



DETEÇÃO EM TEMPO REAL DE ROTURAS RECORRENDO A DADOS DE MONITORIZAÇÃO DA REDE DE DISTRIBUIÇÃO DE ÁGUA

Bruno Ferreira¹, Nelson Carriço², Dídia Covas³

^{1,2}INCITE, Escola Superior de Tecnologia do Barreiro, Instituto Politécnico de Setúbal,
Rua Américo da Silva Marinho, s/n, 2839-001 Lavradio, Portugal

³CERIS, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais 1,
1049-001, Lisboa, Portugal

¹bruno.s.ferreira@estbarreiro.ips.pt

RESUMO

A presente comunicação apresenta a aplicação de uma abordagem estatística para a deteção de roturas em redes de distribuição de água. A abordagem estatística avalia continuamente os dados de caudal, recebidos a partir de um caudalímetro instalado à entrada da rede, e de forma a identificar períodos de medições anómalas. Os eventos de rotura são detetados quando os períodos de medições anómalas excedem uma determinada duração pré-definida. A abordagem estatística é aplicada a dados de caudal de dois casos de estudo. Devido à inexistência de dados reais sobre roturas, o estudo incidirá sobre a introdução artificial de roturas nos dados de para posterior deteção utilizando a abordagem estatística. O objetivo é determinar os parâmetros da abordagem estatística que permitem minimizar (para cada caso de estudo) o tempo necessário para deteção e o número de falsos positivos. Os resultados obtidos permitem concluir que: i) a abordagem estatística pode ser parametrizada para minimizar o tempo necessário para deteção, ou, alternativamente, o número de alarmes falsos positivos; ii) os melhores parâmetros dependem da dimensão da rotura a ser detetada; e iii) para o mesmo caso de estudo, os melhores parâmetros para uma rotura de uma determinada magnitude variam ao longo do ano.

Palavras-Chave

Abordagem estatística, Deteção automática de roturas, Rede de abastecimento de água

1. INTRODUÇÃO

A deteção automática de roturas consiste em determinar se o fluxo de dados de pressão ou de caudal indica que uma rotura está, ou não, presente na rede. O principal desafio consiste em distinguir os sinais de uma rotura (usualmente caracterizado por um aumento no consumo ou numa redução da pressão) dos sinais associados ao normal funcionamento da rede,

nomeadamente, operação de bombas e válvulas, abertura de hidrantes ou consumos de água inesperados [1].

Nas últimas duas décadas têm-se assistido ao desenvolvimento de diversos métodos para a deteção automática de roturas que recorrem apenas dados de monitorização da rede em tempo real. Estes métodos têm a grande vantagem de não necessitarem de dados sobre características ou condição das condutas ou de simulação hidráulica e podem ser implementados em entidades que disponham apenas de um único sensor de caudal (nomeadamente, o caudalímetro de entrada). Dependendo das técnicas utilizadas, estes métodos podem ser divididos em quatro grandes grupos, nomeadamente, métodos de classificação, métodos de previsão, abordagens estatísticas e técnicas de agrupamento [2], [3].

Nos métodos de classificação pretende-se treinar um classificador com dados de caudal sobre o sistema tanto em situação de operação normal como “anormal” (i.e., sob efeito de rotura ou manobra de válvulas). Uma vez treinado, o classificador deve conseguir distinguir se o fluxo de dados em tempo real está ou não relacionado com uma rotura. Os métodos de previsão operam ao estimar o comportamento hidráulico em tempo real e a comparar este valor com o valor real medido. A deteção da rotura é então conseguida ao identificarem-se desvios significativos do comportamento real face ao comportamento estimado. As técnicas de agrupamento comparam, em termos de similaridades, o fluxo de medições de caudal em tempo real com as respetivas medições passadas. O racional é que, caso uma rotura esteja a ocorrer, as medições apresentarão uma baixa similaridade com as medições passadas (sob normal funcionamento da rede). Finalmente, as abordagens estatísticas não necessitam de técnicas adicionais de previsão ou classificação e utilizam análises estatísticas com gráficos de controlo sobre o fluxo de dados de caudal recebido. Comparativamente aos grupos de métodos previamente apresentados, as abordagens estatísticas têm a vantagem de serem usualmente mais simples, intuitivas e interpretáveis [4]

A presente comunicação apresenta a aplicação de uma abordagem estatística a dois casos de estudo para a deteção automática de roturas recorrendo apenas dados de monitorização do consumo total da rede. O objetivo é determinar os parâmetros da abordagem estatística que permitem minimizar (para cada caso de estudo) o tempo necessário para deteção e o número de falsos positivos.

