



Universidade do Minho
Escola de Ciências

Séries Temporais

Trabalho de Análise/Modelação de Dados
Temporalmente Correlacionados

Docente: Raquel Menezes

Trabalho efetuado por:

Eduardo Dias A95467

Tiago Monteiro A94905

João Cunha A95756

Conteúdo

1. Introdução
2. Bacia hidrográfica do rio Douro e do rio Tejo
3. Volume afluente
4. Análise descritiva
5. Albufeira da Valeira
 - a. Modelação SARIMA
 - b. Análise dos resíduos
 - c. Previsão SARIMA
6. Conclusão
7. Anexo

Introdução

Nos tempos atuais, a degradação dos recursos hídricos devido à poluição por atividades antropogénicas bem como a sua escassez devido às secas que têm vindo a ser agravadas pelo aquecimento global é inegável. Há falta de métodos, planos de gestão e ferramentas que modelem e prevejam eventos críticos com o objetivo de preservação do Ambiente. Os métodos de modelação são instrumentos importantes para auxiliar a tomada de decisão em diferentes áreas.

Como tal este trabalho teve como principal objetivo modelar e prever diversas séries temporais ambientais relacionadas com o volume de água presente em albufeiras portuguesas.

As observações recolhidas são referentes às bacias hidrográficas do rio Douro e do rio Tejo, localizadas no Norte e Centro de Portugal, respetivamente. A sua monitorização é prioritária, na medida em que existem zonas sujeitas a grandes períodos de seca, nestas bacias hidrográficas.

A variável hidrométrica selecionada para estabelecer o processo de modelação com o objetivo de avaliação e monitorização da quantidade de água foi o volume afluente mensal (dam^3).

Os dados utilizados para a realização deste estudo foram obtidos do repositório do Sistema Nacional de Informação de Recursos Hídricos (SNIRH). Foram recolhidos mensalmente, no período de agosto de 2001 até dezembro de 2008, em estações de monitorização das bacias.

2. Bacia hidrográfica do rio Douro e do rio Tejo

A água é um elemento essencial à vida e é parte integrante do planeta, desde a crosta à atmosfera. Muitas das atividades humanas e estratégias de desenvolvimento na qualidade de vida têm como impulsionador a água

Como referido anteriormente, este estudo recai sobre a análise do volume afluente mensal de estações das bacias hidrográficas do rio Douro e Tejo, no território português.

O rio Douro nasce nos picos da serra da Urbión e desagua no Oceano Atlântico junto às cidades do Porto e Vila Nova de Gaia. Sendo o terceiro rio mais extenso da Península Ibérica.

A sua bacia hidrográfica está localizada a Norte de Portugal e tem uma área aproximada de 79 000 km², dos quais 19 000 km² se encontram em Portugal.

Quanto ao rio Tejo este é o rio mais extenso da Península Ibérica. Nasce na serra de Albarracim e desagua em Lisboa, também no Oceano Atlântico.

A sua bacia hidrográfica tem uma área de 80 600 km², sendo que 24 850 km² são em Portugal.

3. Volume afluente

Um afluente é um rio ou curso de água menor que desagua em rios principais. Um afluente não flui diretamente para um oceano, mar ou lago. Os afluentes e o rio principal servem para drenar uma determinada bacia hidrográfica.

O volume afluente é o volume de água que chega a um dado rio principal, ou, neste caso, a uma determinada albufeira de um rio principal vindo dos seus afluentes.

O volume da água, num determinado local, é o reflexo das condições dominantes da bacia de alimentação desse local, nomeadamente de fatores hidrometeorológicos.

4. Análise descritiva

Nesta secção, pretende-se realizar uma análise inicial dos dados, onde se calculam as estatísticas descritivas básicas das três séries em estudo.

Na Figura 1 encontram-se representadas as 3 séries em estudo bem como os respetivos *decompose*.

É visível a não estacionariedade nas três séries, bem como a existência de sazonalidade em todas elas.

