

# **IDENTIFICAÇÃO DE NOVIDADES NO COMPORTAMENTO OBSERVADO EM BARRAGENS DE BETÃO COM RECURSO A MODELOS DE *MACHINE LEARNING***

**JOSÉ MIGUEL DA COSTA MENDES**

Dissertação submetida para satisfação parcial dos requisitos do grau de  
**MESTRE EM ENGENHARIA CIVIL — ESPECIALIZAÇÃO ESTRUTURAS E GEOTECNIA**

---

Orientador: Professor Doutor Álvaro Alberto de Matos Ferreira da  
Cunha

---

Coorientador: Doutor Sérgio Bouça Pereira

Coorientador: Doutor Juan Tomé Caires da Mata

JUNHO DE 2023

## **MESTRADO EM ENGENHARIA CIVIL 2022/2023**

DEPARTAMENTO DE ENGENHARIA CIVIL

Tel. +351-22-508 1901

✉ [m.ec@fe.up.pt](mailto:m.ec@fe.up.pt)

*Editado por*

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Rua Dr. Roberto Frias

4200-465 PORTO

Portugal

Tel. +351-22-508 1400

✉ [feup@fe.up.pt](mailto:feup@fe.up.pt)

🌐 <http://www.fe.up.pt>

Reproduções parciais deste documento serão autorizadas na condição que seja mencionado o Autor e feita referência a *Mestrado em Engenharia Civil - 2022/2023 - Departamento de Engenharia Civil, Faculdade de Engenharia da Universidade do Porto, Porto, Portugal, 2023*.

As opiniões e informações incluídas neste documento representam unicamente o ponto de vista do respetivo Autor, não podendo o Editor aceitar qualquer responsabilidade legal ou outra em relação a erros ou omissões que possam existir.

Este documento foi produzido a partir de versão eletrónica fornecida pelo respetivo Autor.

À minha família e amigos,

*“Tudo é considerado impossível, até acontecer”*

*Nelson Mandela*



## **AGRADECIMENTOS**

A realização desta dissertação marca o desfecho da minha vida académica, não teria sido possível sem o apoio de algumas pessoas especiais, às quais dedico estas palavras de profundo agradecimento.

Agradeço à Faculdade de Engenharia da Universidade do Porto, entidade que me acolheu durante o meu percurso académico.

Ao Laboratório Nacional de Engenharia Civil (LNEC), pela disponibilização de todos os meios necessários para a realização desta dissertação.

Ao meu orientador, Professor Doutor Álvaro Cunha pela oportunidade de realização desta dissertação, pelos conselhos e pela partilha de conhecimento.

Ao meu coorientador Doutor Sérgio Pereira pela ajuda, conselhos, disponibilidade e prontidão.

Ao meu coorientador Doutor Juan Mata pela orientação, pela forma positiva de transmissão de conhecimento, pela calma ao longo desta dissertação.

Aos meus pais e irmã por todo o esforço e apoio que foi feito diariamente, mas sobretudo por acreditarem e me ajudarem a concretizar este sonho.

Ao meu grupo de amigos que me acompanham desde a escola primária que estão sempre lá para apoiar e partilhar experiências boas e más.

Aos amigos pelo qual tive o prazer de conhecer neste percurso académico da FEUP, Mafalda Ramos, Ana Cabo, Pepe, Hélder, Pedro, entre outros, pelo apoio e pelas muitas horas passadas em companhia.



## RESUMO

As barragens de betão são estruturas críticas que desempenham um papel fundamental no armazenamento e fornecimento de água, na produção de energia e controlo de enchentes. É essencial garantir a segurança e a integridade dessas estruturas para evitar colapsos ou falhas que possam resultar em consequências catastróficas. Nesse contexto, a aplicação de modelos de *Machine Learning* oferece uma abordagem promissora para a monitorização e controlo contínuo das barragens. Esses modelos são treinados com base em dados históricos que descrevem o comportamento normal das barragens, como medições de deformações, deslocamentos, pressão da água e outros parâmetros relevantes.

Esta dissertação começa por uma breve revisão bibliográfica sobre a história das barragens, dos tipos de barragens de betão e ações nelas aplicadas, referindo a importância do controlo de segurança e monitorização. Posteriormente são apresentados métodos de *Machine Learning* relevantes para a identificação de valores atípicos como o DBSCAN e o LOF.

A deteção antecipada de valores atípicos no comportamento observado em barragens de betão é de extrema importância, pois permite que os engenheiros e responsáveis pela gestão das barragens tomem medidas preventivas antes que problemas mais graves ocorram. Com a identificação desses valores atípicos, podem ser realizadas inspeções mais detalhadas, reparações ou ajustes necessários para garantir a segurança e o funcionamento adequado das barragens.

O objetivo principal deste trabalho é utilizar modelos de *Machine Learning* com o intuito de identificar valores atípicos no comportamento da barragem, recorrendo a dados provenientes do sistema de monitorização da mesma.

Selecionou-se para caso de estudo a barragem do Feiticeiro. Esta barragem apresenta um sistema de recolha automática de dados e um histórico de dados em que é vantajosa a aplicação dos modelos de *Machine Learning* com o objetivo de identificar valor atípicos.

**PALAVRAS-CHAVE:** Barragem de betão, Valores atípicos, *Machine Learning*, Barragem do Feiticeiro, Recolha automática de dados.





## **ABSTRACT**

Concrete dams are critical structures that play a key role in water storage and supply, energy production, and flood control. It is essential to ensure the safety and integrity of these structures to avoid collapse or failure that could result in catastrophic consequences. In this context, the application of Machine Learning models offers a promising approach for the continuous monitoring and control of dams. These models are trained based on historical data describing the normal behavior of dams, such as measurements of stress, strain, water pressure, and other relevant parameters.

This dissertation begins with a brief literature review on the history of dams, the types of concrete dams and actions applied to them, referring to the importance of safety control and monitoring. Subsequently, Machine Learning methods relevant for the identification of outliers, such as DBSCAN and LOF, are presented.

The early detection of outliers in the behavior observed in concrete dams is extremely important, as it allows engineers and those responsible for dam management to take preventive measures before more severe problems occur. With the identification of these atypical values, more detailed inspections, repairs, or necessary adjustments can be made to ensure the safety and proper functioning of the dams.

The main objective of this work is to use Machine Learning models to identify potential outliers' values in the observed dam behaviour, using data from the dam's monitoring system.

The Feiticeiro dam was selected as a case study. This dam has an automatic data acquisition system and a history of data in which is advantageous for the application of Machine Learning models with the objective of identifying outliers.

**KEY WORDS:** Concrete Dam, Outliers, Machine Learning, Feiticeiro Dam, Automatic Data Acquisition.

## Índice Geral

|                                                                              |             |
|------------------------------------------------------------------------------|-------------|
| <b>AGRADECIMENTOS .....</b>                                                  | <b>I</b>    |
| <b>RESUMO .....</b>                                                          | <b>III</b>  |
| <b>ABSTRACT .....</b>                                                        | <b>V</b>    |
| <b>ÍNDICE GERAL.....</b>                                                     | <b>VI</b>   |
| <b>ÍNDICE DE FIGURAS.....</b>                                                | <b>VIII</b> |
| <b>ÍNDICE DE TABELAS .....</b>                                               | <b>X</b>    |
| <b>SÍMBOLOS, ACRÓNIMOS E ABREVIATURAS .....</b>                              | <b>XI</b>   |
| <br>                                                                         |             |
| <b>1. INTRODUÇÃO.....</b>                                                    | <b>1</b>    |
| 1.1 ENQUADRAMENTO GERAL .....                                                | 1           |
| 1.2 TEMA E OBJETIVOS .....                                                   | 2           |
| 1.3 ESTRUTURA DA DISSERTAÇÃO .....                                           | 2           |
| <br>                                                                         |             |
| <b>2. CONTROLO DE SEGURANÇA DE BARRAGENS DE BETÃO .....</b>                  | <b>3</b>    |
| 2.1 INTRODUÇÃO .....                                                         | 3           |
| 2.1.1 TIPOS DE BARRAGENS DE BETÃO .....                                      | 4           |
| 2.1.2 IMPORTÂNCIA DAS ATIVIDADES DE CONTROLO DE SEGURANÇA DE BARRAGENS ..... | 5           |
| 2.1.3 ACIDENTES E INCIDENTES .....                                           | 6           |
| 2.1.4 REGULAMENTO DE SEGURANÇA DE BARRAGENS .....                            | 8           |
| 2.2 OBSERVAÇÃO DAS PRINCIPAIS AÇÕES .....                                    | 8           |
| 2.2.1 PESO PRÓPRIO .....                                                     | 9           |
| 2.2.2 AÇÕES HIDROSTÁTICAS .....                                              | 9           |
| 2.2.3 AÇÕES TÉRMICAS .....                                                   | 11          |
| 2.2.4 AÇÕES DINÂMICAS.....                                                   | 12          |
| 2.3 OBSERVAÇÕES DO COMPORTAMENTO ESTRUTURAL .....                            | 12          |
| 2.3.1 DESLOCAMENTOS .....                                                    | 12          |
| 2.3.2 MOVIMENTOS DE JUNTAS .....                                             | 14          |
| 2.3.3 TENSÕES E EXTENSÕES.....                                               | 15          |
| 2.3.4 CAUDAIS .....                                                          | 16          |
| 2.3.5 ACELERAÇÕES .....                                                      | 16          |
| 2.3.6 ERROS E INCERTEZAS DE MEDIÇÃO DE GRANDEZAS MEDIDAS .....               | 17          |
| 2.4 ANÁLISE E INTERPRETAÇÃO DO COMPORTAMENTO OBSERVADO .....                 | 18          |

|           |                                                                                            |           |
|-----------|--------------------------------------------------------------------------------------------|-----------|
| 2.5       | MODELOS DE INTERPRETAÇÃO QUANTITATIVA .....                                                | 18        |
| 2.6       | IDENTIFICAÇÃO DE VALORES ATÍPICOS .....                                                    | 20        |
| <b>3.</b> | <b>MODELOS DE MACHINE LEARNING UTILIZADOS PARA IDENTIFICAÇÃO DE VALORES ATÍPICOS .....</b> | <b>23</b> |
| 3.1       | INTRODUÇÃO .....                                                                           | 23        |
| 3.2       | APRENDIZAGEM SUPERVISIONADA .....                                                          | 24        |
| 3.3       | APRENDIZAGEM NÃO SUPERVISIONADA .....                                                      | 25        |
| 3.4       | MÉTODO DE REGRESSÃO LINEAR MÚLTIPLA .....                                                  | 26        |
| 3.5       | MÉTODO K-MEANS .....                                                                       | 27        |
| 3.6       | DBSCAN .....                                                                               | 29        |
| 3.7       | MÉTODO LOF .....                                                                           | 31        |
| 3.8       | METODOLOGIAS PROPOSTAS PARA A IDENTIFICAÇÃO DE VALORES ATÍPICOS .....                      | 33        |
| <b>4.</b> | <b>CASO DE ESTUDO: BARRAGEM DO FEITICEIRO .....</b>                                        | <b>37</b> |
| 4.1       | INTRODUÇÃO .....                                                                           | 37        |
| 4.2       | BARRAGEM DO FEITICEIRO .....                                                               | 37        |
| 4.2.1     | SISTEMA DE MONITORIZAÇÃO E GRANDEZA EM ANÁLISE .....                                       | 38        |
| 4.3       | IDENTIFICAÇÃO DE VALORES ATÍPICOS COM OS MODELOS DE CLUSTERING ADOTADOS .....              | 43        |
| 4.4       | ANÁLISE DE SENSIBILIDADE DO PARÂMETRO MINPTS DOS MODELOS DBSCAN E LOF .....                | 53        |
| <b>5.</b> | <b>CONSIDERAÇÕES FINAIS .....</b>                                                          | <b>57</b> |
| 5.1       | TRABALHO REALIZADO .....                                                                   | 57        |
| 5.2       | DISCUSSÃO DOS DADOS OBTIDOS .....                                                          | 58        |
| 5.3       | DESENVOLVIMENTOS FUTUROS .....                                                             | 58        |

## ÍNDICE DE FIGURAS

|                                                                                                                                                                       |    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figura 1 - Tipos de barragens de betão: a) b) barragem gravidade, c) barragem de contrafortes e d) barragem abóbada, adaptado de (Tanchev, 2014). ....                | 5  |
| Figura 2 - Fases da vida da obra, retirado de (Cardoso, 2021). ....                                                                                                   | 5  |
| Figura 3 - Barragem Malpasset no final da construção, à esquerda, e depois de ocorrer a rotura, à direita, adaptado de (Duffaut, 2013). ....                          | 7  |
| Figura 4 - Barragem de Gleno antes e depois da catástrofe, adaptado de (Luino <i>et al</i> , 2014). ....                                                              | 7  |
| Figura 5 - Principais ações nas barragens, adaptado de (Prata <i>et al</i> , 2012). ....                                                                              | 9  |
| Figura 6 – Escalas limnimétricas a montante e a jusante da barragem, retirado de (LNEC, 2020). ....                                                                   | 10 |
| Figura 7 - Piezómetro, retirado de (Mata, 2013). ....                                                                                                                 | 11 |
| Figura 8 - Estação meteorológica, retirado de (LNEC, 2016). ....                                                                                                      | 11 |
| Figura 9 - Estação sísmica, retirado de (Mata, 2013). ....                                                                                                            | 12 |
| Figura 10 - Base de coordenómetro do fio de prumo (figura à esquerda) e telecoordinómetro (figura à direita), retirado de (LNEC, 2020). ....                          | 13 |
| Figura 11 - Antena GNSS localizada no coroamento da barragem, retirado de (LNEC, 2020). ....                                                                          | 14 |
| Figura 12 - Base de alônametro para medição manual, retirado de (Martins <i>et al</i> , 2017). ....                                                                   | 15 |
| Figura 13 - Controlador de juntas tridimensional, retirado de (LNEC, 2020). ....                                                                                      | 15 |
| Figura 14 - Grupo de extensómetros de resistência elétrica antes da betonagem, retirado de (Mata, 2013). ....                                                         | 16 |
| Figura 15 - Tensómetro de resistência antes da betonagem, retirado de (Mata, 2013). ....                                                                              | 16 |
| Figura 16 - Acelerómetro, retirado de (Mata, 2013). ....                                                                                                              | 17 |
| Figura 17- Alguns tipos de aprendizagem em <i>Machine Learning</i> . ....                                                                                             | 24 |
| Figura 18- Esquema de um problema de classificação, adaptado de (Escovedo, Koshiyama, 2020). ....                                                                     | 25 |
| Figura 19 - Representação de possíveis agrupamentos de clustering O1, O2, O3 com respetivo raio de agrupamento R1, R2 e R3 retirado de (Xu <i>et al</i> , 2022). .... | 26 |
| Figura 20 - Identificação do ponto do cotovelo, adaptado de (4 Agrupamento   Introdução ao Machine Learning (dataat.github.io)). ....                                 | 29 |
| Figura 21 - Ilustração do método de cluster DbSCAN, retirado de (Schubert <i>et al</i> , 2017). ....                                                                  | 30 |
| Figura 22 - Exemplo do parâmetro reach-distance para k =4, retirado de (Breunig <i>et al</i> , 2000). ....                                                            | 31 |
| Figura 23- Etapas do algoritmo de cluster LOF. ....                                                                                                                   | 33 |
| Figura 24 - Fluxograma da metodologia proposta para identificação de valores atípicos. ....                                                                           | 35 |
| Figura 25 - Foto e esquema de um bloco da barragem do Feiticeiro, adaptado de (Cunha <i>et al</i> , 2018). ....                                                       | 38 |
| Figura 26 - Telecoordinómetro tipo C1-5.1 e o seu <i>modus operandi</i> , adaptado de (Cunha <i>et al</i> , 2017). ....                                               | 40 |
| Figura 27 - Observações do nível da água na albufeira. ....                                                                                                           | 41 |

|                                                                                                                                  |    |
|----------------------------------------------------------------------------------------------------------------------------------|----|
| Figura 28 - Temperaturas observadas nos termómetros T11, T13 e T15.....                                                          | 42 |
| Figura 29 - Representação da posição dos fios de prumo da barragem do Feiticeiro. ....                                           | 42 |
| Figura 30 - Valores do deslocamento radial obtidos pelos instrumentos de medição automática no FP1. ....                         | 43 |
| Figura 31 - Valores do deslocamento após remoção de erros grosseiros. ....                                                       | 44 |
| Figura 32 - Limite da variação aplicado na mediana de janela móvel. ....                                                         | 44 |
| Figura 33 - Identificação de valores atípicos com base na aplicação da mediana numa janela móvel. ....                           | 45 |
| Figura 34 - Valores dos deslocamentos após aplicação da mediana numa janela móvel. ....                                          | 46 |
| Figura 35 - Efeito do nível da água e efeito térmico nos deslocamentos radiais. ....                                             | 47 |
| Figura 36 - Resíduos obtidos no modelo de Regressão Linear Múltipla.....                                                         | 47 |
| Figura 37 - Resíduos positivos e resíduos negativos representados em coordenadas polares.....                                    | 48 |
| Figura 38 - Valores atípicos identificados pelo DBSCAN (eps=0,1 e minPts=50) representados em coordenadas polares. ....          | 49 |
| Figura 39 - Valores atípicos identificados pelo método DBSCAN (eps=0,1 e minPts=50) representados nos deslocamentos radiais..... | 49 |
| Figura 40 - Valores atípicos identificados pelo método DBSCAN (eps=0,1 e minPts=50). ....                                        | 50 |
| Figura 41 - Critério de escolha do fator LOF (minPts=150). ....                                                                  | 51 |
| Figura 42 - Valores atípicos identificados pelo método LOF (minPts=150) representados em coordenadas polares. ....               | 52 |
| Figura 43 - Valores atípicos identificados pelo método LOF (minPts=150) representados nos deslocamentos radiais. ....            | 52 |
| Figura 44 - Valores atípicos identificados pelo método LOF (minPts=150).....                                                     | 53 |
| Figura 45 - Resultados numéricos DBSCAN (para eps=0,1).....                                                                      | 54 |
| Figura 46 - Resultados numéricos LOF.....                                                                                        | 54 |
| Figura 47 - Valores atípicos identificados pelo método DBSCAN (eps=0,1 e minPts=90). ....                                        | 56 |
| Figura 48 - Valores atípicos identificados pelo método LOF (minPts=10). ....                                                     | 56 |

## ÍNDICE DE TABELAS

|                                                                                                                             |    |
|-----------------------------------------------------------------------------------------------------------------------------|----|
| Tabela 1 - Métodos ou equipamentos de medição e precisões de recolha dos dados, adaptado de (Silva,2022a). ....             | 17 |
| Tabela 2 - Sistema de monitorização da barragem do Feiticeiro, adaptado de (Cunha <i>et al</i> , 2018). ....                | 39 |
| Tabela 3 - Características dos telecoordinómetros da barragem do Feiticeiro, adaptado de (Cunha <i>et al.</i> , 2017). .... | 41 |
| Tabela 4 - Valores numéricos da aplicação da mediana com janela móvel.....                                                  | 45 |
| Tabela 5 - Resultados obtidos DBSCAN.....                                                                                   | 55 |
| Tabela 6 - Resultados obtidos LOF. ....                                                                                     | 55 |

## SÍMBOLOS, ACRÓNIMOS E ABREVIATURAS

FEUP - Faculdade de Engenharia da Universidade do Porto  
DEC - Departamento de Engenharia Civil  
LNEC - Laboratório Nacional de Engenharia Civil  
APA - Agência Portuguesa do Ambiente  
ANEPC - Autoridade Nacional de Emergência e Proteção Civil  
CSB - Comissão de Segurança de Barragens  
DBSCAN - Density-Based Spatial Clustering of Applications Noise  
LOF - Local Outlier Factor  
RAD - Recolha automática de dados  
HST - Hydrostatic, Seasonal, Time  
HTT - Hydrostatic, Thermal, Time  
RSB - Regulamento de Segurança de Barragens  
NPA - Nível pleno de armazenamento  
NmE - Nível mínimo de exploração  
NMC - Nível máxima de cheia  
LVDT - *Linear Variable Differential Transformer*  
a.C. - Antes de Cristo  
d.C. - Depois de Cristo  
PP- Peso próprio  
PH - Pressão hidrostática  
S - Subpressão  
SMP - Sismo máximo de projeto  
h - Altura de água na albufeira  
kN - Quilo-Newton  
m - Metros  
mm - Milímetros  
mA - miliampere  
°C - Graus Celsius  
s - Segundos  
 $\gamma$  - Peso específico do material  
V - Volume

$\delta(h_i, S_i, t_i)$  - Valor observado da grandeza em análise na observação  $i$

$U_h(h_i), U_\theta(S_i), U_t(t_i)$  - Parcelas das grandezas correspondentes ao efeito elástico do nível de água na albufeira, ao efeito elástico da variação sazonal da temperatura e ao efeito do tempo na  $i^{ésima}$  observação

$k$  - Constante que corresponde à diferença entre valores observados e calculados no início do período de calibração

$\varepsilon_i$  - Resíduo da  $i^{ésima}$  observação dado pela diferença entre o valor estimado e o valor observado

$t_d$  - Número de dias decorridos desde o início do ano até à data da observação

$t_m$  - Número de minutos decorridos desde o início do dia até à data da observação

$t$  - Número de dias entre a campanha e o início da observação

$\alpha$  - Número de dias entre o primeiro enchimento e a data do início da análise

$X_j$  -  $j^{ésimo}$  termo associado às ações

$\beta_j$  - Coeficiente de regressão correspondente ao termo  $X_j$

SSE - soma total do quadrado dos resíduos

FP - Fio de prumo







# 1

## INTRODUÇÃO

### 1.1 ENQUADRAMENTO GERAL

A segurança e a estabilidade das barragens de betão são de extrema importância para evitar tragédias e danos ambientais. A monitorização contínua dessas estruturas é essencial para identificar possíveis anomalias ou comportamentos não esperados, tais como falhas estruturais, desgastes excessivos ou alterações nas propriedades físicas dos materiais, permitindo um prolongamento da sua vida útil. Nas últimas décadas, o avanço da tecnologia tem possibilitado a aplicação de técnicas de *Machine Learning* no campo da engenharia civil. A utilização desses modelos computacionais permite extrair informações valiosas e identificar padrões complexos a partir de grandes conjuntos de dados, auxiliando na análise e tomada de decisões relacionadas à segurança das barragens.

Neste contexto, a aplicação de técnicas de *Machine Learning* pode desempenhar um papel fundamental na identificação precoce de anomalias que possam aparecer em barragens de betão, permitindo uma resposta rápida e eficaz para prevenir acidentes e minimizar riscos. A utilização dessas técnicas pode contribuir para uma melhoria dos métodos tradicionais de monitorização e inspeção visual, fornecendo uma abordagem complementar baseada em análise de dados e algoritmos de aprendizagem automática.

A utilização de técnicas de *Machine Learning* não substitui as inspeções visuais e outras formas de monitorização tradicionais, mas sim complementa esses métodos, fornecendo uma visão abrangente e baseada em dados relativos ao comportamento estrutural das barragens. A combinação dessas abordagens pode resultar num sistema de monitorização mais robusto e fiável, capaz não só de garantir a segurança da barragem, mas também de proteger vidas e preservar o meio ambiente envolvente.

No campo da análise de dados, a deteção e a identificação de novidades apresentam um papel crucial na descoberta de informações valiosas e na tomada de decisões fundamentadas. Essas novidades são observações que se desviam significativamente do padrão esperado num conjunto de dados, e sua presença pode distorcer análises estatísticas e prejudicar a qualidade da análise dos resultados.

## 1.2 TEMA E OBJETIVOS

Esta dissertação apresenta como objetivo o estudo de diferentes abordagens de técnicas de *Machine Learning* na identificação de valores atípicos no comportamento de barragens de betão, considerando dados obtidos através de instrumentação instalada. Serão considerados diferentes algoritmos de *Machine Learning* e métodos de deteção de anomalias, com o intuito de desenvolver modelos capazes de identificar com precisão possíveis erros nas medições ou comportamentos atípicos nas barragens. O tema desta dissertação surge como resultado dos desenvolvimentos observados em estudos anteriores sobre a aplicação de métodos desta área no controlo de segurança estrutural de barragens, sendo esta uma área em constante evolução, não apenas dentro do campo da engenharia.

