

Análise de falhas em rede de distribuição de água

2.^a PARTE

Francisco José Valente Ferreira, José Torres Farinha, Mateus Mendes
Instituto Politécnico de Coimbra – ISEC
la2019113494, tfarinha, mmendes@isec.pt

3. DADOS EM ANÁLISE

Os dados utilizados no presente trabalho são de uma zona de Portugal, onde foram inicialmente obtidos numa Zona de Medição e Controlo (ZMC). Incluem caudal medido à saída da central, registos de intervenções técnicas na rede desde 2018 e características da rede de tubagens. O processo de análise seguiu a metodologia CRISP-DM, sendo esta secção dedicada à vertente de *data understanding*. O objetivo é permitir encontrar características relevantes para a deteção de falhas ao longo do tempo no caudal, bem como uma análise geral.

Durante o processo de análise foi preciso procurar ferramentas que permitissem a fácil gestão e processamento de dados. Uma das ferramentas mais conhecidas para desenvolver trabalho de análise de dados é a linguagem python, com ambiente de desenvolvimento *jupyter notebook*. O python é usado tanto pela capacidade de visualização e de processamento como pela versatilidade, quando se usam bibliotecas clássicas e mais recentes, nomeadamente o *pandas*, o *matplotlib* e o *sk-learn*. *Pandas* é uma biblioteca de *software* criada para a linguagem Python, para manipulação e análise de dados. Em particular, oferece estruturas e operações para manipular tabelas numéricas e séries temporais. *Matplotlib* é uma biblioteca de *software* para criação de gráficos e visualizações de dados em geral, feita para a linguagem de programação Python e sua extensão de matemática *NumPy*. *Sk-learn* é uma biblioteca que engloba os mais populares modelos de *machine learning* clássico.

A ferramenta *jupyter notebook* serve para desenvolver *software* de uma forma interativa, em dezenas de linguagens de programação. Isto significa que permite rapidez de visualização e alteração de código através de memórias visuais de *outputs* de código, o que permite a partilha e atualização desses *outputs* sem correr novamente todo o código desenvolvido.

A ZMC em análise caracteriza-se por ser uma planície. Os dados obtidos de caudal correspondem a medições em m³ por hora. O período de amostragem é de 1 h, por razões de rendimento das baterias associadas aos sensores de medição. O *dataset* contém as medições que foram registadas desde janeiro 2020 até agosto de 2022.

A análise mensal do caudal mostra uma variação de sinal ao longo do ano, claramente similar ao longo dos anos em análise. Dá para ver que durante os meses de inverno e primavera, até ao início do verão, há uma estabilização do valor do caudal, pouco crescente, mas que em maio se constata um aumento significativo do caudal.

A Figura 1 mostra a variação do caudal ao longo do ano. São visíveis os aumentos significativos nos meses de verão, onde há variações mais acentuadas, que serão mais propensas a fugas nas ligações. O ano de 2021 mostra uma redução, que pode ser explicada pela redução do consumo. Essa redução do consumo pode ser

comprovada pela Figura 13, que mostra que o ano 2021 foi bastante chuvoso. Logo, o uso da água armazenada em poços e furos garante a pouca utilização da água da ZMC em análise. Sendo assim, foi possível concluir que o principal fator de influência do caudal, anualmente, é a precipitação.

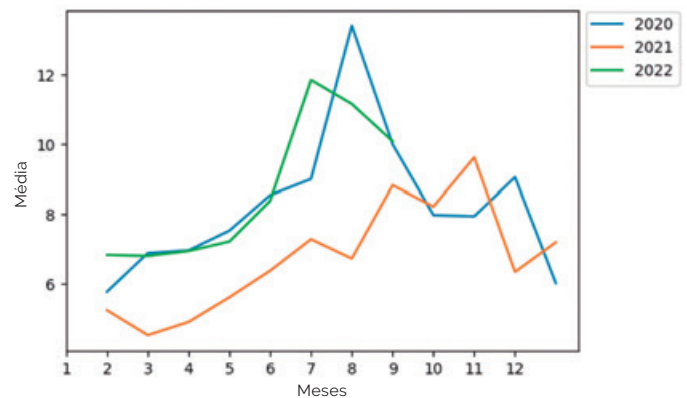


Figura 1. Média mensal do caudal.



Figura 2. Boletim climatológico inverno 2022 [7].

A Figura 3 mostra o sinal do caudal, agregado por mês, que também tem uma variação previsível. Dependendo do dia da semana, é possível observar a queda da média diária na transição de domingo para segunda, que se mantém estável durante a semana e que no sábado