



Universidade De Évora
Área Departamental de Ciências Exactas



“Este trabalho não inclui as observações e críticas feitas pelo júri.”

**Dissertação no âmbito do Mestrado em Análises Químicas
Ambientais**

Ana Mafalda Filipe de Matos Pinto

Orientador: Prof. Dr. Henrique Vicente

Aplicação de Sistemas Inteligentes à Optimização de um Sistema de Tratamento de Água

Departamento de Química

Évora

Dissertação no âmbito do Mestrado em Análises Químicas Ambientais
submetido à Universidade de Évora sob a orientação do Professor Doutor
Henrique Agostinho de Oliveira Moiteiro Vicente



171355

“ Este trabalho não inclui as observações e críticas feitas pelo júri.”

2009

Universidade De Évora

Agradecimentos

A realização deste trabalho contou com a colaboração, directa ou indirecta, de um conjunto amplo de familiares e amigos, aos quais gostaria de agradecer especialmente.

O meu mais profundo agradecimento é dirigido ao Dr. António Ventura, Administrador delegado da empresa Águas do Centro Alentejo, pelo apoio e estímulo que me transmitiu ao saber da minha intenção na realização deste trabalho.

Às Engenheiras Isabel Bronze, Susana Santos e Catarina Eusébio pelo carinho e apoio que sempre demonstraram.

À Engenheira Ana Marcão um agradecimento especial por ter acreditado em mim e pela confiança depositada no meu trabalho.

Sem esquecer nunca, quero agradecer aos meus colegas de trabalho, Rui Oliveira, Hugo Vieira, Vasco Castor, José Berrucho, José Coelho, José Machado, Nuno Neves, Bruno Dionísio, Ricardo Valente, Hélder Amante e Maria Helena Troupa pelo apoio sempre demonstrado.

Um agradecimento especial ao Professor Doutor Henrique Agostinho de Oliveira Moiteiro Vicente, pelo apoio, disponibilidade e entusiasmo que demonstrou em orientar a minha tese de Mestrado.

Aos meus avós e irmão “Tó” o meu sincero obrigado por toda a ajuda e carinho demonstrado.

Ao meu namorado João um obrigado sincero pelo apoio, paciência e carinho que sempre demonstrou nos momentos mais complicados.

Aos meus pais, apenas tenho a dizer:

“Obrigado por serem os meus pais”

Resumo

Um dos principais problemas que estação de Tratamento de Água do Monte Novo tem vindo a apresentar é o aparecimento de teores em manganês na água tratada, que por vezes ultrapassam o valor paramétrico estabelecido no Decreto-Lei 306/07, 27 de Agosto ($50 \mu\text{g dm}^{-3}$).

Este trabalho permitiu relacionar resultados de várias determinações analíticas efectuadas no laboratório da empresa Águas do Centro Alentejo e, através deles construir modelos fundamentados em técnicas e Descoberta de Conhecimento em Base de Dados que permitiram responder ao problema identificado.

Foi ainda possível estabelecer a época do ano em que é mais provável o aparecimento de teores elevados manganês na água tratada. Além disso, mostrou-se que a tomada de água desempenha um papel relevante no aparecimento deste metal na água tratada.

Os modelos desenvolvidos permitiram também estabelecer as condições em que é provável o aparecimento de turvação na cisterna de água tratada. Estas estão relacionadas com o pH, o teor em manganês e o teor em ferro.

Foi ainda realçada a importância da correcção do pH na fase final do processo de tratamento. Por um lado, o pH deve ser suficientemente elevado para garantir uma água incrustante e, por outro, deve ser baixo para evitar problemas de turvação na cisterna da água tratada.

Abstract

"Application of Intelligent Systems to the Optimization of a Water Treatment System"

The present study took place in the water treatment plant of Monte Novo. This study aimed for solutions to the problem of high values of manganese concentration in the treated water, in some periods of the year.

The present work reports models for manganese concentration and for turbidity using Knowledge Discovery Techniques in Data Bases.

Índice

AGRADECIMENTOS	III
RESUMO	V
ABSTRACT	VII
ÍNDICE.....	IX
ÍNDICE DE FIGURAS	XIII
ÍNDICE DE TABELAS	XVII
NOTAÇÃO.....	XIX
1 - Acrónimos.....	xix
2 – Símbolos e Abreviaturas.....	xix
INTRODUÇÃO	21
1.1 – A Gestão do Conhecimento	25
1.2 – Inovação e Valor Acrescentado do Trabalho.....	27
1.3 – Estrutura do Trabalho	29
CARACTERIZAÇÃO DO LOCAL DE TRABALHO	31
2.1 – Albufeira do Monte Novo	32
2.1.1 – Caracterização da Qualidade da Água Captada.....	35
2.1.2 – Caracterização da Qualidade da Água para Consumo Humano	38
2.2 – Estação de Tratamento de Águas do Monte Novo.....	42

2.2.1 – Captação	44
2.2.2 – Elevação da água bruta.....	45
2.2.3 – Câmara de Ozonização	45
2.2.4 – Câmara de Contacto com Dióxido de Carbono	46
2.2.5 – Câmara de Mistura Rápida.....	48
2.2.6 – Canal de Distribuição.....	50
2.2.7 – Decantadores ou Flotadores	51
2.2.8 – Câmaras de Desmanganização	53
2.2.9 – Filtros de Areia.....	54
2.2.10 – Cisterna de Contacto com Cloro	55
2.2.11 – Cisternas de Armazenamento	56
2.2.12 – Reservatórios	57
2.2.13 – Fase Sólida	57
DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS – MODELOS E TÉCNICAS	59
3.1 – Descoberta de Conhecimento em Base de Dados.....	60
3.1.1 – Etapas do Processo de DCBD	61
3.2 – Data Mining.....	63
3.3 – Modelos e Técnicas de Data Mining.....	65
3.3.1 – Árvores de Decisão.....	66
3.3.2 – Indução de Regras	69
3.3.3 – Aproximação de Vizinhanças.....	71
3.3.4 – Avaliação dos Modelos.....	71

PROCEDIMENTO EXPERIMENTAL	73
4.1 – Amostragem e Preservação das Amostras.....	74
4.2 – Procedimento Analítico.....	77
4.2.1 – Determinação do pH e da Temperatura.....	78
4.2.2 – Determinação da Condutividade.....	79
4.2.3 – Determinação da Turvação	80
4.2.4 – Oxigénio Dissolvido	81
4.2.5 – Determinação do Cloro Residual e do Cloro Combinado.	82
4.2.6 – Determinação da Alcalinidade	83
4.2.7 – Determinação da Dureza	84
4.2.8 – Determinação da Oxidabilidade	85
4.2.9 – Determinação do Teor em Ferro	86
4.2.10 – Determinação do Teor em Alumínio	87
4.2.11 – Determinação do Teor em Manganês	88
4.3 – Base de Dados.....	90
APRESENTAÇÃO, INTERPRETAÇÃO E ANÁLISE DE RESULTADOS E MODELOS	91
5.1 – Pré-Processamento e Transformação dos Dados	92
5.2 – Modelação	96
5.3 – Apresentação, Interpretação e Avaliação dos Resultados e dos Modelos.....	96
CONCLUSÕES FINAIS E PERSPECTIVAS FUTURAS	113
REFERÊNCIAS BIBLIOGRÁFICAS.....	115

Índice de Figuras

Figura 1: O Ciclo Hidrológico.....	23
Figura 2: Localização da bacia hidrográfica da albufeira de Monte Novo.	33
Figura 3: Usos do solo na área da bacia hidrográfica da albufeira de Monte Novo.....	34
Figura 4: Fontes de poluição pontual na bacia hidrográfica da albufeira de Monte Novo.....	34
Figura 5: Anexo I do Decreto-Lei nº236/98, de 1 de Agosto.	36
Figura 6: Anexo IV do Decreto-Lei nº236/98, de 1 de Agosto.	37
Figura 7: Anexo V Decreto-Lei nº236/98, de 1 de Agosto.	37
Figura 8: Anexo I do Decreto-Lei 306/07, de 27 de Agosto, referente a parâmetros microbiológicos.	38
Figura 9: Anexo I do Decreto-Lei 306/07, de 27 de Agosto, referente a parâmetros químicos.	39
Figura 10: Anexo I do Decreto-Lei 306/07, de 27 de Agosto, referente a parâmetros indicadores.	40
Figura 11: Parâmetros a determinar no controlo de rotina 1 no controlo de rotina 2 e no controlo de inspecção.	41
Figura 12: Anexo II do Decreto-Lei 306/07, de 27 de Agosto.....	41
Figura 13: Diagrama linear da ETA do Monte Novo.....	43
Figura 14: Vista da torre de captação (a) e interior da mesma (b).....	44
Figura 15: Conduta da elevação da água bruta.....	45
Figura 16: Câmara de ozonização.....	46
Figura 17: Câmara de contacto com o dióxido de carbono.	47
Figura 18: Câmara de mistura rápida.....	48
Figura 19: Linhas de tratamento da câmara de mistura rápida.....	49
Figura 20: Localização das várias picagens de reagentes e dos agitado- res da câmara de mistura rápida.	50
Figura 21: Canal de distribuição.	51
Figura 22: Comportas direccionais no canal de distribuição.	51

Figura 23: Decantadores.	52
Figura 24: Flotador	52
Figura 25: Representação das linhas de tratamento na ETA do Monte Novo.	54
Figura 26: Filtros de areia.	55
Figura 27: Cisterna de contacto com o cloro.	56
Figura 28: Estações elevatórias existentes na ETA.	56
Figura 29: Sistemas de tratamento das lamas.	57
Figura 30: Filtro de banda usado na desidratação de lamas.	58
Figura 31: Processo de DCBD.	61
Figura 32: Algoritmo de indução de árvores de decisão.	68
Figura 33: Validação cruzada com k iterações.	72
Figura 34: Folha de registo que acompanha as amostras.	76
Figura 35: Curva de calibração para a determinação do teor em ferro.	87
Figura 36: Curva de calibração para a determinação do teor em alumínio.	88
Figura 37: Curva de calibração para determinação do teor em manganês.	89
Figura 38: Excerto da base de dados usada.	90
Figura 39: Evolução do teor em manganês na água captada.	93
Figura 40: Distribuição do número de casos por classe de qualidade para o manganês na água tratada.	94
Figura 41: Variação do teor em manganês na água captada e na água tratada ao longo do tempo.	97
Figura 42: Variação do teor em manganês na água tratada após a entrada em funcionamento das câmaras de desmanganização.	99
Figura 43: Variação dos teores em manganês e em oxigénio dissolvido, ao longo do tempo na água captada.	99
Figura 44: Variação dos teores em ferro e em oxigénio dissolvido, ao longo do tempo na água captada.	100
Figura 45: Variação da temperatura, do pH e do oxigénio dissolvido na água captada, ao longo do tempo.	100

Figura 46: Variação do teor em manganês e da turvação na água captada, ao longo do tempo.....	102
Figura 47: Variação do teor em ferro e da turvação na água captada, ao longo do tempo.	103
Figura 48: Variação do pH, da turvação e da temperatura na água captada, ao longo do tempo.....	103
Figura 49: Segmentos identificados no estudo do comportamento da turvação presente na água tratada.	105
Figura 50: Árvore de decisão obtida para o modelo 1.	107
Figura 51: Árvore de decisão obtida para o modelo 2.	109
Figura 52: Árvore de decisão obtida para o modelo 3.	111

Índice de Tabelas

Tabela 1: Características fundamentais do complexo do Monte Novo. .	32
Tabela 2: Caracterização dos objectivos do processo Data Mining	64
Tabela 3: Tarefas e técnicas usadas no processo de <i>Data Mining</i>	66
Tabela 4: Recipientes utilizados e volumes de amostras recolhidos para a determinação dos vários parâmetros.....	75
Tabela 5: Técnicas analíticas utilizadas no laboratório de controlo da ETA do Monte Novo.	77
Tabela 6: Matriz de coincidência do modelo 1.	106
Tabela 7: Matriz de coincidência do modelo 2.	108
Tabela 8: Matriz de coincidência do modelo 3.	111

Notação

1 - Acrónimos

ADCA – Águas do Centro Alentejo

BD – Base de Dados

CMR – Câmara de Mistura Rápida

DCBD – Descoberta de Conhecimento em Base de Dados

DM – *Data Mining*

ETA – Estação de Tratamento de Água

NPA – Nível de Pleno Armazenamento

NTU – Unidade Nefelométrica de Turvação

OMS – Organização Mundial de Saúde

SGBD - Sistemas de Gestão de Bases de Dados

SMEWW – Standard Methods for Examination of Water & Wastewater

2 – Símbolos e Abreviaturas

Al1 – Teor em alumínio na água captada

Al9 – Teor em alumínio na água para consumo humano (cisterna)

C – Cota da albufeira

Fe9 – Teor em ferro na água para consumo humano (cisterna)

Mn1 – Teor em manganês na água captada

Mn9 – Teor em manganês na água para consumo humano (cisterna)

NP – Não permitido

OD – Oxigénio Dissolvido

OD1 – Oxigénio Dissolvido para a água captada

P - Permitido

TeMax – Temperatura máxima do ar

TeMin – Temperatura mínima do ar

Tu9 – Turvação na água para consumo humano (cisterna)

V – Volume da albufeira

Capítulo 1

Introdução

Neste capítulo apresentam-se os principais objectivos do trabalho e descreve-se também o modo como este está estruturado.

A água é um recurso essencial à vida uma vez que todos os seres vivos necessitam dela para viver. A vida nasceu no chamado oceano primordial, ou seja na água. As diversas espécies foram-se adaptando ao meio aquático tendo algumas delas evoluídas de tal forma que acabaram por se adaptar à vida terrestre e à vida aérea.

A água assume ainda um papel relevante na distribuição das populações, nas comunicações e transportes, nos equipamentos e maquinarias, nas construções, na produção e no fornecimento de energia, no tratamento e prevenção de doenças, nos diferentes climas que proporciona, na modelação da superfície terrestre, na agricultura e na alimentação dos seres vivos. Está ainda na base de inúmeras

actividades profissionais, culturais e de recreio. Proporciona muitas paisagens diferentes, quer naturais quer resultantes do engenho humano. Inspirou poetas e artistas e foi tema de acontecimentos como, por exemplo, a EXPO/98 e, mais recentemente, a EXPO/Saragoça.

Como foi referido anteriormente, a água tem sido um dos principais factores de desenvolvimento e de fixação das comunidades humanas ao longo da História. Extremamente dependentes da disponibilidade hídrica, as civilizações foram-se fixando 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 [Serralheiro, 2001] [Lehr & Keeley, 2005b]. A explosão demográfica e o crescente nível de desenvolvimento tecnológico, ao apelarem a uma maior utilização da água, contribuíram para a sua crescente degradação. 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 [Lehr & Keeley, 2005a]. 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, conseqüentemente, dos ecossistemas aquáticos. A crescente intervenção humana nas linhas de água tem originado alterações morfológicas. Estas alterações diminuem a diversidade dos ecossistemas aquáticos e alteram as características ecológicas repercutindo-se nos processos naturais de auto-depuração [Mendes & Oliveira, 2004].

A água existente no planeta Terra não constitui uma realidade estática. Tal como outros elementos essenciais à vida estão integrados em ciclos biogeoquímicos, mantendo-se em circulação permanente. Estabelece-se, assim, um sistema fechado, no qual a quantidade total se mantém praticamente constante e que se denomina ciclo hidrológico, Figura 1 [Mendes & Oliveira, 2004] [URL1].

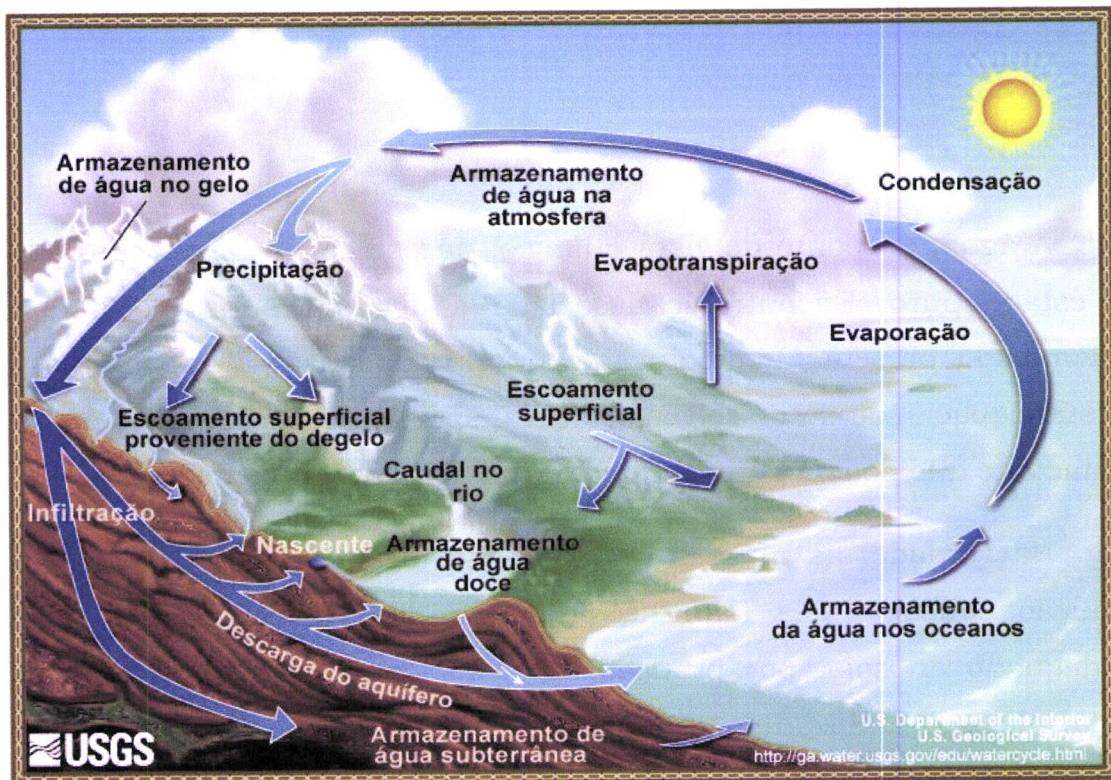


Figura 1: O Ciclo Hidrológico Adaptado de [URL1].

No mecanismo do ciclo hidrológico podem ser identificados os seguintes fenômenos:

- 1 – Transferência da água, na forma de vapor, da superfície para a atmosfera, por evaporação a partir dos estados líquido, sólido e por evapotranspiração;
- 2 – Transporte de água, entre as diversas fases, através de diferentes processos de circulação locais/regionais e pela circulação global;
- 3 – Condensação parcial do vapor de água presente na atmosfera sob a forma líquida e/ou sólida, ficando em suspensão na atmosfera, formando aerossóis (i.e. nuvens e nevoeiros);
- 4 – Transferência de água da atmosfera para a superfície, por precipitação, quer no estado líquido quer no estado sólido.

5 – Escoamento superficial.

6 – Infiltração no solo.

A água, do ponto de vista químico, não se encontra na natureza no estado puro. Encontra-se associada a diversas substâncias que podem estar presentes em solução e/ou em suspensão e que resultam da sua capacidade de dissolver e/ou transportar os materiais com que contacta, e que podem vir a condicionar os seus possíveis usos.

Na verdade, as massas de água encontram-se ameaçadas pela poluição causada pela falta de protecção das nascentes, pelo uso inadequado do solo, por práticas agrícolas com recurso excessivo a fitofármacos e fertilizantes, por actividades pecuárias que recorrem ao uso excessivo de fármacos e pela falta de investimento em tratamento de efluentes [Lehr & Keeley, 2005a].

O tratamento de água tem como principal objectivo a eliminação de impurezas prejudiciais e nocivas à saúde. Quanto mais poluída estiver a água, mais complexo será o processo de tratamento desta e, portanto, mais cara será.

A pressão colocada sobre as reservas de água, quer sejam reservas subterrâneas quer sejam reservas superficiais, condiciona o aproveitamento que se pode fazer deste recurso natural. Deste modo os problemas subjacentes à temática da modelação da qualidade da água situam-se no âmbito das preocupações e necessidades de natureza ambiental que se tem vindo a afirmar como das mais relevantes para um desenvolvimento sustentado e harmonioso das populações.

Portugal não é excepção, constituindo as estações de tratamento de água um caso em que a resolução e a prevenção de problemas relacionados com a qualidade da água se tem posto e continuará a por com grande acuidade. É por conseguinte necessário prevenir e actuar de forma pró-activa, pelo que o recurso a métodos de previsão que

possam ser integrados em sistemas de apoio à tomada de decisão, pode contribuir para solucionar ou averiguar o problema [Neves et al., 1998] [Vicente, 2004] [Santos et al., 2005].

Há assim que recorrer a outros mecanismos computacionais, oriundos da Área Científica da Inteligência Artificial e a Sistemas Inteligentes [Dzeroski, 2002] [Chau, 2006], onde se destaca a utilização de Redes Neurais Artificiais (RNAs) [Lek & Guégan, 1999] [Karul et al., 2000] [Jeong et al., 2001] [Scardi, 2001] [Wilson & Recknagel, 2001] [Lee et al., 2003] [Recknagel, 2003] [Kuo et al., 2007] [Tyagi et al., 2008] [Yang et al., 2008], Algoritmos Genéticos (AG) [Bobbin & Recknagel, 2001] [Jeong et al., 2003] [Chen, 2003] [Liu et al., 2007], Árvores de Decisão (AD) [Atkins et al., 2007], Indução de Regras (IR) [Santos et al., 2005] e, mais recentemente, combinações destas técnicas [Kuo et al., 2006] [Chaves & Kojiri, 2007].

1.1 – 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 [Abelha, 2004] [Vicente, 2004].

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, para 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 extracção 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 [Abelha, 2004] [Vicente, 2004].

1.2 – 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 da Água, porém algumas das suas características estão finalmente a emergir. Tais ferramentas devem dar resposta a questões deveras complexas 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 decisor e explicar-lhe as suas opções, 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 decisores 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

decisor a posicionar-se na sua intersecção, não tendo contudo que prestar atenção à tecnologia que está a ser utilizada.

Neste trabalho aplicam-se ferramentas e técnicas oriundas da Área Científica da Inteligência Artificial ao desenvolvimento de modelos de previsão das alterações na qualidade da água tratada. Com este estudo pretende-se dar uma contribuição para o esforço que há a fazer ao nível do funcionamento da ETA, procurando dar resposta rápida a quaisquer problemas que eventualmente possam ocorrer durante o processo de tratamento da água. Há, no entanto, a noção que é um trabalho sempre inacabado e que deverá ter continuação futura a fim de permitir prever as situações que possam vir a ocorrer e estabelecer os melhores procedimentos tendo em vista os elevados padrões de qualidade exigidos à saída da ETA.

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

- 1 – Descrever os problemas e fenómenos associados à água captada que possam influenciar a qualidade da água tratada;
- 2 – Determinar formas de actuar quando existem perturbações da água ao nível da origem e que possam influenciar a qualidade desta no final do processo de tratamento;
- 3 – Compreensão e resolução de problemas que possam ocorrer ao nível do armazenamento da água numa cisterna;
- 4 – Conhecer os eventuais problemas que possam ocorrer dentro do processo de tratamento e que num trabalho futuro possam ser usados para resolução e entendimento destas dificuldades;
- 5 – Realçar a capacidade dos modelos na descrição do estado da qualidade da água em tempo real e de fazerem previsões sobre cenários futuros;

1.3 – Estrutura do Trabalho

A dissertação encontra-se estruturada e organizada em seis capítulos, não se contabilizando as referências bibliográficas.

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 descreve-se e a caracteriza-se o local de trabalho. O referido capítulo encontra-se dividido em duas partes. Na primeira, descreve-se e caracteriza-se a albufeira do Monte Novo e a sua bacia hidrográfica. A segunda parte é composta pela descrição e pela caracterização tanto da estação de tratamento de água do Monte Novo, como das várias etapas do processo de tratamento da água.

O terceiro capítulo aborda a problemática da Descoberta de Conhecimento em Bases de Dados, apresentando-se modelos e técnicas de *Data Mining*, destacando os objectivos e o tipo de abordagens.

Quanto ao quarto capítulo, numa primeira parte abordam-se os procedimentos analíticos que incluem a recolha e a preservação das amostras e as determinações no laboratório dos vários parâmetros de qualidade da água. Na segunda parte descreve-se a construção da base de dados que irá ser utilizada para o desenvolvimento dos modelos.

No quinto capítulo fundamentam-se as opções tomadas durante o Processo de Descoberta de Conhecimento em Base de Dados e apresentam-se, discutem-se e avaliam-se quer os resultados quer os modelos construídos.

Por fim, no sexto e último capítulo, apresentam-se as conclusões finais e perspectiva-se o trabalho que se pretende desenvolver no futuro.

Capítulo 2

Caracterização do Local de Trabalho

Neste capítulo é descrito e caracterizado o local de trabalho. Este engloba a albufeira e a Estação de Tratamento de Água do Monte Novo. No capítulo também se aborda tanto a qualidade da água captada como a qualidade da água para consumo.

O presente estudo recai sobre o complexo do Monte Novo, constituído pela albufeira e respectiva bacia hidrográfica. Em seguida apresenta-se primeiro uma breve caracterização da albufeira do Monte Novo e, posteriormente, a caracterização da ETA.

2.1 – Albufeira do Monte Novo

A albufeira do Monte Novo localiza-se no concelho de Évora (a sudoeste desta cidade) entre Évora e Reguengos de Monsaraz, a norte da estrada nacional 256 e IP2. Esta albufeira é um aproveitamento do rio Degebe, que é um efluente do rio Guadiana. Iniciou o seu funcionamento em 1982 e destina-se à produção de água para consumo humano, servindo actualmente cerca de 70.000 habitantes dos municípios de Évora, de Reguengos de Monsaraz, parte do município de Mourão e, caso seja necessário pode abastecer o município de Redondo. Na Tabela 1 encontram-se as características fundamentais da albufeira do Monte Novo [Matoso et al., 2004].

Tabela 1: Características fundamentais do complexo do Monte Novo.

Características da Albufeira	Características da Barragem
Dimensão de ocupação – 267 km ²	Comprimento de coroamento – 160 m
Nível de Máxima Cheia – 196 m	Largura de coroamento – 30 m
Área inundada ao NPA – 2,77 km ²	Cota do coroamento – 226,25 m
Capacidade total – 15,28 × 10 ⁶ m ³	Altura acima da fundação – 18 m
Capacidade útil – 14,78 × 10 ⁶ m ³	Fundação – arenitos, quartzodioritos, xistos e grauvaques.

A bacia hidrográfica localiza-se nos concelhos de Évora e Arraiolos, Figura 2. Situa-se numa área de relevo ondulado a plano, com declives predominantemente suaves a moderados, encontrando-se algumas excepções pontuais onde os declives vão de acentuados a muito acentuados [Matoso et al., 2004].