Espera-se que este trabalho contribua para o avanço da identificação de novidades em barragens de betão, fornecendo uma abordagem complementar à monitorização tradicional e ajudando nas atividades de controlo e segurança e na redução dos riscos associados a essas estruturas críticas. Os resultados obtidos podem ser aplicados na implementação de sistemas de alerta precoce e manutenção preditiva, proporcionando um maior nível de segurança e confiabilidade a essas infraestruturas essenciais.

## 1.3 ESTRUTURA DA DISSERTAÇÃO

A presente dissertação está organizada em cinco capítulos, ordenados de forma a serem apresentados conceitos a serem usados ao longo da sua extensão. O capítulo 1 faz uma breve introdução ao tema, onde se apresenta o enquadramento geral e os objetivos que motivaram a realização deste trabalho, estando presente também neste capítulo a estrutura da dissertação.

No capítulo 2 realiza-se um breve enquadramento histórico da evolução e importância das barragens, assim como os tipos de barragens de betão mais usuais, apresentando a suas características principais. É feita uma referência à importância da monitorização e do controlo da segurança das barragens, apresentando os regulamentos de segurança presentes atualmente em Portugal e alguns casos históricos de acidentes com barragens. Segue-se uma abordagem às principais ações presentes numa barragem de betão bem como o seu comportamento estrutural, assim como técnicas e aparelhos de observação das grandezas presentes. Neste capítulo é ainda abordada a análise e interpretação do comportamento observado na barragem, fazendo referência a modelos de interpretação quantitativa e a identificação de valores atípicos.

O capítulo 3 introduz os métodos de *Machine Learning* aplicados neste estudo, introduzindo os conceitos fundamentais para o funcionamento dos mesmos, o seu modo de funcionamento, e explicando o seu *modus operandi* e modo de aplicação. Este capítulo apresenta também a metodologia proposta para a identificação de valores atípicos, descrevendo as etapas utilizadas neste estudo, baseadas nos métodos de *Machine Learning*.

No capítulo 4 introduz-se o caso de estudo, faz-se uma descrição da barragem onde vai ser aplicado o estudo, apresentando as suas características e os aparelhos de monitorização utilizados. Neste capítulo é retratada a metodologia proposta, apresentando graficamente os modelos obtidos pelos métodos de *Machine Learning*, realçando aspetos relevantes nas etapas utilizadas.

Por último, o capítulo 5 apresenta as considerações finais desta dissertação, fazendo uma retrospectiva dos resultados obtidos e apresenta potenciais propostas para trabalhos futuros.

## 2

### CONTROLO DE SEGURANÇA DE BARRAGENS DE BETÃO

#### 2.1 INTRODUÇÃO

A água é um elemento imprescindível na vida humana, quer para consumo, quer para produção de alimentos e de energia. Desde cedo as civilizações adaptaram o seu quotidiano em volta deste recurso, tentando armazenar e potenciar o seu uso, minimizando os riscos associados. A solução encontrada foi a construção de estruturas com capacidade de armazenamento e proteção contra inundações.

Acredita-se que a primeira barragem do mundo tenha sido construída na antiga mesopotâmia, em cerca de 3000 a.C., com cerca de 5 metros de altura e 100 metros de comprimento, usando terra compactada e argila, com o intuito de controlar as inundações do rio Tigre e de armazenar água para irrigação. Ao longo da história, a capacidade e a tecnologia construtiva das barragens foram evoluindo dando origem a barragens como a de Sadd-el-Kafara, construída no Egito em 2600 a.C., a barragem de Zeugma construída na Turquia em 300 d.C pelos romanos, a barragem de Hoover em 1936, nos Estados Unidos da América e mais recentemente a barragem das Três Gargantas, na China em 2003, com 2335 m de extensão e 181 m de altura, sendo considerada a maior barragem hidroelétrica alguma vez construída. (Schnitter, 1994; Gleick *et al*, 2008).

Foram catástrofes, como as sucedidas em Banqiao (Zipper, 2020) e Malpasset (Carmo, 2013), que destacam a importância das atividades de controlo de segurança e monitorização das barragens. Em todas as obras em exploração, as exigências regulamentares de segurança devem ser verificadas, tornando-se cada vez mais importante a utilização de rigorosas técnicas de observação e ainda o desenvolvimento de modelos de referência credíveis para a interpretação e a previsão do seu comportamento com o intuito de se tomarem, quando necessárias, decisões referentes às medidas a implementar, em termos de exploração, conservação, reabilitação ou demolição (Mata, 2007).

A monitorização e controlo da segurança numa barragem é um processo multidisciplinar e multivariado, uma vez que são diversas as grandezas monitorizadas e as análises realizadas envolvem o conhecimento de múltiplas disciplinas como engenharia de estruturas, engenharia de materiais, geologia, hidráulica, química, etc. Atualmente a automatização das medições, embora tenha sido implementada há vários anos em algumas obras em Portugal, ainda não está plenamente estabelecida, existindo erros de medição que devem ser eliminados previamente para garantir a utilidade dos dados, o que compromete a eficiência da resposta no controlo de segurança em tempo útil.

### 2.1.1 TIPOS DE BARRAGENS DE BETÃO

As barragens podem ser classificadas conforme o material usado na sua construção e a sua forma. Dependendo das suas características, as barragens de betão podem ser classificadas, por exemplo, como barragens de gravidade, barragens de contrafortes ou barragens abóbada, conforme apresentado na Figura 1.

As barragens de gravidade são essencialmente uma estrutura de betão que é capaz de assegurar um bom comportamento devido à sua grande massa, uma vez que os principais mecanismos de rotura (deslizamento e derrubamento), são evitados, com um nível de segurança adequado, graças ao peso próprio da barragem e da respetiva mobilização do atrito. É uma estrutura com geometria trapezoidal, com a parte inferior da barragem mais espessa que a parte superior, em que o eixo geralmente corresponde a uma linha reta, ainda que por vezes, para tirar vantagem das condições topográficas locais podem apresentar curvatura no sentido de montante (barragens de arco gravidade) ou uma combinação de curvas e retas. No geral, as barragens de betão são dimensionadas de forma que as tensões sejam maioritariamente de compressão, uma vez que o betão é um material que apresenta um bom comportamento para esse tipo de tensões. As barragens gravidade são usualmente adotadas em vales suaves que apresentam boas fundações rochosas (Asthana, Khare, 2022).

Nas barragens de contrafortes a estrutura resistente é constituída por blocos isolados, concretizando um tipo de aligeiramento da estrutura gravidade maciça mais acentuado que o vazamento; teoricamente esses blocos funcionam isoladamente, permitindo a adaptação a condições geotécnicas diferenciadas ao longo da fundação. Geralmente, as barragens de contrafortes são adequadas para regiões com topografia suave, caracterizadas por vales amplos e elevações relativamente baixas (Rijo, 2007).

As barragens abóbadas apresentam um eixo curvo desenvolvendo um arco. Os esforços provocados pela pressão exercida pela água da albufeira são transmitidos aos encontros e às fundações através de forças de compressão desenvolvidas nos arcos. As barragens abóbadas são usadas normalmente em vales estreitos em que o maciço rochoso da encosta do vale funciona também como apoio. A vantagem deste tipo de barragem é a redução na quantidade de betão em relação às barragens de gravidade e contrafortes, mas em contrapartida requiere uma técnica construtiva mais complexa (Asthana, Khare, 2022). Os efeitos de deterioração do betão podem apresentar uma importância ainda maior nas barragens abóbadas em relação às barragens de gravidade, uma vez que a integridade do betão é fundamental para garantir a estabilidade e a transmissão de cargas da barragem para a fundação (Rijo, 2007; Cardoso, 2021).

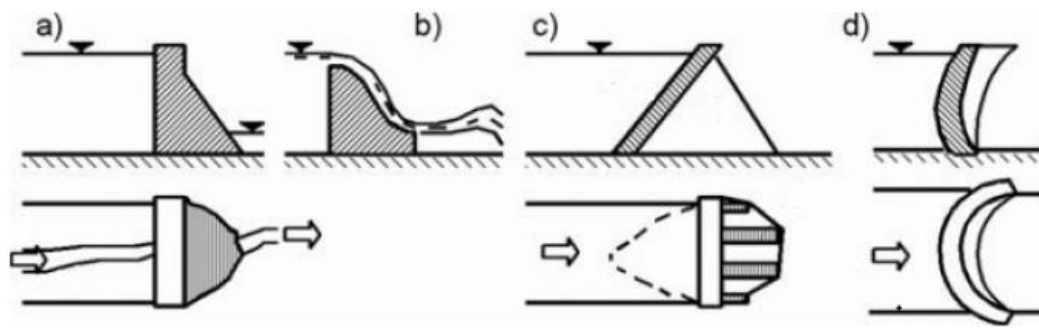


Figura 1 - Tipos de barragens de betão: a) b) barragem gravidade, c) barragem de contrafortes e d) barragem abóbada, adaptado de (Tanchev, 2014).

### 2.1.2 IMPORTÂNCIA DAS ATIVIDADES DE CONTRO DE SEGURANÇA DE BARRAGENS

Como qualquer obra de arte em engenharia civil, as barragens passam por uma sequência de processos até ficarem totalmente concluídas (Figura 2). Durante todos esses processos até ao fim de vida da estrutura existem atividades de monitorização e controlo de segurança. Não existem registos em Portugal de tragédias ligadas a colapsos de barragens, muito devido ao importante trabalho destas atividades, capaz de identificar atempadamente possíveis cenários de deterioração.

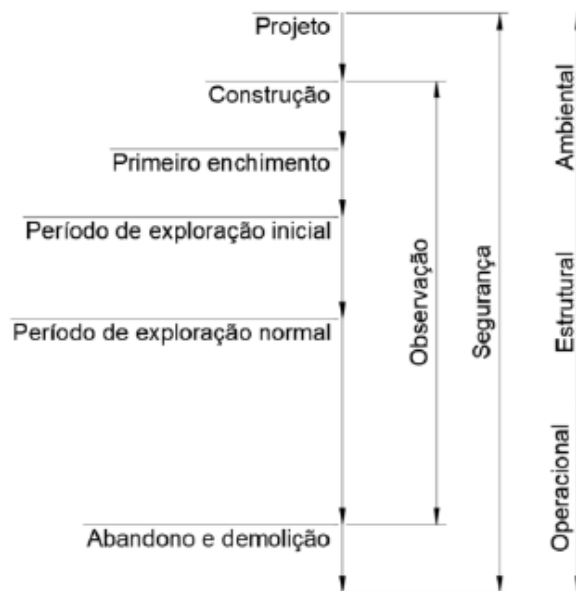


Figura 2 - Fases da vida da obra, retirado de (Cardoso, 2021).

A monitorização conta com diversos instrumentos de medição que permitem obter dados sobre a temperatura do betão e do ar, nível da água na albufeira e a jusante da barragem, deslocamentos e rotações, movimentos de juntas, deformações, tensões, pressões, acelerações e caudais na fundação. O controlo de segurança de uma barragem requer uma ampla gama de informações importantes provenientes do sistema de monitorização e de inspeções visuais. As atividades de controlo de

segurança devem permitir a deteção oportuna de qualquer comportamento anómalo da barragem, que a possa deteriorar, a fim de implementar medidas corretivas (Mata *et al*, 2014).

O controlo de segurança deve ser verificado nos seguintes aspetos:

- Segurança ambiental;
- Segurança estrutural;
- Segurança hidráulico/operacional.

A segurança ambiental refere-se aos efeitos ecológicos e ambientais que a construção de uma barragem pode introduzir no ecossistema local. Devem ser feitos controlos ambientais como a qualidade da água da albufeira, do solo e garantir o caudal ecológico no leito do rio a jusante da barragem. A segurança estrutural está relacionada com a estabilidade estrutural e o comportamento da barragem ao longo do seu tempo de vida. É expectável que a obra seja estável e tenha um bom comportamento face a todas as ações que está sujeita. O controlo de segurança incluía assim, a monitorização da barragem bem como todas as medidas e intervenções aplicadas. A segurança hidráulico/operacional relaciona-se com a adequada exploração e com o bom funcionamento dos órgãos de segurança, assim como com a implementação de sistemas de aviso e alerta para os casos de emergência (Mata, 2007), de modo a garantir a segurança de pessoas e bens no vale a jusante.

### 2.1.3 ACIDENTES E INCIDENTES

Em Portugal existem atualmente cerca de 260 grandes barragens, não existindo registo de incidentes graves em barragens de betão (apambiente.pt). Considerando os elevados danos que uma eventual rotura pode provocar, em termos de perdas de vidas humanas e de bens económicos e ambientais, as barragens são classificadas como estruturas de alto risco, onde as atividades de monitorização e controlo de segurança são muito importantes.

Ao longo da história foram registados alguns acidentes com rotura de barragens com diversas origens, tais como:

- Deficiências construtivas;
- Atos de sabotagem e operações militares;
- Exploração inadequada da albufeira;
- Ocorrência de catástrofes naturais;
- Degradação da estabilidade ou resistência do corpo da barragem, das fundações e/ou ainda das encostas da albufeira.

Estudos estatísticos revelam que o período crítico da vida de uma barragem ocorre durante a sua construção ou no primeiro enchimento completo da albufeira, prolongando-se até 5 a 7 anos após o enchimento da albufeira. Contudo, importa notar que esta conclusão é limitada, pois refere-se fundamentalmente a deficiências construtivas, pois qualquer uma das outras situações pode ocorrer com igual probabilidade durante a vida útil de uma barragem (Lourenço, Mateus, 2013).

Um dos casos de rotura mais conhecidos é a rotura da barragem de Malpasset (Figura 3), em França. Nos primeiros anos após a sua construção, o enchimento da albufeira foi comprometido devido à expropriação de uma mina de fluorite a montante da barragem, tendo posteriormente esta albufeira ficado sem utilização devido à não conclusão da rede de distribuição de água. Em 1959 após dias de chuvas intensas, o nível de água na albufeira da barragem de Malpasset aumentou muito rapidamente,



levando-a ao colapso, provavelmente causada por uma falha tectônica na base do maciço de fundação, insuficientemente caracterizada, e pela consequente instalação de subpressões excessivas na fundação. A construção de estradas nas proximidades, com recurso a explosivos, pode também ter contribuído para esse incidente. Neste incidente foram contabilizadas 433 mortes e inúmeros feridos e grandes danos materiais (Carmo, 2013; Duffaut, 2013).

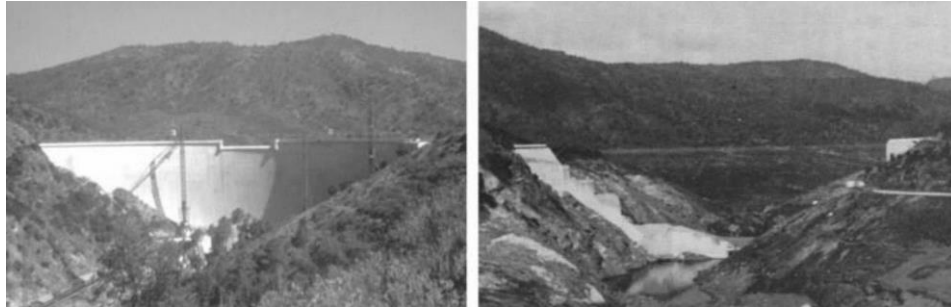


Figura 3 - Barragem Malpasset no final da construção, à esquerda, e depois de ocorrer a rotura, à direita, adaptado de (Duffaut, 2013).

Outra ocorrência foi o caso de Gleno, Itália em 1923, ilustrada na Figura 4. A barragem de Gleno é um caso raro de uma barragem que foi aumentada em altura com uma estrutura constituída por múltiplos arcos/gravidade. Era composta por uma parede de pedra fechando o estreito desfiladeiro e uma serie de 25 arcos de betão armado, apoiados em 26 contrafortes. O projeto inicial previa a construção de uma barragem de gravidade com parede de 30 m a 40 m de altura, no entanto, durante a construção foi adotada, uma mistura entre gravidade e sistema de múltiplos arcos. Esta escolha significou poupança no volume de material utilizado, com considerável redução de custos. Foram verificadas, várias fugas de água e inspetores do Conselho de Engenharia Civil foram chamados para uma investigação, mas durante uma verificação de rotina foi sentido o primeiro tremor da barragem, levando ao colapso da mesma pouco tempo depois. Foram contabilizadas 356 mortes nessa catástrofe (Luino *et al*, 2014).



Figura 4 - Barragem de Gleno antes e depois da catástrofe, adaptado de (Luino *et al*, 2014).

Estes dois exemplos demonstram que mesmo com regulamentos e controlos, os desastres podem acontecer, tanto de forma repentina como de forma mais previsível. Casos como este são cada vez

mais raros, pois a tecnologia e a metodologia de construção foi evoluindo ao longo do tempo, permitindo um melhor conhecimento do comportamento da barragem e dos possíveis tipos de rotura que esta pode ter.

#### 2.1.4 REGULAMENTO DE SEGURANÇA DE BARRAGENS

De modo a minimizar ou mesmo evitar os acidentes, existem regulamentos que devem ser cumpridos em todas as fases da vida da obra: projeto, construção, exploração e demolição. Em Portugal existem dois regulamentos a cumprir, o Regulamento de Segurança de Barragens e o Regulamentos (RSB) de Pequenas Barragens, publicados em anexo ao Decreto-Lei n.º 21/2018, de 28 de março.

Estes regulamentos estabelecem regras para o projeto, construção e exploração das barragens, programas de inspeção, observação e manutenção, planos de emergência e a qualificação técnica adequada dos agentes envolvidos.

As principais entidades públicas e privadas reconhecidas no RSB são as seguintes:

- Dono de obra;
- Autoridade Nacional de Segurança de Barragens;
- Laboratório Nacional de Engenharia Civil;
- Autoridade Nacional de Emergência e Proteção Civil;
- Comissão de Segurança de Barragens.

Os donos de obra são as entidades responsáveis pela aplicação dos regulamentos de segurança de barragens. É sob a sua responsabilidade que atuam outros agentes, como é o caso dos projetistas, empreiteiros, empresas de fiscalização, consultores, etc. A APA, como Autoridade Nacional de Segurança de Barragens, tem a competência genérica do controlo de segurança das barragens, fiscalizando e promovendo o cumprimento dos regulamentos. Outras entidades da Administração Pública têm um papel neste âmbito regulamentar: o Laboratório Nacional de Engenharia Civil (LNEC), que apoia o controlo de segurança das barragens, como consultor da Autoridade; a Autoridade Nacional de Emergência e Proteção Civil (ANEPC), que orienta e coordena atividades de proteção civil nacionais e a Comissão de Segurança de Barragens (CSB), que de forma independente analisa a segurança das barragens portuguesas, com o suporte da APA ([apambiente.pt](http://apambiente.pt)).

## 2.2 OBSERVAÇÃO DAS PRINCIPAIS AÇÕES

As barragens de betão estão sujeitas a muitas ações em simultâneo durante o seu tempo de vida, com especial destaque para as seguintes (Figura 5):

- Peso próprio;
- Ações hidrostáticas;
- Ações térmicas;
- Ações dinâmicas.

Existem ainda outras ações, que são menos comuns devido à sua excecionalidade e à baixa probabilidade de acontecerem em Portugal, como é o caso de explosões, da ação do gelo, incluindo o impacto de placas de gelo no corpo da barragem e o deslizamento de terra (Silva, 2022a).

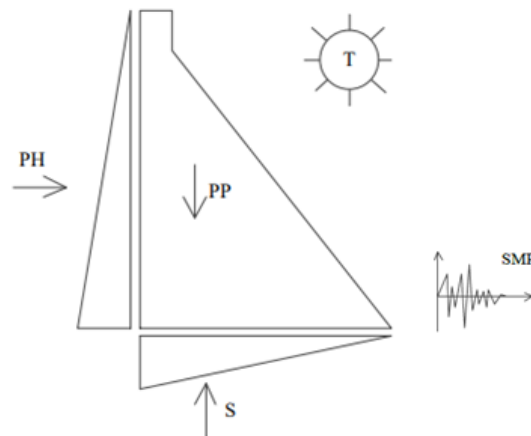


Figura 5 - Principais ações nas barragens, adaptado de (Prata et al, 2012).

### 2.2.1 PESO PRÓPRIO

O peso próprio é devido à força da gravidade, sendo a sua magnitude determinada pela densidade dos materiais que constituem a estrutura. No caso de barragens de betão, como o nome indica, o material predominante é o betão, que apresenta um peso específico entre  $\gamma_c = 23$  a  $25 \text{ kN/m}^3$ , dependendo do tipo de agregado (Chen e Chen, 2015).

O peso próprio de um elemento estrutural pode ser determinado, pode ser determinado pela multiplicação do peso específico do material usado pelo volume desse mesmo material (Equação (1)).

$$PP = \gamma * V \quad (1)$$

Onde:

- $PP$  é o peso próprio [kN]
- $\gamma$  é o peso específico do material [ $\text{kg/m}^3$ ]
- $V$  é o volume do material [ $\text{m}^3$ ]

### 2.2.2 AÇÕES HIDROSTÁTICAS

Designam-se por ações hidrostáticas as ações mecânicas resultantes da presença da água na albufeira. Dependendo da zona onde estas são aplicadas, podem ser denominadas por pressão hidrostática, caso sejam aplicadas sobre os paramentos de montante e de jusante, e por subpressão, no caso das pressões resultantes do escoamento hidráulico pela fundação das barragens.

A ação da pressão hidrostática ao longo do tempo de vida da barragem, varia com a cota da água da albufeira, medidas em limnímetros ou em escalas limnimétricas (Figura 6). A variação entre os níveis de água a montante e a jusante depende essencialmente dos regimes de exploração da albufeira e das condições pluviométricas. A ação da água tem uma importância fulcral na segurança estrutural, podendo dar origem a ações mecânicas ou químicas, como, por exemplo:

- Pressões hidrostáticas nos paramentos da barragem;
- Pressões internas em fissuras das barragens de betão e em superfícies de compartimentação da fundação, dando origem ao aparecimento de subpressões;

- Forças devidas à percolação da água na fundação da barragem;
- Modificação química dos materiais constituintes da fundação;
- Erosão dos materiais e de dispositivos presentes na barragem;
- Deterioração dos maciços rochosos devido à percolação da água pelas descontinuidades;

A ação da pressão hidrostática deve ser considerada para as seguintes quatro situações que enquadram a exploração da albufeira (da Silva, 2013):

- Nível de Máxima Cheia;
- Nível de Pleno Armazenamento;
- Nível mínimo de Exploração Normal;
- Nível mínimo de Exploração Extraordinário.

As subpressões são esforços exercidos numa estrutura ou na sua fundação, devido à percolação da água através do maciço de betão. As situações em que a subpressão é mais gravosa ocorrem em secções fendilhadas das barragens, na interface entre a barragem e o maciço de fundação ou em descontinuidades da fundação. No interior do betão essas pressões são medidas em medidores de pressão embebidos no betão. No caso dos maciços de fundação e da interfase entre a barragem e a fundação, são instalados piezómetros (Figura 7) em furos especificamente efetuados para esse efeito. A pressão aplicada sobre a estrutura varia ao longo do tempo, dependendo fundamentalmente da cota da água a montante e a jusante e da permeabilidade dos materiais (Loureiro, 2018).

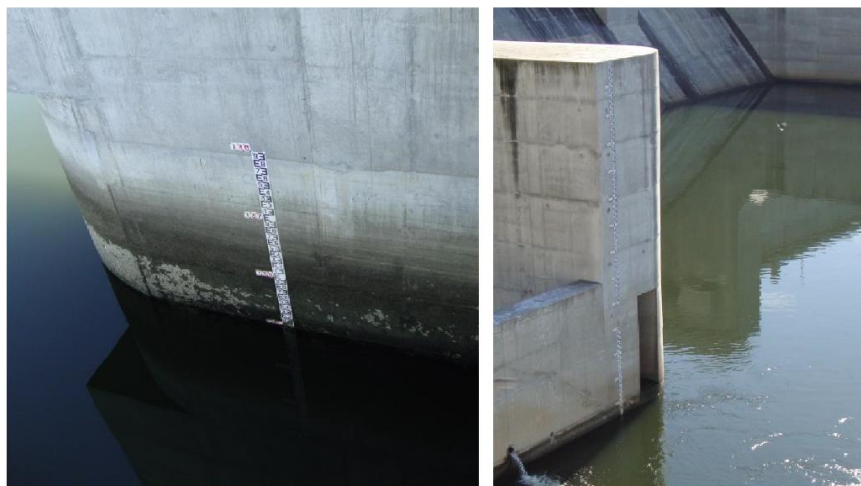


Figura 6 – Escalas limnimétricas a montante e a jusante da barragem, retirado de (LNEC, 2020).



Figura 7 - Piezómetro, retirado de (Mata, 2013).

