

CARACTERIZAÇÃO, MODELAÇÃO E PREVISÃO DE UTILIZAÇÕES DE ÁGUA UTILIZANDO FERRAMENTAS ESTATÍSTICAS

Júlio F. FERREIRA da SILVA

*Doutor em Engenharia Civil – Hidráulica Professor Auxiliar do Departamento de Eng^a Civil da Universidade do Minho,
Azurém 4800-058 Guimarães, Portugal
253510200, juliofs@civil.uminho.pt*

Marta Alexandra Quintã PINTO
Licenciada em Engenharia Civil

RESUMO

Neste trabalho são apresentadas e aplicadas ferramentas estatísticas que podem ser usadas para a caracterização, modelação e previsão de utilizações de água. Os dados históricos são agrupados em quatro tipos característicos: os dias úteis, os sábados, os domingos e os feriados. Faz-se a simulação da aplicação destes modelos à previsão em tempo real. Este tipo de ferramenta apresenta grande facilidade na sua actualização com a incorporação dos registos entretanto ocorridos e, portanto, adaptam-se às eventuais variações. Recorrendo aos resultados destes modelos é possível enviar mensagens de alerta quando as previsões se afastam para além da tolerância admitida, e assim, contribuir para a redução das perdas de água nos sistemas de abastecimento e distribuição.

Palavras-chave: Modelação das utilizações de água, Previsão e Gestão em tempo real.

1 - INTRODUÇÃO

A gestão racional e otimizada da água constitui na actualidade um dos temas de grande importância a nível mundial. As ferramentas que permitem a caracterização, modelação e previsão das utilizações de água constituem uma componente indispensável para a adequada concepção e gestão dos sistemas de utilização da água. A caracterização quantitativa das utilizações de água possibilita um melhor conhecimento dos elementos base de concepção e dimensionamento, bem como de informação necessária à definição de estratégias de desenvolvimento a longo prazo dos sistemas de abastecimento de água e de drenagem de águas residuais. Os programas de controlo dos sistemas de abastecimento de água necessitam de uma adequada gestão da informação, nomeadamente das utilizações. Neste âmbito, é fundamental o conhecimento dos dados históricos, devidamente tratados, e de modelos que sejam capazes de se ajustar às séries cronológicas e de antecipar as necessidades futuras.

O estudo e previsão das utilizações de água têm como principal aplicação a gestão dos sistemas de abastecimento de água, designadamente, para melhorar a garantia da continuidade do serviço e para que as tarefas de exploração sejam executadas da forma mais económica possível. A indústria da água para fazer face aos novos desafios relacionados com a adequada gestão dos sistemas de utilização da água deve dar maior atenção à programação das operações de exploração, onde os estudos para a caracterização e previsão das utilizações de água são fundamentais, por exemplo, na definição dos volumes de água perdidos e no desenvolvimento das respectivas ferramentas de detecção e controlo, na definição de estratégias para a minimização dos incómodos motivados por obras de reparação e, sobretudo, na definição de regras de funcionamento das estações elevatórias para se evitar as horas de energia eléctrica mais caras.

Assim, os principais objectivos deste trabalho são: 1- Caracterização dos comportamentos típicos das utilizações de água em função dos dias tipo (dias de trabalho, Sábados, Domingos e feriados) recorrendo ao conceito de índice sazonal; 2 - Selecção de modelos estocásticos que melhor representem as utilizações; 3 - Análise e selecção de modelos estocásticos de previsão das utilizações de água e estudo do respectivo desempenho.

2 - CARACTERIZAÇÃO DOS CONSUMOS DIÁRIOS/HORÁRIOS

A caracterização das utilizações de água tem por objectivo descrever e quantificar a evolução das utilizações de água ao longo do tempo. Na análise da evolução das utilizações, pretende identificar-se os valores extremos, a tendência e a sazonalidade. As variações periódicas ocorrem ao longo dos meses do ano, dias da semana e dentro de cada dia ao longo das horas.

Os índices sazonais reflectem essas variações periódicas. Um comportamento anómalo, isto é, uma evolução das utilizações diferente daquela que é *característica* do instante em análise (uma fuga, por exemplo) é facilmente detectado possibilitando a sua correcção, o que contribui para a redução das perdas de água. Refira-se que é habitual os sistemas de distribuição de água perderem em fugas cerca de 20 a 30 % da água captada e tratada.

2.1 - Identificação de anomalias

A análise do cronograma das utilizações de água é uma tarefa elementar para se caracterizar a evolução padrão e se detectar os comportamentos que destoam do que é habitual. A observação dos registos de água relativos a períodos que apresentem valores superiores aos usuais ou aos valores dos mesmos instantes dos últimos períodos, podem revelar a existência de caudais perdidos numa ruptura.

Para tipificar a evolução das utilizações horárias, seja numa rede de distribuição, seja num consumidor importante, deve ser executada a tarefa preliminar relativa à remoção dos comportamentos

aberrantes, sob pena da modelação subsequente incluir erradamente a influência das avarias na rede como se tratando de uma componente do fenómeno.

