.
- [Neves e Machado, 1997] Neves J.e Machado J., Formalizing Context in Knowledge Management Systems. *Proceedings of the International Workshop of Distributed Artificial Intelligence and Multi-Agent Systems*, St Petersburg, Russia, 1997.
- [Nielsen, 1992] Nielsen, S.N., Strategies for structural-dynamical modelling. *Ecol. Modelling*, 63: 323-336, 1992.
- [Novais, 2003] Novais, P., Teoria dos Processos de Pré-Negociação em Ambientes de Comércio Electrónico, Tese de Doutoramento, Departamento de Informática, Universidade do Minho, 2003.
- [Nwana, 1996] Nwana, H., Software Agents: An Overview. *Knowledge Engineering Review*, 1 - 40, 1996.
- [Odum e Pinkerton, 1955] Odum, H.T. e Pinkerton, R.C., Time's speed regulator: The optimum efficiency for maximum power output in physical and biological systems. *Amer. Sci.*, 43: 331-343, 1955.
- [O'Hare e Jennings, 1996] O'Hare, G. e Jennings, N., Foundations of Distributed Artificial Intelligence. John Wiley & Sons, New York, USA, 1996.
- [O'Neill et al., 1986] O'Neill, R.V., DeAngelis, D.L. Waide , J.B. e Allen, T.F.H., A hierarchical concept of ecosystems. Princeton University Press, USA, 1986.
- [O'Neill et al., 1975] O'Neill, R.V., Hanes, W.F., Ausmus, B.S. e Reichle, D.E., A theoretical basis for ecosystems analysis with particular reference to element cycling, em F.G. Howell, J.B. Gentry e Smith, M.H. eds., Mineral Cycling in Southeastern Ecosystems. *NTIS pub. CONF-740513*, 28-40, 1975.
- [Papadourakis et al., 1993] Papadourakis, G., Spanoudakis, G. e Gotsias, A., Application of neural networks in short-term stock price forecasting. *Proceedings of 1<sup>st</sup> International Workshop on Neural Networks in the Capital Markets*, Londres, UK. 1993.

- [Patten, 1983] Patten, B.C., On the quantitative dominance of indirect effects in ecosystems, em W.K. Lauenroth, G.V. Skogerboe e M. Flug eds., *Analysis of Ecological Systems: State of Art in Ecological Modelling*. Elsevier, Amesterdão, 1983.
- [Patten, 1991] Patten, B.C., Network ecology: indirect determination of the life-environment relationship in ecosystems, em M. Higashi e T.P. Burns eds., *Theoretical Studies of Ecosystems: The Network Perspective*. Cambridge University Press, 288-351, 1991.
- [Patten, 1997] Patten, B.C., Synthesis of chaos and sustainability in a nonstationary linear dynamic model of the American black bear (*Ursus americanus* Pallas) in the Adirondack mountains of New York. *Ecol. Modelling*, 100: 11-42, 1997.
- [Patterson, 1996] Patterson, D., *Artificial Neural Networks - Theory and Application*. Prentice Hall, Singapura. 1996.
- [Poston, e Stewart, 1978] Poston, T. e Stewart, I., *Catastrophe Theory and Its Applications*. Pitman, Londres, 1978.
- [Prechelt, 1994] Prechelt, L., PROBEN1 - A Set of Neural Networks Benchmark Problem and Benchmarking Rules. Research Report, Fakultät für Informatik, Universität Karlsruhe, Alemanha. 1994.
- [Quinlan, 1997] Quinlan, J., Simplifying Decisions Trees, *International Journal of Man-Machine Studies*, 27: 221-234, 1997.
- [Quinlan, 1998] Quinlan, J., *C4.5 Programs for Machine Learning*. Morgan Kaufmann Publishers, Inc. USA, 1998.
- [Ramakrishman, 1998] Ramakrishman, R., *Databases Management Systems*. McGraw-Hill International Editions. 1998.
- [Rao e Georgeff, 1995] Rao, A. e Georgeff, P., BDI Agents: from Theory to Practice. *Proceedings of the First International Conference on Multi-Agent Systems - ICMAS95*, San Francisco, USA, 1995.
- [Ray et al., 2001] Ray, S., Berec, L., Straškraba, M. e Jørgensen, S.E., Optimization of exergy and implications of body sizes of Phytoplankton and Zooplankton in an aquatic ecosystem model. *Ecological Modelling*, 140: 219-234, 2001.
- [Recknagel, 2001] Recknagel, F., Applications of machine learning to ecological modelling. *Ecological Modelling*, 146: 303-310, 2003.