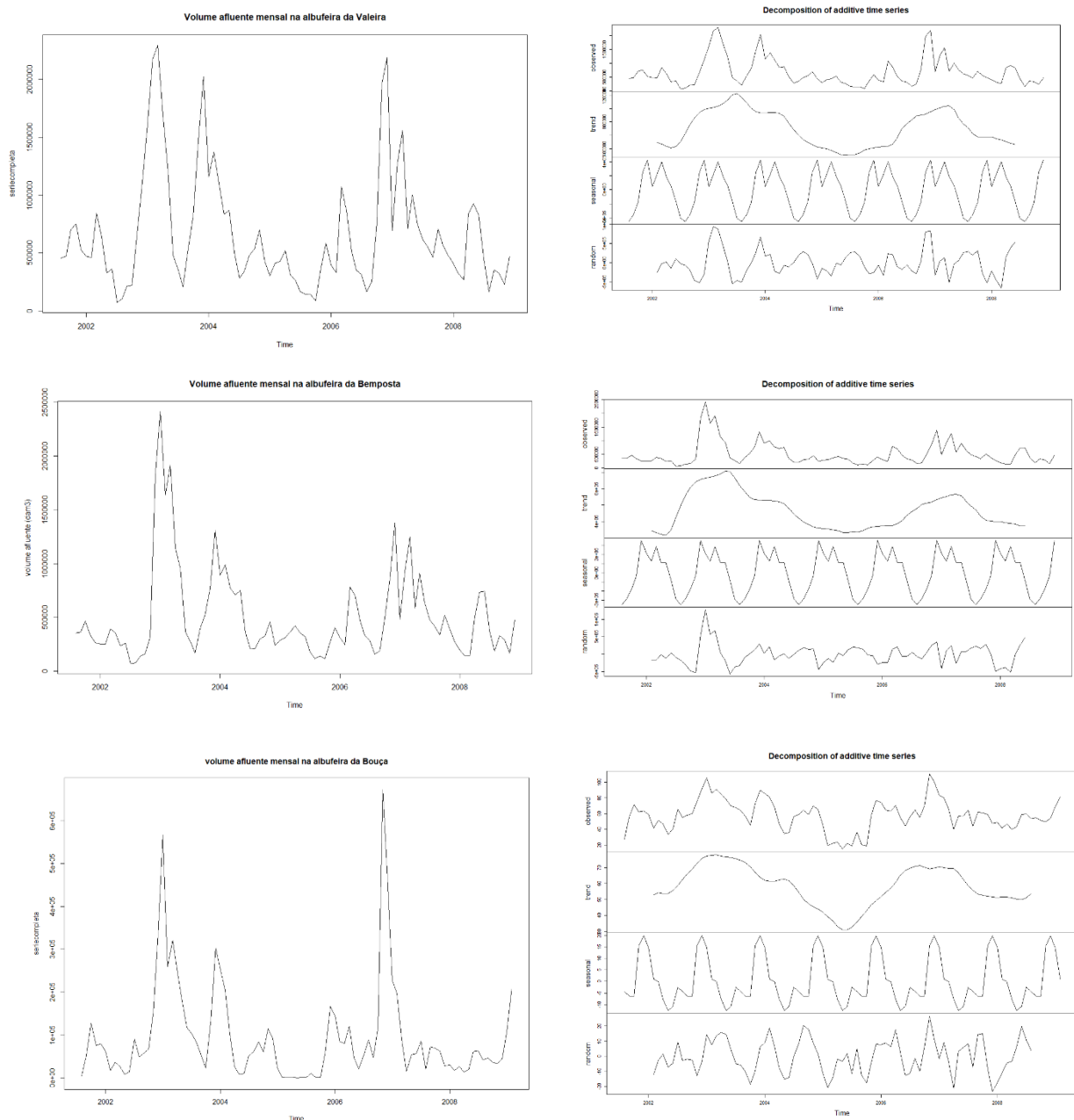


Figura 1: Representação gráfica das séries temporais das três albufeiras em estudo, no período observado.

O maior valor de Volume afluente observado (2412971 dam³) é identificado na albufeira da Bemposta, em julho de 2002. O menor valor de Volume afluente observado (536 dam³) é identificado na albufeira da Bouça, em maio de 2005.

Os dados da albufeira da Valeira e da albufeira da Bouça têm dados omissos (ambas 2), como tal, estes foram preenchidos recorrendo à função *na.interp()* da *library forecast*.

Tabela1: Medidas descritivas da variável do Volume afluente, no período observado

ALBUFEIRAS	MIN.	1ºQUARTIL	MEDIANA	MÉDIA	3ºQUARTIL	MAX.	DADOS OMISSOS
VALEIRA	74460	329651	480505	656040	824896	2295924	2
BEMPOSTA	70442	253757	356918	507175	639446	2412971	0
BOUÇÃ	536	26179	60800	94715	109011	672883	2

Foram também realizados os histogramas e *boxplots* dos dados das três albufeiras (Figura 2). Constata-se que todos os histogramas são assimétricos à direita principalmente o da albufeira da Bemposta e da Bouçã. São também visíveis diversos outliers, ou seja, valores discrepantes relativamente aos restantes. Estas observações não foram excluídas, pois mostram os valores extremos que o Volume afluente pode atingir.