2.2 - Identificação da sazonalidade

A componente sazonal ou estacional (E_t) pode, de acordo com MURTEIRA (1993), 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 cosenos. Uma alternativa, mais simples e muito utilizada para a definição de factores de ponta, consiste na utilização do método das médias (simples ou móveis). 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 nível da sucessão. Os factores sazonais correspondem a um coeficiente de ponderação do nível da sucessão de utilizações. Quando o período sazonal é S e para qualquer inteiro k

$$\sum_{t=k+1}^{k+S} E_t = S \quad (1)$$

Com $S = 2k$ o período sazonal par, define-se a média móvel de S termos bilateral e centrada

$$M_t = \frac{1}{S} \left(\frac{1}{2} C_{t-\frac{S}{2}} + C_{t-\frac{S}{2}+1} + \dots + C_t + \dots + C_{t-\frac{S}{2}-1} + \frac{1}{2} C_{t+\frac{S}{2}} \right) \quad t = \frac{S}{2} + 1, \dots, N - \frac{S}{2} \quad (2)$$

Com t o instante médio das observações incluídas na média móvel e C_t a utilização de água no instante t .

Calculando-se os valores

$$E_j^* = \frac{1}{N-1} \sum_{i=2}^N \left[\frac{C_{S(i-1)+j}}{M_{S(i-1)+j}} \right] \quad j = 1, 2, \dots, k \quad k = S/2 \quad (3)$$

$$E_j^* = \frac{1}{N-1} \sum_{i=1}^{N-1} \left[\frac{C_{S(i-1)+j}}{M_{S(i-1)+j}} \right] \quad j = k+1, k+2, \dots, S \quad (4)$$

obtém-se a estimativa da componente sazonal

$$E_j = E_j^* \left[\frac{S}{\sum_{i=1}^S E_i^*} \right] \quad j = 1, 2, \dots, S \quad (5)$$

3 - MODELAÇÃO E PREVISÃO ESTOCÁSTICA USANDO MODELOS ARIMA

De acordo com MURTEIRA (1993) os modelos estocásticos caracterizam-se pela sua flexibilidade, já que têm a capacidade de responder às variações da sucessão cronológica de utilizações com pesos diferentes para os valores registados em instantes diferentes.

As utilizações de água são um fenómeno dinâmico, fortemente influenciado por variáveis de ocorrência aleatória, como sejam a temperatura e a precipitação, e para a sua modelação será de experimentar modelos probabilísticos, isto é modelos que incluam o tratamento residual dos impulsos ou choques aleatórios já que tais impulsos podem não se esgotar logo que ocorrem mas incorporar-se na futura evolução.

3.1 - Descrição geral dos processos ARIMA

Para tentar explicar a relativa complexidade das utilizações de água, neste caso, dá-se importância aos modelos que reconhecem a existência de variações que não conseguem caracterizar-se por completo, o que dá origem a um erro ou resíduo, ε_t .

Dispondo como dados uma sucessão cronológica de utilizações com observações até ao instante t , $\{C_t, C_{t-1}, C_{t-2}, \dots\}$ é nosso objectivo caracterizar o valor no instante t , C_t que se considera ser função dos valores históricos, ou seja,

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

Um processo estacionário e invertível pode representar-se na forma autoregressiva ou na forma médias móveis, no entanto os processos mistos (autoregressivo e médias móveis) ARMA (p, q) surgem como modelos parcimoniosos, isto é, com poucos parâmetros, o que é uma característica a procurar. Estes processos representados na forma de equação às diferenças permitem apresentar a expressão geral das utilizações

$$C_t = \phi_1 C_{t-1} + \dots + \phi_p C_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (7)$$

onde ϕ é um número real

A equação anterior pode ser escrita de uma forma compacta:

$$\phi_p(B)C_t = \theta_q(B)\varepsilon_t \quad (8)$$

$\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ é o polinómio autoregressivo estacionário

$\theta_q(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ é o polinómio médias móveis invertível

As utilizações de água apresentam, muitas vezes, uma evolução ao longo do tempo e para que se possam transformar num processo estacionário é necessário proceder ao cálculo de diferenças, isto é, se C_t for um processo não estacionário depois de diferenciado d vezes transformar-se-á num processo estacionário e invertível.

Um processo C_t será um ARIMA (p, d, q) se

$$\phi_p(B)(1-B)^d C_t = \theta_q(B)\varepsilon_t \quad (9)$$

As utilizações de água ao longo dos meses, das semanas e das horas dos diversos dias apresentam sempre uma componente sazonal ou periódica de período S , identificada pela correlação significativa entre os valores registados em intervalos de tempo múltiplos de S .

Os processos ARIMA (p, d, q) são muito úteis para modelar a relação entre observações vizinhas da sucessão, $C_{t-p}, C_{t-p+1}, \dots, C_t$ e observações vizinhas do ruído branco, $\varepsilon_{t-q}, \varepsilon_{t-q+1}, \dots, \varepsilon_t$, mas quando a componente cíclica com período S está, também, presente, é de esperar que $C_{t-2S}, C_{t-S}, \dots, C_t$ e $\varepsilon_{t-2S}, \varepsilon_{t-S}, \dots, \varepsilon_t$ estejam relacionados e os modelos estendem-se a:

$$\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D C_t = \theta_q(B)\Theta_Q(B^S)\varepsilon_t \quad (10)$$

$\Phi_P(B^S) = 1 - \Phi_1 B^S - \dots - \Phi_P B^{PS}$ é o polinómio autoregressivo sazonal de grau P em B^S

$\Theta_Q(B^S) = 1 - \Theta_1 B^S - \dots - \Theta_Q B^{QS}$ é o polinómio médias móveis sazonal de grau Q em B^S

O modelo anterior é designado abreviadamente por SARIMA $(p, d, q)(P, D, Q)_s$

3.2 - Modelação ARIMA e previsão das utilizações de água

Box e Jenkins (1970) apresentaram uma metodologia de modelação de sucessões cronológicas com as componentes: identificação, estimação e avaliação do diagnóstico.

O esquema geral para a modelação e previsão estocástica está representado na figura seguinte:

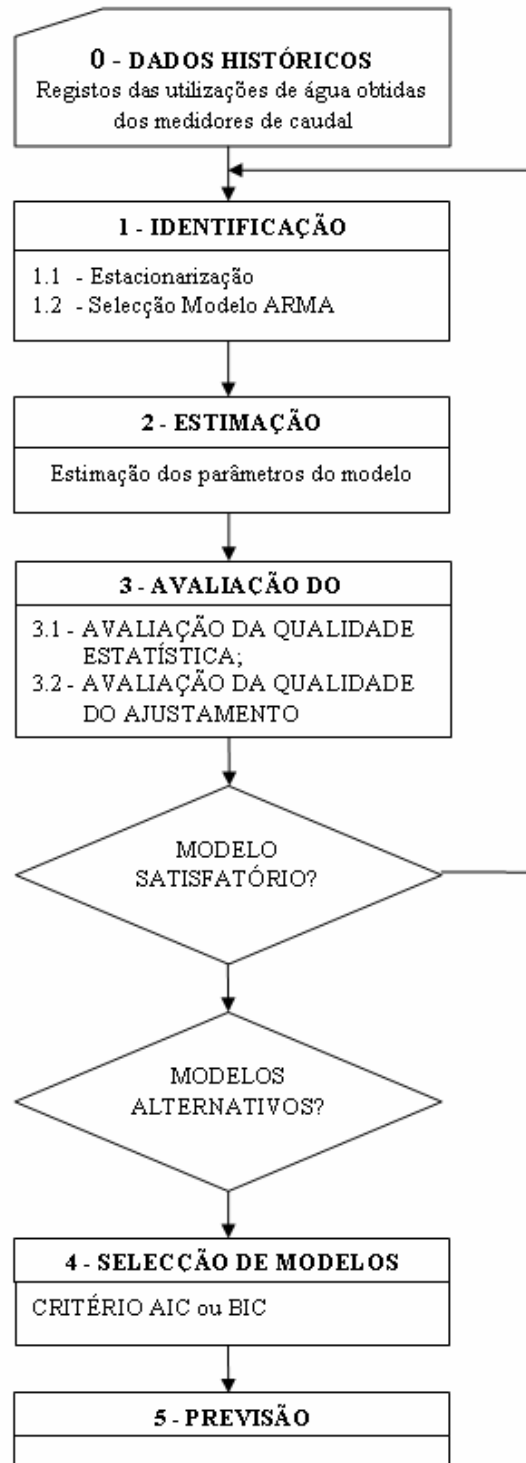


Figura 1 - Modelação ARIMA

Na primeira etapa, a identificação, o objectivo é procurar um modelo que descreva a sucessão cronológica e decompõe-se nas duas tarefas prévias: a estacionarização e a selecção de um modelo ARMA. Como os modelos ARMA dizem respeito a processos estacionários e dado que, em geral, as sucessões cronológicas das utilizações de água não se encontram nestas condições, a primeira subetapa consiste em actuar sobre os dados históricos com um conjunto de transformações de modo a estabilizar a variância, a neutralizar a tendência e a eliminar movimentos de carácter estritamente periódico. A segunda subetapa consiste na comparação das Funções de Autocorrelação (FAC) e

Funções de Autocorrelação Parcial (FACP) estimadas com o aspecto teórico destas funções correspondentes aos diversos modelos ARMA.

Na etapa de estimação são definidos os parâmetros $(\phi_1, \phi_2, \dots, \phi_p)$, $(\theta_1, \theta_2, \dots, \theta_q)$, $\sigma_\varepsilon^2 = \mathcal{V}\{\varepsilon_t\}$, bem como os relativos à componente sazonal.

Na terceira fase, a avaliação do diagnóstico, verifica-se a adequação do modelo já identificado e estimado e devem seguir-se as duas subetapas: controlo da qualidade estatística do modelo e da qualidade do ajustamento.

A quarta fase será a procura de modelos alternativos, eventualmente vizinhos do primeiro modelo seleccionado, já que muitas vezes existem alguns modelos a descrever a sucessão cronológica com igual elevado grau de qualidade.

Para a escolha do modelo a adoptar haverá o recurso a critérios como os apresentados por Akaike (AIC) ou a extensão Bayesiana (BIC).

Tendo por objectivo 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.

4 - APLICAÇÕES

Neste capítulo serão aplicados os fundamentos descritos anteriormente aos dados recolhidos num reservatório na freguesia de Urgezes em Guimarães. Este estudo engloba a caracterização quantitativa e o estudo de modelos representativos da sucessão das utilizações registadas.

4.1 - Identificação de comportamentos anómalos (fugas)

A observação do cronograma das utilizações registadas da figura 1, concretamente dos caudais registados na noite do dia 24.11.93, e na madrugada do dia seguinte, sábado 25.11.93, revela a existência de caudais superiores aos valores habituais naquelas horas e que resultam do agravamento motivado pelos caudais desperdiçados numa ruptura.

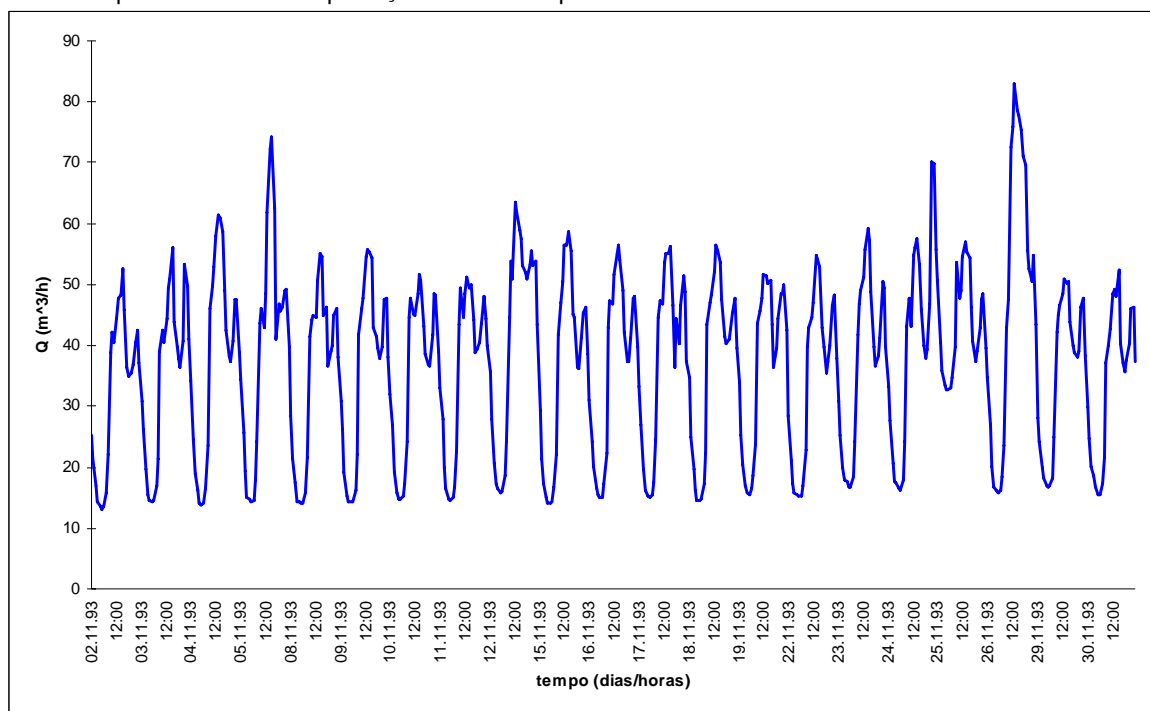


Figura 2 - Cronograma das utilizações de água registadas

4.2 - Caracterização da sazonalidade

Para ilustrar a metodologia de caracterização da sazonalidade através de índices, seguir-se-ão os registos das utilizações representadas na figura anterior. Na figura seguinte podem verificar-se os índices sazonais calculados pelo software SPSS (SPSS (2005)) relativos aos dias de trabalho, aos Sábados e aos Domingos. Note-se que os índices são calculados e referem-se à relação média entre a utilização num dado instante e a média do tipo de dia.