- 
- [Recknagel, 2003] Recknagel, F., Simulation of aquatic food web and species interactions by adaptive agents embodied with evolutionary computation: a conceptual framework. *Ecological Modelling*, 170: 291-302, 2003.
- [Recknagel e Wilson, 2000] Recknagel, F. e Wilson, H., Elucidation and Prediction of Aquatic Ecosystems by Artificial Neuronal Networks, em S. Lek, J.-F. Guégan eds., *Artificial Neuronal Networks: Application to Ecology and Evolution*. Springer-Verlag. Berlim. 143-155. 2000.
- [Rich et al., 1991] Rich, E., and Knight, K. *Artificial Intelligence*. McGraw-Hill Inc., 1991.
- [Rodrigues, 1992] Rodrigues, A., *Modelação Matemática da Qualidade da Água em Albufeiras*, Tese de Doutoramento, Universidade Nova de Lisboa, Portugal, 1992.
- [Rojas, 1996] Rojas, R., *Neural Networks - A Systematic Introduction*. Springer-Verlag. Alemanha. 1996.
- [Rud, 2001] Rud, O., *Data Mining Cookbook*. John Wiley & Sons, Inc. USA. 2001.
- [Rumelhart, Hinton e Williams, 1986] Rumelhart, D., Hinton, G. e Williams, R., Learning Internal representation by error propagation, em D. Rumelhart e J. McClelland eds. *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*. Vol. 1: 318 - 362, MIT Press Cambridge MA. 1986.
- [Rumelhart et al., 1994] Rumelhart, D., Widrow, B. e Lehr, M., The basic ideas in neural networks. *Communications of the ACM*, 37: 87, 1994.
- [Russel e Norvig, 1995] Russel, S. e Norvig, P., *Artificial Intelligence - A Modern Approach*. Prentice Hall, 1995.
- [Salomonsen e Jensen, 1996] Salomonsen, J. e Jensen J., Use of lake model to examine exergy response to changes in phytoplankton growth parameters and species composition. *Ecol. Modelling*, 87: 41-49, 1996.
- [Salski, 1992] Salski, A., Fuzzy knowledge-based models in ecological research. *Ecol. Modelling*, 63: 102-112, 1992.
- [Santos, 1999] Santos M., *Sistemas de Classificação em Ambientes Distribuídos*, Tese de Doutoramento, Departamento de Informática, Universidade do Minho, 1999.
- [Santos e Neves, 1998] Santos, M. e Neves, J., Modelling Learning Classifiers as Multiagent Systems, *Proceedings of II Iberoamerican Workshop on DAI and MAS*, Toledo, Espanha, 1998.