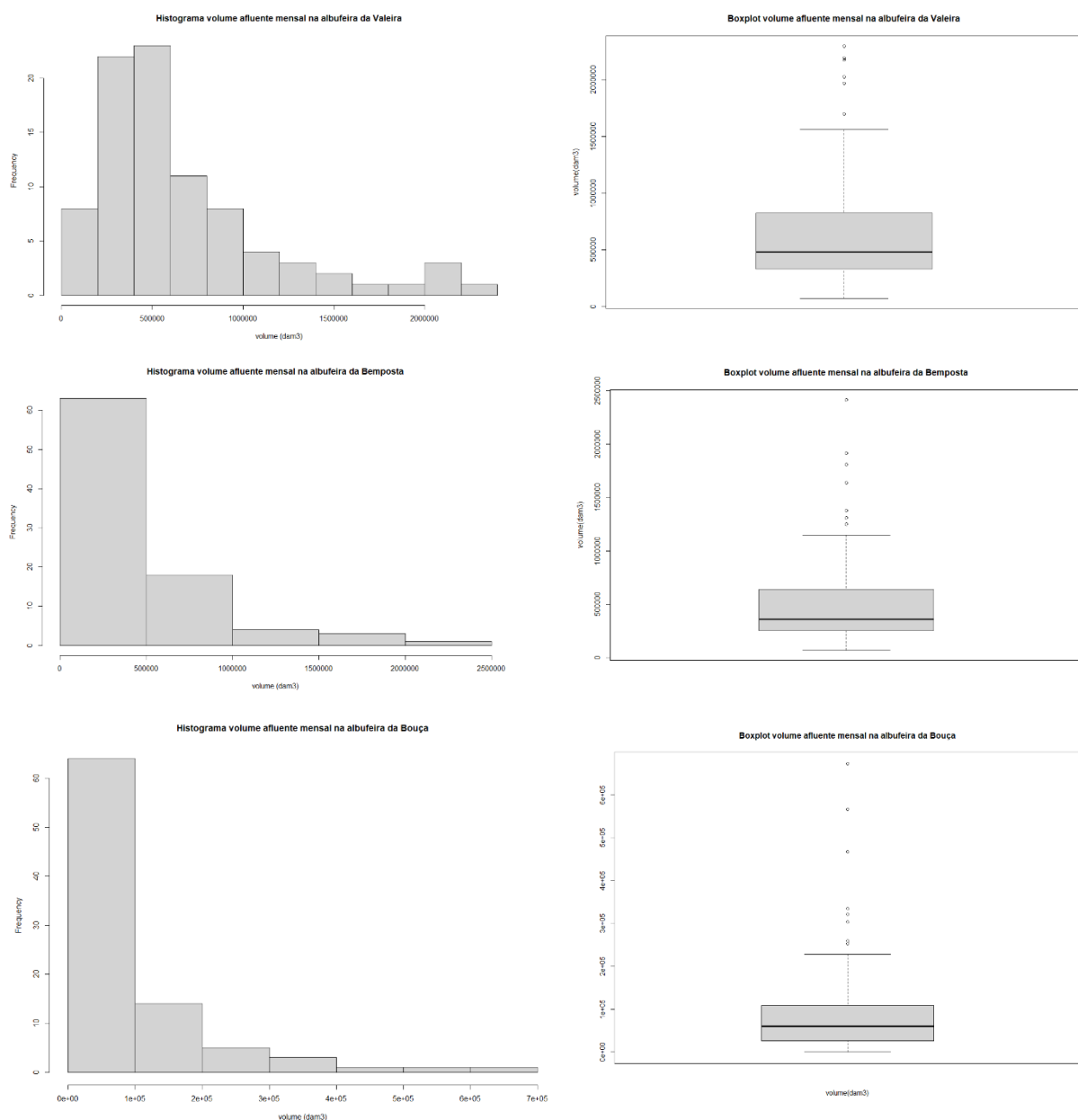


Figura 2: Histogramas e *Boxplots* das séries temporais das três albufeiras em estudo, no período observado.

5. Albufeira da valeira

Nesta Secção descreve-se, a título de exemplo, o processo de modelação da série temporal da albufeira da Valeira. Os principais resultados dos processos de modelação para as séries da albufeira da Bemposta e da Bouçã serão apresentados posteriormente (Anexo).

5.a. Modelação SARIMA

O primeiro passo para a identificação de um modelo SARIMA passa pela estacionarização da série em estudo. Desta forma, depois da estabilização da variância, através da transformação de *Box-Cox* ($\lambda=0.05$), definimos a ordem de diferenciação para a estabilização da média

Foi aplicada uma diferenciação de 1ª ordem ($d=1$) porém concluímos que neste caso não seria benéfico essa diferenciação tendo então optado por $d=0$.

Depois de se transformar a série original numa série estacionária em média e em variância, o próximo passo é detetar a sazonalidade do modelo, através da análise do periodograma. Este determinou s igual a 3.75, porém, a periodicidade aparenta ser anual, e, portanto, considera-se $s = 12$. Para determinar a ordem de D foi definido um modelo com $P=1$ e analisado o valor do coeficiente \hat{v}_1 , como este não se encontra próximo de 1 (0.2355), concluímos que $D=0$. A identificação das restantes ordens da parte sazonal (P e Q) é explorada através da combinação das várias possibilidades, fazendo variar P e Q entre 0 e 1.

De acordo com os resultados obtidos (Tabela 2), o modelo que resulta no menor AIC é o que considera $P=1$ e $Q = 1$ e, portanto, estas são as ordens escolhidas para a parte sazonal do modelo (juntamente com $D=0$ referido anteriormente).

Tabela 2: Ajustamento de vários modelos para a parte sazonal, após escolha da ordem de diferenciação regular, à série transformada da albufeira da Valeira (a tabela apresentada é uma tabela simplificada, foram realizados mais modelos).

MODELO	\hat{v}_1	$\hat{\eta}_1$	AIC
SARIMA(0,0,0)(1,0,0)₁₂	0.2355	-	314.4
SARIMA(0,0,0)(0,0,1)₁₂	-	0.5901	310.87
SARIMA(0,0,0)(1,0,1)₁₂	-0.3767	0.9999	306.47

Por fim, resta identificar as ordens p e q , que, tal como P e Q , são determinadas através da tentativa de diversas combinações.

O modelo que obteve menor AIC (Tabela 3) e com coeficientes significativos (Tabela 5) foi o modelo com as ordens $p=1$ e $q=0$ (Tabela 4).

Assim, o modelo escolhido é **SARIMA(1,0,0)(1,0,1)₁₂**.

Tabela 3: Ajustamento de vários modelos para a parte regular, após escolha da ordem de diferenciação regular e das ordens da parte sazonal, para a série transformada da albufeira da Valeira (a tabela apresentada é uma tabela simplificada, foram realizados mais modelos).

MODELO	$\hat{\Phi}_1$	$\hat{\Phi}_2$	$\hat{\Theta}_1$	$\hat{\nu}_1$	$\hat{\eta}_1$	AIC
SARIMA(1,0,0)(1,0,1)₁₂	0.7268	-	-	0.9952	-0.9477	244.95
SARIMA(0,0,1)(1,0,1)₁₂	-	-	0.7546	-0.5333	1.0000	255.6
SARIMA(1,0,1)(1,0,1)₁₂	0.5717	-	0.2907	-0.6141	0.9999	246.68
SARIMA(2,0,1)(1,0,1)₁₂	-0.0446	0.4687	0.9153	-0.5719	0.9999	244.88

Tabela 4: Resultados da estimação do modelo SARIMA para a série transformada da albufeira da Valeira

MODELO FINAL:	SARIMA(1,0,0)(1,0,1)₁₂	AIC=244.95	$\hat{\sigma} = 0.85337$
	$\hat{\Phi}_1$	$\hat{\nu}_1$	$\hat{\eta}_1$
ESTIMATIVA	0.7268	0.9952	-0.9477
ERRO PADRÃO	0.0711	0.0347	0.1883

Tabela 5: Intervalo de confiança dos coeficientes do modelo SARIMA para a série transformada da albufeira da Valeira.