### 2.2.3 AÇÕES TÉRMICAS

As barragens operam sob condições de temperatura ambiente variável. Como resultado, a tensão de expansão ou retração ocorre por mudanças de temperatura e humidade dentro do betão. As mudanças de temperatura são chamadas de ação térmica ou carga térmica, pois indiretamente provoca deformações e tensões no betão das estruturas. A ação térmica depende do tipo de estrutura, propriedades do material e do modo de construção incluindo o procedimento de aquecimento ou arrefecimento artificial, troca térmica entre a estrutura e a fundação, e flutuações da temperatura ambiente (água, ar) durante o período de construção e operação (Chen e Chen, 2015).

Os valores da temperatura do ar podem ser medidos através de termómetros de máxima e mínima, ou através de estações meteorológicas (Figura 8). A temperatura da água da albufeira é medida através de sondas termométricas, sendo feitas medições a várias profundidades. No caso das temperaturas no interior do betão, são medidas através de instrumentos de resistência elétrica embebidos no betão (Mata, 2007).



Figura 8 - Estação meteorológica, retirado de (LNEC, 2016).

#### 2.2.4 AÇÕES DINÂMICAS

Entendem-se por ações dinâmicas numa barragem de betão, ações como sismos e as vibrações causadas pelo funcionamento da barragem. Estas ações são devem ser consideradas no projeto de estruturas e avaliadas em regiões com risco sísmico, e monitorizadas através de acelerómetros e estações sísmicas (Figura 9).

Até aos dias de hoje não houve registo, em Portugal, de uma rotura devida a ações dinâmicas, mas estas não podem ser ignoradas, pois em casos que apresentam acelerações elevadas conseguem causar danos significativos à estrutura.



Figura 9 - Estação sísmica, retirado de (Mata, 2013).

### 2.3 OBSERVAÇÕES DO COMPORTAMENTO ESTRUTURAL

A observação das grandes estruturas é uma atividade planificada, onde a partir do comportamento observado durante a fase de entrada em serviço, deverão ser estabelecidos valores de referência. Durante o período de exploração, nos primeiros cinco anos, esses valores são reajustados caracterizando assim o comportamento normal da estrutura, tornando-se os valores de referência a longo prazo (Marcelino *et al*, 2008).

As respostas estruturais que são usualmente observadas nas barragens de betão são:

- Deslocamentos;
- Movimentos em juntas;
- Tensões e extensões no betão;
- Caudais drenados e infiltrados;
- Acelerações.

#### 2.3.1 DESLOCAMENTOS

Os deslocamentos podem ser verticais e horizontais. Estes são medidos em pontos previamente identificados ao longo da barragem, sendo obtidos valores absolutos ou relativos. Ao serem obtidos deslocamentos em vários pontos, podemos retirar uma perspetiva dos deslocamentos globais da barragem.

Para a determinação dos deslocamentos horizontais em barragens são usados fios de prumo (Figura 10) e métodos geodésicos (Figura 11). Os métodos geodésicos são usados como meio de

complementar a medição de deslocamentos, geralmente obtidos por fios de prumo, fornecendo um panorama geral da relação entre a obra, a sua fundação e o restante maciço. No caso da aplicação em barragens de betão, os sistemas de observação geodésicos são normalmente compostos por dois sistemas independentes, planimétrico e altimétrico. O planeamento de um sistema de observação geodésica engloba diversas etapas, desde a seleção com base na morfologia do terreno, quantidade, localização e acessibilidade dos pontos-alvo. Em seguida, determina-se a localização e a forma de materialização desses pontos, bem como as grandezas a serem observadas e os equipamentos utilizados (Fernandes, 2016). Os sistemas permitem obter deslocamentos verticais (altimetria) e horizontais (planimetria), com precisão de 0,2 mm e 1 mm, respetivamente (Morais, 2017). O método mais usado para a medição de deslocamentos horizontais é a triangulação, que consiste na medição de direções horizontais, ângulos verticais e distâncias dos pontos de interesse em relação a uma direção conhecida, tomada como referência (Siguel *et al*, 2013).

Os fios de prumo são instalados em furos verticais que atravessam o corpo da obra, onde um fio de aço de alta resistência é fixado num ponto a cota elevada da estrutura (fio de prumo direito), ou fixado num ponto profundo da fundação (fio de prumo invertido). No primeiro caso, o fio, é fixado num ponto da obra, a uma cota elevada, enquanto na extremidade oposta tem um peso de cerca de 600 N, a tensionar o fio. Os deslocamentos obtidos nos vários pontos de acesso ao fio de prumo (por exemplo, em galerias ou plataformas) são definidos relativamente ao ponto de fixação a cota elevada. No segundo caso, o fio é tensionado por ação da impulsão da água contida num reservatório sobre um flutuador nela mergulhado e permite obter nos pontos onde é acessível os deslocamentos relativos ao ponto profundo da fundação que, desde que tal ponto seja considerado fixo, correspondem a deslocamentos absolutos. A instalação dos fios de prumos no corpo da barragem é efetuada em zonas onde a informação é mais pertinente em termos de avaliação do comportamento como, por exemplo, no caso das barragens tipo abóbada, próximo dos encontros, no fecho do arco e nas zonas dos rins (Martins *et al*, 2011) (Mata, 2007).



Figura 10 - Base de coordinómetro do fio de prumo (figura à esquerda) e telecoordinómetro (figura à direita), retirado de (LNEC, 2020).





Figura 11 - Antena GNSS localizada no coroamento da barragem, retirado de (LNEC, 2020).

### 2.3.2 MOVIMENTOS DE JUNTAS

A medição do movimento de abertura-fecho e deslizamento em descontinuidades estruturais, por exemplo, juntas associadas a blocos de barragens de betão, fissuras e descontinuidades em maciços rochosos, é feita com recurso a instrumentos de medição embebidos no betão ou à superfície.

A abordagem clássica para a medição do movimento de abertura-fecho e deslizamento recorre a um alongâmetro, que é um instrumento de natureza mecânica que assegura a medição manual da distância entre pontos de referência em obra, onde se forma um triângulo equilátero com o intuito de determinar os deslocamentos relativos entre os vértices, através do estabelecimento de relações geométricas (Silva, 2022a), Figura 12. Para medições manuais são usados equipamentos fixos à superfície como bases tridimensionais que permitem a leitura da abertura ou fecho das juntas e de deslizamentos relativos nas direções horizontal e vertical.

Nos últimos anos assistiu-se à crescente utilização adicional de sistemas automáticos (Figura 13) suportados em cadeias de medição elétricas compostas por sensores de posicionamento linear associados a elementos de alimentação, condicionamento, conversão e transmissão de sinal.

No método de medição automática são usadas cadeias de medição compostas usualmente por sensores de posicionamento linear, que são sensores elétricos do tipo indutivo, dispondo de uma relação de transdução associada ao fenómeno de variação da indução mútua designando-se, para estes casos, por LVDT – *Linear Variable Differential Transformer*. Este método possibilita uma observação permanente da obra, sem ser necessário dispor de recursos humanos (Martins *et al*, 2017).





Figura 12 - Base de alônametro para medição manual, retirado de (Martins *et al*, 2017).



Figura 13 - Controlador de juntas tridimensional, retirado de (LNEC, 2020).

### 2.3.3 TENSÕES E EXTENSÕES

As extensões podem ser medidas em extensómetros (extensómetros de resistência elétrica tipo Carlson), estes aparelhos obtêm informação sobre o estado de deformação do betão em pontos interiores da obra. Os extensómetros de resistência elétrica tipo Carlson (Figura 14) operam com base na variação da resistência elétrica de um conjunto de condutores quando o instrumento é submetido a uma deformação (Portela, Silva, 1996).

A observação do estado de tensão é feita normalmente de forma indireta, usando as extensões obtidas através dos extensómetros. As componentes de compressão das tensões podem ser diretamente obtidas através de tensómetros (Figura 15), que são usualmente colocados perto dos extensómetros para comparação de valores com o cálculo das tensões provenientes dos extensómetros (Moura, 2005).



Figura 14 - Grupo de extensómetros de resistência elétrica antes da betonagem, retirado de (Mata, 2013).



Figura 15 - Tensómetro de resistência antes da betonagem, retirado de (Mata, 2013).

#### 2.3.4 CAUDAIS

A medição de caudais pode ser usada na deteção de aumento dos caudais drenados ou infiltrados no corpo da barragem ou na fundação. A medição é feita através de uma rede de drenagem, onde são instalados tubos em furos abertos na fundação a partir de galerias de drenagem que recolhem águas que escoam pelo maciço rochoso. O valor do caudal é obtido através de um cálculo simples contabilizando o tempo que demora a encher um reservatório de volume conhecido. O encaminhamento dos caudais é feito para bicas totalizadoras que permitem a quantificação dos caudais drenados e infiltrados em zonas previamente estabelecidas.

O controlo dos caudais drenados e infiltrados é de grande importância, sendo um dos principais elementos do controlo de segurança, e em conjugação com os piezómetros, permite avaliar a eficiência dos sistemas de impermeabilização e drenagem da obra (Mata, 2007).

#### 2.3.5 ACELERAÇÕES

As vibrações presentes nas barragens são causadas por condições do ambiente circundante, como sismos e pequenos tremores induzidos pelo peso da albufeira, a ação do vento e tráfego rodoviário próximo ou no coroamento da barragem, assim como pelo normal funcionamento do aproveitamento hidroelétrico subjacente à albufeira, quando existente (Pereira, 2019). Embora não se tenham registado acidentes devido a este tipo de ações, os sistemas de monitorização baseado em vibrações estão a ser cada vez mais usados em grandes estruturas de engenharia civil, com o intuito de caracterizar a evolução do seu comportamento ao longo do tempo e de identificar alterações resultantes da degradação da estrutura.

Os sistemas de monitorização contínua, são normalmente compostos por acelerómetros uniaxiais (Figura 16) dispostos na horizontal ao longo das galerias da barragem, permitindo o registo de séries temporais de acelerações e posterior caracterização de níveis de vibração e de parâmetros modais (frequências naturais, razões de amortecimento e formas modais). Por sua vez, a monitorização da ação sísmica é efetuada através de sismógrafos triaxiais, colocados na zona circundante da albufeira, permitindo não só a caracterização da ação sísmica, como a sua propagação ao longo do maciço rochoso em várias direções (Pereira *et al*, 2018).



Figura 16 - Acelerómetro, retirado de (Mata,2013).

### 2.3.6 ERROS E INCERTEZAS DE MEDIÇÃO DE GRANDEZAS MEDIDAS

A medição de cada grandeza está sujeita a erros, sejam eles aleatórios, globais ou sistemáticos. Existem várias causas possíveis para esses erros, desde erros humanos de leitura a erros devidos ao equipamento de medição, como erros relacionados com as escalas das medidas ou calibração.

Uma forma de minimizar ou mesmo evitar esses erros de leitura é estabelecer bons procedimentos de recolha dos dados por meio da repetição, verificação ou aplicação de relações entre quantidades observáveis diferentes, calibrações regulares e conservação adequada dos instrumentos (Silva,2022a).

Na Tabela 1, estão representados os métodos ou equipamentos de medição usados para quantificar as principais grandezas avaliadas em barragens, bem como as suas incertezas de medição.

Tabela 1 - Métodos ou equipamentos de medição e precisões de recolha dos dados, adaptado de (Silva,2022a).

| Grandeza                        | Método ou equipamento de medida      | Precisão Aproximada.                |
|---------------------------------|--------------------------------------|-------------------------------------|
| Temperatura do ar e água        | Termómetros                          | $\pm 0,15 \text{ }^{\circ}\text{C}$ |
|                                 | Termómetro de máxima e mínima        | $\pm 0,5 \text{ }^{\circ}\text{C}$  |
| Temperatura do betão            | Termómetro de resistência            | $\pm 0,15 \text{ }^{\circ}\text{C}$ |
| Deslocamento horizontal         | Triangulação                         | Depende de caso para caso           |
|                                 | Fio de prumo                         | $\pm 0,2 \text{ mm}$                |
| Deslocamento vertical           | Nivelamento                          | $0,5 \text{ mm/km}$                 |
| Deslocamento das fundações      | Extensómetros de varas               | $\pm 0,01 \text{ mm}$               |
| Movimentos de juntas e fissuras | Base de alongâmetro                  | $\pm 0,02 \text{ mm}$               |
|                                 | Base tridimensional                  | $\pm 0,01 \text{ mm}$               |
| Extensão                        | Extensómetro de resistência elétrico | -                                   |

|                        |                                    |                                     |
|------------------------|------------------------------------|-------------------------------------|
| Tensão                 | Tensómetro de resistência elétrica | -                                   |
| Nível da água          | Medidor de nível                   | $\pm 0,01$ m                        |
| Altura de sedimentação | Medidor da profundidade da água    | $\pm (0,2 \text{ a } 0,5)$ m        |
| Pluviosidade           | Pluviómetro                        | $\pm 10$ %                          |
| Caudal                 | Dreno                              | Variável, usualmente inferior a 5 % |
|                        | Bica totalizadora                  |                                     |

## 2.4 ANÁLISE E INTERPRETAÇÃO DO COMPORTAMENTO OBSERVADO

Como foi visto anteriormente as barragens são consideradas estruturas com alto risco potencial. O período crítico da vida de uma barragem acontece durante o período de construção e no primeiro enchimento completo da albufeira, expandindo-se durante um intervalo de 5 a 7 anos no início do período de exploração normal da barragem. Durante essa fase crítica não existem muito dados sobre o comportamento da estrutura, uma vez que coincide com o período de construção e início da exploração, então são usados modelos determinísticos e modelos ditos estatísticos com dados recolhidos até ao momento, tentando caracterizar o comportamento observado.

Quando existem dados disponíveis em quantidade suficiente um dos métodos usados é o Método de Separação de Efeitos, que é usado para analisar resultados obtidos pela observação e monitorização de grandezas que afetam a barragem. Tal como o nome indica, permite analisar os dados de uma dada grandeza ao longo do tempo, separando os efeitos sobre a grandeza em particular devido a diversas ações tendo em conta que se conhece a variação dessas ações ao longo do tempo. Este modelo estabelece relações funcionais semi-empíricas entre as grandezas e as solicitações que as originam (Morais, 2017).

Os modelos estatísticos usados para prever a resposta estrutural são baseados nas relações entre as cargas e a resposta da estrutura. Esses modelos estatísticos, também conhecidos como modelos de interpretação quantitativa, são baseados no estabelecimento de modelos de regressão linear múltipla de dados de monitorização obtidos durante a operação das barragens. No período normal de exploração de uma barragem de betão o efeito térmico está diretamente relacionado com as variações de temperatura ambiental, ou seja, do ar e da água da albufeira.

## 2.5 MODELOS DE INTERPRETAÇÃO QUANTITATIVA

Os modelos de interpretação quantitativa definem termos independentes através de conhecimento da obra, onde são usados no relacionamento entre a variação de cada solicitação e a resposta correspondente. Essa relação é estabelecida por uma expressão analítica paramétrica envolvendo funções associadas a coeficientes determinados por calibração estatística. Existem duas abordagens para a escolha dos parâmetros, os modelos HST e HTT.

A abordagem mais comum para desenvolver modelos expeditos no campo da engenharia de barragens é chamada HST (Hydrostatic, Season, Time). Esta abordagem baseia-se na hipótese de que a grandeza em estudo, como são os deslocamentos horizontais, depende da combinação dos efeitos da carga hidrostática, da variação sazonal da temperatura representada de forma simplificada através de funções

sinusoidais, e do efeito do tempo. A abordagem do tipo HTT (Hydrostatic, Temperature, Time) difere da abordagem HST na componente referente ao efeito térmico. Neste caso o efeito da temperatura é representado através das temperaturas medidas no corpo da obra, ou de registos obtidos noutros locais, como a temperatura do ar medida numa estação meteorológica. Ambas as abordagens têm como base o princípio da separação de efeitos, embora se reconheça a existência de alguma correlação entre as variáveis (Silva *et al*, 2022b).

Quando comparados com os modelos HST, os modelos HTT podem apresentar um coeficiente de determinação  $R^2$  maior e desvios-padrão dos resíduos  $\sigma_\varepsilon$  menores (Mata *et al*, 2014).

Este método estatístico, só é aplicável caso exista um número de observações suficiente e as funções geradas consigam prever o comportamento da grandeza. Os métodos consistem em aproximar efeitos associados a um período limitado num ponto específico através da Equação (2):

$$\delta(h_i, S_i, t_i) = U_h(h_i) + U_s(S_i) + U_t(t_i) + k + \varepsilon_i \quad (2)$$

em que,

- $\delta(h_i, S_i, t_i)$  – Valor observado da grandeza em análise na observação  $i$ , que depende da pressão hidrostática, da temperatura e do instante em que se realiza a observação;
- $U_h(h_i), U_s(S_i), U_t(t_i)$  – São, respetivamente, as parcelas das grandezas correspondentes ao efeito elástico do nível de água na albufeira, ao efeito elástico da variação sazonal da temperatura e ao efeito do tempo na  $i^{ésima}$  observação;
- $k$  – Constante que corresponde à diferença entre valores observados e calculados no início do período de calibração;
- $\varepsilon_i$  - Resíduo da  $i^{ésima}$  observação, dado pela diferença entre o valor estimado e o valor observado.

### Efeito do nível da água na albufeira

Os efeitos do nível da água na albufeira podem ser representados por uma função polinomial. A aproximação pode ser feita pela Equação (3) (Silva *et al*, 2022b):

$$U_h(h) = a_1 * h^4 + a_2 * h^3 + a_3 * h^2 + a_4 * h \quad (3)$$

com,

- $h$  - Altura de água na albufeira (m);
- $a_1, a_2, a_3, a_4$ - Coeficientes a ajustar.

\* Em muitas aplicações não é necessário utilizar o polinómio completo.

### Efeito da temperatura

Como referido, os efeitos da temperatura estão relacionados com as variações da temperatura do ar e da água da albufeira e dos seus efeitos no campo térmico da estrutura. No caso dos modelos HTT, este termo é representado através das temperaturas medidas no corpo da obra ou de registo noutros locais, como a temperatura do ar medida numa estação meteorológica. Por sua vez, na abordagem HST, o efeito da temperatura é aproximado por funções sinusoidais de período anual e diário, à semelhança do que acontece com as variações de temperatura ambientais. Estas funções são obtidas pelas expressões que se apresentam Equação (4) (Mata, 2007).

A onda térmica anual é representada por,

$$U_{\theta}(\theta_1)^{anual} = b_1 * \cos(\theta_1) + b_2 * \sin(\theta_1) + b_3 * \sin^2(\theta_1) + b_4 * \cos(\theta_1) * \sin(\theta_1) \quad (4)$$

com,

$$\theta_1 = \frac{2 * \pi * t_d}{365} \quad 1 \leq t_d \leq 365 \quad (5)$$

Resultando da Equação (4) e da Equação (5),

$$U_{\theta}(\theta_1, \theta_2) = b_1 * \cos(\theta_1) + b_2 * \sin(\theta_1) + b_3 * \sin^2(\theta_1) + b_4 * \cos(\theta_1) * \sin(\theta_1) \quad (6)$$

onde,

- $t_d$ - Número de dias decorridos desde o início do ano até à data da observação, representado através de um número real de forma a incorporar também a informação sobre a hora do dia;
- $b_1, b_2, b_3, b_4$  - Coeficientes a ajustar.

Tal como no caso de efeito da pressão hidrostática, nem sempre é necessário utilizar todos os termos. No caso de Portugal, em que os períodos do ano e do dia mais quentes e mais frios têm durações semelhantes, é usualmente suficiente utilizar os dois primeiros termos de cada onda.

### Efeito do tempo

Neste caso estão envolvidos vários efeitos que variam ao longo do tempo, como efeitos de fluência e/ou relaxação do betão, processos de deterioração e ações inelásticas. Tal como o efeito do nível da água da albufeira, pode ser aproximado através de funções polinomiais, como representado na Equação (7) (Silva *et al*, 2022b):

$$U_t(t) = c_1 * t^3 + c_2 * t^2 + c_3 * t + c_4 * \ln\left(1 + \frac{t}{a}\right) \quad (7)$$

com,

- $t$ - Número de dias entre a campanha e o início da observação;
- $a$ - Número de dias entre o primeiro enchimento e a data do início da análise;
- $c_1, c_2, c_3, c_4$ - Coeficientes a ajustar.

## 2.6 IDENTIFICAÇÃO DE VALORES ATÍPICOS

Qualquer analista que tenha analisado um conjunto de dados, é provável que se tenha deparado com valores atípicos. A definição de um valor atípico (*outlier*) pode ser apresentada como uma observação que se desvia tanto das outras observações que levanta suspeitas de que foi gerada por um mecanismo diferente (Hawkins, 1980).

Uma das possibilidades para o aparecimento de valores atípicos é a ocorrência de um erro. Um erro pode ser proveniente de vários fatores como de medição devido a algum engano no processo de medição ou uma falha num instrumento, uma resposta baseada numa má interpretação de uma pergunta ou uma pergunta mal formulada, entre outros. Por vezes, os aparecimentos de valores fora do normal são originados por situações nunca antes presenciadas pela estrutura, como o acontecimento de desastres naturais, que aplicam novas solicitações nunca antes registada que levam ao aparecimento de valores atípicos. Por vezes esses dados pontuais não são contabilizados em análises.

Os valores atípicos podem ser identificados de diversas formas, dependendo do tamanho e propriedades da base de dados. Em bases de dados de menor dimensão muitas vezes, estes valores são identificados através da análise visual dos resultados, sendo facilmente identificados e assinalados. Em bases de dados mais complexas, normalmente é muito difícil distinguir os valores fora do comum, sendo usual recorrer a métodos estatísticos e programas informáticos para fazer essa identificação.





# 3

## MODELOS DE MACHINE LEARNING UTILIZADOS PARA IDENTIFICAÇÃO DE VALORES ATÍPICOS

### 3.1 INTRODUÇÃO

Os avanços tecnológicos são visíveis todos os dias, e dos mais influentes recentemente são na área da inteligência artificial. A aprendizagem automática ou *Machine learning* é uma das principais subáreas da inteligência artificial, e é composta por uma coleção de métodos criados a partir de modelos matemáticos baseados na teoria estatística que permitem aos computadores automatizar tarefas com base na descoberta sistemática de padrões nos conjuntos de dados disponíveis ou em experiências passadas (Bhavsar *et al*, 2017; Alpaydin 2020).

Os métodos de *Machine learning* são usados em programas e aplicações permitindo que estes possam aprender através dos dados recolhidos e identificar padrões, de modo a tomar decisões com uma reduzida intervenção humana com uma capacidade de adaptação e melhoria dos comportamentos e respostas, levando o algoritmo a ajustar as respostas de modo iterativo para ter um melhor desempenho. Apesar da existência de algoritmos de *Machine Learning* não ser algo novo, a capacidade de aplicar automaticamente cálculos matemáticos complexos com resultados fiáveis e repetíveis cada vez mais rápido, é um dos avanços recentes na área. Podemos dividir a aprendizagem em *Machine Learning* em dois grandes tipos: supervisionada e não-supervisionada. Conforme representado na Figura 17, podemos observar essa divisão e os seus consequentes subtipos.

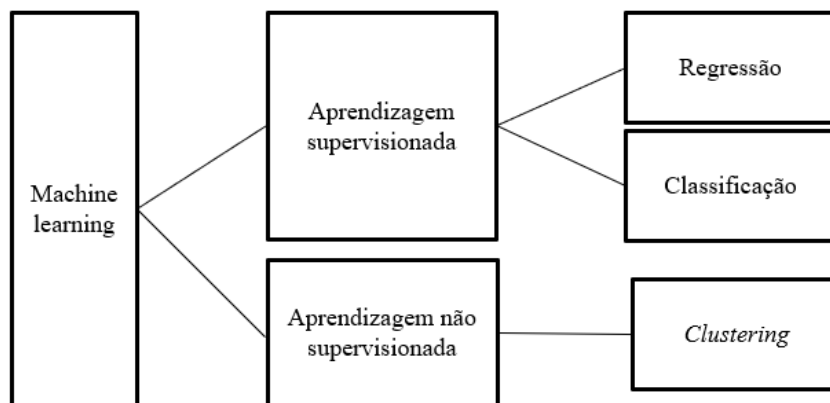


Figura 17- Alguns tipos de aprendizagem em *Machine Learning*.

Nas secções seguintes serão apresentados os conceitos gerais sobre aprendizagem supervisionada e aprendizagem não supervisionada, assim como os principais conceitos relacionados com os métodos de *Machine Learning* adotados no âmbito desta tese.