- [Saraiva et al., 1994] Saraiva, M., Pinto, P., Rabaça, J., Ramos, A. e Revez, M., Protection, reclamation and improvement of small urban streams in Portugal. The Ecological Bases of River Management, em Harper e Ferguson eds., John Willey e Sons, NY, 275-288. 1994.
- [Scardi e Harding Jr., 1999] Scardi, M. e Harding Jr., L., Developing an empirical model of phytoplankton primary production: a neural network case study. *Ecol. Modelling*, 120: 213-223, 1999.
- [Schindler, 1988] Schindler, D.W., Effects of acid rain on freshwaters ecosystems. *Science*, 239: 149-157, 1988.
- [Schleiter et al., 1999] Schleiter, I., Borchardt, D., Wagner, R., Dapper, T., Schmidt, K.-D., Schmidt, H.-H. e Werner, H., Modelling water quality, bioindication and population dynamics in lotic ecosystem using neural networks. *Ecol. Modelling*, 120: 271-286, 1999.
- [Schoffeniels, 1976] Schoffeniels, E., Anti-Chance. Pergamon Press, USA, 1976.
- [Schwarz, 1978] Schwarz, G., Estimating the dimension of a model. *The Annals of Statistics*, 6: 461-464, 1978.
- [Sequeira et al., 1991] Sequeira, R.A., Sharpe, P.J.H., Stone, N.D., El-Zik, K.M. e Makela, M.E., Object-oriented simulation: plant growth and discrete organ to organ interactions. *Ecol. Modelling*, 58: 55-89, 1991.
- [Sequeira et al., 1997] Sequeira, R.A., Olsen, R.L. e McKinion, J.M., Implementing generic, object-oriented models in biology. *Ecol. Modelling*, 94: 17-31, 1997.
- [Silvert, 1993] Silvert, W., Object-oriented ecosystem modelling. *Ecol. Modelling*, 68: 37-53, 1993.
- [Sharda e Rampal, 1996] Sharda, R. e Rampal R., Neural Networks and Management Science/Operations Research: A Bibliographic Essay. *Encyclopedia of Library and Information Science*, 61: 247-259, 1996.
- [Sarle, 1999] Sarle, W., Neural network frequently asked questions. Available from <ftp://ftp.sas.com/pub/neural/FAQ.html>.
- [Shi et al., 1999] Shi, Z., Tamura, Y. e Ozaki, T., Nonlinear time series modelling with the radial basis function-based state-dependent autoregressive model. *Internacional Journal of Systems Science*, 30: 717-727, 1999.
- [Shoneberg, 1990] Shoneberg, E., Price prediction using neural networks: a project report. *Neurocomputing*, 2: 17-27, 1990.

- 
- [Søndergaard, 1989] Søndergaard, M., Phosphorus release from a hypertrophic lake sediments; experiments with intact sediment cores in a continuous flow system. *Arch. Hydrobiol.*, 116: 45-59, 1989.
- [Sousa, 2004] Sousa, C., Data Mining: Metodologias, Tecnologias, Modelos e Aplicações. Dissertação de Mestrado, Universidade do Minho, 2004.
- [Stenseth, 1986] Stenseth, N.C., Darwinian evolution in ecosystems: a survey of some ideas and difficulties together with some possible solutions, em J.L. Casti e A. Karlqvist eds., Complexity, Language and Life: Mathematical Approaches. Springer-Verlag, Berlim, 105-129, 1986.
- [Straškraba, 1979] Straškraba, M., Natural control mechanisms in models of aquatic ecosystems. *Ecol. Modelling*, 6: 305-322, 1979.
- [Straškraba, 1994] Straškraba, M., Ecotechnological models for reservoir water quality management. *Ecol. Modelling*, 74: 1-38, 1994.
- [Straškraba, 2001] Straškraba, M., Natural control mechanisms in models of aquatic ecosystems. *Ecol. Modelling*, 140: 195-205, 2001.
- [Streeter e Phelps, 1925] Streeter, H. e Phelps, E., A study of pollution and the natural purification of the Ohio River. Public Health Bulletin n° 146, USA Public Health Service. 1925.
- [Suárez, 1999] Suárez, I., Mastering chaos in ecology. *Ecol. Modelling*, 117: 305-314, 1999.
- [Tang e Fishwick, 1993] Tang, Z. e Fishwick, F., Feed-forward neural nets as models for time series forecasting. *ORSA Journal of Computing*, 5: 374-386, 1993.
- [Thuraisingham, 1999] Thuraisingham, B., Data Mining Technologies, Techniques, Tools and Trends. CRC Press LLC. USA. 1999.
- [Traylor e Gelfond, 1993] Traylor, B. e Gelfond, M., Representing null values in Logic Programming. *Proceedings of the ILPS'93 Workshop on Logic Programming with Incomplete Information*, Vancouver, Canada, 1993.
- [Turban e Aronson, 1998] Turban, E. e Aronson, J., Decision Support Systems and Intelligent Systems. Prentice Hall, 1998.
- [Ulanowicz, 1986] Ulanowicz, R.E., Growth and Development, Ecosystems Phenomenology. Springer-Verlag, 1986.