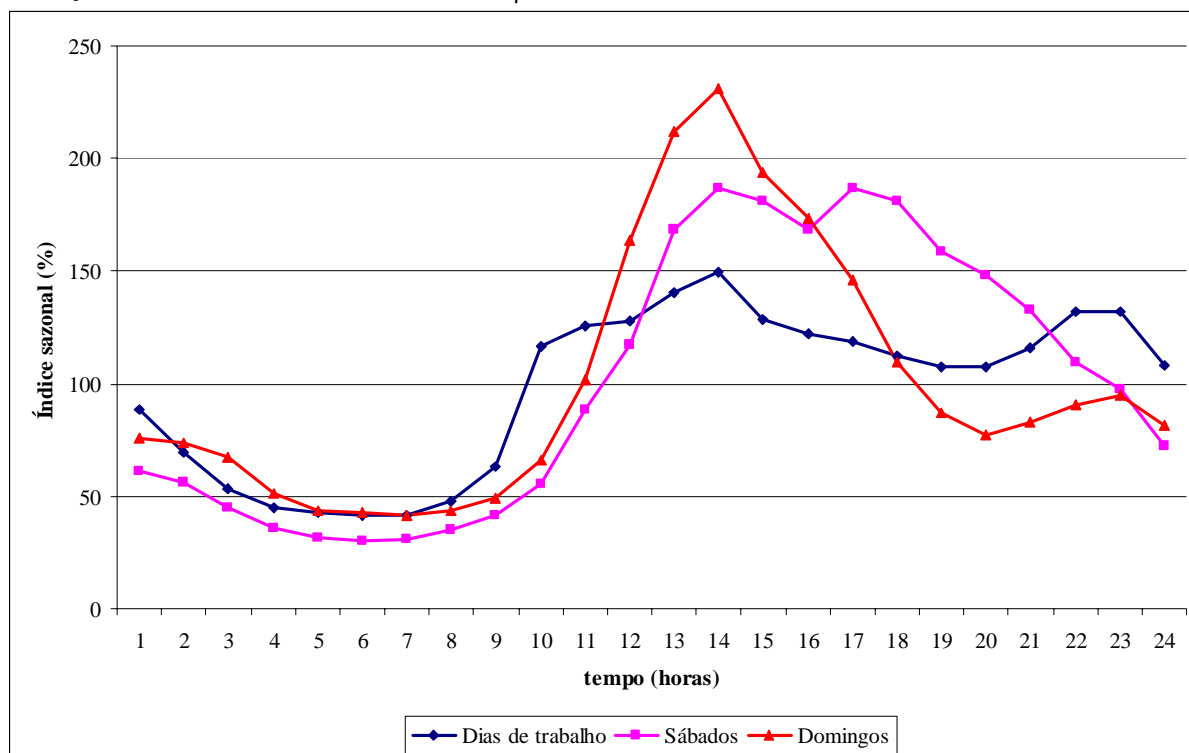


Figura 3 – Índices sazonais das utilizações de água registadas

4.3 - Modelação e previsão horária usando os processos ARIMA

O conhecimento do comportamento das funções FAC e FACP é fundamental para uma boa selecção do modelo.

A representação gráfica das funções FAC e FACP encontram-se na figura seguinte:

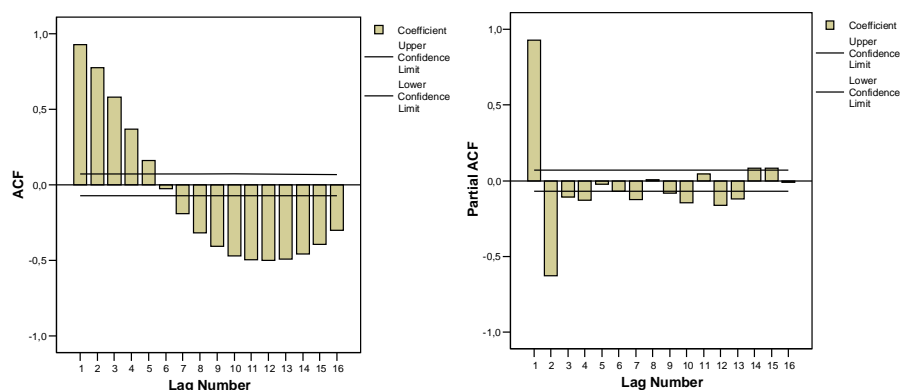


Figura 4 - Funções FAC e FACP das utilizações de água registadas

Na análise do comportamento da FAC estimada é de notar o decaimento para zero sob a forma exponencial. A análise da FACP mostra um decaimento brusco para zero logo no primeiro *lag* (valor significativo). Estes factos sugerem um processo misto, autoregressivo e de médias móveis.

Na óptica do utilizador dum pacote de software a etapa Estimação não requer, em geral, grande intervenção por parte do analista, uma vez identificado o modelo, as estimativas dos respectivos parâmetros são feitas automaticamente recorrendo a um dos métodos: máxima verosimilhança, mínimos quadrados e momentos.

A avaliação dos modelos alternativos é 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.

Para os valores dos caudais registados foram estudados os vários modelos

Quadro 1 - Selecção do modelo para as utilizações de água

Modelo	Coefficiente de correlação	AIC
(1,0,0)(0,0,0)	0,928579056	5404,4101
(0,0,1)(0,0,0)	0,914975872	6096,1301
(1,0,1)(0,0,0)	0,949851767	5135,3509
(1,0,1)(1,0,0)	0,964977038	4864,5880
(1,0,1)(1,0,1)	0,971051764	4708,4638

Assim, o “melhor modelo” será o SARIMA (1,0,1) (1,0,1). Na figura seguinte representam-se os dados históricos e os valores do modelo seleccionado.

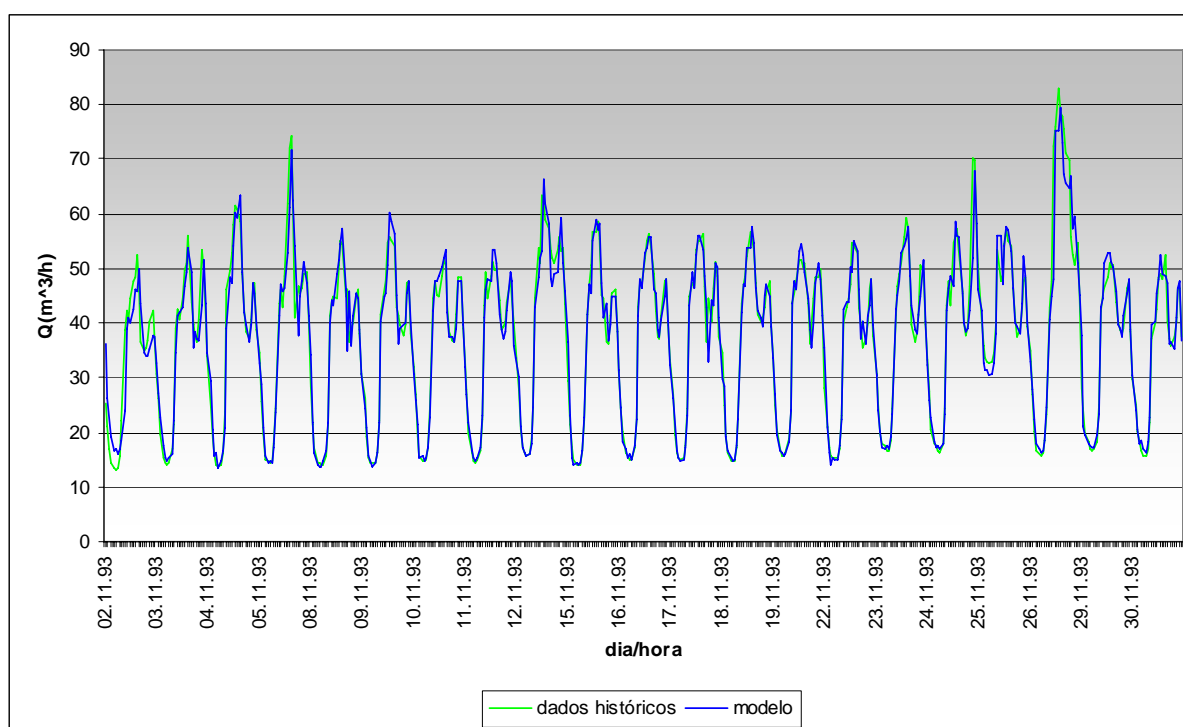


Figura 5 – Modelo ARIMA (1,0,1)(1,0,1)

Uma vez validado o modelo, este pode ser usado para a simulação de previsões em tempo real. Nas figuras seguintes estão representados os valores das previsões hora a hora, incluindo no processo as medições que entretanto vão estando disponíveis.

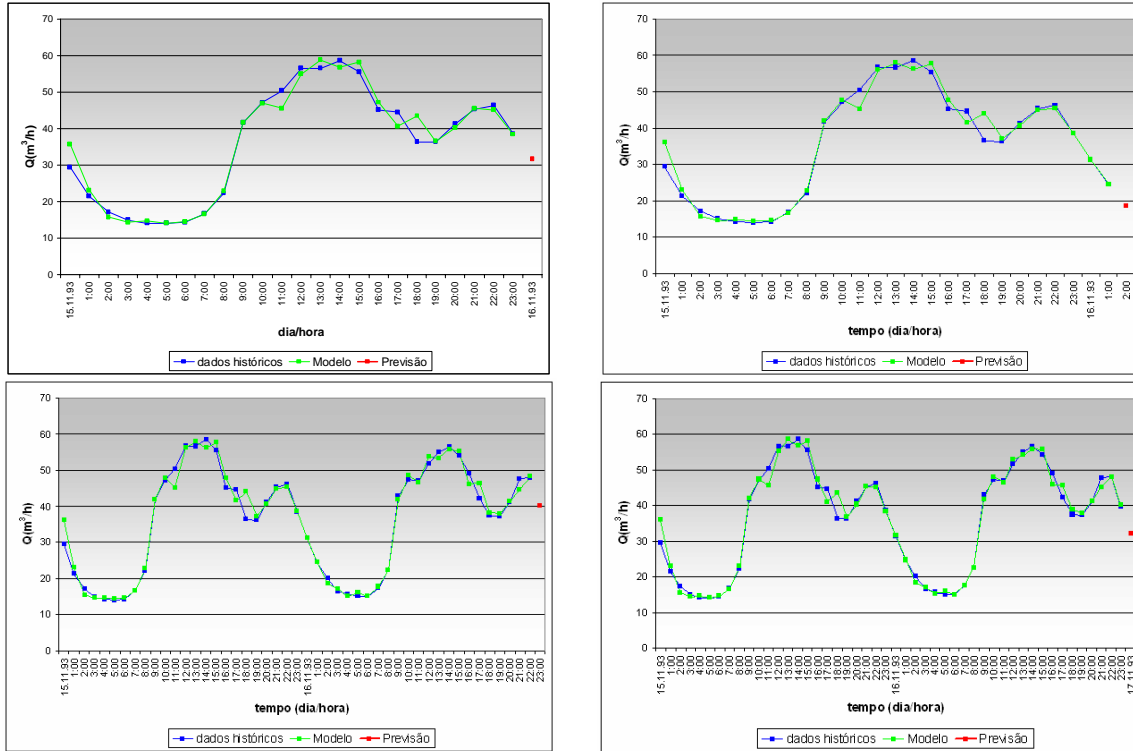


Figura 6 - Previsão para as primeiras e últimas horas do dia de trabalho 16 de Novembro

Na figura seguinte comparam-se os valores das previsões hora a hora com os registos dos caudais ocorridos no dia 16 de Novembro. O coeficiente de correlação entre os dois conjuntos de valores vale $r=0,995554$.