### 3.2 APRENDIZAGEM SUPERVISIONADA

A aprendizagem supervisionada é uma tarefa de *Machine Learning*, que consiste num método de aprendizagem a partir de exemplos, ou seja, nesse tipo de aprendizagem um aprendiz (normalmente, um programa de computador), recebe um conjunto de dados de treino e um conjunto de dados de teste. O objetivo é que o aprendiz "aprenda" com os exemplos rotulados do conjunto de treino, de forma a poder identificar corretamente os exemplos não rotulados do conjunto de teste. O conjunto de treino é constituído por  $n$  pares ordenados  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ , em que cada  $x$  é uma medida ou um conjunto de medidas de um único ponto dos dados e  $y$  é o rótulo correspondente a esse ponto dos dados. Por exemplo,  $x$  pode representar um grupo de cinco características de um paciente num hospital, incluindo altura, peso, temperatura, nível de açúcar no sangue e pressão arterial. O correspondente  $y$  pode corresponder a uma classificação do paciente como "saudável" ou "não saudável". Mais formalmente, o modelo baseado na aprendizagem supervisionada procura encontrar uma função  $h(x_i)$ , denominada hipótese, que se aproxime da função  $f(x_i)$ , onde  $f(x_i)$ , é a saída da  $i$ -ésima entrada de  $x$  (Russell, Norvig, 2002; Learned-Miller, 2014).

O algoritmo visa aprender uma regra geral que faça o mapeamento correto das entradas e das saídas. A utilização de métodos como a classificação e regressão na aprendizagem supervisionada tem como finalidade a identificação de padrões para posteriormente prever os valores do rótulo em dados adicionais ainda por rotular. Esta aprendizagem é frequentemente usada em aplicações onde são usados dados históricos como ferramenta para previsão de possíveis dados futuros.

A classificação é o processo de identificação da categoria ou classe de uma observação através de um método de aprendizagem supervisionada, onde as categorias dos objetos já são conhecidas para fins de treino e teste. Após a conclusão do processo de treino para uma determinada tarefa de classificação, o algoritmo pode atribuir uma categoria a um objeto recém-observado (Lee, Shin, 2020).

Este processo é feito rotulando os dados de entrada numa classe, o que pode ser definido como uma função matemática que procura associar corretamente dados de entrada a um rótulo categórico (classe), que são divididos em dois subconjuntos: base de treino e base de teste. A base de treino é

submetida a uma calibração dos dados feita pelo modelo, de modo a obter uma maior precisão de classificações corretas. Na Figura 18 temos um exemplo de processamento da classificação.

De modo similar com a classificação, a regressão utiliza dados históricos para prever uma saída a partir de uma faixa contínua de valores possíveis. Na regressão o resultado é numérico, enquanto na classificação é categórico. O algoritmo faz uma relação funcional entre os dados de entrada e das saídas, sendo a exatidão calculada com base na diferença entre os valores da saída observada e da saída prevista pelo modelo adotado.

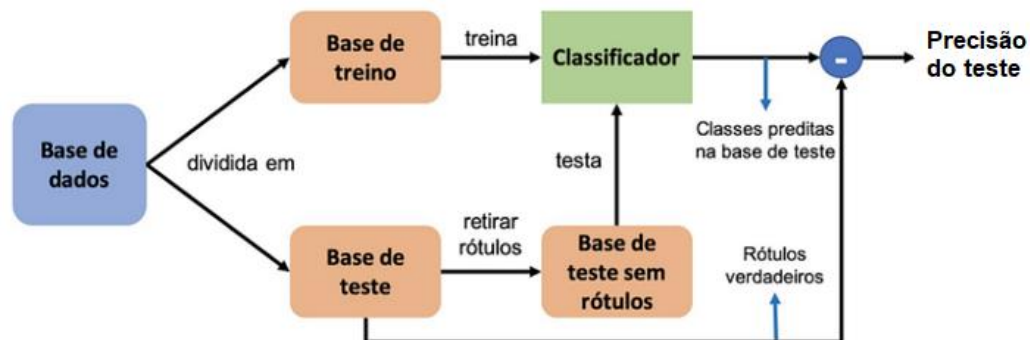


Figura 18- Esquema de um problema de classificação, adaptado de (Escovedo, Koshiyama ,2020).

### 3.3 APRENDIZAGEM NÃO SUPERVISIONADA

Na aprendizagem não supervisionada o modelo aprende padrões com o conjunto de dados de entrada enquanto os dados de saída não são conhecidos. O algoritmo aglomera os dados de entrada em grupos ou cria *clusters* de dados para os classificar/identificar e quando um novo conjunto de dados de entrada é fornecido o modelo procura o melhor grupo de dados a que este novo conjunto pode pertencer de modo a obter a saída (Haykin, 2009).

A aprendizagem não supervisionada procura identificar alguma(s) característica(s) de similaridade nos dados de modo a agrupá-los ou organizá-los entre si.

As ferramentas de *Clustering* são usadas para agrupar conjuntos de objetos com base em semelhanças num espaço multidimensional. Objetos no mesmo cluster são mais semelhantes entre si do que em clusters diferentes. O *clustering* é considerado como *Machine Learning* não supervisionada, uma vez que os rótulos de classe dos objetos não são conhecidos de antemão. No *clustering*, os objetos são agrupados num espaço de dimensão  $n$  e a similaridade entre dois objetos é medida com uma função de distância de similaridade (Lee, Shin, 2020).

Quando agrupamos observações (Figura 19), queremos que estas sejam o mais semelhante entre si dentro do seu grupo, e que as observações noutros grupos sejam diferentes. Isto, permite-nos identificar quais as observações que são semelhantes entre si e potencialmente categorizá-las, embora não haja critérios específicos para um bom agrupamento. Estes serão definidos pelo utilizador, dependendo das suas necessidades específicas. As ferramentas de *clustering* podem ser usadas para encontrar pontos de dados incomuns/*outliers* ou identificar propriedades dentro dos dados.

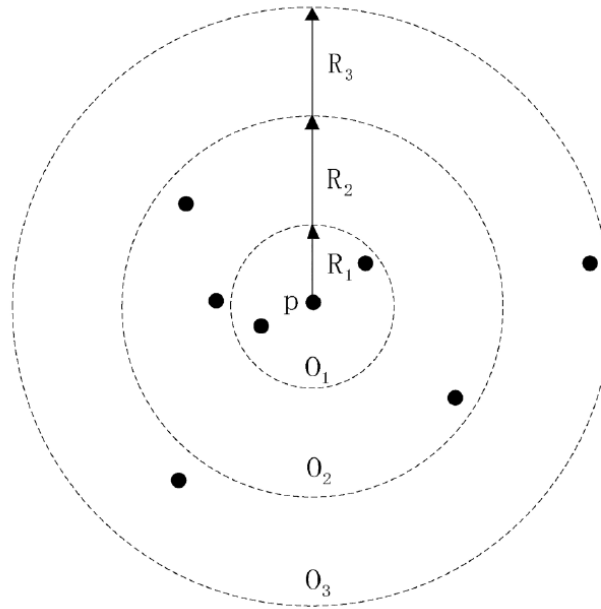


Figura 19 - Representação de possíveis agrupamentos de clustering O1, O2, O3 com respetivo raio de agrupamento R1, R2 e R3 retirado de (Xu *et al*, 2022).

### 3.4 MÉTODO DE REGRESSÃO LINEAR MÚLTIPLA

Muitos problemas em engenharia ou ciência surgem no momento de investigar as relações entre duas ou mais variáveis. A regressão é uma técnica estatística muito útil para tais problemas (Montgomery, Runger, 2003).

A regressão estuda o relacionamento entre uma variável chamada dependente e outras variáveis chamadas independentes. Esta relação é representada por um modelo matemático, isto é, por uma equação que associa a variável dependente com as variáveis independentes. Este modelo é designado por modelo de regressão linear simples se define uma relação linear entre a variável dependente e uma variável independente. Se em vez de uma, forem incorporadas várias variáveis independentes, o modelo passa a denominar-se modelo de regressão linear múltipla (Henriques, 2011).

A regressão linear múltipla como o nome indica é uma regressão, logo entra dentro do tipo de aprendizagem supervisionada, em que se utilizam ferramentas estatísticas. Estimam-se os valores numéricos da relação linear através da equação (8) (Mata, 2007):

$$U = \sum_{j=1}^P X_j * \beta_j + k + \varepsilon \quad (8)$$

onde,

- $U$  – Variável dependente, nomeadamente o valor observado da grandeza em análise;
- $P$  – Número total de variáveis independentes;
- $X_j$  –  $j^{ésimo}$  termo associado à variável independente  $j$ ;
- $\beta_j$  – Coeficiente da regressão da variável independente  $X_j$ ;
- $k$  – Constante que corresponde ao termo independente;

- $\varepsilon$  – Resíduo da observação, dado pela diferença entre o valor estimado e o valor observado.

Para a estimativa dos parâmetros da regressão linear múltipla é usualmente adotado o método dos mínimos quadrados, onde se obtém uma estimativa de  $Y$ . No caso de o modelo ser representado em forma matricial com  $n$  equações temos que:

$$Y = X \cdot \beta + \varepsilon \quad (9)$$

em que,

$$Y = \begin{Bmatrix} y_1 \\ \vdots \\ y_n \end{Bmatrix}; X = \begin{bmatrix} x_{11} & \vdots & x_{pn} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_{1n} & \vdots & x_{pn} & 1 \end{bmatrix}; \beta = \begin{Bmatrix} \beta_1 \\ \vdots \\ \beta_p \\ \beta_n \end{Bmatrix}; \varepsilon = \begin{Bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{Bmatrix} \quad (10)$$

De modo a minimizar o somatório do quadrado dos resíduos, para se chegar a valores mais assertivos dos coeficientes usamos:

$$\sum_{i=1}^n \varepsilon_i^2 = \varepsilon^T * \varepsilon = (Y - X * \beta)^T * (Y - X * \beta) \quad (11)$$

No caso de  $X^T * X$  ser invertível, os valores estimados dos parâmetros  $\beta_j$  são dados por:

$$\hat{\beta} = (X^T * X)^{-1} * X^T * Y \quad (12)$$

Para esta equação ser válida e calculável temos de garantir que  $X^T * X$  tenha um determinante não nulo, e que o número de observações  $n$  seja superior ao número de variáveis independentes  $P$ .

### 3.5 MÉTODO K-MEANS

O K-means é um método de *clustering*, de aprendizagem não supervisionada, popularmente usado em análise de *clusters*, *data mining* e reconhecimento de padrões (Faraoun, Boukelif, 2007; Hegde *et al* , 2015).

A ideia básica deste método consiste em definir agrupamentos de modo que a variação total intra-clustering seja minimizada. O algoritmo padrão é o Hartigan-Wong 1979, que define a variação total dentro do cluster como a soma dos quadrados das distâncias euclidianas entre os pontos e o centróide do agrupamento, sendo representadas pela Equação (13):

$$W(C_k) = \sum_{x_i \in C_k} (x_i - \mu_k)^2 \quad (13)$$

em que,

- $x_i$  é um ponto dos dados pertencente ao *cluster*  $C_k$ ;
- $\mu_k$  é o valor médio dos pontos atribuídos ao *cluster*  $C_k$ .

Para cada observação ( $x_i$ ) é atribuído um *cluster* de modo que a distância da soma dos quadrados de  $x_i$  aos seus centros de cluster atribuídos ( $\mu_k$ ) é minimizada. A variação total dentro do *cluster* é definida da seguinte forma, Equação (14):

$$tot.withiness = \sum_{k=1}^K w(C_k) = \sum_{k=1}^K \sum_{x_i \in C_k} (x_i - \mu_k)^2 \quad (14)$$

Em que a soma total do quadrado das distâncias dentro do *cluster* (*tot.withiness* ou *SEE*) mede a compacidade do agrupamento e pretende-se que esta seja o menor possível.

Cada observação é associada a um *cluster*, onde nenhum ponto dos dados é conhecido, no entanto o ponto localizado do centro do cluster é denominado centróide. O processo de agrupamento ocorre ao reduzir a distância euclidiana entre os dados e o centróide correspondente a cada *cluster*. Inicialmente, é definido o número de centróides a utilizar (*k*), um para cada cluster, assumindo que os centróides estejam inicialmente distantes uns dos outros. Os pontos do cluster mais próximos de cada centróide são selecionados e associados a esse centróide, criando clusters e consequentemente novos centróides. Após a obtenção desses *k* novos centróides, uma nova iteração é realizada entre os mesmos pontos do conjunto de dados e o novo ponto mais próximo ao centróide do cluster. Este processo continua até que nenhuma alteração na localização dos centróides seja possível (Ajith *et al*, 2018).

O algoritmo de agrupamento k-means pode ser definido pelas seguintes etapas:

- Seleção de *n* clusters (centróides);
- O método da distância euclidiana é usado para calcular a distância entre os objetos e o centróide;
- Identificação do centróide mais próximo para cada ponto dos dados, agrupando os pontos com base na distância mínima ao centróide;
- Repetição das etapas 2 e 3 até que nenhum objeto seja deixado sem *cluster*.

Para dar início ao processo de *clustering* do k-means é necessário definir o número de clusters em que este irá agrupar os dados, valor este que pode afetar a qualidade do resultado. Escolher um valor muito pequeno pode levar a resultados subajustados e a escolha de um *k* muito grande pode levar a resultados sobreajustados. O número ideal de agrupamentos depende das medidas de similaridade e dos parâmetros usados para o mesmo (Gasch, Eisen, 2002). Para obter o melhor número de clusters é necessário executar o k-means para um intervalo de valores e comparar os resultados. Ainda não existe nenhum método para determinar o valor exato do número de cluster ótimo, mas pode ser estimado usando algumas técnicas como por exemplo o método do cotovelo (*Elbow method*).

Este método é uma heurística usada para determinar o valor de *k*, número de clusters, numa base de dados. O método consiste em fazer o gráfico da soma total do quadrado das distâncias dentro do *cluster* (SSE) em função do número de *clusters*, e escolher o ponto onde a inclinação muda mais drasticamente da curva exibida pelo gráfico. Esse ponto é chamado de cotovelo, em que o valor de *k* correspondente define o número de *clusters* ótimo, como representado no exemplo da Figura 20.

Podemos adotar os seguintes passos para definir os clusters ideais (Bholowalia, Kumar, 2014; Syakur *et al*, 2018):

- Cálculo do algoritmo de agrupamento (por exemplo, agrupamento k-means) para diferentes valores de *k*. Por exemplo, variando *k* de 1 a 10 clusters;
- Para cada *k*, calcular a soma total do quadrado das distâncias dentro dos *clusters* (SSE);
- Traçar a curva de SSE de acordo com o número de clusters *k*;
- A localização da maior variação na inclinação da curva é geralmente considerada como um indicador do número apropriado de grupos.

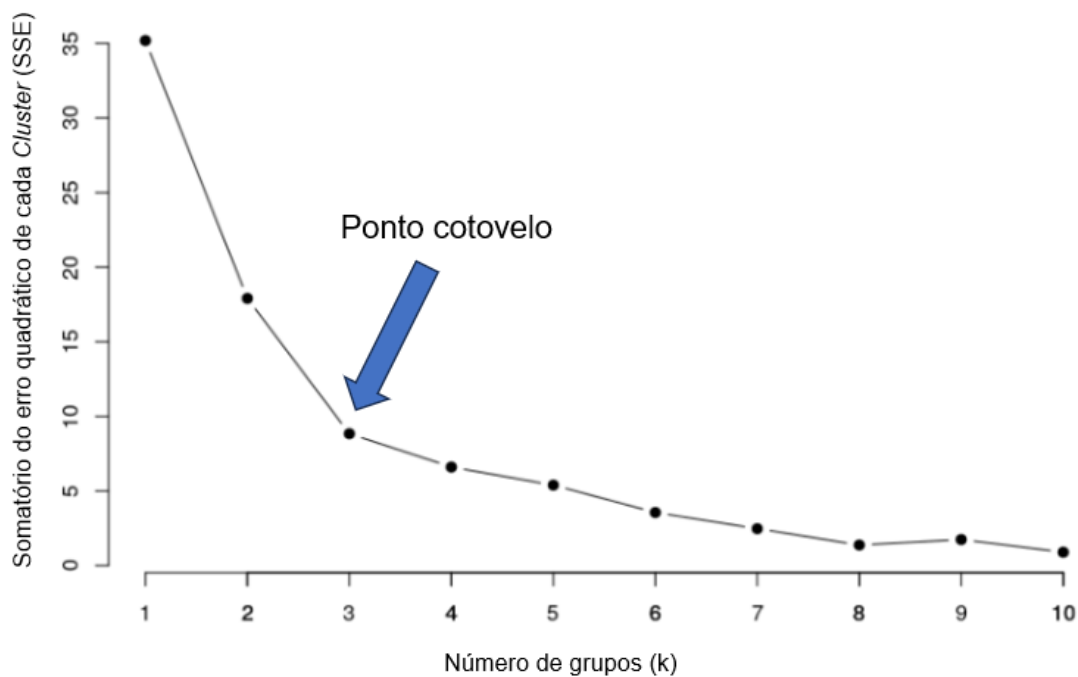


Figura 20 - Identificação do ponto do cotovelo, adaptado de ([4 Agrupamento | Introdução ao Machine Learning \(dataat.github.io\)](#)).

Analisando o exemplo representado na Figura 20, podemos sugerir que o número de cluster a usar para esta base de dados é 3, uma vez que demonstra uma variação grande do declive da reta. É sempre o utilizador que especifica o número de clusters a usar no K-means podendo este variar de analista para analista para a mesma base de dados.

### 3.6 DBSCAN

Normalmente, todos os métodos de *clustering* usam a mesma abordagem: calcular as semelhanças e depois agrupar os pontos dos dados em *clusters* através das suas semelhanças. No exemplo do *K-means*, anteriormente analisado, cada ponto dos dados em análise pertence a um cluster, mesmo que estes estejam muito dispersos no espaço vetorial. Como os clusters dependem no valor médio dos pontos, uma alteração nos dados pode levar a uma nova reorganização dos clusters. No caso do Dbscan essa possibilidade é mais reduzida, devido à forma como este algoritmo forma os clusters.

DBSCAN (*Density-Based Spatial Clustering of Applications Noise*) é um algoritmo de *clustering* baseado em densidade, usando métodos de aprendizagem não supervisionados. O DBSCAN tem como ideia base a criação de um *cluster* no espaço com uma região contígua de alta densidade de pontos, separada de outros clusters regiões contíguas de baixa densidade de pontos (Ester *et al*, 1996).

O algoritmo DBSCAN requer basicamente dois parâmetros: o *eps* e o *minPts*.

O parâmetro *eps* é uma distância máxima definida para o qual os pontos devem estar uns dos outros para serem considerados parte de um *cluster*. Isso significa que se a distância entre dois pontos for menor ou igual a esse valor (*eps*), esses pontos são considerados vizinhos.

Se o valor de  $\epsilon$  escolhido for muito pequeno, grande parte dos dados não serão agrupados, consequentemente irão ser considerados atípicos por não satisfazerem esta distância mínima. Por outro lado, se o valor escolhido for muito alto, os clusters fundir-se-ão e a maioria dos objetos estará no mesmo cluster. O  $\epsilon$  deve ser escolhido com base na distância do conjunto de dados.

O parâmetro  $\text{minPts}$  corresponde ao número mínimo de pontos pertencentes à área abrangente pelo parâmetro  $\epsilon$  para formar uma região densa. Por exemplo, se definirmos o parâmetro  $\text{minPts}$  como 5, precisaremos de pelo menos 5 pontos abrangidos para formar uma região densa, isto é, um *cluster* (Hahsler *et al*, 2019).

O modelo introduzido pelo DBSCAN utiliza uma estimativa simples do nível de densidade mínima, baseada num limiar para o número de vizinhos,  $\text{minPts}$ , dentro da distância definida pelo parâmetro  $\epsilon$ . Os objetos com mais de  $\text{minPts}$  vizinhos dentro deste raio (incluindo o ponto de consulta) são considerados como um ponto central. O DBSCAN começa por encontrar as áreas que satisfazem esta densidade mínima e que estão separadas por áreas de menor densidade. Por razões de eficiência, o DBSCAN não efetua a estimativa da densidade entre os pontos, em vez disso, todos os vizinhos situados dentro do  $\epsilon$  de um ponto central são considerados como fazendo parte do mesmo *cluster* que o ponto central. Os pontos considerados não centrais, ou seja, que não cumpram a condição implementada pelo parâmetro  $\text{minPts}$ , mas contêm pelo menos um ponto central são designados por pontos de fronteira. Os pontos de fronteira são também considerados pertencentes ao *cluster*. Os pontos que não são alcançáveis por densidade a partir de qualquer ponto central são considerados atípicos e não pertencem a nenhum *cluster* (Schubert *et al*, 2017).

Na Figura 21, está ilustrado um exemplo do método de *clustering* DBSCAN com um parâmetro  $\text{minPts}=4$ . Os círculos envolventes dos pontos representam a distância máxima definida pelo parâmetro  $\epsilon$ , enquanto as setas representadas na Figura 21 indicam os pontos considerados pertencentes ao cluster. Os pontos representados a vermelho são os pontos considerados centrais, enquanto os pontos B e C estão representados a amarelo, uma vez que são considerados pontos de fronteira por apenas terem um ponto dentro do seu círculo. No caso do ponto N, representado a azul, este não pertence ao cluster sendo considerado um ponto atípico.

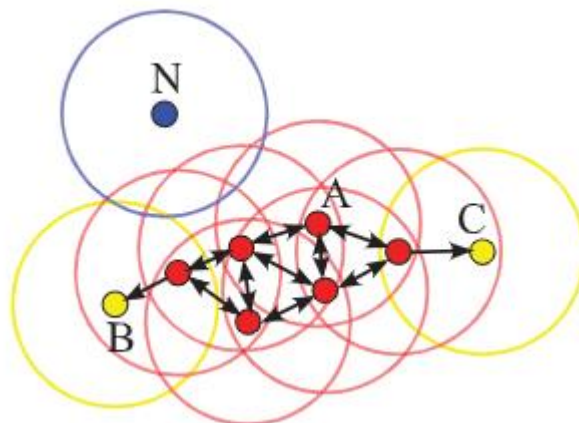


Figura 21 - Ilustração do método de cluster Dbscan, retirado de (Schubert *et al*, 2017).



### 3.7 MÉTODO LOF

*Local Outlier Factor* (LOF) é um modelo de aprendizagem não supervisionado usado para detetar anomalias em base de dados com múltiplas variáveis, é baseado no conceito de densidade local. O *Local Outlier Factor* (LOF) é um algoritmo de deteção de valores atípicos, baseado em densidade, com metodologia representa na Figura 23, que realiza a comparação entre a densidade local de um ponto com as densidades dos seus pontos vizinhos (*k-neighbors*). A técnica original do LOF foi apresentada por Breuning e consiste em calcular um fator LOF para todos os pontos dos dados existentes na base de dados. O cálculo deste fator é dependente de três parâmetros: *k-distance*, *reach-distance* e *local reachability density* (Ajith et al, 2018).

A *k-distance* é a distância entre um ponto dos dados em análise e seu ponto vizinhos mais próximo. O número de pontos vizinhos vai ser arbitrado pelo utilizador, ou seja, se o valor de *k* for igual a 3, o método aplica a distância aos 3 pontos mais próximos do ponto em análise.

A *reach-distance* é o valor máximo entre a distância Euclidiana entre o ponto em estudo e o ponto alvo da análise e a *k-distance*. Na Equação (15) é apresentada a fórmula de calculo do parâmetro *reach-distance*, baseado no exemplo ilustrado na Figura 22.

$$reach-dist(q, p) = \max\{k-dist(q); d(p, q)\} \quad (15)$$

Na Figura 22 está representado um exemplo em que o *k* toma o valor de 4 e a *k-distance* está representada pela nomenclatura *k-dist(p)*, em que *p* é o ponto a ser analisado, e a distância Euclidiana representada pela nomenclatura *d(p, q)*, em que *q* representa o ponto alvo. O círculo representa a distância máxima do ponto *p* aos pontos vizinhos para este exemplo.

Analisando o exemplo da figura, *reach-dist(q, p)* é igual a *k-dist(p)*, uma vez que *k-dist(q)* é igual a *d(p, q)*. Caso o ponto alvo seja o ponto *r*, temos que a distância ao ponto vizinho mais próximo, *reach-dist(r, p)* é inferior à distância euclidiana entre os pontos *r* e *p*, *d(r, p)*. Neste caso o parâmetro *reach-distance* é igual a *d(p, r)*.