- 
- [Ulbricht, 1994] Ulbricht, C., Multi-recurrent networks for traffic forecasting. *Proceedings of AAAI'94 Conference*, Seattle, USA. 1994.
- [URL 1] <http://www.inag.pt/>
- [Vayenas e Pavlou, 2001] Vayenas, D. e Pavlou, S., Chaotic dynamics of a microbial system of coupled food chains. *Ecological Modelling*, 136: 285-295, 2001.
- [Volterra, 1926] Volterra, V., Fluctuations in the abundance of a species considered mathematically. *Nature*, 188: 558-560, 1926.
- [Weigend e Gershenfeld, 1994] Weigend, A., e Gershenfeld N., Time Series Prediction: Forecasting the Future and Understanding the Past. Addison-Wesley. USA. 1994.
- [Weiss, 1999] Weiss, G., Multiagent Systems - A Modern Approach to Distributed Artificial Intelligence. MIT Press, 1999.
- [Wen e Lee, 1998] Wen, C. e Lee, C., A neural network approach to multiobjective optimization for water quality management in a river basin. *Water Resources Research*, 34: 427-436, 1998.
- [Werner e Obach, 2001] Weerner, H. e Obach, M., New neural networks types estimating the accuracy of response for ecological modelling. *Ecological Modelling*, 146: 289-298, 2001.
- [Whittaker et al., 1971] Whittaker, R.H. e Woodwell, G. M., Evolution of natural communities, em J.A. Weins ed., Ecosystem Structure and Function. Oregon State University Press, Corvallis, 137-159, 1971.
- [Winstansley, 1991] Winstansley, G., Artificial Intelligence in Engineering. John Wiley & Sons, 1991.
- [Wolfram, 1984a] Wolfram, S., Computer software in science and mathematics. *Scient. Amer.*, 251: 140-151, 1984.
- [Wolfram, 1984b] Wolfram, S., Cellular automata as models of complexity. *Nature*, 311: 419-424, 1984.
- [Wooldridge, 1999] Wooldridge, M., Intelligent Agents, em G. Weiss, ed. Multiagent Systems - A modern Approach to Distributed Artificial Intelligence. MIT Press, 1999.
- [Wooldridge e Jennings, 1995] Wooldridge, M. e Jennings, N., Intelligent agents: theory and practice. *The Knowledge Engineering Review*, 115-152, 1995.
- [Yao e Poh, 1995] Yao, J. e Poh, H., Forecasting the KLSE index using neural networks. *ICNN'95 - IEEE International Conference on Neural Networks Proceedings*. Perth, Austrália. 1995.

[Zeeman, 1978] Zeeman, E.C., Catastrophe Theory: Selected Papers 1972-1977. Addison-Wesley, Londres, 1978.

[Zupan e Gasteiger, 1993] Zupan, J. e Gasteiger, J., Neural Networks for Chemists an Introduction. VCH. NY. 1993.



## Apêndice A

# Conjunto de Regras Gerado a Partir dos Dados de Qualidade da Água para as Albufeiras do Divôr, Monte Novo e Vigia - Modelo 1

Neste anexo apresenta-se o conjunto de regras gerado a partir dos dados de qualidade da água para as três albufeiras sobre as quais recaiu o presente estudo (i.e. Divôr, Monte Novo e Vigia) e que dão corpo ao modelo 1 (Capítulo 5). No final da apresentação de cada uma das regras destaca-se o **suporte** (i.e. o número de casos onde a referida regra é encontrada) e a **confiança** (i.e. é a probabilidade condicional dessa regra).

(i) - Regras para a classe D:

### Regra 1

Se  $oD \leq 61,8$

e  $CBO_5 > 2,6$

e  $P_{total} > 0,03$

Então  $\rightarrow D$

(44, 0,977)

### Regra 2

Se  $oD > 61,8$

e  $CQO \leq 19,8$

e  $N_{Kjeldahl} \leq 1,01$

e  $P_{\text{total}} \leq 0,124$

e  $\text{pH} > 8,37$

Então  $\rightarrow$  D

**(3, 0,667)**

**Regra 3**

Se  $\text{oD} > 61,8$

e  $\text{CQO} \leq 19,8$

e  $N_{\text{Kjeldahl}} \leq 1,01$

e  $P_{\text{total}} > 0,171$

Então  $\rightarrow$  D

**(4, 1,0)**

**Regra 4**

Se  $\text{oD} > 61,8$

e  $\text{CQO} \leq 19,8$

e  $N_{\text{Kjeldahl}} > 1,01$

e  $\text{fosfato}_{\text{total}} > 0,042$

e  $I_{\text{radiação}} \leq 93,958$

Então  $\rightarrow$  D

**(3, 1,0)**

**Regra 5**

Se  $\text{oD} > 61,8$

e  $\text{CQO} > 19,8$

e  $\text{transparência} \leq 0,3$

Então  $\rightarrow$  D

**(2, 1,0)**

**Regra 6**

Se  $\text{oD} > 61,8$

e  $\text{CQO} > 19,8$

e  $\text{transparência} > 0,3$

e  $\text{amoníaco} \leq 0,002$

e  $\text{CBO}_5 \leq 7,0$

e  $\text{nitritos} \leq 0,021$



e  $N_{\text{Kjeldahl}} > 0,86$

e  $N_{\text{Kjeldahl}} \leq 1,02$

Então  $\rightarrow$  D

**(4, 1,0)**

### **Regra 7**

Se  $\text{oD} > 61,8$

e  $\text{CQO} > 19,8$

e  $\text{transparência} > 0,3$

e  $\text{amoníaco} \leq 0,002$

e  $\text{CBO}_5 \leq 7,0$

e  $\text{nitritos} > 0,021$

Então  $\rightarrow$  D

**(3, 1,0)**

### **Regra 8**

Se  $\text{oD} > 61,8$

e  $\text{CQO} > 19,8$

e  $\text{transparência} > 0,3$

e  $\text{amoníaco} \leq 0,002$

e  $\text{CBO}_5 > 7,0$

Então  $\rightarrow$  D

**(2, 1,0)**

### **Regra 9**

Se  $\text{oD} > 61,8$

e  $\text{CQO} > 19,8$

e  $\text{transparência} > 0,3$

e  $\text{amoníaco} > 0,009$

e  $\text{amoníaco} \leq 0,016$

e  $N_{\text{amoniaco}} \leq 0,25$

Então  $\rightarrow$  D

**(3, 1,0)**

(ii) - Regras para a classe C:

**Regra 1**

Se  $oD \leq 61,8$

e  $CBO_5 \leq 2,6$

Então  $\rightarrow C$

(7, 0,857)

**Regra 2**

Se  $oD \leq 61,8$

e  $CBO_5 > 2,6$

e  $P_{total} \leq 0,03$

Então  $\rightarrow C$

(4, 0,75)

**Regra 3**

Se  $oD > 61,8$

e  $CQO \leq 19,8$

e  $N_{Kjeldahl} \leq 1,01$

e  $P_{total} > 0,124$

e  $P_{total} \leq 0,171$

Então  $\rightarrow C$

(2, 1,0)

**Regra 4**

Se  $oD > 61,8$

e  $CQO \leq 19,8$

e  $N_{Kjeldahl} > 1,01$

e  $\text{fosfato}_{total} \leq 0,042$

Então  $\rightarrow C$

(7, 1,0)