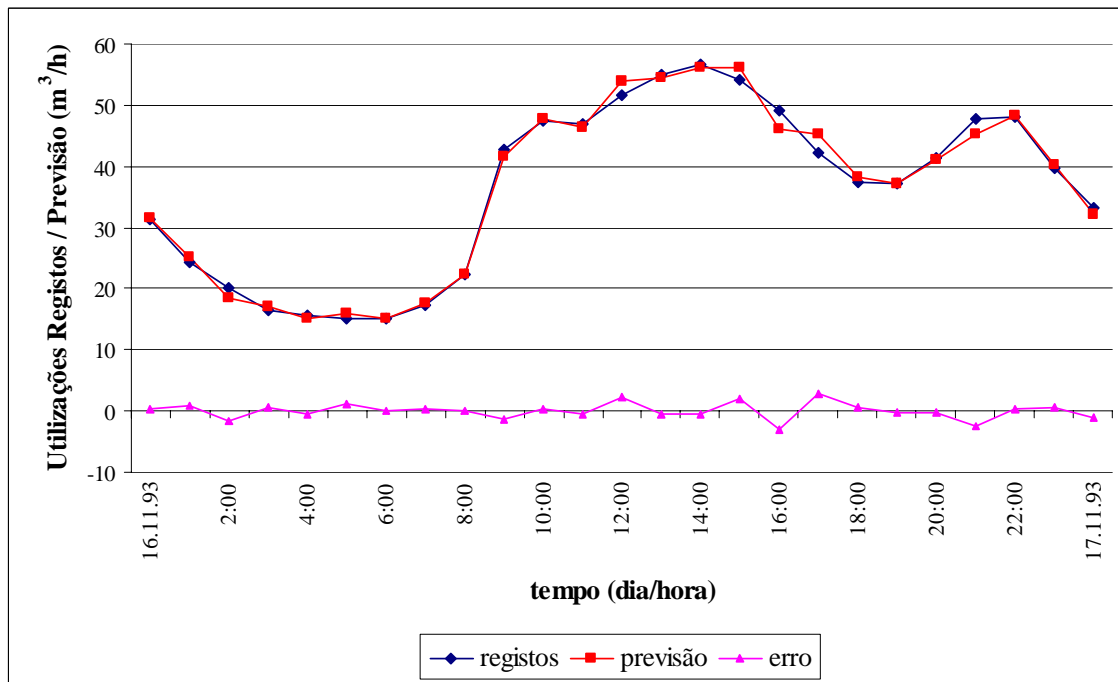


Figura 7 - Previsão hora a hora para dia de trabalho 16 de Novembro

Nas figuras seguintes apresenta-se a previsão a 24 horas para o dia 17 de Novembro e um pormenor onde se representam os valores dessa previsão versus valores registados.

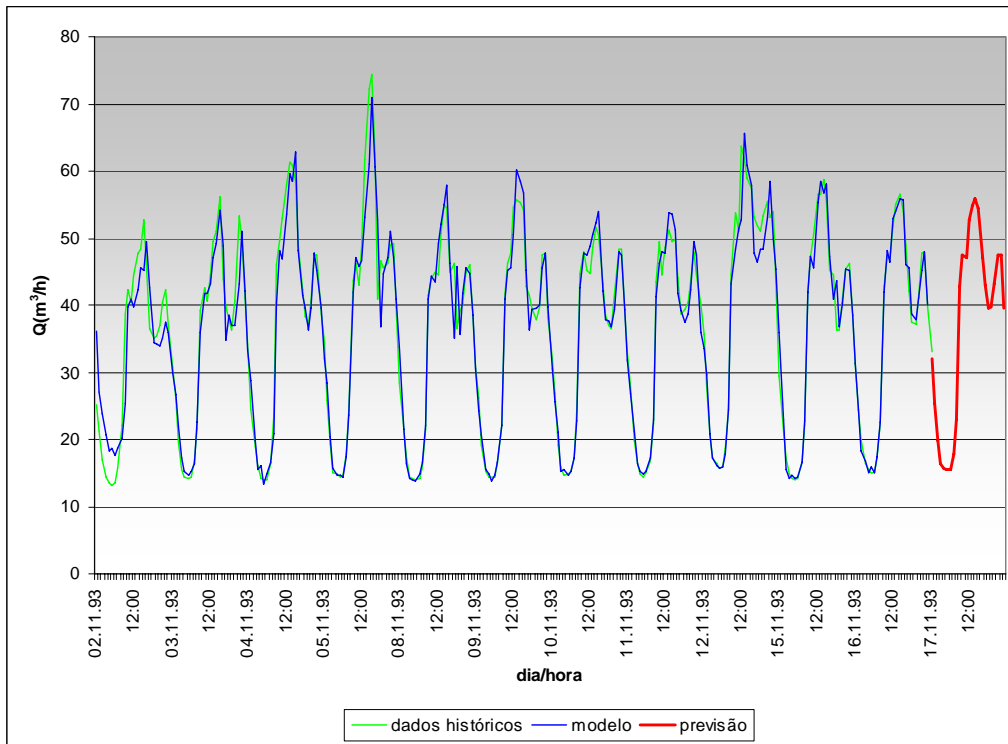


Figura 8 - Previsão para todo o dia de trabalho 17 de Novembro

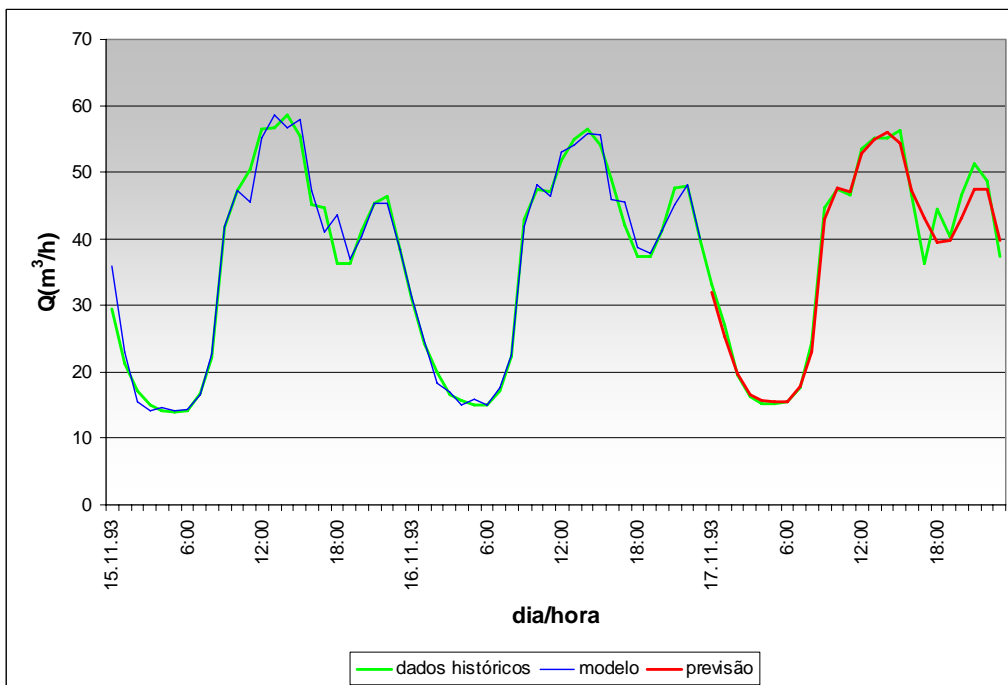


Figura 9 - Pormenor da previsão para todo o dia de trabalho 17 de Novembro

5 - CONCLUSÕES

A caracterização, modelação e previsão das utilizações de água são tarefas indispensáveis na concepção e gestão de sistemas de utilização da água. Este trabalho pode ser facilmente concretizado com o recurso a softwares de estatística. Por exemplo, índices sazonais que indicam a evolução das utilizações ao longo dum determinado período, designadamente ao longo das horas dum dia, podem ser facilmente determinados recorrendo ao conceito de médias móveis que incluem, dinamicamente, os registos mais recentes. A caracterização do comportamento típico dos caudais registados num determinado ponto de medição ajuda na identificação de comportamentos anómalos e, subsequentemente, na detecção de eventuais fugas de água das condutas. As metodologias devem ser definidas para que, desde logo, detectem erros ou a falta de coerência nos dados, já que a eficiência dum modelo de previsão está muito dependente da informação base usada. Assim, recomenda-se a separação prévia dos dados históricos consoante o tipo de dia: dias de trabalho, Sábados, Domingos e Feriados.

A análise e modelação estocástica através dos processos ARIMA tem a capacidade de se ajustar bem a sucessões cronológicas de caudais horários, conforme se constatou pelas aplicações a dados concretos.

As operações de gestão em tempo real de sistemas de abastecimento de água exigem que os modelos de previsão apresentem como característica principal a sua capacidade de adaptação, de resposta imediata às variações nas utilizações, mas em caso de alterações bruscas, para além da tolerância definida com recurso aos valores passados mais habituais para o instante em causa, dar sinais de alerta para que a eventual anomalia seja reparada. Os modelos apresentados neste trabalho possuem estas características. Isto fica provado através dos elevados coeficientes de determinação obtidos que indicam o bom ajuste dos modelos aos dados reais.

AGRADECIMENTOS

O primeiro autor regista o apoio dado pela Fundação para a Ciência e Tecnologia ao Centro de Engenharia Civil da Universidade do Minho.

REFERÊNCIAS

1. BOX, G.E.P., G.M.Jenkins, "Time series analysis forecasting and control", 1970;
2. FERREIRA DA SILVA, Júlio F., "Modelos Base de Apoio à Decisão no planeamento de Sistemas de Abastecimento de água", Trabalho de Síntese submetido à Universidade do Minho no Âmbito de provas de Aptidão Pedagógica e Capacidade Científica; Guimarães, Dezembro 1994;
3. FERREIRA DA SILVA, Júlio F., J. M. P. Vieira e Naim Haie, - "Caracterização Quantitativa de Consumos Horários de Água". in 3º Congresso Nacional da Água, Março de 1996;
4. FERREIRA DA SILVA, Júlio F., J. M. P. Vieira e Naim Haie, - "Modelação e Previsão de Consumos de Água em Tempo Real Utilizando Séries Matemáticas". in 3º Congresso Nacional da Água, Março de 1996;
5. MURTEIRA, Bento J. F. *et al*, "Análise de sucessões cronológicas", McGraw-Hill, Lisboa, 1993;
6. SPSS – Manual do Software SPSS V14, SPSS Inc., 2005;
7. TAVARES, L. Valadares *et al*, "Investigação Operacional"; Instituto Superior Técnico; McGraw-Hill; Lisboa, 1996.