2. METODOLOGIA

A abordagem estatística utilizada no presente estudo deteta roturas através da avaliação contínua de dados de caudal recebidos a partir de um caudalímetro instalado à entrada da rede de distribuição. As medições de caudal são registadas e transmitidas com uma determinada frequência fixa de aquisição, por exemplo, a cada 10 minutos. Cada medição de caudal, y_t , é caracterizada por um determinado valor (e.g., 20 m³/h) associado a um determinado instante temporal t (e.g., 04/05/2023 16:20:00). As medições recebidas podem ser classificadas como anómalas ou não anómalas. Este processo de classificação é realizado com base na normalização das medições (utilizando-se um registo histórico de um determinado número de dias anteriores). Um evento de rotura é finalmente detetado (e é emitido um alarme) quando o período de medições anómalas excede uma determinada duração pré-definida.

SMART WATER:

Transición hacia sistemas inteligentes, sostenibles y resilientes

Assim, o primeiro passo consiste na definição dos três parâmetros relevantes para a utilização da abordagem, nomeadamente:

- a *dimensão do registo histórico*, h , a ser utilizada na análise estatística e expresso em número de dias;
- o *valor limite*, k , que estabelece se uma medição (após normalização) é, ou não, anómala, expresso em número de desvios padrão;
- a *duração limite para medições anómalas consecutivas*, d , que, quando excedida, é emitido um alarme de que uma rotura foi detetada, expressa em horas.

A aplicação da abordagem estatística começa com a normalização Z-score da medição y_t a ser avaliada. Esta normalização pode ser conseguida com recurso à seguinte expressão:

$$z_t = \frac{y_t - \bar{y}}{\sigma} \quad (1)$$

Em que y_t é a medição a ser avaliada; z_t é o valor normalizado para a medição y_t ; $\bar{y}$ é o valor de caudal médio para as últimas h medições não anómalas para a mesma hora do dia que t ; σ é o desvio padrão para as últimas h medições não anómalas para a mesma hora do dia que t . Note-se que o valor de $\bar{y}$ e de σ varia com a hora do dia em t . Esta abordagem considera que a diferença entre dias úteis e fins-de-semana são mínimas e, como tal, a normalização não diferencia entre este tipo de dias. Caso a diferença entre dias úteis e fins-de-semana seja relevante, a abordagem pode ser adaptada ao considerar que as últimas h medições não anómalas não são apenas referentes à mesma hora que t , mas também ao mesmo dia da semana.

O valor normalizado, z_t , representa o número de desvios padrão que a medição de caudal y_t está acima ou abaixo em relação ao valor médio, tendo em conta os registos históricos $\bar{y}$, e para a mesma hora que t . Para determinar se uma medição é, ou não, anómala, o respetivo valor normalizado é comparado com o limite pré-estabelecido v . Este limite pode ser definido, por exemplo, igual a 2. Neste caso, uma medição y_t é considerada anómala quando o seu valor é superior ao valor médio $\bar{y}$ por uma margem superior a duas vezes o valor do desvio padrão σ .

Consoante o resultado do processo de classificação, a medição classificada é adicionada ao registo histórico como medição anómala (e pode ser utilizada na normalização de medições futuras) ou não anómala (e não poderá ser utilizado no processo de normalização). Caso a medição seja classificada como não anómala, o processo de avaliação da medição y_t dá-se por concluído e a abordagem é recomeçada para uma nova medição. Se, por outro lado, a medição é classificada como anómala, é necessário avaliar a duração de medições de caudal anómalas consecutivas.

Embora uma medição anómala não confirme a existência de uma rotura, um longo período com medições anómalas consecutivas pode indiciar que um consumo anómalo (nomeadamente uma rotura) está presente na rede. Assim, é avaliado se as últimas d horas de medições são anómalas. Um evento de rotura só é considerado detetado (e só é emitido um alarme) quando o período de medições anómalas consecutivas excede a duração d . Após emissão do alarme, são idealmente enviadas equipas para o terreno de forma a localizar-se o evento detetando.