**Regra 5**

Se  $oD > 61,8$

e  $CQO \leq 19,8$

e  $N_{Kjeldahl} > 1,01$

e  $\text{fosfato}_{\text{total}} > 0,042$

e  $I_{\text{radiação}} > 93,958$

Então  $\rightarrow C$

**(6, 0,833)**

### **Regra 6**

Se  $\text{oD} > 61,8$

e  $\text{CQO} > 19,8$

e  $\text{transparência} > 0,3$

e  $\text{amoníaco} \leq 0,002$

e  $\text{CBO}_5 \leq 7,0$

e  $\text{nitritos} \leq 0,021$

e  $N_{\text{Kjeldahl}} \leq 0,86$

Então  $\rightarrow C$

**(4, 1,0)**

### **Regra 7**

Se  $\text{oD} > 61,8$

e  $\text{CQO} > 19,8$

e  $\text{transparência} > 0,3$

e  $\text{amoníaco} \leq 0,002$

e  $\text{CBO}_5 \leq 7,0$

e  $\text{nitritos} \leq 0,021$

e  $N_{\text{Kjeldahl}} > 1,02$

Então  $\rightarrow C$

**(2, 1,0)**

### **Regra 8**

Se  $\text{oD} > 61,8$

e  $\text{CQO} > 19,8$

e  $\text{transparência} > 0,3$

e  $\text{amoníaco} > 0,002$

e  $\text{amoníaco} \leq 0,009$

Então  $\rightarrow C$

**(21, 1,0)**

**Regra 9**

Se  $oD > 61,8$   
 e  $CQO > 19,8$   
 e  $transparência > 0,3$   
 e  $amoníaco > 0,009$   
 e  $amoníaco \leq 0,016$   
 e  $N_{amoniacal} > 0,25$   
 Então  $\rightarrow C$

**(2, 1,0)**

**Regra 10**

Se  $oD > 61,8$   
 e  $CQO > 19,8$   
 e  $transparência > 0,3$   
 e  $amoníaco > 0,016$   
 Então  $\rightarrow C$

**(6, 1,0)**

*(iii)* - Regras para a classe B:

**Regra 1**

Se  $oD > 61,8$   
 e  $CQO \leq 19,8$   
 e  $N_{Kjeldahl} \leq 1,01$   
 e  $P_{total} \leq 0,124$   
 e  $pH \leq 8,37$   
 Então  $\rightarrow B$

**(10, 1,0)**

## Apêndice B

# Conjunto de Regras Gerado a Partir de um Grupo Restrito de Parâmetros de Qualidade da Água para as Albufeiras do Divôr, Monte Novo e Vigia - Modelo 2

Neste anexo apresenta-se o conjunto de regras gerado a partir de um grupo restrito de parâmetros de qualidade da água para as três albufeiras sobre as quais recaiu o presente estudo (i.e. Divôr, Monte Novo e Vigia) e que dão corpo ao modelo 2 (Capítulo 5). No final da apresentação de cada uma das regras destaca-se o **suporte** (i.e. o número de casos onde a referida regra é encontrada) e a **confiança** (i.e. é a probabilidade condicional dessa regra).

(i) - Regras para a classe D:

### Regra 1

Se  $oD \leq 61,8$

e  $pH > 7,17$

e  $transparência \leq 1,3$

e  $cloretos \leq 45$

Então  $\rightarrow D$

(29, 0,895)

### Regra 2

Se  $oD \leq 61,8$

e pH > 7,17  
 e transparência > 1,3

Então → D

(22, 0,951)

**Regra 3**

Se oD > 61,8  
 e transparência ≤ 0,3

Então → D

(4, 1,0)

**Regra 4**

Se oD > 61,8  
 e transparência > 0,3  
 e  $v_{\text{vento}} > 1,925$   
 e  $v_{\text{vento}} \leq 2,836$   
 e cloretos > 18  
 e precipitação ≤ 1,74

Então → D

(5, 0,796)

**Regra 5**

Se oD > 61,8  
 e transparência > 0,3  
 e  $v_{\text{vento}} > 2,836$   
 e cloretos > 18  
 e precipitação ≤ 0,925  
 e  $T_{\text{água}} \leq 23$

Então → D

(17, 0,578)

**Regra 6**

Se oD > 61,8  
 e oD ≤ 75,9  
 e transparência > 0,3  
 e  $v_{\text{vento}} > 2,836$

e cloretos  $> 18$   
 e cloretos  $\leq 25$   
 e precipitação  $> 0,925$   
 e precipitação  $\leq 1,74$   
 e  $T_{\text{água}} \leq 23$   
 Então  $\rightarrow D$

