



TRATAMENTO DE DADOS ANÓMALOS DE CAUDAL EM SISTEMAS DE ABASTECIMENTO DE ÁGUA

Bruno Ferreira¹, Nelson J. G. Carriço¹

1. Escola Superior de Tecnologia do Barreiro, Instituto Politécnico de Setúbal, Rua Américo da Silva Marinho, 2839-001 Lavradio, Portugal

bruno.s.ferreira@estbarreiro.ips.pt

nelson.carrico@estbarreiro.ips.pt

RESUMO

Os sistemas de telemetria associados a sistemas de abastecimento de água recolhem e armazenam, de forma sistemática e contínua, medições de caudal e pressão com uma dada frequência de aquisição. Estas medições constituem séries temporais que apresentam, frequentemente, valores anómalos, muitas vezes com origem na falta de fiabilidade do sensor, como por exemplo, problemas com o sistema de transmissão de dados ou de armazenamento, ou ainda funcionamento do sistema fora do padrão normal (e.g., ocorrência roturas, manobras em válvulas).

No âmbito do presente trabalho entende-se por dado anómalo uma observação que apresenta inconsistência com as restantes observações e/ou que apresenta uma baixa probabilidade de ter sido originado a partir da mesma distribuição estática das restantes observações (Barnet e Lewis, 1994). Desta forma, a sua inclusão na análise poderá comprometer a determinação de diversos parâmetros, conduzindo a resultados enviesados podendo, por isso, comprometer a avaliação do desempenho do sistema. É, portanto, de extrema importância a identificação destes dados anómalos previamente à fase de modelação e análise do sistema (Liu *et al.*, 2004).

O interesse no tratamento dos dados anómalos de caudal que deu origem a este trabalho está na pretensão de aplicar técnicas avançadas (e.g., aprendizagem automática) para a deteção de eventos anómalos (e.g., rotura) em tempo quase real. No entanto, para a aplicação destas técnicas é necessário que as séries temporais de caudal sejam validadas através do tratamento dos dados anómalos e que a frequência temporal seja normalizada, dado que esta é, sistematicamente, irregular.

Desta forma, o tratamento incide sobre os diversos problemas resultantes do processo de recolha, transmissão e armazenamento dos dados sem, no entanto, manipular possíveis eventos anómalos. Apesar do tratamento e validação de dados ser um ramo de investigação bastante consolidado, ainda não é uma questão prioritária para muitas entidades gestoras (Kirstein *et al.*, 2019).

No âmbito do projeto de investigação “Water Intelligence System Data” (WISDom) foi desenvolvida uma metodologia para o tratamento e validação de séries temporais de caudal. O projeto conta com a participação de três entidades gestoras (EG) bastante diferentes em termos de número de clientes, nível socioeconómico, dimensão da rede (e.g., extensão em quilómetros, número de ZMC), volume anual fornecido, e localização geográfica. Desta forma, é possível abranger diversos tipos de anomalias nos dados, tendo em conta os diferentes tipos de sistemas de recolha, transmissão e armazenamento presente nas EG parceiras, assim como a respetiva maturidade digital e práticas já em curso para o tratamento destes dados.

Assim, o primeiro passo consistiu na **identificação dos diversos tipos de anomalias** presentes nas séries das três EG parceiras, nomeadamente:

- Irregularidade na frequência de aquisição;
- Duplicação de leituras, apresentando uma medição igual;
- Duplicação de leituras, apresentando uma medição diferente;



- Valores anormalmente baixos entre valores altos;
- Valores anormalmente altos entre valores baixos;
- Valores negativos;
- Zeros pontuais entre medições elevadas;
- Patamares estáticos.

De seguida, e para cada tipo de anomalia identificado, são **aplicados técnicas e testes específicos para a sua identificação automática**. Esta identificação é de extrema importância pois permitirá o posterior tratamento destes dados, tendo em conta a duração da anomalia.

O próximo passo consiste na **remoção de anomalias e na imputação de valores pontuais**. As diversas anomalias detetadas nas séries cuja duração seja inferior a 30 minutos são eliminadas e são imputados diretamente valores para o seu lugar. As anomalias que apresentem duração superior a 30 minutos são removidas das séries, deixando no seu lugar uma falha que será preenchida posteriormente através de métodos regressivos.

Uma vez removidos os dados anómalos que poderiam comprometer a análise, é possível proceder-se à **normalização da frequência temporal das séries**. Esta normalização consiste na transformação da série com frequência de aquisição irregular para uma série com um passo de tempo regular. Esta normalização é realizada com recurso à regra dos trapézios, de forma a calcular-se o caudal médio associado a um passo de tempo pré-estabelecido. Face ao desconhecimento sobre a verdadeira distribuição do consumo em falhas com duração superiores a 30 minutos, a normalização da frequência temporal é interrompida, reiniciando-se após a falha.

Uma vez normalizada a série é possível proceder-se ao **preenchimento das falhas** (i.e., superiores a 30 minutos). Este preenchimento é efetuado com base no histórico de medições (estando este histórico já normalizado) e através de métodos autorregressivos.

O objetivo do presente artigo é a aplicação da metodologia desenvolvida para o tratamento de dados anómalos em séries de caudal, associados a problemas durante o processo de recolha, transmissão e armazenamento, e tendo em vista a posterior utilização para a deteção de eventos anómalos recorrendo a técnicas de aprendizagem automática. Para tal, serão identificados os diversos problemas recorrentes existentes nas séries temporais de caudal e, para cada um destes, serão utilizadas diferentes técnicas. As séries temporais analisadas serão de três casos de estudo reais, bastante distintos entre si.

Palavras-Chave: Caudal; Dados anómalos; Séries temporais; Tratamento de dados

REFERÊNCIAS BIBLIOGRÁFICAS

Barnet, V., Lewis, T. (1994). *Outliers in statistical data*. John Wiley

Kirstein, J. K., Høgh, K., Rygaard, M., & Borup, M. (2019). A semi-automated approach to validation and error diagnostics of water network data. *Urban Water Journal*, 16(1), 1–10.
<https://doi.org/10.1080/1573062X.2019.1611884>

Liu, H., Shah, S., & Jiang, W. (2004). On-line outlier detection and data cleaning. *Computers and Chemical Engineering*, 28(9), 1635–1647. <https://doi.org/10.1016/j.compchemeng.2004.01.009>