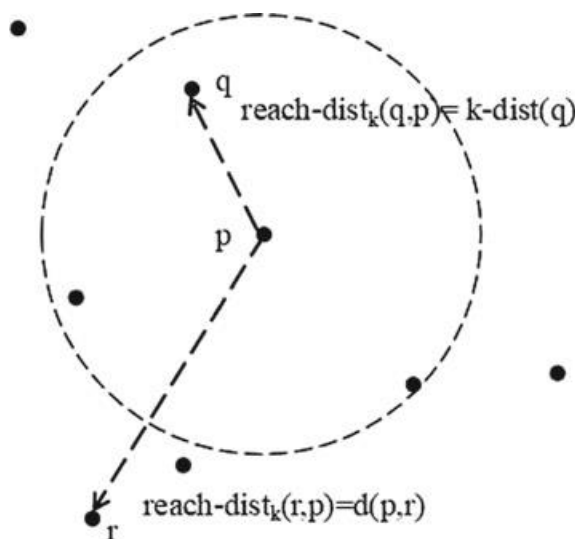


Figura 22 - Exemplo do parâmetro reach-distance para *k* =4, retirado de (Breunig et al, 2000).

O *local reachability density* (LRD) de um ponto pode ser definido como o inverso de distâncias médias de alcance de todos os pontos na vizinhança  $k$  do ponto em análise. Esta média é obtida pela divisão da soma das *reach-distances*, em que o ponto alvo pertence á vizinhança do ponto de partida, pelo número  $k$  escolhido. Vão ser escolhidos novamente o ponto  $p$  como ponto de partida e o ponto  $q$  como ponto alvo, esta operação está representada na Equação (16) como:

$$LRD_k(p) = \frac{1}{k} \left( \sum_{q \in N(p,k)} reach\ dist_k(p, q) \right)^{-1} \quad (16)$$

onde  $N(p, k)$  é o conjunto de pontos de *k-neighbors* mais próximos de  $p$ .

O fator LOF de um ponto dos dados é a média das razões de  $LRD(o)$  a  $LRD(p)$  para todos os pontos  $o$  na  $k$  vizinhança mais próxima do ponto  $p$ . Neste caso o nosso ponto  $o$  irá ser o ponto em que se está a focar o cálculo e os pontos  $p$  os pontos vizinhos do ponto  $o$ . Está média é calculada pela divisão do LRD do ponto  $o$  pela soma dos valores de LRD dos pontos vizinhos dividindo pelo número de pontos  $k$ , obtendo então a seguinte expressão:

$$LOF_k(p) = \frac{1}{k} \sum_{o \in N(p,k)} \frac{LRD_k(o)}{LRD_k(p)} \quad (17)$$

Assim, pode-se concluir que o fator LOF de quaisquer objetos depende de um único parâmetro que é o número mínimo de pontos na vizinhança mais próximos. O fator LOF deverá seguir a seguinte interpretação (Breunig *et al*, 2000; Xu *et al*, 2022):

- se o  $LOF \approx 1$ , então é provável que não seja um valor atípico;
- Se o  $LOF \ll 1$ , então é provável que não seja um valor atípico;
- Se o  $LOF \gg 1$ , então é provável que seja um valor atípico;

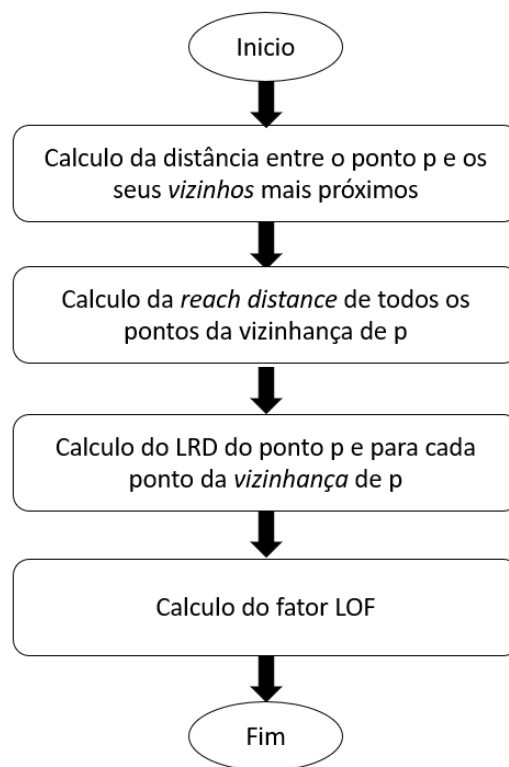


Figura 23- Etapas do algoritmo de cluster LOF.

### 3.8 METODOLOGIAS PROPOSTAS PARA A IDENTIFICAÇÃO DE VALORES ATÍPICOS

Nesta secção é apresentada uma metodologia para a identificação de potenciais valores atípicos com recurso aos métodos de *Machine Learning* descritos nas seções anteriores. A metodologia proposta, que será aplicada a um caso de estudo conforme apresentado no capítulo 4, consiste nos seguintes passos, Figura 24:

- Seleção da grandeza a analisar (eg. medições de deslocamento radial observado com recurso a sistemas de recolha automática (RAD)), e das variáveis representativas das principais ações (nível água da albufeira e temperatura do ar).
- Remoção de erros grosseiros nas medições.
- Remoção de valores atípicos locais, através da aplicação de métodos estatísticos robustos, nomeadamente de uma mediana com janela móvel e de um limite de variação aceitável (relativamente ao valor mediano da janela móvel estabelecida), com a finalidade de identificar e eliminar valores atípicos locais. Nesse método, uma janela móvel é usada para analisar um subconjunto contínuo de dados em um determinado intervalo de tempo. A escolha do tamanho da janela depende da frequência dos dados e da natureza da grandeza observada. Ao estabelecer um limite de variação aceitável, o especialista pode levar em consideração o comportamento esperado da grandeza observada, por exemplo, para frequências de observação horárias espera-se que a variação da resposta da obra em uma janela móvel de 5 horas seja relativamente suave em comparação com a variação observada em uma janela de 1 ano.

- Aplicação do método de separação de efeitos, baseado na Regressão Linear Múltipla, estabelecendo relação entre a altura da água no reservatório e o efeito térmico definido como variação sazonal com o deslocamento radial (modelo HST), gerando os respetivos coeficientes de regressão e resíduos,  $\varepsilon$ . Criação de dois subconjuntos de resíduos, os que apresentam valores positivos ou nulos e os que apresentam valores negativos. Os resíduos positivos indicam os casos em que o modelo está subestimando os valores observados, enquanto os resíduos negativos indicam casos em que o modelo está superestimando os valores observados, permitindo analisar e compreender melhor o impacto de cada tipo de valor atípico no desempenho da regressão.
- Conversão dos resíduos de coordenadas cartesianas para coordenadas polares, convertendo os dados para uma estrutura circular tornando os pontos representados por uma distância radial e um ângulo. A transformação em coordenadas polares pode facilitar a deteção de valores anómalos, através da sua representação circular, em função da época do ano, uma vez que podem ser observados como pontos que se desviam significativamente da maioria dos dados no plano polar.
- Aplicação do algoritmo de *clustering* DBSCAN, utilizando os resíduos já previamente separados em valores positivos e negativos, convertidos em coordenadas polares. Os parâmetros da distância de vizinhanças ( $\epsilon$ ) e número mínimo de pontos para forma um *cluster* ( $\text{minPts}$ ), são escolhidos adequadamente pelo especialista. Os resultados são apresentados de forma gráfica, comparando os valores iniciais e os valores finais após a identificação e remoção dos valores atípicos, fazendo visualmente uma avaliação do impacto da remoção desses valores.
- Em alternativa à aplicação do algoritmo DBSCAN, adotar o algoritmo de *clustering* LOF, utilizando os mesmos dados, escolhendo devidamente o número de *k-neighbours* ( $\text{minPts}$ ), de modo a serem obtidos os fatores LOF. Em seguida é feita uma análise aos fatores LOF, onde o especialista através de um método empírico estabelece o valor limite para o qual os pontos com fator superior ao valor estabelecido são considerados valores atípicos.
- Apresentação dos resultados obtidos através de representação gráfica, comparando os valores iniciais e os valores finais após a identificação e remoção dos valores atípicos.

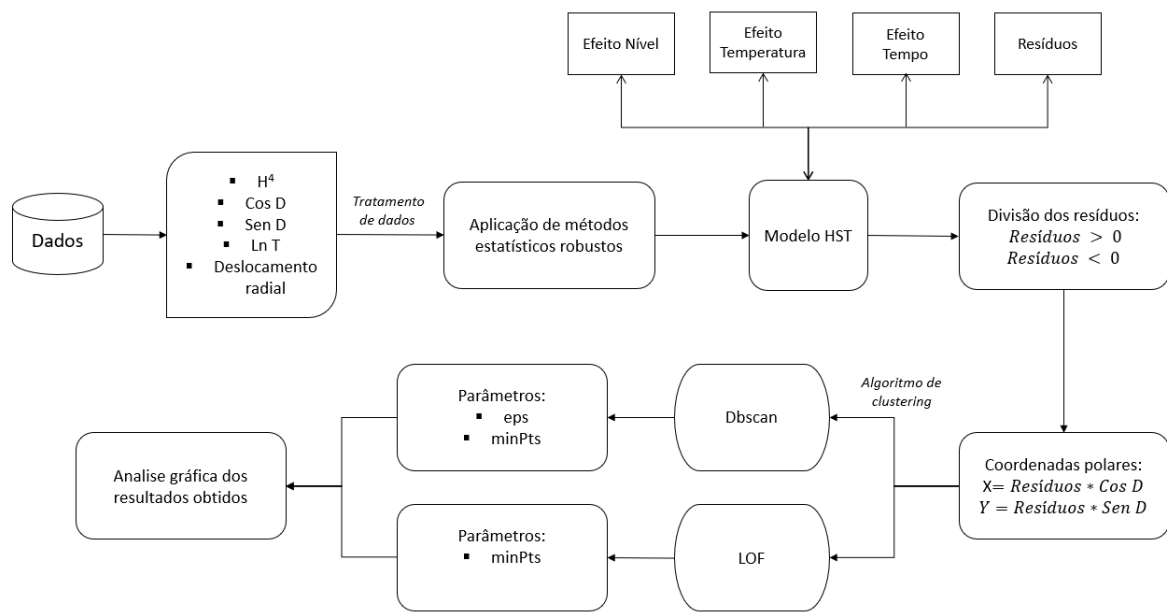


Figura 24 - Fluxograma da metodologia proposta para identificação de valores atípicos.

Para a realização das metodologias propostas é utilizado o *software Rstudio* (R Core Team, 2022), que é um programa de linguagem de programação em R, onde são inseridos os dados e desenvolvido todos os processos descritos de modo a obter os resultados. Usfruindo das funcionalidades disponíveis na biblioteca foram usados pacotes como o *Stats* (Rdocumentation.org), em particular a função *lm()* (Rdocumentation.org) para o cálculo dos coeficientes e resíduos do modelo de Regressão Linear Múltipla. Foi utilizado também o pacote *Roll* (Rdocumentation.org) para a execução da mediana de janela móvel e o pacote *DbSCAN*, em particular a função *dbSCAN()* (Rdocumentation.org) e *lof()* (Rdocumentation.org) para aplicação dos métodos para obter os resultados finais.



# 4

## CASO DE ESTUDO: BARRAGEM DO FEITICEIRO

### 4.1 INTRODUÇÃO

Neste capítulo pretende-se apresentar resultados aplicados a um caso prático, dos processos e metodologias apresentados nos capítulos anteriores, no âmbito de identificação e análise de valores atípicos. O caso prático escolhido foi a barragem do Feiticeiro, devido ao histórico de dados de recolha automática que esta barragem apresenta. É feita uma descrição geral da barragem e do sistema de recolha de dados instalado.

Neste trabalho são analisados de entre todos os dados recolhidos, os deslocamentos do fio de prumo FP1-131,31 m. Para tratamento dos resultados foi utilizado o *software Rstudio* (R Core Team, 2022), onde são inseridos os dados e desenvolvido todos os processos de modo a obter os resultados. Através das observações dos dados fio de prumo foi feito um tratamento de dados inicial, de modo a retirar dados que não têm significado para o estudo que está a ser realizado, sendo usados métodos mais grosseiros e uma mediana móvel.

Para a deteção e análise de valores atípicos dos deslocamentos no fio de prumo, foram efetuadas abordagens através das metodologias DBSCAN, LOF, Regressão Linear Múltipla e utilização de coordenadas polares. Por fim efetua-se a comparação dos resultados de cada metodologia, através da análise dos levantamentos gráficos apresentados ao longo deste capítulo.

### 4.2 BARRAGEM DO FEITICEIRO

O aproveitamento hidroelétrico do Baixo Sabor, no rio Sabor, inclui duas grandes barragens que distam cerca de 12,6 e 3,3 km da foz desse mesmo rio no rio Douro. A barragem do Feiticeiro localizada a 3,3 km da confluência do rio Sabor com o rio Douro, tem o intuito de criar um contraembalse para a barragem do Baixo Sabor, localizada a 12,6 km, permitindo regularizar caudais e garantir condições de bombagem a partir do rio Douro.

A barragem do Feiticeiro (Figura 25) é uma barragem gravidade de betão, com traçado retilíneo, com uma altura máxima de 45m, que permite criar uma albufeira com uma capacidade total de 31 hm<sup>3</sup> e capacidade útil de 12 hm<sup>3</sup>. A capacidade total corresponde a um nível de pleno armazenamento (NPA) de 138m, o nível mínimo de exploração (NmE) é de 130 m e o nível de máxima cheia (NMC) coincide com o NPA, 138 m. O coroamento da barragem, situado à cota 140 m, tem um comprimento de 315

m, em alinhamento reto e uma espessura de 8,5 m. O descarregador de superfície tem perfil do tipo WES, com crista à cota 127,00 m, é formado por 4 vãos de 16,00 m de largura, controlados por comportas, garantindo um caudal máximo de descarregamento de 5000 m<sup>3</sup>/s. A dissipação de energia é efetuada numa bacia de dissipação tipo *roller bucket* a jusante. A descarga de fundo, incorporada no muro lateral da margem direita do descarregador de superfície, tem uma capacidade máxima de vazão de 51,5 m<sup>3</sup>/s. A barragem tem ainda dois túneis independentes que conectam o reservatório às duas unidades de energia reversíveis, que estão instaladas na margem direita a jusante da barragem (Monteiro *et al*, 2020; apambiente.pt).

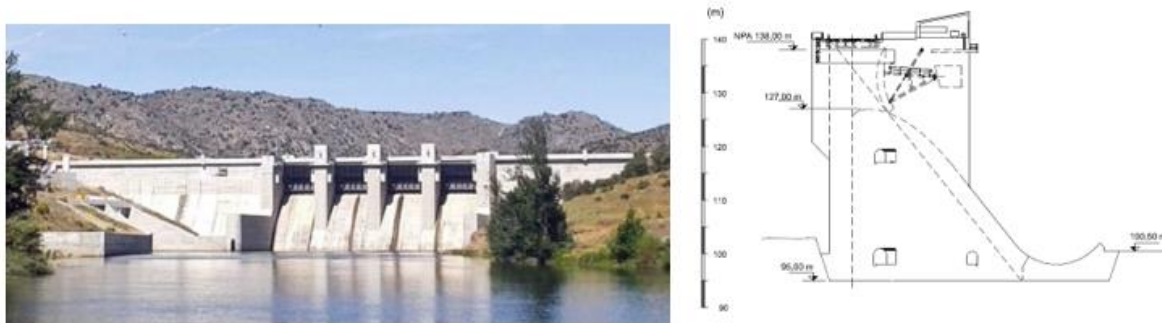


Figura 25 - Foto e esquema de um bloco da barragem do Feiticeiro, adaptado de (Cunha *et al*, 2018).

#### 4.2.1 SISTEMA DE MONITORIZAÇÃO E GRANDEZA EM ANÁLISE

O sistema de monitorização da barragem do Feiticeiro é composto por sistemas de monitorização tradicional, com recolha de dados manual, e automática. Segundo a regulamentação portuguesa de segurança de barragens, concretamente o Regulamento de Segurança de Barragens (RSB) (RSB, 2018) e os Documentos técnicos de apoio ao Regulamento de Segurança de Barragens (APA, 2018), nas barragens de gravidade com as características da barragem do Feiticeiro devem ser monitorizadas as seguintes grandezas:

- Deslocamentos de pontos da barragem e da fundação;
- Movimentos de juntas e fissuras;
- Temperaturas no betão;
- Tensões ou deformações;
- Caudais infiltrados;
- Subpressões;
- Nível da água na albufeira;
- Precipitação e temperatura do ar.

Na barragem do Feiticeiro os deslocamentos são medidos usando um sistema integrado que inclui fios de prumo, extensómetros de varas e observações geodésicas. Os movimentos relativos entre os blocos são medidos em bases tridimensionais e em medidores de movimentos de junta. A deformação do betão é medida com extensómetros elétricos, dispostos em grupos, distribuídos em seções radiais, permitindo a determinação do estado de tensão através do conhecimento do estado de deformação e da deformabilidade do betão.

Os caudais que percolam pela fundação são medidas individualmente em drenos do sistema de drenagem instalado na fundação da barragem e as quantidades de águas drenadas e infiltradas são medidas em bicas, que diferenciam a quantidade total de água que escoar na galeria de drenagem em



várias zonas da barragem. O sistema de drenagem é composto por um conjunto de 57 drenos, distribuídos pela galeria de drenagem com 3 drenos por bloco. A medição da subpressão na fundação é realizada por uma rede piezométrica constituída por 26 piezómetros.

A temperatura do betão é medida com recurso a 76 termómetros de resistência elétrica distribuídos pela espessura da barragem em vários blocos. A localização dos termómetros foi definida levando em consideração o conjunto dos outros dispositivos de resistência elétrica (extensómetros e medidores de movimento de junta embebidos) que também permitem a medição da temperatura do betão.

O sistema utilizado para a medição do nível de água do reservatório é constituído por um medidor de pressão de alta precisão, que fornece um registo da altura da água ao longo do tempo, e uma escala de nível que permite a realização de medições manuais. A temperatura e a humidade do ar são medidas em uma estação meteorológica automatizada localizada na margem direita, aproximadamente a 100 m de distância da crista da barragem (Cunha *et al*, 2017).

Tabela 2 - Sistema de monitorização da barragem do Feiticeiro, adaptado de (Cunha *et al*, 2018).

| Grandezas                           | Equipamentos de observação               | Número de dispositivos de observação |                    |
|-------------------------------------|------------------------------------------|--------------------------------------|--------------------|
|                                     |                                          | Recolha manual                       | Recolha automática |
| Temperatura do ar                   | Estação meteorológica                    | 1                                    | Sim                |
| Humidade relativa do ar             |                                          | 1                                    | Sim                |
| Radiação solar                      |                                          | 1                                    | Sim                |
| Temperaturas no betão               |                                          | 1                                    | Sim                |
| Níveis de água                      | Escala de nível montante                 | 1                                    | Sim                |
|                                     | Escala de nível jusante                  | 1                                    | Sim                |
| Temperaturas no Betão               | Termómetros de resistência elétrica      | 76                                   | 21                 |
| Deslocamentos absolutos e relativos | Bases de coordenómetro (3 fios de prumo) | 9                                    | 4                  |
|                                     | GNSS no coroamento (mais estação remota) | 4                                    | 4                  |
|                                     | Nivelamento do coroamento e nas galerias | 2                                    | -                  |
|                                     | Extensómetros de varas na fundação       | 9                                    | 6                  |
|                                     | Medidores de movimentos de junta         | 38                                   | -                  |
|                                     | Bases tridimensionais                    | 47                                   | 16                 |
| Extensões no betão                  | Extensómetros de resistência elétrica    | 45                                   | 9                  |

|                                |                     |    |    |
|--------------------------------|---------------------|----|----|
| Subpressões na fundação        | Piezómetros         | 26 | 15 |
| Caudais drenados e infiltrados | Drenos de fundação  | 57 | -  |
|                                | Bicas totalizadoras | 4  | 4  |

Dado a barragem possuir um sistema de recolha automática de dados e dada a importância dos deslocamentos no controlo de segurança das barragens de betão, a sua observação é efetuada com uma frequência relativamente elevada, particularmente as observações efetuadas com fios de prumo. As grandezas a analisar no âmbito desta dissertação foram as principais solicitações, tais como o nível de água do reservatório (Figura 27) e a temperatura em alguns pontos representativos do campo térmico no corpo da barragem (Figura 28), e uma resposta estrutural da barragem do Feiticeiro, nomeadamente os deslocamentos horizontais (direção radial), com frequência de medição horária.

A medição automatizada dos deslocamentos horizontais é realizada por telecoordinómetros tipo C1-5.1, representado na Figura 26. Este tipo de aparelho, tem como *modus operandi* a iluminação do fio do prumo por dois díodos emissores de luz pulsante, gerando duas sombras, que são detetadas do outro lado por um sensor de imagem linear, permitindo a determinação da posição do fio através de cálculos matemáticos. As principais características do telecoordinómetro, apresentadas na tabela, são adequadas para as condições ambientais presentes na barragem e o erro máximo esperado devido ao instrumento representado na Tabela 3, igual a 0,1 mm, é adequado para a grandeza dos deslocamentos que ocorrem nesta barragem (Cunha *et al.*, 2017).

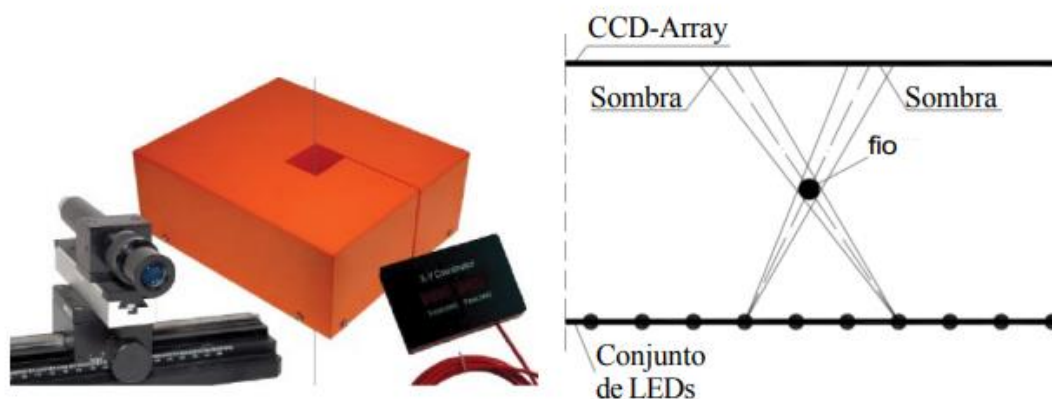


Figura 26 - Telecoordinómetro tipo C1-5.1 e o seu *modus operandi*, adaptado de (Cunha *et al.*, 2017).

Tabela 3 - Características dos telecoordinómetros da barragem do Feiticeiro, adaptado de (Cunha *et al.*, 2017).

|                          |                                          |
|--------------------------|------------------------------------------|
| Gama de medição          | 50x50mm<br>(direção radial e tangencial) |
| Precisão                 | 0,1 mm                                   |
| Saída de sinal analógico | 4-20 mA                                  |
| Humidade relativa        | Até 95 %                                 |

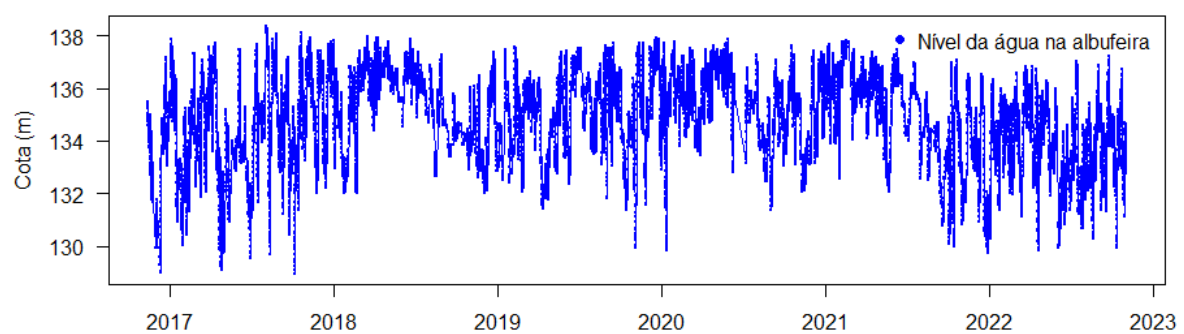


Figura 27 - Observações do nível da água na albufeira.