**(3, 1,0)**

(ii) - Regras para a classe C:

**Regra 1**

Se  $oD \leq 61,8$   
 e  $pH \leq 7,17$   
 Então  $\rightarrow C$

**(4, 1,0)**

**Regra 2**

Se  $oD \leq 61,8$   
 e  $pH > 7,17$   
 e transparência  $\leq 1,3$   
 e cloretos  $> 45$   
 Então  $\rightarrow C$

**(2, 1,0)**

**Regra 3**

Se  $oD > 61,8$   
 e transparência  $> 0,3$   
 e transparência  $\leq 1,1$   
 e  $v_{\text{vento}} \leq 1,925$   
 e  $T_{\text{água}} \leq 15,8$   
 Então  $\rightarrow C$

**(2, 1,0)**



**Regra 4**

Se  $oD > 61,8$   
 e  $\text{transparência} > 1,1$   
 e  $v_{\text{vento}} \leq 1,925$   
 Então  $\rightarrow C$

**(6, 1,0)**

**Regra 5**

Se  $oD > 61,8$   
 e  $\text{transparência} > 0,3$   
 e  $v_{\text{vento}} > 1,925$   
 e  $\text{cloretos} \leq 18$   
 Então  $\rightarrow C$

**(19, 0,895)**

**Regra 6**

Se  $oD > 75,9$   
 e  $\text{transparência} > 0,3$   
 e  $v_{\text{vento}} > 2,836$   
 e  $\text{cloretos} > 18$   
 e  $\text{cloretos} \leq 25$   
 e  $\text{precipitação} > 0,925$   
 e  $\text{precipitação} \leq 1,74$   
 e  $T_{\text{água}} \leq 23$   
 Então  $\rightarrow C$

**(3, 0,758)**

**Regra 7**

Se  $oD > 61,8$   
 e  $\text{transparência} > 0,3$   
 e  $v_{\text{vento}} > 2,836$   
 e  $v_{\text{vento}} \leq 3,376$   
 e  $\text{cloretos} > 25$   
 e  $\text{precipitação} > 0,925$   
 e  $\text{precipitação} \leq 1,74$



e  $T_{\text{água}} \leq 23$

Então  $\rightarrow C$

(5, 0,508)

**Regra 8**

Se  $oD > 61,8$

e  $\text{transparência} > 0,3$

e  $v_{\text{vento}} > 2,836$

e  $\text{cloretos} > 18$

e  $\text{precipitação} \leq 1,74$

e  $T_{\text{água}} > 23$

Então  $\rightarrow C$

(7, 1,0)

**Regra 9**

Se  $oD > 61,8$

e  $\text{transparência} > 0,3$

e  $v_{\text{vento}} > 1,925$

e  $\text{cloretos} > 18$

e  $\text{precipitação} > 1,74$

e  $\text{pH} > 7,5$

Então  $\rightarrow C$

(6, 0,833)

(iii) - Regras para a classe B:

**Regra 1**

Se  $oD > 61,8$

e  $\text{transparência} > 0,3$

e  $\text{transparência} \leq 1,1$

e  $v_{\text{vento}} \leq 1,925$

e  $T_{\text{água}} > 15,8$

Então  $\rightarrow B$

(2, 1,0)

**Regra 2**

Se  $oD > 61,8$   
e  $\text{transparência} > 0,3$   
e  $v_{\text{vento}} > 3,376$   
e  $\text{cloretos} > 25$   
e  $\text{precipitação} > 0,925$   
e  $\text{precipitação} \leq 1,74$   
e  $T_{\text{água}} \leq 23$   
Então  $\rightarrow B$

**(2, 1,0)**

**Regra 3**

Se  $oD > 61,8$   
e  $\text{transparência} > 0,3$   
e  $v_{\text{vento}} > 1,925$   
e  $\text{cloretos} > 18$   
e  $\text{precipitação} > 1,74$   
e  $\text{pH} \leq 7,5$   
Então  $\rightarrow B$

**(2, 1,0)**