I.C (95%)	LIM.INFERIOR	LIM.SUPERIOR
$\hat{\Phi}_1$	0.5874742	0.8661153
$\hat{\nu}_1$	0.9272565	1.0631615
$\hat{\eta}_1$	-1.3166952	-0.5786298

5.b. Análise dos resíduos

Para validar o modelo escolhido é necessário realizar uma análise dos resíduos. Idealmente, estes devem apresentar um comportamento que se aproxima ao de um ruído branco. Ou seja, é esperado que estejam garantidas, além da média nula e variância constante, a normalidade e a independência dos erros.

O histograma da Figura 3 sugere que os resíduos têm uma distribuição muito ligeiramente assimétrica à direita, mas Normal, o que é confirmado pelo teste de *Kolmogorov-Smirnov* que não rejeita a hipótese de normalidade (valor de prova de 0,988). Além disso, de acordo com a representação gráfica da série dos resíduos (Figura 3), esta parece apresentar uma distribuição uniforme em torno do resíduo zero, o que leva a crer que os erros têm média nula e variância constante. De facto, o teste-t para o valor esperado confirma a não rejeição da hipótese de média nula, resultando num valor de prova de 0.8617. Assim, conclui-se que não há evidência estatística para rejeitar a hipótese de que os erros seguem uma distribuição Normal de média nula e variância constante. Quanto à independência, o teste de Box-Pierce foi aplicado à série dos resíduos, sendo que o valor de prova é 0.5016, ou seja, segundo os resultados do teste, a hipótese de independência não é rejeitada. Mais, as FAC e FACP estimadas dos resíduos (Figura 3) assemelham-se às FAC e FACP de um ruído branco e, portanto, pode admitir-se a independência dos

erros. Conclui-se então que os resíduos apresentam um comportamento semelhante ao de um ruído branco.

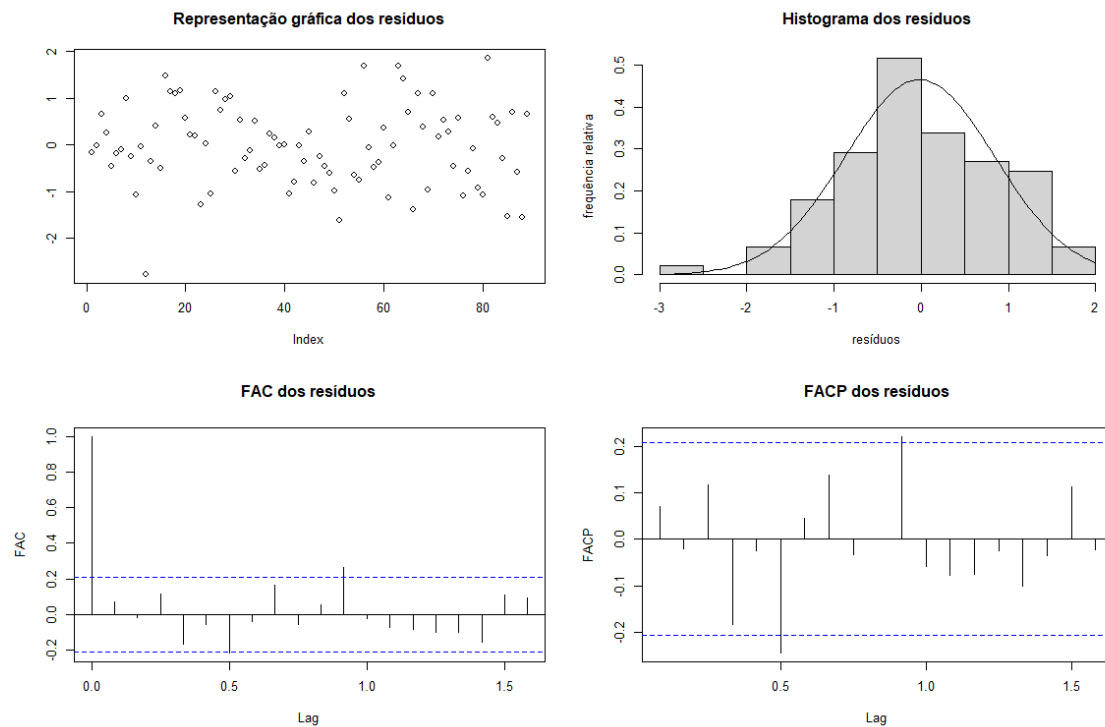


Figura 3: Série dos resíduos para a série transformada da albufeira da Valeira, após ajustamento do modelo SARIMA, e respetivo histograma, FAC e FACP estimadas.

5.c. Previsão SARIMA

Antes de realizar a previsão calculamos as estimativas pontuais as quais representamos graficamente na Figura 4.

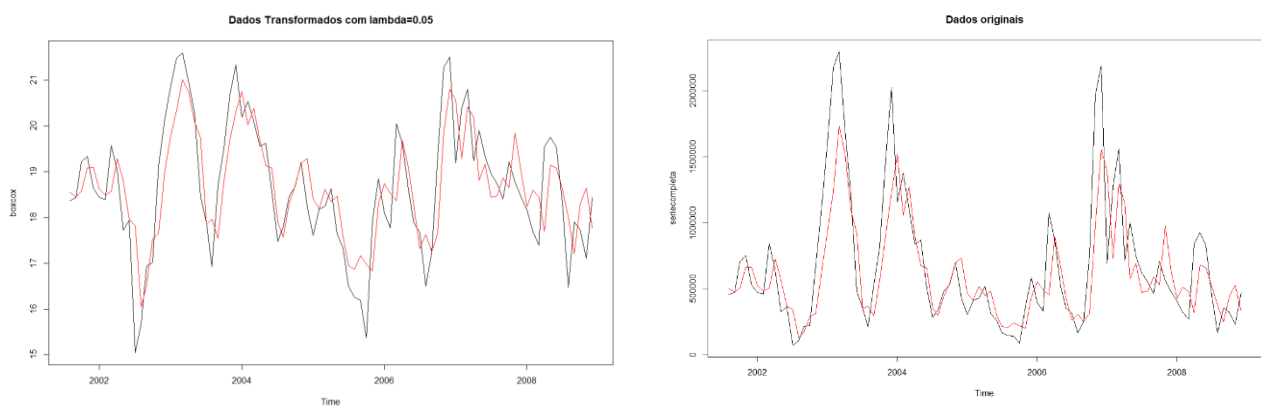


Figura 4: Estimativas pontuais obtidas através do modelo SARIMA, sobrepostas à série da albufeira da Valeira transformada (esquerda) e original (direita).

Na Figura 5 encontram-se representadas as previsões (no período de teste, i.e., de janeiro de 2007 a dezembro de 2008), pontuais e intervalares, obtidas através do modelo final, nas unidades originais, sobrepostas à série em estudo.

A Figura 5 sugere que a qualidade preditiva do modelo é melhor na série de treino do que na série de teste.

Relativamente aos intervalos de previsão, pode dizer-se que a sua taxa de cobertura é, neste caso, de 100%, uma vez que todas as 24 observações da série de teste pertencem ao interior dos mesmos.

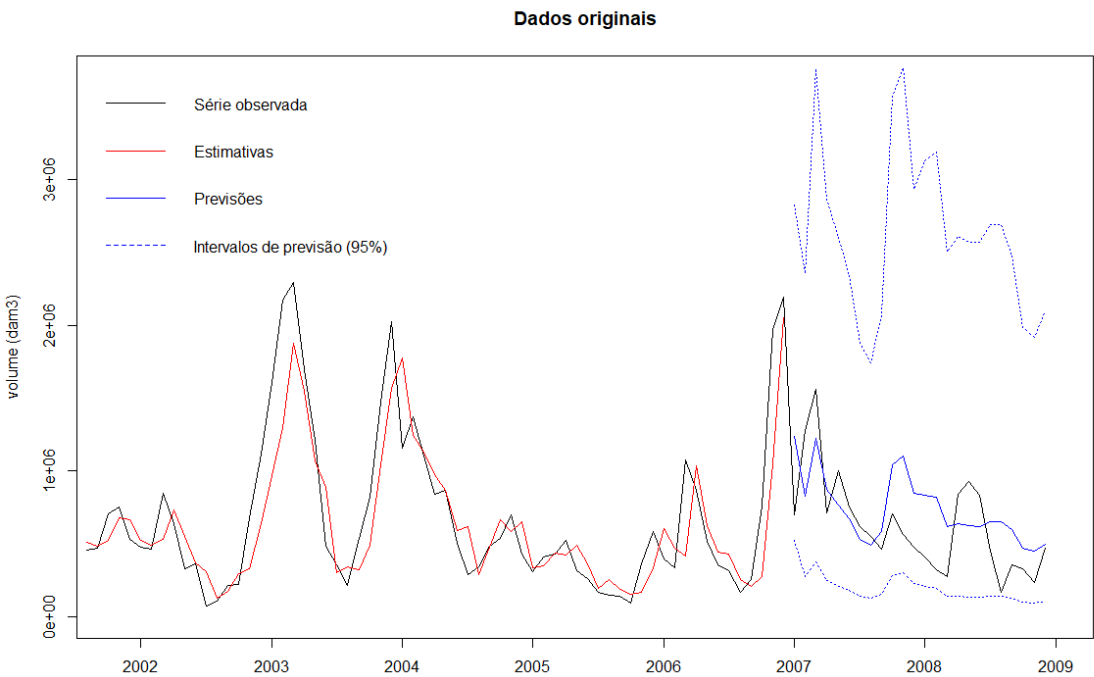


Figura 5: Previsões (no período de teste), pontuais e intervalares (95%), e estimativas pontuais (no período de agosto de 2001 a dezembro de 2006) obtidas através do modelo SARIMA, sobrepostas à série da albufeira da Valeira.

Tabela 6: Previsões pontuais (no período de teste) obtidas através do modelo SARIMA.

PREVIS ÕES	JANEI RO	FEVER EIRO	MAR ÇO	ABR IL	MA IO	JUN HO	JUL HO	AGOS TO	SETEM BRO	OUTU BRO	NOVEM BRO	DEZEM BRO
2007	123583 1	826697	12232 5	8708 31	7646 79	6724 23	5279 06	48738 7	584188	104162 3	1102905	843590
2008	835262	817875	61710 5	6378 19	6228 13	6190 66	6508 76	65038 3	593976	467304	451032	496094

Os dados utilizados para este estudo possuíam observações até dezembro de 2022, porém grande parte dos valores pós 2008 foram calculados pelo SNIRH. Um dos nossos objetivos para este trabalho era fazer previsões para todos os anos depois de 2008 e compará-los com os valores calculados pelo SNIRH. Porém os valores e gráficos obtidos não foram satisfatórios e como tal serão apenas apresentados no Anexo.

Conclusão

Este trabalho foi dedicado à análise de séries temporais de variáveis ambientais, nomeadamente o volume afluente, observado em albufeiras da bacia hidrográfica do rio Douro e Tejo.

Obtivemos dados ambientais através do Sistema Nacional de Informação de Recursos Hídricos, tendo posteriormente identificado o modelo SARIMA que melhor se ajustou a estes.

Seguidamente dividimos a nossa série em série de treino e de teste e utilizamos o modelo obtido para obter previsões pontuais e intervalares para a série de teste.

O nosso modelo mostrou-se eficiente, tendo apresentado excelentes estimativas pontuais, boas previsões pontuais para a série de teste e uma taxa de cobertura de 100%.

ANEXO

Albufeira da Valeira

Previsão 2009-2021

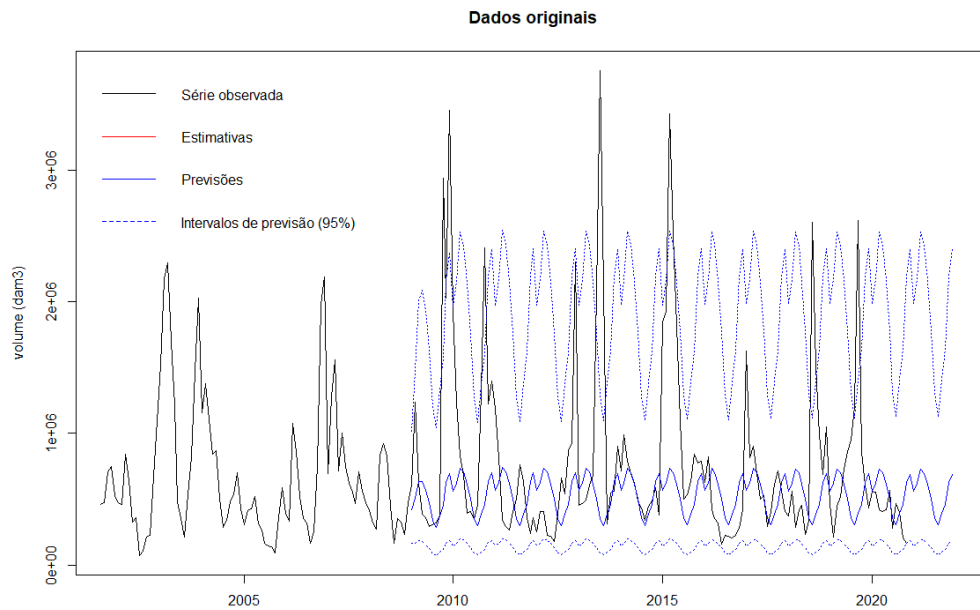


Figura A.1: Previsões pontuais e intervalares (95%) (no período de janeiro de 2009 a dezembro de 2021) obtidas através do modelo SARIMA, sobrepostas à série da albufeira da Valeira.

Albufeira de Bemposta

Análise dos resíduos

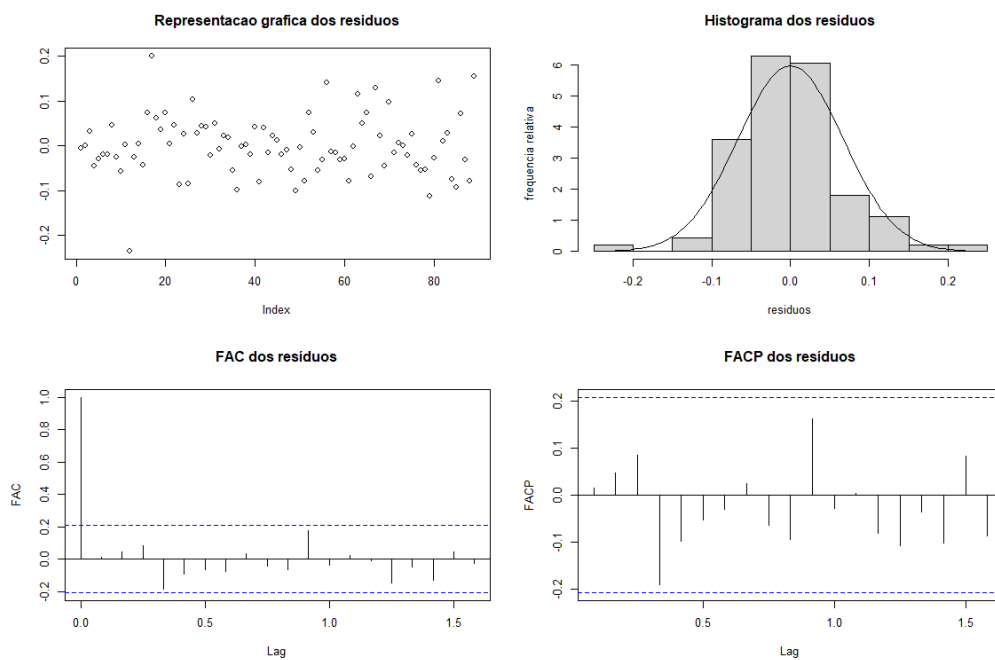


Figura A.2: Série dos resíduos para a série transformada da albufeira da Bemposta, após ajustamento do modelo SARIMA, e respetivo histograma, FAC e FACP estimadas.

Previsão SARIMA

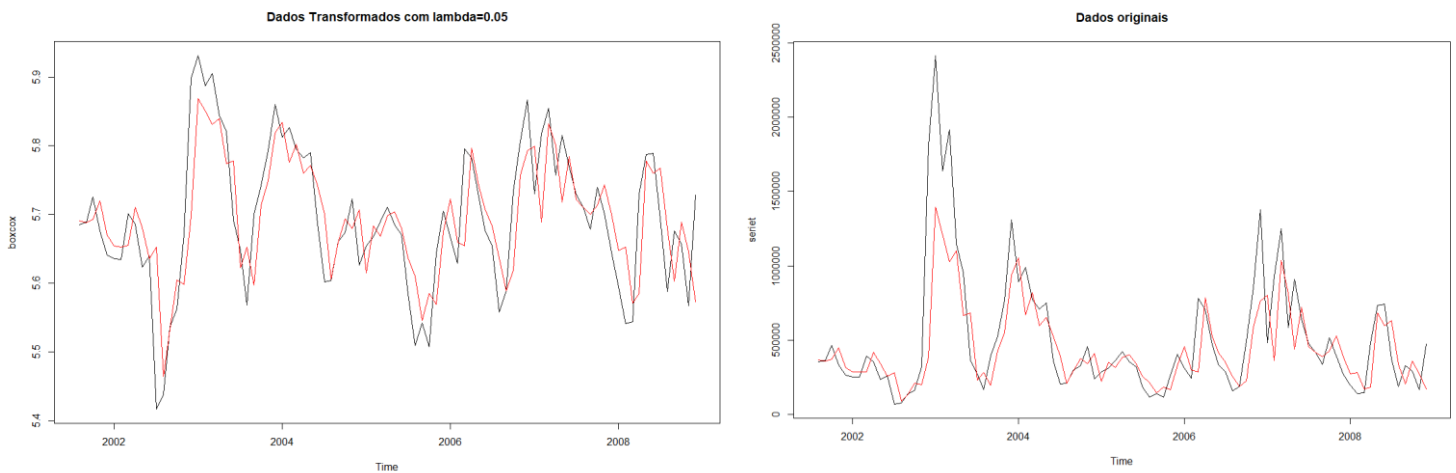


Figura A.3: Estimativas pontuais obtidas através do modelo SARIMA, sobrepostas à série da albufeira da Bemposta transformada (esquerda) e original (direita).

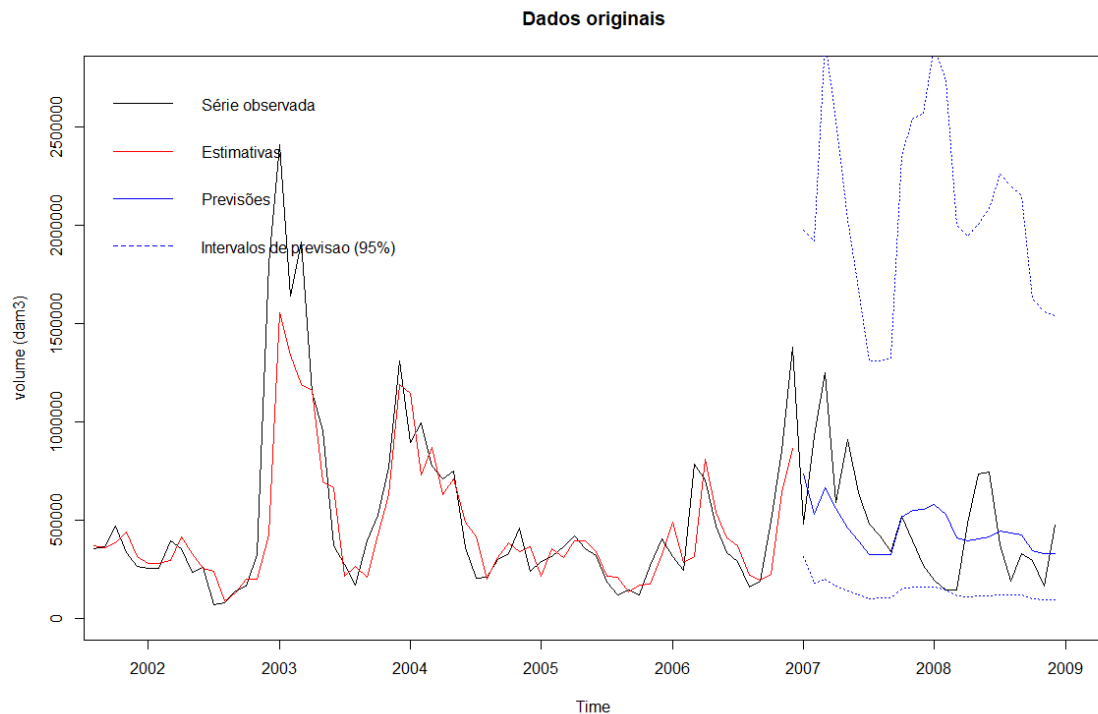


Figura A.4: Previsões (no período de teste), pontuais e intervalares (95%), e estimativas pontuais (no período de agosto de 2001 a dezembro de 2006) obtidas através do modelo SARIMA, sobrepostas à série da albufeira da Bemposta.

Albufeira de Bouçã

Análise dos resíduos

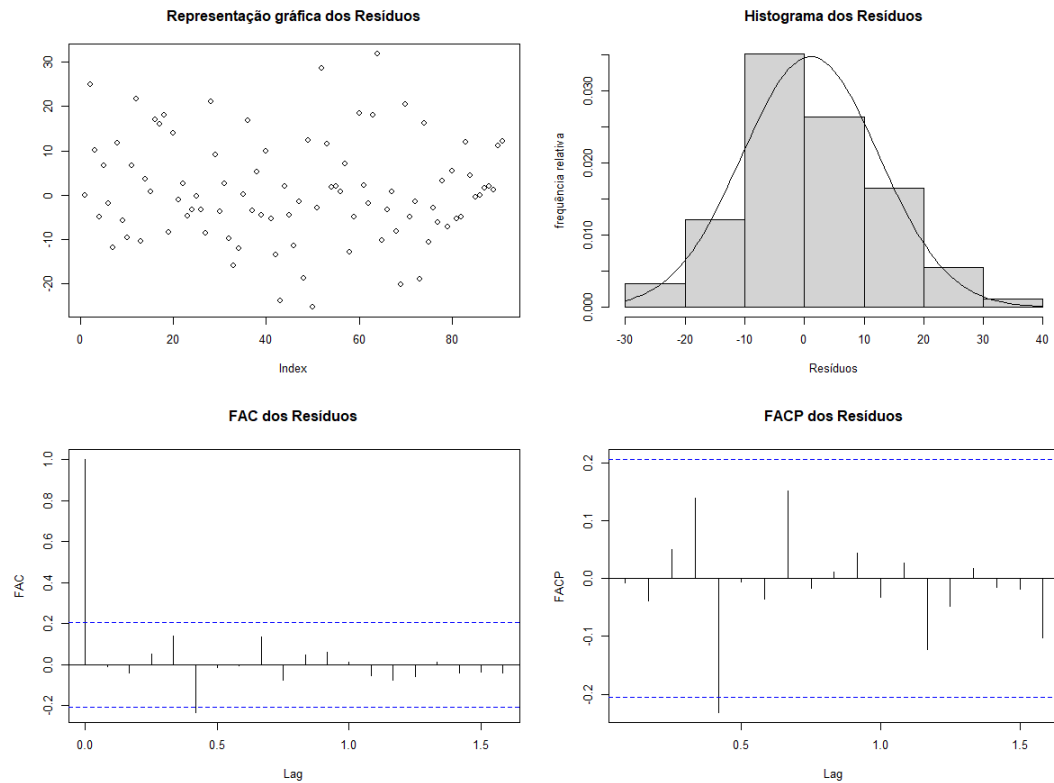


Figura A.5: Série dos resíduos para a série transformada da albufeira da Bouçã, após ajustamento do modelo SARIMA, e respetivo histograma, FAC e FACP estimadas.

Previsão SARIMA

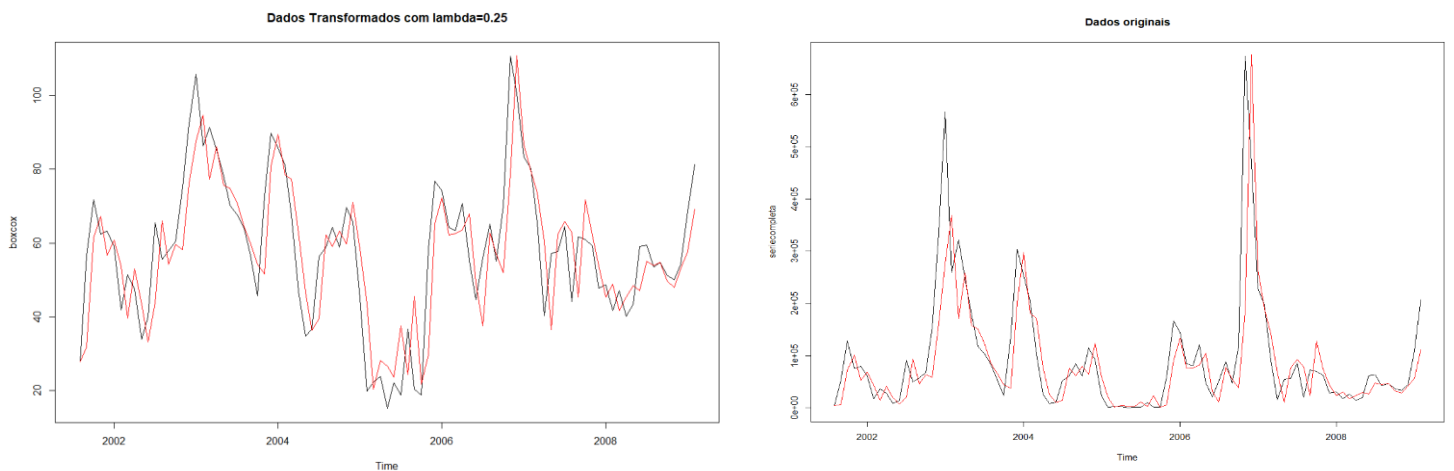


Figura A.6: Estimativas pontuais obtidas através do modelo SARIMA, sobrepostas à série da albufeira da Bouçã transformada (esquerda) e original (direita).

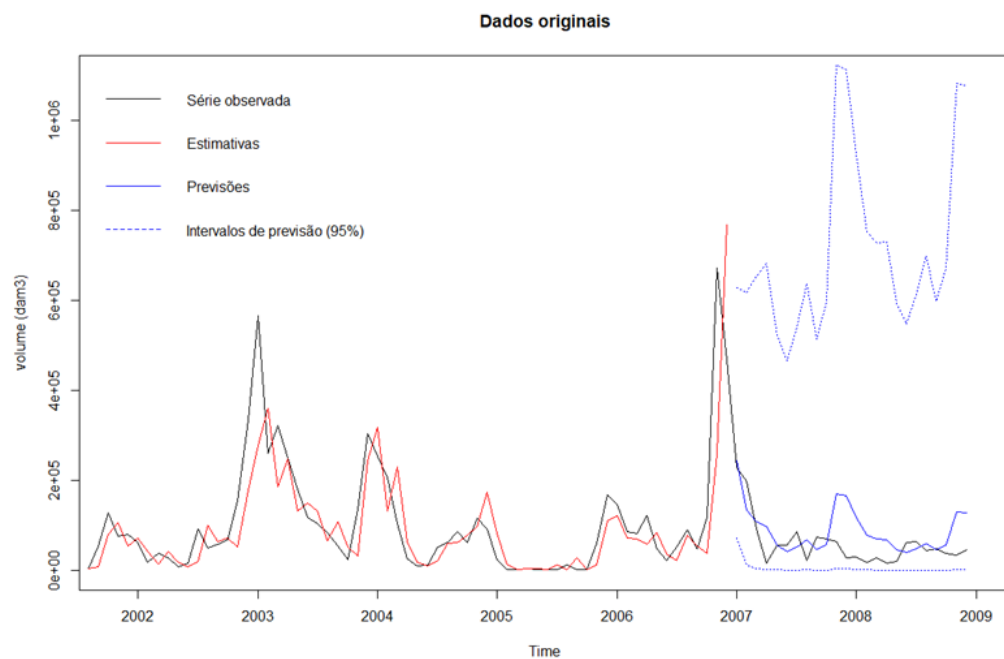


Figura A.7: Previsões (no período de teste), pontuais e intervalares (95%), e estimativas pontuais (no período de agosto de 2001 a dezembro de 2006) obtidas através do modelo SARIMA, sobrepostas à série da albufeira da Bouça.