A deteção de um evento de rotura é, idealmente, realizada no menor tempo possível face ao momento no qual este evento teve início. Neste sentido, qualquer técnica de deteção de roturas deve ser parametrizada para minimizar o tempo necessário à deteção destes eventos. Existe, no entanto, o risco de que a o evento detetado não se trate de um evento de rotura, mas sim de um consumo inesperado que foi mal interpretado pela técnica de deteção. Assim, qualquer técnica de deteção de roturas deve ser também parametrizada para minimizar o número de alarmes falsos positivos.

3. RESULTADOS

A abordagem estatística para deteção de roturas é aplicada a dados de caudal de dois casos de estudo. O primeiro caso de estudo está localizado numa zona turística do sul de Portugal com elevado efeito de sazonalidade. O número de consumidores varia entre 3.000 no inverno e 14.000 no verão, sendo a área maioritariamente composta por moradias com grandes espaços ajardinados. O consumo num dia médio de verão é de cerca de 235 m³/h, bastante superior ao consumo médio num dia de inverno (60 m³/h). O segundo caso de estudo está localizado na área metropolitana de Lisboa e serve uma população de cerca de 3.300 habitantes. A área é composta maioritariamente por residências unifamiliares e edifícios até dois pisos. O consumo num dia médio de verão é de cerca de 55 m³/h e num dia médio de inverno é de cerca de 25 m³/h. Ambos os casos de estudo dispõem de um medidor de caudal instalado à saída do reservatório que recolhe os dados com uma frequência de 10 minutos.

As roturas artificiais são introduzidas nas séries temporais como um incremento constante ao caudal. São consideradas roturas com duas magnitudes (15% e 40% do consumo médio ao longo dos h dias de registo histórico e doravante designadas por rotura de pequena e grande magnitude) e em diferentes períodos do ano (i.e., no verão e no inverno). No total são avaliadas oito roturas (dois casos de estudo, duas magnitudes, dois períodos do ano). Para cada rotura são avaliadas quatro combinações de parâmetros da abordagem estatística (como resultado de duas alternativas para cada um dos parâmetros d e k). O objetivo é determinar, para cada caso de estudo, os parâmetros ideais que minimizem tanto o número de falsos positivos como o tempo necessário para a deteção. Considera-se que a duração do registo histórico, h , é igual a 14 dias.

A deteção de cada evento de rotura é realizada de forma a simular a aplicação da abordagem em contexto real. Neste sentido, avalia-se uma semana de medições de caudal com a rotura introduzida artificialmente a partir da marca dos 3,5 dias (i.e., a partir dos 3,5 dias a contar do tempo 0). Na aplicação da abordagem estatística foram consideradas as seguintes regras:

- Se a rotura é detetada antes de ter realmente acontecido (i.e., antes da marca dos 3,5 dias), considera-se como um alarme falso positivo.
- Se a rotura é detetada depois de ter realmente acontecido (i.e., depois da marca dos 3,5 dias) considera-se que a localização e reparação demora 24 h e, a partir desse momento, remove-se o efeito da rotura artificialmente introduzida.
- Se a rotura é detetada depois de ser sido reparada, considera-se como um alarme falso positivo.

SMART WATER:

Transición hacia sistemas inteligentes, sostenibles y resilientes

Para demonstrar o procedimento utiliza-se o CE1 de forma a detetar uma rotura que ocorreu no inverno (especificamente a semana entre 11 e 18 de fevereiro). Os valores dos parâmetros d e k são ambos considerados igual a 1.

A Figura 2 apresenta os resultados obtidos com recurso à abordagem estatística. A Figura 2a apresenta as medições de caudal a serem avaliadas y_t (em linha preta), enquanto que o valor médio para cada passo de tempo (obtido com base nos últimos 14 valores não anómalos para essa hora) é apresentado a linha laranja. As duas linhas azuis verticais correspondem, respetivamente, ao início e ao fim do evento de rotura. A Figura 2b apresenta o valor Z-score associado a cada medição de caudal obtido com recurso à Equação 1. A linha horizontal cinza corresponde ao limite k (neste caso igual a 1) e, como tal, medições com valor Z-score inferior a este limite são consideradas como medições não anómalas (e são apresentadas a linha verde). Por outro lado, medições com Z-score superior ao limite k são consideradas como medições anómalas (e são apresentadas com a linha vermelha).

Finalmente, na Figura 2c apresenta-se o resultado da aplicação da abordagem estatística. Os períodos com medições anómalas ou não anómalas são representados, respetivamente, como uma área vermelha ou verde. Os alarmes de deteção de rotura (representados com linhas tracejadas verticais) são emitidos quando os períodos com medições anómalas consecutivas excedem uma duração limite d . Neste caso, foram emitidos 10 alarmes ao longo da semana, cinco dos quais após ocorrência e durante a reparação da rotura. Os restantes cinco alarmes são considerados como alarmes falsos positivos. Adicionalmente, também é possível avaliar o tempo necessário para a deteção do evento, e que corresponde ao intervalo de tempo compreendido entre o momento em que a rotura teve início (primeira linha vertical azul) e o alarme emitido após início de tal evento. Para esta situação, o tempo necessário à deteção foi de 4 horas e 20 minutos.

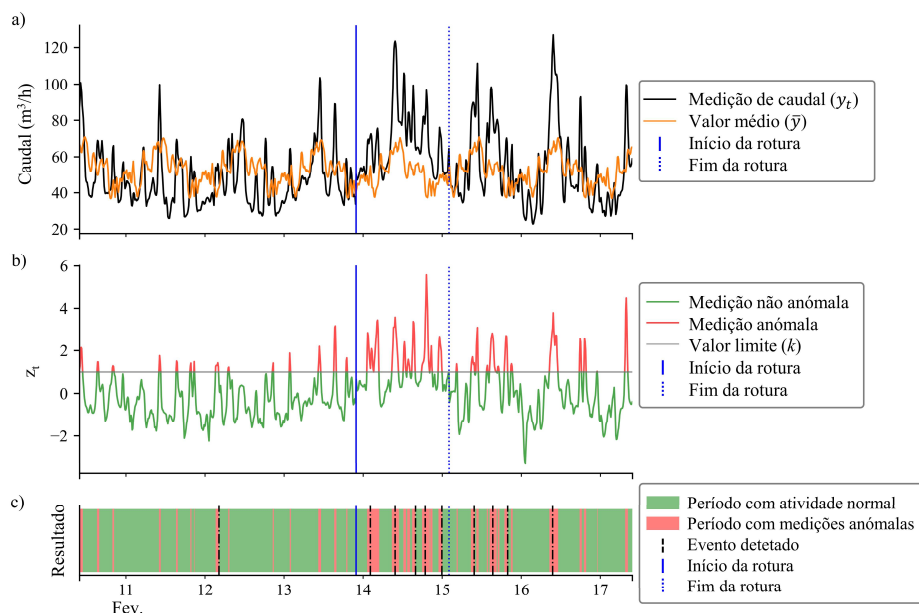


Figura 1. Deteção de uma rotura de grande magnitude, no inverno, para o CE1 (considera-se $k=1$ e $d=1$): a) medições de caudal e valor médio para cada passo de tempo; b) valores Z-score para cada medição de caudal; c) resultados com indicação dos períodos com medições anómalas e dos alarmes emitidos.

SMART WATER:

Transición hacia sistemas inteligentes, sostenibles y resilientes

A abordagem estatística foi aplicada ao CE1 para detecção de roturas de pequena ou grande magnitude, a ocorrer no inverno, e considerando quatro alternativas de parâmetros (i.e., k de 0,5 e 1 e d de 1 e 3). O Quadro 1 apresenta os resultados obtidos, nomeadamente, o número de alarmes falsos positivos e o tempo necessário à detecção de cada rotura. É possível concluir que, para roturas de igual magnitude, valores menores para os parâmetros (i.e., $k=0,5$ ou $d=1$) levam a um maior número de falsos positivos, mas com um menor tempo necessário para detecção. Por outro lado, valores mais elevados para os parâmetros (i.e., $k=1$ ou $d=3$) levam a um menor número de falsos positivos, mas a um maior tempo para detecção. Também é possível concluir que roturas de pequena magnitude (e quando comparadas com as de grande magnitude) apresentam um maior tempo para detecção.

Quadro 1. Resultados da detecção para o CE1, para uma rotura de pequena e de grande magnitude a ocorrer no inverno, e considerando quatro alternativas de parâmetros.

Número de alarmes falsos positivos [tempo necessário à detecção]				
	Rotura de grande magnitude		Rotura de pequena magnitude	
	$k = 0.5\sigma$	$k = 1\sigma$	$k = 0.5\sigma$	$k = 1\sigma$
$d = 1 \text{ h}$	10 [4h10]	5 [4h20]*	11 [6h40]	4 [12h10]
$d = 3 \text{ h}$	2 [6h10]	0 [6h20]	0 [42h10]	0 [43h30]

Nota: *Resultados apresentados na Figura 2.

O Quadro 2 apresenta os resultados da detecção para o CE1, para roturas ocorridas no verão, e considerando quatro alternativas de parâmetros (i.e., k de 1 e 2 e d de 1 e 3). Para roturas de igual magnitude, valores menores para os parâmetros levam a um maior número de falsos positivos, mas com um menor tempo necessário para detecção. Note-se que os melhores parâmetros (que resultam em zero falsos positivos) não são os mesmos para o inverno e para o verão. Para roturas ocorridas no verão, é interessante notar que não existe diferença no tempo necessário para detecção de roturas de pequena ou de grande magnitude. Neste caso de estudo o consumo de água no verão é dominado por regas com sistemas automáticos que garantem a rega periódica. Assim, e apesar do consumo ser muito maior no verão do que no inverno, a variabilidade dos consumos é menor. Como tal, o efeito de uma rotura de pequena ou grande magnitude apresenta um desvio significativo ao consumo expectável e pode ser detetada de forma célere.

Quadro 2. Resultados da detecção para o CE1, para uma rotura de pequena e de grande magnitude a ocorrer no verão, e considerando quatro alternativas de parâmetros.

Número de alarmes falsos positivos [tempo necessário à detecção]				
	Rotura de grande magnitude		Rotura de pequena magnitude	
	$k = 1\sigma$	$k = 2\sigma$	$k = 1\sigma$	$k = 2\sigma$
$d = 1 \text{ h}$	28 [1h]	3 [1h]	28 [1h]	3 [1h]
$d = 3 \text{ h}$	8 [3h]	0 [3h]	8 [3h]	0 [3h]

SMART WATER:

Transición hacia sistemas inteligentes, sostenibles y resilientes

O Quadro 3 apresenta os resultados da detecção para o CE2, para roturas ocorridas no inverno, e considerando quatro alternativas de parâmetros (i.e., k de 1 e 2 e d de 1 e 3). Por um lado, os resultados obtidos são os esperados: valores menores para os parâmetros levam a um maior número de falsos positivos, mas com um menor tempo necessário para detecção. Por outro lado, e de forma similar ao CE1 (Quadro 2), não existe diferença no tempo necessário para detecção de roturas de pequena ou de grande magnitude. A rede abastece uma pequena área maioritariamente residencial com rotinas relativamente bem definidas e baixa variabilidade nos consumos durante o inverno. Desta forma, uma rotura (de pequena ou grande magnitude) causa um desvio significativo ao consumo expectável e pode ser detetada de forma célere.

Quadro 3. Resultados da detecção para o CE2, para uma rotura de pequena e de grande magnitude a ocorrer no inverno, e considerando quatro alternativas de parâmetros.

Número de alarmes falsos positivos [tempo necessário à detecção]				
	Rotura de grande magnitude		Rotura de pequena magnitude	
	$k = 1\sigma$	$k = 2\sigma$	$k = 1\sigma$	$k = 2\sigma$
$d = 1 \text{ h}$	10 [1h]	3 [1h]	10 [1h]	1 [1h]
$d = 3 \text{ h}$	1 [3h]	0 [3h]	2 [3h]	0 [3h]

O Quadro 4 apresenta os resultados da detecção para o CE2, para roturas ocorridas no verão, e considerando quatro alternativas de parâmetros (i.e., k de 1 e 2 e d de 1 e 3). Uma vez mais, valores menores para os parâmetros levam a um maior número de falsos positivos, mas com um menor tempo necessário para detecção. É importante realçar que, para este caso de estudo, os melhores parâmetros (que resultam em zero falsos positivos) são os mesmos para o verão e para o inverno (i.e., $d=3$ e $k=2$). A variabilidade dos consumos é relativamente alta no verão (e especialmente quando comparada com a variabilidade dos consumos no inverno). Esta diferença está relacionada com atividades imprevisíveis da população servida, tais como, ir ou regressar de férias, o enchimento de piscinas, ou a rega de espaços verdes. Esta variabilidade é diretamente refletida na capacidade do método, que apresenta um maior tempo necessário para a detecção de roturas de pequena magnitude (e em igualdade dos restantes parâmetros).

Quadro 4. Resultados da detecção para o CE2, para uma rotura de pequena e de grande magnitude a ocorrer no verão, e considerando quatro alternativas de parâmetros.

Número de alarmes falsos positivos [tempo necessário à detecção]				
	Rotura de grande magnitude		Rotura de pequena magnitude	
	$k = 1\sigma$	$k = 2\sigma$	$k = 1\sigma$	$k = 2\sigma$
$d = 1 \text{ h}$	11 [1h20]	2 [3h20]	11 [5h40]	2 [7h40]
$d = 3 \text{ h}$	3 [3h20]	0 [5h20]	2 [7h40]	0 [56h40]

4. DISCUSSÃO E CONCLUSÕES

A aplicação da abordagem estatística para detecção de roturas a dois casos de estudo permitiu demonstrar que:

- A abordagem estatística pode ser parametrizada para minimizar o tempo necessário à detecção (com valores de parâmetros baixos) ou o número de alarmes falsos positivos (com valores de parâmetros altos).
- Os melhores parâmetros dependem da magnitude da rotura a ser detetada. Para o CE2 no verão, os melhores parâmetros para uma rotura de grande magnitude são $k=2$ e $d=3$, enquanto que para uma rotura de pequena magnitude são $k=1$ e $d=3$. Por outras palavras, a abordagem estatística pode apenas ser idealmente parametrizada para detetar uma certa gama de magnitudes de rotura.
- Para o mesmo caso de estudo, os melhores parâmetros para uma certa magnitude de rotura variam durante o ano. Isto pode ser visto no CE1, onde os melhores parâmetros para uma rotura de grande magnitude são $k=1$ e $d=3$ no inverno e $k=2$ e $d=3$ no verão.

Concluindo, a abordagem estatística fornece um método prático e eficaz para a detecção de roturas e pode ser aplicada em entidades que disponham de um único caudalímetros de entrada na rede de distribuição. Contudo, é importante notar que a correta parametrização da abordagem estatística é crucial para evitar um número elevado de falsos alarmes positivos, que podem afetar negativamente a confiança nos resultados da abordagem. Os trabalhos futuros incluem a utilização de algoritmos de inteligência artificial para apoio no processo de parametrização.

AGRADECIMENTOS

Os autores agradecem à Fundação para a Ciência e Tecnologia (FCT) pelo financiamento da investigação através da bolsa de doutoramento do Bruno Ferreira (referência n.º SFRH/BD/149392/2019) e da unidade de investigação CERIS (referência n.º UIDB/04625/2020).

REFERÊNCIAS

- [1] S. El-Zahab and T. Zayed, "Leak detection in water distribution networks: an introductory overview," *Smart Water*, vol. 4, no. 1, p. 5, Dec. 2019, doi: 10.1186/s40713-019-0017-x.
- [2] Z. Hu, B. Chen, W. Chen, D. Tan, and D. Shen, "Review of model-based and data-driven approaches for leak detection and location in water distribution systems," *Water Supply*, vol. 21, no. 7, pp. 3282–3306, Nov. 2021, doi: 10.2166/ws.2021.101.
- [3] Y. Wu and S. Liu, "A review of data-driven approaches for burst detection in water distribution systems," *Urban Water J.*, vol. 14, no. 9, pp. 972–983, Oct. 2017, doi: 10.1080/1573062X.2017.1279191.
- [4] X. Wan, P. K. Kuhanestani, R. Farmani, and E. Keedwell, "Literature Review of Data Analytics for Leak Detection in Water Distribution Networks: A Focus on Pressure and Flow Smart Sensors," *J. Water Resour. Plan. Manag.*, vol. 148, no. 10, Oct. 2022, doi: 10.1061/(ASCE)WR.1943-5452.0001597.