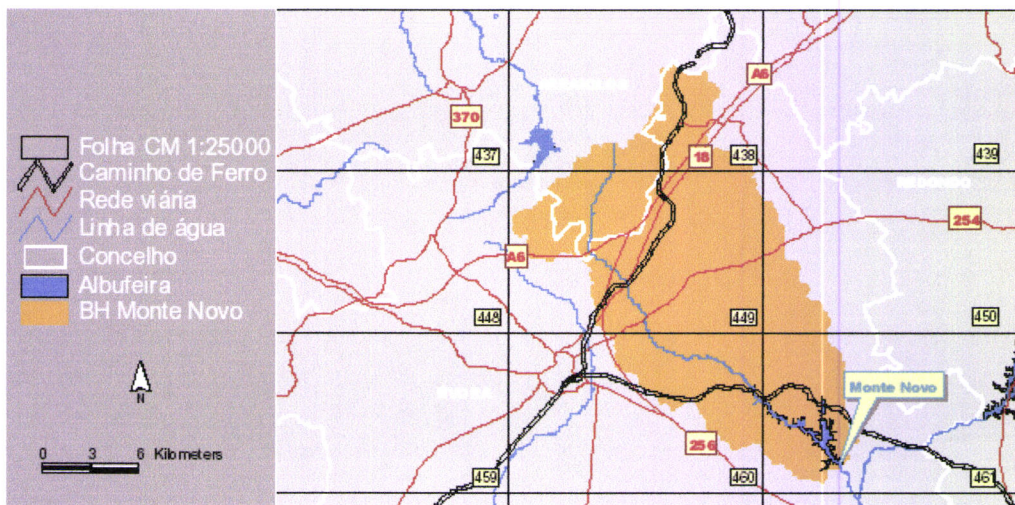


Figura 2: Localização da bacia hidrográfica da albufeira de Monte Novo.
Adaptada de Matoso [Matoso, et al., 2004].

O clima é mediterrânico, com forte influência continental, 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). O período chuvoso regista precipitações médias anuais na ordem de 650 a 700 mm. Em anos de seca, contudo, pode não ultrapassar os 50 mm [Cunha et al., 1980].

A unidade geológica fundamental da área em estudo é a zona de Ossa Morena. Do ponto de vista litológico, as formações predominantes são xistos, grauvaques, arenitos, calcários e quartzodioritos.

Na área da bacia hidrográfica, como se pode observar na Figura 3, existem poucas zonas urbanas e praticam-se, sobretudo, as culturas arvenses de sequeiro, embora existindo também explorações agro-silvo-pastoris, culturas arbóreas-arbustivas, arvenses de regadio e pequenas manchas de eucaliptais [Matoso et al., 2004].

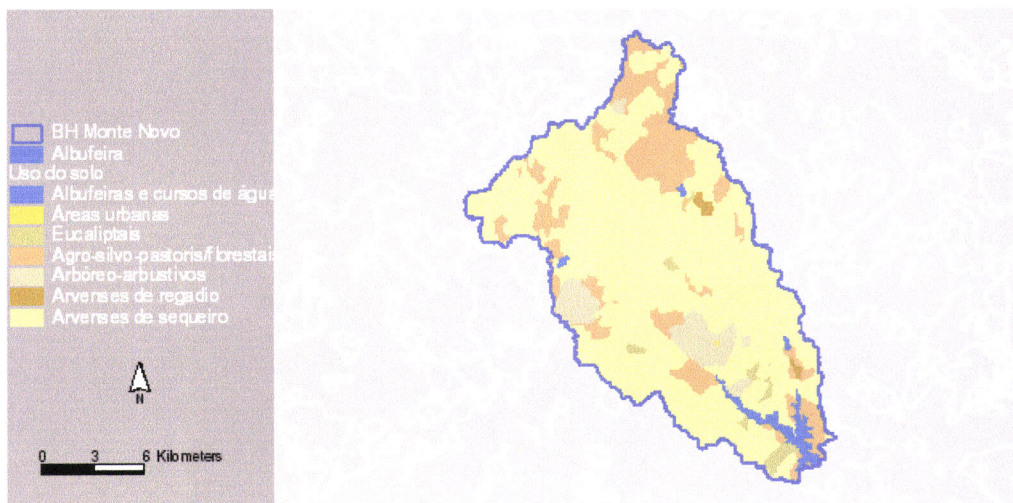


Figura 3: Usos do solo na área da bacia hidrográfica da albufeira de Monte Novo. Adaptado de Matoso [Matoso et al., 2004].

As principais fontes de poluição pontual, Figura 4, resultam de bovinicultura, da suinicultura, das áreas urbanas, de um lagar de azeite e da indústria de curtumes [DRAOTA, 1999] [Matoso et al., 2004].

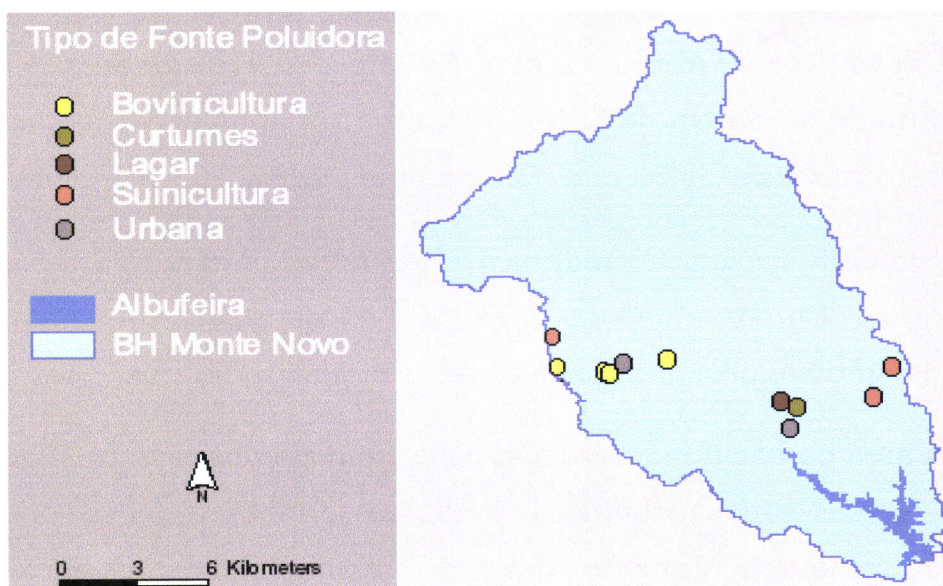


Figura 4: Fontes de poluição pontual na bacia hidrográfica da albufeira de Monte Novo. Adaptado de Matoso [Matoso et al., 2004].

As fontes de poluição pontual, apresentadas na Figura 4, 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 as lançar na natureza [DRAOTA, 1999].

As fontes poluidoras da bacia hidrográfica da albufeira do Monte Novo produzem uma carga diária de $502,9 \text{ kg}_{\text{CBO}}, \text{ dia}^{-1}$, corresponde a 8382 equivalentes de população. Contudo, em virtude de existirem sistemas de tratamento, a carga rejeitada, que entra nas linhas de água, é estimada em $23,4 \text{ kg}_{\text{CBO}}, \text{ dia}^{-1}$, ou seja, correspondente a 390 equivalentes de população [DRAOTA, 1999].

2.1.1 – Caracterização da Qualidade da Água Captada

A qualidade da água da albufeira pode ser avaliada de acordo com o critério de classificação das águas brutas destinadas à produção de água para consumo humano. De acordo com esta metodologia, a qualidade da massa de água obtêm-se parâmetro a parâmetro, de acordo com os intervalos estipulados no anexo I do Decreto-Lei n° 236/98, de 1 de Agosto (Figura 5). É atribuída à massa de água a classificação do parâmetro mais desfavorável.

Neste método, 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).

ANEXO I

Qualidade das águas doces superficiais destinadas à produção de água para consumo humano

Parâmetros	Expressão dos resultados	A1		A2		A3	
		VMR	VMA	VMR	VMA	VMR	VMA
pH, 25°C	Escala de Sorensen	6.5-8.5	-	5.5-9.0	-	5.5-9.0	-
Cor (após filtração simples)	mg/l, escala Pt-Co	10	(O) 20	50	(O) 100	50	(O) 200
Sólidos suspensos totais	mg/l	25	-	-	-	-	-
Temperatura	°C	22	(O) 25	22	(O) 25	22	(O) 25
Condutividade	µS/cm, 20°C	1 000	-	1 000	-	1 000	-
Cheiro	Factor de diluição, a 25°C	3	-	10	-	20	-
Nitratos (*)	mg/l NO ₃	25	(O) 50	-	(O) 50	-	(O) 50
Fluoretos (1)	mg/l F	0.7-1.0	1.5	0.7-1.7	-	0.7-1.7	-
Cloro orgânico total extraível	mg/l Cl	-	-	-	-	-	-
Ferro dissolvido (*)	mg/l Fe	0.1	0.3	1.0	2.0	1.0	-
Manganês (*)	mg/l Mn	0.05	-	0.10	-	1.00	-
Cobre	mg/l Cu	0.02	(O) 0.05	0.05	-	1.00	-
Zinco	mg/l Zn	0.5	3.0	1.0	5.0	1.0	5.0
Boro	mg/l B	1.0	-	1.0	-	1.0	-
Berílio	mg/l Be	-	-	-	-	-	-
Cobalto	mg/l Co	-	-	-	-	-	-
Níquel	mg/l Ni	-	-	-	-	-	-
Vanádio	mg/l V	-	-	-	-	-	-
Arsénio	mg/l As	0.01	0.05	-	0.05	0.05	0.10
Cádmio	mg/l Cd	0.001	0.005	0.001	0.005	0.001	0.005
Crómio total	mg/l Cr	-	0.05	-	0.05	-	0.05
Chumbo	mg/l Pb	-	0.05	-	0.05	-	0.05
Selénio	mg/l Se	-	0.01	-	0.01	-	0.01
Mercurio	mg/l Hg	0.0005	0.0010	0.0005	0.0010	0.0005	0.0010
Bário	mg/l Ba	-	0.1	-	1.0	-	1.0
Cianetos	mg/l CN	-	0.05	-	0.05	-	0.05
Sulfatos	mg/l SO ₄	150	250	150	(O) 250	150	(O) 250
Cloretos	mg/l Cl	200	-	200	-	200	-
Substâncias tensoactivas (que reagem com o azul-de-metileno)	mg/l, sulfato de lauril e sódio	0.2	-	0.2	-	0.5	-
Fosfatos (*) (2)	mg/l P ₂ O ₅	0.4	-	0.7	-	0.7	-
Fenóis	mg/l C ₆ H ₅ OH	-	0.001	0.001	0.005	0.010	0.100
Hidrocarbonetos dissolvidos ou emulsionados	mg/l	-	0.05	-	0.20	0.50	1.00
Hidrocarbonetos aromáticos polinucleares	µg/l	-	0.2	-	0.2	-	1.0
Pesticidas totais (paratíão, hexaclorociclo-hexano, dieldrina e outros)	µg/l	-	1.0	-	2.5	-	5.0
Carência química de oxigénio (CQO) (*)	mg/l O ₂	-	-	-	-	30	-
Oxigénio dissolvido (*) (3)	‰ saturação de O ₂	70	-	50	-	30	-
Carência bioquímica de oxigénio a (CBO ₅ , 20°C) (*)	mg/l O ₂	3	-	5	-	7	-
Azoto Kjeldahl (excluindo o azoto de NO ₂ e NO ₃)	mg/l N	1	-	2	-	3	-
Azoto amoniacal	mg/l NH ₄	0.05	-	1.00	1.50	2.00	(O) 4.00
Substâncias extraíveis com clorofórmio	mg/l	0.1	-	0.2	-	0.5	-
Carbono orgânico total (COT)	mg/l C	-	-	-	-	-	-
Carbono orgânico residual após floculação e filtração através de membrana (5µm)	mg/l C	-	-	-	-	-	-
Coliformes totais	/100 ml	50	-	5 000	-	50 000	-
Coliformes fecais	/100 ml	20	-	2 000	-	20 000	-
Estreptococos fecais	/100 ml	20	-	1 000	-	10 000	-
Salmonelas		Ausência em 5 000 ml	-	Ausência em 1 000 ml	-	-	-

(O) Os limites podem ser excedidos em caso de condições geográficas ou meteorológicas excepcionais (n.º 1 do artigo 10.º).
 (*) Os limites podem ser excedidos para os parâmetros marcados com * em lagos de pouca profundidade e baixa taxa de renovação.
 (1) Os valores indicados constituem os limites inferior e superior das concentrações, determinados em função da média anual das temperaturas máximas diárias.
 (2) Este parâmetro é incluído para satisfazer as exigências ecológicas de certos meios.
 (3) Refere-se a um VMR.

VMR — valor máximo recomendado.
 VMA — valor máximo admissível.

Figura 5: Anexo I do Decreto-Lei nº236/98, de 1 de Agosto.

A frequência de amostragem e de análise das águas superficiais está regulamentada no Anexo IV do Decreto-Lei nº236/98, de 1 de Agosto, sendo função da classe a que a água pertence e do parâmetro em causa (Figura 6).

ANEXO IV

Frequência mínima de amostragem e de análise de águas superficiais (*)

	Classe de água								
	A1			A2			A3		
	Grupo de parâmetros								
	G1	G2	G3	G1	G2	G3	G1	G2	G3
Frequência mínima (numero/ano)	4	2	1	8	4	2	12	6	3

(*) A determinação dos parâmetros correspondentes ao grupo G2 implica, em simultâneo, a determinação dos parâmetros contidos em G1 e, idênticamente para G3, implica G2 e G1.

Figura 6: Anexo IV do Decreto-Lei nº236/98, de 1 de Agosto.

Os parâmetros de qualidade das águas doces superficiais destinadas à produção de água para consumo humano dividem-se em três grupos, G1, G2 e G3. Os parâmetros considerados em cada um dos grupos encontram-se no anexo V do referido Decreto-Lei, Figura 7.

ANEXO V

Classificação dos parâmetros de qualidade de águas superficiais em grupos (G1, G2 e G3) segundo a frequência de amostragem e de análise

G1	G2	G3
<p>pH Cor Sólidos suspensos totais Temperatura Condutividade eléctrica Cheiro Nitratos Cloratos Fosfatos Carência química de oxigénio (CQO) Oxigénio dissolvido Carência bioquímica de oxigénio (CBO₅) Azoto amoniacal Coliformes totais Coliformes fecais</p>	<p>Ferro dissolvido Manganés Cobre Zinco Sulfatos Substâncias tensoactivas Fenóis Azoto Kjeldahl Estreptococos fecais</p>	<p>Fluoretos Boro Arsénio Cádmio Crómio total Chumbo Selénio Mercúrio Bário Cianetos Hidrocarbonetos dissolvidos e emulsionados Hidrocarbonetos aromáticos polinucleares Pesticidas totais Substâncias extraíveis com clorofórmio Salmonelas</p>

Figura 7: Anexo V Decreto-Lei nº236/98, de 1 de Agosto.

A água da albufeira do Monte Novo pertence à classe A3, i.e. uma água bastante poluída que obriga para além do tratamento físico, a um tratamento químico de afinação e desinfecção.

2.1.2 – Caracterização da Qualidade da Água para Consumo Humano

A qualidade das águas destinadas ao consumo humano obtém-se parâmetro a parâmetro, de acordo com os valores estipulados no anexo I do Decreto-Lei n.º 306/07, de 27 de Agosto. Os referidos parâmetros estão divididos em três grupos, i.e. microbiológicos (Figura 8), químicos (Figura 9) e indicadores (Figura 10).

ANEXO I

(a que se refere o n.º 1 do artigo 6.º)

Parâmetros e valores paramétricos

Parte I — Parâmetros microbiológicos

Parâmetro	Valor paramétrico	Unidade
<i>Escherichia coli</i> (<i>E. coli</i>)	0	Número/100 ml.
Enterococos	0	Número/100 ml.
<i>Escherichia coli</i> (<i>E. coli</i>)	0	Número/150 ml.
Enterococos	0	Número/150 ml.
<i>Pseudomonas aeruginosa</i>	0	Número/150 ml.
Número de colónias a 22°C	100	Número/ml.
Número de colónias a 37°C	20	Número/ml.

Figura 8: Anexo I do Decreto-Lei 306/07, de 27 de Agosto, referente a parâmetros microbiológicos.

ANEXO I
Parâmetros e valores paramétricos
Parte II — Parâmetros químicos

Parâmetro	Valor paramétrico	Unidade	Observações
Acetilaminas	0,10	µg/l	V. n. 1.
Antimônio	5,0	µg/l Sb	
Arsênio	10	µg/l Ar	
Benzeno	1,0	µg/l	
Benzo(a)pireno	0,010	µg/l	
Boro	1,0	mg/l B	
Bromatos	25 (de 25 de Dezembro de 2003 até 25 de Dezembro de 2008). 10 (após 25 de Dezembro de 2008).	µg/l BrO ₃	V. n. 2.
Cádmio	5,0	µg/l Cd	V. n. 3.
Crômio	50	µg/l Cr	V. n. 3.
Cobre	1,0	mg/l Cu	V. n. 3.
Cianetos	50	µg/l Cy	
1,2 dicloroetano	3,0	µg/l	
Epicletridina	0,10	µg/l	V. n. 1.
Fluoretos	1,5	mg/l F	
Chumbo	25 (de 25 de Dezembro de 2003 até 25 de Dezembro de 2013). 10 (após 25 de Dezembro de 2013).	µg/l Pb	V. n. 3 e 4.
Mercurio	1	µg/l Hg	
Níquel	20	µg/l Ni	V. n. 3.
Nitratos	50	mg/l NO ₃	V. n. 5.
Nitritos	0,5	mg/l NO ₂	V. n. 5.
Pesticida individual	0,10	µg/l	V. n. 6 e 7.
Pesticida — total	0,50	µg/l	V. n. 6 e 8.
Hidrocarbonetos aromáticos policíclicos (HAP)	0,10	µg/l	Soma das concentrações dos compostos especificados. V. n. 9.
Selênio	10	µg/l Se	
Tetracloroetano e tetracloreto	10	µg/l	Soma das concentrações dos compostos especificados.
Trihalometanos — total (THM)	150 (de 25 de Dezembro de 2003 até 25 de Dezembro de 2008). 100 (após 25 de Dezembro de 2008).	µg/l	Soma das concentrações dos compostos especificados. V. n. 10.
Cloro de vanilo	0,50	µg/l	V. n. 1.

Figura 9: Anexo I do Decreto-Lei 306/07, de 27 de Agosto, referente a parâmetros químicos.

ANEXO I
Parâmetros e valores paramétricos
Parte III — Parâmetros indicadores

Parâmetro	Valor paramétrico	Unidade	Observações
Alumínio	200	µg/Ld'	
Amónio	0,50	mg/L NH ₃	
Cálcio	-	mg/l Ca	V. n. 1, 3 e 5.
Cloratos	250	mg/l Cl	V. n. 1.
<i>Clostridium parfringens</i> (incluindo esporos)	0	N/100 ml	V. n. 6.
Cob	20	mg/l PbCo	
Condutividade	2 500	µS/cm a 20°C	V. n. 1.
Dureza total	-	mg/l CaCO ₃	V. n. 1, 4 e 5
pH	≥ 6,5 e ≤ 9	unidades de pH	V. n. 1 e 7.
Ferro	300	µg/l Fe	
Magnésio	-	mg/l Mg	V. n. 1, 3 e 5
Manganes	50	µg/l Mn	
Microcistinas — LR total	1	µg/l	V. n. 16.
Chenro. a 25°C	3	Fator de diluição	
Oxalibilidade	5	mg/l O	V. n. 8.
Sulfatos	250	mg/l SO ₄	V. n. 1.
Sódio	200	mg/l Na	
Sabor, a 25°C	3	Fator de diluição	
Número de colónias	Sem alteração anormal	N/ml a 22°C	V. n. 17 e 18.
Número de colónias	Sem alteração anormal	N/ml a 37°C	V. n. 17 e 18.
Bactérias coliformes	0	N/100 ml	V. n. 9.
Carbono orgânico total (COT)	Sem alteração anormal	mg/l C	V. n. 10 e 18.
Turbacão	4	UNT	V. n. 11.
α-total	0,5	Bq/l	V. n. 12 e 14.
β-total	1	Bq/l	V. n. 12 e 14.
Trítio	100	Bq/l	V. n. 12 e 14.
Dose indicativa total	0,10	mSv/ano	V. n. 13, 14 e 15.
Derivação residual	-	mg/l	V. n. 19.

Figura 10: Anexo I do Decreto-Lei 306/07, de 27 de Agosto, referente a parâmetros indicadores.

Os parâmetros atrás referidos encontram-se divididos em três grupos que se denominam por controlo de rotina 1, controlo de rotina 2 e controlo de inspecção.

O controlo de rotina tem visa fornecer regularmente informações sobre a qualidade organoléptica e microbiológica da água destinada ao consumo humano, bem como aferir e regular a eficácia dos tratamentos existentes, especialmente a desinfecção, tendo em vista determinar a conformidade da água com os valores paramétricos estabelecidos pelo Decreto-Lei 306/07 27 de Agosto.

O controlo de inspecção tem como objectivo obter as informações necessárias para verificar o cumprimento dos valores paramétricos estabelecidos pelo referido Decreto-Lei. A Figura 11 mostra quais o parâmetros que pertencem a cada um dos grupos.

Tipo de controlo (v. n. 1)	Faizimato	Tipo de controlo (v. n. 1)	Faizimato
Controlo de rotina 1	<i>Escherichia coli</i> (<i>E. coli</i>) Bactérias coliformas Decontaminante residual	Controlo de inspeção	Antimônio Arsénio Benzeno Benzol(a)pireno Boro Bromatos Cádmio Cálcio Cromo Cianatos Cobre Cromo 1,2-dicloroetano Dureza total Enterococos Fluoretos Magnésio Mercúrio Níquel HAP Pesticidas individuais Pesticidas (total) Selénio Cianetos Tetracloroetano e tricloreto Trihalometanos Sódio Carbono orgânico total Sulfatos Clorato de sódio Epiclondrina Acilaurida
Controlo de rotina 2	Alumínio Amónio Número de colónias a 21°C Número de colónias a 37°C Conductividade <i>Clostridium perfringens</i> , incluindo esporos Cérebro pH Ferro Manganésio Nítritos Nítratos Oxidabilidade Chumbo Saber Turvação		

Figura 11: Parâmetros a determinar no controlo de rotina 1 no controlo de rotina 2 e no controlo de inspeção.

A frequência mínima anual de amostragem e de análise das águas para consumo humano está regulamentada no Anexo II do Decreto-Lei nº 306/2007, de 27 de Agosto, Figura 12.

ANEXO II

(a que se refere o n.º 1 do artigo 10.º)

Controlo da qualidade da água

QUADRO B2

Frequência mínima de amostragem e de análise da água destinada para consumo humano fornecida por uma entidade gestora em alta

Volume de água fornecida em alta (metros cúbicos por dia)	Tipos de controlo da água (para consumo humano)			Volume de água fornecida em alta (metros cúbicos por dia)	Tipos de controlo da água (para consumo humano)		
	Controlo de rotina 1	Controlo de rotina 2 (v. n. 1)	Controlo de inspeção (v. n. 1)		Controlo de rotina 1	Controlo de rotina 2 (v. n. 1)	Controlo de inspeção (v. n. 1)
≤ 250	2	2	1	> 2000 e ≤ 5000	18	6	1
> 250 e ≤ 1000	6	2	1	> 5000 e ≤ 15000	24	8	2
> 1000 e ≤ 2000	12	4	1	> 15000 e ≤ 25000	72	24	4
				> 25000 e ≤ 50000	104	36	4
				> 50000 e ≤ 100000	156	52	6
				> 100000	365	104	12

Figura 12: Anexo II do Decreto-Lei 306/07, de 27 de Agosto.

2.2 – Estação de Tratamento de Águas do Monte Novo

A ETA do Monte Novo tem como finalidade a produção de água para consumo humano. Como foi referido anteriormente, a água da albufeira do Monte novo pertence à classe A3, ou seja uma água bastante poluída que obriga para além do tratamento físico, a um tratamento químico de afinação e desinfecção.

O sistema de tratamento de águas aplicado na ETA do Monte Novo é bastante complexo e envolve diversas etapas até à distribuição para a rede pública, e que passam, como referido anteriormente, por processos de tratamento físico, químico de afinação e de desinfecção, cujo diagrama linear se apresenta na figura 13 e que se passam a descrever.

A primeira etapa do sistema de tratamento de águas consiste na captação e elevação da água bruta até à ETA do Monte Novo. A fase de tratamento propriamente dita inclui a pré-oxidação da água na câmara de ozonização, na sua remineralização na câmara de contacto com o dióxido de carbono e o processo de coagulação/floculação que decorre na câmara de mistura rápida.

A clarificação da água pode ser feita quer por decantação quer por flotação. Após esta etapa, a água passa pela câmara de desmanganização, pelos filtros de areia sendo finalmente conduzida para a cisterna de contacto com o cloro.

Após o processo de tratamento a água é encaminhada, através de condutas adutoras, para os reservatórios de destino onde o cloro residual é de novo ajustado.

Na Figura 13 esquematiza-se também o tratamento das lamas que, após tratadas e desidratadas, são enviadas para aterro sanitário.

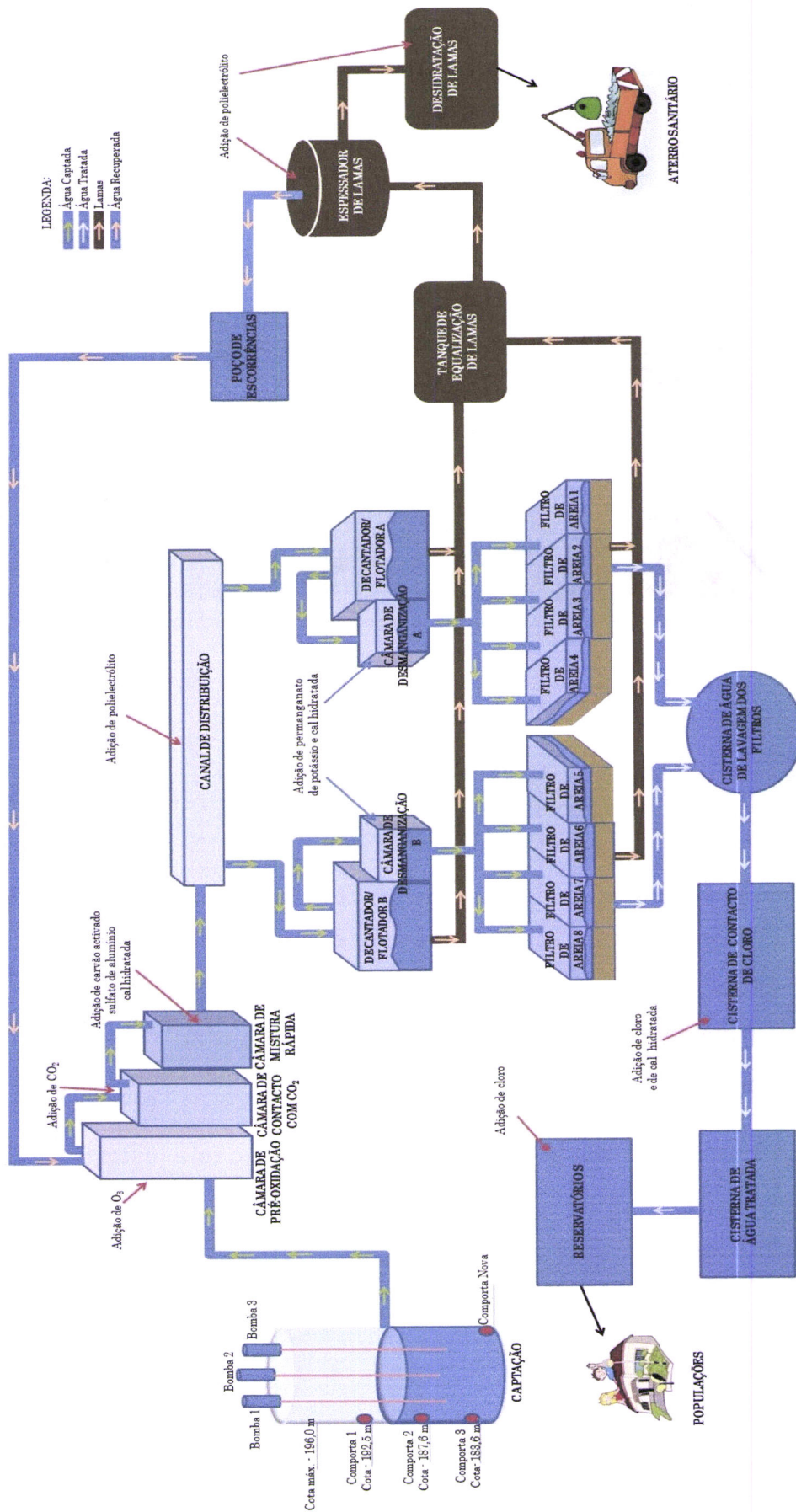


Figura 13: Diagrama linear da ETA do Monte Novo.

2.2.1 – Captação

O sistema de captação e elevação da água bruta está localizado na torre de captação existente na albufeira do Monte Novo, Figura 14.

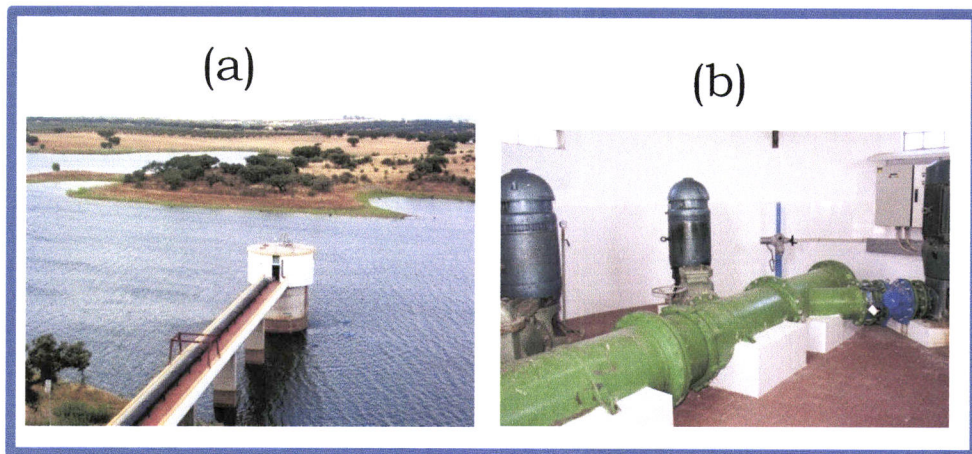


Figura 14: Vista da torre de captação (a) e interior da mesma (b).

A torre de captação encontra-se equipada com três comportas murais de entrada de água, cada entrada a uma cota diferente, Figura 13. As comportas murais alimentam o poço de bombagem, onde estão instaladas três electrobombas que são as responsáveis pela elevação da água até à estação propriamente dita. Duas das electrobombas têm capacidade para elevar $650 \text{ m}^3/\text{h}$ e a terceira possui uma capacidade para elevar $360 \text{ m}^3/\text{h}$.

Com o aumento de capacidade de tratamento desta estação foi construída, em 2005, uma quarta comporta, situando-se esta no fundo do poço de captação.

2.2.2 – Elevação da água bruta

A água captada é elevada e transportada através de uma conduta desde a captação até à estação de tratamento (Figura 15). A ETA do Monte Novo encontra-se preparada para tratar 1100 m³/h, estando dividida em 2 linhas de tratamento, cada uma com a capacidade máxima de 550 m³/h.



Figura 15: Conduto da elevação da água bruta.

2.2.3 – Câmara de Ozonização

A água bruta é encaminhada para uma câmara de contacto onde, por acção do ozono, se realizará a pré-oxidação.

A câmara de contacto com o ozono, Figura 16, possui duas células com chicanas verticais, operando em regime de contracorrente, sendo o tempo de contacto de 30 minutos. O ozono é aplicado através de difusão na massa de água a tratar.

O ozono é usado no tratamento de água como agente oxidante e desinfectante. Devido à sua instabilidade química, este reagente terá que ser produzido no local onde irá ser aplicado. Este reagente é um gás tóxico, extremamente corrosivo e pouco solúvel em água.

O ozono é usado para promover a oxidação de micropoluentes orgânicos, permitindo, para além da remoção da cor e do odor da água, o controlo de agentes precursores de tri-halometanos. Oxida também poluentes fenólicos, alguns tipos de pesticidas e poluentes inorgânicos tais como o ferro, o manganês e o sulfureto de hidrogénio. É ainda responsável pela desinfecção primária que consiste na inactivação do fitoplâncton e do zooplâncton.



Figura 16: Câmara de ozonização.

2.2.4 – Câmara de Contacto com Dióxido de Carbono

Depois de sofrer a pré-oxidação com ozono a água captada é encaminhada para a câmara de contacto com o dióxido de carbono, Figura 17.

O dióxido de carbono é aplicado à água por difusão, no tanque de contacto com o dióxido de carbono. Esta difusão é efectuada através da válvula motorizada, sendo regulada em função o caudal que entra na ETA e do pH que a água apresenta.

O dióxido de carbono é injectado na água bruta através de um conjunto de mangueiras microporosas que se encontram fixadas no fundo do tanque de contacto.



Figura 17: Câmara de contacto com o dióxido de carbono.

Para atingir o equilíbrio calco – carbónico da água é necessário proceder-se à sua remineralização, sendo esta operação realizada através da adição conjunta de dióxido de carbono e de cal hidratada.

A remineralização tem como objectivo corrigir a agressividade da água pré-oxidada, com o enriquecimento da sua concentração em sais minerais para permitir a realização das fases seguintes do tratamento. Este passo tem como objectivo reduzir o teor em dióxido de carbono livre visando, por um lado, criar as condições de tratabilidade da água a potabilizar e, por outro, minimizar a corrosão dos órgãos e equipamentos de tratamento.

O dióxido de carbono é armazenado no estado líquido, num depósito onde é mantido a pressão e temperatura constante. À saída do depósito, o dióxido de carbono passa por um conjunto de vaporizadores

onde, de forma controlada, se dá passagem do estado líquido ao estado gasoso.

2.2.5 – Câmara de Mistura Rápida

A Câmara de Mistura Rápida (CMR), Figura 18, possui duas linhas de tratamento, Figura 19, cada uma com capacidade máxima de 550 m³/h. Este órgão possui chicanas, sendo constituído por vários compartimentos.



Figura 18: Câmara de mistura rápida.

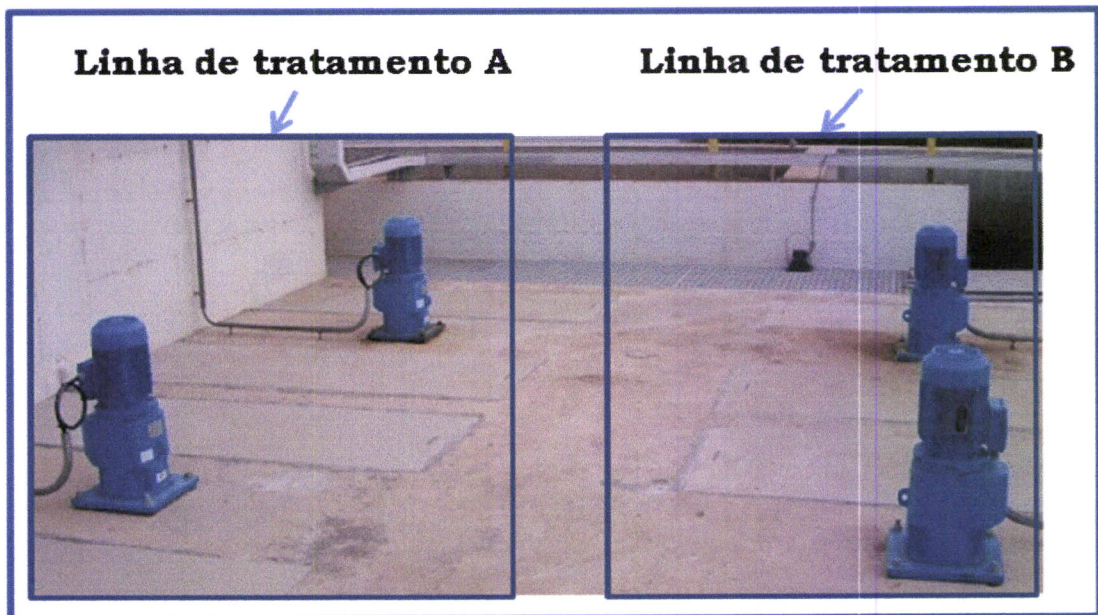


Figura 19: Linhas de tratamento da câmara de mistura rápida.

Na CMR inicia-se processo de coagulação/floculação. Cada linha de tratamento possui, à entrada, uma picagem de sulfato alumínio (que para promover a coagulação), outra de cal hidratada (que visa a remineralização) e um agitador para garantir a homogeneização. Ao passar para o segundo compartimento é adicionado carvão activado e sofre novamente agitação. A Figura 20 mostra a localização das picagens e dos agitadores.

O hidróxido de cálcio é aplicado nas câmaras de mistura rápida sob a forma de suspensão aquosa denominada leite de cal. O seu doseamento é feito em função do caudal circulante e do valor de pH da água à saída da câmara de mistura rápida.

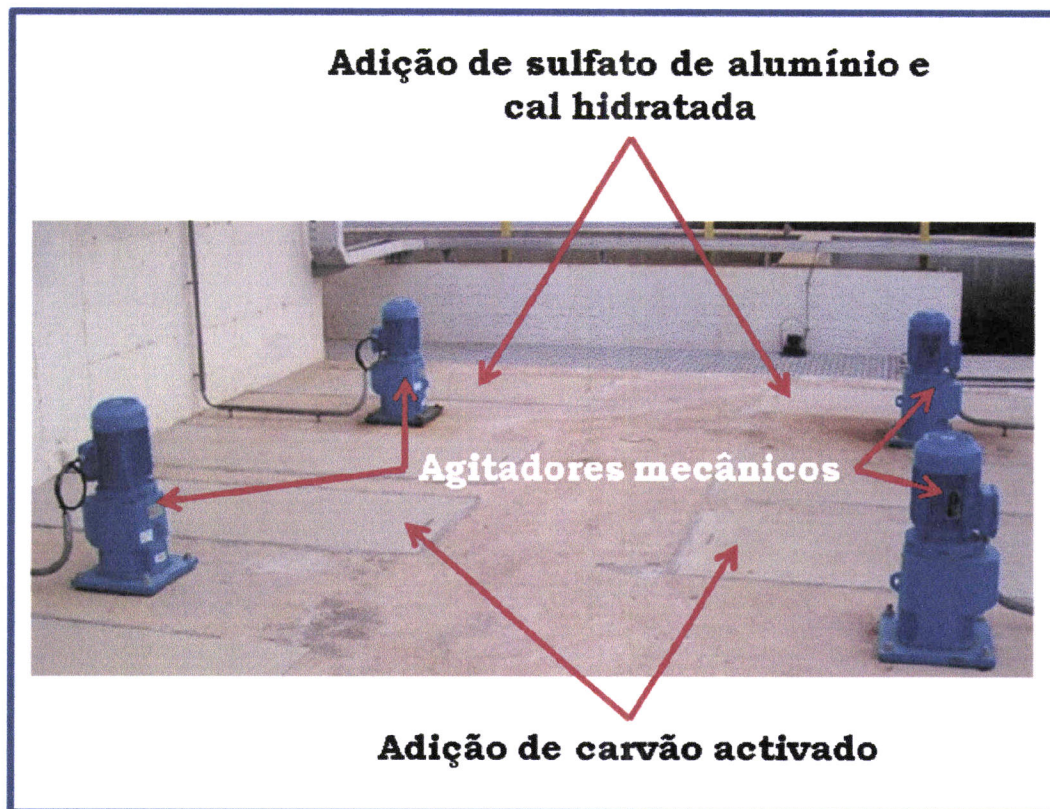


Figura 20: Localização das várias picagens de reagentes e dos agitadores da câmara de mistura rápida.

2.2.6 – Canal de Distribuição

O canal de distribuição (Figura 21) destina-se ao encaminhamento da água proveniente das duas linhas de tratamento da CMR para os decantadores ou para os flotadores, como mostra a Figura 22.

Na caleira é adicionado um floculante que irá ser responsável pela formação de flocos volumosos e pesados resultantes do agrupamento de material coagulado.



Figura 21: Canal de distribuição.

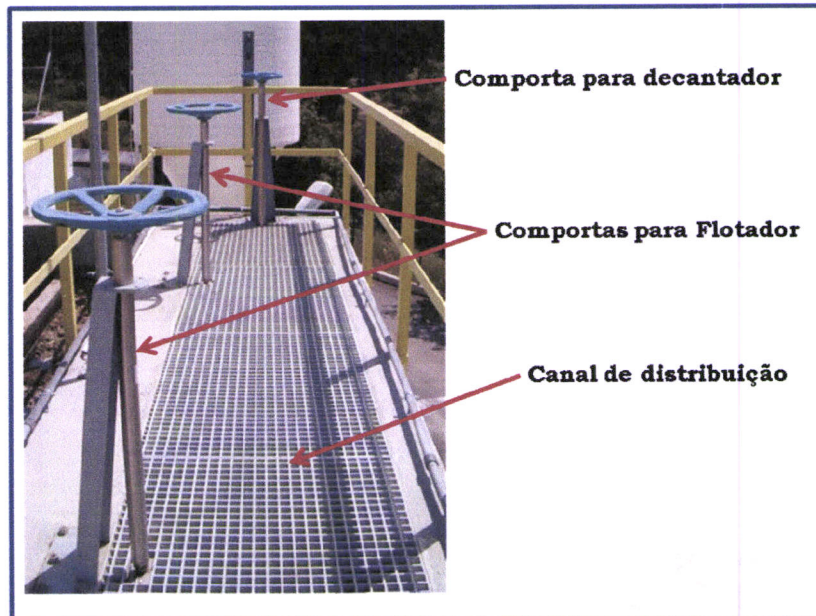


Figura 22: Comportas direccionais no canal de distribuição.

2.2.7 – Decantadores ou Flotadores

Na ETA do Monte Novo a água pode sofrer clarificação quer por decantação quer por flotação.

No canal de distribuição a água divide-se por duas linhas de tratamento, possibilitando a utilização dos dois decantadores (Figura 23) ou dos dois flotadores (Figura 24). Cada linha de tratamento é constituída por um decantador ou por um flotador, uma câmara de desmanganação e uma bateria de quatro filtros de areia.



Figura 23: Decantadores.



Figura 24: Flotador

A decantação consiste na separação gravítica da matéria sólida floculada presente na água visando a sua clarificação. A matéria residual, obtida por decantação gravítica durante o processo de clarificação da água, é designada por lamas decantadas.

A flotação consiste na separação em fluxo ascendente da matéria floculada por injeção de água saturada com ar.

O funcionamento deste órgão de tratamento baseia-se na separação de uma fase sólida de uma fase líquida, através do transporte das partículas sólidas até à superfície do plano de água arrastadas por bolhas finas de ar, geradas num sistema de pressurização de ar.

O material flotado (lamas) é removido de forma mecânica e enviado para tratamento posterior.

Cada tipo de clarificação da água funciona separadamente, uma vez que podem necessitar de dosagens diferentes de coagulante e/ou floculante.

Os decantadores devem funcionar quando a água bruta apresenta uma turvação igual ou superior a 50 NTU (normalmente na época de chuvas), caso contrário, funcionam os flotadores.

2.2.8 – Câmaras de Desmanganização

As duas câmaras de desmanganização encontram-se entre os decantadores/flotadores e os filtros de areia.

Neste local irá ocorrer a oxidação de manganês por intermédio de injeção de permanganato de potássio. A reacção é favorecida a pH 7 podendo ser necessário ajustar o pH usando-se, para o efeito, leite de cal.

Para que esta reacção ocorra sem originar a típica coloração rosa na água tratada, a mistura é efectuada numa câmara de reacção agitada mecanicamente.

2.2.9 – Filtros de Areia

Após a desmanganação, a água é encaminhada graviticamente para a bateria de filtros correspondente.

A filtração em areia é um processo físico de remoção da matéria sólida em suspensão e de impurezas coloidais, obtido pela passagem da água por um leito de composição granular, por exemplo, areia siliciosa. Ocorre de acordo com dois processos simultâneos, i.e. retenção mecânica e adsorção física. No primeiro caso, as partículas maiores ficam retidas nos poros intersticiais da areia, e no segundo caso, os colóides são adsorvidos à superfície dos grãos de areia que constituem o leito filtrante.

Na Figura 25 está esquematizado o percurso que a água toma desde o canal de distribuição até à saída dos filtros na estação de tratamento de águas do Monte Novo.

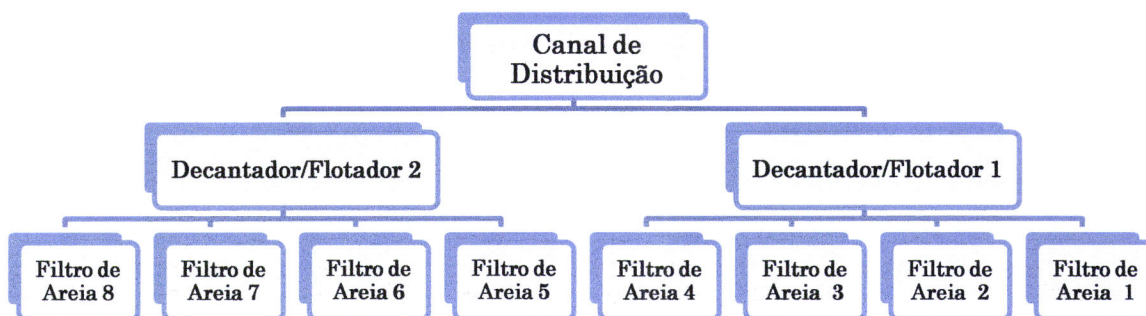


Figura 25: Representação das linhas de tratamento na ETA do Monte Novo.

O processo de filtração é importante para melhor a qualidade da água ao nível dos sólidos. A Figura 26 mostra os filtros de areia que se encontram na ETA Monte Novo.



Figura 26: Filtros de areia.

2.2.10 – Cisterna de Contacto com Cloro

Depois de sofrer o processo de filtração a água é encaminhada para uma cisterna (Figura 27) onde irá sofrer um ajuste de pH, com adição de cal, tendo em vista a protecção das tubagens. Para além desta correcção de pH, irá também haver uma desinfeccção com cloro, permitindo assim que a água chegue aos consumidores sem qualquer problema de contaminação.

O uso deste desinfectante tem como objectivo principal a permanência de um desinfectante residual estável ao longo de todo o sistema de abastecimento.



Figura 27: Cisterna de contacto com o cloro.

2.2.11 – Cisternas de Armazenamento

Na estação de tratamento do Monte Novo existem três cisternas de armazenamento que se encontram interligadas entre si. A Figura 28 mostra as estações elevatórias existentes na ETA Monte Novo.

Estas cisternas recebem a água proveniente da cisterna de contacto com o cloro e, a partir delas, a água é encaminhada por elevação.

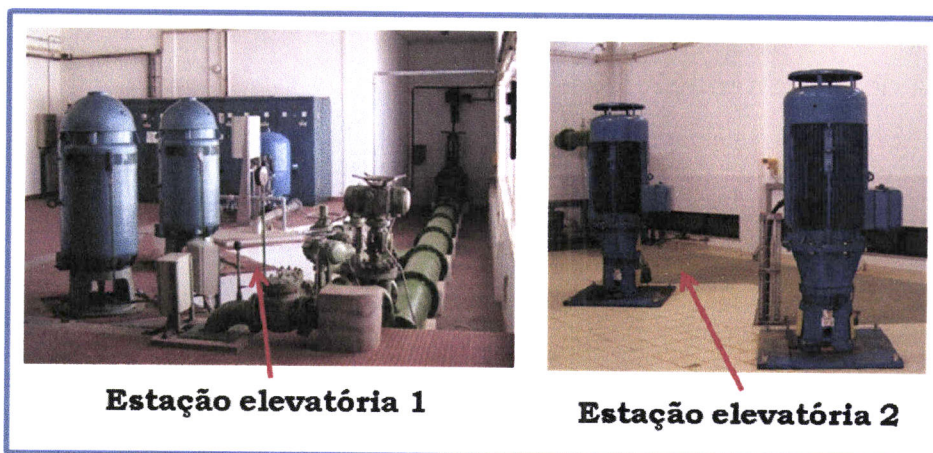


Figura 28: Estações elevatórias existentes na ETA.

2.2.12 – Reservatórios

Os reservatórios são o destino da água proveniente da estação elevatória do Monte Novo e daqui é distribuída, através de condutas adutoras, pelas populações dos municípios de Évora, de Reguengos de Monsaraz, parte do município de Mourão e, caso seja necessário, do município Redondo.

Nestes locais é efectuado um novo ajuste de cloro de modo garantir desinfectante residual até às habitações.

2.2.13 – Fase Sólida

As lamas provenientes quer dos decantadores/flotadores, quer da lavagem dos filtros são posteriormente equalizadas, espessadas e desidratadas por intermédio de um filtro de bandas e enviadas para aterros sanitários.

Inicialmente estas lamas irão, por gravidade, para uma lagoa de equalização onde permanecem em constante agitação. Em seguida serão elevadas para um espessador onde será adicionado polímero permitindo assim a sua concentração (Figura 29).

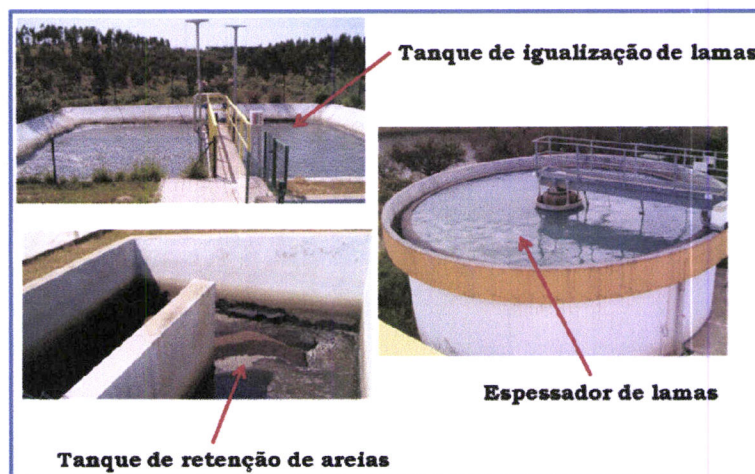


Figura 29: Sistemas de tratamento das lamas.

As lamas espessadas serão enviadas para o filtro de banda (Figura 30) sofrendo desidratação. Os sobrenadantes do espessador são enviados para a câmara de mistura rápida para reutilização.

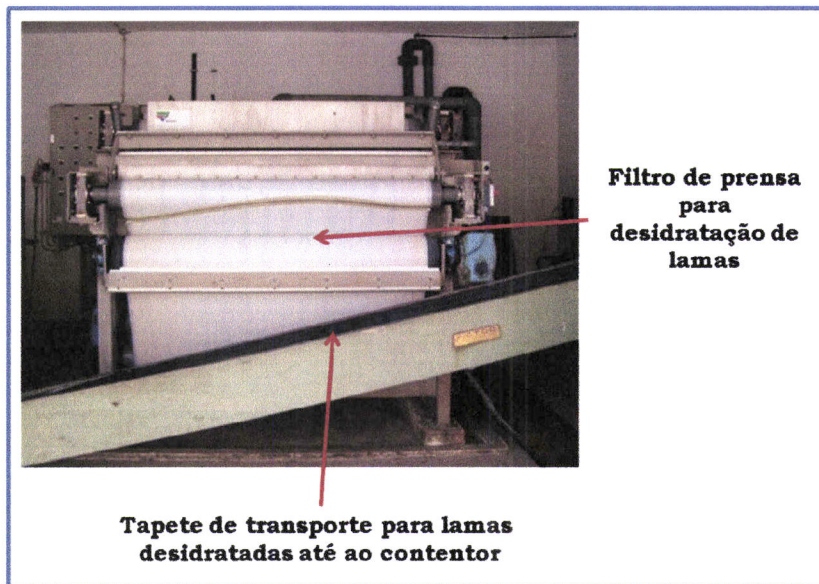


Figura 30: Filtro de banda usado na desidratação de lamas.

Capítulo 3

Descoberta de **C**onhecimento em Base de **D**ados – **M**odelos e **T**écnicas

São apresentados objectivos, conceitos, tipos de abordagem, áreas relacionadas, metodologia e especificações relacionados com o processo de Descoberta de Conhecimentos em Bases de Dados. São apresentados também os principais modelos e técnicas usadas na etapa de *Data Mining*.

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.

O desenvolvimento tecnológico tem proporcionado, no que concerne ao armazenamento de dados, um crescimento exponencial, tanto em número de registos como em complexidade. Como consequência deste aumento efectivo de informação, o seu processamento através dos métodos tradicionais tornou-se cada vez mais difícil e complexo [Michalski et al., 1998] [Goebe et al., 1999] [Sousa, 2004] [Teixeira, 2005].

O desenvolvimento de processos de análise de dados implicou o aparecimento de aplicações vocacionadas para a tarefa de Descoberta de Conhecimento em Bases de Dados incorporando ferramentas de *Data Mining* para extrair informação, padrões de comportamento e tendências desconhecidas de forma autónoma e porventura automática [Thuraisingham, 1999] [Apte et al., 2002].

3.1 – Descoberta de Conhecimento em Base de Dados

O termo Descoberta de Conhecimento em Base de Dados (DCBD) foi instituído em 1989 e refere-se a um processo que abrange a identificação e o reconhecimento de padrões numa Base de Dados (BD) de forma automática, i.e. a obtenção, sem nenhuma formulação prévia de hipóteses, de 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, como se pode observar na Figura 31, envolve diversas etapas [Adriaans & Zantinge, 1996],

[Fayyad et al., 1996], [Michalski et al., 1998], [Sousa, 2004], [Teixeira, 2005].

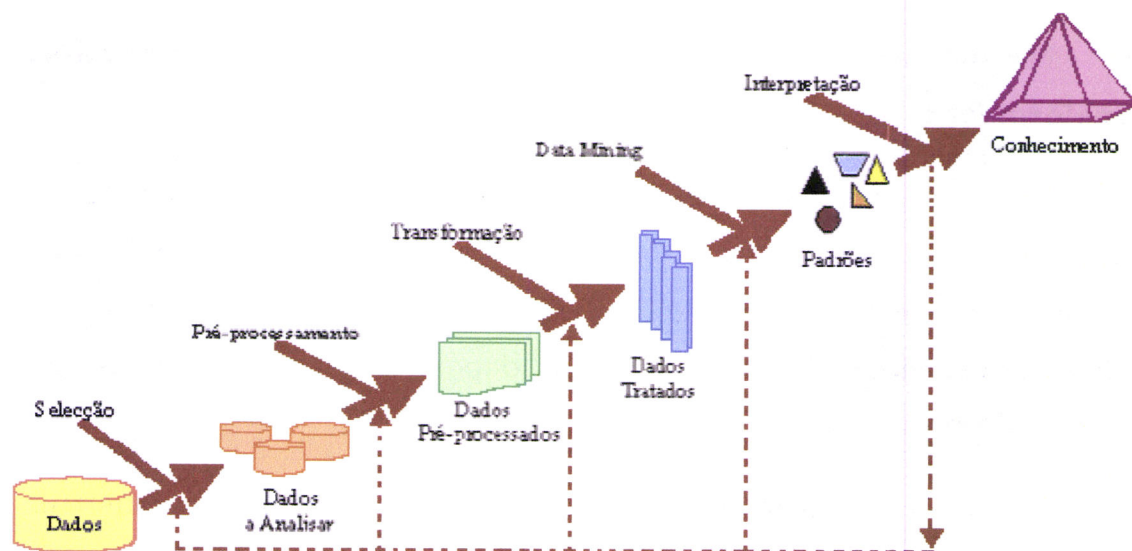


Figura 31: Processo de DCBD. Adaptado de Sousa [Sousa, 2004].

3.1.1 – Etapas do Processo de DCBD

Na resolução de qualquer problema é necessário definir o domínio sobre o qual se pretende executar o estudo, proceder à recolha dos dados e/ou variáveis necessárias à consecução dos objectivos, i.e. a construção da BD sobre a qual se irá actuar.

Na etapa de **selecção** é feita uma segmentação dos dados de acordo com critérios que determinam subconjuntos (e.g. pessoas que possuem um só carro). A maioria das organizações já possuem BD contudo, nem sempre todos os dados necessários estão disponíveis sendo, portanto, necessário um trabalho de compatibilização.

A etapa de **pré-processamento** (*data cleaning*) visa a eliminação de ruído, de erros, de dados incompletos e/ou inconsistentes. São estabelecidas as estratégias para minimizar os problemas relacionados com dados omissos e alguns dados podem ser reconfigurados para

assegurar uma base de dados uniforme e consistente (e.g. valores de um determinado parâmetro expressos em unidades diferentes).

Assegurada uma BD uniforme e consistente é necessário armazenar os dados num formato que facilite a sua utilização por parte das ferramentas de *Data Mining*. Esta etapa denomina-se **transformação**.

A etapa de **data mining** constitui o núcleo do processo de DCBD e, muitas vezes, é confundida com ele. Consiste na escolha e utilização dos métodos e técnicas que melhor se enquadram no cumprimento das metas estabelecidas. A referida escolha depende fundamentalmente do objectivo do processo de *Data Mining*.

Na etapa de **interpretação e avaliação**, pretende-se construir uma representação do conhecimento que seja acessível e clara para os utilizadores finais. 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.

Antes de se poder dar o processo de DCBD como concluído deve fazer-se a avaliação da qualidade da informação obtida. Este passo pode ser encarado como a avaliação do desempenho do mecanismo de aprendizagem. Inclui-se nesta etapa porque a determinação dos valores de confiança e de suporte são extremamente importantes para a validação dos modelos.

A Figura 31, que esquematiza a DCBD, pode sugerir um percurso linear, no entanto, isso geralmente não se verifica, uma vez que em cada etapa pode ser identificada a necessidade de retorno para uma das fases anteriores. Por exemplo, se na fase *Data Mining* se concluir que os dados não são plenamente coerentes, ou for verificada a necessidade da inclusão de um novo dado, isso pode levar ao retorno para fases anteriores como a de transformação ou a de selecção. Nos projectos de

DCBD, normalmente 80% do tempo é dispendido nas etapas de preparação dos dados, enquanto os restantes 20% referem-se à etapa de *Data Mining* propriamente dita [Adriaans & Zantinge, 1996].

3.2 – Data Mining

Encontram-se várias definições para o termo *Data Mining*, consoante o autor, a abordagem e a área de especialização [Fayyad et al., 1996] [Fayyad, 1997] [Chang et al., 2001] [Cortez, 2004] [Sousa, 2004] [Teixeira, 2005]. Contudo, existem características comuns que se podem sintetizar na seguinte definição: aplicação de métodos e técnicas em grandes Bases de Dados (BD), para encontrar tendências ou padrões com o intuito de descobrir conhecimento.

Como foi anteriormente referido, a fase de *Data Mining* (DM) é apenas uma etapa de um processo mais abrangente, denominado de Descoberta de Conhecimento em Bases de Dados (DCBD).

A etapa de *Data Mining*, como foi referido anteriormente, consiste na escolha e utilização dos métodos e técnicas que melhor se enquadram no cumprimento dos objectivos estabelecidos. Na Tabela 2 apresentam-se e caracterizam-se alguns desses objectivos divididos em duas tarefas, i.e. previsão e descrição. No caso da previsão pretende-se inferir comportamentos futuros com base nas experiências do passado, enquanto que no caso da descrição pretende-se, de um modo geral, descrever de forma compacta conjuntos de dados ou associações variáveis.

Tabela 2: Caracterização dos objectivos do processo Data Mining

Objectivo		Caracterização
Previsão	Classificação	Consiste em encontrar, a partir de exemplos, uma função que associe um caso a uma classe, dentro de diversas classes discretas de classificação, de forma a atribuir uma classe de classificação a um novo objecto.
	Regressão	Consiste em encontrar, a partir de exemplos, uma função que represente o comportamento de uma variável, de modo a prever valores futuros ou desconhecidos dessa variável.
Descrição	Associação ou Dependência	Consiste em encontrar modelos que descrevam dependências significativas entre variáveis, através da identificação de grupos de dados fortemente associados.
	Sumariação	Consiste em encontrar uma descrição compacta para um subconjunto de dados. É aplicada à análise exploratória de dados e à geração automática de relatórios.
	Segmentação	Consiste em identificar um conjunto finito de categorias ou segmentos para descrever os dados. Dentro do mesmo segmento os objectos são semelhantes e, entre segmentos, são discrepantes. Pode ser um objectivo intermédio, podendo ser realizada numa fase inicial para encontrar segmentos homogéneos de dados para posterior aplicação, a cada um deles, de algoritmos de aprendizagem para classificação ou para previsão.
	Visualização	Consiste em apresentar os resultados (finais ou intermédios) obtidos com as ferramentas de DM. Pretende dar a conhecer ao utilizador informações complexas através de um formato acessível (e.g. gráficos e/ou diagramas).

3.3 – Modelos e Técnicas de Data Mining

Um modelo é definido como uma função que atribui a cada exemplo, no domínio definido pelos atributos de entrada, um valor contido no domínio dos atributos de saída. Cada modelo contém um conjunto de parâmetros que têm de ser ajustados (ou estimados) a partir de um conjunto de dados, através de um algoritmo, na fase de aprendizagem. Após a aprendizagem, é possível extrapolar novas saídas, alimentando o modelo com novas entradas (utilização de um modelo) [Cortez, 2004] [Teixeira, 2005].

Na construção de um modelo definem-se as principais características do sistema, que devem representar o mais fielmente possível a realidade e recolhem-se os dados necessários para a sua construção e para a sua validação. Torna-se portanto necessária uma divisão da BD em dois subconjuntos, mutuamente exclusivos, um para gerar o modelo, designado como **conjunto de treino**, e outro para a sua validação, denominado **conjunto de teste**. A cardinalidade dos referidos subconjuntos é variável e dependente de vários factores (e.g. a dimensão da base de dados).

Deve realçar-se que não existe um modelo universal de DM que resolva, de forma eficiente, todos os problemas. A escolha de um determinado algoritmo é de certa forma uma arte, uma vez que existem diferentes modelos para as mesmas tarefas de DM com vantagens e desvantagens intrínsecas [Fayyad et al., 1996]. Na Tabela 3 é apresentado um mapeamento entre as diversas tarefas e as técnicas de DM, que são normalmente utilizadas.

Tabela 3: Tarefas e técnicas usadas no processo de *Data Mining*

Tarefa \ Técnica	Previsão		Descrição			
	Classificação	Regressão	Associação	Sumariação	Segmentação	Visualização
Árvores de Decisão	✓	✓		✓	✓	✓
Indução de Regras	✓	✓	✓		✓	
Redes Neurais Artificiais	✓	✓	✓		✓	
Algoritmos Genéticos	✓	✓		✓	✓	
Aproximação de Vizinhanças				✓	✓	

Como foi referido anteriormente, os modelos e as técnicas utilizadas na etapa de *Data Mining* dependem da natureza do problema e dos objectivos que se pretendem atingir. No caso do presente trabalho, as técnicas utilizadas foram as árvores de decisão e a indução de regras na tarefa de classificação e a aproximação de vizinhanças na tarefa de segmentação.

3.3.1 – Árvores de Decisão

As Árvores de Decisão tiveram a sua origem na área da Aprendizagem Automática (i.e. *Automatic Interaction Detection*). Esta análise testa automaticamente todos os valores de um determinado atributo de forma a identificar aqueles que têm uma forte associação com os registos de saída.

Uma árvore de decisão é uma representação de um conjunto de regras, que classificam as instâncias desde o nó raiz até aos nós

terminais ou folhas. Cada nó da árvore especifica um teste para os atributos da instância e, cada ramo descendente desse nó, correspondente 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 até ao nó terminal. 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 [Quinlan, 1993] [Kufirin, 1997] [Han & Kamber, 2006].

Existem dois tipos de árvores de decisão, i.e. árvores de classificação e árvores de regressão. As primeiras, aplicam-se a problemas de classificação, i.e. têm por objectivo qualificar os registos e associá-los com uma determinada classe. As segundas, por seu lado, aplicam-se a problemas de regressão i.e. visam estimar o valor de uma determinada variável [Quinlan, 1996] [Berry & Linoff, 2000]. Estes dois tipos de árvores apresentam a mesma estrutura. Quando os dados são apresentados ao modelo, cada registo percorre um determinado caminho para uma série de testes, até que o registo alcança uma folha ou nó puro da árvore, onde lhe é atribuída uma classe ou um valor numérico consoante se tratem, respectivamente, de árvores de classificação ou de árvores de regressão [Berry & Linoff, 2000].

Na construção de uma árvore de decisão existem dois problemas com que se pode deparar, i.e. que atributo seleccionar para teste num determinado nó e, quando parar a divisão dos exemplos. No que diz respeito à escolha dos atributos, existem várias medidas para avaliar a capacidade de um dado atributo para discriminar as classes, no entanto todas convergem em dois pontos, i.e. uma divisão que mantém as proporções de classes em todas as partições é inútil, enquanto uma divisão onde em cada partição todos os exemplos são da mesma classe tem utilidade máxima [Sousa, 2004] [Teixeira, 2005] [Han & Kamber, 2006].

No que concerne à paragem da divisão dos exemplos, esta deve acontecer quando todos os exemplos pertencem à mesma classe, ou todos os exemplos têm os mesmos valores dos atributos mas classes diferentes, ou o número de exemplos é inferior a um certo limite, ou o mérito de todos os possíveis testes de partição dos exemplos é muito baixo [Sousa, 2004] [Teixeira, 2005] [Han & Kamber, 2006].

Os algoritmos de indução de árvores de decisão constroem árvores a partir dos dados de treino, de uma forma recursiva subdividindo este conjunto de dados até que seja formada apenas por nós “puros”, em que cada nó represente apenas uma única classe ou a satisfação de um critério. Utilizam a aprendizagem supervisionada, i.e. são fornecidas as respostas correctas ao sistema, a partir de um conjunto de exemplos, cada um deles formado por um vector de entrada e um vector de saída ou resposta [Rich & Knight, 1991]. O algoritmo ID3, Figura 32, desenvolvido em 1983 por Ross Quinlan, foi um dos algoritmos pioneiros que permitiu o desenvolvimento das árvores de decisão. Desde então têm sido introduzidas novas funcionalidades que resultaram no aparecimento de evoluções deste algoritmo (e.g. algoritmos C4.5 e C5.0).

Seja T a árvore de decisão a induzir
 Seja S o conjunto de exemplos para aprendizagem
 Se todos os exemplos de S pertencem à mesma classe C
 então a árvore tem um só nodo C
 Senão seleccionar o atributo A mais “informativo” cujos valores são $V_1,$
 \dots, V_n
 Particionar S em n subconjuntos S_1, \dots, S_n um para cada valor V_i de A
 construir (recursivamente) sub - árvores T_1, \dots, T_n para cada S_1, \dots, S_n

Figura 32: Algoritmo de indução de árvores de decisão.

Uma das limitações inerentes ao algoritmo ID3 consiste na possibilidade de gerar árvores demasiado ajustadas aos dados de treino

(i.e. sobreajustamento), com um desempenho quase perfeito neste conjunto de dados, mas com um baixo desempenho nos novos dados (conjunto de teste). No algoritmo C4.5 [Quinlan, 1993] foram introduzidos melhoramentos que permitem ultrapassar o sobreajustamento, através da introdução da poda da árvore (i.e. redução de algumas subárvores a folhas, tendo em conta os erros nesse nó e a soma dos erros nos nós que dele descendem). Além do sobreajustamento o algoritmo C4.5 permite ultrapassar problemas concretos e comuns do mundo real (e.g. atributos com valores numéricos, valores omissos, dados com ruído), disponibilizando a possibilidade de realizar validação cruzada, incrementando assim a qualidade da estimativa do erro cometido pelo classificador.

O algoritmo C5.0 [URL2] é o sucessor mais recente do algoritmo C4.5, melhorado para lidar com as exigências do mundo real, através do incremento da eficiência ao nível do tempo de processamento e de memória utilizada. Uma das características mais importantes deste algoritmo é a utilização da técnica de *Boosting* [Schapire, 2002], que consiste em gerar vários classificadores a partir do mesmo conjunto de treino e depois combiná-los num único classificador final, no qual cada classificador inicial participa com um certo peso. Este peso é ajustado durante o processo de treino [Quinlan, 1996]. Em determinados casos a redução dos erros de classificação pode atingir os 40% [URL2].

3.3.2 – Indução de Regras

A Indução de Regras refere-se à detecção de tendências e padrões em grupos de dados. Esta técnica consiste na descoberta de regras de previsão, do tipo **Condição... Acção**, onde a condição da regra especifica alguns atributos, e a acção da regra prevê um valor para um determinado atributo cujo conhecimento é desejado [Quinlan, 1993] [Fayyad et al., 1996] [Berson et al., 2000].

As regras devem, segundo Kufrin [Kufrin, 1997] e Quinlan [Quinlan, 1993], satisfazer três condições, nomeadamente previsões correctas (i.e. se o **Se** da regra é verdadeiro, o **Então** da regra também deve ser verdadeiro); compreensíveis (i.e. devem ser de fácil compreensão para o utilizador); e úteis para a tomada de decisão (i.e. expressar conhecimento que é ou não esperado por parte do utilizador).

A indução de regras é uma das técnicas de *Data Mining* mais populares e permitiu fazer uma das descobertas mais famosas na cadeia de supermercados *Wall-Mart*: “às sextas-feiras, quem compra fraldas também compra cerveja” [Quinlan, 1993] [Berson et al., 2000].

A Indução de Regras surge, na maior parte dos casos, associada às Árvores de Decisão, sendo usadas para representar o conhecimento nelas contido. Na apresentação das regras, a regra mais importante corresponde ao primeiro nó da árvore e, as regras menos relevantes correspondem aos nós subsequentes de acordo com a sua relevância.

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 pela divisão do número de casos onde X e Y aparecem juntos (i.e. o suporte absoluto), pelo número de casos onde somente X aparece [Quinlan, 1993].

As vantagens da técnica de Indução de Regras prendem-se com o modo directo de lidar com os dados, o desempenho, a facilidade de explicação e compreensão das regras. Em contrapartida, a dificuldade no manuseamento da informação incompleta ou de valores fora do contexto e serem altamente heurísticas constituem as suas principais desvantagens [Langley & Simon, 1995].

3.3.3 – Aproximação de Vizinhanças

A aproximação de vizinhanças baseia-se no princípio que registos semelhantes estão próximos uns dos outros, quando analisados numa perspectiva espacial [Han & Kamber, 2006] [Bramer, 2007].

A verificação da localização dos registos, interpretados como pontos no espaço, permite a identificação de regiões, denominadas classes ou segmentos, que definem características comuns para os registos que representam. A complexidade na utilização desta técnica aumenta à medida que cresce o número de registos a analisar, uma vez que cada registo é comparado com os restantes registos da amostra.

Esta técnica assenta na construção de partições dos objectos armazenados numa BD, num conjunto de k classes, sendo k um parâmetro de entrada. O processo de obtenção das classes é iniciado com centróides em posições aleatórias, as quais são optimizadas iterativamente através da movimentação dos centros [Han & Kamber, 2006].

Cada classe pode ser representada através do seu centro de gravidade, i.e. pela localização média de todos os membros do segmento (estratégia *k-means*) ou por um dos objectos da classe próximo do seu centro (estratégia *k-medoid*) [Teixeira, 2005] [Han & Kamber, 2006].

Para determinação das classes, cada registo é transformado num ponto no espaço, apresentando este tantas dimensões quantos os atributos em análise. O valor de cada campo é interpretado como a distância da origem até à sua localização num dado eixo.

3.3.4 – Avaliação dos Modelos

Após a geração dos modelos, é necessário avaliar o desempenho dos mesmos. Existem vários métodos para estimar a capacidade de generalização de um modelo (e.g. Estatística Simples, Validação com

Divisão da Amostra, Validação Cruzada e *Bootstrapping*) [Han & Kamber, 2006] [Bramer, 2007].

Um dos métodos mais utilizados para estimar a capacidade de generalização é a validação com divisão da amostra, que se baseia numa divisão dos dados do problema em casos de treino, para a aprendizagem do modelo, e casos de teste para avaliar a sua capacidade de generalização. As principais vantagens deste método prendem-se com a sua simplicidade e rapidez, a sua principal desvantagem está relacionada com a redução efectiva dos dados disponíveis para treino [Han & Kamber, 2006] [Bramer, 2007].

A Validação Cruzada, Figura 33, é um melhoramento do método de Validação com Divisão da Amostra, que permite a utilização de todos os casos disponíveis. Na validação cruzada k -desdobrável, os dados (P) são divididos em k subconjuntos mutuamente exclusivos (P_1, P_2, \dots, P_k) de cardinalidade semelhante. Os modelos são treinados e testados k vezes. O erro final de generalização é dado pela média dos erros de validação obtidos durante os k testes. Os valores de k podem variar entre 2 e n , embora o valor 10 seja o mais utilizado. A Validação Cruzada é notavelmente superior à validação com divisão da amostra para pequenos conjuntos de exemplos de treino. Contudo, é conseguida à custa de um grande esforço computacional [Han & Kamber, 2006] [Bramer, 2007].

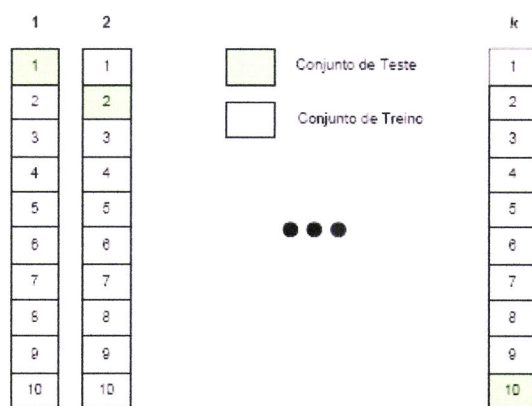


Figura 33: Validação cruzada com k iterações.

Capítulo 4

Procedimento **E**xperimental

Neste capítulo são apresentados os passos do procedimento seguido na realização deste trabalho. É descrito o laboratório de controlo da Empresa Águas do Centro Alentejo (local onde foi realizada a componente analítica), assim como o processo de colheitas adoptado. Descreve-se também a construção da base de dados que foi utilizada na construção dos modelos.

O laboratório da empresa Águas do Centro Alentejo (ADCA) fica situado na ETA do Monte Novo e é responsável pelo controlo interno da qualidade da água.

No laboratório da ETA do Monte Novo são determinados vários parâmetros, nomeadamente a alcalinidade, a dureza, o pH, a temperatura, a turvação, o alumínio, o cloro livre, o cloro combinado, o manganês, a oxidabilidade, a condutividade, o oxigénio dissolvido e o

ferro, quer na água captada quer na água tratada. Alguns dos parâmetros referidos são ainda determinados em pontos intermédios da estação, nomeadamente à saída da câmara de mistura rápida, dos decantadores/flotadores e dos filtros de areia.

4.1 – Amostragem e Preservação das Amostras

O procedimento de amostragem é um elemento importante do programa de controlo da qualidade da água para consumo humano. Não se pode depositar confiança no resultado da uma análise laboratorial, mesmo que seja seguido um método analítico aconselhado e o equipamento mais adequado, se a amostra não for representativa da água a controlar.

O principal objectivo da amostragem é recolher uma amostra de água de volume suficientemente pequeno que possa ser convenientemente transportado e manuseado no laboratório mas, que seja representativo da massa de água. Deste modo, as proporções relativas e as concentrações das espécies que se querem dosar não devem variar significativamente até à conclusão das análises [Mendes & Oliveira, 2004] [Eaton et al., 2005] [Zhang, 2007].

O controlo analítico da qualidade da água inicia-se com a colheita da amostra, sendo esta efectuada diariamente durante o período da manhã. Neste trabalho foram seguidas as recomendações do SMEWW [Eaton et al., 2005], encontrando-se na Tabela 4 o tipo de recipiente utilizado e o volume recolhido em função do parâmetro a analisar. Houve o cuidado de passar os recipientes pela água a colher e o enchimento foi total, garantindo a inexistência de espaços livres para evitar a alteração das características da amostra.

Tabela 4: Recipientes utilizados e volumes de amostras recolhidos para a determinação dos vários parâmetros.

Determinação	Recipiente	Volume de amostra / cm ³
Alcalinidade	Polietileno	100
Dureza	Polietileno	10
Oxidabilidade	Polietileno	100
pH	Polietileno	200
Temperatura	Polietileno	200
Turvação	Polietileno	10
Cloro Livre e Cloro Combinado	Polietileno	10
Condutividade	Polietileno	200
Oxigénio Dissolvido	Polietileno	200
Ferro	Polietileno	50
Alumínio	Polietileno	5
Manganês	Polietileno	5

A recolha das amostras foi efectuada duas horas após a entrada em funcionamento da estação para permitir a renovação da água nos vários corpos de tratamento. Os pontos de amostragem foram a conduta de elevação (água captada), no canal de distribuição, à saída dos decantadores/flotadores, à saída dos filtros das linhas de tratamento 1 e 2 e à saída de estação (água tratada).

A periodicidade da realização destas determinações depende do local de amostragem. No caso da água captada (i.e. água colhida na conduta de elevação) e da água tratada (i.e. água colhida à saída da estação de tratamento) são realizadas diariamente. No caso dos restantes pontos de colheitas são realizadas diariamente as determinações dos parâmetros pH, temperatura, turvação, oxigénio dissolvido, condutividade, alumínio e manganês, enquanto as restantes análises são efectuadas semanalmente.

Cada amostra é acompanhada pela folha de registo, Figura 34, onde consta a descrição do ponto de amostragem, data e hora da colheita, data e hora de entrada das amostras no laboratório, o registo dos resultados dos parâmetros determinados no local (e.g. temperatura, oxigénio dissolvido, pH), identificação do técnico responsável pela colheita e outros aspectos, tais como a conservação da amostra e as condições ambientais.

Águas do Centro Alentejo **Análises de Processo - Controlo Interno**

LOCAL: _____ DATA: ____/____/____

PARÂMETROS	CÓDIGO DE AMOSTRA										ASSINATURA
	HORA DE COLHEITA										
	DESCRIÇÃO DO PONTO DE COLHEITA										
Acidez (mg/L)											
Dureza Total (mg/L)											
pH (escala linear)											
Temperatura (°C)											
Turbidez (NTU)											
Amónio (mg/L)											
Cl ₂ (mg/L)											
Cl ₂ liberado (mg/L)											
Manganês (µg/L)											
Oxigénio dissolvido (mg/L)											
Condutividade (µS/cm)											
O ₃ liberado (mg/L)											
Ferro (µg/L)											

OBSERVAÇÕES

Responsável pela colheita: _____ Tipo de acondicionamento: _____

Data de Receção no Laboratório: ____/____/____ Hora Receção no Laboratório: _____

responsável de laboratório

Figura 34: Folha de registo que acompanha as amostras.

No caso do presente trabalho, como o laboratório se situa na ETA do Monte Novo, as amostras são colhidas e direccionadas rapidamente para análise não sendo necessário a conservar as amostras.

4.2 – Procedimento Analítico

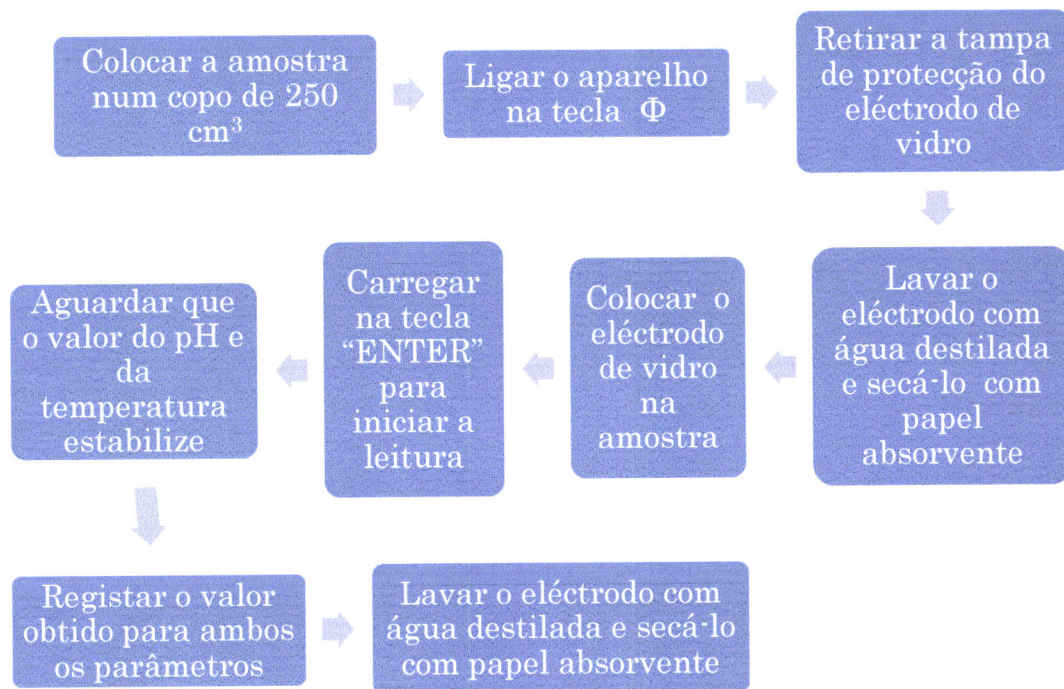
Os parâmetros determinados no laboratório de controlo da empresa Águas do Centro Alentejo seguem os métodos apresentados na Tabela 5.

Tabela 5: Técnicas analíticas utilizadas no laboratório de controlo da ETA do Monte Novo.

Parâmetro	Técnica	Método	Procedimen to Analítico
pH/Temperatura	Potenciometria	SMEWW - 4500-H ⁺ B	1
Condutividade	Condutimetria	SMEWW - 2510 B	2
Turvação	Fotometria	SMEWW - 2130 B	3
Oxigénio dissolvido	Potenciometria	SMEWW - 4500-O B	4
Cloro Residual e Cloro Combinado	Fotometria	SMEWW - 4500-Cl G	5
Alcalinidade	Titulação Ácido-Base	SMEWW - 2320 B	6
Dureza	Titulação complexiométrica	SMEWW - 2340 C	7
Oxidabilidade	Titulação por retorno	Kubel - Tiemann	8
Ferro	Espectroscopia UV/Vis.	SMEWW - 3500-Fe	10
Alumínio	Espectroscopia UV/Vis.	Kit Fotométrico (MerK)	11
Manganês	Espectroscopia UV/Vis.	Kit Fotométrico (MerK)	12

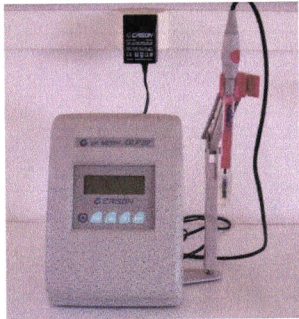
Em seguida apresentam-se os procedimentos analíticos seguidos para a determinação dos parâmetros referidos, o equipamento utilizados e os métodos de quantificação.

4.2.1 – Determinação do pH e da Temperatura

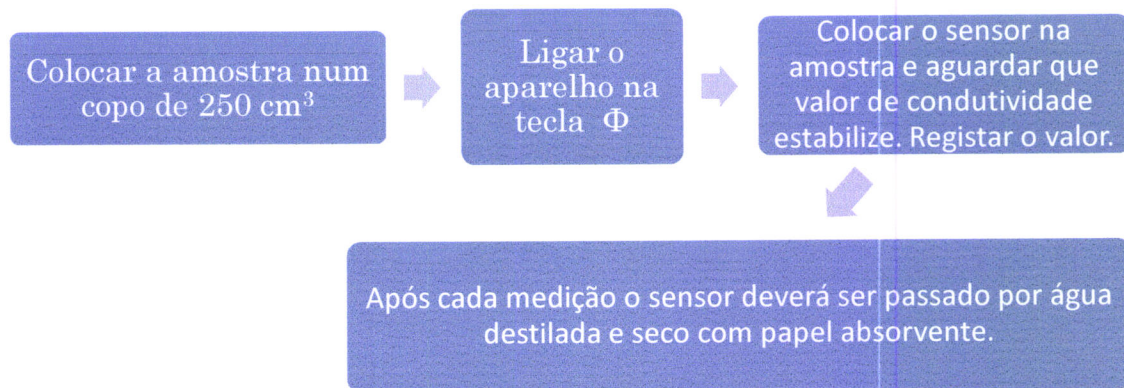


Procedimento Analítico 1: Determinação do pH e da temperatura.

Quadro 1: Equipamento utilizado na determinação do pH/temperatura.

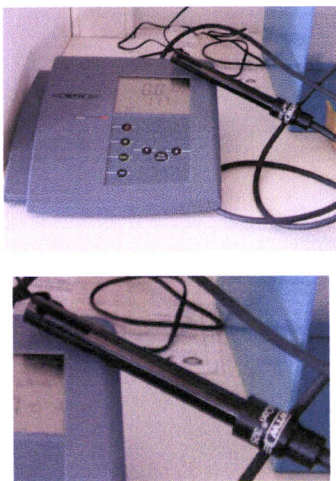
Equipamento	Marca	Observações
	pH Meter GLP22	<p>Fabricante: Crison Instruments, S.A.</p> <p>País/Fabricao: Espanha</p> <p>Nº de série: 507007</p>

4.2.2 – Determinação da Condutividade

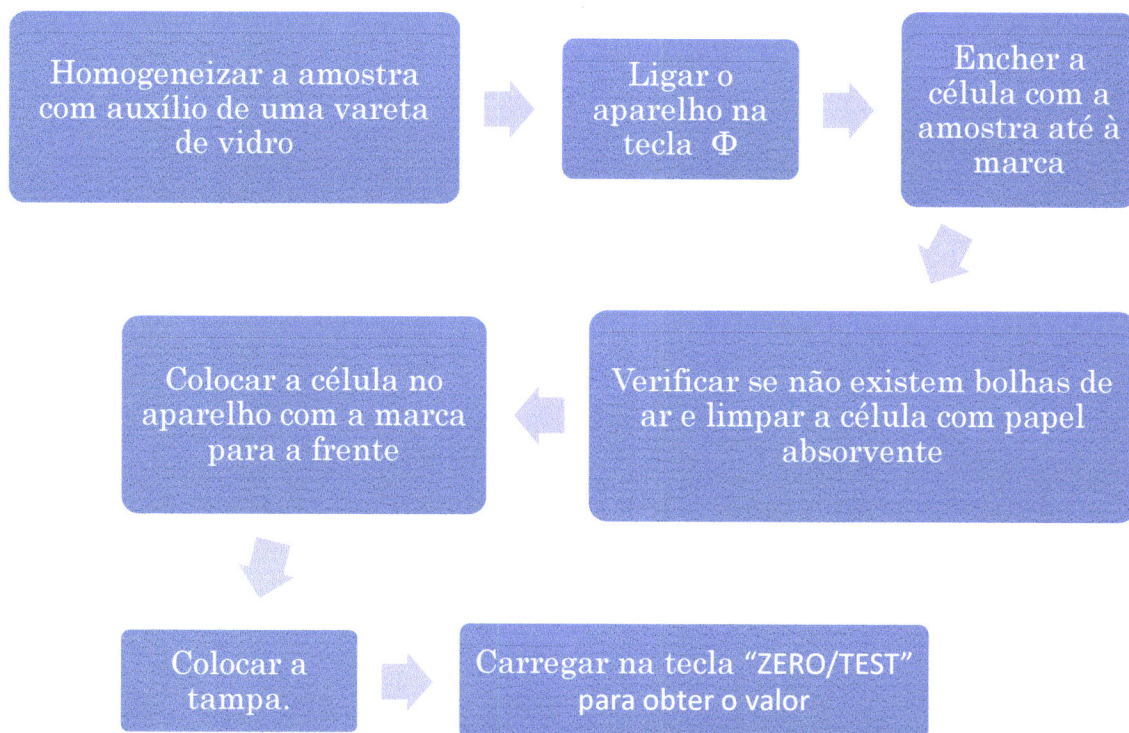


Procedimento Analítico 2: Determinação da condutividade.

Quadro 2: Equipamento utilizado na determinação da condutividade.


Equipamento	Marca	Observações
	inolab cond 720	<p>Fabricante: W.T.W</p> <p>País/Fabricao: Alemanha</p> <p>Nº de série: 05100018</p> <p>Eléctrodo: Tetracon 325</p>

4.2.3 – Determinação da Turvação

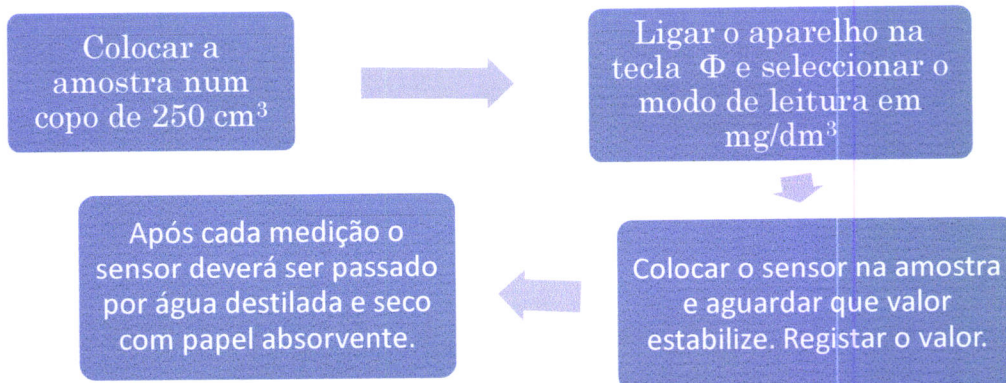


Procedimento Analítico 3: Determinação da turvação.

Quadro 3: Equipamento utilizado na determinação da turvação.


Equipamento	Marca	Observações
	PC CHECKIT TURBIDITY	<p>Fabricante: Lovibond</p> <p>País/Fabrico: Alemanha</p> <p>Nº de série: 52201</p>

4.2.4 – Oxigénio Dissolvido

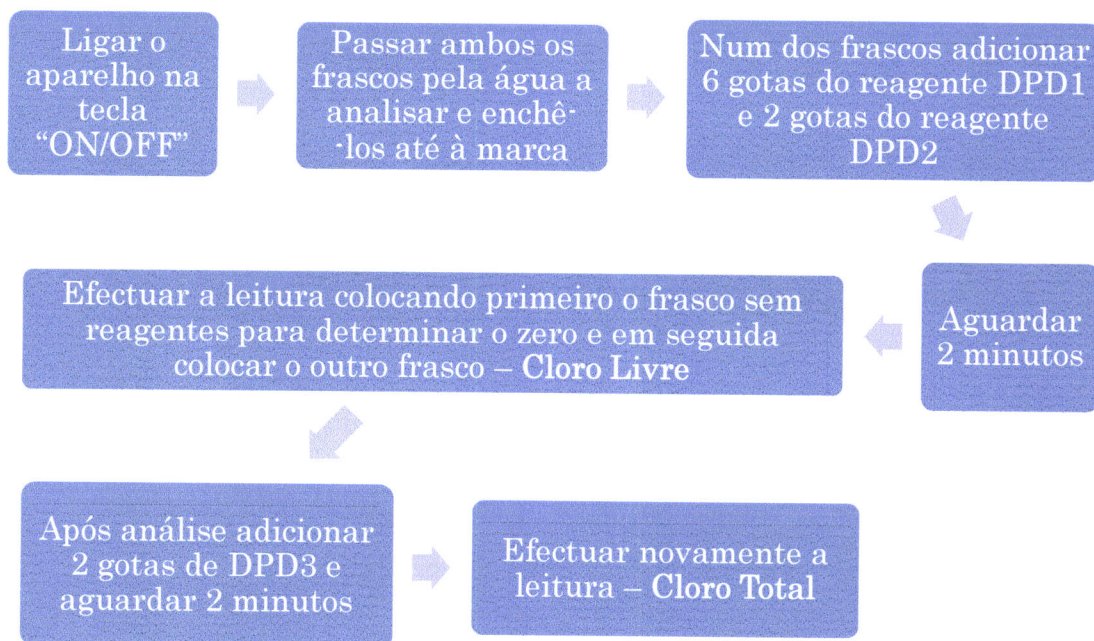


Procedimento Analítico 4: Determinação do oxigénio dissolvido.

Quadro 4: Equipamento utilizado para a determinação do oxigénio dissolvido.


Equipamento	Marca	Observações
	OXI 45	<p>Fabricante: Crison Instruments, S.A.</p> <p>País/Fabrico: Espanha</p> <p>Eléctrodo: DurOx</p>

4.2.5 – Determinação do Cloro Residual e do Cloro Combinado



Procedimento Analítico 5: Determinação do cloro livre e cloro total.

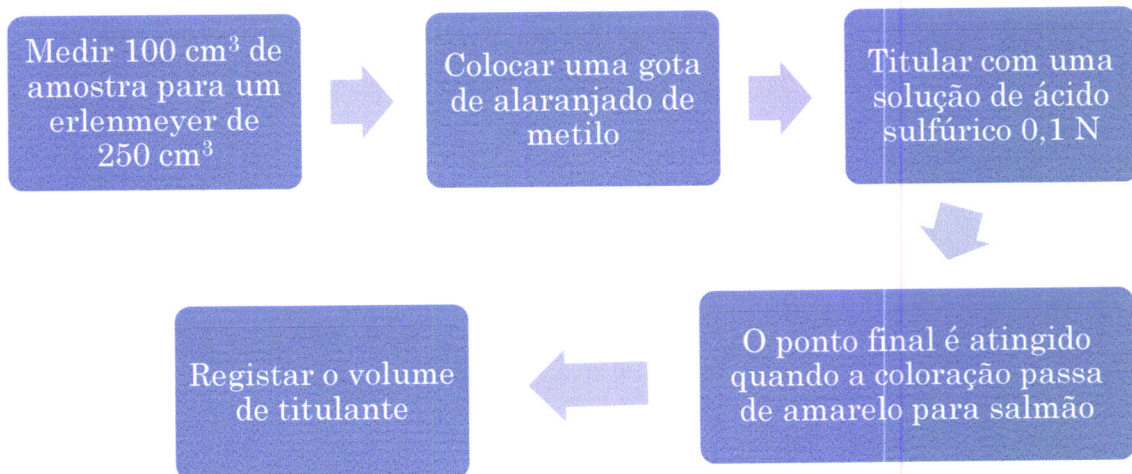
Quadro 5: Equipamento e reagentes usados na determinação do cloro residual e combinado

Equipamento	Marca	Observações
	PC CHECKIT CHLORINE	<p>Fabricante: Lovibond</p> <p>País/Fabrico: Alemanha</p>

O teor em cloro combinado é obtido por diferença:

$$\text{Cloro}_{\text{Combinado}} = \text{Cloro}_{\text{Total}} - \text{Cloro}_{\text{Livre}}$$

4.2.6 – Determinação da Alcalinidade



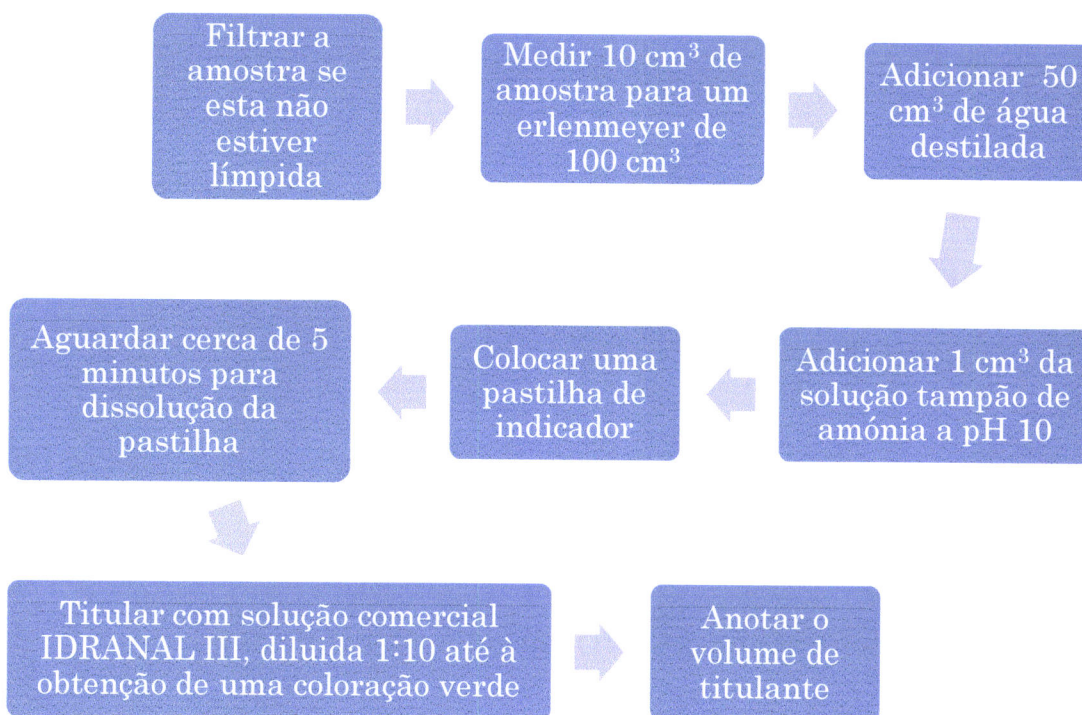
Procedimento Analítico 6: Determinação da alcalinidade.

A alcalinidade é dada pela expressão:

$$\text{Alcalinidade (mg}_{CaCO_3}/dm^3) = \frac{A \times N \times 50000}{V}$$

onde A denota o volume de titulante expresso em cm³, N denota a concentração ácido sulfúrico expressa na normalidade e V denota o volume de amostra expresso em cm³.

4.2.7 – Determinação da Dureza



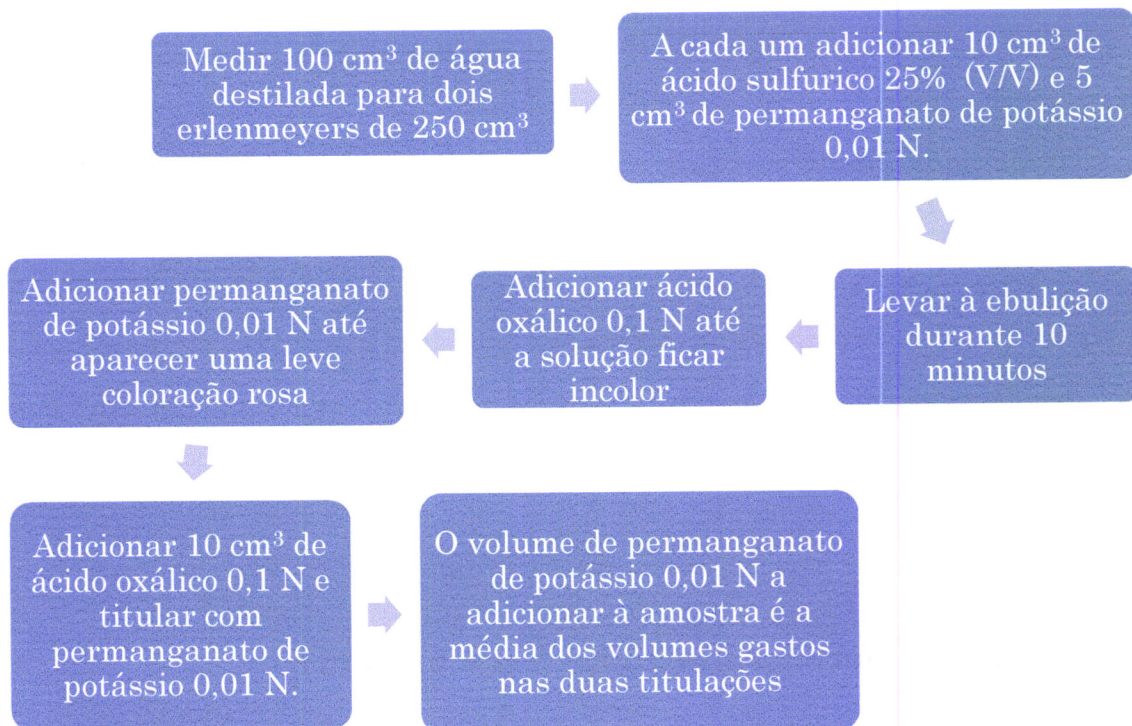
Procedimento Analítico 7: Determinação da dureza.

A dureza é determinada pela a expressão:

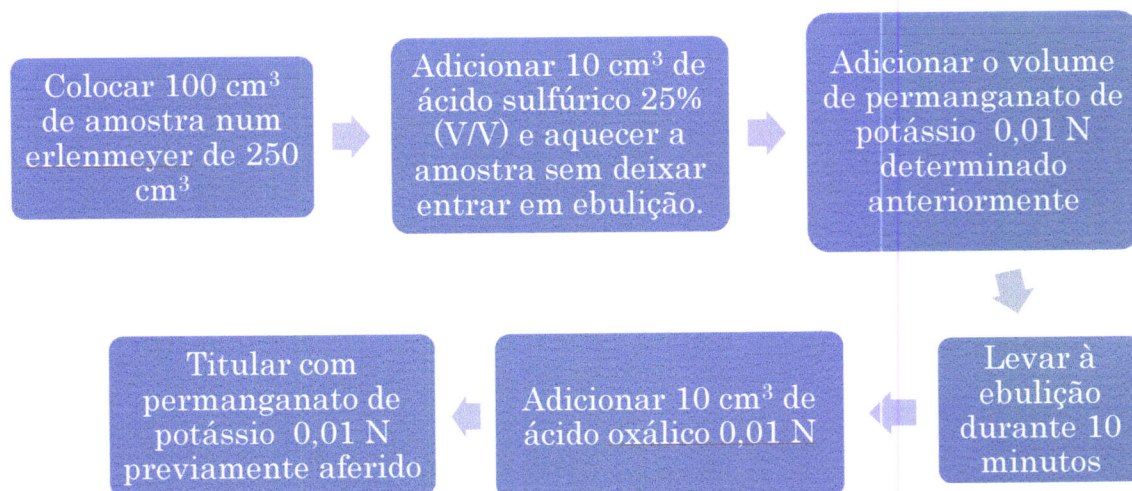
$$Dureza (mg_{CaCO_3}/dm^3) = 1000 \times \frac{V'}{V}$$

onde V denota o volume de amostra expresso em cm³ e V' denota o volume de titulante expresso em cm³.

4.2.8 – Determinação da Oxidabilidade



Procedimento Analítico 8: Determinação do título de permanganato de potássio.



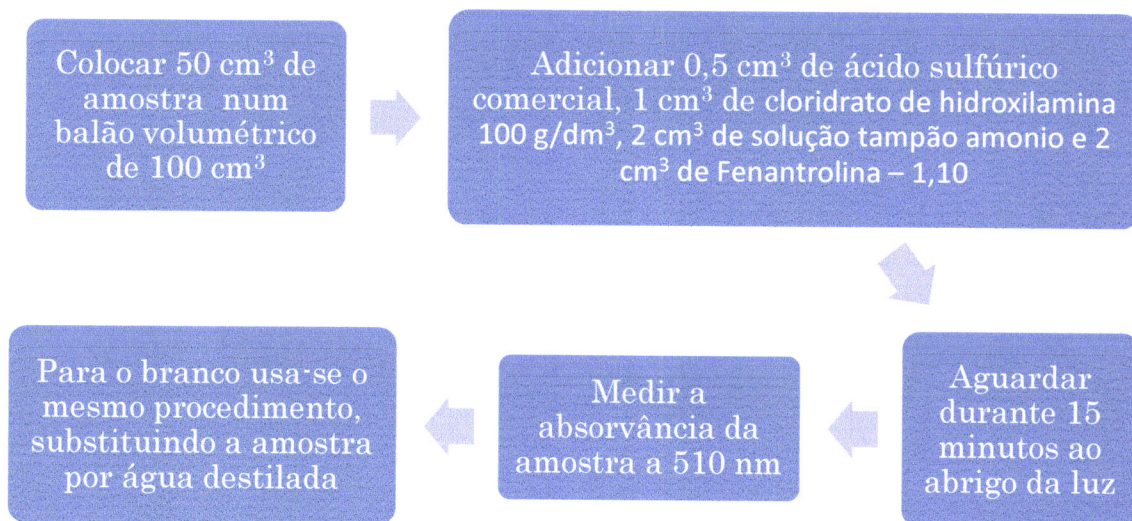
Procedimento Analítico 9: Determinação da oxidabilidade.

A oxidabilidade ao permanganato de potássio é obtida utilizando a expressão:

$$\text{Oxidabilidade ao Permanganato de Potássio (mg O}_2\text{/dm}^3\text{)} = 8 \times \frac{V_i}{V} \times f$$

onde V_i denota o volume de titulante expresso em cm^3 , V denota o volume de solução de permanganato de potássio expresso em cm^3 e f denota o factor de diluição da amostra.

4.2.9 – Determinação do Teor em Ferro



Procedimento Analítico 10: Determinação do teor em ferro.

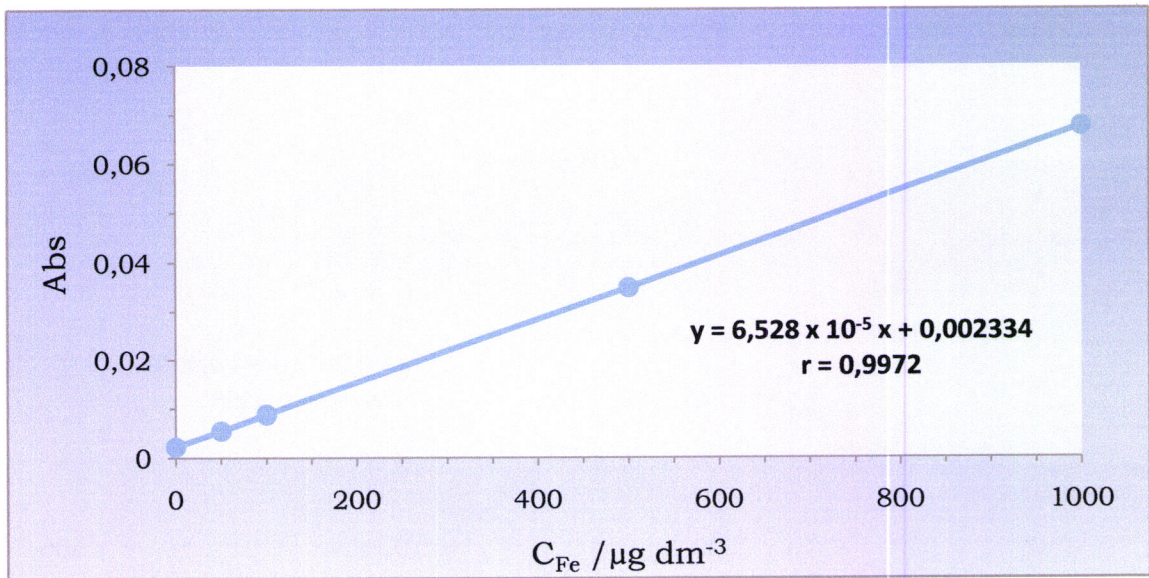
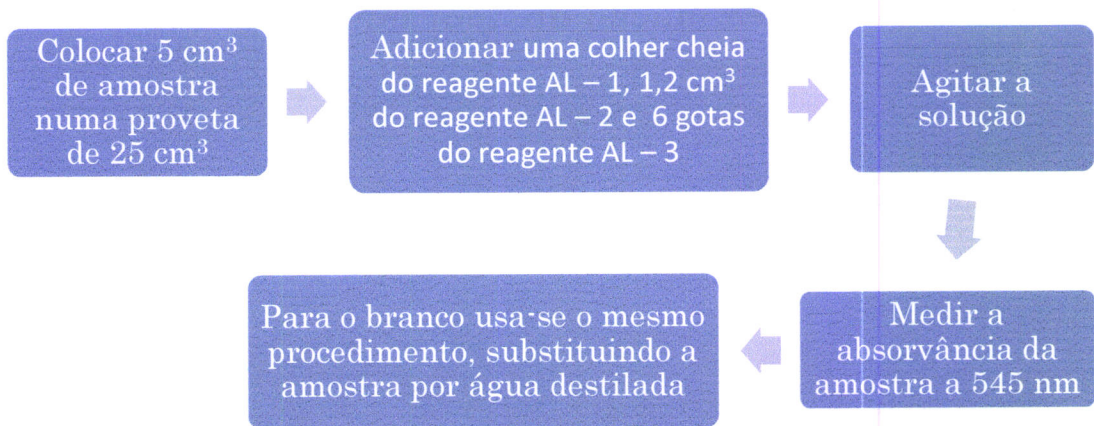


Figura 35: Curva de calibração para a determinação do teor em ferro.

4.2.10 – Determinação do Teor em Alumínio



Procedimento Analítico 11: Determinação do teor em alumínio.

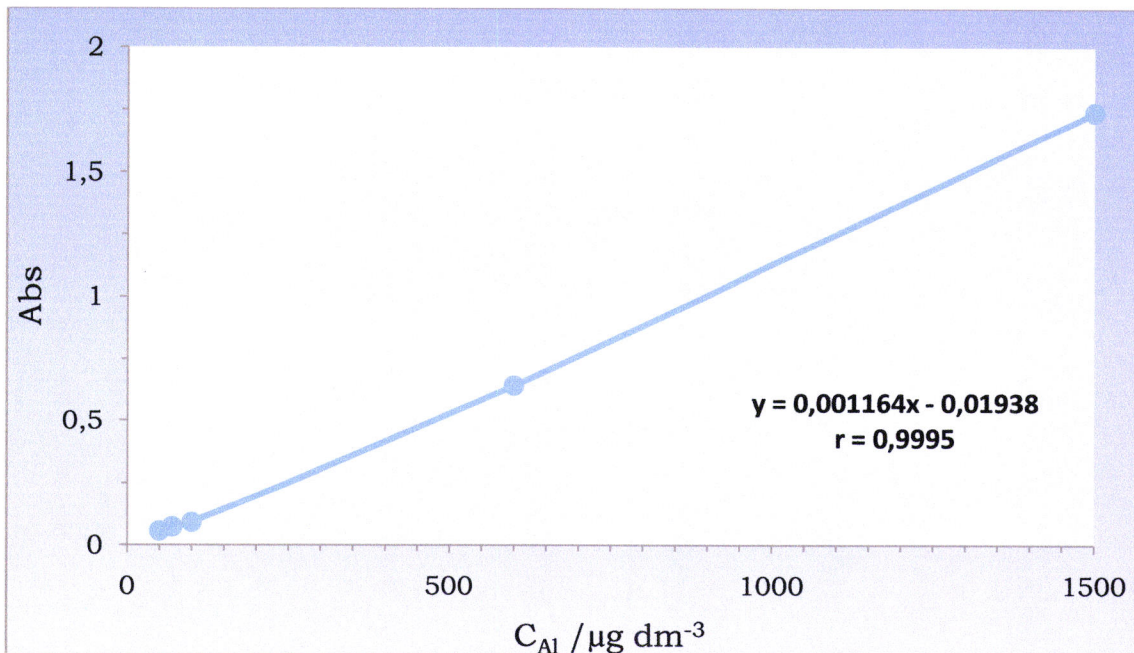
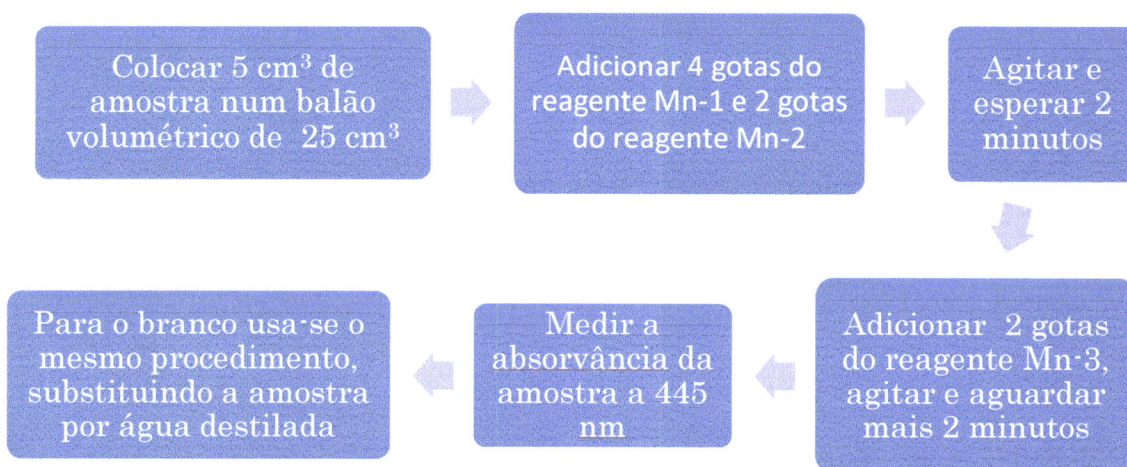


Figura 36: Curva de calibração para a determinação do teor em alumínio.

4.2.11 – Determinação do Teor em Manganês



Procedimento Analítico 12: Determinação do teor em manganês.

4.2.11.2 - CURVA DE CALIBRAÇÃO

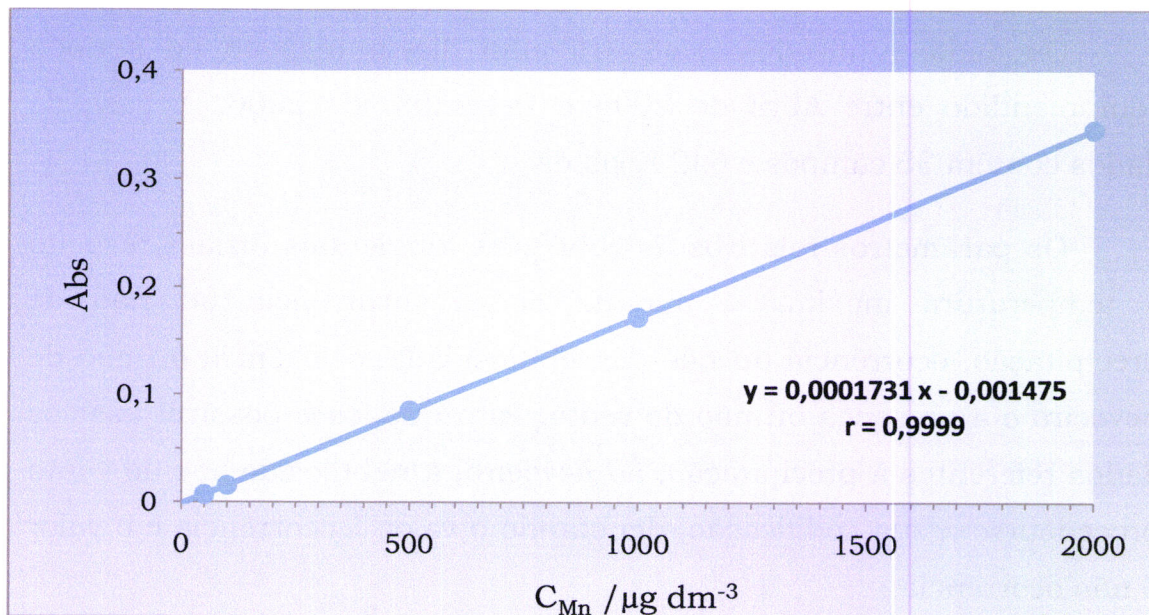



Figura 37: Curva de calibração para determinação do teor em manganês.

Quadro 6: Equipamento utilizado na determinação do teor em ferro, alumínio e manganês

Equipamento	Marca	Observações
	Hélios Alpha	<p>Fabricante: Thermo Electron Corporation</p> <p>País/Fabrico: Inglaterra</p> <p>Nº de série: UVA 131724</p>

4.3 – Base de Dados

Os dados utilizados neste trabalho dizem respeito ao período compreendido entre Abril de 2005 e Dezembro de 2006. A base de dados contém 35 campos e 642 registos.

Os parâmetros relativos às condições ambientais diziam respeito à temperatura máxima e mínima do ar, ocorrência ou não de precipitação, ocorrência ou não de nebulosidade, ocorrência ou não de nevoeiro e a presença ou não de vento. Para que fosse possível usar os dados referentes à precipitação, ao nevoeiro, à nebulosidade e ao vento procedeu-se à sua codificação, denotando o valor 1 ocorrência e o valor 0 não ocorrência.

Os dados em falta foram substituídos pelo valor 99999 para agilizar o pré-processamento dos dados. Na Figura 38 mostra-se um excerto da BD. Foi construída utilizando o *Microsoft Excel* e guardado em formato *txt*.

	A	B	C	D	Q	R	S	T	U	V	W	X	Y
1	Dia	Mês	Ano	Alc1	P9	Te9	Tu9	Al9	CIL9	CIC9	Mn9	O9	C9
105	1	9	2005	130,0	7,62	24,2	0,0	64,81	0,30	0,30	2,25	4,59	513
106	2	9	2005	150,0	7,87	23,6	0,0	56,12	0,49	0,34	18,82	4,08	514
107	5	9	2005	147,5	7,57	24,7	0,0	38,39	vest.	0,29	45,32	4,47	539
108	6	9	2005	147,5	7,62	22,6	0,0	40,31	0,33	0,36	5,57	4,39	510
109	7	9	2005	145,0	7,58	23,7	1,3	92,36	vest.	0,39	22,13	4,16	509
110	8	9	2005	147,5	7,62	22,2	0,2	49,68	0,45	0,40	0,00	4,20	513
111	9	9	2005	140,0	7,64	22,2	0,8	71,36	0,45	0,29	23,45	4,31	507
112	12	9	2005	145,0	7,45	23,1	0,3	56,12	0,51	0,38	30,08	4,39	511
113	13	9	2005	145,0	7,31	22,3	0,3	28,01	0,78	0,30	34,72	4,16	508
114	14	9	2005	145,0	7,96	22,5	0,0	39,41	0,46	0,29	7,55	4,16	518
115	15	9	2005	145,0	7,45	19,4	0,2	16,15	0,41	0,30	0,00	4,23	512
116	16	9	2005	147,5	7,40	22,4	0,0	46,30	0,48	0,34	0,00	3,88	514
117	19	9	2005	102,5	7,67	21,7	0,0	39,75	0,48	0,36	36,70	4,39	521
118	20	9	2005	135,0	7,42	22,0	1,2	96,99	0,22	0,36	6,23	4,74	520
119	21	9	2005	140,0	7,39	22,0	0,0	68,20	0,49	0,24	52,60	4,63	517
120	22	9	2005	130,0	7,60	22,6	0,0	64,14	0,70	0,32	0,00	4,50	522
121	23	9	2005	152,5	7,30	22,6	0,0	67,75	0,14	0,28	12,19	4,63	507
122	26	9	2005	147,5	7,34	20,5	0,0	82,99	0,34	0,32	32,06	4,55	511
123	27	9	2005	127,5	7,62	22,0	0,0	64,48	0,45	0,29	14,18	5,41	519
124	28	9	2005	147,5	7,40	19,8	0,0	79,27	vest.	0,36	0,00	5,33	520
125	29	9	2005	152,5	7,40	22,3	0,0	94,62	0,58	0,24	28,09	4,02	514
126	30	9	2005	152,5	7,80	22,6	0,0	58,04	vest.	0,32	6,89	4,33	518

Figura 38: Excerto da base de dados usada.

Capítulo 5

Apresentação, **I**nterpretação e **A**nálise de **R**esultados e **M**odelos

Apresentam-se, interpretam-se e avaliam-se os resultados obtidos e os modelos desenvolvidos. Justificam-se as opções tomadas no Processo de Descoberta de Conhecimento em Base de Dados.

O processo de DCBD, como foi referido anteriormente, comporta várias etapas, como se pode observar na Figura 31. Neste trabalho utilizou-se a aplicação *Microsoft Excel* na fase de pré-processamento dos dados e a aplicação *Weka* na fase de modelação. Esta última disponibiliza diversas técnicas de aprendizagem automática para problemas de classificação e de segmentação.

5.1 – Pré-Processamento e Transformação dos Dados

A forma como os dados estão estruturados condiciona o sucesso e a prossecução de uma análise inteligente dos mesmos. As técnicas de pré-processamento de dados têm por objectivo a melhoria da qualidade dos dados a utilizar, contribuindo para que a fase de *Data Mining* seja efectuada da forma mais eficiente e precisa.

Um dos problemas mais comuns associados a uma BD prende-se com a existência de dados desconhecidos e/ou incompletos e/ou com ruído. Existem várias técnicas de lidar com atributos que apresentem valores omissos/desconhecidos. As técnicas mais comuns consistem ou em eliminar esses registos ou em substituir cada valor omissos pelo valor mais comum do atributo (i.e. a moda), ou pelo valor médio ou pela mediana. A escolha da técnica adequada resulta da análise do problema em estudo [Pyle, 1999] [Han & Kamber, 2006] [Bramer, 2007]. No caso do presente trabalho, atendendo ao elevado número de registos, optou-se por descartar os que se apresentavam incompletos.

No estudo da influência das características da água captada sobre o teor em manganês presente na água tratada verificou-se, em seguida, como se comportava o manganês existente na água captada ao longo do período em estudo, Figura 39.

Exceptuando o período compreendido entre Abril e Maio de 2006, a concentração de manganês na água captada raramente ultrapassa $800 \mu\text{g}/\text{dm}^3$.

Durante o período referido, os valores anormalmente elevados foram devidos a um problema ocorrido ao nível da captação, resolvido em Junho do mesmo ano. Deste modo optou-se por descartar os registos cujo valor de manganês na água captada fosse superior a $800 \mu\text{g}/\text{dm}^3$.

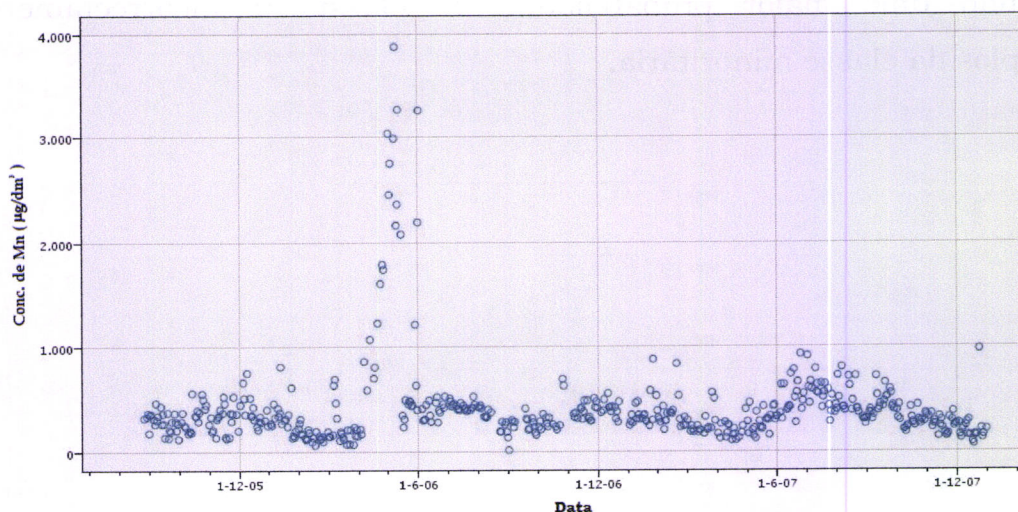


Figura 39: Evolução do teor em manganês na água captada.

Após se ter descartado os registos incompletos e os referentes à situação anómala atrás descrita, fez-se a segmentação dos dados tendo em conta o teor em manganês presente na água tratada. Deste modo constituíram-se duas classes NP e P, com base no valor paramétrico indicado no Decreto-Lei 306/07, de 27 de Agosto (i.e. $50 \mu\text{g dm}^{-3}$), onde NP e P denotam, respectivamente teores de manganês na água tratada não permitidos (i.e. superiores a $50 \mu\text{g dm}^{-3}$) e permitidos (i.e. não superiores a $50 \mu\text{g dm}^{-3}$).

Na Figura 40 apresenta-se a distribuição dos valores por classe para o parâmetro manganês.

A análise da distribuição mostra que existe desproporção entre o número de casos de cada classe. Este desequilíbrio, de acordo com Weiss e Provost [Weiss & Provost, 2001], pode originar modelos com pior desempenho na previsão das classes minoritárias. As razões para esta situação são duas. A primeira deriva do facto das regras geradas para a classe minoritária serem baseadas em menos exemplos e, consequentemente, mais sobre-ajustadas, i.e. o classificador tenderá a aprender limites mais rígidos do conceito. A segunda razão tem a ver com o facto de a classe mais frequentemente prevista ser a maioritária,

existindo uma maior probabilidade de classificar incorrectamente exemplos da classe minoritária.

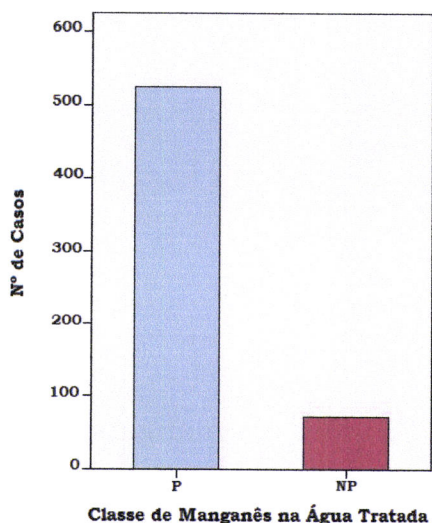


Figura 40: Distribuição do número de casos por classe de qualidade para o manganês na água tratada.

Existem, fundamentalmente, dois métodos para tornar a distribuição de classes mais equilibrada. Um dos métodos, denominado *under-sampling*, consiste em criar uma amostra mais pequena do conjunto de exemplos da classe maioritária. O outro, designado por *over-sampling*, consiste em gerar casos a partir dos casos iniciais do conjunto, de forma a aumentar o número de casos da(s) classe(s) minoritária(s). Estes dois métodos têm associadas desvantagens, como o desprezar de dados potencialmente úteis no primeiro método e, no segundo, o aumento do tamanho do conjunto de treino e, conseqüentemente, o tempo de computação, e poder proporcionar sobre-ajustamento aos dados. Este balanceamento acontece apenas no conjunto de treino, devendo ser respeitada a distribuição original no conjunto de teste [Pyle, 1999] [Han & Kamber, 2006] [Bramer, 2007].

No presente trabalho, atendendo ao número de exemplos pertencentes a cada uma das classes, para não correr o risco de descartar casos potencialmente úteis, optou-se pelo método de *over-sampling*.

Outro aspecto importante, que deve ser atendido durante a fase de pré-processamento e transformação dos dados, prende-se com o escalonamento dos dados. Este procedimento visa transformar os dados de modo a acelerar e a melhorar o processo de aprendizagem. Por exemplo, se o atributo $X \in [5, 9]$ e o atributo $Y \in [9500, 18000]$ o algoritmo de aprendizagem pode atribuir uma importância inadequada ao atributo Y em consequência dele apresentar um domínio mais alargado [Pyle, 1999] [Han & Kamber, 2006] [Bramer, 2007].

Neste trabalho os dados foram escalonados para o domínio $[0, 1]$ tendo sido utilizada a equação:

$$X_{esc} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

onde, X , X_{esc} , X_{max} e X_{min} denotam, respectivamente, o valor do atributo, o valor do atributo escalonado para o domínio $[0, 1]$, o valor máximo que o atributo apresenta e o valor mínimo que o atributo exhibe.

No estudo exploratório do comportamento da turvação presente na água captada começou-se por descartar os registos que apresentavam dados omissos no que respeita à turvação, ao pH e os teores em manganês e em ferro. Em seguida, fez-se a segmentação dos dados tendo em conta a turvação presente na água tratada. Deste modo constituíram-se duas classes NP e P, com base no valor paramétrico indicado no Decreto-Lei 306/07, de 27 de Agosto (i.e. 4 NTU), onde NP e P denotam, respectivamente turvação não permitida (i.e. superiores a 4 NTU) e permitida (i.e. não superiores a 4 NTU).

5.2 – Modelação

A previsão do teor em manganês na água tratada foi definida com um problema de classificação. A indução de regras e as árvores de decisão são técnicas de *Data Mining* mais utilizadas neste tipo de problemas, permitindo uma representação expressiva do conhecimento, de fácil utilização e compreensão.

No presente trabalho utilizou-se o algoritmo J.48, versão do algoritmo C4.5, disponibilizado pelo *software* do domínio público *Weka* para a indução de árvores de decisão.

Cada simulação foi executada 10 vezes de forma a assegurar significância estatística dos resultados obtidos. Em cada uma delas os dados disponíveis foram particionados de forma aleatória, em dois conjuntos mutuamente exclusivos (i.e. o conjunto de treino, constituído por 2/3 dos exemplos e o conjunto de teste formado pelos restantes exemplos).

O estudo exploratório do comportamento da turvação presente na água captada foi definido como sendo um problema de segmentação. A aproximação de vizinhanças é uma das técnicas mais utilizadas em tarefas deste tipo, tendo-se optado pela estratégia *k-means* e utilizado o *software* de domínio público *Weka*.

5.3 – Apresentação, Interpretação e

Avaliação dos Resultados e dos Modelos

A empresa Águas do Centro Alentejo iniciou a actividade de exploração da ETA do Monte Novo em Abril de 2005. No início realizou acções de reabilitação da ETA, a fim de solucionar problemas antigos como é o caso de aparecimento de manganês, em certas alturas do ano, na água tratada. Para tentar resolver definitivamente o problema, foram

cridas zonas de oxidação específicas, designadas por **Câmaras de Desmanganização**, onde é adicionada à água em agitação, solução de permanganato de potássio e, em certas alturas do ano, hidróxido de cálcio, para facilitar a remoção do manganês e do ferro.

A Figura 41 mostra a variação do teor em manganês ao longo do tempo quer na água captada, quer na água tratada.

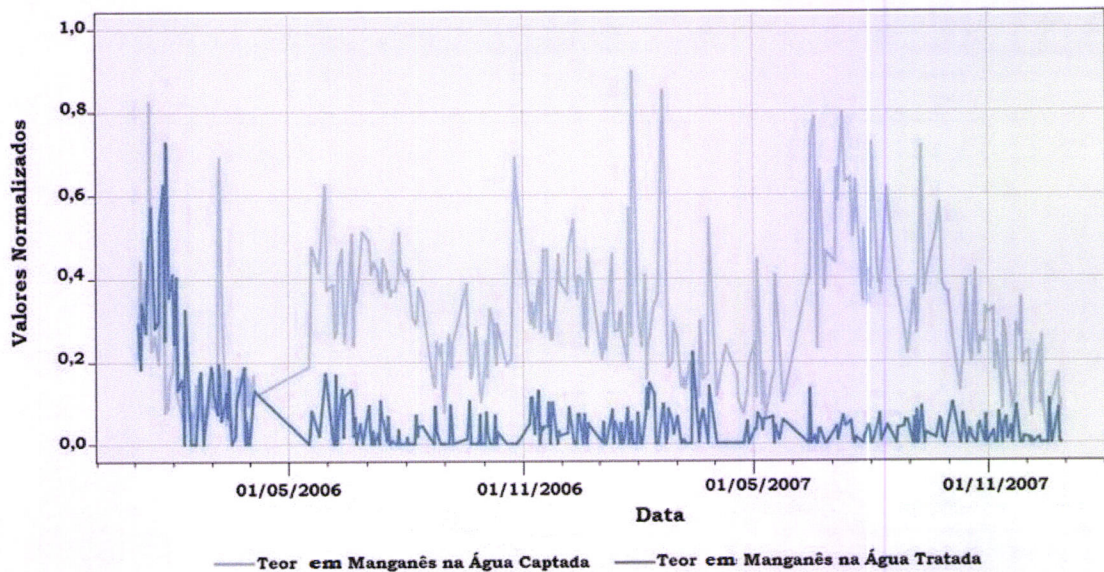


Figura 41: Variação do teor em manganês na água captada e na água tratada ao longo do tempo.

O teor de manganês na água tratada varia de forma semelhante e apresenta valores próximos do teor em manganês na água captada até Maio de 2006. A partir desta data o teor em manganês na água tratada é quase sempre inferior ao apresentado pela água captada, mantendo-se a abaixo do valor paramétrico.

A Figura 41 mostra ainda que no período compreendido entre Novembro de 2006 e Maio de 2007 o teor em manganês, quer na água tratada, quer na água captada aumentou. Contudo, na água tratada, os valores exibidos não excederam o valor paramétrico.

A Figura 42 mostra que as épocas do ano onde é mais elevado o teor em manganês na água tratada correspondem ao Outono e ao início do Inverno. Deste modo, para prevenir eventuais problemas, neste período deverá ser adicionado hidróxido de cálcio nas câmaras de desmanganização. O período correspondente à Primavera e ao Verão é menos crítico, não sendo de esperar, em situação normal, a existência de problemas, não devendo ser necessário hidróxido de cálcio nas câmaras de desmanganização.

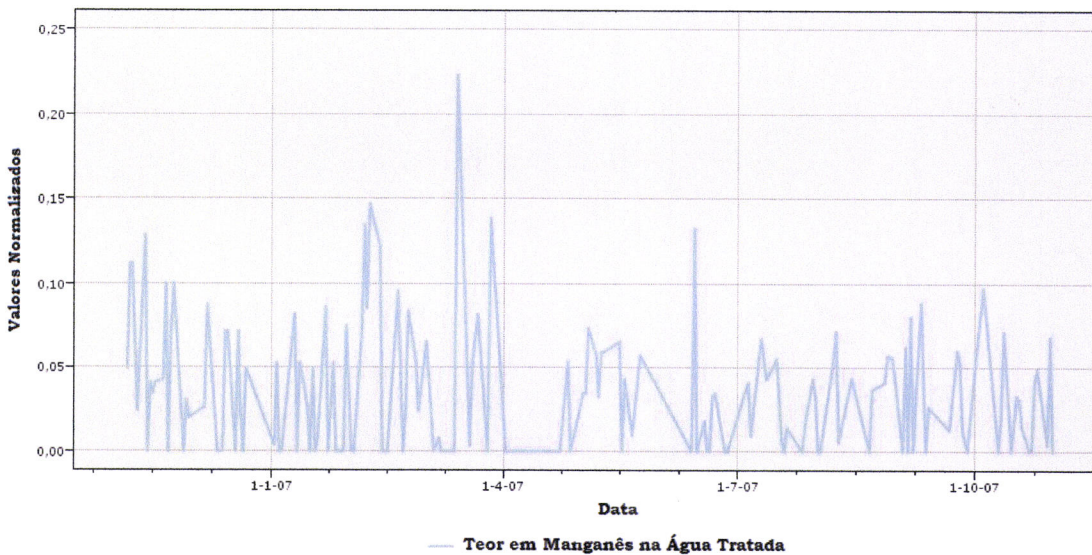


Figura 42: Variação do teor em manganês na água tratada após a entrada em funcionamento das câmaras de desmanganização.

A Figura 43 e a Figura 44 mostram que, na albufeira, a variação dos teores em manganês e em ferro é oposta à variação do teor em oxigénio dissolvido, i.e. quando o teor em oxigénio dissolvido é elevado os teores em manganês e em ferro são baixos.

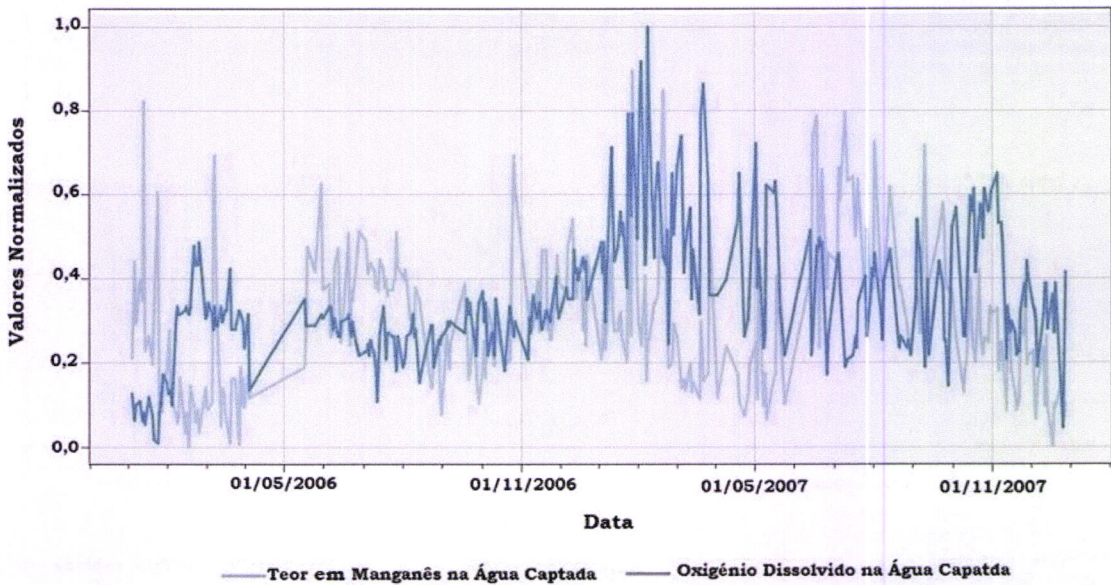


Figura 43: Variação dos teores em manganês e em oxigênio dissolvido, ao longo do tempo na água captada.

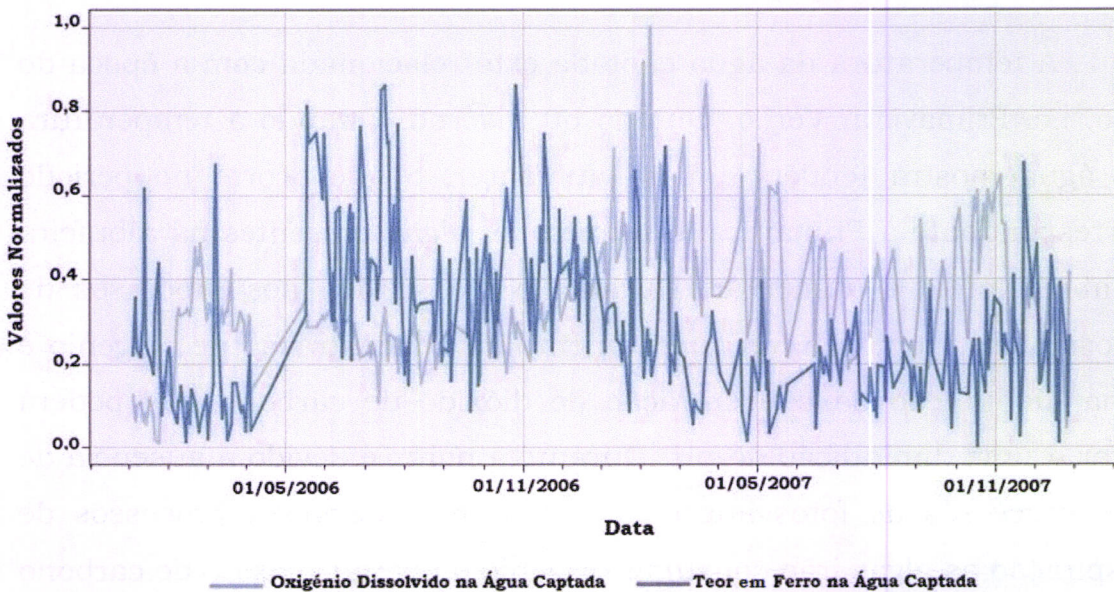


Figura 44: Variação dos teores em ferro e em oxigênio dissolvido, ao longo do tempo na água captada.

A Figura 45 mostra que a variação do pH e a variação da temperatura, em geral, se sobrepõe, enquanto o oxigênio dissolvido tem um comportamento oposto.

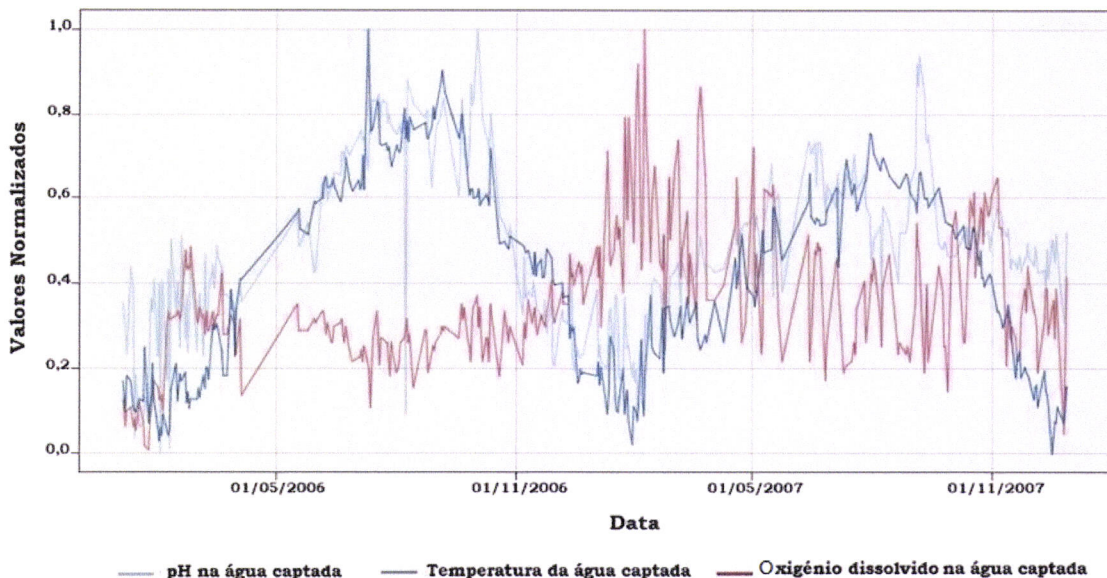


Figura 45: Variação da temperatura, do pH e do oxigênio dissolvido na água captada, ao longo do tempo.

A temperatura da água captada está relacionada com a época do ano, i.e. Primavera, Verão, Outono ou Inverno. Quando a temperatura da água mostra tendência para aumentar, o que ocorre no período correspondente à Primavera, o número de algas presentes na albufeira também tende a aumentar. Durante o dia, devido aos processos de fotossíntese, poderá haver um aumento da concentração de oxigênio e uma diminuição da concentração de dióxido de carbono que poderá levar a uma diminuição do pH. Durante a noite, e devido à ausência de luz, processos de fotossíntese não ocorrem e devido a processos de respiração as algas irão consumir oxigênio e libertar dióxido de carbono que poderá conduzir a um aumento de pH [Sturm & Morgan, 1996].

Para além do consumo de oxigênio provocado pela respiração há que ainda ter em conta o consumo de oxigênio devido a processos de degradação das algas por intermédio de microrganismos, uma vez que há um aumento da matéria orgânica devido ao facto destas possuírem tempos de vida muito curtos [Sturm & Morgan, 1996].

Um outro factor que poderá influenciar os valores de pH e de oxigénio dissolvido prende-se com o facto da colheita da amostra ser realizada logo ao início da manhã. Está-se na transição da noite para o dia e, pelas razões expostas, os níveis de oxigénio dissolvido encontram-se mais baixos e o pH mais elevado.

A Figura 46 e a Figura 47 mostram que no período de Junho a Novembro os teores em manganês e em ferro são elevados e a turvação é baixa. Esta situação poderá indicar, que nesta altura do ano, quer o manganês quer o ferro presentes na água captada se encontram na forma livre.

No período de Novembro a Junho verifica-se que a variação da turvação e a variação dos teores em manganês e em ferro são semelhantes, apresentando a turvação valores relativamente elevados. Estes factos poderão indicar que o manganês e o ferro presentes na água captada estão precipitados sob a forma de óxidos e/ou de hidróxidos, facilitando a sua remoção da água durante o processo de tratamento.

Estes resultados estão em consonância com os apresentados anteriormente, que apontam o período correspondente ao Outono e ao início do Inverno como o mais problemático no que diz respeito à presença de manganês na água tratada. Neste período o pH da água captada apresenta tendência a baixar, Figura 45, tornando-se necessário adicionar hidróxido de cálcio na câmara de desmanganização para aumentar o pH da água e, deste modo, precipitar o manganês e o ferro criando condições para a sua remoção. No período correspondente à Primavera e ao Verão o pH da água captada apresenta tendência a aumentar, Figura 45, estando criadas as condições para a precipitação do manganês e do ferro.



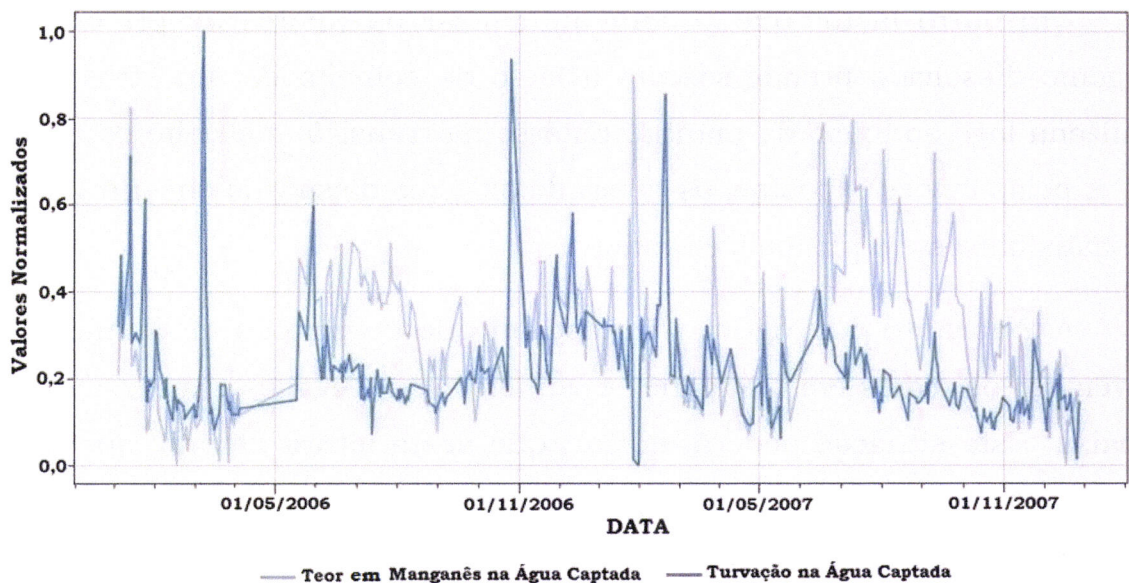


Figura 46: Variação do teor em manganês e da turvação na água captada, ao longo do tempo.

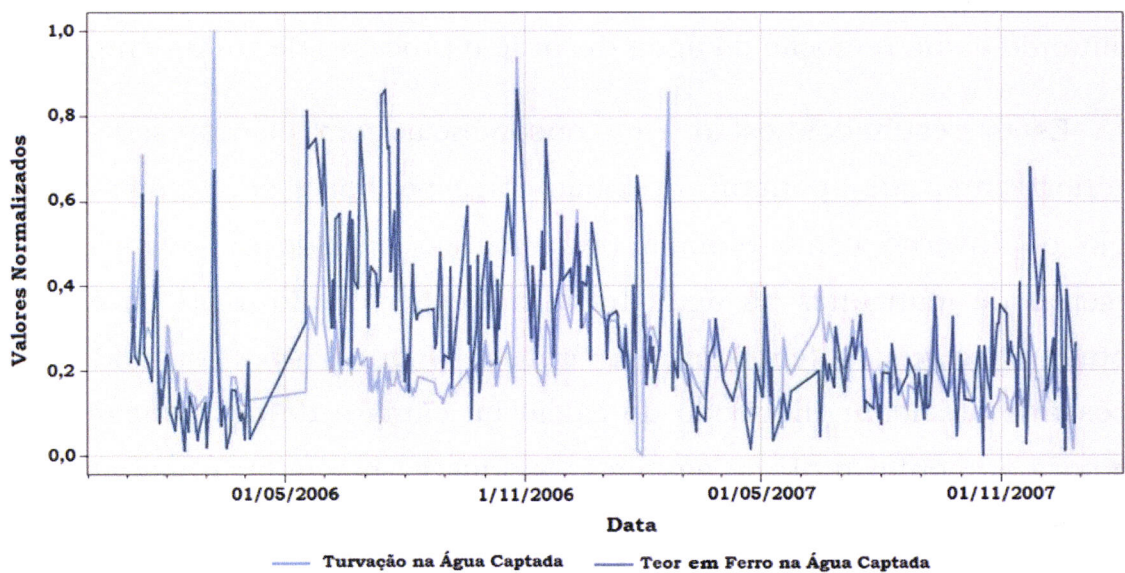


Figura 47: Variação do teor em ferro e da turvação na água captada, ao longo do tempo.

A Figura 48 mostra, para água captada, a variação da turvação, do pH e a temperatura. Verifica-se que no geral, a turvação varia de modo inverso relativamente ao pH e à temperatura.

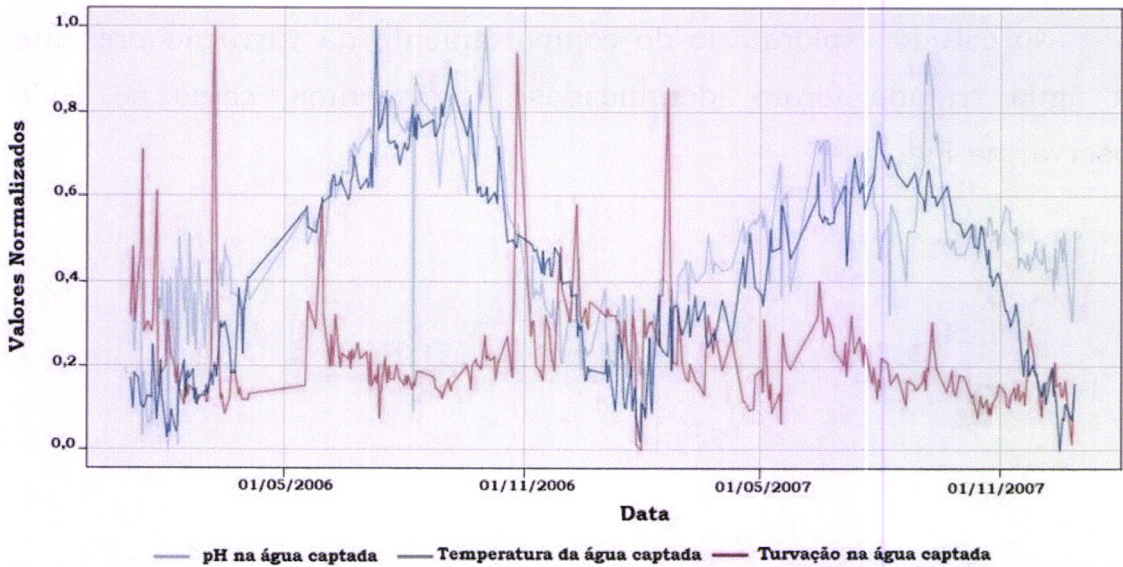


Figura 48: Variação do pH, da turvação e da temperatura na água captada, ao longo do tempo.

Este comportamento poderá ser devido a movimentos de convecção na massa de água. No Outono/Inverno a temperatura do ar diminui, o que provoca uma diminuição da temperatura da água à superfície. Esta torna-se mais densa relativamente à do fundo, originando movimentos de convecção e o transporte de sedimentos que se vão influenciar os valores de turvação na água captada. Quando a temperatura do ar aumenta não estão reunidas as condições para ocorrer movimentos de convecção, a matéria em suspensão tem tendência a sedimentar e os valores de turvação são baixos.

As variações ao longo do tempo exibidas pela turvação e pelo pH parecem mostrar que, no Outono/Inverno, a principal causa para a turvação exibida pela água captada está relacionada com o transporte de sólidos uma vez que, como foi referido anteriormente, não estão

reunidas as condições para a precipitação do ferro e do manganês. Seria interessante no futuro confrontar estes resultados meteorológicos, principalmente de precipitação, no sentido de tentar perceber quando é que a turvação é devida à convecção e quando é devida ao transporte pelas águas pluviais.

No estudo exploratório do comportamento da turvação presente na água tratada foram identificados 4 segmentos, como se pode observar na Figura 49.

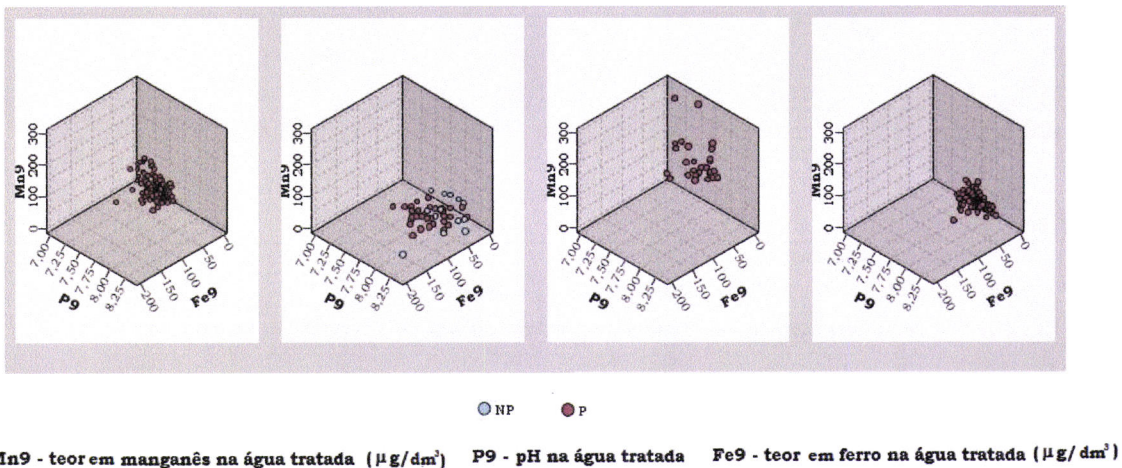


Figura 49: Segmentos identificados no estudo do comportamento da turvação presente na água tratada.

Os segmentos 1, 3 e 4 são constituídos maioritariamente por casos em que a turvação pertence à classe P. O segmento 1 é caracterizado por valores de pH baixos, inferiores a 7,5 e em teores em manganês e ferro igualmente baixos, inferiores $50 \mu\text{g dm}^{-3}$. Por sua vez, o segmento 3 é caracterizado por valores de pH baixos inferiores a 7,8 e por teores em ferro muito baixos. O teor em manganês neste segmento é muito variável. O segmento 4, por seu lado, é caracterizado por valores de pH intermédios, compreendidos entre 7,3 e 8,1 e em teores em manganês e ferro baixos, inferiores $50 \mu\text{g dm}^{-3}$.

O segmento 2 contém a grande maioria dos casos em que a turvação pertence à classe NP. É caracterizado por apresentar teores em

manganês baixos, inferiores $50 \mu\text{g dm}^{-3}$ enquanto o pH e o teor em ferro são variáveis.

Estes resultados mostram a influência do pH sobre a turvação da água tratada e sublinham a importância de ajustar o pH da água no final do processo de tratamento. Por um lado o pH deve ser suficientemente elevado para garantir que a água seja incrustante e, por outro lado, o pH deverá ser suficientemente baixo para garantir que a turvação não exceda o valor paramétrico.

Seria interessante no futuro, para cada um dos segmentos identificados, desenvolver modelos de classificação e/ou de previsão para verificar em que condições a turvação apresenta valores acima do valor paramétrico. Um outro aspecto a ter em conta é a avaliar se a adição de cloro como desinfectante influencia de algum modo a turvação.

A previsão do teor em manganês na água tratada foi definida como um problema de classificação tendo-se recorrido à indução de regras e às árvores de decisão. Em seguida passa-se à apresentação e avaliação dos modelos obtidos, utilizando-se para o efeito a matriz de coincidências [Kohavi & Provost, 1998]. Trata-se de uma matriz $L \times L$, onde L denota o número de classes possíveis. A matriz de coincidências avalia o resultado de uma classificação, mapeando os valores previstos pelo modelo com os valores desejados.

Os modelos obtidos para a previsão do teor em manganês na água tratada foram bastante simples o que denota que o conceito foi facilmente aprendido. As acuidades dos modelos, medidas em termos da percentagem de acertos relativamente ao número de casos apresentados, foram elevadas. Em todos eles, a percentagem de acerto relativo quer para o conjunto de treino, quer para o conjunto de teste, em cada uma das classes, é semelhante. Deste modo, a capacidade de generalização dos modelos é boa, não havendo indícios de sobre-ajustamento (i.e. a percentagem de acertos ser muito menor para o conjunto de teste face ao conjunto de treino).

Quando foi fornecido ao algoritmo todos os parâmetros disponíveis na BD obteve-se a árvore de decisão que é mostrada na Figura 50, modelo 1.

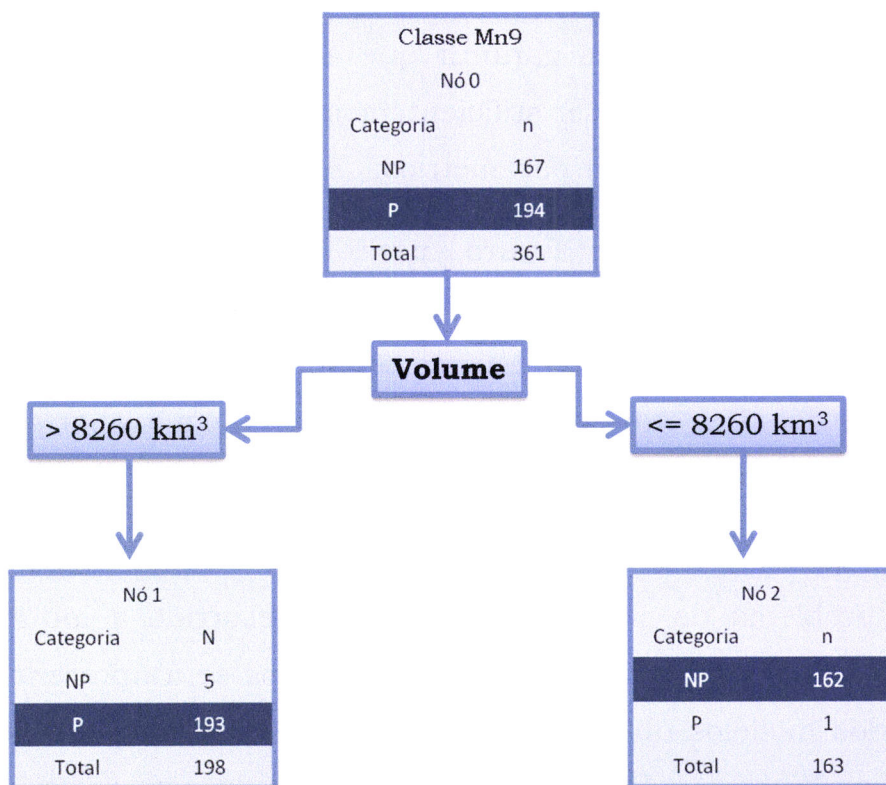


Figura 50: Árvore de decisão obtida para o modelo 1.

A análise do modelo 1 mostra que o volume é suficiente para classificar os exemplos. A acuidade do modelo é de 98,34% (355 acertos/361 casos) para o conjunto de treino e de 97,16% (137 acertos/141 casos) para o conjunto de teste, Tabela 6.

Tabela 6: Matriz de coincidência do modelo 1.

Classe	Conjunto de treino		Conjunto Teste	
	NP	P	NP	P
NP	162	5	45	2
P	1	193	2	92

O conjunto de regras gerado a partir deste modelo apresenta-se em seguida, destacando-se em cada uma delas o suporte (i.e. número de casos onde a regra é encontrada) e a confiança (i.e. a probabilidade condicional dessa regra) que, como se pode verificar são elevados.

Regras para a classe NP:

Regra 1 (198; 0,975)

se VB \leq 8260

então NP

Regras para a classe P:

Regra 1 (163; 0,994)

se VB $>$ 8260

então P

O passo seguinte foi impedir que o algoritmo tivesse acesso ao campo “volume”. Obteve-se a árvore de decisão que é mostrada na Figura 51, modelo 2.

A análise do modelo 2 mostra que a cota é suficiente para classificar os exemplos. A acuidade do modelo é de 98,35% (357 acertos/363 casos) para o conjunto de treino e de 97,16% (137 acertos/141 casos) para o conjunto de teste, Tabela 7.

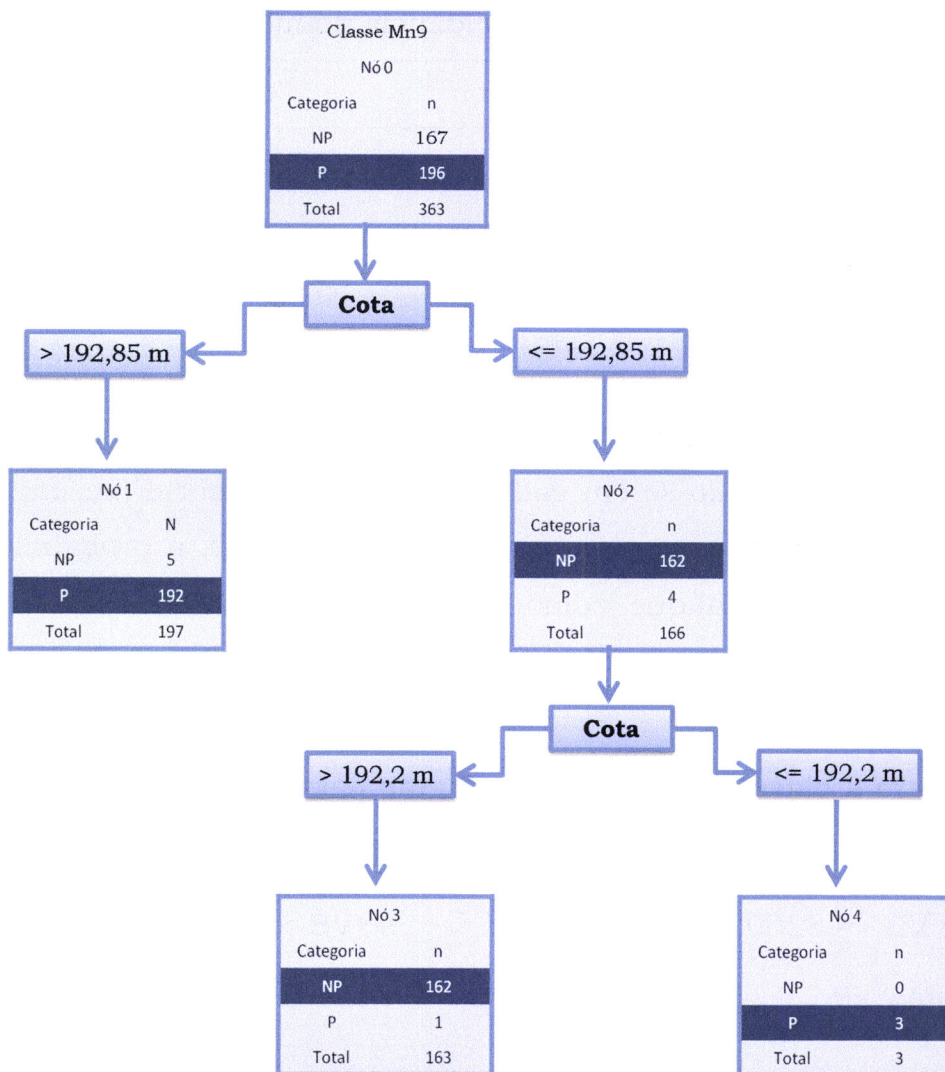


Figura 51: Árvore de decisão obtida para o modelo 2.

Tabela 7: Matriz de coincidência do modelo 2.

Classe	Conjunto de treino		Conjunto Teste	
	NP	P	NP	P
NP	162	5	45	2
P	1	195	2	92

O conjunto de regras gerado a partir modelo 2 apresenta-se em seguida, destacando-se em cada uma delas o suporte e a confiança que, como se pode verificar são elevados.

Regras para a classe NP:

Regra 1 (163; 0,994)

se $C \leq 192,850$

e $C > 192,200$

então NP

Regras para a classe P:

Regra 1 (3; 1,0)

se $C \leq 198,850$

e $C \leq 192,200$

então P

Regra 2 (197; 0,976)

se $C > 198,850$

então NP

O modelo 1 e o modelo 2 são semelhantes, uma vez existe uma correspondência entre a cota da albufeira e volume água armazenado, correspondendo, no caso da albufeira do Monte Novo, à cota 192,85 m o volume 8260 km³.

O modelo 2 prevê que quando a água é captada a cotas situadas entre 192,2 e 192,85 m, o teor em manganês na água tratada excede o limite permitido pelo Decreto-Lei 306/07, 27 de Agosto, o que não acontece se a captação estiver a ser feita acima deste intervalo.

Se atendermos ao facto que a primeira e a segunda comporta de captação se situam, respectivamente, à cota de 192,5 m e 187,6 m o modelo revela que os problemas que surgem com o teor em manganês na água tratada, podem estar relacionados com a mudança do nível de captação. É, portanto, necessário que a ETA esteja preparada para dar resposta aos problemas que podem surgir.

Foi feita uma nova tentativa de modelação não permitindo que o algoritmo tivesse acesso aos campos relativos à albufeira. O modelo obtido, designado modelo 3, encontra-se na Figura 52.

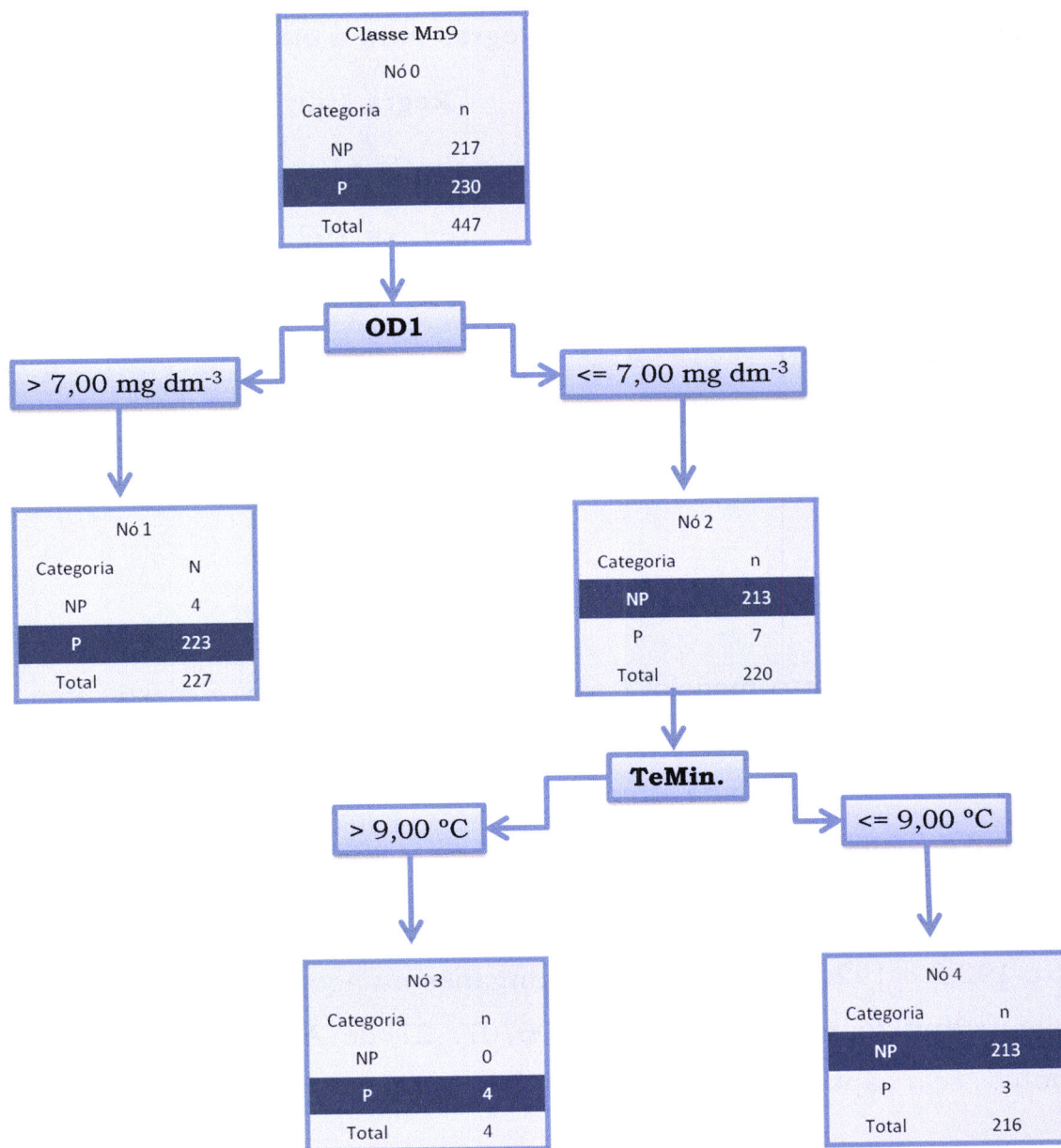


Figura 52: Árvore de decisão obtida para o modelo 3.

A análise do modelo 3 mostra que o teor em oxigénio dissolvido na água captada é o parâmetro que mais influencia a classificação, secundado pela temperatura mínima do ar.

A acuidade do modelo é de 98,43% (440 acertos/447 casos) para o conjunto de treino e de 96,13% (149 acertos/155 casos) para o conjunto de teste, Tabela 8.

Tabela 8: Matriz de coincidência do modelo 3.

Classe	Conjunto de treino		Conjunto Teste	
	NP	P	NP	P
NP	213	4	56	4
P	3	227	2	93

O conjunto de regras gerado a partir modelo 3 apresenta-se em seguida destacando-se, tal com anteriormente, o suporte e a confiança que, como se pode verificar, são elevados.

Regras para a classe NP:

Regra 1 (216; 0,986)

se OD1 \leq 7

e TeMin. \leq 9

então NP

Regras para a classe P:

Regra 1 (4; 1,0)

se OD1 \leq 7

e TeMin. $>$ 9

então P

Regra 2 (227; 0,982)

se OD1 $>$ 7

então P

O modelo 3 mostra que quando o teor de oxigênio dissolvido na água captada é superior a 7 mg dm^{-3} não é de esperar problemas ao nível do manganês na água tratada. No caso do oxigênio dissolvido ser inferior a 7 mg dm^{-3} e a temperatura mínima do ar ser inferior a $9 \text{ }^\circ\text{C}$, o modelo prevê situações de incumprimento. Provavelmente esta situação corresponde ao período referente ao final do Outono/início do Inverno onde, como foi discutido anteriormente, ocorrem problemas ao nível de manganês na água tratada.

Foi feita uma outra tentativa de modelação, vedando o acesso aos campos relativos à albufeira e à temperatura mínima do ar. Os resultados obtidos foram semelhantes, uma vez que o algoritmo de aprendizagem utilizou a temperatura máxima do ar.

Foram ainda feitas outras tentativas, barrando o acesso aos campos relativos à albufeira e à temperatura mínima e máxima do ar. Neste caso, o algoritmo apenas utilizou o teor em oxigênio dissolvido na água captada e os resultados obtidos, foram inferiores. No caso de se impedir o acesso simultâneo aos campos relativos à albufeira e ao teor em oxigênio dissolvido, o algoritmo J.48 não foi capaz de realizar a tarefa de classificação.

Capítulo 6

Conclusões **F**inais e **P**erspectivas **F**uturas

Neste capítulo são apresentadas as principais conclusões que se podem tirar com este trabalho. É também perspectivado o trabalho que se poderá desenvolver no futuro.

Após a realização do trabalho é chegado o momento de analisar o caminho percorrido e abrir novos caminhos.

A produção de água para consumo humano, a partir de águas superficiais, é uma tarefa que lida com ecossistemas que apresentam características e dinâmicas muito próprias, sujeitos a alterações que, muitas vezes, surgem como resposta a factores externos que normalmente são difíceis de controlar.

O cumprimento das metas definidas por qualquer organização requer um comportamento pró-activo, que passa por antecipar proble-

mas de modo a estar preparado para os enfrentar, sabendo exactamente como actuar nessa contingência.

No caso da ETA do Monte Novo torna-se de grande importância antecipar os problemas relacionados com o aparecimento de manganês e de turvação na água tratada.

O presente trabalho dá uma contribuição para a resolução destes problemas tendo-se recorrido a técnicas e metodologias da Área Científica da Inteligência Artificial e da Descoberta de Conhecimentos em Bases de Dados utilizando, como ferramentas, a Indução de Regras, Árvores de Decisão e Aproximação de Vizinhanças. Desenvolveram-se modelos que contribuem para o desejável comportamento pró-activo potenciando a previsão, dando resposta aos principais desafios lançados, constituindo-se como exemplo da aplicação das ferramentas e técnicas *Data Mining*, à resolução de problemas de uma indústria.

Os modelos desenvolvidos permitiram relacionar os problemas ao nível do manganês na água tratada com a mudança do nível da captação, identificar as épocas do ano em que a ETA deverá estar preparada para responder a eventuais oscilações no teor em manganês na água tratada. Para além do exposto, foi ainda possível perceber a importância do ajuste do pH na fase final do tratamento.

No entanto, há a noção de que este é um trabalho inacabado. Para além dos desafios já mencionados no texto, outros há que merecem ser equacionados. Um deles, está relacionado como estudo do comportamento dos vários órgãos de tratamento da ETA, tentando avaliar de que forma é que as características da água no final de uma etapa de tratamento, poderá influenciar o funcionamento e/ou as condições de operação da fase seguinte. Este desafio poderá conduzir ao desenvolvimento de Sistemas de Conhecimento baseado em *Data Mining* para a optimização, em tempo real, do funcionamento de uma estação de tratamento de água.

Referências Bibliográficas

- [Abelha, 2004] Abelha, A., Sistemas Multiagente como Suporte a Trabalho Cooperativo em Unidades Hospitalares, Tese de Doutoramento, Departamento de Informática, Universidade do Minho, Braga, Portugal, 2004.
- [Adriaans & Zantinge, 1996] Adriaans, P. & Zantinge, D., Data Mining. Addison-Wesley, Edimburgo, 1996.
- [Apte et al., 2002] Apte, C., Liu, B., Pednault, E. & Smith, P., Business applications of Data Mining. *Communications of the ACM*, 45: 49-53, 2002.
- [Atkins et al., 2007] Atkins, J. P., Burdon, D. & Allen, J. H., An application of contingent valuation and decision tree analysis to water quality improvements. *Marine Pollution Bulletin*, 55: 591-602, 2007.
- [Berry & Linoff, 2000] Berry, M. & Linoff, G., Mastering Data Mining: The Art and Science of Customer Relationships Management. John Wiley & Sons, USA, 2000.
- [Berson et al., 2000] Berson, A., Smith, S. & Thearling, K., Building Data Mining Applications. Mc-Graw Hill, USA, 2000.
- [Bobbin & Recknagel, 2001] Bobbin, J. & Recknagel, F., Inducing explanatory rules for the prediction of algal blooms by genetic algorithms. *Environment International*, 27: 237-242, 2001.
- [Bramer, 2007] Bramer, M., Principles of Data Mining. Springer, Londres, 2007.
- [Chang et al., 2001] Chang, G., Healey, M., McHugh, J. & Wang, J., Mining the Word Wide Web - An Information Search Approach. Kluwer Academic Publishers, USA, 2001.

- [Chau, 2006] Chau, K.-W., A review on integration of artificial intelligence into water quality modelling. *Marine Pollution Bulletin*, 52: 726-733, 2006.
- [Chaves & Kojiri, 2007] Chaves, P. & Kojiri, T., Deriving reservoir operational strategies considering water quantity and quality objectives by stochastic fuzzy neural networks. *Advances in Water Resource*, 30: 1329-1341, 2007.
- [Chen, 2003] Chen, L., A study of applying genetic programming to reservoir trophic state evaluation using remote sensor data. *International Journal of Remote Sensing*, 24: 2265-2275, 2003.
- [Cortez, 2004] Cortez, P., Aprendizagem e Avaliação de Modelos – Apontamentos Pedagógicos. Departamento de Sistemas de Informação, Universidade do Minho, Portugal, 2004.
- [Cunha et al., 1980] Cunha, L. V., Gonçalves, A. S., Figueiredo, V. A. & Lino, M., A Gestão da Água. Princípios fundamentais e sua aplicação em Portugal. Fundação Calouste Gulbenkian, Lisboa, Portugal, 1980.
- [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, Direcção Regional do Ambiente e Ordenamento do Território, 1999.
- [Dzeroski, 2002] Dzeroski, S., Environmental Sciences. Em W. Klösgen & J. M. Zytkow Eds., *Handbook of Data Mining and Knowledge Discovery*, Vol. pp. 817-830, Oxford University Press, Oxford, U. K., 2002.
- [Eaton et al., 2005] Eaton, A., Clesceri, L., Rice, E. & Greenberg, A. Eds., *Standard Methods for the Examination of Water and*

Capítulo 6

Conclusões **F**inais e **P**erspectivas **F**uturas

Neste capítulo são apresentadas as principais conclusões que se podem tirar com este trabalho. É também perspectivado o trabalho que se poderá desenvolver no futuro.

Após a realização do trabalho é chegado o momento de analisar o caminho percorrido e abrir novos caminhos.

A produção de água para consumo humano, a partir de águas superficiais, é uma tarefa que lida com ecossistemas que apresentam características e dinâmicas muito próprias, sujeitos a alterações que, muitas vezes, surgem como resposta a factores externos que normalmente são difíceis de controlar.

O cumprimento das metas definidas por qualquer organização requer um comportamento pró-activo, que passa por antecipar proble-

mas de modo a estar preparado para os enfrentar, sabendo exactamente como actuar nessa contingência.

No caso da ETA do Monte Novo torna-se de grande importância antecipar os problemas relacionados com o aparecimento de manganês e de turvação na água tratada.

O presente trabalho dá uma contribuição para a resolução destes problemas tendo-se recorrido a técnicas e metodologias da Área Científica da Inteligência Artificial e da Descoberta de Conhecimentos em Bases de Dados utilizando, como ferramentas, a Indução de Regras, Árvores de Decisão e Aproximação de Vizinhanças. Desenvolveram-se modelos que contribuem para o desejável comportamento pró-activo potenciando a previsão, dando resposta aos principais desafios lançados, constituindo-se como exemplo da aplicação das ferramentas e técnicas *Data Mining*, à resolução de problemas de uma indústria.

Os modelos desenvolvidos permitiram relacionar os problemas ao nível do manganês na água tratada com a mudança do nível da captação, identificar as épocas do ano em que a ETA deverá estar preparada para responder a eventuais oscilações no teor em manganês na água tratada. Para além do exposto, foi ainda possível perceber a importância do ajuste do pH na fase final do tratamento.

No entanto, há a noção de que este é um trabalho inacabado. Para além dos desafios já mencionados no texto, outros há que merecem ser equacionados. Um deles, está relacionado como estudo do comportamento dos vários órgãos de tratamento da ETA, tentando avaliar de que forma é que as características da água no final de uma etapa de tratamento, poderá influenciar o funcionamento e/ou as condições de operação da fase seguinte. Este desafio poderá conduzir ao desenvolvimento de Sistemas de Conhecimento baseado em *Data Mining* para a optimização, em tempo real, do funcionamento de uma estação de tratamento de água.

- Wastewater. 21st Edition Ed., American Public Health Association, USA, 2005.
- [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. & Uthurusamy, R., Advances in Knowledge Discovery and Data Mining. AAAI Press/MIT Press, Massachusetts, USA, 1996.
- [Goebel & Gruenwald, 1999] Goebel, M. & Gruenwald, L., A survey of Data Mining and knowledge discovery software tools. *ACM SIGKDD Explorations*, 1: 20-33, 1999.
- [Han & Kamber, 2006] Han, J. & Kamber, M., Data Mining: Concepts and Techniques. 2^a Ed., Morgan Kaufmann Publishers, USA, 2006.
- [Jeong et al., 2003] Jeong, K.-S., Kim, D.-K., Whigham, P. & Joo, G.-J., Modelling Microcystis aeruginosa bloom dynamics in the Nakdong River by means of evolutionary computation and statistical approach. *Ecological Modelling*, 161: 67-78, 2003.
- [Jeong et al., 2001] Jeong, K.-S., Joo, G.-J., Kim, H.-W., Ha, K. & Recknagel, F., Prediction and elucidation of phytoplankton dynamics in the Nakdong River (Korea) by means of a recurrent artificial neural network. *Ecological Modelling*, 146: 115-129, 2001.
- [Karul et al., 2000] Karul, C., Soyupak, S., Çilesiz, A., Akbay, N. & Germen, E., Case studies on the use of neural networks in eutrophication modelling. *Ecological Modelling*, 134: 145-152, 2000.

- [Kufrin, 1997] Kufrin, R., Generating C4.5 Productions Rules in Parallel; 4th National Conference on Artificial Intelligence (AAAI-97). 1997.
- [Kuo et al., 2006] Kuo, J.-T., Wang, Y.-Y. & Lung, W.-S., A hybrid neural-genetic algorithm for reservoir water quality management. *Water Research*, 40: 1367-1376, 2006.
- [Kuo et al., 2007] Kuo, J.-T., Hsieh, M.-H., Lung, W.-S. & She, N., Using artificial neural network for reservoir eutrophication prediction. *Ecological Modelling*, 200: 171-177, 2007.
- [Langley & Simon, 1995] Langley, P. & Simon, H., Applications of Machine Learning and Rule Induction. *Communications of the ACM*, 38, 1995.
- [Lee et al., 2003] Lee, J. H., Huang, Y., Dickman, M. & Jayawardena, A. W., Neural network modelling of coastal algal blooms. *Ecological Modelling*, 159: 179-210, 2003.
- [Lehr & Keeley, 2005b] Lehr, J. H. & Keeley, J. Eds., *Water Encyclopedia: Oceanography; Meteorology; Physics and Chemistry; Water Law; and Water History, Art and Culture*. 1^a Ed., John Wiley & Sons, Hoboken, New Jersey, USA, 2005a.
- [Lehr & Keeley, 2005a] Lehr, J. H. & Keeley, J. Eds., *Water Encyclopedia: Surface and Agricultural Water*. 1^a Ed., John Wiley & Sons, Hoboken, New Jersey, USA, 2005b.
- [Lek & Guégan, 1999] Lek, S. & Guégan, J. F., Artificial neural network as a tool in ecological modelling, an introduction. *Ecological Modelling*, 120: 65-73, 1999.
- [Liu et al., 2007] Liu, S., Butler, D., Brazier, R., Heathwaite, L. & Khu, S.-T., Using genetic algorithms to calibrate a water quality model. *Science of The Total Environment*, 374: 260-272, 2007.

- [Matoso et al., 2004] Matoso, A., Rasga, M., Santana, M. & Murteira, M., Principais albufeiras do Alentejo monitorizadas. Comissão de Coordenação da Região Alentejo, Direcção de Serviços de Monitorização Ambiental, Alentejo, Portugal, 2004.
- [Mendes & Oliveira, 2004] Mendes, B. & Oliveira, F. J., Qualidade da Água para Consumo Humano. Lidel - Edições Técnicas, Lisboa, Portugal, 2004.
- [Michalski et al., 1998] Michalski, R., Bratko, I. & Miroslav, K., Machine Learning and Data Mining Methods and Applications. John Wiley and Sons, Inc., Londres, 1998.
- [Neves et al., 1988] Neves, J., Machado, J., Cortez, P., Arteiro, J. & Vicente, H., A distributed Web-based water quality diagnosis and simulation in lakes and reservoirs. Em A. Bargiela & E. Kerckhoffs Eds., *Simulation Technology: Science and Art, Proceedings of ESS'98 - 10th European Simulation Symposium and Exhibition*, 296-301, SCS, Nottingham, U.K., 1998.
- [Pyle, 1999] Pyle, D., Data Preparation for Data Mining. Morgan Kaufmann, San Francisco, USA, 1999.
- [Quinlan, 1993] Quinlan, J., C4.5 Programs for Machine Learning. Morgan Kaufmann Publishers Inc, USA, 1993.
- [Quinlan, 1996] Quinlan, J. R., Bagging, boosting and C4.5. Em Eds., *Proceedings of AAAI'96 National Conference on Artificial Intelligence*, 725-730, AAAI Press, Menlo Park, USA., 1996.
- [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.

- [Rich et al., 1991] Rich, E. & Knight, K., Artificial Intelligence. McGraw-Hill Inc., 1991.
- [Santos et al., 2005] Santos, M. F., Cortez, P., Quintela, H., Neves, J., Vicente, H. & Arteiro, J., Ecological Mining - A Case Study on Dam Water Quality. Em A. Zanasi, C.A. Brebbia & N.F.F. Ebecken Eds., Data Mining VI - Data Mining, Text Mining and their Business Applications, WIT Transactions of Information and Communication Technologies, Vol. 35, pp. 523-531, WIT Press, Southampton, UK, 2005.
- [Scardi, 2001] Scardi, M., Advances in neural network modeling of phytoplankton primary production. *Ecological Modelling*, 146: 33-45, 2001.
- [Schapire, 2002] Schapire, E. R., The boosting approach to machine learning: An overview. MSRI Workshop on Nonlinear Estimation and Classification, 2002.
- [Serralheiro, 2001] Serralheiro, R., A Sustentabilidade do Regadio e a Conservação do Solo e da Água em Condições Mediterrâneas. *Anais da Universidade de Évora* 10-11: 209-248, 2001.
- [Sousa, 2004] Sousa, C., Data Mining: Metodologias, Tecnologias, Modelos e Aplicações, Dissertação de Mestrado, Universidade do Minho, 2004.
- [Stumm & Morgan, 1996] Stumm, W. & Morgan, J. J., Aquatic Chemistry - Chemical Equilibria and Rates in Natural Waters. 3^a Ed., John Wiley & Sons, Inc., New York, 1996.
- [Teixeira, 2005] Teixeira, H. Q., Sistemas de Conhecimento Baseados em Data Mining: Aplicação à Análise da Estabilidade de Estruturas Metálicas, Dissertação de Mestrado, Universidade do Minho, 2005.

- [Thuraisingham, 1999] Thuraisingham, B., *Data Mining Technologies, Techniques, Tools and Trends*. CRC Press LLC, USA, 1999.
- [Tyagi et al., 2008] Tyagi, P., Chandramouli, V., Lingireddy, S. & Buddhi, D., Relative Performance of Artificial Neural Networks and Regression Models in Predicting Missing Water Quality Data. *Environmental Engineering Science*, 25: 657-668, 2008.
- [URL1] <http://ga.water.usgs.gov/edu/watercycle.html> (Visitado em 4 de Janeiro de 2009).
- [URL2] <http://www.rulequest.com/see5-info.html> (Visitado em 4 de Janeiro de 2009).
- [Vicente, 2004] Vicente, H., *Especificação e Prototipação da Sistemas de Gestão e Controlo da Qualidade da Água de Albufeiras*, Departamento de Química, Universidade de Évora, Évora, Portugal, 2004.
- [Weiss & Provost, 2001] The effect of class distribution on classifier learning: na empirical study. Technical Report ML-TR-44, Department of Computer Science, Rutgers University, 2001.
- [Wilson & Recknagel, 2001] Wilson, H. & Recknagel, F., Towards a generic artificial neural networks model for dynamic predictions of algal abundance in freshwater lakes. *Ecological Modelling*, 146: 69-84, 2001.
- [Yang et al., 2008] Yang, W., Nan, J. & Sun, D., An online water quality monitoring and management system developed for the Liming River basin in Daqing, China. *Journal of Environmental Management*, 88: 318-325, 2008.
- [Zhang, 2007] Zhang, C., *Fundamentals of Environmental Sampling and Analysis*. John Wiley & Sons, Inc., Hoboken, New Jersey, USA, 2007.