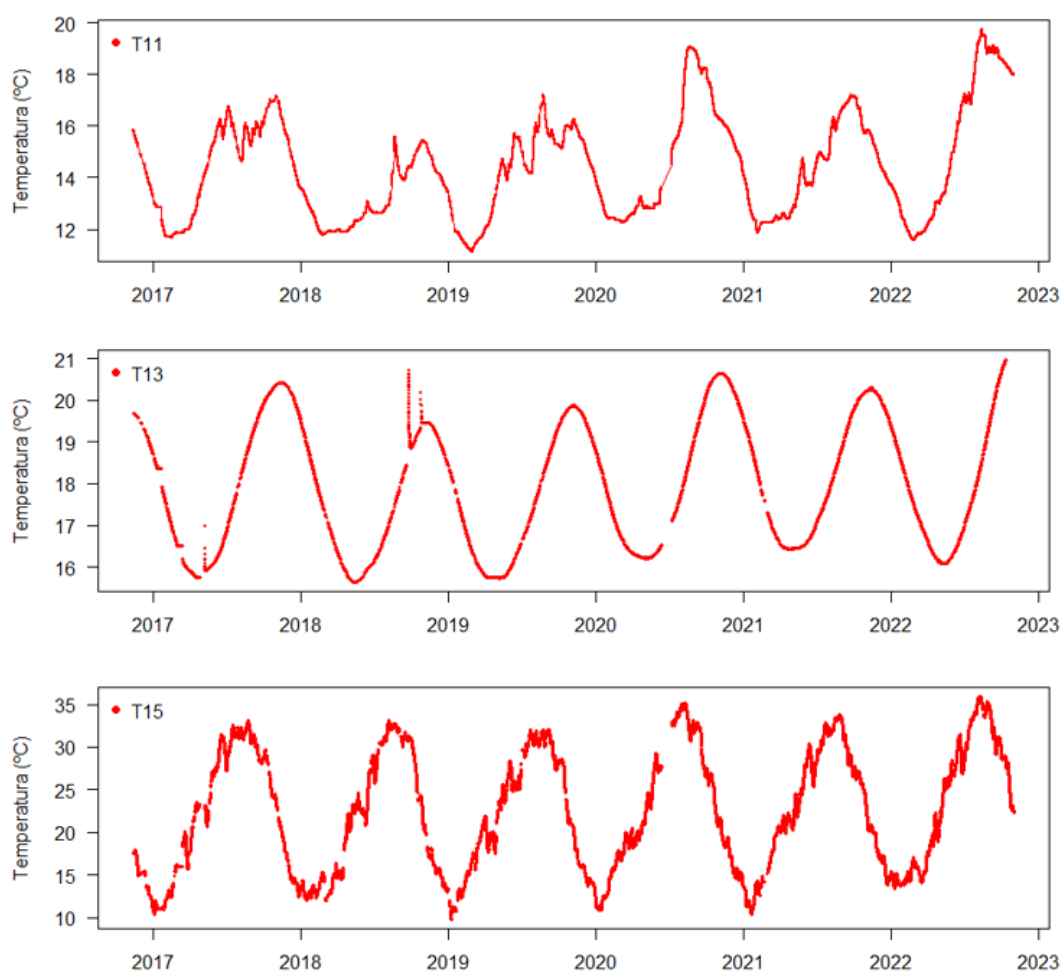


Figura 28 - Temperaturas observadas nos termómetros T11, T13 e T15.

A análise efetuada no presente trabalho compreende as observações obtidas entre o dia 14 de novembro de 2016 e o dia 31 de outubro de 2022, utilizando-se apenas os valores dos deslocamentos, registados na base do fio de prumo FP1 (Figura 29) situada à cota 131,31 m (FP1-131,31 m).

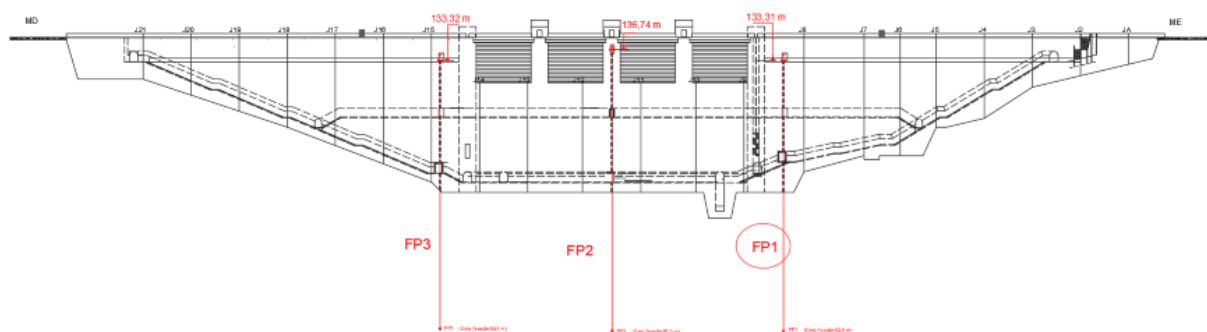


Figura 29 - Representação da posição dos fios de prumo da barragem do Feiticeiro.

#### 4.3 IDENTIFICAÇÃO DE VALORES ATÍPICOS COM OS MODELOS DE CLUSTERING ADOTADOS

Neste estudo de identificação de valores atípicos através de modelos de clustering foram definidas as grandezas a analisar dos valores recolhidos pelos aparelhos de medição, sendo definida como grandeza alvo deste estudo os deslocamentos radiais. Os dados foram introduzidos no *software Rstudio* (R Core Team, 2022), para se dar início ao tratamento dos dados RAD e gerar uma representação gráfica destes valores.

Na Figura 30, estão representados os valores dos deslocamentos radiais observados na base do fio de prumo FP1 à cota 131,31 m, representados pelo eixo das ordenadas em que os valores positivos e negativos representam deslocamentos nas direções de montante e de jusante, respetivamente. No eixo das abcissas estão representadas as datas das medições durante o período de avaliação. Observa-se a existência de valores de deslocamentos com uma grande discrepância, o que não é expectável, uma vez que a recolha automática dos dados é feita com um período horário, as variações de valores deveriam ser mais “suaves”. Essas ocorrências podem estar associadas a erros de medição, como são exemplo os provocados por partículas que se encontram no ar, interferindo com a leitura do sensor de imagem do aparelho.

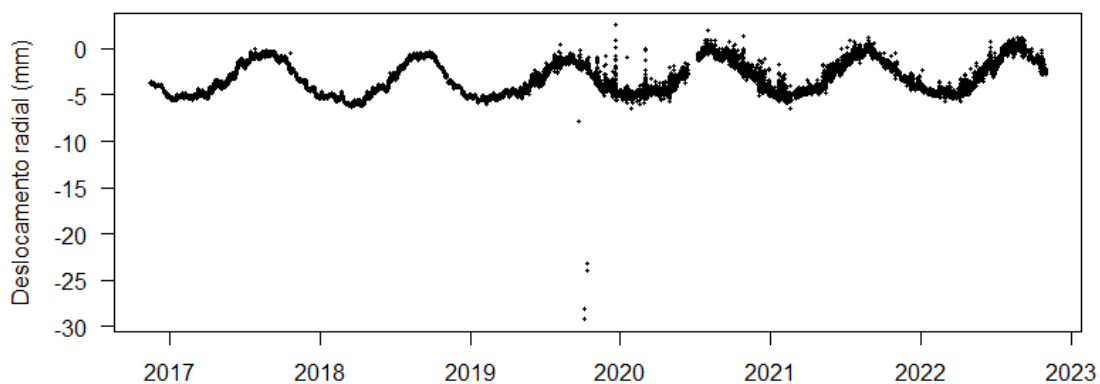


Figura 30 - Valores do deslocamento radial obtidos pelos instrumentos de medição automática no FP1.

Fez-se um primeiro processo de tratamento de dados, onde foram eliminados valores gerados devido a erros de medição grosseiros, tendo sido eliminadas as medições com valores absolutos superiores a  $|d| > 10$  mm, tanto a montante como a jusante. Os resultados deste primeiro procedimento podem ser observados na Figura 31, onde é possível verificar que os valores extremos do deslocamento radial se encontram compreendidos dentro do intervalo imposto.

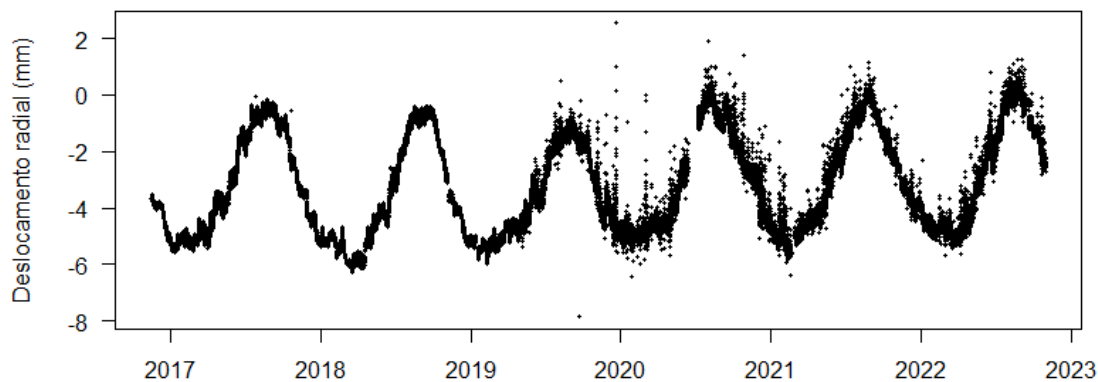


Figura 31 - Valores do deslocamento após remoção de erros grosseiros.

O procedimento seguinte é a aplicação de métodos estatísticos robustos, nomeadamente uma mediana com janela móvel, com a finalidade de identificar e eliminar valores atípicos locais. Foi aplicada a função *Roll\_median()* (Rdocumentation.org), disponibilizada no *software Rstudio* (R Core Team, 2022), de modo a gerar uma mediana com janela móvel de 5 medições. O limite da variação dos valores foi estabelecido utilizando o valor da incerteza de leitura do telecoordinómetro, sendo aplicado o limite de três vezes o valor da incerteza. Esse limite foi aplicado aos valores da diferença entre os deslocamentos radiais medidos pelo aparelho e os valores estimados pela mediana, conforme representado na Figura 32. É possível verificar a representação através de linhas vermelhas, o intervalo dos limites estabelecidos,  $[-0,3;0,3]$  (mm), uma vez que a incerteza do aparelho de medição é de 0,1 mm. A escolha do tamanho da janela móvel a aplicar na mediana, depende da frequência dos dados e da natureza da variável observada, neste caso os deslocamentos radiais. Na sequência dos vários testes realizados, foram obtidas diferentes soluções conforme o valor da janela móvel adotado (Tabela 4). Em termos numéricos, os resultados obtidos foram semelhantes; em comparação com o tamanho da amostra, no entanto, em termos visuais nos gráficos obtidos eram visíveis as diferenças, em que quanto maior o valor da janela móvel, maior o número de valores atípicos identificados. Embora o número de valores atípicos seja maior, também pontos considerados aceitáveis para o comportamento da barragem eram identificados como valores atípicos e retirados da amostra.

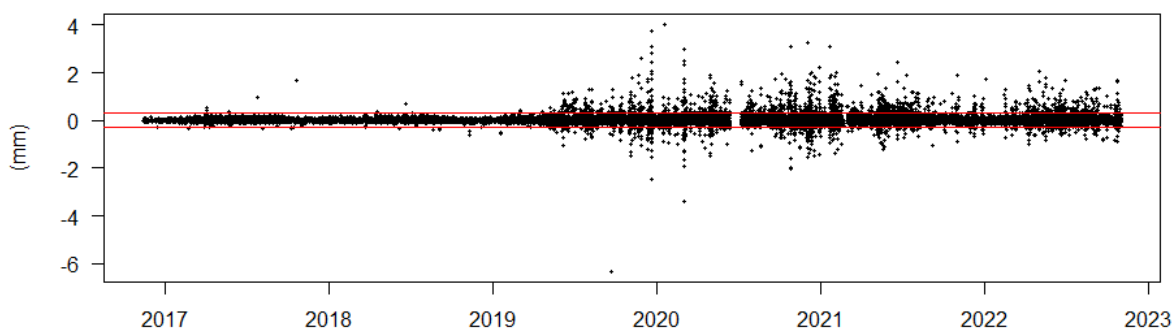


Figura 32 - Limite da variação aplicado na mediana de janela móvel.

Tabela 4 - Valores numéricos da aplicação da mediana com janela móvel.

| Número de dados<br>Total | Tamanho da<br>janela móvel<br>[número de<br>medições] | Número de medições<br>consideradas aceitáveis |
|--------------------------|-------------------------------------------------------|-----------------------------------------------|
| 44161                    | 3                                                     | 43357                                         |
| 44161                    | 4                                                     | 43221                                         |
| 44161                    | 5                                                     | 42983                                         |
| 44161                    | 6                                                     | 42893                                         |
| 44161                    | 7                                                     | 42776                                         |
| 44161                    | 8                                                     | 42684                                         |
| 44161                    | 9                                                     | 42490                                         |
| 44161                    | 10                                                    | 42328                                         |

Foi estabelecida uma mediana com uma janela móvel de 5 medições para analisar grandezas observadas com frequência horária, sendo obtido como resultado o gráfico apresentado na Figura 33. Os pontos representados a preto representam as observações pertencentes à mediana móvel de janela móvel de 5 medições, enquanto os pontos a vermelho representam os pontos que foram considerados pontos atípicos locais. Assumiu-se que as observações fora dos limites definidos, passariam a ser considerados como valores atípicos sendo retirados da amostra, conforme representado na Figura 34.

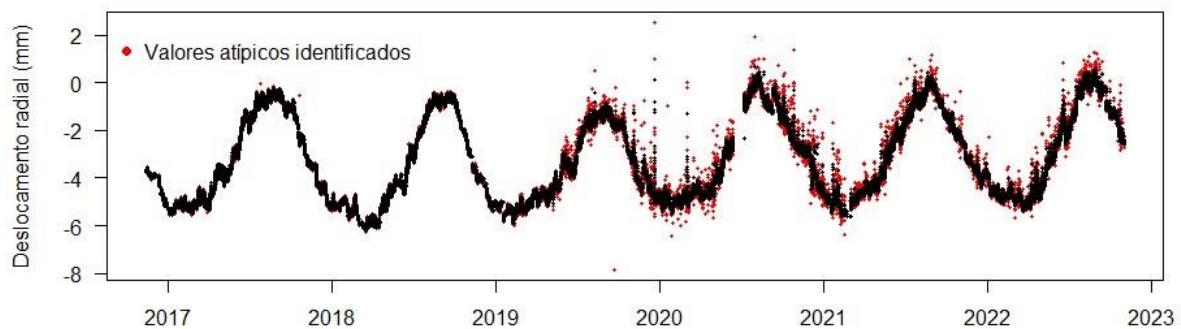


Figura 33 - Identificação de valores atípicos com base na aplicação da mediana numa janela móvel.

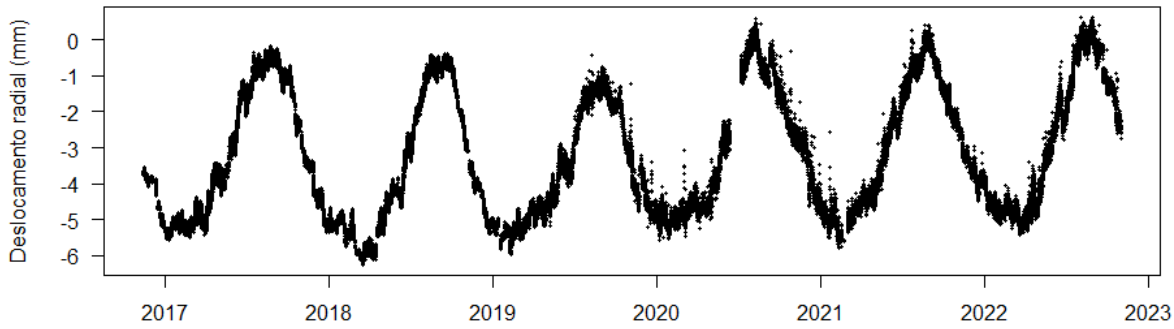


Figura 34 - Valores dos deslocamentos após aplicação da mediana numa janela móvel.

Realizou-se um modelo de regressão do tipo HST, utilizando o método de Regressão Linear Múltipla com os dados resultantes do procedimento anterior, a aplicação mediana de janela móvel de 5 medições. Nesta abordagem foram considerados o efeito do nível da água da albufeira e o efeito térmico, onde para efeito do nível da água da albufeira foi usada a 4ª potência da altura da água representado por,  $\beta_1 \times h^4$ , em que  $h$  representa a altura de água na albufeira (variando entre 0 e 45 m). Para efeitos de cálculo o valor de  $h$  foi normalizado de forma a variar entre zero ( $h=0$  corresponde ao cenário de albufeira vazia) e um ( $h=1$  corresponde ao cenário de a albufeira apresentar uma altura de 45 m). O efeito térmico apresenta a forma de duas funções sinusoidais, nomeadamente  $\beta_2 \times \sin(d) + \beta_3 \times \cos(d)$ , para representar o efeito da variação térmica anual da temperatura, onde  $d = \frac{2 \times \pi \times D}{365}$ , em que  $D$  é o número de dias entre a observação e o início do ano, tendo sido considerando a hora do dia no cálculo de  $D$ . Os parâmetros  $\beta_1$ ,  $\beta_2$  e  $\beta_3$  são os coeficientes a serem estimados pelo modelo, no que resulta na Equação (18):

$$\delta_{HST} = \beta_1 \times h^4 + \beta_2 \times \sin(d) + \beta_3 \times \cos(d) + k + \varepsilon \quad (18)$$

Para a estimativa dos coeficientes, foi utilizada a função de regressão linear disponibilizada no software *Rstudio* (R Core Team, 2022), a função *lm()* (Rdocumentation.org), ao qual recorre ao método dos mínimos quadrados. Através do processo mencionado obteve-se a Equação (19):

$$\delta_{HST} = -2,272897 \times h^4 - 1,729263 \times \sin(d) - 1,464927 \times \cos(d) + 0,5054 \quad (19)$$

Na Figura 35 estão apresentados os gráficos com o resultado do efeito do nível da água e efeito térmico dos deslocamentos radiais. No gráfico representativo do efeito do nível da água (esquerda), o eixo das abcissas representa a altura de água na albufeira, no eixo das ordenadas está representado o deslocamento relativo do fio de prumo correspondente ao efeito do nível. Note-se que a função do deslocamento do fio de prumo, representada pela linha azul, aumenta em valores negativos quanto maior a altura da água na albufeira. Isso traduz-se num maior deslocamento radial no sentido de jusante da barragem para níveis mais elevados de água na albufeira.

O efeito térmico, representado na Figura 35 à direita, apresenta no eixo das abcissas os meses do ano relativamente a um ano e no eixo das ordenadas o deslocamento relativo observado no fio de prumo. A vermelho, esta representada a variação do deslocamento radial ao longo do ano, sendo possível ver que há uma variação sazonal, em correspondência com as estações do ano, devido as variações de temperatura do ar que nas mesmas se fazem sentir.



Na Figura 36 estão representados os valores dos resíduos, que resultam da diferença entre os valores previstos pelo modelo e os valores observados.

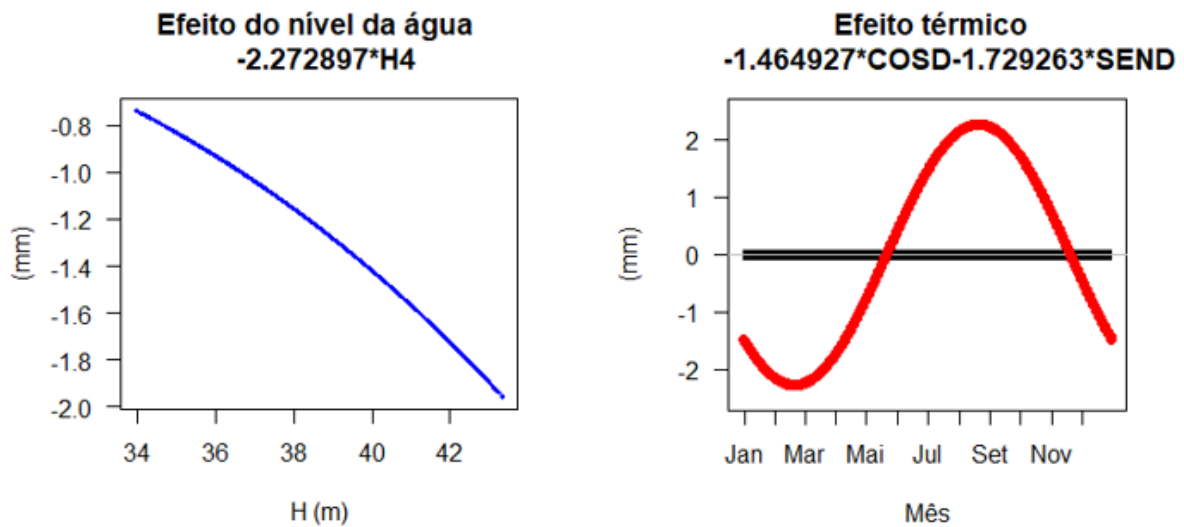


Figura 35 - Efeito do nível da água e efeito térmico nos deslocamentos radiais.

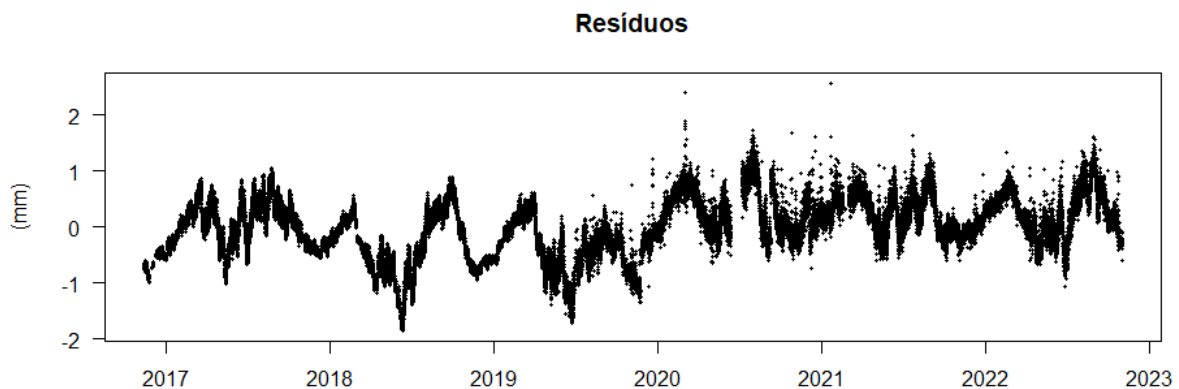


Figura 36 - Resíduos obtidos no modelo de Regressão Linear Múltipla.

De modo a definir padrões únicos e aumentar o desempenho dos métodos de *Machine Learning* a serem utilizados, os resíduos são divididos em dois subconjuntos, um com valores positivos e nulos e consequentemente outro com valores negativos. Fez-se também a conversão de coordenadas cartesianas para coordenadas polares, de modo a caracterizar o padrão dos resíduos ao longo de ciclos de um ano, conforme representado na Figura 37. Essa transformação permite em vez de lidar com as coordenadas cartesianas em duas dimensões (x e y), a representação polar permite uma descrição mais concisa com apenas duas informações: valor dos resíduos e dia do ano da observação. A transformação em coordenadas polares facilita a deteção de valores atípicos que se afastam dessa estrutura circular, já que eles podem ser observados como pontos que se desviam significativamente da maioria dos dados (centrados em zero).

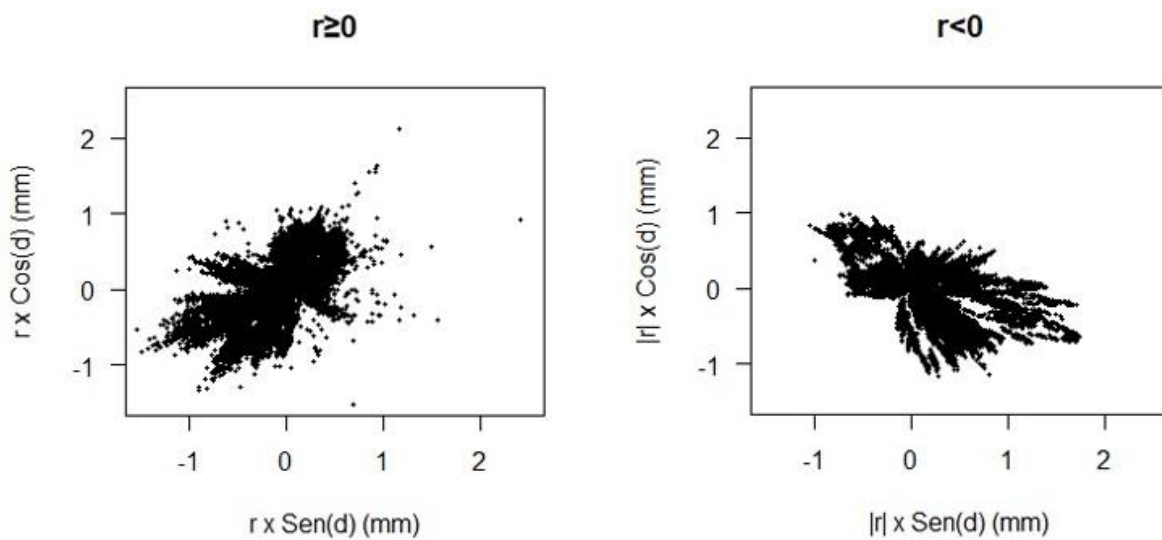


Figura 37 - Resíduos positivos e resíduos negativos representados em coordenadas polares.

Deu-se início à aplicação do métodos de *clustering* DBSCAN, anteriormente abordado no capítulo 3.6, utilizando a função *dbscan()* (Rdocumentation.org), disponibilizado pelo *software Rstudio* (R Core Team, 2022). Os dados inseridos no algoritmo foram os resíduos provenientes da Regressão Linear Múltipla convertidos em coordenadas polares, tendo sido adotado um parâmetro *eps* de 0,1 e um parâmetro de *minPts* de 50. Na Figura 38 estão ilustrados os valores atípicos identificados pelo algoritmo DBSCAN com representação em coordenadas polares, à esquerda estão representados os resíduos positivo e nulos e à direita os resíduos negativos. No caso dos resíduos positivos e nulos o método DBSCAN identificou dois *clusters*, escolhendo-se o *cluster* com os pontos mais próximos do centro como valores aceitáveis (representado pela cor salmão) e o *cluster* com valores mais distantes do centro como valores atípicos (representado pela cor preta). Para os resíduos negativos foram identificados quatro *clusters*, tendo sido adotado o mesmo critério utilizado no caso dos resíduos positivos e nulos: o *cluster* com os pontos mais próximos do centro é considerado como valores aceitáveis (representado pela cor salmão) e os *clusters* com os valores mais afastados são valores atípicos (representado pelas cores preta, verde e azul). Na Figura 39 é feita a representação dos valores atípicos identificados pelo algoritmo DBSCAN nos deslocamentos radiais, sendo apresentado em cima os casos dos resíduos positivos e nulos e em baixo os casos dos resíduos negativos.

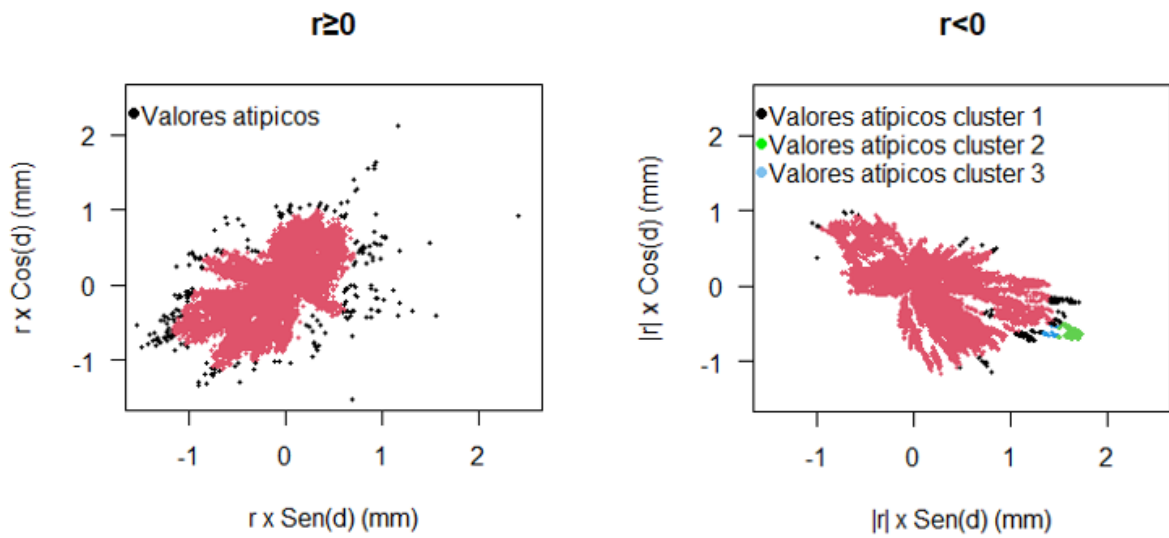


Figura 38 - Valores atípicos identificados pelo DBSCAN (eps=0,1 e minPts=50) representados em coordenadas polares.

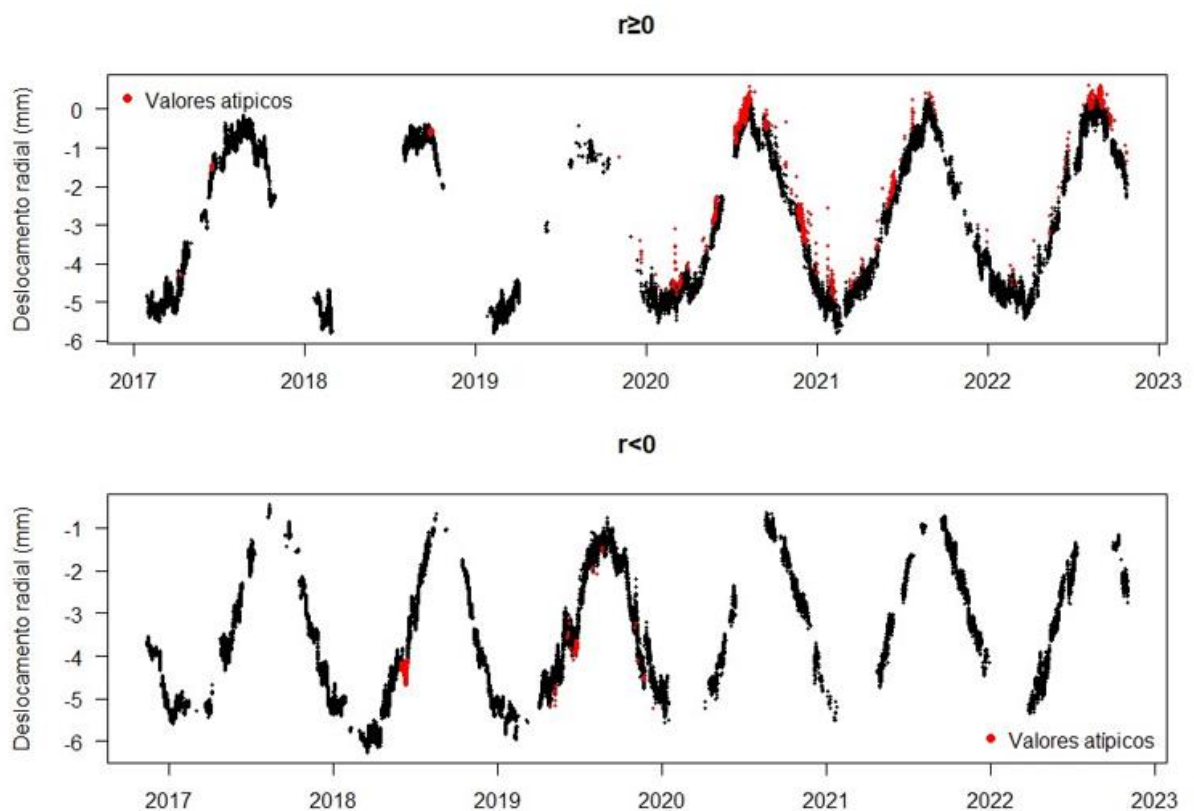


Figura 39 - Valores atípicos identificados pelo método DBSCAN (eps=0,1 e minPts=50) representados nos deslocamentos radiais.

De forma a facilitar a análise dos resultados obtidos, é feita uma representação gráfica do método com a junção dos valores atípicos identificados pelo algoritmo DBSCAN nos deslocamentos radiais, na Figura 40.

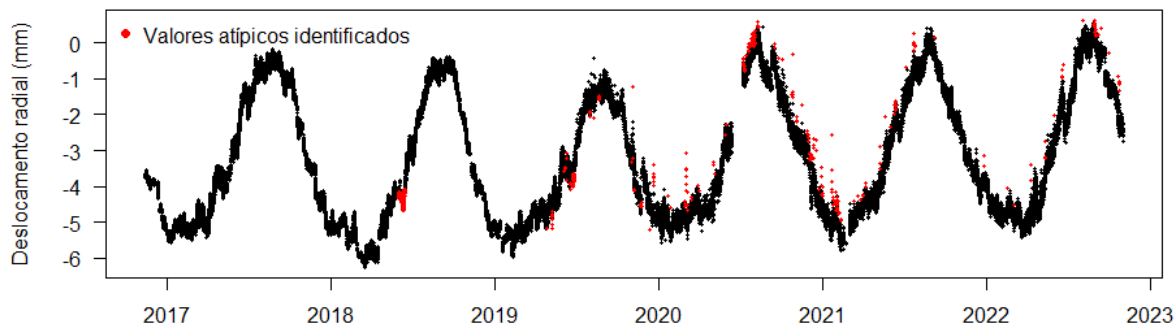


Figura 40 - Valores atípicos identificados pelo método DBSCAN (eps=0,1 e minPts=50).

Testou-se também o método LOF tendo como ponto de partida a etapa prévia à da utilização do método DBSCAN. Deste modo, procedeu-se início à aplicação do método de *clustering* LOF, introduzido no capítulo 3.7, com recurso ao *software Rstudio* (R Core Team, 2022), utilizando a função *lof()* (Rdocumentation.org). Da mesma forma que no algoritmo DBSCAN, foram utilizados os resíduos provenientes da Regressão Linear Múltipla convertidos em coordenadas polares, adotando o parâmetro minPts igual a 150. Para a escolha dos fatores LOF mais adequados para a identificação de valores atípicos é utilizado um método empírico, utilizando-se um critério semelhante ao método do cotovelo, onde é escolhido o valor onde existe uma variação grande do declive da reta tangente ao gráfico, como representado na Figura 41. Para os valores dos resíduos positivos e nulos é adotado um fator LOF de 2 (gráfico de cima), e para os valores dos resíduos negativos um fator LOF de 1,5 (gráfico de baixo).

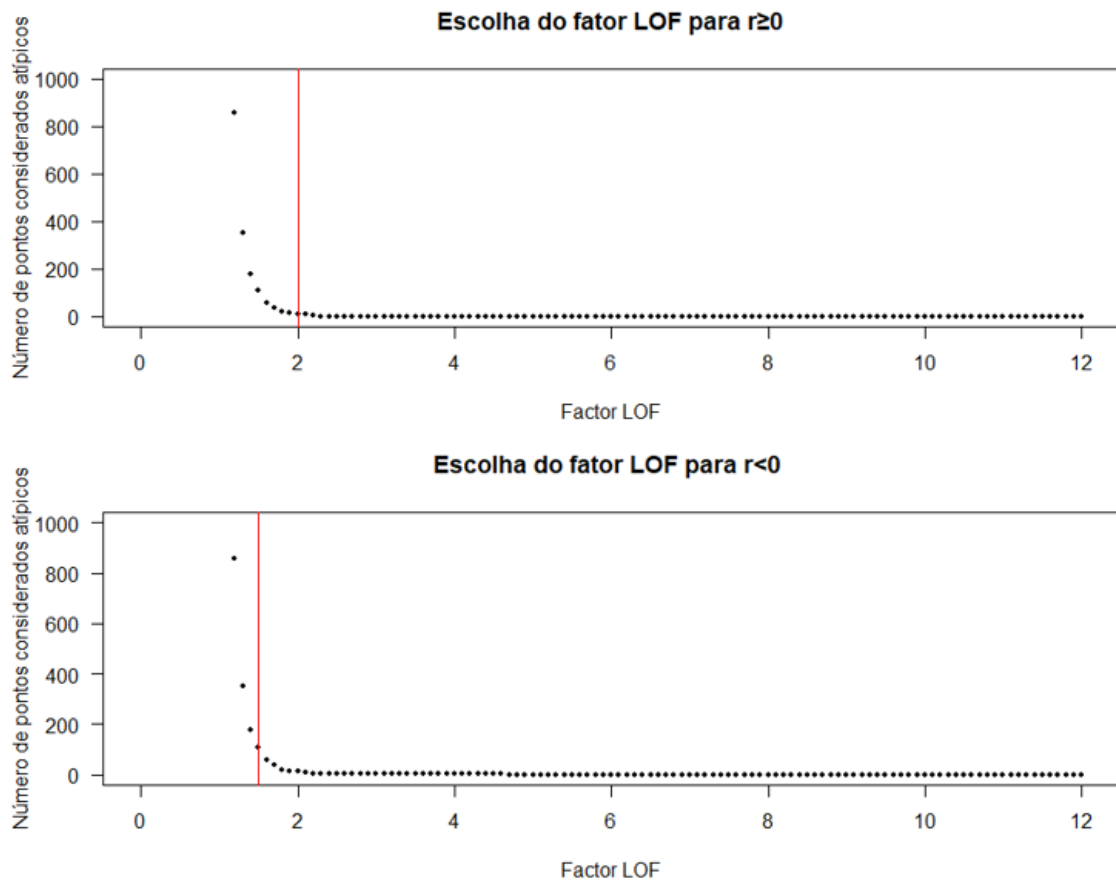


Figura 41 - Critério de escolha do fator LOF (minPts=150).

Utilizando-se os valores do fator LOF provenientes da etapa anterior, os valores atípicos identificados pelo algoritmo LOF com representação em coordenadas polares estão representados pela Figura 42, onde à esquerda são apresentados os resíduos positivo e nulos e à direita os resíduos negativos, com os valores atípicos estão representados pela cor vermelha. Na Figura 43 são apresentados os valores atípicos identificados pelo algoritmo LOF nos deslocamentos radiais, sendo apresentado em cima os casos dos resíduos positivos e nulos e em baixo os casos dos resíduos negativos.

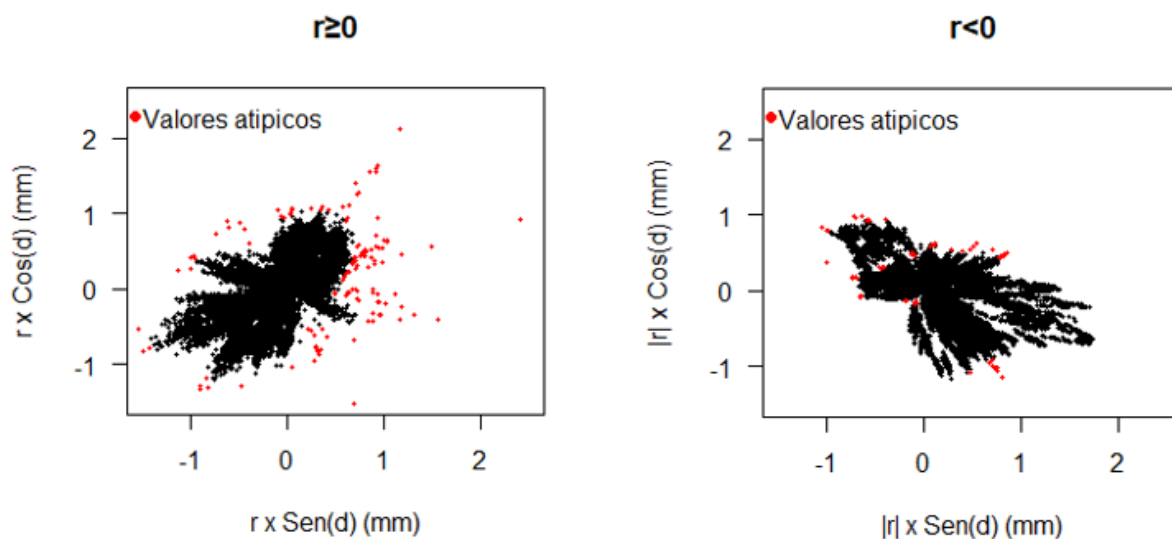


Figura 42 - Valores atípicos identificados pelo método LOF (minPts=150) representados em coordenadas polares.

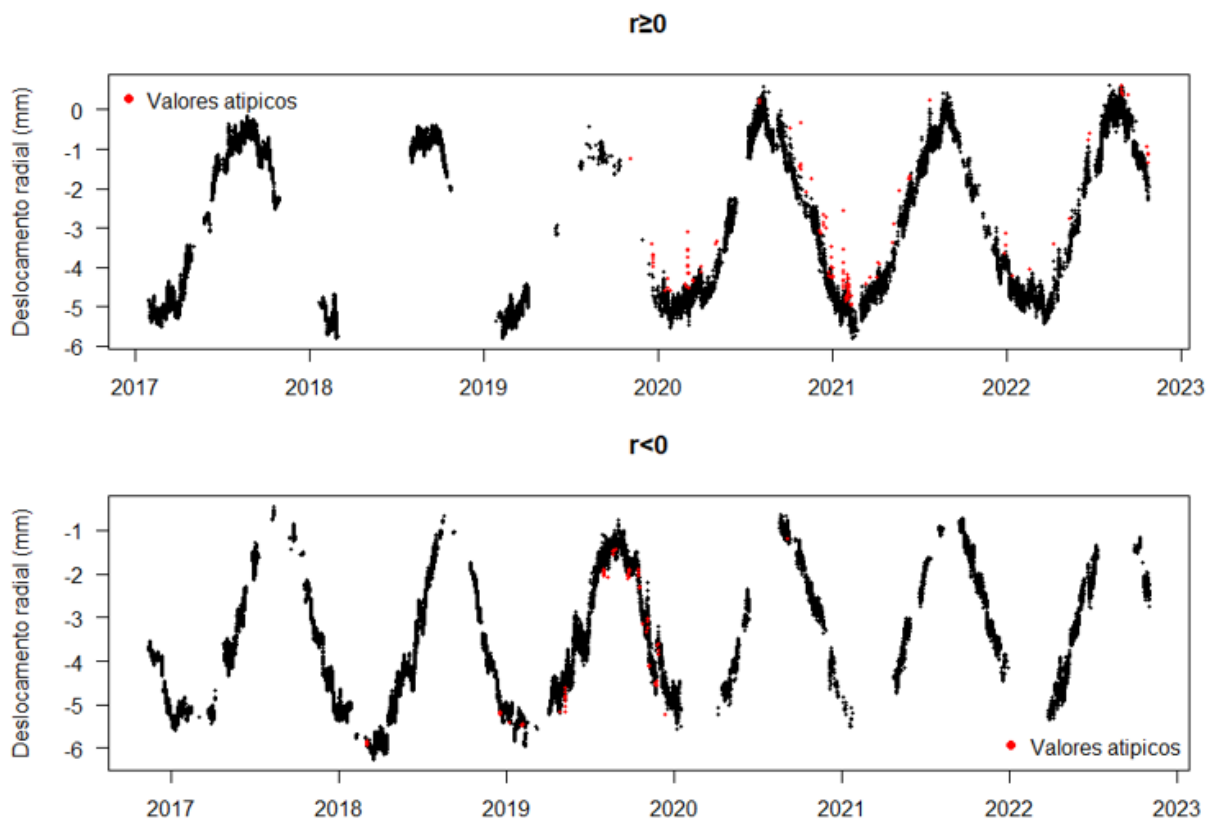


Figura 43 - Valores atípicos identificados pelo método LOF (minPts=150) representados nos deslocamentos radiais.

Novamente, de forma a facilitar a análise dos resultados, é feita uma representação gráfica do método com a junção dos valores atípicos nos deslocamentos radiais identificados pelo algoritmo LOF (Figura 44).

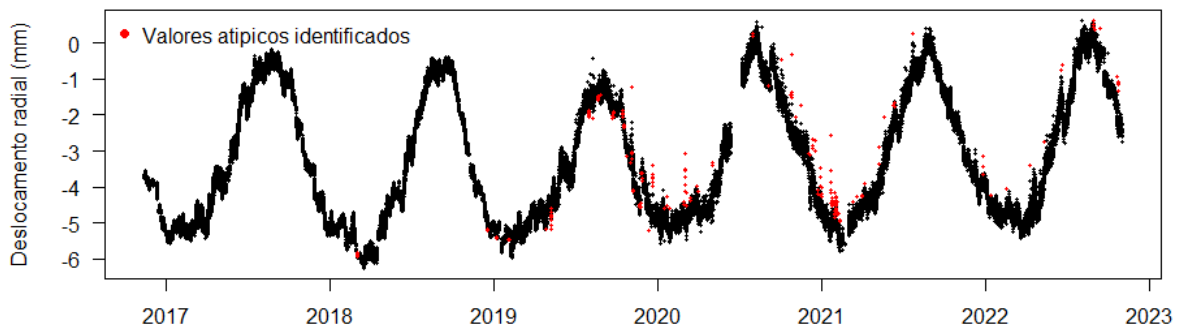


Figura 44 - Valores atípicos identificados pelo método LOF (minPts=150).

#### 4.4 ANÁLISE DE SENSIBILIDADE DO PARÂMETRO MINPTS DOS MODELOS DBSCAN E LOF

Como já foi referido anteriormente, os métodos de *clustering* DBSCAN e LOF são ambos baseados em aprendizagem não supervisionada e no conceito de densidade, no entanto, como apresentam metodologias diferentes, obtêm-se resultados diferentes.

Na Tabela 5 e Tabela 6 estão apresentados, respetivamente, o número de medições consideradas válidas com recurso aos métodos DBSCAN e LOF utilizando-se vários valores para os parâmetros minPts. Note-se que no caso do caso do DBSCAN o valor numérico de observações, consideradas como valores atípicos aumenta com o aumento do parâmetro minPts. O contrário é verificado no método LOF, onde o número de valores atípicos identificados, aumenta com o aumento do parâmetro minPts, até que chega a um pico e volta a diminuir. É possível verificar graficamente estes factos na Figura 45 e Figura 46.

Embora para o método DBSCAN, em princípio, pareça benéfico utilizar um parâmetro minPts maior, após a análise gráfica dos valores atípicos é possível concluir que o aumento de valores identificados não acarreta grandes benefícios. Como é apresentado na Figura 47, o aumento do número de valores atípicos para o caso dos parâmetros  $\text{eps}=0,1$  e  $\text{minPts}=90$ , verifica-se, em regra, maioritariamente nas zonas circundadas a vermelho que representa uma zona de pontos considerados aceitáveis para o comportamento da barragem. A posterior remoção desses pontos pode causar uma falta de dados durante o período considerado.

No caso do método LOF, a escolha de um parâmetro minPts que apresenta um número de pontos atípicos superior, pode também, revelar-se não ser benéfico. Tendo como exemplo a escolha do parâmetro  $\text{minPts}=10$ , que apresenta um número de valores atípicos superior ao utilizado ao apresentado no capítulo 4.3, analisando graficamente (Figura 48) podemos constatar que o mesmo acontece como previamente no método DBSCAN. Para este parâmetro o aumento de valores atípicos corresponde a pontos considerados aceitáveis para o comportamento da barragem.

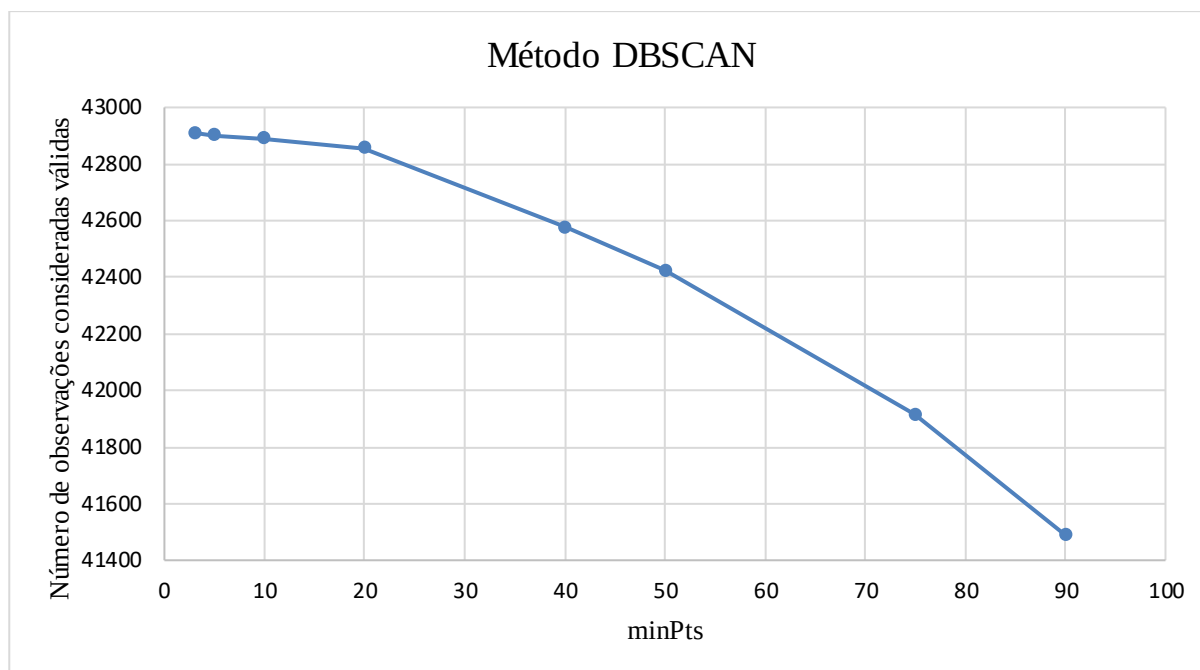


Figura 45 - Resultados numéricos DBSCAN (para  $\text{eps}=0,1$ ).

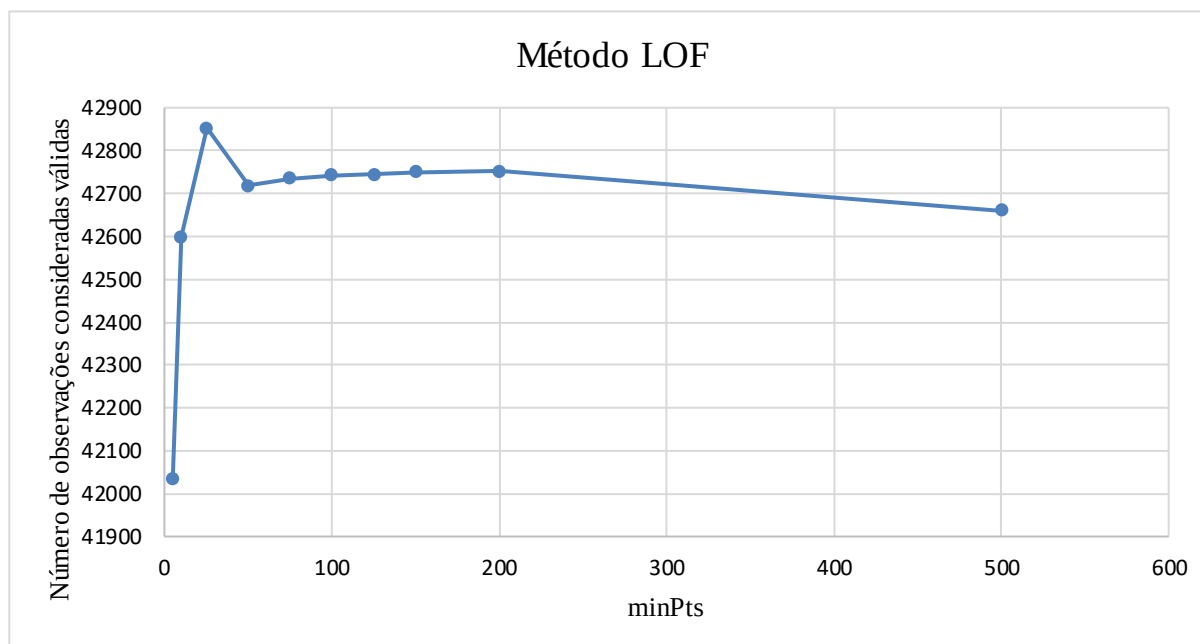


Figura 46 - Resultados numéricos LOF.



Tabela 5 - Resultados obtidos DBSCAN.

| Número de dados de entrada <sup>(1)</sup> | eps | minPts | Número de dados saída do Dbscan | Número de valores atípicos identificados |
|-------------------------------------------|-----|--------|---------------------------------|------------------------------------------|
| 42983                                     | 0,1 | 3      | 42910                           | 73                                       |
| 42983                                     | 0,1 | 5      | 42902                           | 81                                       |
| 42983                                     | 0,1 | 10     | 42889                           | 94                                       |
| 42983                                     | 0,1 | 20     | 42857                           | 126                                      |
| 42983                                     | 0,1 | 40     | 42577                           | 406                                      |
| 42983                                     | 0,1 | 50     | 42424                           | 559                                      |
| 42983                                     | 0,1 | 75     | 41914                           | 1059                                     |
| 42983                                     | 0,1 | 90     | 41489                           | 1494                                     |

Tabela 6 - Resultados obtidos LOF.

| Número de dados de entrada <sup>(1)</sup> | minPts | Número de dados de saída do LOF | Número de valores atípicos identificados |
|-------------------------------------------|--------|---------------------------------|------------------------------------------|
| 42983                                     | 5      | 42036                           | 947                                      |
| 42983                                     | 10     | 42598                           | 385                                      |
| 42983                                     | 25     | 42853                           | 130                                      |
| 42983                                     | 50     | 42718                           | 265                                      |
| 42983                                     | 75     | 42735                           | 248                                      |
| 42983                                     | 100    | 42742                           | 241                                      |
| 42983                                     | 125    | 42745                           | 238                                      |
| 42983                                     | 150    | 42749                           | 234                                      |
| 42983                                     | 200    | 42751                           | 232                                      |
| 42983                                     | 500    | 42660                           | 323                                      |

Nota: (1) Após remoção dos valores atípicos grosseiros.

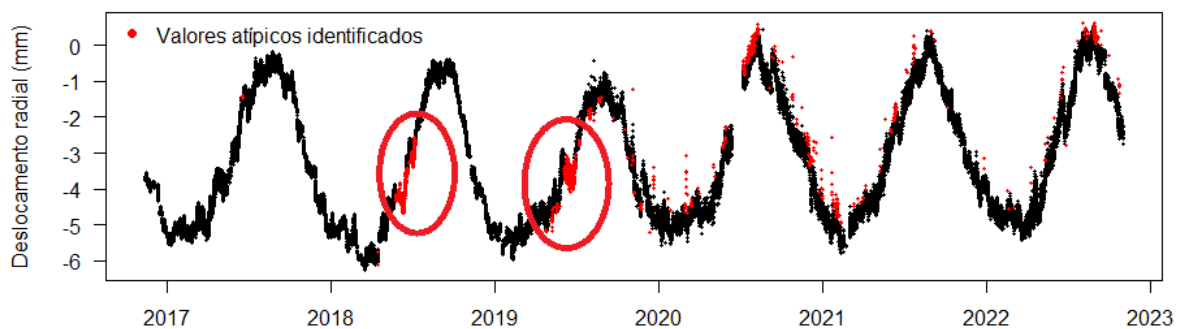


Figura 47 - Valores atípicos identificados pelo método DBSCAN (eps=0,1 e minPts=90).

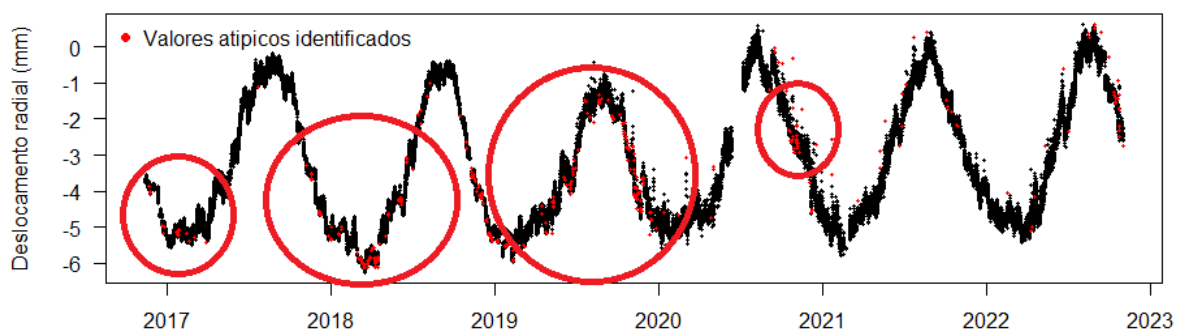


Figura 48 - Valores atípicos identificados pelo método LOF (minPts=10).

# 5

## CONSIDERAÇÕES FINAIS

### 5.1 TRABALHO REALIZADO

A aplicação de modelos de *Machine Learning* oferece uma abordagem promissora para o controlo e monitorização contínua das barragens. Neste trabalho foram desenvolvidos métodos de *Machine Learning*, em particular na identificação de valores atípicos, para analisar numa das principais grandezas observadas em barragens de betão.

No capítulo 2 realizou-se uma introdução histórica da evolução e importância das barragens para as populações. Apresentou-se os tipos mais comuns de barragens de betão e suas características principais. É destacada a importância da monitorização e controlo da segurança das barragens, com a apresentação dos regulamentos de segurança atualmente em vigor em Portugal e alguns casos históricos de acidentes e incidentes envolvendo barragens. Realizou-se também uma análise das principais ações presentes numa barragem de betão, bem como do comportamento estrutural esperado. São descritas técnicas e dispositivos de observação das grandezas envolvidas. Além disso, é abordada a análise com referência a modelos de interpretação quantitativa e a identificação de valores atípicos.

O capítulo 3 concentra-se nos métodos de *Machine Learning*, apresentando os tipos e os conceitos fundamentais para o seu funcionamento. Foram abordados métodos de Machine Learning dos vários tipos, no entanto focou-se nos métodos utilizados neste trabalho: A Regressão Linear Múltipla, DBSCAN e LOF. Explicou-se o seu funcionamento e quais são as suas aplicações. Foi também apresentada a metodologia proposta para a identificação de valores atípicos em grandezas observadas em barragens de betão.

No capítulo 4 é apresentado o caso de estudo. O estudo foi aplicado à barragem do Feiticeiro, sendo feita uma descrição geral da mesma e do seu sistema de monitorização. Foram apresentados os dados resultantes do sistema de monitorização da barragem que foram utilizados para o desenvolvimento do trabalho, nomeadamente os deslocamentos radiais observados num fio de prumo. Posteriormente fez-se a remoção de valores dos erros de medições grosseiros e aplicaram-se métodos estatísticos robustos, mais propriamente uma mediana de janela móvel para remoção de valores atípicos locais. Foi elaborado um modelo de Regressão Linear Múltipla com abordagem HST com a finalidade de se obterem os resíduos. Seguidamente esses mesmos resíduos foram usados como dados de entrada para a aplicação dos métodos de DBSCAN e LOF, sendo identificados valores atípicos, efetuando-se a sua representação de forma gráfica.

Por fim, realizaram-se análises aos resultados obtidos pelos dois métodos, fazendo a comparação entre eles.

## 5.2 DISCUSSÃO DOS DADOS OBTIDOS

Este trabalho teve como objetivo principal investigar o uso de modelos de *Machine Learning* para identificar valores atípicos no comportamento observado em barragens de betão. A aplicação destes modelos de identificação de valores atípicos permite concluir que, por um lado foram obtidos resultados vantajosos, mas também foram identificadas algumas desvantagens.

A utilização de recolha de dados automáticos apresenta um potencial positivo, como um maior número de dados recolhidos, garantindo uma frequência uniforme em todos os tipos de instrumentos de medição, tornando possível a sincronização das medições dos vários aparelhos de medição. Leva também à redução de erros de medição cometidos por técnicos. Este tipo de recolha de dados também apresenta desvantagens como o custo dos sistemas de medição, a necessidade de manutenção contínua de modo a garantir a qualidade das medidas e a ocorrência de erros de leitura.

Os métodos de tratamento de dados iniciais revelaram ser positivos para a aplicação dos métodos de identificação de valores atípicos, fazendo uma limpeza de valores grosseiros na base de dados inicial que podiam ser prejudiciais, para a qualidade do modelo HST desenvolvido e por consequência obter-se-iam resíduos de maior valor (e dispersão).

Os valores dos resíduos obtidos através do método de Regressão Linear Múltipla com abordagem HST foram estratégicos, permitindo serem organizados em subconjuntos e convertidos em coordenadas polares.

A aplicação do algoritmo de *clustering* DBSCAN aos valores dos resíduos obtidos através do método de Regressão Linear Múltipla com abordagem HST, revelou-se positivo, quando são utilizados os parâmetros adequados. Quando são usados parâmetros inadequados, o resultado final na identificação de valores atípicos não se revelou muito vantajosa, uma vez que são identificados e classificados como valores atípicos muitos valores considerados normais no comportamento da barragem. Com os parâmetros corretos obteve-se um bom desempenho ajudando a complementar os primeiros tratamentos de dados realizados pela mediana de janela móvel.

A utilização do algoritmo de *clustering* LOF obteve um desempenho semelhante ao do algoritmo de *clustering* DBSCAN, sendo também uma mais-valia quando aplicado com parâmetros adequados e com pior desempenho quando usados parâmetros inadequados.

No entanto estes métodos também abrangem valores considerados normais no comportamento da barragem, precedendo à sua eliminação. Embora isto aconteça o número de valores considerados normais no comportamento da barragem são de uma escala pouco significativa comparado com a amostra total e os valores considerados atípicos.

De uma forma geral, os métodos propostos revelaram-se positivos quando usados adequadamente, contribuindo para uma identificação mais precisa de valores atípicos de deslocamentos radiais, tendo sempre de haver cautelas em relação aos parâmetros utilizados.

## 5.3 DESENVOLVIMENTOS FUTUROS

Como desenvolvimentos futuros, propõe-se a aplicação dos modelos desenvolvidos nesta dissertação, focando-se em outras grandezas observadas na barragem, tais como subpressões e ações dinâmicas.

Considera-se pertinente a exploração de diferentes técnicas de *Machine Learning* além das abordadas nesta dissertação, como redes neurais convolucionais (CNN) e redes neurais recorrentes (RNN) e comparar com os resultados obtidos nesta dissertação.

Sugere-se ainda a implementação de sistemas de alerta precoce com base nos modelos de *Machine Learning* que possam notificar os engenheiros responsáveis pela segurança estrutural de barragens de betão sobre possíveis comportamentos atípicos ou anomalias, permitindo uma ação rápida e eficaz para prevenir falhas ou danos.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Ajith, A., Dutta, P., Mandal, J. K., Bhattacharya, A., & Dutta, S. (2018). *Emerging Technologies in Data Mining and Information Security*, Volume 2, Springer, Singapura.
- Alpaydin E. (2020). *Introduction to Machine Learning*. Fourth ed, MIT Press, Cambridge Massachusetts.
- APA. (2018). *Documentos técnicos de apoio ao Regulamento de Segurança de Barragens (RSB)*. Comissão dos Regulamentos de Barragens (2008-2016). 1ª Edição. Agência Portuguesa do Ambiente, I.P., abril de 2018.
- apambiente.pt, [Barragens de Portugal | Agência Portuguesa do Ambiente \(apambiente.pt\)](https://apambiente.pt/Barragens-de-Portugal-Agência-Portuguesa-do-Ambiente-apambiente.pt), consultado em 15 de março de 2023.
- Asthana, B.N., Khare, D. (2022). *Dam Safety*. In: *Recent Advances in Dam Engineering*. Springer, Cham.
- Bhavsar, P., Safro, I., Bouaynaya, N., Polikar, R., & Dera, D. (2017). *Machine learning in transportation data analytics*. In *Data analytics for intelligent transportation systems*, Transportation Systems, 31 de dezembro de 2017, pág. 283–307, Elsevier, Clemson.
- Bholowalia, P., & Kumar, A. (2014). *EBK-means: A clustering technique based on elbow method and k-means in WSN*. International Journal of Computer Applications, Volume 105, No.:9, novembro de 2014, pág. 17-24, Gallen.
- Breunig, M. M., Kriegel, H. P., Ng, R. T., & Sander, J. (2000). *LOF: identifying density-based local outliers*. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 16 a 18 de maio de 2000, Dallas, pág. 93-104.
- Cardoso, J. (2021). *Modelação do comportamento observado em barragens de betão com base em técnicas de aprendizagem automática*. Dissertação de Mestrado em Engenharia de Estruturas, Instituto Superior Técnico Lisboa.
- Carmo, J. (2013). *Grandes barragens: vulnerabilidades e riscos*. In *Riscos naturais, antrópicos e mistos*, pág. 441-461, Departamento de Geografia. Faculdade de Letras, Universidade de Coimbra.
- Chen, S. H., & Chen, M. L. (2015). *Hydraulic structures*. Springer, Berlin, Heidelberg.
- Cunha, J., Gomes, J., Moura, G., Batista, A. (2018). *A monitorização estrutural das barragens do Baixo Sabor e do Feiticeiro*. Encontro Nacional Betão Estrutural - BE2018, LNEC, 7 a 9 de novembro de 2018, Lisboa.
- Cunha, J., Mata, J., & Gonzalo, O. (2017). *Structural safety control of the Feiticeiro dam based on an automated data acquisition system*. 85th Annual Meeting of International Commission on Large Dams, 3 a 7 de julho de 2017, Praga, pág. 3-7.
- da Silva, J. M. R. M. (2013). *Injeção de juntas de contração de barragens abóbada*. Dissertação de Mestrado em Engenharia de Estruturas, Universidade do Porto.
- Duffaut, P. (2013). *The traps behind the failure of Malpasset arch dam, France, in 1959*. Journal of Rock Mechanics and Geotechnical Engineering, Vol. 5, outubro de 2013, pág. 335-341.
- Escovedo, T., & Koshiyama, A. S (2020). *Introdução a Data Science - Algoritmos de Machine Learning e métodos de análise*. Casa do Código, São Paulo.

- Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). *A density-based algorithm for discovering clusters in large spatial databases with noise*. 2nd International Conference on Knowledge Discovery and Data Mining, Volume 96, 2 a 4 de agosto de 1996, Portland, pág. 226-231.
- Faraoun, K. M., & Boukelif, A. (2007). *Neural networks learning improvement using the k-means clustering algorithm to detect network intrusions*. International Journal of Computer, Electrical, Automation, Control and Information Engineering, Volume 1, No.:10, janeiro de 2007, pág. 3151-3158.
- Fernandes, A. B. G. (2016). *Novos Sistemas de Observação Geodésica de Barragens*. Dissertação de Mestrado em Engenharia Geográfica, Universidade do Porto.
- Gasch, A. P., Eisen, M. B. (2002) *Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering*. Genome Biology, Volume 3, No.:11, 10 de outubro de 2002, pág. 1-22, Berkeley.
- Gleick, P., Palaniappan, M., Morikawa, M. & Morrison, J. (2009). *Three Gorges Dam Project, Yangtze River, China*. The World's Water 2008-2009: The Biennial Report on Freshwater Resources. Bibliovault OAI Repository, the University of Chicago Press., pág. 139-150.
- Hahsler, M., Piekenbrock, M., & Doran, D. (2019). *dbscan: Fast density-based clustering with R*. Journal of Statistical Software, Volume 91, outubro de 2019, pág. 1-30, Austria.
- Haykin, S. (2009). *Neural Networks and Learning Machines*, 3rd Edition, Pearson Education Inc., Nova Jersey.
- Hawkins, D. M. (1980). *Identification of outliers (Vol. 11)*. Chapman and Hall, London.
- Hegde, N.N., Nagananda, M.S., Harsha, M. (2015). *EEG signal classification using K-means and FCM clustering methods*. IJSTE - International Journal of Science Technology & Engineering, Volume 2, Edição 1, julho de 2015, Ahmedabad.
- Henriques, C. (2011). *Análise de regressão linear simples e múltipla*. Departamento de Matemática. Escola Superior de Tecnologia de Viseu, Viseu.
- Learned-Miller, E. G. (2014). *Introduction to supervised learning*. Department of Computer Science, University of Massachusetts, Amherst.
- Lee, I.; Shin, Y. (2020). *Machine learning for enterprises: Applications, algorithm selection, and challenges*. Business Horizons, Volume 63, Edição 2, março-abril de 2020, pág. 157-170, Indiana.
- LNEC. (2016). *Barragem do Feiticeiro: Inspeção à estrutura e ao sistema de observação efetuada em março de 2015, após o primeiro enchimento da albufeira*. Nota Técnica 22/2016 LNEC, março de 2016, Laboratório Nacional de Engenharia Civil, Lisboa.
- LNEC. (2020). *Barragem do Feiticeiro: Inspeção à estrutura e ao sistema de observação efetuada em março de 2020*. Nota Técnica 149/2020 LNEC, novembro de 2020, Laboratório Nacional de Engenharia Civil, Lisboa.
- Loureiro, M. A. O. (2018). *Estudo de subpressões em barragens de concreto no contexto da segurança de barragens*. Dissertação de Mestrado de Engenharia Civil. Universidade Tecnológica Federal do Paraná.
- Lourenço, F. L., Mateus, M. A. (2013). *Riscos naturais, antrópicos e mistos: Homenagem ao Professor Doutor Fernando Rebelo*. Departamento de Geografia, Faculdade de letras, Universidade de Coimbra, pág. 441-461, Coimbra.

- Luino, F., Tosatti, G., & Bonaria, V. (2014). *Dam failures in the 20th century: nearly 1000 avoidable victims in Italy alone*. Journal of Environmental Science and Engineering, Volume 3, No.:1, 20 de janeiro de 2014, pág. 19-31, Nova Iorque.
- Pereira, G. J., Magalhães, F., Monteiro, G., Palma, J., Pereira, S., & Silva Matos, D. (2018). *Seismic Monitoring System of Baixo Sabor Scheme for Structural Dynamic Behaviour Monitoring and Risk Management*. ICOLD 26th Congress - 86th Annual Meeting, Viena, 1 a 7 de julho de 2018.
- Pereira S., (2019). *Structural Condition Assessment of Dams Based on Continuous Dynamic Monitoring*. Dissertação de Doutoramento em Engenharia Civil, Universidade do Porto.
- Prata, J. T., Bretas, E. M., Lemos, J. V. (2012). *Avaliação da segurança estrutural de barragens gravidade. Análise comparativa de regulamentos internacionais*. Encontro Nacional Betão Estrutural - BE2012, Faculdade de Engenharia da Universidade do Porto, 24 a 26 de outubro de 2012, Porto.
- Portela A. & Silva A. (1996). *Mecânica dos Materiais*. Plátano-Edições Técnicas, Lisboa.
- Marcelino, J., Ramos, J. M., & Santos, L. O. (2008). *Observação do comportamento estrutural de barragens e de pontes. Critérios e métodos*. 5º Congresso Luso-Moçambicano de Engenharia, 2º Congresso de Engenharia de Moçambique, Maputo, 2 a 4 de setembro de 2008.
- Martins, L., Mata, J., & Ribeiro, A. (2011). *A Qualidade das Medições de Deslocamento Radial e Tangencial pelo Método do Fio-de-Prumo no Controlo da Segurança Estrutural de Barragens de Betão. Medições E Ensaios*. CONFMET, 4 e 5 de novembro de 2010, Lisboa.
- Martins, L. L., Ribeiro, A. S., Mata, J., & de Castro, A. T. (2017). *Medição do Movimento de abertura-fecho e deslizamento de juntas em barragens de betão*. SPMed, junho de 2017, pág. 24-36.
- Mata, J. (2007). *Aplicação de redes neuronais ao controlo de segurança de barragens de betão*. Dissertação de Mestrado em Engenharia Civil, Instituto Superior Técnico Lisboa, Lisboa.
- Mata, J. (2013). *Structural safety control of concrete dams aided by automated monitoring systems*. Dissertação de Doutoramento em Engenharia Civil, Instituto Superior Técnico Lisboa, Lisboa.
- Mata, J., Tavares de Castro, A., & Sá da Costa, J. (2014). Constructing statistical models for arch dam deformation. *Structural Control and Health Monitoring*, Volume 21, Edição 3, pág. 423-437.
- Monteiro, G. S., Neves, J. P., & Gomes, J. P. (2020). *Seismic monitoring system of Baixo Sabor and feiticeiro dams*. Civil Engineering Journal, Volume 6, No.:11, pág 2072-2085.
- