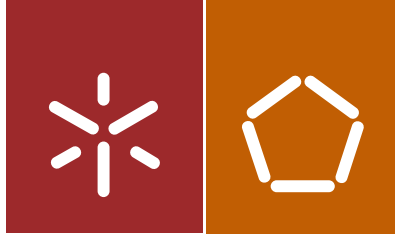




Universidade do Minho
Escola de Engenharia

Ana Rita Baptista Brandão

Modelos estatísticos para a
caracterização e previsão de parâmetros
indicadores da qualidade da água.
Aplicação ao rio Ave.



Universidade do Minho
Escola de Engenharia

Ana Rita Baptista Brandão

Modelos estatísticos para a
caracterização e previsão de parâmetros
indicadores da qualidade da água.
Aplicação ao rio Ave.

Dissertação de Mestrado
Ciclo de Estudos Integrados Conducentes ao
Grau de Mestre em Engenharia Civil

Trabalho efetuado sob a orientação do
Professor Doutor Júlio F. Ferreira da Silva

AGRADECIMENTOS

Após o término desta dissertação, fica a necessidade de prestar o meu profundo agradecimento a todas as pessoas sem as quais não teria sido possível a concretização da mesma.

Em primeiro lugar, quero expressar o meu sincero agradecimento ao meu orientador, Professor Doutor Júlio F. Ferreira da Silva, por toda a disponibilidade, apoio e simpatia que sempre demonstrou, por todos os conhecimentos transmitidos e também por todas as sugestões que tornaram a realização desta dissertação possível.

Em segundo lugar, quero agradecer aos meus amigos por sempre me incentivarem e por todos os bons momentos proporcionados.

Finalmente, resta-me deixar um agradecimento especial à minha família, por todo o apoio, incentivo e carinho que sempre me transmitiram ao longo de todo o meu percurso académico.

RESUMO

A presente dissertação dedica-se ao estudo de modelos estatísticos de previsão e caracterização da qualidade da água superficial, a curto prazo, recorrendo à aplicação dos *softwares SPSS e Microsoft Excel*. O estudo foi aplicado ao rio Ave, onde tendo por base os dados recolhidos nos anos de 2005 e 2006, na estação de monitorização automática das Taipas (sonda INAG 05G/06), se efetuou a modelação das sucessões cronológicas das variáveis: condutividade elétrica, oxigénio dissolvido, temperatura e pH. Para o efeito, consideraram-se os valores horários, registados ao longo de 24 horas. Adicionalmente foram usados dados recolhidos por uma sonda no ano de 2012.

Recorrendo ao *software SPSS*, foram ajustados modelos estatísticos como o Sazonal Simples, Aditivo de *Winters* e ARIMA, às sucessões cronológicas de cada parâmetro. Com o intuito de compreender a sazonalidade de determinados comportamentos, foram comparados os modelos gerados semana a semana com os modelos gerados para a totalidade do período de estudo. Efetuou-se também a comparação entre modelos gerados para o mesmo período, mas para anos distintos. Importa ainda referir que foram criados modelos que incluíram o caudal do rio como variável independente. Por fim, efetuaram-se previsões a curto prazo, tendo por base modelos ARIMA.

Previamente ao tratamento dos dados, fez-se um breve estudo dos modelos hidrológicos matemáticos, séries temporais, suas componentes e modelos que as caracterizam.

Relativamente ao caso de estudo, efetuou-se uma pesquisa relativa às possíveis fontes de poluição das suas águas, analisando também a evolução da sua qualidade ao longo do tempo.

Os modelos estatísticos de previsão apresentam-se como uma ferramenta bastante útil no que respeita ao estudo da qualidade da água, permitindo prever e avaliar antecipadamente o seu estado, possibilitando desta forma uma otimização tanto ao nível de controlo como de vigilância de alguns dos parâmetros indicadores de qualidade da água.

Palavras-chave: Parâmetros indicadores de qualidade da água; Modelos de previsão; *SPSS*;

ABSTRACT

This dissertation is dedicated to the study of statistical prediction models and characterization of surface water quality in the short term, applying SPSS software and Microsoft Excel. The study of water quality parameters was applied to river Ave and focused on the following variables: electrical conductivity, dissolved oxygen, temperature and pH. For this purpose were considered the hourly values recorded over 24 hours, collected since 2005 until 2006, by the automatic monitoring station located in Taipas. Further information collected in 2012, was already used.

Statistical models like Simple Seasonal, Winters' Additive and ARIMA, were adjusted to the time series of each parameter by using *SPSS* software. In order to understand the seasonality of certain behaviors, the models generated from week to week were compared with the models generated for the entire period. In addition, a comparison between models generated for the same period but for different years was made, as well as models including the river flow as an independent variable. Finally, there were also made short-term forecasts, based on ARIMA models.

A brief study of mathematical hydrologic models, time series models and their components was carried. Regarding the case study, it was taken a research of possible pollution sources on the river water and its quality evolution over time.

Statistical forecasting models present themselves as a very powerful tool to study water quality, allowing predictions and evaluations of water status in advance. This way, an optimization control as well as surveillance can improve water quality.

Keywords: Water quality parameters; Forecasting models; SPSS;

ÍNDICE

1. INTRODUÇÃO.....	1
2. ESTADO DA ARTE	5
2.1. Parâmetros que interferem na qualidade da água	5
2.1.1. Parâmetros em análise.....	8
2.1.2. Métodos utilizados na formulação de índices de qualidade da água.....	12
2.2. Tipos de modelos matemáticos hidrológicos	16
2.3. Séries temporais	18
2.3.1. Componentes das séries temporais.....	19
2.3.2. Modelos que caracterizam as séries temporais	20
2.4. Rio Ave e modelação estatística	34
2.5. Estudos recentes.....	36
3. METODOLOGIA.....	39
3.1. Caso de estudo 1: Variação semanal dos parâmetros <i>versus</i> modelo ajustado pelo SPSS.....	45
3.2. Caso de estudo 2: Variação anual dos parâmetros <i>versus</i> modelo ajustado pelo SPSS	46
3.3. Caso de estudo 3: Caudal do rio como variável independente	46
3.4. Caso de estudo 4: Modelação ARIMA na realização de previsões	46
3.5. Caso de estudo 5: Modelos gerados para o ano de 2005 <i>versus</i> modelos gerados para o ano de 2012.....	46
3.6. Caso de estudo 6: Variação dos parâmetros <i>versus</i> variação do caudal	47
4. CASO DE ESTUDO: RIO AVE	49
4.1. Indústrias locais e poluição associada.....	50
4.2. Evolução da qualidade da água do rio Ave.....	52
5. RESULTADOS E RESPETIVA ANÁLISE	57
5.1. Caso de estudo 1: Variação semanal dos parâmetros <i>versus</i> modelo ajustado pelo SPSS.....	62
5.2. Caso de estudo 2: Variação anual dos parâmetros <i>versus</i> modelo ajustado pelo SPSS	67
5.3. Caso de estudo 3: Caudal do rio como variável independente	72
5.4. Caso de estudo 4: Modelação ARIMA na realização de previsões	77

5.5. Caso de estudo 5: Modelos gerados para o ano de 2005 <i>versus</i> modelos gerados para o ano de 2012	81
5.6. Caso de estudo 6: Variação dos parâmetros <i>versus</i> variação do caudal.....	86
6. CONCLUSÕES	99
6.1. Sugestões para desenvolvimentos futuros	102
REFERÊNCIAS BIBLIOGRÁFICAS.....	103

ÍNDICE DE FIGURAS

Figura 1 - Definição das variáveis.....	40
Figura 2 - Preenchimento das variáveis com os dados históricos.	41
Figura 3 - Criação de Modelos de Previsão.....	42
Figura 4 - Seleção da variável a estudar.	43
Figura 5 - Definição de critérios do " <i>Expert Modeler</i> ".	44
Figura 6 - Definição de critérios ARIMA.	44
Figura 7 - Bacia hidrográfica do rio Ave (Fonte: Agência Portuguesa do Ambiente, 2014). ..	49
Figura 8 - Classificação da qualidade da água na bacia hidrográfica do rio Ave no ano de 2005 (Fonte: SNIRH, 2014).	53
Figura 9 - Classificação da qualidade da água na bacia hidrográfica do rio Ave no ano de 2012 (Fonte: SNIRH, 2014).	53
Figura 10 - Dados históricos de Oxigénio Dissolvido para o período de 07-06-2005 a 13-06-2005.	58
Figura 11 - Dados históricos de pH para o período de 23-05-2006 a 29-05-2006.	59
Figura 12 - Exemplo de dados históricos de Condutividade Elétrica com uma perturbação, para o período de 29-11-2005 a 05-12-2005.	61
Figura 13 - Dados históricos de Condutividade Elétrica <i>versus</i> modelo SARIMA (0,1,1)(1,0,0) para o período de 30-05-2006 a 05-06-2006.	63
Figura 14 - Dados históricos de Oxigénio Dissolvido <i>versus</i> modelo SARIMA (1,1,0)(1,0,1) para o período de 30-05-2006 a 05-06-2006.	64
Figura 15 - Dados históricos de Temperatura <i>versus</i> modelo SARIMA (1,1,2)(2,1,1) para o período de 30-05-2006 a 05-06-2006.	65
Figura 16 - Dados históricos de pH <i>versus</i> modelo Sazonal Simples para o período de 30-05-2006 a 05-06-2006.....	66
Figura 17 - Dados históricos de Condutividade Elétrica <i>versus</i> modelo SARIMA (1,1,1)(1,0,1) para o período de 07-07-2005 a 15-06-2006.	68
Figura 18 - Dados históricos de Oxigénio Dissolvido <i>versus</i> modelo SARIMA (1,1,1)(0,1,1) para o período de 07-07-2005 a 15-06-2006.	69
Figura 19 - Dados históricos de Temperatura <i>versus</i> modelo SARIMA (0,1,1)(1,1,1) para o período de 07-07-2005 a 15-06-2006.	70
Figura 20 - Dados históricos de pH <i>versus</i> modelo SARIMA (1,1,1)(1,0,1) para o período de 07-07-2005 a 15-06-2006.	71

Figura 21 - Dados históricos de Condutividade Elétrica <i>versus</i> modelo ARIMA (0,1,1) para o período de 30-05-2006 a 05-06-2006, considerando o Caudal a montante como variável independente.	73
Figura 22 - Dados históricos de Oxigénio Dissolvido <i>versus</i> modelo ARIMA (1,1,0)(0,0,0) para o período de 30-05-2006 a 05-06-2006, considerando o Caudal a montante como variável independente.	74
Figura 23 - Dados históricos de Temperatura <i>versus</i> modelo SARIMA (0,1,1)(1,1,0) para o período de 30-05-2006 a 05-06-2006, considerando o Caudal a montante como variável independente.	75
Figura 24 - Dados históricos de pH <i>versus</i> modelo Sazonal Simples para o período de 30-05-2006 a 05-06-2006, considerando o Caudal a montante como variável independente.	76
Figura 25 - Previsão para as 24 h do dia 29-05-2006 usando modelo ARIMA (2,1,1)(0,0,0) para a sucessão cronológica de oxigénio dissolvido <i>versus</i> dados históricos registados pela sonda.	78
Figura 26 - Pormenor dos valores previstos para as 24 h do dia 29-05-2006, usando o modelo ARIMA (2,1,1)(0,0,0), <i>versus</i> valores registados para o mesmo período.	80
Figura 27 - Dados históricos de Condutividade Elétrica <i>versus</i> modelo SARIMA (0,1,2)(1,0,1), para o período de 21-06-2005 a 27-06-2005.	82
Figura 28 - Dados históricos de Condutividade Elétrica <i>versus</i> modelo SARIMA (0,1,2)(2,0,1), para o período de 21-06-2012 a 27-06-2012.	83
Figura 29 - Dados históricos de Temperatura <i>versus</i> modelo SARIMA (1,1,0)(1,0,1), para o período de 21-06-2005 a 27-06-2005.	84
Figura 30 - Dados históricos de Temperatura <i>versus</i> modelo SARIMA (1,1,0)(1,0,1), para o período de 21-06-2012 a 27-06-2012.	85
Figura 31 - Dados históricos de Condutividade Elétrica e Caudal a montante para o período de 14-02-2006 a 20-02-2006.	87
Figura 32 - Dados históricos de Oxigénio Dissolvido e Caudal a montante para o período de 14-02-2006 a 20-02-2006.	88
Figura 33 - Dados históricos de Temperatura e Caudal a montante para o período de 14-02-2006 a 20-02-2006.	89
Figura 34 - Dados históricos de pH e Caudal a montante para o período de 14-02-2006 a 20-02-2006.	90
Figura 35 - Funções de autocorrelação e autocorrelação parcial da sucessão residual do modelo SARIMA (0,1,1)(1,0,0) para a variável Condutividade Elétrica.	94

Figura 36 - Funções de autocorrelação e autocorrelação parcial da sucessão residual do modelo SARIMA (0,1,1)(1,1,1) para a variável Condutividade Elétrica.....97

ÍNDICE DE TABELAS

Tabela 1 - Média dos valores registados de 07-07-2005 a 15-06-2006 em comparação com VMR e VMA.	54
Tabela 2 - Média dos valores registados de 03-01-2012 a 06-07-2012 em comparação com VMR e VMA.	55
Tabela 3 - Ajuste do modelo SARIMA (0,1,1)(1,0,0), gerado a partir da opção " <i>Expert Modeler</i> ".	92
Tabela 4 - Parâmetros do modelo SARIMA (0,1,1)(1,0,0).	93
Tabela 5 - Ajuste do modelo SARIMA (0,1,1)(1,1,1), gerado manualmente.	95
Tabela 6 - Parâmetros do modelo SARIMA (0,1,1)(1,1,1).	96

LISTA DE ACRÓNIMOS

ACP	<i>Análise de Componentes Principais</i>
AIC	<i>Akaike Information Criteria</i>
AR	<i>Auto Regressivos</i>
ARIMA	<i>Modelos Auto Regressivos Integrados e de Médias Móveis</i>
ARMA	<i>Modelos Auto Regressivos e de Médias Móveis</i>
BIC	<i>Bayesian Information Criteria</i>
ETAR	<i>Estação de Tratamento de Águas Residuais</i>
FAC	<i>Função de Autocorrelação</i>
FACP	<i>Função de Autocorrelação Parcial</i>
INAG	<i>Instituto Nacional da Água</i>
MA	<i>Médias Móveis</i>
OD	<i>Oxigénio Dissolvido</i>
PAR	<i>Auto Regressivos Periódicos</i>
REQM	<i>Erro Quadrático Médio</i>
RH	<i>Região Hidrográfica</i>
SARIMA	<i>Modelos Auto Regressivos Integrados e de Médias Móveis Sazonais</i>
SARMA	<i>Modelos Auto Regressivos e de Médias Móveis Sazonais</i>
SI	<i>Sistema Internacional de Unidades</i>
SNIRH	<i>Sistema Nacional de Informação de Recursos Hídricos</i>
SPSS	<i>Statistical Package for the Social Sciences</i>
STD	<i>Sólidos Totais Dissolvidos</i>
VMA	<i>Valor Máximo Admissível</i>
VMR	<i>Valor Máximo Recomendado</i>

1. INTRODUÇÃO

São diversos os recursos naturais necessários para assegurar a qualidade da saúde humana, sendo a água um dos mais importantes. Para além de todos os benefícios que proporciona aos seres humanos, a água contribui ainda para o desenvolvimento económico e social, devido ao papel que exerce no transporte marítimo, agricultura, pesca, no lazer em praias e na riqueza paisagística e valorização da qualidade de vida urbana e do turismo, em cidades banhadas por cursos de água, (Yassuda, 1993). Porém, o crescimento da população muito tem contribuído para o aumento da procura e contaminação dos recursos hídricos, (World Health Organization, 2014; Peters & Meybeck, 2000). A falta de condições sanitárias e de tratamento de resíduos nas zonas com maior densidade populacional dos países em desenvolvimento pode também ser um fator de peso, no que respeita à degradação da qualidade da água. Tais comportamentos, associados aos diferentes usos competitivos deste recurso, contribuem de forma significativa para a sua acentuada diminuição em termos de quantidade e qualidade, podendo levar a graves problemas de saúde pública, (Gibbons, 1987).

A água é um elemento essencial à vida, porém pode servir de veículo para várias doenças, trazendo sérios riscos à saúde em consequência da sua má qualidade. Ao longo da história, o progresso que se verificou nos serviços públicos de abastecimento de água refletiu-se numa acentuada melhoria da saúde da população, diminuindo/eliminando as doenças por aí transmitidas e baixando também o índice de mortalidade. A questão da qualidade da água foi assim ganhando grande evidência, sendo que assegurar a disponibilidade deste recurso à geração atual e às gerações futuras, em padrões de qualidade adequados aos respetivos usos, constitui, atualmente, a principal meta a assegurar por parte das entidades responsáveis pela sua gestão.

Uma vez que a gestão integrada de todos os recursos hídricos se tornou numa questão tão pertinente, o estudo e exploração de modelos estatísticos para a caracterização e previsão de parâmetros indicadores da qualidade da água, tal como de outras ferramentas, podem revelar-se bastante úteis para essa mesma gestão. Uma otimização a nível de controlo e vigilância de alguns dos parâmetros indicadores de qualidade da água pode, assim, ser a chave para uma gestão racional, e sobretudo sustentável, dos recursos hídricos.

O principal objetivo da presente dissertação consiste então na caracterização, através de ferramentas estatísticas, da evolução temporal de parâmetros indicadores da qualidade da água. Para tal, serão desenvolvidos modelos matemáticos de sucessões cronológicas, tendo em vista a modelação e a previsão de parâmetros de qualidade da água do rio Ave.

Em concreto, serão analisados valores registados ao longo dos últimos anos numa estação de monitorização automática do rio Ave e também valores recolhidos através de uma sonda no ano de 2012, de modo a avaliar o comportamento/tendências dos parâmetros indicadores da qualidade da água ao longo do tempo e também de modo a verificar qual o modelo que melhor se ajusta a cada uma das variáveis em estudo.

A análise das séries temporais das variáveis em estudo e o uso de um modelo adequado a cada uma delas serão assim essenciais para posteriormente se poderem efetuar previsões com boa precisão. Será também importante tirar conclusões acerca da origem das variações ou da sazonalidade de determinados comportamentos das variáveis que caracterizam a qualidade da água.

Em suma, para a concretização do objetivo principal, será fundamental passar pelas seguintes etapas:

- ❖ Definição dos parâmetros a analisar;
- ❖ Estudo dos modelos que caracterizam as séries temporais;
- ❖ Enquadramento do caso de estudo e revisão do estado atual dos conhecimentos;
- ❖ Aplicação do *software SPSS* e dos modelos de previsão;
- ❖ Análise crítica dos resultados obtidos.

Considera-se que através de uma abordagem temporal, aplicando ferramentas estatísticas em função do tempo, será possível realizar previsões e simulações de cenários que permitam a avaliação das condições presentes e futuras dos recursos hídricos, contribuindo assim para a gestão e conservação dos mesmos e também para a construção de políticas que permitam a melhoria da qualidade da água.

Nos capítulos que se seguem serão descritos com maior detalhe os parâmetros que interferem na qualidade da água, alguns modelos hidrológicos matemáticos, séries temporais e modelos que as caracterizam. Relativamente ao caso de estudo, far-se-á a revisão do estado

atual dos conhecimentos, uma análise da qualidade das suas águas ao longo do tempo e em específico a aplicação de modelos estatísticos, dando especial ênfase ao *software SPSS*. Por fim, apresentar-se-ão as conclusões retiradas do estudo realizado.

2. ESTADO DA ARTE

2.1. Parâmetros que interferem na qualidade da água

O primeiro passo para assegurar a qualidade da água passa por controlar todos os fatores que possam interferir, direta ou indiretamente, para a sua degradação. A gestão dos recursos hídricos interliga-se, deste modo, com a gestão ambiental.

De modo a avaliar se um determinado meio hídrico apresenta condições satisfatórias para assegurar os seus usos potenciais, é necessário efetuar a caracterização físico-química e bacteriológica da água, ou seja, verificar se se cumprem os valores paramétricos estipulados por lei.

A avaliação da qualidade de qualquer sistema ambiental depende, fundamentalmente, da escolha dos parâmetros que serão representativos do seu estado, no momento da recolha da amostra. É portanto essencial avaliar e identificar os parâmetros mais importantes para a qualidade do meio hídrico, definindo, numa fase inicial do processo, quais os mais relevantes para que posteriormente se efetue uma adequada monitorização do mesmo.

A metodologia utilizada para a definição de parâmetros de qualidade da água relevantes para a monitorização de águas consiste primeiramente na caracterização da área de estudo, que envolve o reconhecimento das características locais que apontem para alguma relação com as fontes de poluição e usos da água, que possam ser responsáveis pela contaminação do meio hídrico. Seguidamente, deve proceder-se a um levantamento dos dados de campo, no qual se escolhem alguns pontos estratégicos, que obedecem a critérios específicos, que possam interferir nos resultados das análises das amostras.

Assim, a escolha preliminar de parâmetros específicos a serem analisados, como o pH, oxigénio dissolvido, sólidos suspensos totais, nutrientes etc., é um passo fundamental em todo o processo de caracterização da qualidade da água.

A periodicidade da recolha de amostras pode também influenciar os resultados obtidos, pelo que adotar um perfil temporal mais detalhado será o ideal para que se possa caracterizar de forma mais realista a qualidade das águas em estudo. Esta será a forma mais coerente de se

efetuar o estudo, visto que há grande variação da composição da água, em função do período (seco ou húmido) em que se efetua o estudo, podendo conduzir a resultados bastante distintos.

No âmbito do estudo dos parâmetros de qualidade de água, o conhecimento das condições ambientais locais revela-se bastante útil para a sua modelação. Desta forma, identificar as fontes de contaminação da água será também uma forma de reduzir os seus eventuais efeitos, uma vez que normalmente existe uma relação direta das atividades antropogénicas com o nível de poluição identificado no meio hídrico, (Oliveira, Campos, & Medeiros, 2010).

Relativamente às fontes de poluição, estas podem ser classificadas em difusas ou pontuais. No grupo das fontes difusas destacam-se os solos agrícolas, o escoamento de águas pluviais e a mineração. No caso das fontes pontuais, as de maior ênfase são a indústria e também os aglomerados, quer sejam urbanos ou rurais (lançamentos de esgotos domésticos, lixo urbano, etc.). Ambos os tipos de fontes poluidoras contribuem com microrganismos, nutrientes, matéria orgânica, sais minerais, cor, espumas, químicos sintéticos, produtos tóxicos, aquecimento, sólidos em suspensão, fertilizantes, pesticidas, metais pesados, resíduos sólidos, substâncias radioativas, etc., que prejudicam gravemente a qualidade da água, podendo pôr em causa a saúde pública. Salienta-se ainda a possível inexistência de ETAR ou descargas de efluentes de ETAR com baixa eficiência de remoção de nutrientes como fonte poluidora, que possuem também um peso bastante significativo, no que respeita à deterioração da qualidade do meio hídrico, (Duarte, 2013).

As fontes difusas de poluição, especialmente a agricultura, têm sido objeto de atenção em vários países, devido à dificuldade de se estabelecer procedimentos de avaliação de impactos ambientais e de se adotar padrões aceitáveis, como outrora ocorreu com as fontes pontuais, (Sims, Simard; & Joern, 1998; Toledo & Nicolella, 2002; Parry, 1998). Embora vários autores enfatizem a complexidade das fontes difusas, no que respeita ao mecanismo de transporte de nutrientes em bacias agrícolas, uma abordagem mais ampla que envolva outras variáveis de qualidade de água deverá ser considerada.

Assim, o uso de indicadores de qualidade de água consiste no emprego de variáveis que se correlacionam com as alterações ocorridas na bacia, sejam estas de origens antropogénicas ou naturais. Cada sistema lótico possui as suas próprias características, pelo

que se torna difícil estabelecer uma única variável como um indicador padrão para qualquer sistema hídrico. Neste sentido, os trabalhos de campo tem como objetivo a obtenção de índices de qualidade de água que reflitam resumidamente e objetivamente as alterações, com ênfase para as intervenções humanas, como o uso agrícola, urbano e industrial, Toledo & Nicolella, 2002; Couillard & Lefebvre, 1985).

As interações entre as diversas variáveis medidas numa amostra de água constituem o ponto de partida para a avaliação da qualidade da água, desde que estas interações sejam obtidas a partir de uma distribuição amostral no espaço e no tempo das variáveis do sistema a ser estudado, (Toledo & Nicolella, 2002; Harmancioglu, Ozkul & Alpaslan, 1998).

Para uma interpretação ecológica da qualidade das águas superficiais e/ou para estabelecer um sistema de monitorização, é necessária a utilização de métodos simples e que forneçam informações objetivas e interpretáveis, partindo para critérios próprios que considerem as características peculiares dos recursos hídricos, (Pineda & Schäfer, 1987). Neste contexto, o uso de índices de qualidade de água tem como função prever e acompanhar todo o programa de monitorização das águas, através de informações sucintas, possibilitando a prevenção da deterioração dos recursos hídricos, ao longo do tempo. Obviamente que todos estes índices contemplam um grau de subjetividade, pois dependem da escolha das variáveis que constituirão os indicadores principais das alterações da qualidade de água. Índices baseados em técnicas estatísticas favorecem a determinação dos indicadores mais característicos do meio hídrico em estudo, embora não permitam generalizações para todos os outros, já que cada sistema hídrico, em princípio, possui sua característica peculiar. Por outro lado, como instrumento de avaliação ao longo do tempo ou do espaço, estes índices permitem acompanhar as alterações ocorridas no meio hídrico, (Toledo & Nicolella, 2002, Haase, Krieger, & Possoli, 1989).

Em Portugal existe uma base de dados - Sistema Nacional de Informação de Recursos Hídricos (SNIRH) - que fornece informação e dados recolhidos nas diversas redes de monitorização sobre vários parâmetros relacionados com o meio hídrico. No que respeita à qualidade da água, os mais comuns são: temperatura, oxigénio dissolvido, pH, turvação, cloreto, carbonato, bicarbonato, dióxido de carbono, sulfato, sódio, potássio, cálcio, magnésio, sólidos suspensos totais, sólidos suspensos voláteis, carência bioquímica de oxigénio, azoto amoniacal, fósforo, metais pesados, óleos, gorduras, coliformes fecais, condutividade elétrica,

clorofila e cor, (Sistema Nacional de Informação de Recursos Hídricos, 2014). No entanto, é de salientar que nem todas as estações de monitorização do país possuem registos para todas as variáveis indicadoras de qualidade da água.

2.1.1. Parâmetros em análise

Como referido anteriormente, são vários os parâmetros que podem ser analisados no âmbito da qualidade da água. Tendo em conta o sistema lótico em estudo (rio Ave), as suas principais características e os dados e meios disponíveis, os parâmetros a analisar no presente estudo são:

- ❖ Oxigénio Dissolvido (OD): trata-se de um indicador de elevada importância no que respeita à avaliação da qualidade da água, uma vez que permite identificar fenómenos ambientais como a eutrofização ou a poluição orgânica. Importa salientar que a eutrofização surge associada à excessiva produção primária e relaciona-se com a proliferação acelerada e excessiva de biomassa algal e redução da biodiversidade resultante de um enriquecimento progressivo do meio aquático com nutrientes (essencialmente azoto e fósforo), causando um desequilíbrio na cadeia trófica aquática. A eutrofização ocorre frequentemente nas estações mais quentes, quando os fatores de temperatura, luz solar e nutrientes são mais elevados. Nestas circunstâncias, uma vez que as condições do meio são propícias à ocorrência deste fenómeno, a qualidade da água degrada-se significativamente, nomeadamente no que diz respeito ao índice de cor, turvação, sabor e cheiro. É ainda de referir que apesar de se tratar de um processo natural, o enriquecimento da água em nutrientes possui uma origem essencialmente antropogénica, como é o caso da atividade agropecuária (culturas de arroz, queijarias, adegas, suiniculturas, boviniculturas, aviculturas, etc.), das estações de tratamento de águas residuais (inexistência de ETAR ou descargas de efluentes de ETAR com baixa eficiência de remoção de nutrientes), atividades de recreio ou mesmo das indústrias, (Duarte, 2013).

No que respeita à existência de resíduos orgânicos que são lançados no meio hídrico, estes são decompostos por microrganismos que utilizam oxigénio na sua respiração, ou seja, ao aumento de matéria orgânica presente na água corresponde um aumento de microrganismos e, conseqüentemente, uma diminuição do oxigénio dissolvido existente, gasto na respiração desses microrganismos.

As principais fontes de nutrientes e matéria orgânica são então os resíduos domésticos, agrícolas e resíduos industriais. Níveis de oxigênio dissolvido abaixo de 5mg/L são considerados perigosos para a vida animal, sendo uma das principais causas que leva à morte de peixes nos rios poluídos, (Silva, 2013).

Por último, importa referir que a concentração de OD nos cursos de água, além das atividades antropogénicas, depende também da temperatura, da pressão atmosférica, da salinidade, de atividades biológicas e ainda de características hidráulicas do próprio curso de água.

O oxigênio dissolvido mede-se em mg/L.

- ❖ Temperatura: é a medida da intensidade de calor expresso numa determinada escala. A escala mais usada é grau centígrado ou grau Celsius (°C). A análise deste parâmetro revela-se essencial, na medida em que este apresenta uma grande influência na velocidade das reações químicas, na solubilidade dos gases e sobretudo na taxa de crescimento dos microrganismos presentes no meio hídrico.

A temperatura serve também para avaliar a estabilidade das águas, uma vez que a sua propagação se dá maioritariamente através da mistura da água. Em relação a essa mistura é de referir que esta depende não só da densidade das massas de água mas também da existência ou não de vento no local. Através da temperatura é então possível verificar se um meio hídrico se encontra ou não estratificado termicamente, podendo ser estável ou não, (Silva, 2013).

A temperatura pode ser medida por diferentes dispositivos, como é o caso do termómetro ou sensor.

- ❖ pH: é um indicador que expressa o grau de acidez ou alcalinidade da água em análise e representa a concentração de iões de hidrogénio existentes nessa mesma água. A escala de pH varia entre os valores 0 e 14, sendo adimensional. Valores abaixo de 7 indicam presença de águas ácidas, já valores maiores do que 7 indicam alcalinidade. O valor de 7 corresponde ao pH neutro, (Duarte, 2013).

Um fator de grande influência no valor do pH na água é o tipo de solo por onde esta circula. No que respeita à vida aquática, o pH pode variar entre os valores de 6 a 9, (Silva, 2013).

- ❖ Condutividade elétrica: traduz a capacidade da água em conduzir a corrente elétrica, variando com a temperatura, a concentração de iões dissolvidos, a mobilidade e a valência dos iões. Assim, dado que os compostos orgânicos e inorgânicos interferem na condutividade elétrica da água, consoante a sua concentração, a um aumento da quantidade de iões dissolvidos na água irá corresponder um aumento da condutividade elétrica. Os principais iões responsáveis pela condutividade elétrica da água são o cálcio, magnésio, ferro, potássio, sódio, carbonatos, sulfatos e cloretos, (Silva, 2013). Este indicador revela-se bastante importante, na medida em que permite alertar para a presença de descargas industriais não tratadas, produtos provenientes de minas, esgotos, entre outros. Por outro lado, materiais orgânicos, como óleos, álcool ou fenóis não possuem a capacidade de conduzir eletricidade. Assim, quando se apresentam na forma dissolvida na água, a condutividade elétrica é severamente reduzida.

Desta forma, quando se mede a condutividade elétrica de uma amostra de água, na realidade está também a quantificar-se uma grande quantidade de compostos nela contidos, sendo a partir daí possível identificar situações anómalas. Na identificação destas situações, deve atender-se que cada meio hídrico tende a ter uma gama relativamente consistente de valores de condutividade elétrica que, uma vez conhecidos, podem ser utilizados como base de comparação para medições regulares de condutividade. Desta forma, alterações significativas na condutividade podem indicar que uma descarga ou alguma outra fonte de contaminação tenha entrado no meio hídrico.

A condutividade elétrica da água permite também avaliar o grau de mineralização da água, pelo que os seus valores representam a carga mineral presente na água e a geologia local ou até regional. Assim, em formações predominantemente de granito a condutividade elétrica é extremamente baixa. Por outro lado, em solos argilosos os valores de condutividade são bastante altos.

Ainda no que respeita à condutividade elétrica, deve dar-se especial importância à influência que a temperatura possui nos valores medidos. Uma vez que se trata de um fator que interfere bastante nos valores de condutividade elétrica, os dados medidos de condutividade elétrica devem ser sempre acompanhados da temperatura na qual foram medidos. Para propósitos comparativos de dados de condutividade elétrica, deve então definir-se uma das temperaturas de referência (20 °C ou 25 °C). Atualmente, os equipamentos disponíveis já são capazes de fornecer a condutividade elétrica já convertida para uma das temperaturas de referência.

Assim, quanto maior for a temperatura da água, maior será a liberdade de movimento dos iões dissolvidos condutores de corrente elétrica, pelo que o valor da condutividade irá também aumentar.

A interferência e necessidade da correção da temperatura é um dos pontos mais críticos da medição da condutividade, uma vez que a sua definição equivocada pode trazer erros bastante elevados, visto que em condições ambientais normais, a cada grau de temperatura, o erro da medição pode atingir 2%, em águas superficiais, até 6%, em águas de condutividade muito baixa, perto da concentração de águas puras. Porém não deve ser esquecido que outra fonte de erro é aquela que se relaciona com o operador. Aqui alerta-se para o facto de não se aguardar a estabilização da temperatura na sonda antes de começar as medições ou mesmo erros no que respeita ao manuseamento da própria sonda sem se tomar as devidas precauções. Neste caso, operar com sondas que ficam expostas ao sol ou armazenadas em veículos quentes pode trazer um grande potencial de erro na medição. Para equipamentos maiores e com águas mais paradas, o tempo de estabilização pode ser crítico para uma boa medição deste parâmetro.

Através da medição da condutividade elétrica da água é então possível obter-se valores de vários parâmetros, tais como: Sólidos Totais Dissolvidos (STD); Salinidade; Resistência elétrica, onde contrariamente à condutividade, esta representa o isolamento da eletricidade que a água propicia, representado pela unidade de Ohms; Densidade da água do mar.

Os valores da condutividade elétrica da água são utilizados há décadas como indicadores da qualidade da água, podendo a sua representação ser expressa por diferentes unidades e, principalmente, pelos seus múltiplos. No Sistema Internacional de Unidades (S.I.), representa-se a condutividade em Siemens por metro (S/m). No entanto, em medições realizadas em amostras de água, utiliza-se preferencialmente microSiemens ($\mu\text{S}/\text{cm}$) ou miliSiemens por centímetro (mS/cm).

2.1.2. Métodos utilizados na formulação de índices de qualidade da água

A modelação e previsão de parâmetros de qualidade da água, através de modelos matemáticos de sucessões cronológicas, pode revelar-se fundamental na caracterização do comportamento típico, ao longo de um determinado período, dos valores registados para determinados parâmetros, numa determinada secção de um rio. Assim, será em muito facilitada a tarefa de identificação de comportamentos anómalos. Uma vez identificados este género de comportamentos, poder-se-á tomar um conjunto de medidas que minimizem ou mitiguem os efeitos da afluência de água ao sistema de abastecimento com qualidade imprópria, (Silva, Salgado, & Menezes, 2009).

Para uma gestão integrada do meio hídrico é fundamental o conhecimento dos processos físicos, químicos e biológicos que nele ocorrem, sobretudo em meios cujo desequilíbrio é fruto da pressão da atividade humana. A maximização dos benefícios e a consequente minimização dos impactos negativos decorrentes da ação antropogénica na utilização dos recursos hídricos é um dos principais objetivos da interação entre a modelação matemática e as tecnologias de informação e comunicação, no sentido de proporcionar a sua aplicação na resolução de problemas relacionados com os sistemas hídricos, contribuindo assim para o desenvolvimento sustentável da sociedade contemporânea, (Pinho, 2000).

A modelação matemática de problemas que ocorrem no meio hídrico permite a organização e o tratamento de um grande volume de informação, simulando o comportamento do sistema natural, através de modelos hidrodinâmicos e de qualidade da água. Em Pinho (2000), são salientados alguns trabalhos relacionados com a modelação da qualidade da água, nomeadamente modelos de qualidade da água em zonas costeiras e também oceânicas.

Ao longo do tempo, vários métodos foram sendo utilizados e testados, tendo em vista a modelação de parâmetros de qualidade de água. Um dos métodos usados na formulação de índices de qualidade de água baseia-se na técnica multivariada da análise fatorial, (Toledo & Nicolella, 2002; Shoji; Yamanoto, & Nakamura, 1966; Shahwan & Odening, 2007; Possoli & Haase, 1993), que possibilita a interpretação dos dados. Esta técnica permite conhecer o comportamento dos dados, a partir de uma dimensão reduzida do espaço original dos parâmetros. Permite ainda selecionar as variáveis mais representativas do meio hídrico, favorecendo a definição de indicadores mais sensíveis, tanto para adoção de um programa de monitorização, como para avaliação das alterações ocorridas no meio hídrico. É então possível estabelecer um índice de qualidade de água resultante da composição das diversas variáveis, de acordo com a importância individual de cada uma delas, de maneira a expressar de forma objetiva e integrada, as alterações da qualidade da água em diferentes bacias.

O modelo estatístico subjacente à análise fatorial, (Toledo & Nicolella, 2002; SAS, 1985), é expresso por:

$$z_{ij} = \sum_{p=1}^m a_{jp}F_{pi} + u_jY_{ji} \quad (i = 1,2, \dots, N; j = 1,2, \dots, n) \quad (1)$$

Em que:

$a_{jp} F_{pi}$: representa a contribuição do fator comum p à combinação linear;

$u_j Y_{ji}$: traduz o erro residual na representação da medida observada z_{ij} .

É ainda de salientar que pode ser assumido que os F 's e os Y 's têm médias zero e variância unitária, visto que na prática eles são desconhecidos. Além disso, os n fatores únicos supostamente são independentes e também independentes dos m fatores comuns. Neste modelo, os F 's são variáveis estatísticas ou variáveis aleatórias, definidos por uma função de densidade de probabilidade, que para certos propósitos, é tomada como sendo normal. A análise fatorial possui assim três etapas distintas, sendo elas:

- ❖ Preparação da matriz de correlação;
- ❖ Extração dos fatores comuns e possível redução do espaço;
- ❖ Rotação dos eixos relativos aos fatores comuns, visando uma solução mais simples e facilmente interpretável.

O uso de ferramentas analíticas, como o sistema de informação geográfica e estatísticas multivariadas que permitem lidar com complexas interações, tem-se tornado comum na gestão de bacias hidrográficas. Contudo, a sua eficácia depende da qualidade e quantidade de dados de campo, que, em geral, tendem a ser dispersos, principalmente quando se tratam de bacias hidrográficas inteiras. As bases de dados geradas por instituições governamentais com a finalidade de monitorização podem então ser úteis para um entendimento inicial dos processos e de interações que venham a ocorrer dentro da bacia hidrográfica e para mostrar pesquisas mais detalhadas, úteis aos planos de gestão, (Sliva & Williams, 2001). Desta forma, sugerem-se, (Ferrier, Edwards, & Hirst, 2001), quatro estágios necessários para estabelecer a gestão sustentável de uma bacia:

1. Determinar o estado atual do ambiente e identificar as forças dominantes de mudanças;
2. Estabelecer um limite específico acima do qual danos ecológicos são prováveis de ocorrer;
3. Prognosticar a possível extensão temporal e espacial do problema, usando características locais;
4. Desenvolver planos de gestão apropriados, através da utilização de cenários múltiplos de avaliação.

Para atingir esses estágios, será necessário utilizar ferramentas que permitam agrupar um grande número de informações decorrentes de estudos da variação qualitativa e quantitativa dos recursos hídricos, envolvendo uma função multivariada dos aspetos climatológicos, geomorfológicos, antropogénicos, entre outros, de forma que se torne possível a interpretação e o reconhecimento das tendências, ao longo do tempo e do espaço.

Segundo alguns autores, as análises multivariadas abrangem uma grande diversidade de técnicas, tais como: análises de *clusters*, de componentes principais (ACP), discriminantes e análises de redes neuronais. Todas estas são técnicas de identificação padrão, largamente aplicadas em estudos de avaliação de contaminações dos recursos hídricos, causadas tanto por fontes pontuais como difusas, (Vidal, López, & Santoalla, 2000).

A análise de componentes principais (ACP) é considerada uma das técnicas mais eficazes e comuns, no que respeita à redução da dimensão de grandes conjuntos de dados, sem perda da informação. Matematicamente, a ACP pode ser definida como combinações lineares

de variáveis aleatórias, com propriedades especiais, em termos da variância. Esta técnica tem como objetivo identificar um novo conjunto de variáveis (Y_1, Y_2, \dots, Y_n) como combinações lineares das variáveis originais (X_1, X_2, \dots, X_n) , de tal modo que os primeiros termos expliquem grande parte da variação dos dados e possuam uma interpretação prática. Com este procedimento pretende-se a obtenção de novas coordenadas, que tenham variância máxima e que não sejam correlacionadas entre si, (Bollmann & Marques, 2000; Lima, 2001).

A aplicação desta técnica tem sido largamente difundida nos últimos anos, devido à necessidade de se ter uma redução drástica dos dados para análise e tomada de decisão. Vários autores, como Berzas, Vidal, Vega, Camargo e Esteves, aplicaram esta técnica para reduzir o número de variáveis utilizadas, avaliando as variações espaciais e temporais da qualidade das águas em diferentes bacias, (Lima, 2001).

Um dos objetivos iniciais da ACP, consiste em verificar se poucos componentes conseguem explicar a maior parte da variância dos dados. Se tal ocorrer, pode reduzir-se a dimensão dos dados, permitindo compreender melhor a aproximação dos dados e as correlações entre as variáveis. Deste modo, consegue selecionar-se e reduzir os dados, através de eixos explicativos e da representação gráfica. Apesar das vantagens mencionadas, estes autores enfatizam que vários problemas podem ocorrer na aplicação desta técnica, principalmente na interpretação dos resultados, já que nem sempre os novos componentes podem apresentar uma explicação prática e clara, (Bollmann & Marques, 2000).

Relativamente à análise de *clusters*, esta engloba uma grande variedade de técnicas para análise exploratória dos dados. Um de seus principais objetivos é agrupar objetos dentro de classes, sendo que os objetos dentro de uma classe são similares, mas diferentes das outras classes. As características das classes só são conhecidas posteriormente, a partir do estudo dos dados. Existem duas grandes categorias utilizadas na análise de *clusters*: métodos hierárquicos e não-hierárquicos. O primeiro compreende uma forma sequencial de agrupamento, iniciando-se com os mais simples pares de objetos, até se formarem conjuntos maiores, passo a passo. Já os métodos não-hierárquicos, avaliam a distribuição global dos pares dos objetos, classificando-os dentro de um dado número de grupos, assumindo um número de *cluster* inicial. Em qualquer um dos métodos o processo de formação e junção dos agrupamentos é repetido, até um simples *cluster* conter todas as amostras. São várias as aplicações de análises de *clusters* para avaliação da qualidade da água, (Lima, 2001).

Além das técnicas multivariadas, outros métodos, como a matriz de correlação e as análises de séries temporais, são muitas vezes aplicados em estudo de dados relativos à área ambiental. Salienta-se ainda que a matriz de correlação avalia a associação entre variáveis, tentando estabelecer a importância de cada uma como determinante para o comportamento da variável dependente que se quer estudar. Os procedimentos mais usuais até agora utilizados são o da correlação duas a duas e o da correlação de uma variável com uma combinação linear das outras. Mais recentemente, o uso da matriz de correlações tem ganho uma nova importância na escolha de indicadores ambientais e por isso estas aplicações podem ser observadas em vários estudos, (Bollmann & Marques, 2000).

2.2. Tipos de modelos matemáticos hidrológicos

Os modelos hidrológicos matemáticos podem dividir-se em dois grupos: os determinísticos e os não determinísticos.

Os modelos determinísticos englobam os empíricos e os baseados em fenómenos físicos, enquanto que os não determinísticos incluem os probabilísticos e os estocásticos.

Quando os valores da série podem ser escritos através de uma função matemática perfeitamente determinada por uma ou mais variáveis, está-se perante um modelo determinístico. Deste modo, os modelos determinísticos podem ser:

- ❖ Empíricos, se a análise dos processos hidrológicos em jogo se reduz à procura de uma relação causa-efeito entre variáveis de entrada (dados) e variáveis de saída (resultados);
- ❖ Baseados em fenómenos físicos, se se procuram reproduzir as leis físicas que regem os fenómenos. Estes modelos podem ainda subdividir-se em agregados e distribuídos, consoante as variáveis hidrológicas integradas na formulação exprimam valores médios no espaço ou valores em pontos desse espaço.

Os modelos determinísticos caracterizam-se então pelo facto de uma dada ação implicar uma resposta. Se tal resposta for aleatória, pelo menos em parte, está-se perante um modelo não determinístico. Assim, os modelos não determinísticos, podem ser:

- ❖ Probabilísticos, quando as variáveis intervenientes possuem um comportamento puramente aleatório, sendo portanto, ignorada a sequência temporal ou a sequência espacial dos valores dessas variáveis;
- ❖ Estocásticos, quando além da componente aleatória se verifica uma outra que permite contemplar a sequência temporal ou a sequência espacial dos valores das variáveis.

A distinção anteriormente descrita é muito frequentemente adotada em Hidrologia, devendo, porém, notar-se que alguns autores incluem os modelos probabilísticos no âmbito geral dos modelos estocásticos, (Quintela & Portela, 2002).

Uma série temporal pode ser encarada como um conjunto de observações de um processo estocástico, isto é, um conjunto ordenado de variáveis aleatórias no tempo, no espaço ou num outro conjunto.

Os modelos temporais podem ter muitas observações, pelo que se tornam sujeitos a perturbações de natureza aleatória, sendo necessário descobrir o modelo que melhor descreve o seu comportamento.

Um processo estocástico, numa análise de séries temporais, é qualquer coleção de variáveis aleatórias $X(t)$, $t \in T$, em que T é um conjunto de índices representando o tempo. O conjunto T é denominado espaço de parâmetros (\mathbb{R} , \mathbb{R}^+ , N ou \mathbb{Z}).

O conjunto S , contradomínio das variáveis $X(t)$, é denominado espaço de estados (\mathbb{R} , \mathbb{Z} , N , \mathbb{R}^2 , ...).

Qualquer sucessão de variáveis aleatórias independentes e identicamente distribuídas, X_n , em que o índice n corresponde à seriação cronológica de ordem n , é um processo estocástico, cujo espaço de parâmetros são os naturais. O espaço de estados tem a ver com o tipo de variáveis da sucessão e, naturalmente, pode ser muito variado. Se as variáveis possuírem distribuição Normal, $N(0; \sigma^2)$, tem-se $S = \mathbb{R}$, (Silva, 2013).

Quanto à descrição e à caracterização do comportamento do processo, dado um processo estocástico $X(t)$, tal que para todo o t , tem-se $E[X(t)^2] < +\infty$:

❖ Função de valor médio:

$$\mu(t) = E[X(t)] \quad (2)$$

❖ Função de Variância:

$$\sigma^2(t) = Var[X(t)] = E[(X(t) - \mu(t))^2] \quad (3)$$

❖ Função de Covariância:

$$\rho(t_1, t_2) = Cov[X(t_1), X(t_2)] = E[(X(t_1) - \mu(t_1))(X(t_2) - \mu(t_2))] \quad (4)$$

❖ Função de Correlação:

$$\rho(t_1, t_2) = \frac{Y(t_1, t_2)}{\sigma(t_1)\sigma(t_2)} = \frac{Cov[X(t_1), X(t_2)]}{\sqrt{Var[X(t_1)]Var[X(t_2)]}} \quad (5)$$

Sendo, $\sigma^2(t) = Cov[X(t), X(t)] = Y(t, t)$.

2.3. Séries temporais

Pode definir-se uma série temporal ou histórica como um conjunto de observações feitas em pontos ou períodos sucessivos de tempo, durante um determinado intervalo, (Murteira, Muller, & Turkman, 1993). Assim, a análise estatística dos modelos de séries temporais tem aplicações em diferentes áreas da ciência, como a engenharia, economia, meteorologia, entre outras.

As séries temporais podem ser estudadas e analisadas de formas distintas, sendo que para se prever o comportamento de determinados fenômenos ao longo do tempo, é necessário determinar o modelo que melhor descreve o comportamento observado, isto é, estabelecer uma relação em que a variabilidade dos dados é explicada a partir de funções determinísticas do tempo e do seu próprio passado já observado. Em geral, a série é composta por uma componente estocástica e uma componente determinística. A primeira está associada a variações irregulares ou aleatórias e a segunda a variações sazonais.

Ajustar um modelo a um determinado acontecimento pode revelar-se uma tarefa bastante difícil, na medida em que podem sempre ocorrer erros de medição ou podem ser

recolhidos dados que foram influenciados por fatores externos, sobre os quais não se tem nem controlo nem conhecimento. Uma previsão a partir da série temporal procura, deste modo, construir um modelo matemático, a partir do qual seja possível prever valores futuros da série com base em acontecimentos passados. Modelos estes, que devem, no entanto, ser ajustados de modo a que se obtenham previsões fiáveis de observações futuras.

As séries temporais podem representar-se por meio de várias funções matemáticas, tais como linear, quadrática, exponencial, etc., podendo portanto depender de mais do que uma variável.

No que respeita aos modelos de regressão linear, a ordem das observações é irrelevante para a análise, já em séries temporais a ordem dos dados é fundamental. Uma característica muito importante deste tipo de dados é que as observações vizinhas são dependentes e o interesse consiste em analisar e modelar esta dependência.

Os modelos para estudar as séries temporais são geralmente construídos a partir de modelos auto regressivos, modelos integrados e modelos de média móvel.

2.3.1. Componentes das séries temporais

Uma série temporal pode ser composta por três componentes não observáveis:

- ❖ Tendência (T_t);
- ❖ Sazonalidade (S_t);
- ❖ Variação aleatória denominada de ruído branco (a_t).

A primeira escolha para a elaboração de um modelo seria um relacionamento adicional destas componentes: $Z_t = T_t + S_t + a_t$. Pode ainda construir-se um modelo multiplicativo ($Z_t = T_t \cdot S_t \cdot a_t$) ou realizar-se a transformação *log*, no modelo multiplicativo, quando ele se transforma no modelo *log-linear*.

Ao analisar uma série temporal, deve estudar-se cada uma destas componentes separadamente, retirando-se o efeito das outras.

Para analisar a tendência, os dois métodos mais utilizados são o ajuste de uma função polinomial do tempo e a análise do comportamento da série em volta de um ponto, estimando a tendência naquele ponto. Na primeira opção, utilizam-se os modelos de regressão polinomial e na segunda os modelos auto regressivos, (Latorre & Cardoso, 2001; Morettin & Tolo, 1985; Diggle, 1992).

2.3.2. Modelos que caracterizam as séries temporais

O objetivo principal da análise de séries temporais consiste em encontrar um bom modelo que tenha em conta as relações existentes entre as observações, permitindo, deste modo, a sua descrição. Um modelo a escolher será aquele que inclui o menor número de parâmetros estimados, mas que a perda de qualidade do ajustamento não seja significativa, (Silva, 2013).

Na análise de uma série temporal, primeiramente deseja-se modelar o fenómeno estudado para, a partir daí, se descrever o comportamento da série, fazer estimativas e, por último, avaliar quais os fatores que influenciaram o comportamento da série, procurando definir relações de causa e efeito entre duas ou mais séries. Para tal, há um conjunto de técnicas estatísticas disponíveis que dependem do modelo definido (ou estimado para a série), bem como do tipo de série analisada e do objetivo do trabalho, (Latorre & Cardoso, 2001).

Para a análise de tendências, podem-se ajustar modelos de regressão polinomial baseados na série inteira, na vizinhança de um determinado ponto ou também com funções matemáticas.

Define-se como um fenómeno sazonal aquele que ocorre regularmente em períodos fixos de tempo e, se existir sazonalidade dita determinística na série, podem-se utilizar modelos de regressão que incorporem funções do tipo seno ou cosseno à variável tempo.

Os modelos auto regressivos formam outra classe de modelos. Na análise do comportamento de uma série temporal, livre de tendência e de sazonalidade, podem ser utilizados modelos auto regressivos (AR) ou que incorporem médias móveis (ARMA). Quando há tendência, utilizam-se os modelos auto regressivos integrados de médias móveis (ARIMA) e, para incorporar a componente de sazonalidade, utilizam-se os modelos SARIMA.

Por último, existem os modelos lineares generalizados. Neste grupo de modelos estatísticos, a variável resposta é um processo de contagem e as variáveis independentes são variáveis que têm como objetivo explicar o comportamento da série ao longo do tempo. Estes modelos são mais indicados para casos em que as variáveis em estudo não têm aderência à distribuição Normal, principalmente pelo fato de serem processos de contagem. Estes modelos constituem um grupo de distribuições de probabilidades, conhecido como família exponencial de distribuições, que englobam diversas funções, como a regressão linear, de *Poisson*, *log-linear*, etc.. Os modelos generalizados são uma extensão desta classe de modelos, nos quais cada variável independente analisada não entra no modelo com o seu valor, mas sim, adotando uma função não paramétrica de forma não especificada, (Latorre & Cardoso, 2001).

Uma série temporal é uma sequência de dados obtidos em intervalos regulares de tempo durante um período específico, (Latorre & Cardoso, 2001; Everitt, 1995; Moretinn & Toloï, 1985). Este conjunto pode ser obtido através de observações periódicas do evento de interesse, ou através de processos de contagem. Se a série temporal for denominada como Z , o valor da série no momento t pode ser escrito como $Z_t (t=1,2,\dots,n)$. Denomina-se trajetória de um processo, a curva obtida no gráfico da série temporal, sendo que o conjunto de todas as possíveis trajetórias é denominado processo estocástico. Considera-se, portanto, que uma série temporal é uma amostra deste processo. O conjunto de observações ordenadas no tempo pode ser discreto ou contínuo. Pode obter-se uma série temporal discreta a partir de uma amostra de pontos de uma série contínua ou por meio de um parâmetro como, por exemplo, a média de períodos fixos de tempo.

2.3.2.1. Modelos de regressão polinomial

Nos modelos de regressão polinomial, os valores da série são considerados como variável dependente (Y) e os períodos do estudo como variável independente (X).

Primeiramente deve fazer-se o diagrama de dispersão de $Z_t(Y)$ em relação ao tempo, a fim de se verificar qual a função que melhor se ajusta à trajetória do processo: linear, parábola, exponencial, etc.. Para se evitar a correlação entre os termos da equação de regressão, recomenda-se fazer a transformação da variável período na variável período-centralizada (período menos o ponto médio da série temporal), estimando-se assim, o modelo de regressão

correspondente. No entanto, por vezes, devido à grande oscilação dos pontos, é necessário suavizar a série, reduzindo o ruído branco, (Draper & Smith, 1981).

As vantagens da estimação da tendência utilizando modelos de regressão polinomial são o grande poder estatístico desta classe de modelos, a fácil elaboração e a interpretação. Porém, algumas vezes não há uma função definida, como a linear ou exponencial, tornando-se necessário o ajuste a outra função matemática ou utilizar outra classe de modelos.

A segunda opção para a análise de séries temporais seria a estimação da tendência, analisando o comportamento da série em volta de um ponto, estimando a tendência para valores da série próximos a ele e não utilizando a série como um todo. A análise utilizando parte da série é mais recomendada quando se deseja avaliar apenas uma parte da trajetória ou quando o comportamento da série é muito instável. Nesta situação, as melhores projeções são feitas a partir de um passado recente da mesma.

Após a estimativa da tendência, para se analisar as outras componentes será necessário construir uma série livre de tendência, através das diferenças da mesma ($D_d Z_t$), onde d é o grau do polinómio obtido na análise da tendência. Por exemplo, se a tendência obtida for linear (primeiro grau), bastaria fazer uma diferença da série Z_t ($Z_t - Z_{t-1}$), para que ela estivesse livre de tendência, (Diggle, 1992; Moretinn & Toloï, 1985).

Os modelos auto regressivos e de médias móveis (ARMA) são bem conhecidos e considerados adequados para analisar séries estacionárias, (Box & Jenkins, 1976). No entanto, nem todas as séries encontradas na natureza podem ser modeladas como processos ARMA.

Em algumas situações práticas, encontram-se séries temporais que apresentam um comportamento sazonal ou periódico.

❖ Sazonalidade

A sazonalidade é uma componente da série temporal difícil de ser estimada, pois é necessário compatibilizar a questão conceitual do fenómeno em estudo com a questão estatística. Define-se um fenómeno sazonal como aquele que ocorre regularmente em períodos fixos de tempo, (Diggle, 1992; Everitt, 1995; Moretinn & Toloï, 1985).

Os índices sazonais servem então para revelar a evolução das utilizações, ao longo dum determinado período. Estes índices podem ser facilmente determinados recorrendo ao conceito de médias móveis que incluem, dinamicamente, os registos mais recentes. A caracterização do comportamento típico dos valores de caudal, ou dum qualquer parâmetro de qualidade da água, ajuda na identificação de comportamentos anómalos, eventualmente, indicadores da presença de poluentes e/ou contaminantes na água, (Silva, Salgado, & Menezes, 2009).

No caso de existir uma sazonalidade dita determinística podem ser utilizados modelos de regressão que incorporem funções do tipo seno ou cosseno à variável tempo. Para se descobrir se existe sazonalidade na série de valores e verificar qual o seu ritmo é importante realizar uma análise espectral. Com este tipo de análise é possível identificar um padrão sazonal, mesmo dentro de uma variabilidade aleatória. A análise espectral utiliza um conjunto de funções que contém seno e cosseno e tenta ajustá-las à variância observada numa série de observações no tempo, tendo em conta a amplitude das "ondas", o período em que elas se repetem e a fase em que se iniciam (frequência), (Diggle, 1992).

Assim, a componente sazonal ou estacional (E_t) pode ser caracterizada usando-se modelos determinísticos através de uma análise de regressão em termos de polinómios trigonométricos, já que muitas funções periódicas podem representar-se por uma combinação linear de senos e cossenos, (Murteira, Muller, & Turkman, 1993).

Uma alternativa, mais simples e muito utilizada para a definição de curvas padrão, consiste na utilização do método das médias simples ou móveis, por exemplo, para sazonalidade anual, seria utilizada a média móvel de 12 meses. Esta metodologia tem como principal vantagem a sua fácil programação ou utilização, já que qualquer *software* comercial dedicado a *time series* tem a possibilidade do cálculo e representação gráfica destes índices sazonais. No modelo multiplicativo, que é o tradicionalmente usado, a componente estacional representa a taxa de acréscimo em relação ao valor médio, (Silva, Salgado, & Menezes, 2009). Salienta-se ainda que trabalhar-se com a diferença entre a série original (Z_t) e o polinómio estimado para a sazonalidade.

Para as séries sazonais, muitas vezes a correlação entre observações vizinhas, Z_t e Z_{t-1} , dependem de t . Nestes casos, é mais adequado ajustar um modelo cujos parâmetros também

variem de forma periódica, em vez de um modelo com parâmetros constantes, para todos os instantes t . Assim, quando a correlação entre os meses se modifica de forma periódica, as séries designam-se periodicamente estacionárias, (Silva, 2013).

Segundo Hurd & Gerr, a correlação periódica pode ser analisada por meio dos gráficos de dispersão de diferentes períodos. Uma classe de processos periódicos muito utilizada é a dos modelos auto regressivos periódicos (PAR), estudados inicialmente por Thomas & Fiering e Jones & Brelsford e amplamente analisados em Pagano, onde foram apresentadas as propriedades destes modelos, assumindo normalidade do processo, e Troutman, que mostra as condições necessárias e suficientes para a estacionariedade periódica. A estacionariedade periódica dos modelos PAR conduz às funções de autocorrelação e autocorrelação periódicas. Nesse sentido, vale ressaltar os trabalhos de Anderson & Vecchia, que desenvolveram alguns resultados para a função de autocorrelação periódica amostral, e Sakai, que estudou a função de autocorrelação parcial periódica dos processos PAR. Outras propriedades dos modelos PAR e suas relações com os modelos SARMA, podem ser encontradas em Silva (2013).

Relativamente à estacionariedade, um processo diz-se estacionário, quando o sistema se apresenta num estado de equilíbrio estatístico, em torno de um nível médio fixo, ou seja, tem propriedades estatísticas que não variam no tempo. Normalmente, as séries temporais não apresentam o comportamento de um processo estacionário. No entanto, recorrendo a transformações que estabilizam a média e/ou a variância pode converter-se uma série temporal não estacionaria, numa série estacionaria, (Silva, 2013).

2.3.2.2. Modelos auto regressivos

Numa fase inicial é essencial definir se a série é estacionária ou não, para, a partir daí, se estabelecer a estrutura do modelo probabilístico que estimará a série. Uma série é considerada estacionária quando as suas observações ocorrem aleatoriamente ao redor de uma média constante, ou seja, não há tendência. Isto significa que $E(Z_t)=E(Z_{t+m})=m$ e $Var(Z_t)=Var(Z_{t+m})=constante$. Para tal, define-se a função de autocorrelação que, em cada período j da série, é calculado o coeficiente de correlação entre as observações t e $t+j$. Neste caso, se t e $t+j$ são independentes, a correlação entre t e $t+j$ são zero.

O modelo mais simples é obtido para a série temporal estacionária, ou seja, livre de tendência e de sazonalidade. Esta série é consequência da variação aleatória do ruído branco em volta de uma grande média, ao longo do tempo. Esta pode ser escrita como a combinação aleatória dos valores anteriores da Z_t ($Z_t = b_1 \cdot Z_{t-1} + b_2 \cdot Z_{t-2} + \dots + b_p \cdot Z_{t-p}$) e, por isso, a série toda pode ser função do ruído branco. Esta classe de modelos é conhecida como modelos auto regressivos - AR - (no caso, de ordem p). Salienta-se ainda que este é um processo iterativo, onde há a identificação da ordem p através da função de autocorrelação, sendo que é a partir daí que se faz a estimativa de um modelo de previsão, bem como a análise ou estimativa de erros. A dificuldade desta técnica reside na identificação do modelo, pois é possível que pessoas diferentes identifiquem modelos de ordem diferentes para a mesma série temporal.

Para muitas séries, a melhor solução consiste em combinar o modelo auto regressivo (AR) com o de médias móveis (MA). Este é composto pela combinação linear de valores próximos da série (AR de ordem p) com uma combinação linear dos ruídos brancos próximos ao valor da série (MA de ordem q).

Tanto o modelo AR, como o MA e o ARMA são utilizados para séries estacionárias.

Por outro lado, quando o processo é não estacionário homogêneo, uma das maneiras de analisá-lo é incorporando um processo de diferenças no modelo ARMA. Este é o modelo conhecido como ARIMA (modelo auto regressivo integrado de médias móveis), onde d é a ordem das diferenças necessárias para tirar a tendência da série. Há duas situações em que a série pode ser considerada não estacionária:

- ❖ Quando durante um período os pontos oscilam em volta de uma média e, depois, mudam de patamar (neste caso basta tomar uma diferença da série);
- ❖ Quando a série é não estacionária em relação à tendência (geralmente, para torná-las estacionárias é necessário tomar a segunda diferença).

Os modelos ARIMA podem dar conta da sazonalidade, porém, quando a sazonalidade ocorre em múltiplos períodos, é necessário que se considere no modelo uma componente de sazonalidade estocástica. Nesta situação, utiliza-se o modelo SARIMA que incorpora as funções trigonométricas (preferencialmente, seno e cosseno) ao modelo ARIMA, e a ordem da sazonalidade vai depender da série, (Diggle, 1992; Moretinn & Toloï, 1985).

❖ Modelo ARIMA

O modelo ARIMA é muito utilizado na modelação e em previsões de séries temporais, tratando-se de um modelo auto regressivo, integrado e de média móvel. Este modelo foi criado pelos estatísticos George Box e Gwilym Jenkins, o que torna o modelo também conhecido como Modelo de Box-Jenkins.

O modelo ARIMA surge então como uma generalização do modelo auto regressivo de média móvel (ARMA). A representação ARIMA (p,d,q) refere-se, respetivamente, às ordens de auto regressão, de integração e de média móvel em que p representa o número de termos auto regressivos, d o número de diferenças e q o número de termos da média móvel.

Existem vários métodos possíveis para ajustar um modelo para uma série temporal, sendo o modelo de Box e Jenkins dos modelos de séries temporais mais importantes e amplamente utilizados, (Shahwan & Odening, 2007). Este modelo revela-se bastante fiável na previsão da qualidade da água, permitindo revelar informações importantes sobre as variações químicas e biológicas e também sobre as intervenções causadas pelo Homem e/ou pela sazonalidade.

Box e Jenkins (1976), apresentaram uma metodologia, de modelação de sucessões cronológicas, que inclui as componentes de identificação, estimação e avaliação do diagnóstico. O esquema geral proposto para a modelação e previsão de sucessões cronológicas de parâmetros de qualidade da água, segundo estes autores, é o seguinte:

1. Análise da sucessão cronológica de parâmetros de qualidade da água e de caudais;
2. Identificação dos intervalos de tempo com medições anómalas, por exemplo valores nulos que resultam de avaria da sonda, e remoção destes valores da série;
3. Caracterização da sazonalidade e determinação dos índices sazonais;
4. Identificação dos parâmetros a incluir no modelo;
5. Estimação dos parâmetros do modelo;
6. Avaliação do diagnóstico e avaliação da qualidade do modelo;
7. Estudo de modelos alternativos;
8. Seleção do modelo;
9. Realização de previsões.

A metodologia deve incluir, desde logo, como tarefa preliminar a deteção de erros ou a falta de coerência nos dados, já que a eficiência de um modelo de previsão está muito dependente da informação base usada. A caracterização quantitativa da sazonalidade (ou periodicidade) pode fazer-se com recurso a um procedimento que determina índices sazonais baseado no conceito de médias móveis, (Silva, Salgado, & Menezes, 2009; Box & Jenkins, 1976).

Deste modo, na etapa da identificação, o objetivo consiste em procurar um modelo que descreva adequadamente a sucessão cronológica. Para tal, o primeiro passo consiste na estacionarização da série temporal. Após se ter a série temporal estacionária, o passo seguinte é o da escolha de um modelo ARMA (auto regressivo e de médias móveis).

Para a identificação de modelos, tomando como exemplo uma sucessão de valores horários de caudais, ou de um parâmetro de controlo da qualidade da água com observações até ao instante t , $\{C_t, C_{t-1}, C_{t-2}, \dots\}$, o objetivo consiste em caracterizar o valor no instante t , C_t que se considera ser função dos valores históricos, ou seja:

$$C_t = f(C_{t-1}, C_{t-2}, \dots, C_{t-24}, \dots, C_{t-48}, \dots) \quad (6)$$

Considerando um modelo ARIMA $(p,d,q)(P,D,Q)_S$ univariado, este pode ser representado por:

$$\Phi(B)[\nabla^d \nabla_S^D (C_t - \mu)] = \Theta(B)a_t \quad (7)$$

Onde:

- ❖ p e P representam a ordem do polinómio auto regressivo da parte não sazonal e sazonal, respetivamente;
- ❖ q e Q são a ordem do polinómio de médias móveis da parte não sazonal e sazonal, respetivamente;
- ❖ ∇^d e ∇^D são os operadores diferença;
- ❖ $\Phi(B) = \phi_p(B)\Phi_P(B)$ com $\phi_p(B)$ e $\Phi_P(B)$ representam os polinómios auto regressivos da parte não sazonal e sazonal, respetivamente;

- ❖ $\theta(B) = \theta_q(B)\theta_Q(B)$ com $\theta_q(B)$ e $\theta_Q(B)$ representam os polinômios de médias móveis da parte não sazonal e sazonal, respectivamente;
- ❖ a_t representa um ruído branco;
- ❖ μ é uma constante opcional que identifica o valor esperado para C_t .

Caso se pretendam incluir variáveis independentes explicativas, o modelo pode ser escrito como:

$$\Phi(B) \left[\nabla \left(C_t - \sum_{i=1}^m c_i x_{it} - \mu \right) \right] = \theta(B) a_t \quad (8)$$

sendo:

- ❖ $\nabla = (1 - B)^d(1 - B^S)^D$;
- ❖ x_i : variável independente i ;
- ❖ c_i : coeficiente de regressão.

Os processos mistos (auto regressivos e médias móveis) têm-se revelado como modelos parcimoniosos, isto é, com poucos parâmetros, o que é uma característica a procurar. Em alternativa, pode procurar-se outros modelos como o seguinte:

$$C_t = (\phi_1 C_{t-1}) \frac{E_t}{E_{t-1}} + \phi_2 C_{t-24} + \theta_1 \left(\frac{C_{t-1} + C_{t-2}}{2} \right) + \theta_2 \left(\frac{C_{t-24} + C_{t-48}}{2} \right) + \varepsilon_t \quad (9)$$

Onde ϕ_1, ϕ_2, θ_1 e θ_2 são números reais e ε_t o erro aleatório, (Silva, Salgado, & Menezes, 2009).

Geralmente nos processos não estacionários a função média exhibe uma forma que pode ser mais ou menos evidente como função do tempo e, pode integrar a tendência ou ter uma componente sazonal que pode ser ou não, susceptível de interpretação direta. Uma transformação muito usual para neutralizar a tendência, ou seja, para estabilizar a média das séries temporais não estacionárias, que faz parte da classe dos filtros lineares, é o "operador diferença". O objetivo passa por determinar a menor diferenciação da série estacionária, dado que, a variância aumenta conforme aumenta a ordem das diferenças. Outra transformação que permite eliminar movimentos estritamente periódicos da série temporal é o "operador de diferença sazonal". Todas estas transformações e sua utilização devem ser feitas com bastante

moderação, para não se introduzirem alterações desnecessárias na variação da série e para que não se perca informação significativa dos dados.

Muitas séries temporais são não estacionárias em variância, quer sejam ou não estacionárias em média, sendo necessário determinar que tipo de transformações a usar para estabilizar a variância. Box e Cox introduziram uma transformação paramétrica para estabilizar a variância. Estas transformações estão apenas definidas para séries de valores positivos. Numa série que não seja estacionária em média, nem em variância, deve-se estabilizar primeiro a variância e só depois a média, dado que, o cálculo das diferenças para estabilizar a média pode introduzir uma diferente variação na série, o que dificulta a escolha da transformação para estabilizar a variância. A análise do gráfico dos valores da série ao longo do tempo, normalmente chamado de cronograma, permite muitas vezes detetar a instabilidade da variância, qual o tipo de tendência e a existência de movimentos periódicos, indicando quais as transformações mais adequadas a utilizar no processo de estacionarização. Para se averiguar sobre a normalidade da série temporal transformada, pode-se analisar através da inspeção visual do seu histograma e da sua Normal *q-q plot*, e efetuar o teste de hipóteses de Shapiro-Wilk. Este teste tem como hipótese nula que a série temporal transformada segue uma distribuição Normal e tem a estatística de teste, (Silva, 2013).

A seleção de um modelo ARMA é o segundo passo dentro da etapa da identificação, sendo essencial para uma correta identificação o cálculo e análise das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) dos modelos mais utilizados. Estas funções medem, para cada valor, a intensidade com que pares de valores do processo separados por um intervalo variam entre si (covariam).

Além de se investigar a correlação total de uma forma global, interessa também investigar a correlação parcial (FACP). Após calcular as FAC e FACP amostrais da série corretamente transformada, é necessário encontrar as ordens de p e q , com vista a ajustar um dos modelos ARMA (p, q) . Esta parte é bastante importante no processo de análise de séries temporais, pois a escolha incorreta dos parâmetros p e q pode levar a erros significativos de previsão, (Silva, 2013).

Muitas vezes a FAC e FACP teóricas dos modelos ARMA não correspondem à estrutura apresentada pelas FAC e FACP amostrais do modelo ajustado. Em alguns modelos

mistos a identificação do modelo é particularmente difícil. Nestes casos é necessário adotar critérios adicionais para selecionar o melhor modelo. Um critério a utilizar pode ser o critério Akaike, AIC – *Akaike Information Criteria*, que é definido de forma a penalizar os modelos com maior número de parâmetros e com maior variância residual. De acordo com este critério, em caso de dúvida entre dois modelos, deve selecionar-se o modelo para o qual o correspondente AIC tenha o menor valor, (Silva, 2013).

A etapa seguinte é a da estimação dos parâmetros do modelo selecionado. Para a estimação dos parâmetros dos modelos auto regressivos e de médias móveis existem três metodologias mais utilizadas que são: a máxima verosimilhança, os mínimos quadrados e os momentos, sendo que o primeiro é normalmente o preferido para estimar os parâmetros de modelos Box-Jenkins. Nesta etapa o analista não tem grande influência, uma vez que a estimação dos parâmetros do modelo pode ser feita por vários programas informáticos.

Assim, na estimação dos parâmetros do modelo procede-se à estimação dos parâmetros: $\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q$. A estimativa dos parâmetros usando o método dos mínimos quadrados pode fazer-se minimizando:

$$S(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q) = \sum_{t=1}^N (C_t \phi_1 C_{t-1} \phi_2 C_{t-2} - \dots)^2 \quad (10)$$

Ora, a minimização da soma do quadrado dos erros pode ser conseguida adotando uma ferramenta de otimização, que recorre ao conceito de gradiente reduzido generalizado, o *GRG2* desenvolvido por Leon Lasdon da Universidade do Texas e Allan Waren da Universidade de Cleveland, incorporado no *Microsoft Excel*. O problema pode ser formulado através de:

Minimizar:

$$\sum_{t=1}^N \varepsilon_t^2 = f(\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q) \quad (11)$$

sujeito a : $0 \leq \phi_i \leq 1; e 0 \leq \theta_i \leq 1$

Onde:

$$\sum_{t=1}^N \varepsilon_t^2 = \sum_{t=1}^N [C_t(\text{real}) - C_t(\text{mod})]^2 \quad (12)$$

Sendo $C_t(\text{real})$ o valor real do parâmetro no instante t e $C_t(\text{mod})$ o valor determinado pelo modelo para o mesmo instante t , (Silva, Salgado, & Menezes, 2009).

A avaliação do diagnóstico representa a última etapa, na qual se pretende avaliar a qualidade estatística do modelo e a qualidade do seu ajustamento. Representa uma etapa bastante importante, na medida em que se vai aceitar ou rejeitar o modelo como adequado para efetuar previsões. No caso de o modelo ser rejeitado, deve voltar-se novamente à etapa da identificação e escolher um modelo alternativo.

Assim, para a avaliação da qualidade do modelo podem seguir-se os diversos procedimentos recomendados em publicações dedicadas à modelação de sucessões cronológicas, como Murteira, (Silva, Salgado, & Menezes, 2009). Os passos são:

1. Verificar se os parâmetros estimados são significativamente diferentes de zero;
2. Verificar se os parâmetros estimados estão afastados das regiões de não estacionaridade e de não invertibilidade (o rácio t em módulo, $|t| \geq 1$);
3. Verificar se as correlações entre estimadores apresentam valores $< 0,7$. Sempre que dois parâmetros tiverem um coeficiente de correlação superior a 0,7 devem procurar-se modelos alternativos, dado que esta situação torna o modelo instável;
4. Verificar se a função de autocorrelação (FAC) residual tem valores dentro dos limites críticos;
5. Verificar se a estatística Q é manifestamente inferior ao quantil 0,95 da distribuição Qui-quadrado com $m-p-q$ graus de liberdade. Esta análise é importante, dado que um modelo com poucos parâmetros, em geral, é capaz de produzir melhores previsões;
6. Verificar se a função de autocorrelação parcial (FACP) residual tem valores dentro dos limites críticos.

A análise da qualidade da aproximação entre o conjunto de valores reais e o determinado pelo modelo pode, também, fazer-se usando como critério o coeficiente de correlação e a raiz do erro quadrático médio:

$$REQM = \sqrt{\frac{[C_i(real) - C_i(mod)]^2}{N}} \quad (13)$$

Uma vez identificado, estimado e avaliado a consistência do modelo efetuam-se as previsões. Os valores previstos das diferentes variáveis em estudo vão ser comparados com os seus valores observados para o mesmo período de tempo, e assim pode observar-se a eficácia das previsões dos modelos elaborados.

Salienta-se que alguns parâmetros de qualidade da água de um rio, onde existem infraestruturas para aproveitamento hidroelétrico, apresentam um carácter periódico com ciclo idêntico às descargas realizadas a montante. Será, deste modo, sensato investigar a hipótese dos valores do parâmetro em análise num dado instante estarem relacionados com os valores ocorridos nos instantes imediatamente anteriores, com os valores do mesmo instante dos ciclos anteriores e com os valores dos caudais descarregados a montante.

Em suma, na etapa de estimação são determinados os parâmetros do modelo, enquanto na avaliação do diagnóstico é feita a verificação e adequação do modelo, já identificado e estimado, devendo seguir-se duas subetapas: controlo da qualidade estatística do modelo e da qualidade do ajustamento, (Silva, Salgado, & Menezes, 2009).

2.3.2.3. Modelos lineares generalizados

Neste grupo de modelos estatísticos a variável dependente ou resposta (Y) é um processo de contagem e as variáveis independentes são variáveis que explicam o comportamento da série ao longo do tempo. Esta classe de modelos é indicada em casos que as variáveis em estudo não têm aderência à distribuição normal, principalmente pelo fato de serem processos de contagem (ou seja, são variáveis quantitativas discretas).

Estes modelos compõem um grupo de distribuições de probabilidades conhecido como família exponencial de distribuições que englobam diversas funções, como a regressão linear, de *Poisson*, *log-linear* etc.. Os modelos generalizados são uma extensão desta classe de modelos, nos quais cada variável independente analisada não entra no modelo com o seu valor, mas sim, adotando uma função não paramétrica de forma não especificada. Sendo assim, não é necessário assumir uma relação linear entre a variável dependente e a variável independente em estudo. A trajetória proporciona a visualização não só da forma, mas também das possíveis não linearidades nas relações estudadas, uma vez que não apresenta uma função paramétrica rígida.

A disponibilidade de diversos modelos estatísticos permite a utilização da melhor técnica estatística, porém o fascínio pela técnica não deve obscurecer o desenvolvimento das várias hipóteses existentes, (Latorre & Cardoso, 2001).

2.3.2.4. Outros modelos

São diversos os modelos e as técnicas, como os métodos estatísticos convencionais, os modelos estocásticos auto regressivos, integrados e de médias móveis ARIMA e os métodos heurísticos, designadamente as redes neuronais artificiais, utilizados para a análise e previsão de sucessões cronológicas. Por exemplo, em Sun, Silva e Anderson, podem ser encontrados exemplos da aplicação de técnicas de análise e previsão de sucessões cronológicas. Srinivasulu, usa em conjunto e de forma integrada diversas ferramentas, entre as quais as redes neuronais artificiais e algoritmos genéticos para fazer a análise e previsão de caudais de rios, (Silva, Salgado, & Menezes, 2009).

A modelação por redes neuronais artificiais consiste em estimar os pesos de cada variável de entrada introduzida na rede, de forma que a variável resposta seja a mais próxima da realidade possível. Existem vários algoritmos que realizam essa tarefa e, basicamente, estes diferem entre si na forma como são realizados os ajustes dos pesos. A escolha pelo algoritmo a ser utilizado deve-se basear nas tarefas que se desejam realizar. Prestam-se, portanto, para classificação ou reconhecimento de padrões, aproximação de funções, previsão e controlo, (Lima, 2006).

A ferramenta de modelação de redes neuronais artificiais é capaz de fazer previsões de concentração de um poluente num determinado local, a partir de variáveis de entrada (por exemplo parâmetros meteorológicos) e de saída (que podem ser, por exemplo, dados históricos de medição de concentração desse poluente recolhidos num determinado período). Na modelação, geralmente é testado um modelo de multicamadas, com diversas configurações. Fazem-se variar a quantidade de neurónios na camada de entrada, na intermediária, na de saída e também os algoritmos. No que respeita à validação do modelo, esta pode ser realizada por meio da comparação entre os valores previstos e os reais medidos em alguns pontos de monitorização. Os resultados deste género de modelos devem, tal como todos os outros, ser comparados com modelos alternativos. Devido à sua versatilidade, as redes neuronais artificiais têm sido bastante aplicadas em diversas áreas do conhecimento, em especial na modelação ambiental, (Lima, 2006).

Alguns autores procuram associar modelos de sucessões cronológicas a modelos determinísticos de simulação dos fenómenos, como o escoamento e transporte de poluentes na água. São exemplos Bruen e Vazquez-Amabile, (Silva, Salgado, & Menezes, 2009).

Em jeito de conclusão, salienta-se que deve sempre ter-se em conta a diversidade e complexidade dos modelos existentes, considerando modelos alternativos, eventualmente vizinhos do primeiro modelo selecionado, já que muitas vezes existem alguns modelos a descrever a sucessão cronológica com igual elevado grau de qualidade. A seleção do modelo a adotar poderá basear-se em critérios como os apresentados por Akaike (AIC) ou a extensão Bayesiana (BIC). O coeficiente de correlação (r) e a raiz do erro quadrático médio (REQM) poderão, também, ser usados para este efeito. Sendo o objetivo último da análise o estabelecimento de previsões, então deve definir-se um período, pós-amostal não utilizado na estimação, para se fazer um estudo comparativo, com o cálculo dos erros de previsão, dos diversos modelos que passaram nos filtros da avaliação do diagnóstico, (Silva, Salgado, & Menezes, 2009).

2.4. Rio Ave e modelação estatística

Relativamente ao rio Ave como caso de estudo, existem já alguns trabalhos que recorreram a modelação estatística para modelar a qualidade da sua água.

Em Silva (2013), o trabalho consistiu na análise e modelação de séries temporais das principais variáveis de qualidade da água, recolhidas na estação de monitorização VALADA (03D/02), situada perto da foz do rio Âncora, no período de Novembro de 2003 a Junho de 2006. Os principais objetivos deste trabalho foram analisar as séries temporais das variáveis em estudo e tirar conclusões sobre o seu comportamento ao longo do tempo, ajustando também um modelo ARIMA, adequado a cada uma dessas variáveis que fosse capaz de efetuar previsões com uma boa precisão.

Os resultados obtidos permitiram concluir que, de uma forma geral, o rio Âncora teve, durante aquele período, uma boa qualidade da água. Relativamente aos modelos ARIMA que foram ajustados às variáveis para efetuar previsões, estes revelaram possuir boa precisão, uma vez que em média os valores observados estavam dentro do intervalo de previsão a 80%. A precisão desses modelos foi verificada comparando os valores previstos com aqueles que foram realmente observados, no mesmo período de tempo. Os resultados obtidos demonstraram assim que os modelos ARIMA podem ser uma boa opção para a previsão de variáveis de qualidade da água.

Ainda no contexto da modelação estatística, em Gonçalves (2005), foram apresentadas e desenvolvidas metodologias no sentido de permitir a modelação estatística da concentração de poluentes nas águas de superfície da bacia hidrográfica do Rio Ave. O conjunto de dados utilizados diz respeito a valores mensais de onze variáveis de qualidade de água, observados entre 1988 e 2003, em 20 estações de amostragem de qualidade, da rede de monitorização dos recursos hídricos superficiais existente nesta bacia hidrográfica.

Foram utilizados métodos de estatística descritiva para se efetuar uma análise preliminar espaço-temporal das variáveis de qualidade da água. Tal procedimento permitiu identificar os trechos do Rio Ave e as variáveis de qualidade com os piores e os melhores comportamentos, para se detetar padrões de comportamento ao longo do tempo. Recorreu-se ainda a procedimentos da estatística multivariada, nomeadamente à análise de *clusters*, para classificar as estações de amostragem de qualidade em grupos homogéneos (no espaço e no tempo), com base nas variáveis de qualidade. Com recurso a processos determinísticos e estocásticos, aplicados aos dados disponíveis de precipitação, foram também estimados valores mensais de precipitação. Por último, ajustaram-se modelos lineares às séries das

variáveis de qualidade observadas nas estações, com o objetivo de se escolher os que melhor descreviam/explicavam a variação da qualidade da água ao longo do tempo.

Em suma, com ambos os trabalhos que foram desenvolvidos no âmbito da modelação estatística, para este caso de estudo, foi possível concluir que através de um bom ajuste dos modelos estatísticos se podem conseguir previsões bastante fiáveis, no que toca às variáveis que caracterizam a qualidade do meio hídrico.

2.5. Estudos recentes

Além dos estudos atrás mencionados e uma vez que a exploração de modelos estatísticos tem sido uma constante no que respeita à hidrologia, neste ponto serão analisados alguns estudos mais recentes que englobam estes modelos.

São diversas as técnicas e métodos estatísticos utilizadas para estudar a qualidade da água. Em, Zhang, Chen, Recknagel, & Li (2014), recorreu-se a modelos de previsão para acompanhar o desenvolvimento de cianobactérias. Os métodos estatísticos utilizados, como a análise do coeficiente de correlação de *Pearson*, permitiram identificar correlações e desfasamentos de dados com relação linear. A metodologia *Wavelets* usada possibilitou não só examinar as correlações e desfasamentos entre o crescimento de cianobactérias mas também as condições de qualidade da água em diferentes escalas de tempo. Este estudo debruçou-se sobre um lago, situado na China, e foi efetuado entre 2010 e 2011. Teve como objetivo compreender melhor os desfasamentos entre as condições de qualidade da água e o crescimento de cianobactérias nesse lago. Os modelos utilizados revelaram-se bastante fiáveis.

Num outro estudo, foram adotadas cinco estratégias de modelação para analisar as sucessões cronológicas do nível de água de seis lagos com diferentes características, tais como a forma, o tamanho, a altitude e a gama de variações. Os modelos incluíram a teoria do caos, modelos ARIMA, SARIMA, redes neuronais artificiais, entre outros. Cada caso foi formulado com uma premissa diferente, ou seja, com diferentes pressupostos subjacentes. A teoria do caos foi elaborada com maior detalhe, sendo que esta é bastante usual para identificar a existência de sinais caóticos por uma série de técnicas. As previsões foram efetuadas usando uma técnica não-linear. Este estudo concluiu que as cinco estratégias de modelação para alguns casos se revelaram boas, no entanto noutros casos possuíam menor

qualidade. Daí muitas vezes ser necessária uma reflexão crítica sobre qual o melhor método estatístico a utilizar para estudar determinadas séries de dados, (Khatibi, Ghorbani, Naghipour, Jothiprakash, Fathima, & Fazelifard, 2014).

Atualmente as técnicas de computação têm sido muito utilizadas na modelação hidrológica e também para efetuar previsões. Em Lohani & Rakesh Kumar (2012), o potencial de redes neuronais artificiais e sistemas neuro-fuzzy foram usados, juntamente com modelos auto regressivos, para estudar a periodicidade mensal do caudal que entrava num reservatório. Trabalhando as séries temporais dos dados recolhidos, foram ajustados várias redes neuronais artificiais e modelos neuro-fuzzy, com diferentes vetores de entrada. Para avaliar o desempenho destes modelos foi ainda feita uma comparação com modelos auto-regressivos. Neste estudo concluiu-se que as redes neuronais artificiais revelavam melhor precisão nas previsões, em comparação com os modelos auto regressivos.

Voltando ao contexto da modelação ARIMA, é de notar que os dados das séries temporais muitas vezes contêm padrões lineares e não lineares. Desta forma, nem os modelos ARIMA nem as redes neuronais podem ser totalmente adequados para a modelação e previsão de dados de séries temporais. O modelo ARIMA não pode lidar com relações não lineares, enquanto o modelo de rede neuronal por si só não é capaz de lidar com ambos os padrões lineares e não lineares, igualmente bem. Em Faruk (2009), utilizou-se um modelo de rede neuronal e um modelo ARIMA, efetuando a comparação entre os dois. A rede neuronal para a previsão de séries temporais foi testada usando observações de dados de qualidade da água de 108 meses, incluindo a temperatura da água, boro e oxigénio dissolvido, desde 1996 até 2004, de um rio situado na Turquia. Os resultados revelaram que as redes neuronais devido à sua capacidade em reconhecer padrões de séries temporais e características não lineares, proporcionam muito mais precisão do que o modelo ARIMA, no que toca a previsões de qualidade da água.

Em Valipour, Banihabib, & Behbahani (2012), o objetivo consistiu em prever a entrada de caudal na barragem do rio Dez, usando modelos ARMA, ARIMA e redes neuronais artificiais. A fim de aumentar a precisão da previsão, foram-se aumentando o número de parâmetros dos modelos. Posteriormente fez-se uma comparação com as redes neuronais artificiais estáticas e dinâmicas. Para este trabalho, foram utilizadas as descargas mensais entre 1960 e 2007. Os primeiros 42 anos foram usados para calibrar os modelos, enquanto os cinco últimos anos foram usados para efetuar previsões. Nos modelos ARMA e

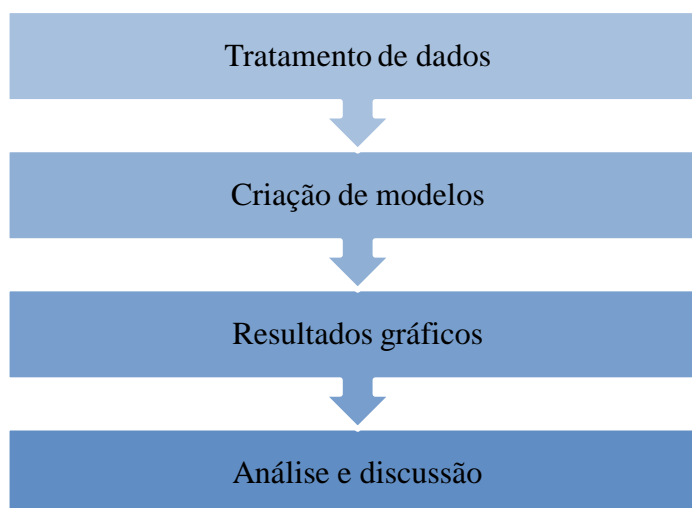
ARIMA, o polinómio foi derivado respetivamente com quatro e seis parâmetros para prever o caudal de entrada. Na rede neuronal artificial, recorreu-se a várias funções e diferentes neurónios nas camadas escondidas. Ao comparar os erros dos modelos, o modelo de rede neuronal artificial dinâmica foi escolhido como o melhor modelo para efetuar a previsão de caudal a entrar nesta barragem. Já entre o modelo ARMA e ARIMA, o modelo ARIMA foi o que revelou um menor erro.

Em jeito de conclusão, salienta-se mais uma vez a clara importância e utilidade dos trabalhos desenvolvidos no âmbito da modelação estatística. Os vários estudos revelam que os modelos estatísticos são bastante credíveis, no que diz respeito à realização de previsões bem como à caracterização da qualidade do meio hídrico. No entanto, tal como atrás mencionado, não deve nunca dispensar-se a possibilidade de utilizar outras técnicas, dependendo do tipo de dados que se pretende tratar.

3. METODOLOGIA

Tendo por base o caso de estudo, abordado no quarto capítulo da presente dissertação, e os dados recolhidos nos anos de 2005 e 2006 na estação de monitorização automática das Taipas (sonda INAG 05G/06), disponibilizados no *site* do Sistema Nacional de Informação de Recursos Hídricos (SNIRH), foi efetuada a modelação das sucessões cronológicas das variáveis em estudo: condutividade elétrica, oxigénio dissolvido, temperatura e pH. Para o efeito, consideraram-se os valores horários, registados ao longo de 24 horas, durante o período compreendido entre 07-07-2005 e 15-06-2006. Importa referir que foram usados os dados relativos a este período devido à suspensão da manutenção das estações de monitorização em 2010, pelo que dados mais recentes poderiam ter a sua qualidade comprometida. Adicionalmente foram usados dados recolhidos por uma sonda no ano de 2012, já usados em outros trabalhos.

Assim, a metodologia utilizada englobou quatro etapas principais:



No que ao tratamento de dados diz respeito, com recurso ao *Microsoft Excel*, inicialmente dividiu-se o período em análise por semanas, de modo a obter intervalos mais curtos que permitirão uma análise mais detalhada da variação dos diversos parâmetros. Uma vez separados os dados, passou-se à construção de gráficos com a variação de cada parâmetro ao longo das várias semanas.

Dado que um dos objetivos desta dissertação passa por modelar as sucessões cronológicas das variáveis em estudo, recorreu-se ao *software IBM SPSS Statistics 22*, que permitiu a criação de diferentes casos de estudo.

O primeiro passo na utilização deste *software*, tal como ilustrado na Figura 1, passa por definir o tipo de variáveis a tratar, na opção "Visualização da variável". Desta forma, é necessário definir as variáveis, sejam elas do tipo Data, como é o caso dos dias do mês e as horas, pois trata-se de valores horários, ou as variáveis do tipo numérico, como é o caso dos valores da condutividade elétrica, oxigénio dissolvido, temperatura e pH.

	Nome	Tipo	Largura	Decimais	Rótulo	Valores	Ausente	Colunas	Alinhar	Medir	Função
1	Data	Data	11	0		Nenhum	Nenhum	8	Centro	Nominal	Entrada
2	Hora	Data	8	0		Nenhum	Nenhum	5	Centro	Nominal	Entrada
3	Condutividade	Numérico	8	2		Nenhum	Nenhum	9	Centro	Nominal	Entrada
4	OD	Numérico	8	2		Nenhum	Nenhum	8	Centro	Nominal	Entrada
5	Temperatura	Numérico	8	2		Nenhum	Nenhum	8	Centro	Nominal	Entrada
6	pH	Numérico	8	2		Nenhum	Nenhum	8	Centro	Nominal	Entrada
7	DAY_	Numérico	8	0	DAY, not periodic	Nenhum	Nenhum	10	Centro	Nominal	Entrada
8	HOUR_	Numérico	8	0	HOUR, period 24	Nenhum	Nenhum	10	Centro	Ordinal	Entrada
9	DATE_	Sequência ...	7	0	Date. Format: ...	Nenhum	Nenhum	9	Centro	Nominal	Entrada
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											
25											

Figura 1 - Definição das variáveis.

Definido o tipo das variáveis, colocam-se os seus valores nas respetivas colunas, sendo que agora se recorre à opção "Visualização de dados", tal como exemplificado abaixo na Figura 2.

	Data	Hora	Condutividade	OD	Temperatura	pH	DAY_	HOUR_	DATE_	var
1	07-Jun-2005	10:00:00	58,00	8,78	18,30	6,73	1	0	1 0	
2	07-Jun-2005	11:00:00	58,00	8,82	18,60	6,72	1	1	1 1	
3	07-Jun-2005	12:00:00	58,00	8,88	18,80	6,72	1	2	1 2	
4	07-Jun-2005	13:00:00	58,00	8,86	19,10	6,74	1	3	1 3	
5	07-Jun-2005	14:00:00	58,00	8,87	19,50	6,74	1	4	1 4	
6	07-Jun-2005	15:00:00	59,00	9,01	19,90	6,80	1	5	1 5	
7	07-Jun-2005	16:00:00	59,00	9,02	20,50	6,87	1	6	1 6	
8	07-Jun-2005	17:00:00	62,00	8,93	20,60	6,89	1	7	1 7	
9	07-Jun-2005	18:00:00	64,00	8,93	20,30	6,92	1	8	1 8	
10	07-Jun-2005	19:00:00	62,00	8,89	20,00	6,95	1	9	1 9	
11	07-Jun-2005	20:00:00	59,00	8,85	19,70	6,95	1	10	1 10	
12	07-Jun-2005	21:00:00	57,00	8,78	19,60	6,93	1	11	1 11	
13	07-Jun-2005	22:00:00	56,00	8,69	19,40	6,90	1	12	1 12	
14	07-Jun-2005	23:00:00	55,00	8,66	19,30	6,86	1	13	1 13	
15	08-Jun-2005	0:00:00	54,00	8,58	19,30	6,83	1	14	1 14	
16	08-Jun-2005	1:00:00	53,00	8,50	19,20	6,81	1	15	1 15	
17	08-Jun-2005	2:00:00	52,00	8,42	19,20	6,79	1	16	1 16	
18	08-Jun-2005	3:00:00	51,00	8,39	19,20	6,78	1	17	1 17	
19	08-Jun-2005	4:00:00	50,00	8,28	19,10	6,76	1	18	1 18	
20	08-Jun-2005	5:00:00	50,00	8,26	19,10	6,73	1	19	1 19	
21	08-Jun-2005	6:00:00	49,00	8,20	19,00	6,69	1	20	1 20	
22	08-Jun-2005	7:00:00	49,00	8,28	19,00	6,69	1	21	1 21	
23	08-Jun-2005	8:00:00	49,00	8,35	19,00	6,70	1	22	1 22	
24	08-Jun-2005	9:00:00	49,00	8,41	19,10	6,68	1	23	1 23	

Figura 2 - Preenchimento das variáveis com os dados históricos.

Desta forma, pode então proceder-se à obtenção dos modelos das sucessões cronológicas. Tal como se pode observar abaixo na Figura 3, em primeiro lugar, na barra de ferramentas acede-se à opção "Analisar", escolhendo a hipótese "Previsão", seguida da opção "Criar Modelos".

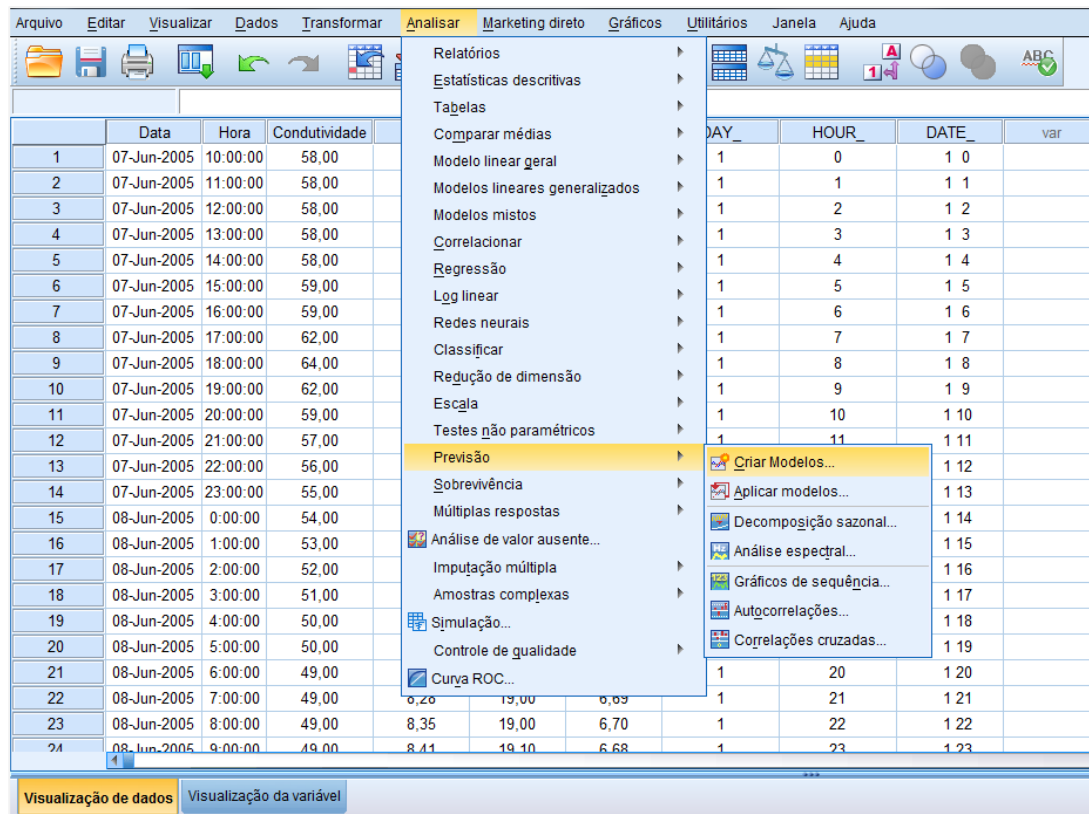


Figura 3 - Criação de Modelos de Previsão.

Para gerar um modelo é necessário, em primeiro lugar, escolher qual a variável dependente a tratar. Neste passo podem gerar-se modelos que possuam mais do que uma variável, sendo que estas serão as variáveis independentes, tal como se pode observar na Figura 4. Nos modelos gerados na presente dissertação, numa primeira fase, cada um dos parâmetros foi analisado individualmente, como uma variável dependente.

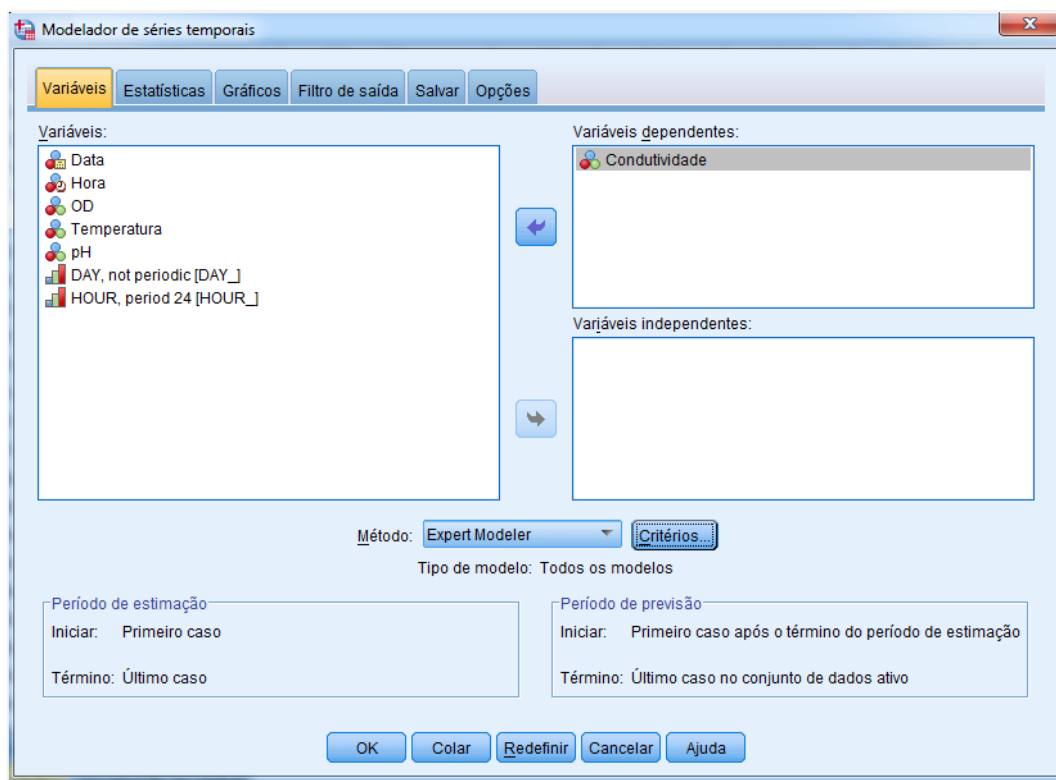


Figura 4 - Seleção da variável a estudar.

Para gerar um modelo existem várias opções, sendo que para este estudo foram gerados modelos automaticamente, através da opção "*Expert Modeler*" e também modelos ARIMA, gerados manualmente. Importa salientar que na opção "*Expert Modeler*" foram escolhidos dois critérios diferentes. Numa primeira fase, ilustrada na Figura 5, consideraram-se todos os modelos possíveis, isto é, o programa através dos valores que lhe foram apresentados selecionou o modelo que melhor se ajustava. Numa segunda fase, escolheu-se como critério do "*Expert Modeler*" a opção que força a criação apenas de modelos ARIMA. Com vista a um melhor ajuste de modelos e, uma vez que normalmente os modelos ARIMA são modelos mais simples, além das duas opções do "*Expert Modeler*", foi também efetuada a construção manual de modelos ARIMA. No caso de gerar modelos ARIMA manualmente foi necessário definir as componentes não sazonais e sazonais, características deste tipo de modelos, tal como exemplificado na Figura 6. A procura de modelos alternativos, vizinhos dos modelos já encontrados, deve-se ao facto de muitas vezes existirem outros modelos que descrevem a mesma sucessão cronológica com igual grau de qualidade.

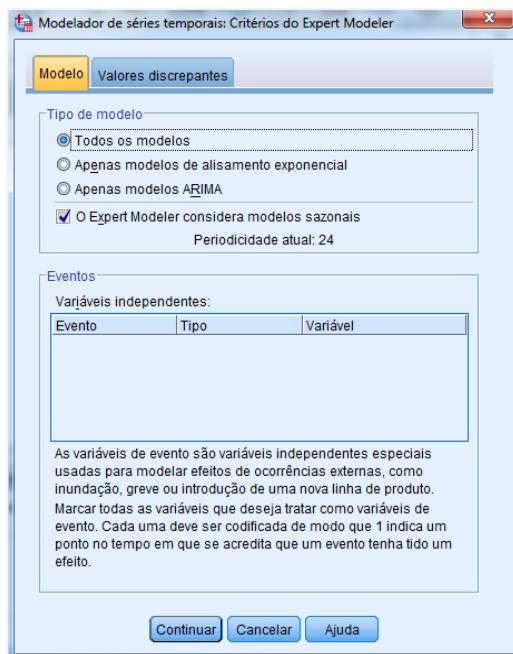


Figura 5 - Definição de critérios do "Expert Modeler".

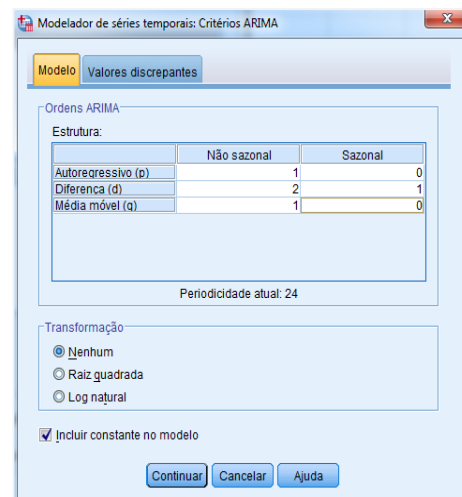


Figura 6 - Definição de critérios ARIMA.

A criação dos modelos para as várias semanas em estudo tratou-se então de um processo iterativo, com vista à obtenção do modelo com melhor ajuste às sucessões cronológicas dos diversos parâmetros. Na fase da seleção do tipo de modelo é fundamental ativar a opção que salva os valores previstos, isto é, os valores do modelo, para posteriormente se construírem gráficos que permitem uma visão mais clara dos modelos obtidos.

Deste modo, foram-se obtendo os vários modelos para cada uma das variáveis e de forma a escolher o melhor modelo, realizaram-se os seguintes procedimentos:

- ❖ Observação dos gráficos que traduzem o comportamento das Funções de Autocorrelação (FAC) e Funções de Autocorrelação Parcial (FACP) dos resíduos. Estas funções medem, para cada valor, a intensidade com que pares de valores do processo separados por um intervalo, variam entre si, ou seja, covariam. A autocorrelação é então usada para descrever a correlação entre dois valores da mesma série temporal, em diferentes períodos de tempo. Já a autocorrelação parcial é utilizada para descrever a correlação entre dois valores da mesma série, porém retirando a influência de valores intermediários.

- ❖ Recurso a critérios como o *Akaike Information Criteria* (AIC), que é definido de forma a penalizar os modelos com maior número de parâmetros e com maior variância residual, ou o *Bayesian Information Criteria* (BIC);
- ❖ Avaliação de modelos alternativos, feita de acordo com o coeficiente de correlação de cada modelo, assim como, pelos valores do AIC. De acordo com os critérios definidos, o melhor modelo é aquele que apresenta o maior coeficiente de correlação e, também, o menor valor de AIC. Relativamente ao coeficiente de correlação de *Pearson*, este mede o grau da correlação existente entre duas variáveis de escala métrica. Varia entre -1 e 1, sendo que 1 representa uma correlação perfeita, 0 significa que as variáveis não dependem linearmente uma da outra e -1 significa uma correlação negativa perfeita entre as variáveis (se uma aumenta, a outra diminui);
- ❖ Verificação do coeficiente de determinação (R ao quadrado): medida de ajuste do modelo. Quanto mais próximo for da unidade, melhor a qualidade do modelo.

É ainda de salientar que uma vez identificado o modelo, as estimativas dos respetivos parâmetros são feitas automaticamente recorrendo a um dos métodos: máxima verosimilhança, mínimos quadrados e momentos. Desta forma, consegue verificar-se a adequação do modelo, controlando a qualidade estatística do modelo e também a qualidade do ajustamento.

Este *software* permitiu, assim, uma análise detalhada dos vários parâmetros, através da formulação de diferentes casos de estudo.

3.1. Caso de estudo 1: Variação semanal dos parâmetros *versus* modelo ajustado pelo SPSS

O primeiro caso de estudo criado contemplou a variação de cada um dos parâmetros ao longo das várias semanas. Selecionando o melhor modelo para cada um dos parâmetros, construíram-se gráficos no *Microsoft Excel*, que permitiram uma análise mais detalhada dos resultados obtidos.

3.2. Caso de estudo 2: Variação anual dos parâmetros *versus* modelo ajustado pelo SPSS

Com o intuito de compreender o tipo de modelos gerado e respetivo coeficiente de correlação, efetuou-se a modelação dos quatro parâmetros para a totalidade do período de estudo (07-07-2005 a 15-06-2006), sem fracionamento dos dados, tal como anteriormente.

3.3. Caso de estudo 3: Caudal do rio como variável independente

Muitas vezes as séries possuem um comportamento sazonal, isto é, que se repete no tempo. Na tentativa de entender este comportamento, no terceiro caso de estudo incluiu-se como variável independente nos modelos o caudal (Q), uma vez que se trata de um fator que condiciona diretamente o comportamento da série.

3.4. Caso de estudo 4: Modelação ARIMA na realização de previsões

Tendo por objetivo o estabelecimento de previsões, no quarto caso de estudo recorreu-se à modelação ARIMA para realizar previsões a 24h. Para tal, definiu-se um período, pós-amostral não utilizado na estimação, para se fazer um estudo comparativo, com o cálculo dos erros de previsão, dos diversos modelos que passaram nos filtros da avaliação do diagnóstico.

3.5. Caso de estudo 5: Modelos gerados para o ano de 2005 *versus* modelos gerados para o ano de 2012

Uma vez que inicialmente a sucessão cronológica dos registos foi dividida por semanas com o intuito de facilitar o estudo e de proporcionar um melhor ajuste dos modelos, e dado que também se encontram disponíveis dados mais atuais, nomeadamente para o ano de 2012, foi feita uma comparação entre os modelos obtidos no *software SPSS* para o ano de 2005 e os modelos obtidos para o ano de 2012. O objetivo consiste em verificar se para o mesmo período do ano, os modelos são idênticos. À partida fará sentido comparar valores em períodos homólogos, uma vez que por exemplo, no caso da temperatura, não terá qualquer utilidade comparar valores de temperatura de um período de Verão com valores de temperatura de um período de Inverno, pois naturalmente a água terá temperaturas bastante distintas, consoante a estação do ano. Em relação aos restantes parâmetros, o mesmo se deverá verificar.

A utilidade da verificação desta hipótese reside assim no facto de, numa situação prática, conhecendo a altura do ano em análise, à partida saber-se-á qual o modelo que melhor se irá ajustar à variação de determinada variável, pelo que em termos de previsões isso será bastante benéfico. Claro está que não se dispensará a análise dos registos imediatamente anteriores a esse momento. Deste modo, numa fase posterior à previsão dos valores que os parâmetros irão adquirir, será então possível, caso necessário, tomar medidas de prevenção ou de mitigação relativas a determinados acontecimentos indesejáveis.

3.6. Caso de estudo 6: Variação dos parâmetros *versus* variação do caudal

Uma vez que a variação dos parâmetros em análise se encontra intimamente ligada com a variação do caudal do rio, foram construídos gráficos que contemplam a variação do caudal a montante da estação de monitorização automática e também a variação de cada um dos parâmetros. Os dados relativos à variação do caudal foram também obtidos através do *site* do SNIRH e são valores registados na estação de monitorização existente na localidade de Garfe.

Com recurso ao *Microsoft Excel*, criou-se então este último cenário, que contemplou a variação do caudal. Neste caso não foi necessário recorrer ao *SPSS*, pois trata-se de séries de dados às quais não se pretendia ajustar nenhum modelo.

Na zona do estuário, o interesse e a riqueza biológicos ainda são reduzidos devido às más condições físico-químicas e biológicas da água que o rio vinha a apresentar, em consequência da elevada quantidade de efluentes urbanos e industriais, canalizados para este rio. No entanto, esta situação tem vindo a ser corrigida ao longo dos últimos anos.

A bacia hidrográfica do rio Ave é constituída principalmente por granitos, sendo que estes constituem sensivelmente 83% da bacia. Os xistos estão presentes na ordem dos 14% e com menos frequência ocorrem areias, cascalheiras e aluviões, na casa dos 2%, (Sistema Nacional de Informação de Recursos Hídricos, 2014).

4.1. Indústrias locais e poluição associada

Através da análise dos vários parâmetros de qualidade da água será possível realizar a modelação da sua variação e posteriormente efetuar previsões, a curto prazo, através de modelos estatísticos. É assim de crucial importância fazer uma boa avaliação qualitativa dos recursos hídricos, tentando perceber quais são as atividades sócio-económicas que exercem pressão sobre as massas de água em determinado território, uma vez que são muitas vezes estas pressões que influenciam o seu estado, causando graves impactos na saúde pública e no ecossistema.

Ao longo dos últimos séculos, o rio Ave e os seus afluentes marcaram, de forma singular, a implantação industrial do setor têxtil na região. Tal facto deve-se às vantagens associadas às facilidades hídricas para a produção de energia e também ao abastecimento de água nas diferentes fases dos processos industriais, como é o caso das lavagens ou das operações de tinturaria. Deste modo, ao longo dos tempos, a distribuição espacial das unidades industriais foram acompanhando muito de perto o traçado das linhas de água, intensificando-se na vizinhança dos aglomerados urbanos.

Relativamente à bacia hidrográfica do rio Ave, a localização da indústria têxtil possui raízes históricas, que remontam a um artesanato ligado ao linho e à implantação do têxtil do algodão a partir de meados do século XIX. Posteriormente, logo no início do século XX, foram construídas fábricas nas margens deste rio e dos seus afluentes. O vale do Ave foi, desta forma, transformando-se num território industrial de grande densidade empresarial.

A poluição de origem industrial nesta bacia revela-se, assim, um problema bastante antigo, com forte expressão espacial e especificidade, quanto às suas fontes.

A concentração de fábricas têxteis foi, indubitavelmente, a principal responsável pelos impactos nocivos em todo o ambiente fluvial, assumindo-se como um problema à escala regional. No entanto, com o passar do tempo e com o desenvolvimento das técnicas industriais dá-se uma maior independência relativamente aos cursos de água. A água deixa então de ser utilizada como um bem essencial na produção industrial, passando a ser utilizada como veículo de condução dos desperdícios resultantes das águas utilizadas nos diferentes processos produtivos da indústria. O lançamento de escórias de carvão e resíduos industriais nas margens deste curso de água foi uma prática comum em algumas fábricas, sendo que o lançamento destas substâncias se tornou fortemente prejudicial para as espécies piscícolas, resultando em situações muito nocivas para a saúde pública.

O elevado número de autos de transgressões nos anos cinquenta relativos ao Vale do Ave esteve intimamente relacionado com a fiscalização exercida nesse período às estruturas de saneamento e também com os impactos resultantes da poluição aquática e degradação do leito e das margens, causadas pelas atividades industriais. A situação era tão grave que em 1987 a autarquia de Vila do Conde admitiu instaurar processos-crime e encerrar as fábricas poluentes no rio Ave, se essas empresas não se mostrassem interessadas em colaborar com a autarquia na defesa daquele que era considerado um dos mais poluídos rios da Europa. Em 1988, um estudo elaborado no âmbito do projeto de Gestão Integrada de Recursos Hídricos do Norte revelou que, em mais de 100 quilómetros dos cursos dos principais rios do litoral norte de Portugal, a qualidade da água era má ou péssima, salientando que parte significativa desses troços se situava na bacia do Ave.

Em suma, pode concluir-se que apesar de terem sido impostas à partida obras e técnicas de tratamento dos efluentes líquidos produzidos, estas revelaram-se insuficientes face ao volume de desperdícios produzido, pelo que a qualidade do rio Ave e seus afluentes começou a degradar-se bastante com o passar dos anos. Em 1997, a situação atingiu níveis inadmissíveis e deste modo tiveram que ser postas em prática novas medidas e mais rígidas por parte do governo, (Costa, 2008).

Atualmente, a melhoria na qualidade da água deste rio pode constatar-se a olho nu, no entanto, existem dados disponíveis, como é o caso do *site* do SNIRH, que comprovam esta melhoria significativa.

4.2. Evolução da qualidade da água do rio Ave

O INAG, em colaboração com as Direções Regionais, implementou uma rede automática de qualidade da água, que permite conhecer em contínuo e em tempo real a evolução de cinco parâmetros de qualidade da água: temperatura, oxigénio dissolvido, pH, condutividade e turvação. A instalação destas estações automáticas teve início em 2001, progredindo de Sul para Norte.

A classificação da qualidade da água ficou desta forma facilitada, permitindo tirar várias conclusões, entre as quais os usos potenciais da massa de água classificada. Assim, para essa classificação são consideradas cinco classes:

- ❖ Excelente (A);
- ❖ Boa (B);
- ❖ Razoável (C);
- ❖ Má (D);
- ❖ Muito Má (E).

Importa referir que para a classificação atrás mencionada foram selecionados os quatro parâmetros estudados ao longo da presente dissertação: temperatura, condutividade elétrica, oxigénio dissolvido e pH.

Assim, além dos dados disponibilizados que serviram de base para o desenvolvimento desta dissertação, o *site* do SNIRH permite, através de uma análise rápida, constatar a evolução claramente positiva da qualidade da água que se verificou na bacia hidrográfica do Ave, ao longo dos últimos anos.

De forma a clarificar esta evolução, apresentam-se abaixo as Figuras 8 e 9, que foram recolhidas diretamente do referido *site*.

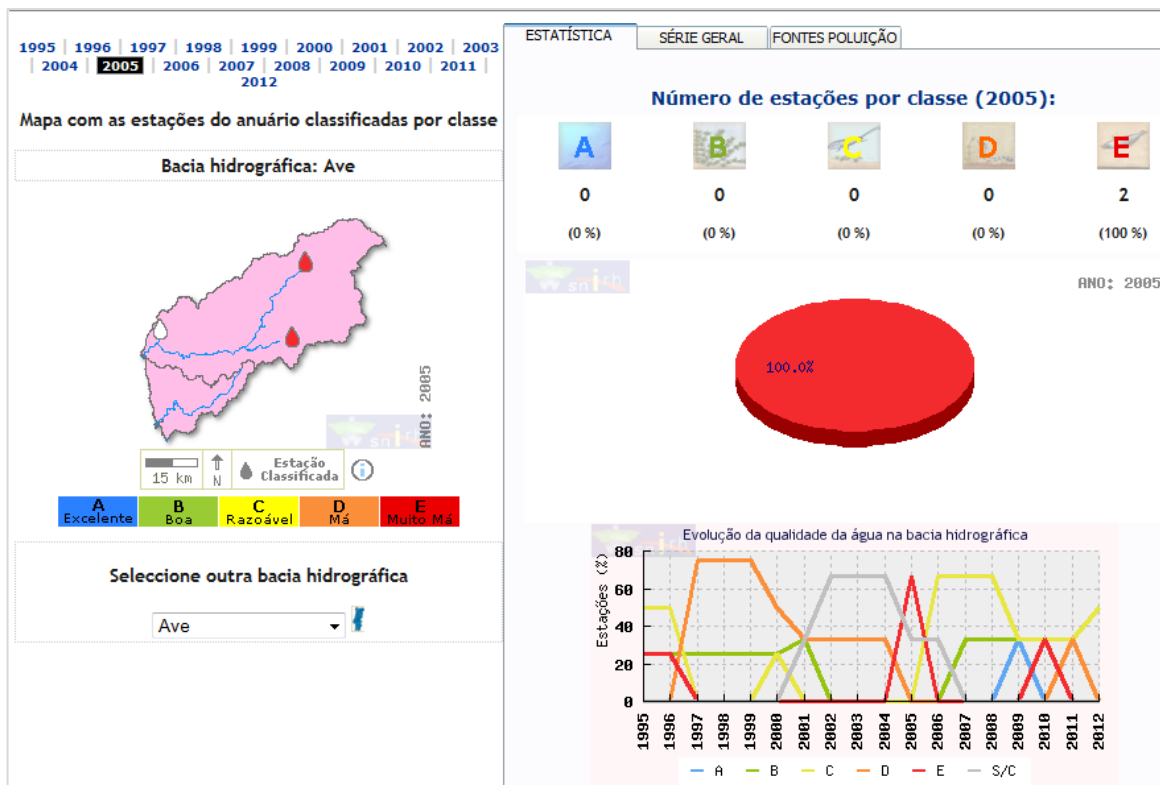


Figura 8 - Classificação da qualidade da água na bacia hidrográfica do rio Ave no ano de 2005 (Fonte: SNIRH, 2014).

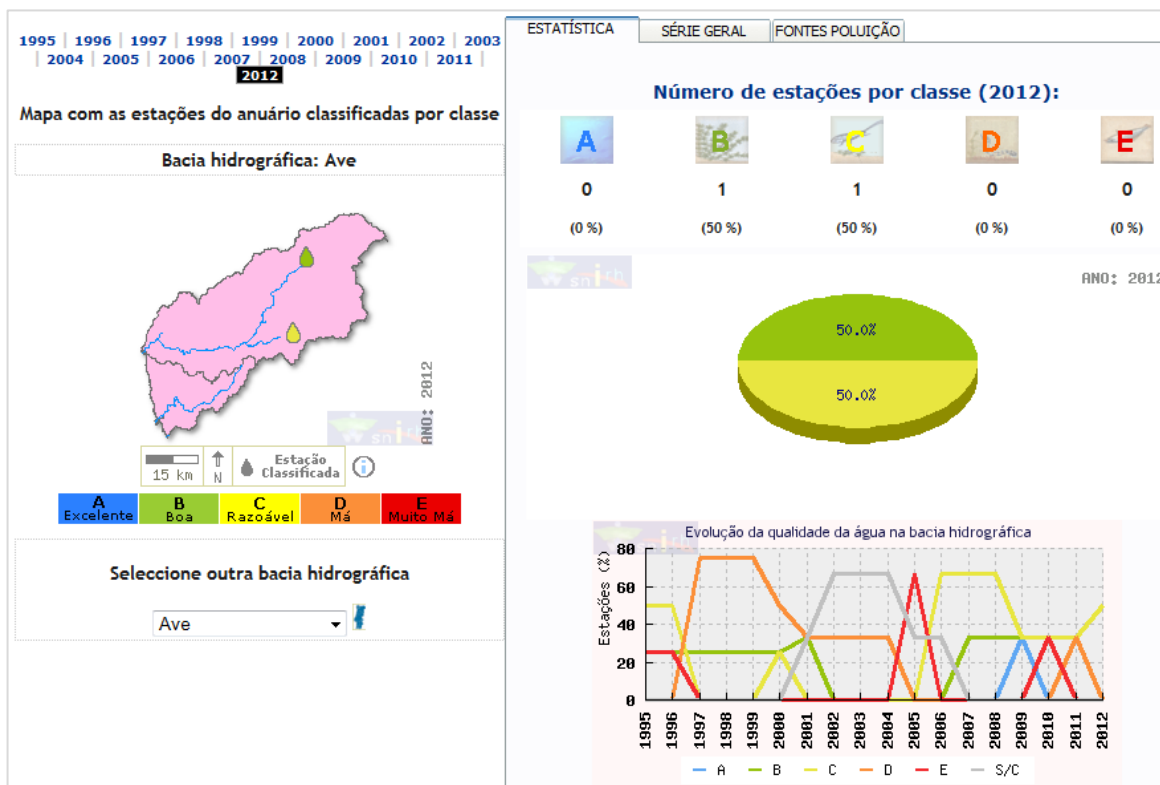


Figura 9 - Classificação da qualidade da água na bacia hidrográfica do rio Ave no ano de 2012 (Fonte: SNIRH, 2014).

Tal como se pode observar nas figuras acima apresentadas, as águas da bacia hidrográfica do rio Ave passaram de uma qualidade classificada como muita má (classe E), no ano de 2005, para uma qualidade classificada entre boa e razoável (classes B e C), no ano de 2012. Daqui se depreende que as medidas adotadas com a finalidade de melhorar a qualidade deste meio hídrico surtiram um efeito bastante positivo.

Em termos mais concretos, poder-se-á comparar as médias dos valores registados em diferentes períodos, para cada um dos parâmetros, com os valores estabelecidos por lei para águas doces superficiais destinadas à produção de água para consumo humano¹. Assim, a fim de se verificar a melhoria significativa que se registou no meio hídrico em estudo, serão analisados os registos efetuados nos anos de 2005/2006 e no ano de 2012.

Para o período de 07-07-2005 a 15-06-2006 a média de valores registados para cada uma das variáveis em estudo apresenta-se na Tabela 1.

Tabela 1 - Média dos valores registados de 07-07-2005 a 15-06-2006 em comparação com VMR e VMA.

Parâmetro	Média	VMR	VMA
Condutividade elétrica, 20°C (µ/cm)	58.84	1000	-
Oxigénio dissolvido (mg/L)	9.76	70% sat. O ₂	-
Temperatura (°C)	14.89	22	25
pH	6.48	6.8 a 8.5	-

Relativamente ao período de 03-01-2012 a 06-07-2012, apenas foi possível analisar a média dos registos de condutividade elétrica e de temperatura, apresentada na Tabela 2.

¹ Os valores registados foram comparados com os Valores Máximos Recomendados (VMR) e Admissíveis (VMA), definidos no Anexo I do Decreto-Lei n.º 236/98, de 1 de Agosto de 1998.

Tabela 2 - Média dos valores registados de 03-01-2012 a 06-07-2012 em comparação com VMR e VMA.

Parâmetro	Média	VMR	VMA
Condutividade elétrica, 20°C (μ/cm)	41.99	1000	-
Temperatura (°C)	13.36	22	25

Uma vez que o vale do Ave é constituído essencialmente por granito, a condutividade elétrica tenderá a adquirir valores mais baixos. No entanto, tal como anteriormente referido, na presença de metais dissolvidos, estes valores terão tendência para aumentar. Assim, e uma vez que com o passar dos anos estes valores foram diminuindo, é possível concluir que os metais presentes na água foram diminuindo. Salienta-se ainda que a média dos valores registados não ultrapassou os valores estipulados por lei em nenhum dos períodos.

O mesmo se passa com a temperatura, que também foi adquirindo valores menores com o passar do tempo. Neste caso não se pode logo concluir que a poluição por aquecimento das indústrias locais diminuiu, uma vez que os valores de temperatura não se devem apenas a esse fator pois poderá estar-se perante anos bastante distintos, no que respeita à intensidade das estações do ano. No entanto, independentemente desse fator, os valores da temperatura adquiriram níveis bastante mais aceitáveis com o passar dos anos e a sua média em ambos os períodos também não ultrapassou os limites legais.

Em relação ao oxigénio dissolvido e ao pH, uma vez que não foi possível ter acesso a dados mais recentes, resta concluir que pelo menos para o referido período entre 2005 e 2006, a média dos valores registados não ultrapassou os valores estipulados por lei. Ainda assim, dada a classificação efetuada pelo SNIRH relativa a este meio hídrico, conclui-se que apesar da média não ter ultrapassado os valores máximos recomendados ou admissíveis, os limites terão sido ultrapassados em momentos pontuais, não só por estes parâmetros mas também pelos outros, pois as águas do rio Ave estavam classificadas como muito más.

Em jeito de conclusão, pode afirmar-se que, de facto, ocorreu uma melhoria significativa da qualidade da água do meio em estudo com o passar dos anos.

5. RESULTADOS E RESPECTIVA ANÁLISE

A presente secção dedica-se à apresentação e análise dos resultados obtidos, de acordo com a metodologia adotada e anteriormente descrita.

Face ao elevado volume de resultados gráficos obtidos, optou-se por incluir nesta secção apenas aqueles que serão mais representativos do estudo realizado. Pelo mesmo motivo, não se incluiu anexos na presente dissertação.

De modo a ilustrar a primeira etapa deste estudo, que consistiu no tratamento dos dados semana a semana e construção dos respetivos gráficos, apresentam-se de seguida, a título de exemplo, as Figuras 10 e 11, que consistem nos gráficos que contém os dados históricos de oxigénio dissolvido e pH para semanas distintas.

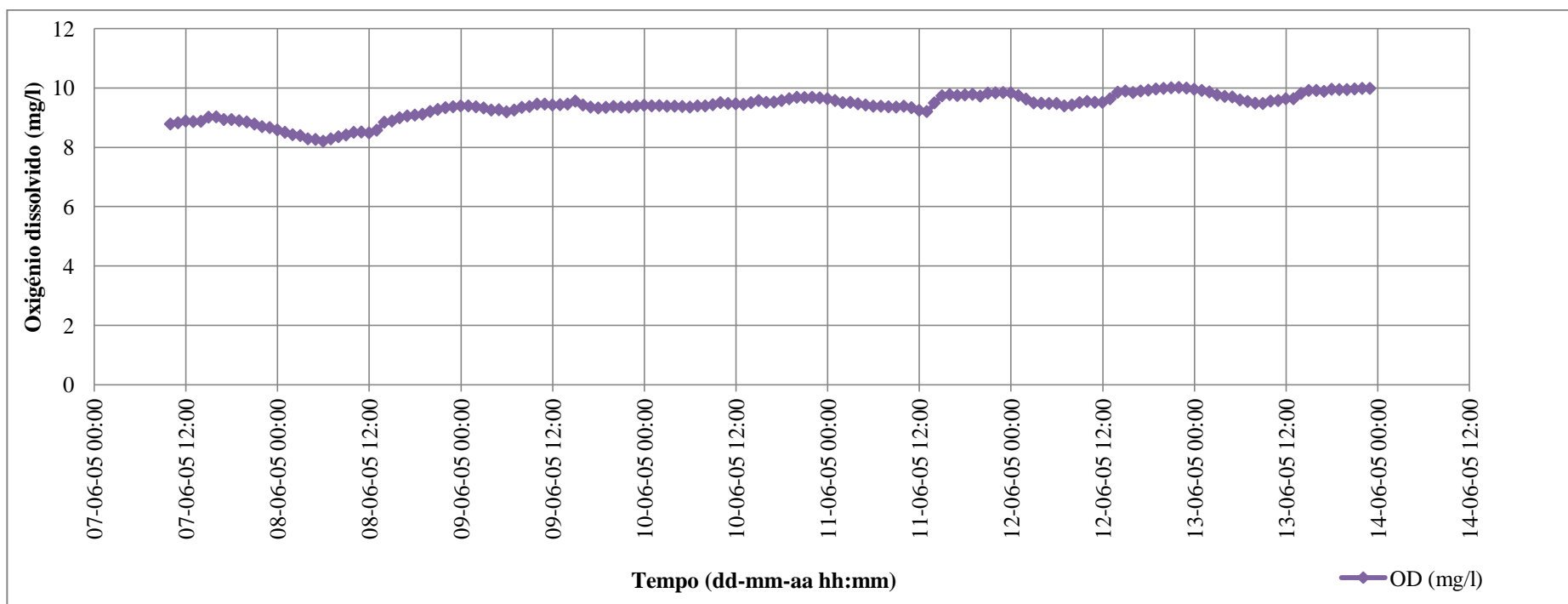


Figura 10 - Dados históricos de Oxigénio Dissolvido para o período de 07-06-2005 a 13-06-2005.

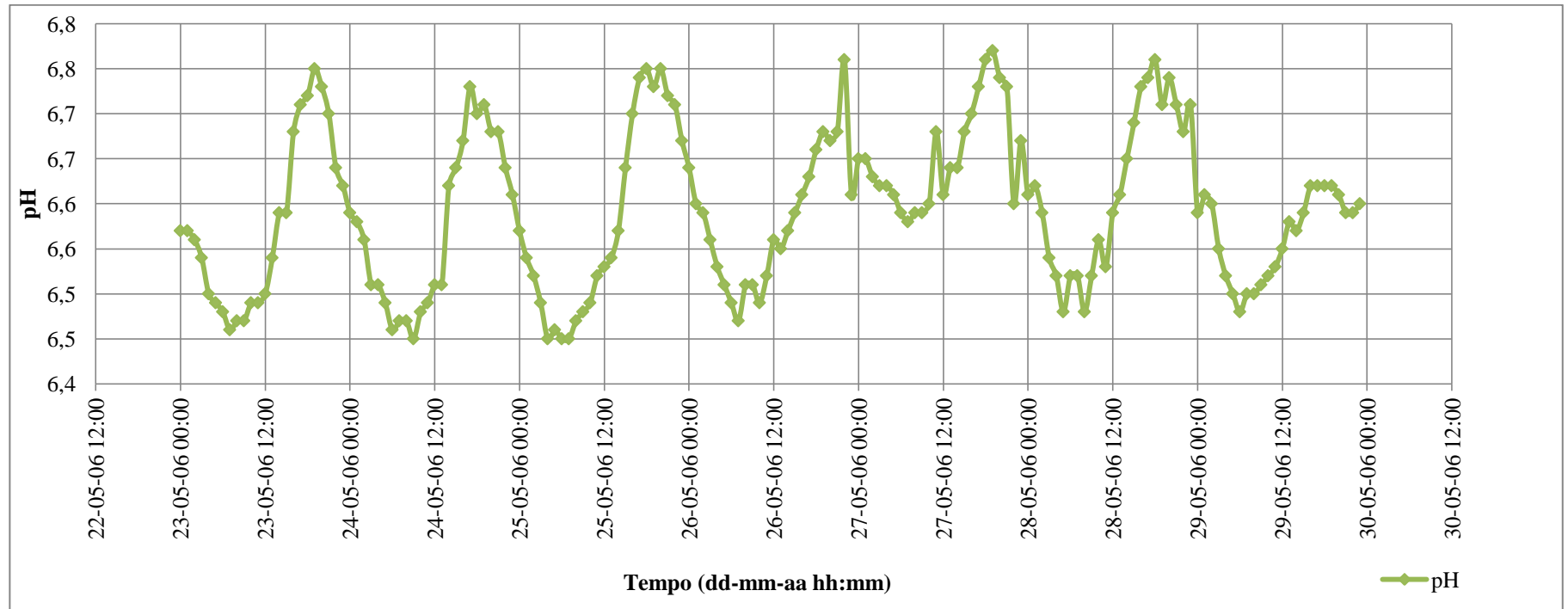
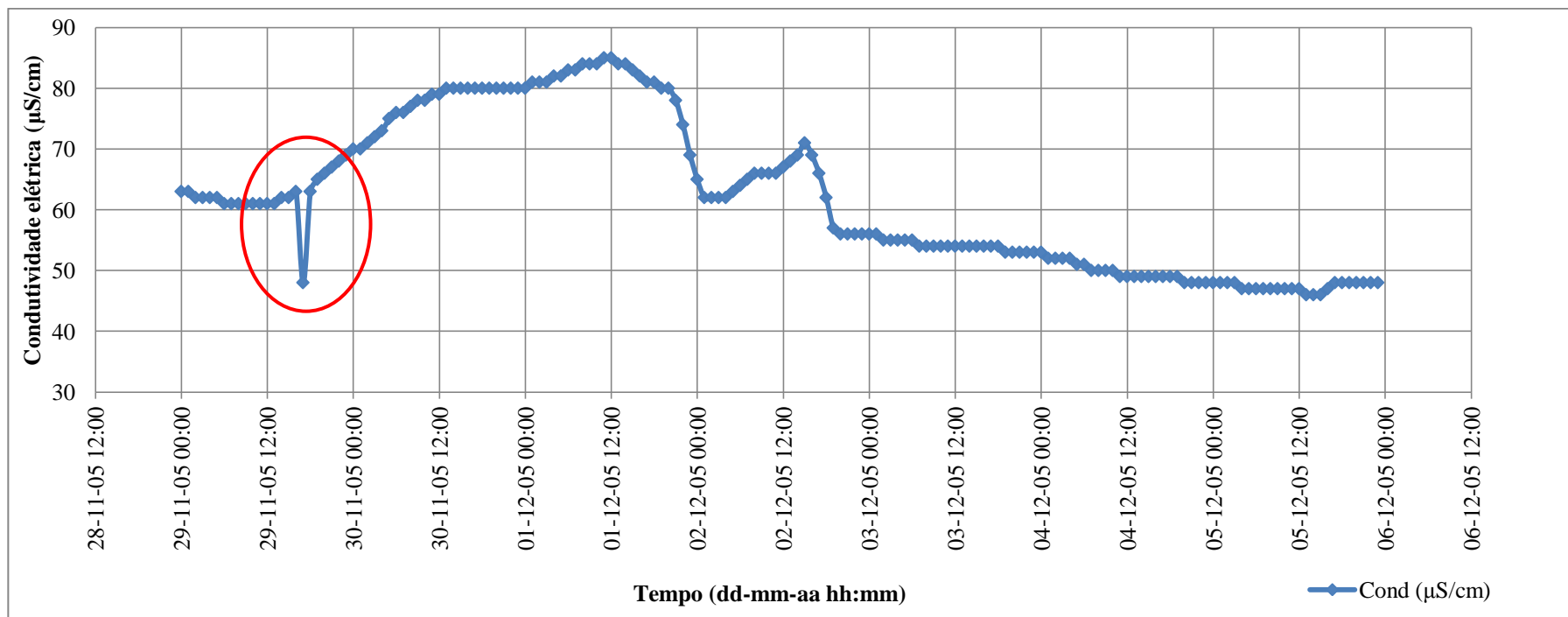


Figura 11 - Dados históricos de pH para o período de 23-05-2006 a 29-05-2006.

Deste modo, através da leitura dos gráficos obtidos semana a semana, pode facilmente detetar-se a presença de alguma perturbação ou alteração brusca de algum dos parâmetros. Trata-se de uma fase importante, na medida em que os modelos gerados posteriormente no *software SPSS* serão tanto melhores quanto menos perturbações existirem. Assim, sempre que um modelo não possua um bom ajuste, pode sempre verificar-se se é devido a alguma dessas perturbações. No entanto, a inspeção visual do comportamento das séries é apenas um indicador, sendo necessárias análises estatísticas para uma conclusão mais verdadeira sobre o comportamento da série.

A título de exemplo, abaixo apresenta-se a Figura 12, que é um gráfico com dados históricos de condutividade elétrica, no qual se pode constatar uma clara perturbação. Uma perturbação deste género pode ser causada por diversos fatores, desde fatores externos até problemas da própria sonda que regista os valores de condutividade elétrica.



5.1. Caso de estudo 1: Variação semanal dos parâmetros *versus* modelo ajustado pelo *SPSS*

Uma vez divididos os dados por semanas, passou-se para o primeiro caso de estudo explicitado na metodologia. Os gráficos apresentados nas Figuras 13, 14, 15 e 16 ilustram a comparação entre os valores registados e os modelos gerados pelo *software SPSS*. A título de exemplo, para este caso selecionaram-se os gráficos dos quatro parâmetros para o período de 30-05-2006 a 05-06-2006. São também apresentados os coeficientes de correlação entre os valores observados e os valores do modelo, que permitem aferir a qualidade do ajustamento do modelo às séries de dados.

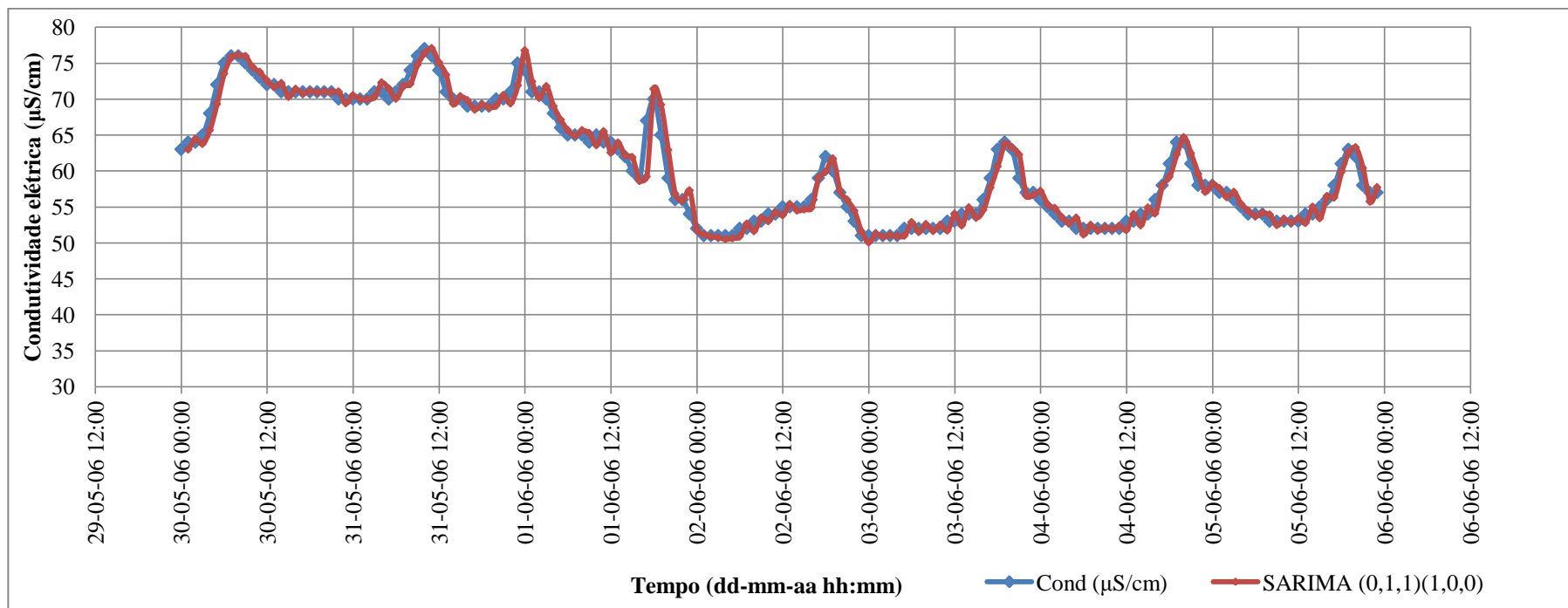


Figura 13 - Dados históricos de Condutividade Elétrica *versus* modelo SARIMA (0,1,1)(1,0,0) para o período de 30-05-2006 a 05-06-2006.

O coeficiente de correlação entre os registos de Condutividade Elétrica e o modelo SARIMA (0,1,1)(1,0,0) foi r=98.7%.

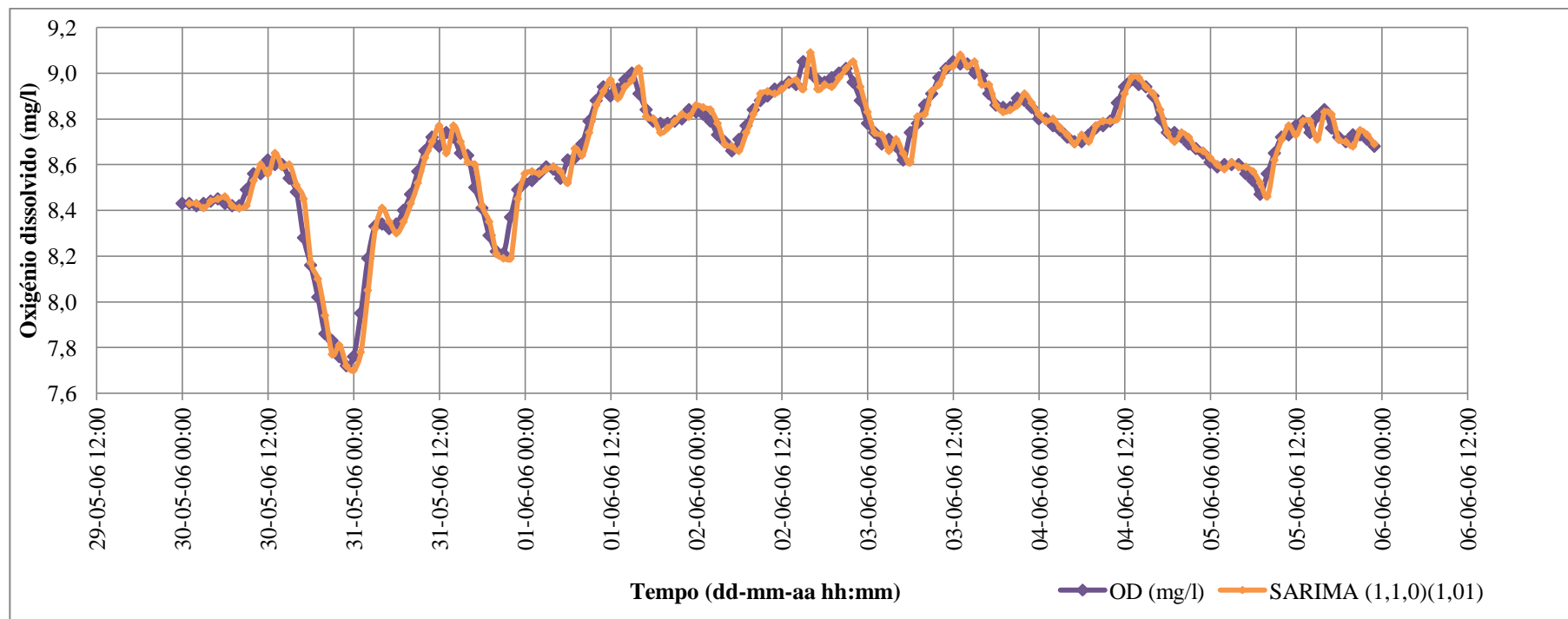


Figura 14 - Dados históricos de Oxigênio Dissolvido *versus* modelo SARIMA (1,1,0)(1,0,1) para o período de 30-05-2006 a 05-06-2006.

O coeficiente de correlação entre os dados históricos de Oxigênio Dissolvido e o modelo SARIMA (1,1,0)(1,0,1) foi $r=98.4\%$.

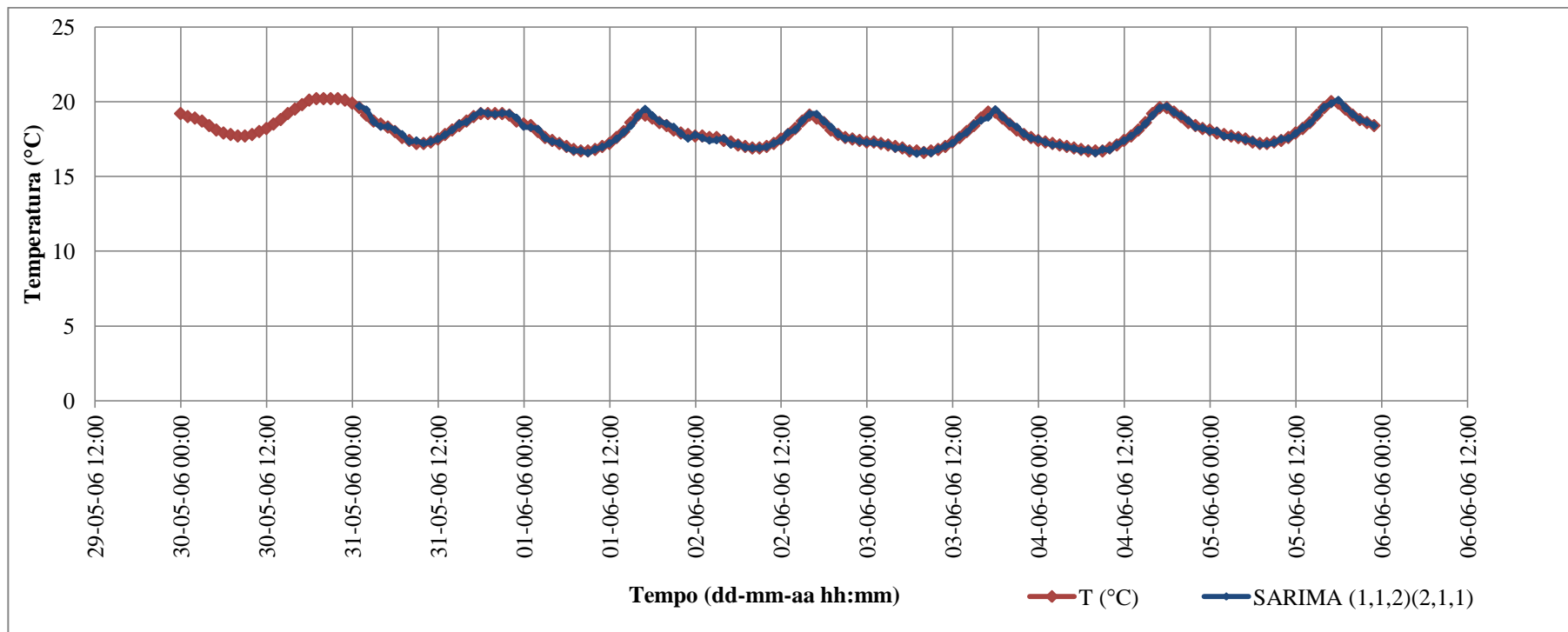


Figura 15 - Dados históricos de Temperatura *versus* modelo SARIMA (1,1,2)(2,1,1) para o período de 30-05-2006 a 05-06-2006.

O coeficiente de correlação entre os dados históricos de Temperatura e o modelo SARIMA (1,1,2)(2,1,1) foi $r=99.3\%$.

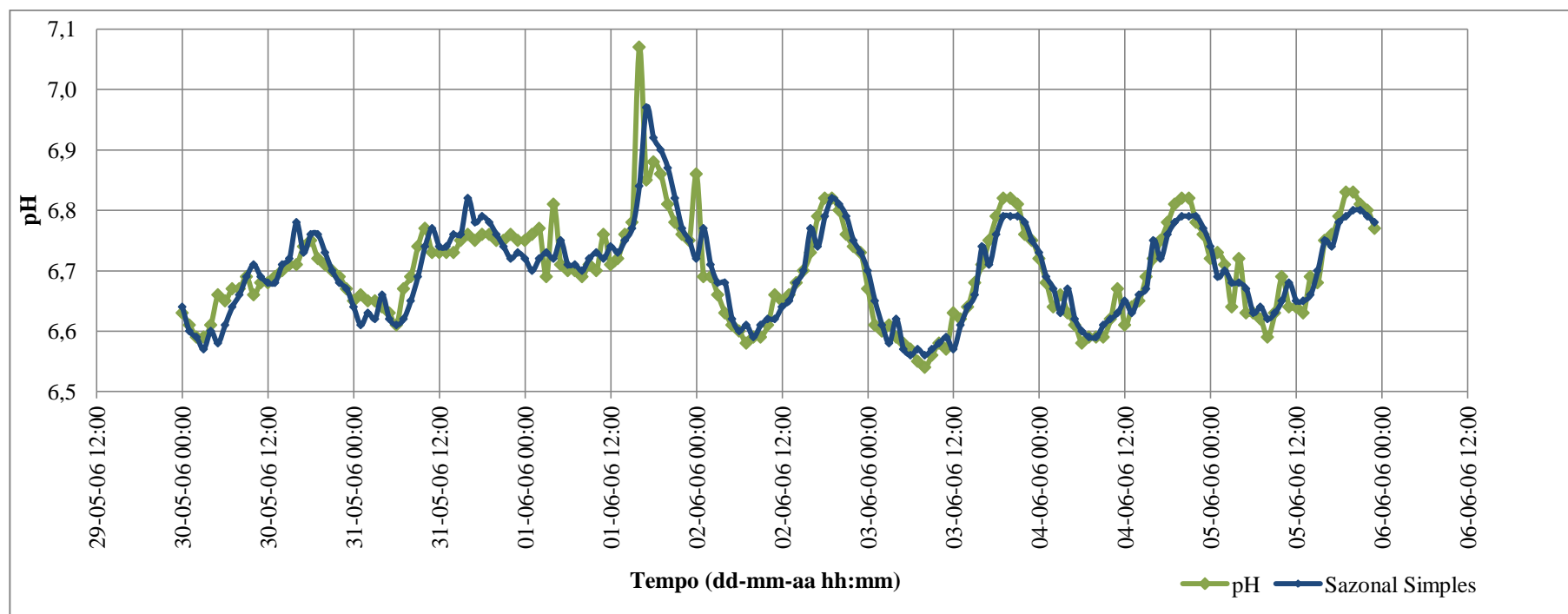


Figura 16 - Dados históricos de pH *versus* modelo Sazonal Simples para o período de 30-05-2006 a 05-06-2006.

O coeficiente de correlação para o caso dos dados históricos de pH e modelo Sazonal Simples foi $r=89.5\%$.

Através da análise dos gráficos acima apresentados, tal como nos restantes também obtidos, pode constatar-se que os modelos gerados pelo *software SPSS* possuem um ótimo ajuste à variação dos diversos parâmetros, pelo que os respetivos coeficientes de correlação entre a variação e o modelo ajustado são bastante elevados.

É ainda de salientar que para o exemplo apresentado os modelos selecionados são na sua maioria modelos do tipo ARIMA (neste caso SARIMA, pois foi tida em conta a componente sazonal), no entanto, para todo o intervalo de estudo foram gerados além dos modelos ARIMA, os modelos Sazonal Simples e Aditivo de *Winters*. Salienta-se ainda que os modelos ARIMA são, geralmente, aqueles que apresentam um maior coeficiente de correlação e um menor número de parâmetros.

5.2. Caso de estudo 2: Variação anual dos parâmetros *versus* modelo ajustado pelo SPSS

Os resultados apresentados para este caso de estudo englobam a totalidade do período de estudo. Assim, abaixo apresentam-se as Figuras 17, 18, 19 e 20 que são os gráficos obtidos para os quatro parâmetros em análise, com os respetivos modelos ajustados pelo *SPSS*.

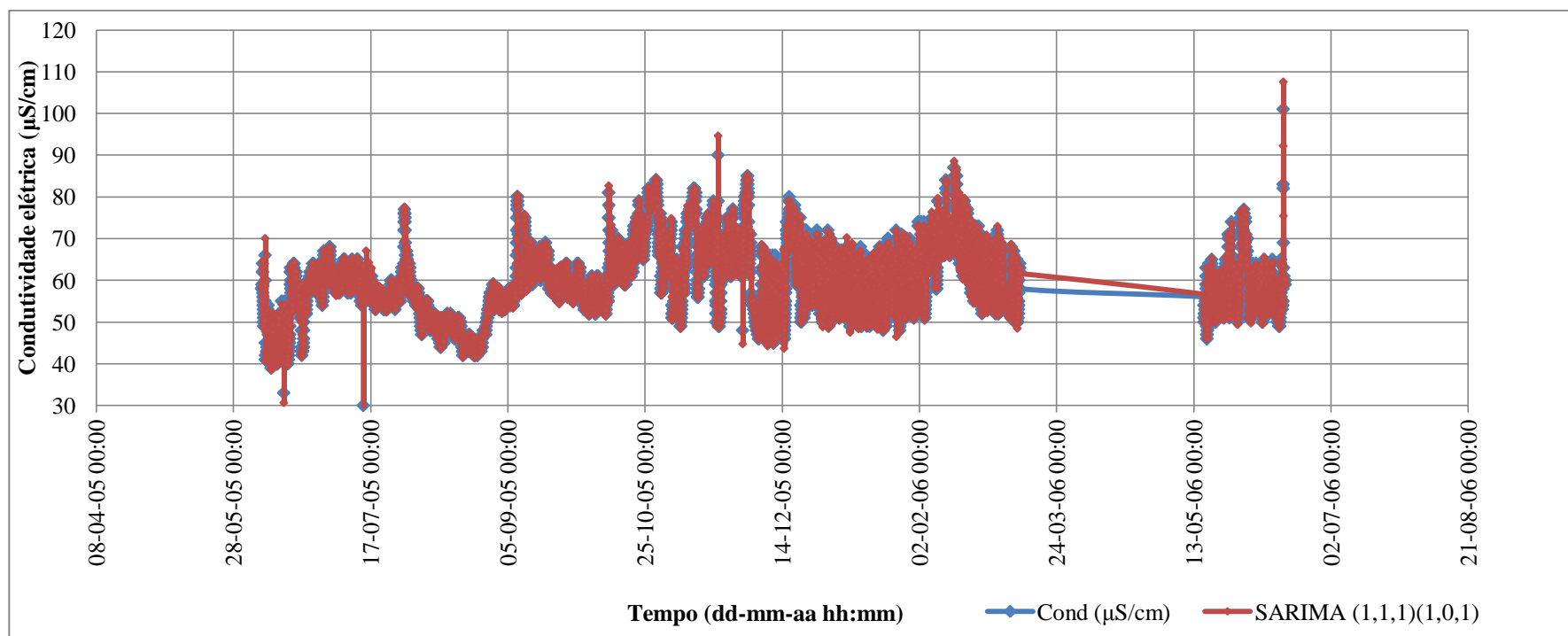


Figura 17 - Dados históricos de Condutividade Elétrica *versus* modelo SARIMA (1,1,1)(1,0,1) para o período de 07-07-2005 a 15-06-2006.

Neste caso o coeficiente de correlação entre os registos de Condutividade Elétrica e o modelo gerado foi $r=98.7\%$.

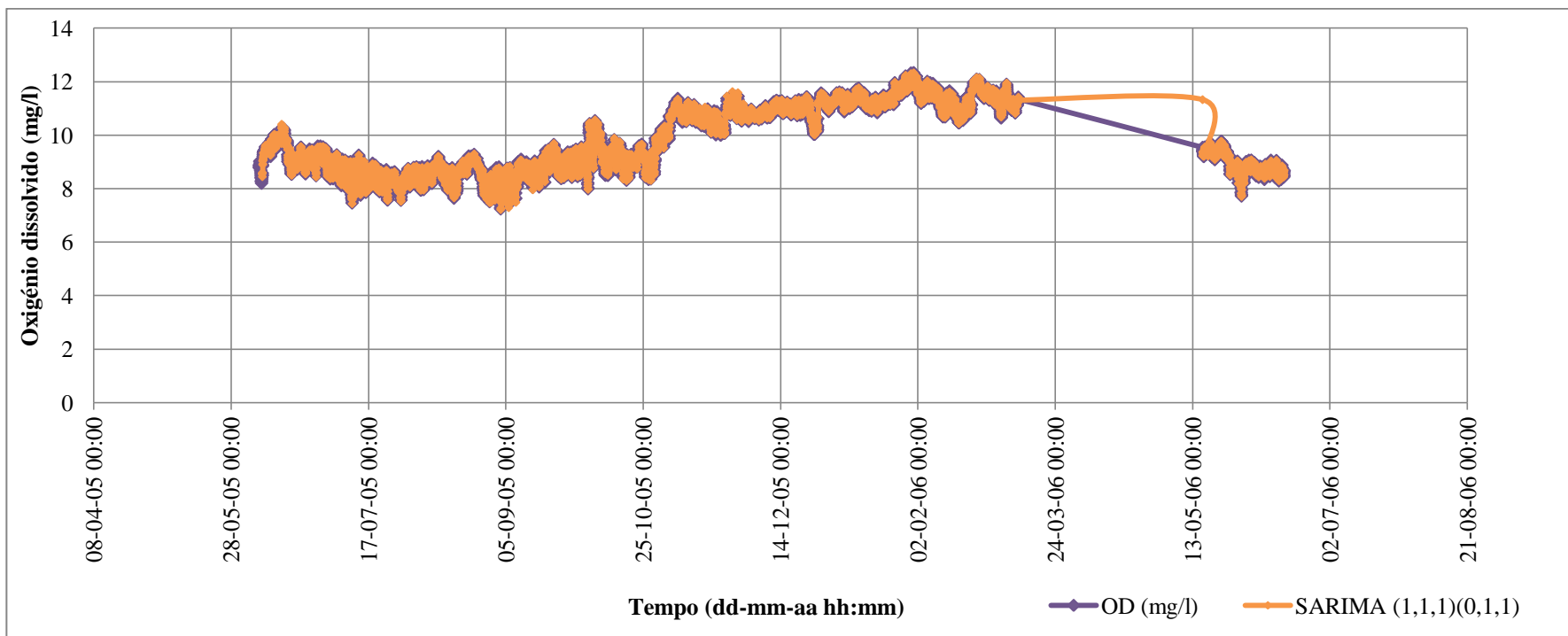


Figura 18 - Dados históricos de Oxigénio Dissolvido *versus* modelo SARIMA (1,1,1)(0,1,1) para o período de 07-07-2005 a 15-06-2006.

O coeficiente de correlação entre os dados históricos de Oxigénio Dissolvido e o modelo gerado foi $r=99.8\%$.

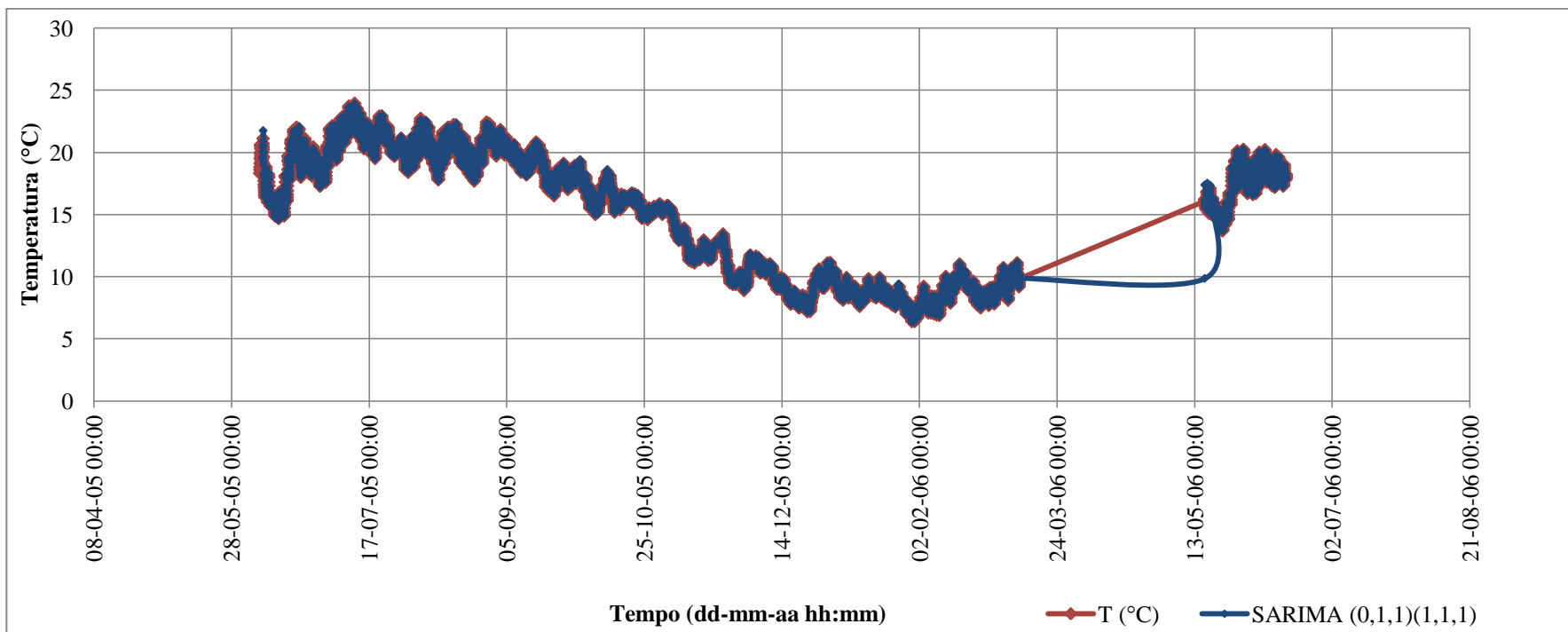


Figura 19 - Dados históricos de Temperatura *versus* modelo SARIMA (0,1,1)(1,1,1) para o período de 07-07-2005 a 15-06-2006.

Para o caso da temperatura, o coeficiente de correlação atingiu o valor perfeito, $r=100\%$.

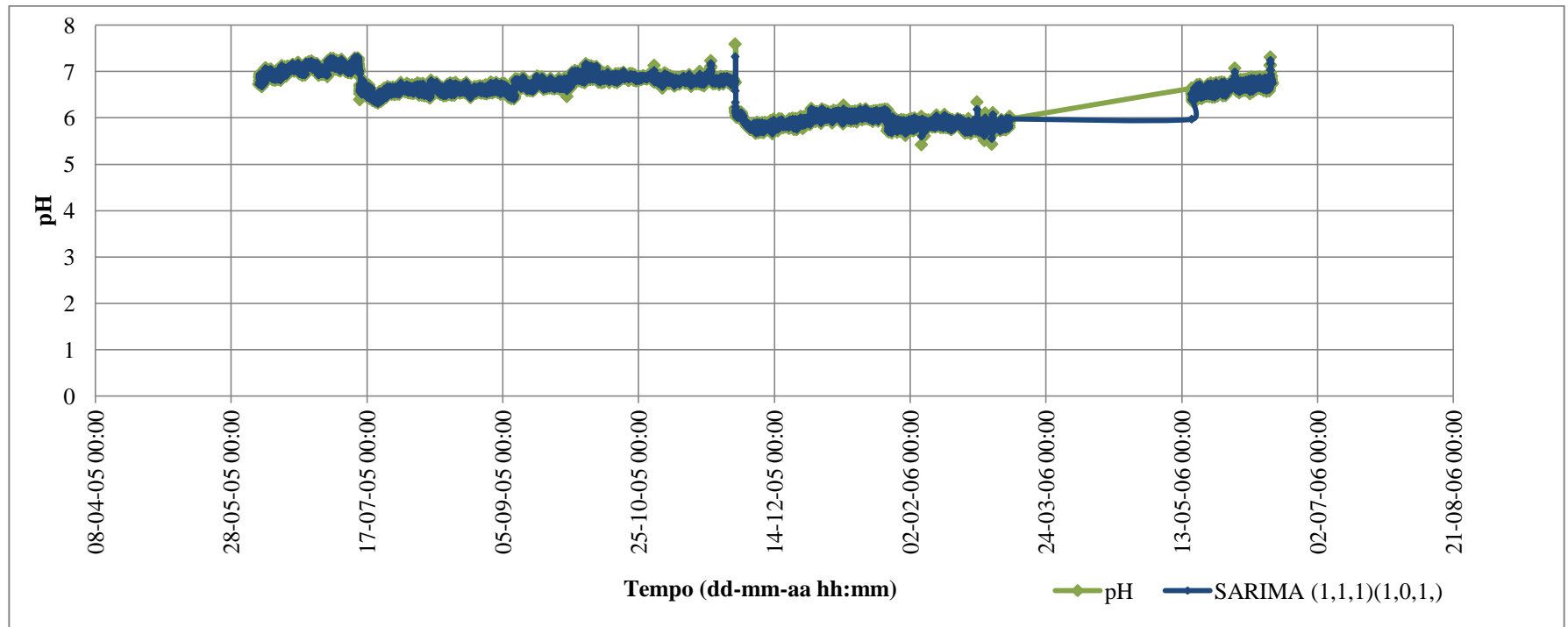


Figura 20 - Dados históricos de pH *versus* modelo SARIMA (1,1,1)(1,0,1) para o período de 07-07-2005 a 15-06-2006.

Para o caso do pH, o coeficiente de correlação entre os registos e o modelo foi $r=99.6\%$.

Através da observação dos gráficos acima apresentados, pode concluir-se que os coeficientes de correlação são um pouco mais elevados para todos os parâmetros do que no caso de estudo anterior.

A diferença mais notável entre os modelos gerados semana a semana e estes modelos, que contém a totalidade do período em estudo, traduz-se no tipo de modelos gerado. Enquanto nos modelos para períodos mais curtos foram gerados todo o tipo de modelos (ARIMA, SARIMA, Sazonal Simples ou Aditivo de *Winters*), neste caso são todos do tipo SARIMA.

Dado que os coeficientes de correlação neste caso de estudo são mais elevados, pode concluir-se que os modelos ajustados para todo o período em análise revelam muita qualidade. No entanto, uma vez que se tratam de valores horários, será mais sensato analisar sempre períodos mais curtos de tempo. Aqui deve também pensar-se no caso das previsões a curto prazo, que farão mais sentido se se analisar um intervalo de tempo mais curto para a estimação dos parâmetros.

5.3. Caso de estudo 3: Caudal do rio como variável independente

Para ilustrar este caso de estudo, abaixo apresentam-se as Figuras 21, 22, 23 e 24 que são os gráficos obtidos para o período de 30-05-2006 a 05-06-2006, que incluem as sucessões cronológicas dos registos e também o modelo gerado no *SSPS* que considerou o caudal do rio (Q) como variável independente nos modelos.

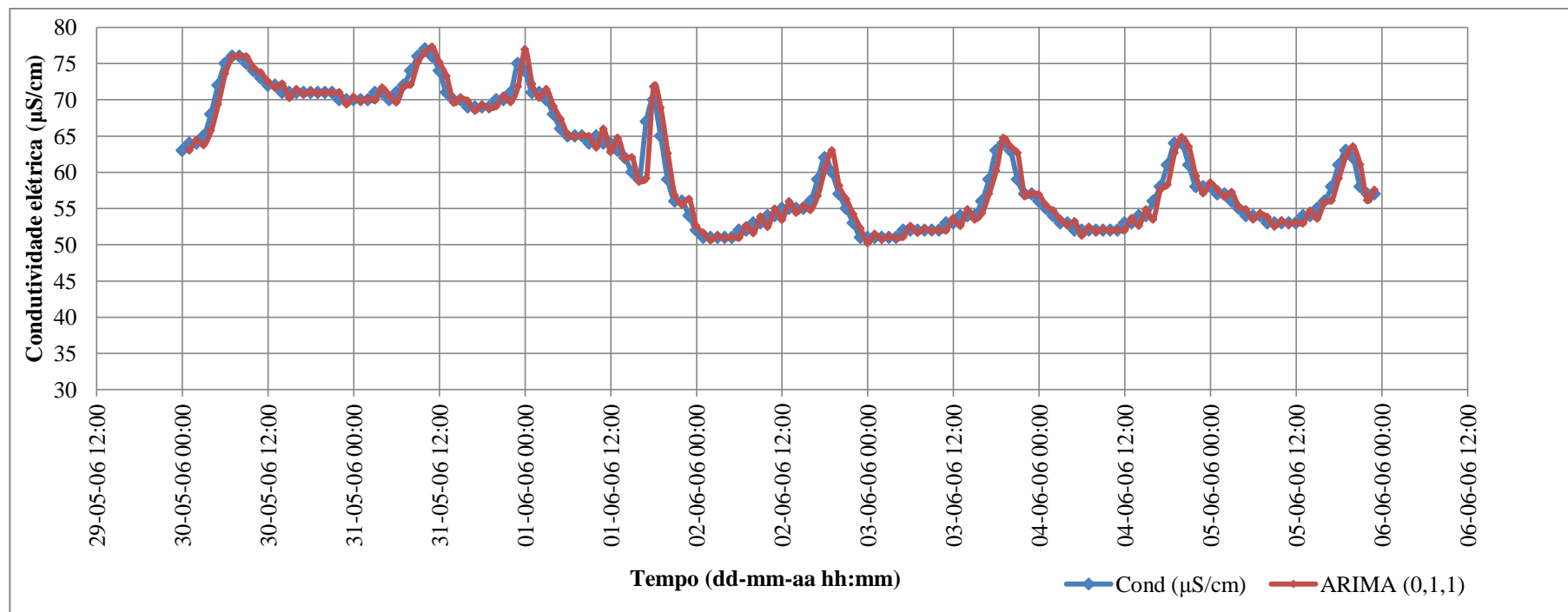


Figura 21 - Dados históricos de Condutividade Elétrica *versus* modelo ARIMA (0,1,1) para o período de 30-05-2006 a 05-06-2006, considerando o Caudal a montante como variável independente.

O coeficiente de correlação entre os dados históricos de Condutividade Elétrica e o modelo ARIMA (0,1,1) foi $r=98.5\%$.

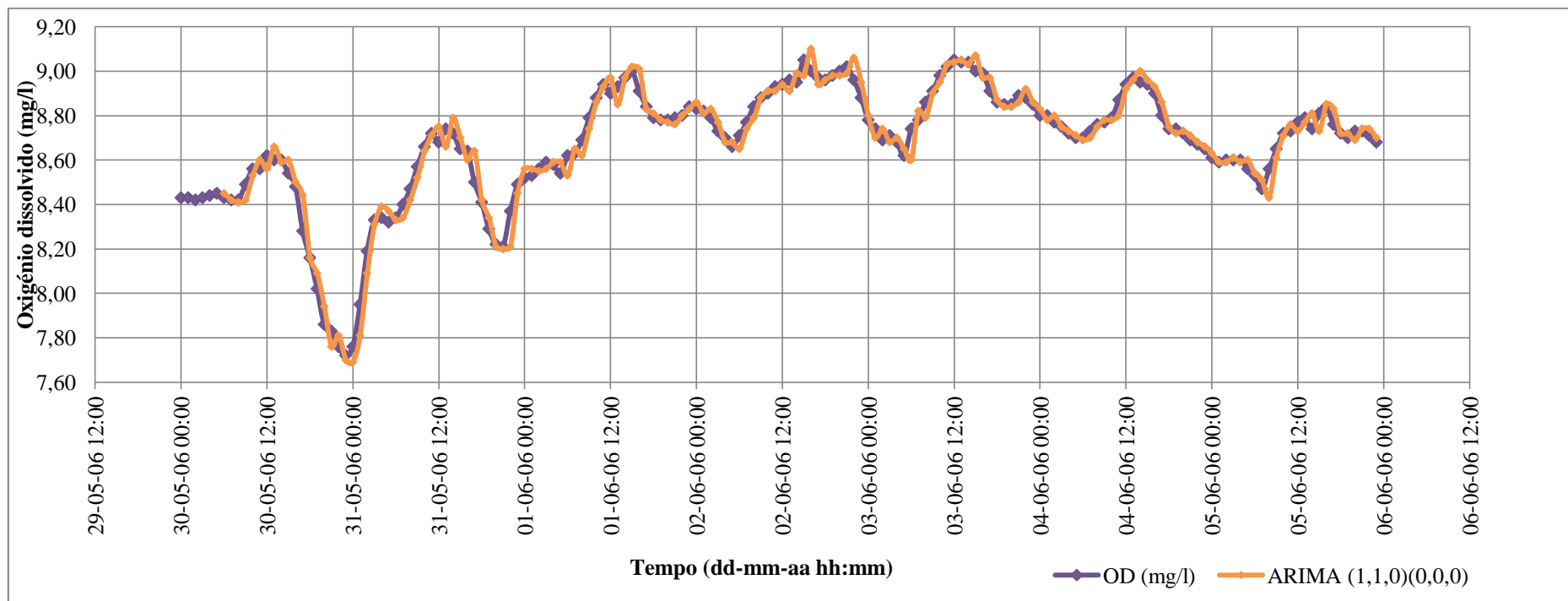


Figura 22 - Dados históricos de Oxigénio Dissolvido *versus* modelo ARIMA (1,1,0)(0,0,0) para o período de 30-05-2006 a 05-06-2006, considerando o Caudal a montante como variável independente.

O coeficiente de correlação entre os dados históricos de Oxigénio Dissolvido e o modelo ARIMA (1,1,0)(0,0,0) foi $r=98.4\%$.

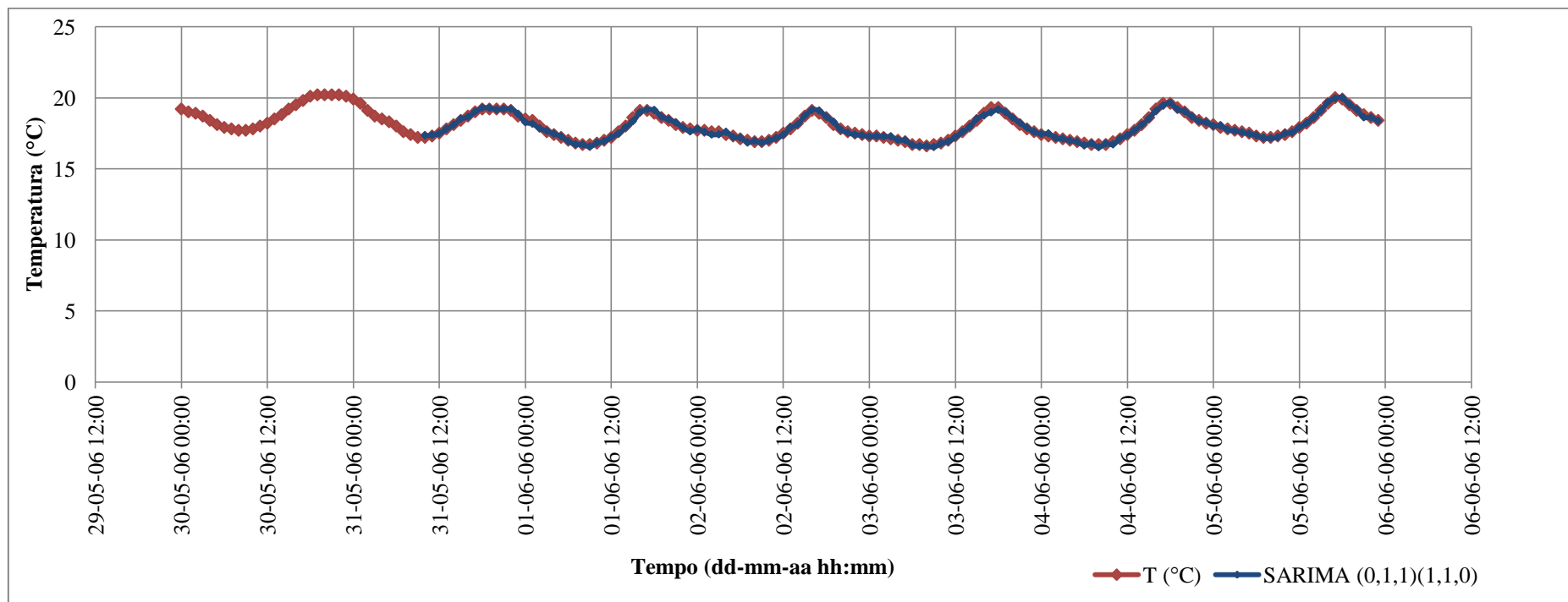


Figura 23 - Dados históricos de Temperatura *versus* modelo SARIMA (0,1,1)(1,1,0) para o período de 30-05-2006 a 05-06-2006, considerando o Caudal a montante como variável independente.

O coeficiente de correlação entre os dados históricos de Temperatura e o modelo SARIMA (0,1,1)(1,1,0) foi $r=99.3\%$.

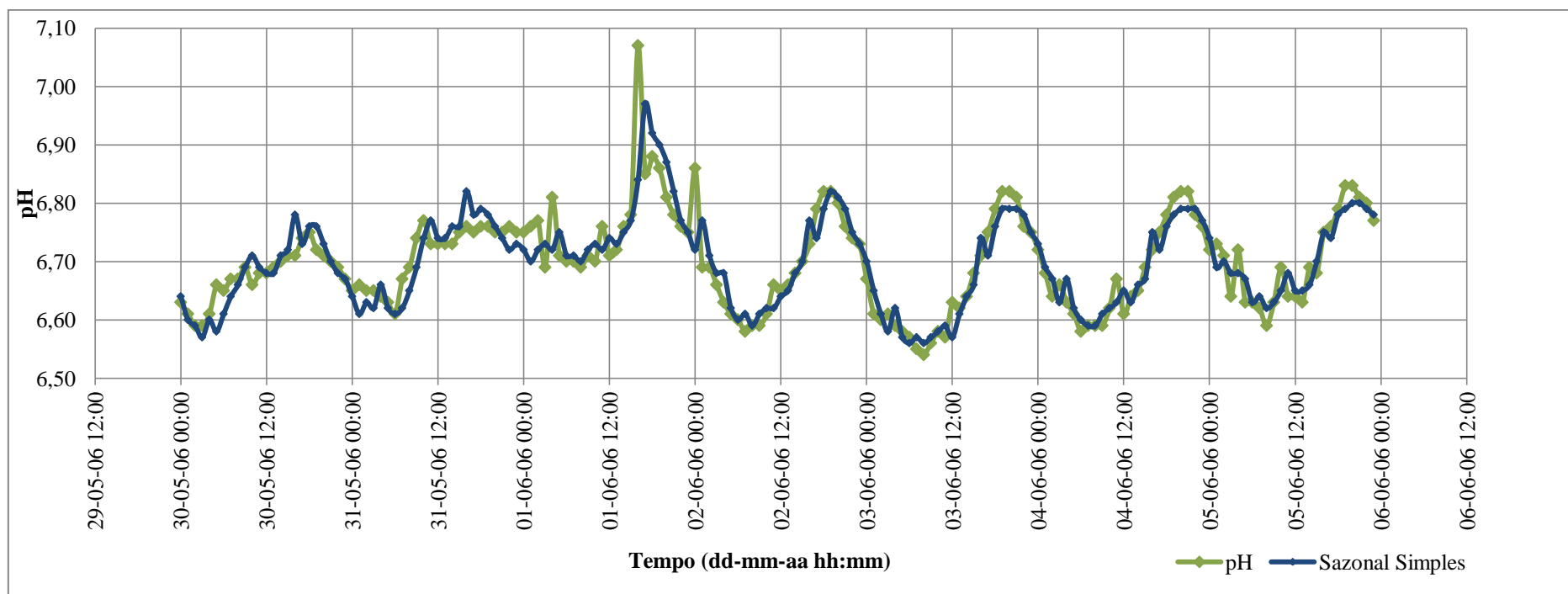


Figura 24 - Dados históricos de pH *versus* modelo Sazonal Simples para o período de 30-05-2006 a 05-06-2006, considerando o Caudal a montante como variável independente.

O coeficiente de correlação entre os dados históricos de pH e o modelo Sazonal Simples foi $r=89.5\%$.

Conforme pode observar-se, os resultados obtidos neste ponto são muito idênticos aos obtidos anteriormente, onde não se considerou o caudal a montante como variável independente. Daqui se conclui que apesar do caudal influenciar os valores adquiridos pelos parâmetros ao longo do tempo, não possui influência no tipo de modelo gerado.

5.4. Caso de estudo 4: Modelação ARIMA na realização de previsões

Para ilustrar o quarto caso de estudo, a título de exemplo, apresenta-se em seguida na Figura 25 a previsão realizada para as 24h do dia 29-05-2006, tendo por base os registos de oxigénio dissolvido entre 23-05-2006 e 28-05-2006.

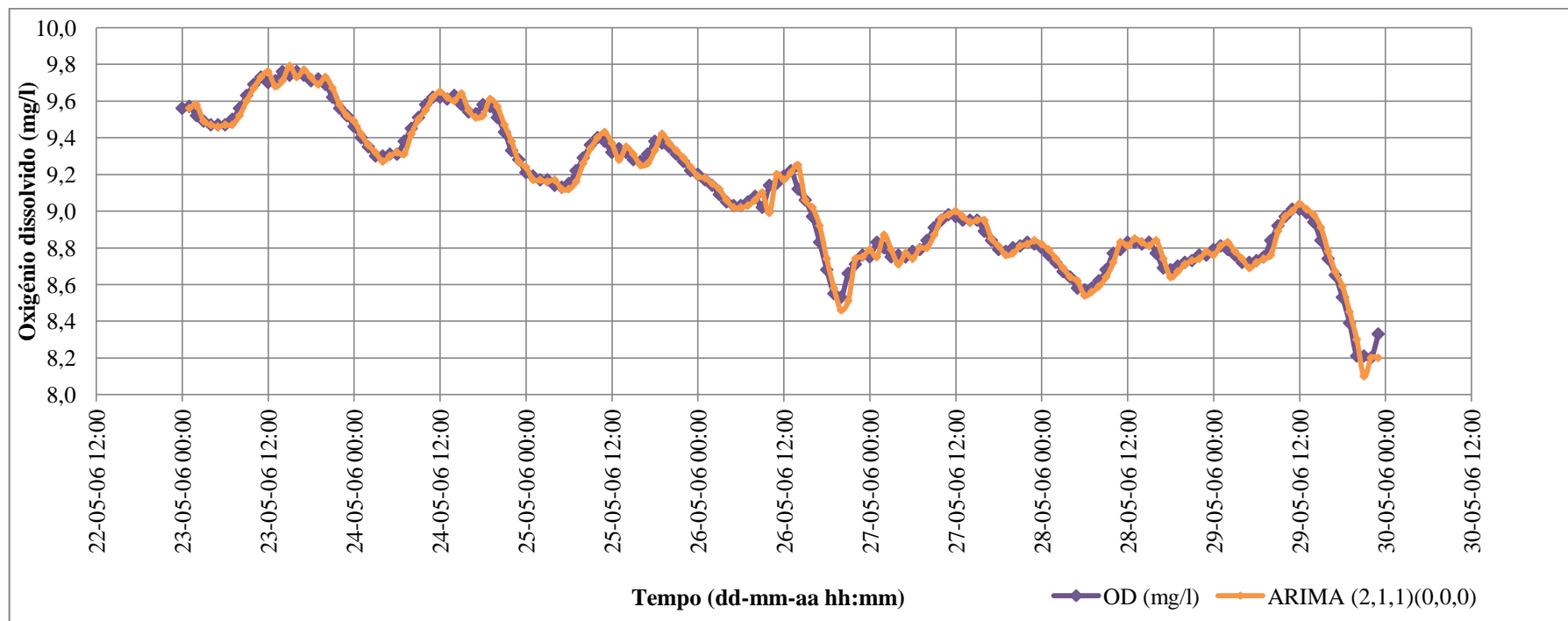


Figura 25 - Previsão para as 24 h do dia 29-05-2006 usando modelo ARIMA (2,1,1)(0,0,0) para a sucessão cronológica de oxigénio dissolvido *versus* dados históricos registados pela sonda.

No caso da previsão efetuada, a vantagem residiu no fato de se ter utilizado dados passados, existindo a possibilidade de efetuar a comparação entre os valores previstos com os valores observados. Neste caso, uma vez que se efetuou a previsão para o dia 29-05-2006 e já se conheciam os valores que realmente foram observados, tornou-se fácil a constatação da qualidade do modelo. Assim, uma vez que o coeficiente de correlação entre os valores registados e os valores previstos para esse dia foi $r=99.4\%$, pode concluir-se que o modelo selecionado é bastante fiável no que respeita a previsões.

De modo a verificar a precisão do modelo selecionado, em seguida apresentam-se em pormenor, na Figura 26, os valores previstos para as 24h do dia 29-05-2006 *versus* valores registados para o mesmo período.

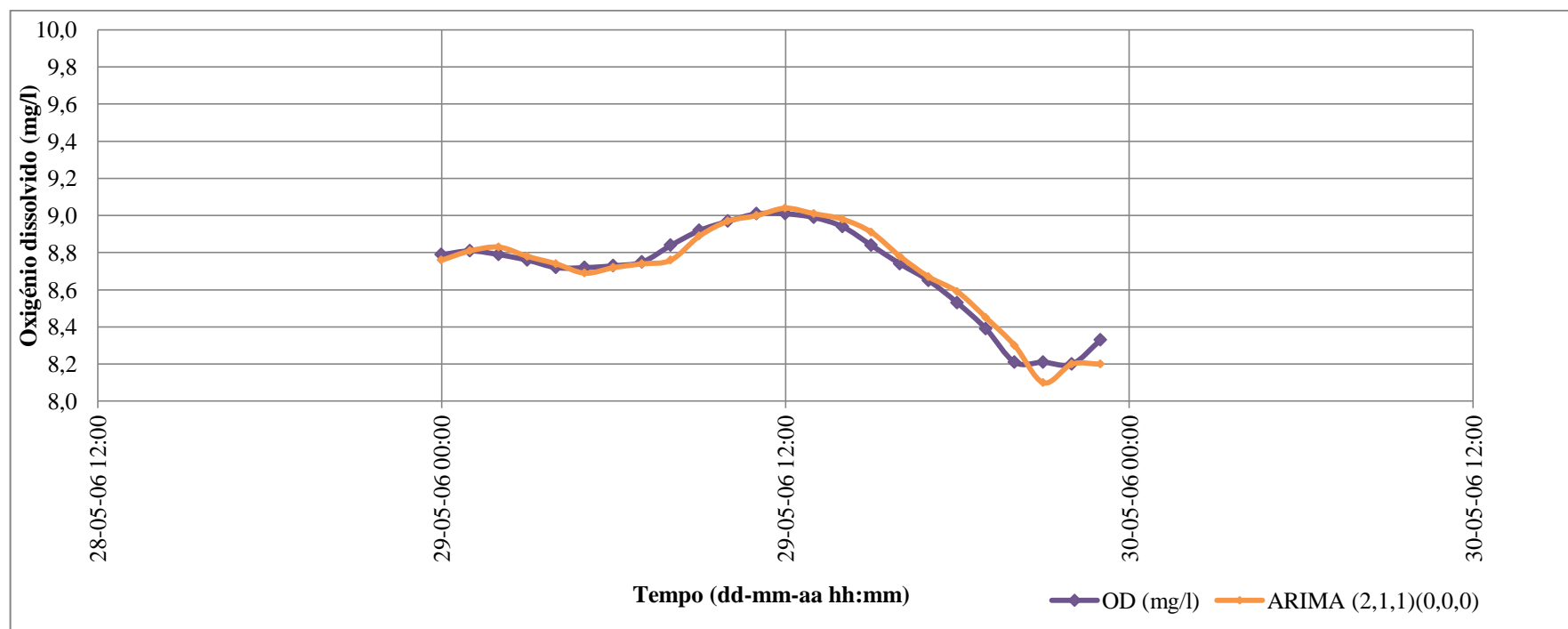


Figura 26 - Pormenor dos valores previstos para as 24 h do dia 29-05-2006, usando o modelo ARIMA (2,1,1)(0,0,0), *versus* valores registados para o mesmo período.

5.5. Caso de estudo 5: Modelos gerados para o ano de 2005 *versus* modelos gerados para o ano de 2012

Para ilustrar este caso de estudo, foi selecionado ao acaso o período de 21-06-2005 a 27-06-2005 e também o período de 21-06-2012 a 27-06-2012. As variáveis em estudo foram a condutividade elétrica e a temperatura, tal como se pode observar na Figura 27, 28, 29 e 30.

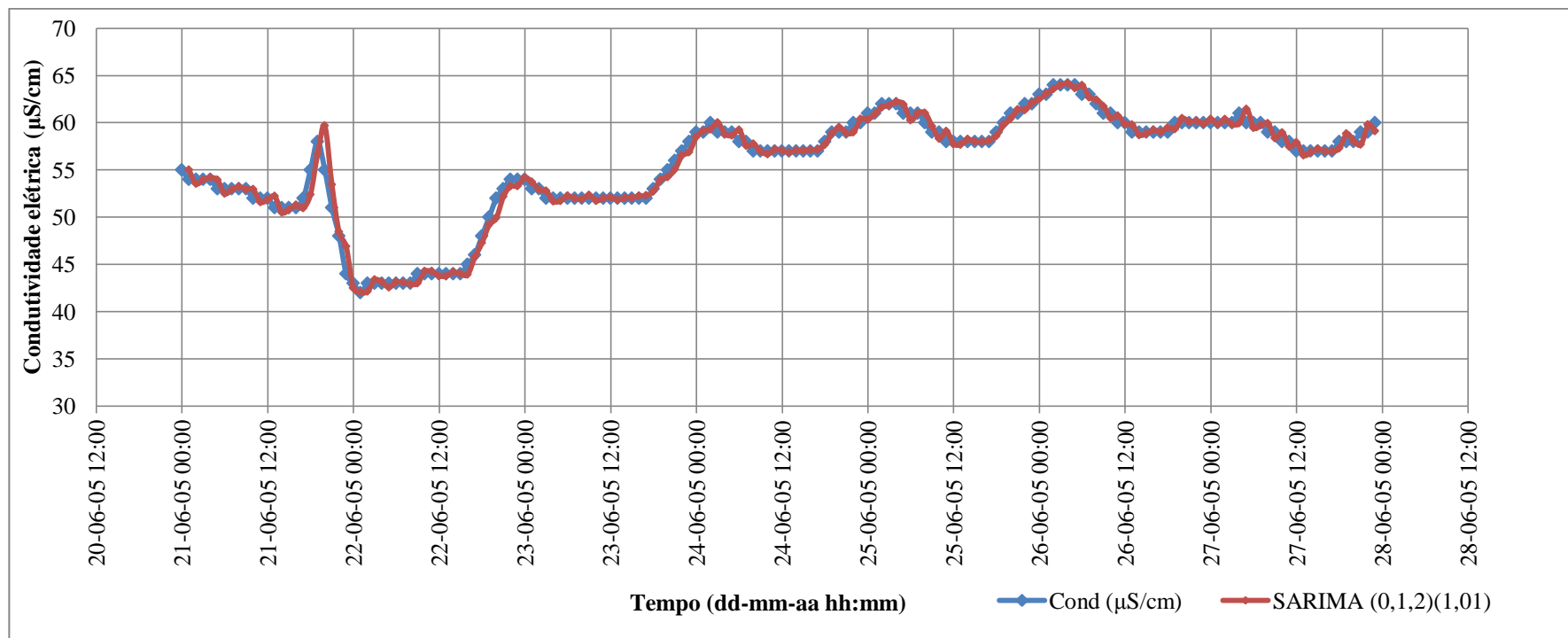


Figura 27 - Dados históricos de Condutividade Elétrica *versus* modelo SARIMA (0,1,2)(1,0,1), para o período de 21-06-2005 a 27-06-2005.

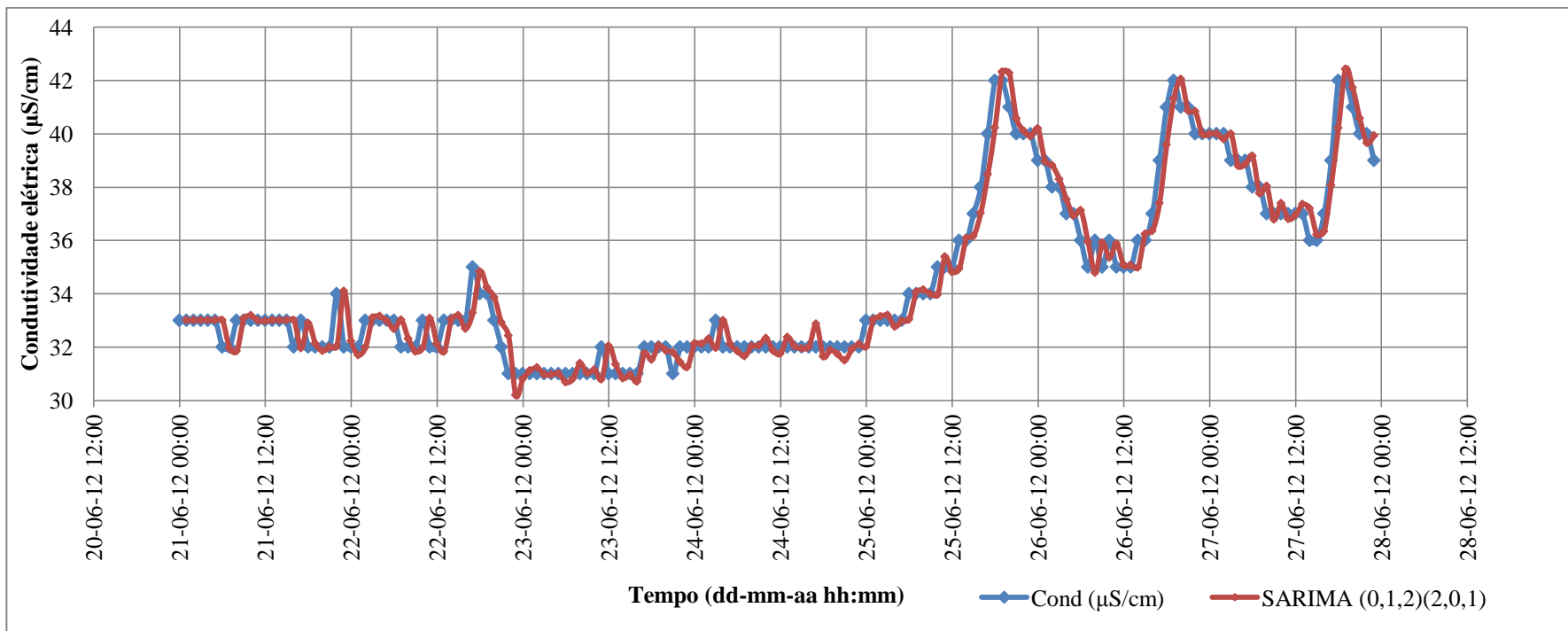


Figura 28 - Dados históricos de Condutividade Elétrica *versus* modelo SARIMA (0,1,2)(2,0,1), para o período de 21-06-2012 a 27-06-2012.

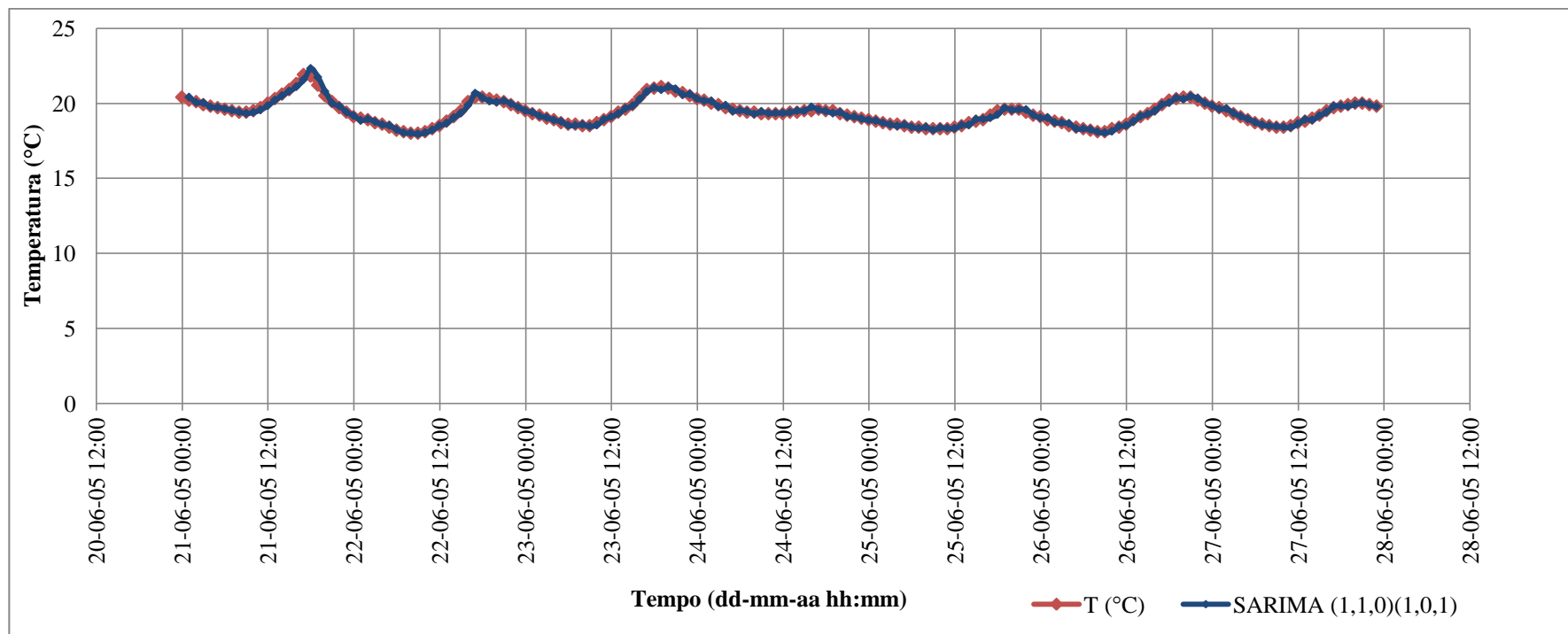


Figura 29 - Dados históricos de Temperatura *versus* modelo SARIMA (1,1,0)(1,0,1), para o período de 21-06-2005 a 27-06-2005.

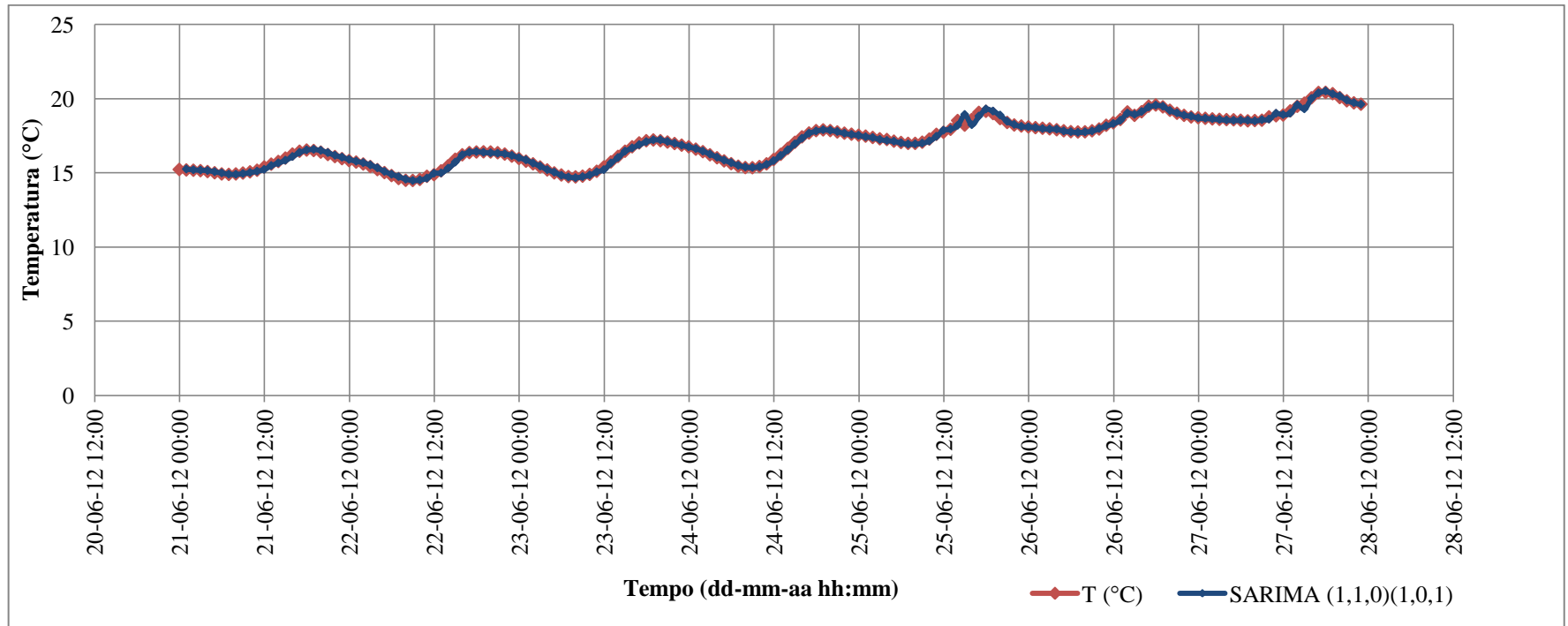


Figura 30 - Dados históricos de Temperatura *versus* modelo SARIMA (1,1,0)(1,0,1), para o período de 21-06-2012 a 27-06-2012.

Tal como esperado, os modelos gerados para o mesmo período do ano são bastante idênticos. Tratam-se todos de modelos do tipo ARIMA, sendo que neste caso os da temperatura são exatamente os mesmos para os dois anos. Relativamente à condutividade elétrica, o coeficiente de correlação para o período de 2005 tomou o valor de 99.1%. Já para o mesmo período mas de 2012, este coeficiente atingiu o valor de 97.7%.

Em relação à temperatura, para o período de 2005 o coeficiente de correlação entre os valores observados e os modelos ajustados foi de 99% e em 2012 foi de 99.8%.

Como já referido, chegar a esta conclusão é de crucial importância, pois para o caso das previsões a curto prazo será bastante útil à partida ter-se logo a noção do tipo de modelo a utilizar ou do intervalo de valores em que os parâmetros do modelo se irão inserir, tendo por base os dados históricos disponíveis.

5.6. Caso de estudo 6: Variação dos parâmetros *versus* variação do caudal

De modo a ilustrar os resultados obtidos para o último caso de estudo, abaixo apresentam-se as Figuras 31, 32, 33 e 34 com os resultados para o período de 14-02-2006 a 20-02-2006, das sucessões cronológicas dos registos *versus* variação do caudal, para os quatro parâmetros em estudo.

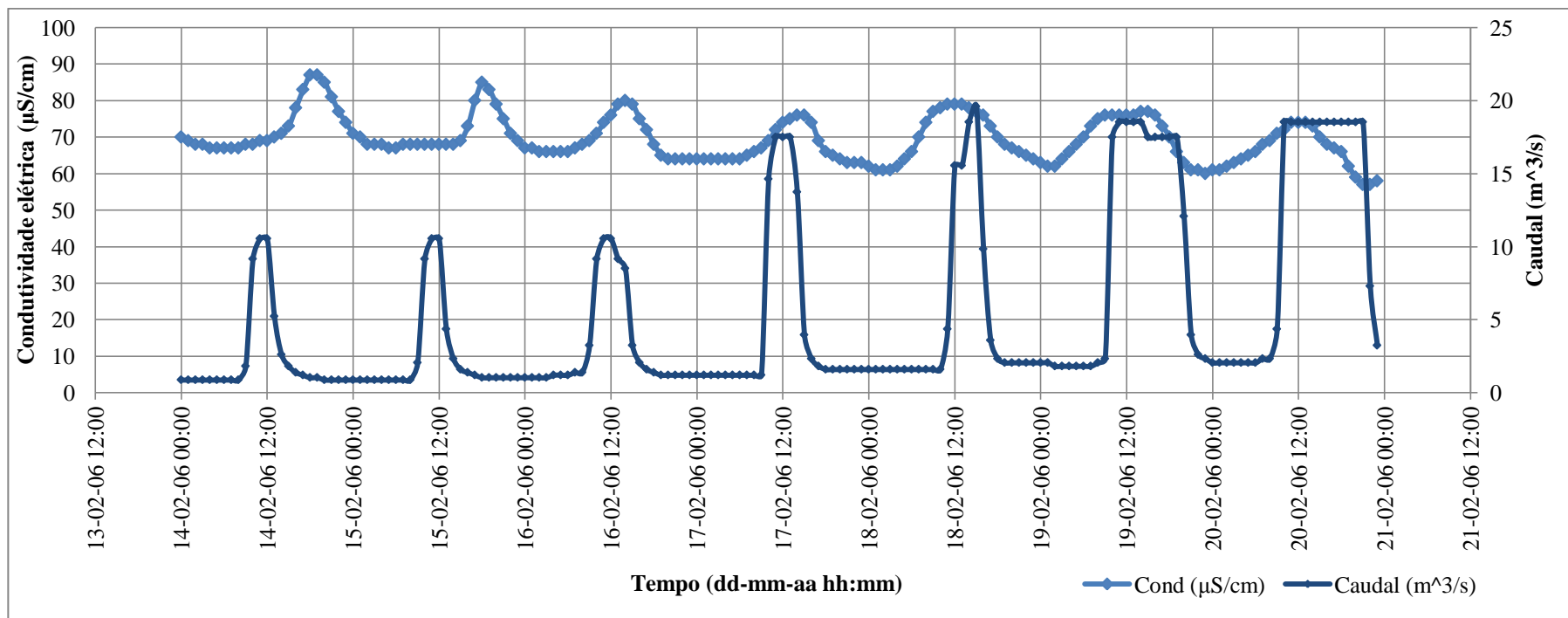


Figura 31 - Dados históricos de Condutividade Elétrica e Caudal a montante para o período de 14-02-2006 a 20-02-2006.

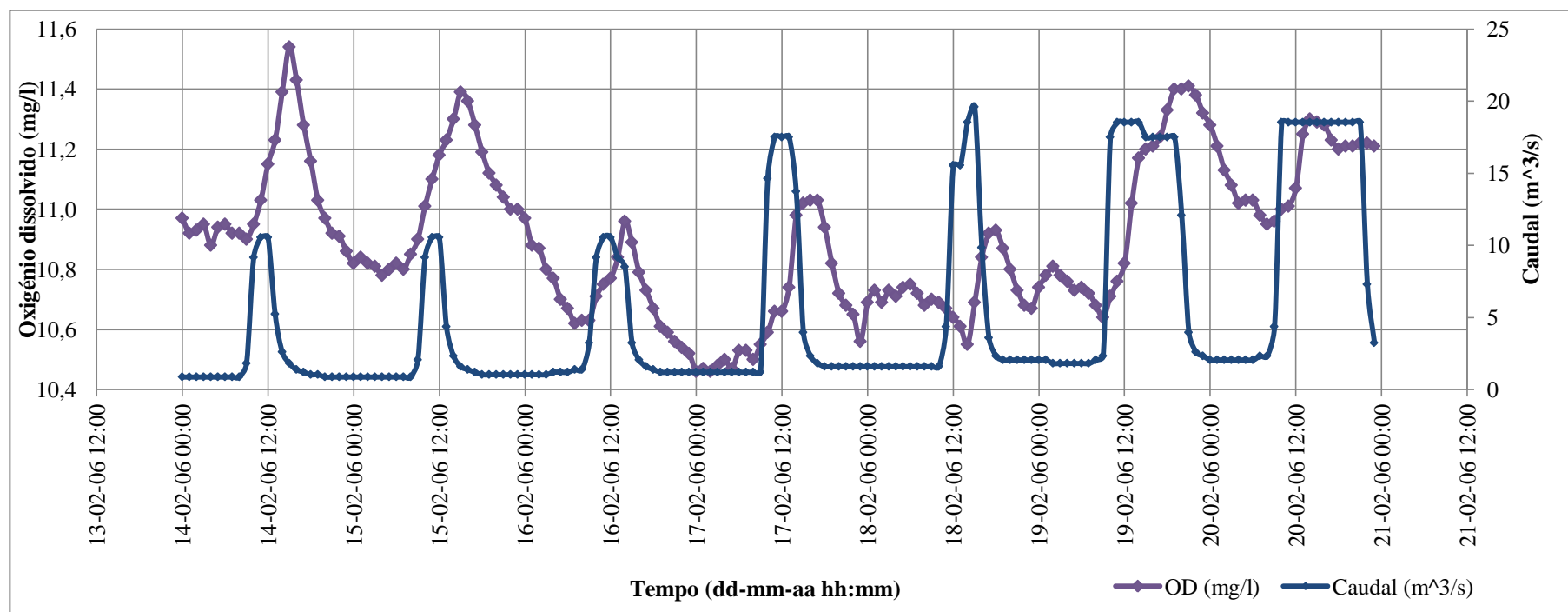


Figura 32 - Dados históricos de Oxigênio Dissolvido e Caudal a montante para o período de 14-02-2006 a 20-02-2006.

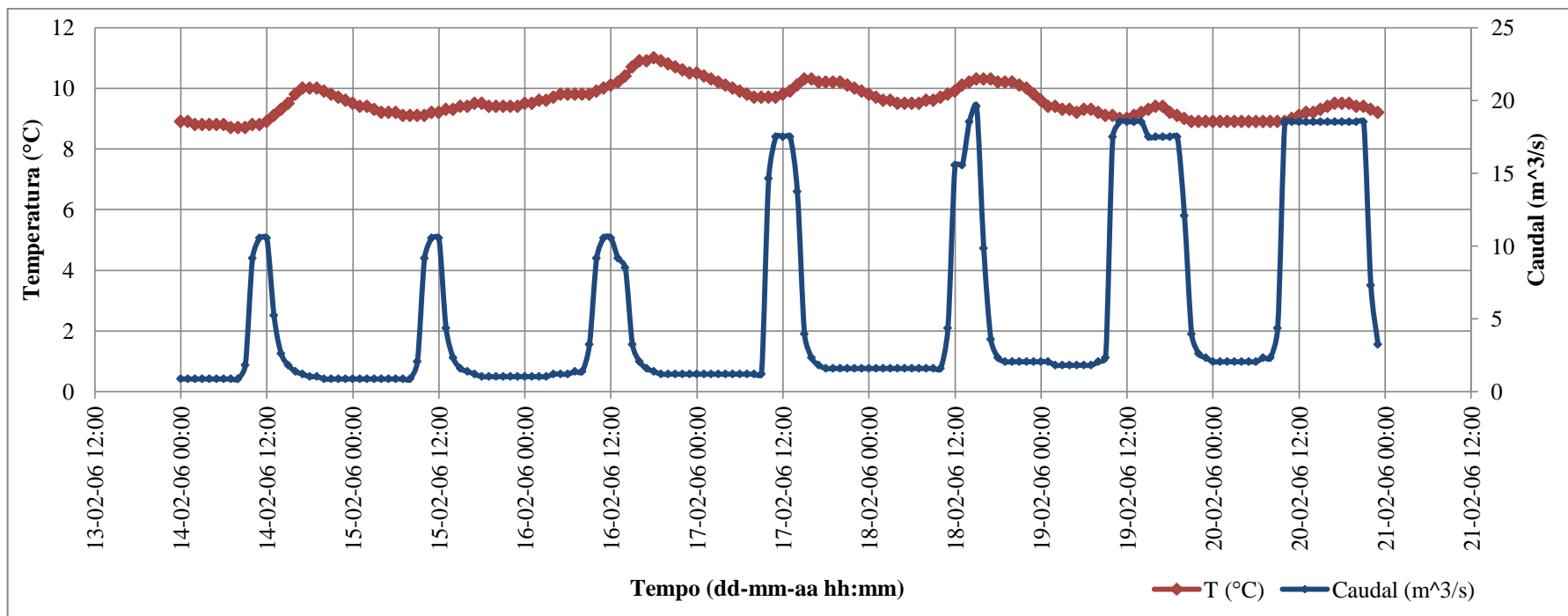


Figura 33 - Dados históricos de Temperatura e Caudal a montante para o período de 14-02-2006 a 20-02-2006.

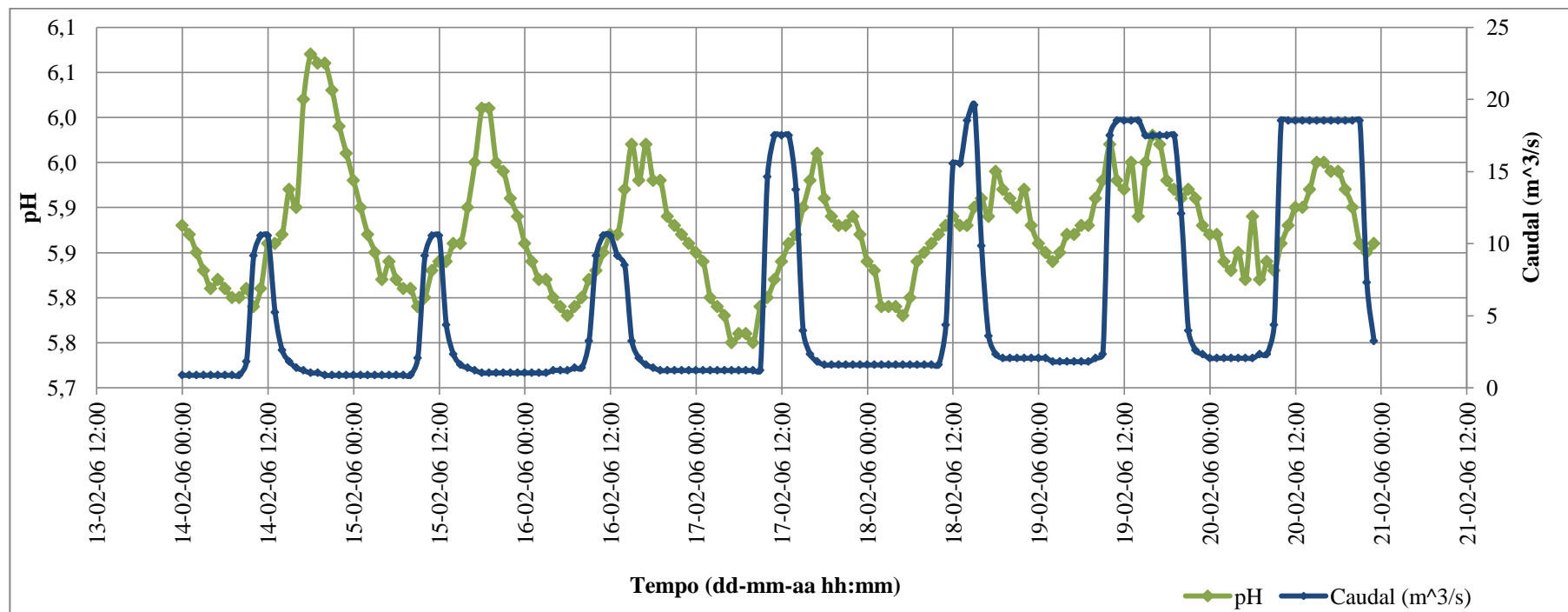


Figura 34 - Dados históricos de pH e Caudal a montante para o período de 14-02-2006 a 20-02-2006.

Deve ainda salientar-se que após a construção dos gráficos para todas as semanas, através de uma análise cuidada e, uma vez que se trata de um longo período, pode verificar-se que há parâmetros que são mais suscetíveis às variações de caudal do que outros. Assim, pode concluir-se que os parâmetros que mais variam em função do caudal são a condutividade elétrica e o oxigénio dissolvido. Obviamente que os restantes parâmetros também variam, mas não de forma tão direta. É no entanto de ressaltar que, em qualquer um dos casos, a variação não é instantânea, uma vez que no caso de existir um aumento ou uma diminuição de caudal resultante das descargas dos aproveitamentos hidroelétricos, esta acontece sempre de montante para jusante e decorre sempre um determinado intervalo de tempo.

Em jeito de conclusão, resta ressaltar que de entre os diversos modelos gerados para as várias séries de dados, apenas foram selecionados para análise os que melhor se ajustavam às variações dos quatro parâmetros, no entanto, não obsta que possam existir modelos que se ajustem melhor.

❖ **Parâmetros de ajustamento dos modelos gerados no SPSS**

A título de exemplo, nas Tabelas 3, 4, 5 e 6 apresenta-se parte da informação e os modelos gerados no *software SPSS*, que melhor se ajustaram às sucessões cronológicas dos valores registados, para a condutividade elétrica, no rio Ave, no período de 30-05-2006 a 05-06-2006. Serão apresentados apenas os melhores modelos e as suas diferenças, que conduziram à escolha do melhor.

No caso específico apresentado, o modelo selecionado foi o SARIMA (0,1,1)(1,0,0), cujo coeficiente de correlação entre os valores reais da sucessão e do modelo SARIMA é $r=98.70\%$. Relativamente ao modelo SARIMA (0,1,1)(1,1,1), o coeficiente de correlação era $r=98.00\%$, daí ser selecionado o primeiro modelo.

Tabela 3 - Ajuste do modelo SARIMA (0,1,1)(1,0,0), gerado a partir da opção "*Expert Modeler*".

Estatística de ajuste	Média	SE	Mínimo	Máximo	Percentil							
					5	10	25	50	75	90	95	
R quadrado estacionário	,377	.	,377	,377	,377	,377	,377	,377	,377	,377	,377	,377
R ao Quadrado	,973	.	,973	,973	,973	,973	,973	,973	,973	,973	,973	,973
RMSE	1,329	.	1,329	1,329	1,329	1,329	1,329	1,329	1,329	1,329	1,329	1,329
MAPE	1,534	.	1,534	1,534	1,534	1,534	1,534	1,534	1,534	1,534	1,534	1,534
MaxAPE	11,628	.	11,628	11,628	11,628	11,628	11,628	11,628	11,628	11,628	11,628	11,628
MAE	,940	.	,940	,940	,940	,940	,940	,940	,940	,940	,940	,940
MaxAE	7,791	.	7,791	7,791	7,791	7,791	7,791	7,791	7,791	7,791	7,791	7,791
BIC normalizado	,630	.	,630	,630	,630	,630	,630	,630	,630	,630	,630	,630

Tabela 4 - Parâmetros do modelo SARIMA (0,1,1)(1,0,0).

				Estimativa	SE	t	Sig.	
Condutividade- Modelo_1	Condutividade	Nenhuma Transformação	Diferença	1				
			MA	Retardo 1	-,565	,065	-8,684	,000
			AR, Sazonal	Retardo 1	,297	,078	3,829	,000

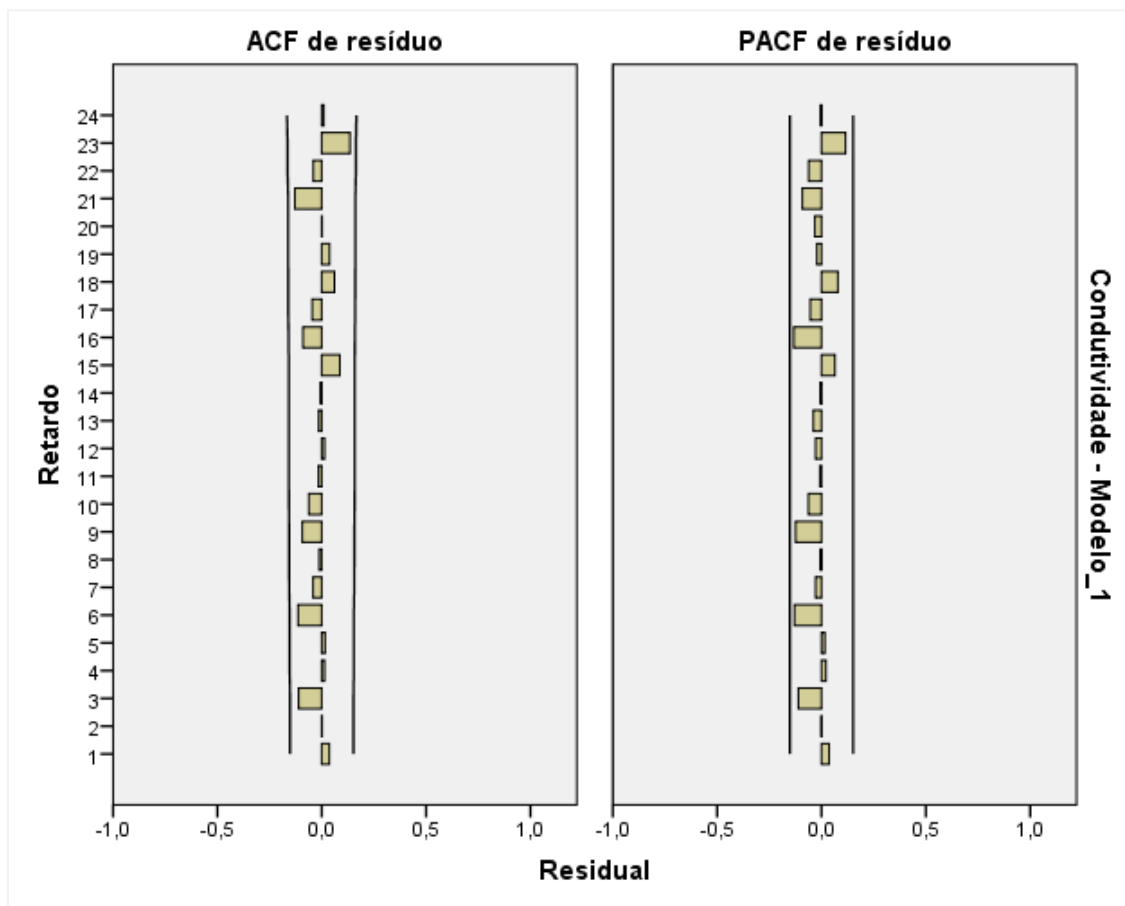


Figura 35 - Funções de autocorrelação e autocorrelação parcial da sucessão residual do modelo SARIMA (0,1,1)(1,0,0) para a variável Condutividade Elétrica.

Tabela 5 - Ajuste do modelo SARIMA (0,1,1)(1,1,1), gerado manualmente.

Estatística de ajuste	Média	SE	Mínimo	Máximo	Percentil							
					5	10	25	50	75	90	95	
R quadrado estacionário	,327	.	,327	,327	,327	,327	,327	,327	,327	,327	,327	,327
R ao Quadrado	,958	.	,958	,958	,958	,958	,958	,958	,958	,958	,958	,958
RMSE	1,527	.	1,527	1,527	1,527	1,527	1,527	1,527	1,527	1,527	1,527	1,527
MAPE	1,769	.	1,769	1,769	1,769	1,769	1,769	1,769	1,769	1,769	1,769	1,769
MaxAPE	11,423	.	11,423	11,423	11,423	11,423	11,423	11,423	11,423	11,423	11,423	11,423
MAE	1,071	.	1,071	1,071	1,071	1,071	1,071	1,071	1,071	1,071	1,071	1,071
MaxAE	7,653	.	7,653	7,653	7,653	7,653	7,653	7,653	7,653	7,653	7,653	7,653
BIC normalizado	,985	.	,985	,985	,985	,985	,985	,985	,985	,985	,985	,985

Tabela 6 - Parâmetros do modelo SARIMA (0,1,1)(1,1,1).

				Estimativa	SE	t	Sig.	
Condutividade- Modelo_1	Condutividade	Nenhuma Transformação	Constante	-,025	,091	-,273	,786	
			Diferença	1				
			MA	Retardo 1	-,525	,073	-7,144	,000
			AR, Sazonal	Retardo 1	,259	,170	1,522	,130
			Diferença sazonal		1			
			MA, Sazonal	Retardo 1	,985	3,423	,288	,774

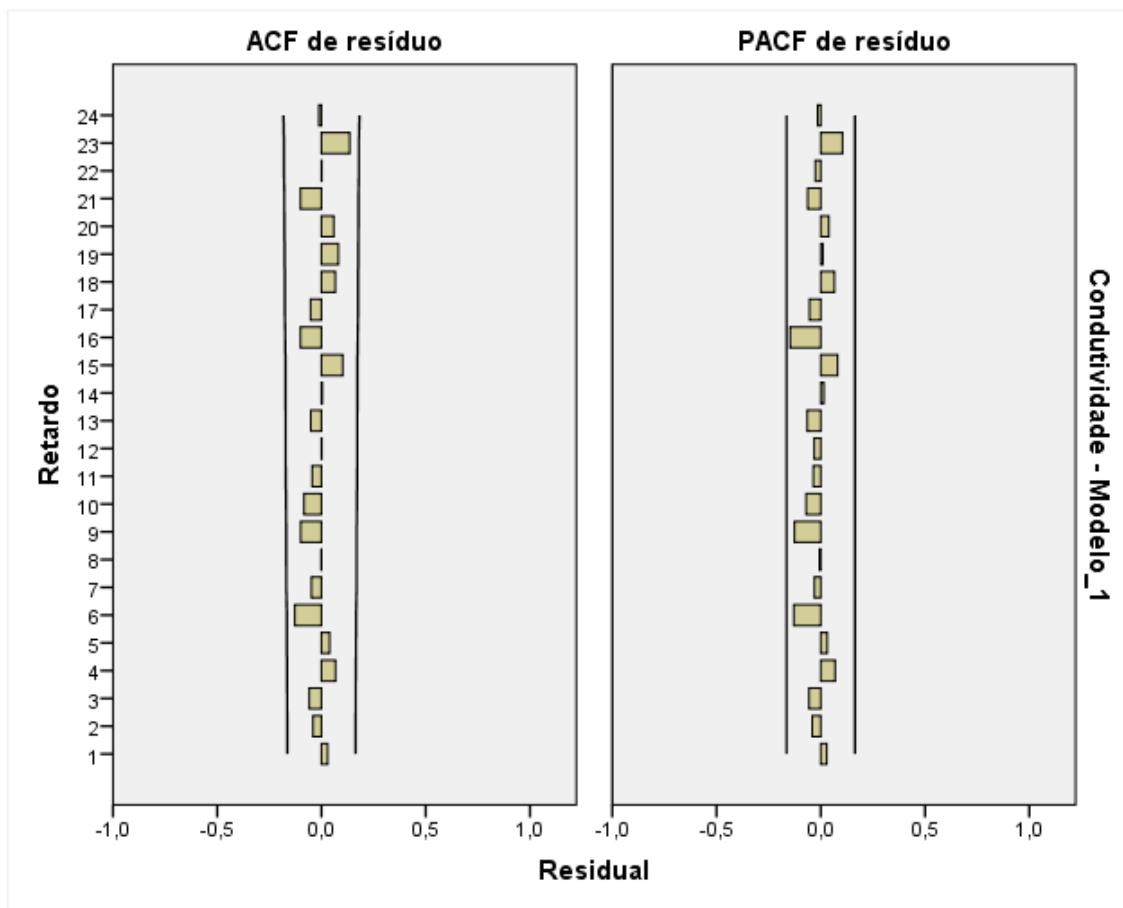


Figura 36 - Funções de autocorrelação e autocorrelação parcial da sucessão residual do modelo SARIMA (0,1,1)(1,1,1) para a variável Condutividade Elétrica.

Relativamente às Figuras 35 e 36 com os gráficos das funções de autocorrelação e autocorrelação parcial de resíduos apresentados, conforme se pode verificar, ambos se encontram dentro dos limites críticos, pelo que a escolha dos modelos se baseou nos restantes critérios mencionados no capítulo 5.

6. CONCLUSÕES

Na presente dissertação efetuou-se a caracterização, através de ferramentas estatísticas, da evolução temporal de parâmetros indicadores da qualidade da água. Para tal, foram desenvolvidos modelos matemáticos de sucessões cronológicas, tendo em vista a modelação e a previsão de parâmetros de qualidade da água do rio Ave.

Numa primeira fase foi essencial definir quais os parâmetros a estudar, para que estes fossem os mais representativos da qualidade do meio hídrico em análise. Esta etapa ficou facilitada uma vez analisados os registos disponíveis da estação de monitorização automática do rio Ave selecionada. Seguidamente procurou-se também entender qual a importância de cada uma das variáveis e de que maneira estas se relacionam com a qualidade da água. Obviamente que existem outras variáveis que poderiam ter sido exploradas e estudadas, no entanto e dadas as características de cada uma, considera-se que as escolhidas serão as mais importantes, uma vez que foram também utilizadas pelo Sistema Nacional de Informação de Recursos Hídricos para caracterizar a qualidade das águas a nível nacional, nas estações de monitorização automáticas. São também estas as variáveis que permitem identificar os problemas mais frequentes da qualidade da água.

Uma vez definidos os parâmetros a analisar, efetuou-se um estudo de alguns dos vários modelos que caracterizam as séries temporais. Esta etapa foi crucial, uma vez que não faria qualquer sentido partir para a aplicação do *software SPSS* sem primeiro adquirir as noções básicas ou os principais conceitos associados aos modelos a que o programa recorre e escolhe automaticamente. Nem todos os modelos estudados nesta fase estarão disponíveis no *software* utilizado, no entanto, considera-se essencial ter a perceção de que existem várias formas de analisar e estudar as séries temporais. Assim, após a pesquisa e o estudo realizados relativamente aos modelos existentes, salienta-se a importância que a escolha de um modelo possui. Nesta etapa deve sempre ter-se em conta modelos alternativos, eventualmente vizinhos do primeiro modelo selecionado, já que muitas vezes existem alguns modelos a descrever a mesma sucessão cronológica com igual grau de qualidade. A seleção do modelo a adotar irá então condicionar todo o processo de previsão, pelo que deverão sempre ser tidos em conta vários critérios de seleção.

Em relação ao caso de estudo, o seu enquadramento no espaço e no tempo, acompanhado da revisão do estado atual dos conhecimentos, permitiram retirar algumas conclusões sobre o comportamento dos parâmetros indicadores da qualidade da água do rio Ave. Aqui vale ressaltar que o vale do Ave em tempos foi um grande centro empresarial nomeadamente na área da indústria têxtil, pelo que a contaminação das suas águas surge em consequência dessas mesmas indústrias. No entanto, conforme se conseguiu apurar, a qualidade do rio Ave tem vindo a aumentar significativamente.

Relativamente ao *software SPSS*, através da sua utilização tornou-se então possível não só caracterizar as sucessões cronológicas dos parâmetros em estudo como efetuar previsões. Neste capítulo, começou-se por analisar a variação de cada parâmetro individualmente, relacionando depois o caudal do rio e por fim foram ajustados modelos estatísticos às variações de cada um dos parâmetros. Os modelos gerados e selecionados para todo o período de estudo foram de vários tipos, no entanto deu-se especial atenção aos modelos ARIMA, por serem modelos simples e por possuírem, regra geral, um bom ajuste às variações registadas.

Face aos resultados obtidos e uma vez que estes revelaram sempre um coeficiente de correlação bastante elevado, considera-se que ficou provado que os modelos estatísticos são bastante fiáveis. Claramente que, tal como outros modelos, os modelos utilizados também possuem as suas limitações, além de que deve ainda referir-se que estes poderão não ter sido os melhores. Esta questão prende-se com o facto de muitas vezes ter sido necessário explorar manualmente novos modelos, daí existir sempre a possibilidade de se encontrar um modelo melhor para determinado período. No entanto, tal como referido anteriormente, os coeficientes de correlação ente os valores observados e os modelos gerados pelo *software SPSS* foram sempre bastante elevados, pelo que se conclui que os modelos estatísticos serão bastante credíveis.

Ainda no contexto da modelação estatística, foi possível concluir que quanto mais curtos forem os períodos em análise, mais precisos serão os resultados. Como exemplo disso, tem-se o caso dos modelos gerados pelo *SPSS* que contemplam todo o período de estudo e a análise feita semana a semana. No primeiro caso os modelos foram sempre do tipo ARIMA, já no segundo nem sempre tal facto se verificou, ou seja, quanto mais curtos forem os períodos, melhor será o ajuste dos modelos. Assim, o conceito de previsão, seja de curto ou de

longo prazo, estará diretamente relacionado com a variável independente dos dados, pelo que para uma previsão de longo prazo, um ano, por exemplo, os dados utilizados não serão em horas ou dias, mas sim meses. Do mesmo modo que, tal como na presente dissertação, utilizando dados horários ou diários, a intenção será fazer-se previsões de curto prazo, como por exemplo, previsão para uma semana.

Uma conclusão interessante também obtida através da aplicação do referido *software*, reside no facto de se registarem sempre o mesmo tipo de modelos ou modelos bastante semelhantes para os mesmos períodos do ano. Nesta fase foi fundamental utilizar dados mais recentes para se efetuar a comparação ente os modelos gerados. Esta conclusão facilitará em muito a tarefa de modelação de problemas deste tipo, pois à partida conhecendo o período em análise, saber-se-á logo qual o intervalo de valores ou qual o tipo de modelo que em princípio se irá utilizar para o período que se quer tratar. Obviamente que nunca se dispensará a observação de registos obtidos no período imediatamente anterior.

As técnicas baseadas em séries temporais procuram modelar matematicamente acontecimentos futuros, relacionando-os com dados passados das variáveis, ou seja, parte-se do princípio que o futuro será uma espécie de projeção do passado. Certamente que existem sempre perturbações que podem ocorrer no meio hídrico que não serão contempladas nos modelos, mas como em qualquer outro método, existirão sempre ocorrências que não estão previstas, que farão parte da margem de erro destes modelos. No entanto, tendo em conta todos os resultados obtidos considera-se que a modelação estatística pode ser uma ferramenta bastante útil na área da gestão dos recursos hídricos, contribuindo para a caracterização da qualidade do meio hídrico e ajudando na definição de políticas sustentáveis.

Em suma, pode concluir-se que os métodos quantitativos proporcionam uma boa aproximação da realidade a curto prazo, desde que devidamente empregues. No entanto, as incertezas vão sempre existir. Assim, não deve nunca esquecer-se que quanto maior o horizonte previsto, maior será a probabilidade de erro.

6.1. Sugestões para desenvolvimentos futuros

Relativamente à metodologia utilizada existem ainda algumas tarefas que poderiam realizar-se como continuidade deste estudo. Deste modo, como propostas para possíveis trabalhos futuros, no âmbito da modelação estatística, sugerem-se as seguintes tarefas:

- ❖ Utilização de modelos que relacionem as variáveis entre si, como por exemplo modelos ARIMA que considerem como variável dependente a condutividade elétrica e como variável independente a temperatura, uma vez que se tratam de dois parâmetros que se encontram intimamente ligados. O mesmo se sugere para os restantes parâmetros;
- ❖ Associar a metodologia desenvolvida a um sistema de alerta que permita fazer previsões a curto prazo em tempo real. A partir daí será desde logo facilitada a tomada de decisões no que respeita, por exemplo, às utilizações possíveis do meio hídrico em questão.

REFERÊNCIAS BIBLIOGRÁFICAS

- Agência Portuguesa do Ambiente*. (s.d.). Obtido em 16 de Agosto de 2014, de www.apambiente.pt.
- Ambiente, M. d. (1998). Diário da República. *Decreto-Lei nº 236/98 de 1 de Agosto*. Lisboa.
- Ashurst, P. R., & Dennis, M. J. (1998). *Analytical Methods of Food Authentication*.
- Berzas, J. J., García, L. F., & Rodríguez, R. C. (2000). *Evolution of the Water Quality of a Managed Natural Wetland. Tablas de Daimiel National Park (Spain)*.
- Bollmann, H. A., & Marques, M. D. (2000). *Bases para Estruturação de Indicadores de Qualidade de Águas*.
- Box, G. E., & Jenkins, G. M. (1976). *Time series analysis forecasting and control*.
- Costa, F. S. (s.d.). A indústria têxtil na bacia hidrográfica do rio Ave. *Uma perspectiva segundo as fábricas de fição e tecidos, numa relação historicamente sustentada pelo domínio público hídrico*.
- Costa, F. S. (2008). Hidro-conflitos na bacia hidrgráfica do rio Ave. *Uma análise a partir das transgressões cometidas no período 1902-1973*.
- Couillard, D., & Lefebvre, Y. (1985). Analysis of water quality indices.
- Da Silva Pinho, J. L. (2000). *Aplicação da modelação matemática ao estudo da hidrodinâmica e da qualidade da água em zonas costeiras*. Guimarães, Portugal: Universidade do Minho.
- Diggle, P. J. (1992). *Time series: a biostatistical introduction*.
- Draper, N. R., & Smith, S. (1981). Applied regression analysis. *Wiley Series in Probability and Mathematical Statistics*.
- Duarte, A. A. (2013). *Apontamentos teóricos de Processos em Hidráulica Ambiental*. Guimarães.
- Everitt, B. S. (1995). *The Cambridge dictionary of statistics in the medical sciences*.

- Faruk, D. O. (Setembro de 2009). *A hybrid neural network and ARIMA model for water quality time*.
- Ferreira, J. G. (s.d.). *Saneamento Básico. Proposta de análise do problema em Portugal 1970-2000*.
- Ferrier, R. C., Edwards, A. C., & Hirst, D. (2001). *Water Quality of Scottish Rivers. Spatial and Temporal Trends*.
- Gibbons, D. C. (1987). *The Economic Value of Water*. New York: Resources for the future.
- Gonçalves, A. M. (2005). *Modelação Estatística da Qualidade das Águas de Superfície da Bacia Hidrográfica do Rio Ave*. Lisboa, Portugal: Universidade de Lisboa.
- Haase, J., Krieger, J. A., & Possoli, S. (1989). *Estudo da viabilidade do uso da técnica fatorial como um instrumento na interpretação de qualidade das águas da bacia hidrográfica do Guaíba, RS, Brasil*.
- Harmancioglu, N. B., Ozkul, S. A., & Alpaslan, M. N. (1998). *Water monitoring and network design*.
- Infopédia*. (s.d.). Obtido em 15 de Junho de 2014, de <http://www.infopedia.pt>.
- Jonhson, R. A., & Wichern, D. W. (1992). *Applied Multivariate Statistical Analysis*.
- Khatibi, R., Ghorbani, M., Naghipour, L., Jothiprakash, V., Fathima, T., & Fazelifard, M. (Janeiro de 2014). *Inter-comparison of time series models of lake levels predicted*.
- Latorre, M. d., & Cardoso, M. R. (2001). *Análise de séries temporais em epidemiologia. Uma introdução sobre os aspetos metodológicos*.
- Lima, E. B. (2001). *Modelação integrada para gestão da qualidade da água na bacia do rio Cuiabá*. Rio de Janeiro, Brasil.
- Lima, M. M. (2006). *Estimativa de concentração de material particulado em suspensão na atmosfera por meio da modelagem de redes neurais artificiais*. Belo Horizonte, Brasil: Universidade Federal de Minas Gerais.
- Lohani, A., & Rakesh Kumar, R. S. (Março de 2012). *Hydrological time series modeling. A comparison between adaptive neuro-fuzzy, neural network and autoregressive techniques*.

- Morettin, P. A., & Toloí, C. M. (1985). *Previsão de séries temporais*.
- Murteira, B., Muller, D., & Turkman, K. (1993). *Análise de Sucessões Cronológicas*.
- Oliveira, C., Campos, V. P., & Medeiros, Y. D. (2010). *Avaliação e identificação de parâmetros importantes para a qualidade de corpos d'água no semiárido Baiano. Estudo de caso: Bacia hidrográfica do Rio Salitre*.
- Oliveira, S. (1994). *Relatório de qualidade ambiental no Estado de São Paulo*.
- Orea, D. G. (1998). *Evaluación de impacto ambiental*.
- Otto, M. (1998). *Analytical Chemistry - Multivariate Methods*.
- Parry, R. (1998). Agriculture phosphorus and water quality. *A U.S. Environmental Protection Agency perspective*.
- Peters, N. E., & Meybeck, M. (2000). *Water Quality Degradation Effects on Freshwater. Impacts of Human Activities*.
- Pineda, M. D., & Schäfer, A. (1987). *Adequação de critérios e métodos de avaliação da qualidade de águas superficiais baseada no estudo ecológico do Rio Gravataí, Rio Grande do Sul, Brasil*.
- Possoli, S., & Haase, J. (1993). *Estudo da utilização da técnica de análise fatorial na elaboração de um índice de qualidade de água*.
- Quintela, A., & Portela, M. M. (2002). *A Modelação Hidrológica em Portugal nos Últimos 25 Anos do Século XX, nas Perspectivas Determinística, Probabilística e Estocástica*.
- SAS, I. (1985). *SAS/SAT user's guide: statistics*.
- Shahwan, T., & Odening, M. (2007). *Computational Intelligence in Economics and Finance*.
- Shoji, H., Yamamoto, T., & Nakamura, T. (1966). *Factor analysis on stream pollution of the Yodo River systems*.
- Silva, C. F. (2013). *Modelação de Sucessões Cronológicas de Microrganismos presentes na Água no Vale do Ave*. Guimarães, Portugal: Universidade do Minho.

- Silva, J. F., Salgado, S. M., & Menezes, R. (2009). 1º Seminário sobre Gestão de Bacias Hidrográficas: As Regiões do Norte e as Perspectivas Futuras de Gestão. *Modelação e previsão de sucessões cronológicas de parâmetros de águas superficiais*.
- Sims, J. T., Simard, R. R., & Joern, B. C. (1998). *Phosphorus loss in agricultural drainage. Historical perspective and current research*.
- Sistema Nacional de Informação de Recursos Hídricos*. (2014). Obtido em 15 de Fevereiro de 2014, de <http://snirh.pt/>.
- Sliva, L., & Williams, D. D. (2001). *Buffer Zone Versus Whole Catchment Approaches to Studying Land Use Impact on River Water Quality*.
- Toledo, L. G., & Nicolella, G. (Janeiro de 2002). *Índice de qualidade de água em microbacia sob uso agrícola e urbano*.
- Valipour, M., Banihabib, M. E., & Behbahani, S. M. (Novembro de 2012). *Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir*.
- Vega, M., Pardo, R., & Barrado, E. (1998). Vega, M., Pardo, R., B. *Assessment of Seasonal and Polluting Effects on the Quality of River Water by Exploratory Data Analysis*.
- Vidal, M., López, A., & Santoalla, M. C. (2000). *Factor Analysis for the Study of Water Resources Contamination Due to the Use of Livestock Shurries as Fertilizer*.
- World Health Organization*. (2014). Obtido em 15 de Fevereiro de 2014, de <http://www.who.int/topics/water/en/>.
- Yassuda, E. R. (1993). *Gestão de Recursos Hídricos. Fundamentos e aspectos institucionais*.
- Zhang, X., Chen, Q., Recknagel, F., & Li, R. (Maio de 2014). *Wavelet analysis of time-lags in the response of cyanobacteria growth to*.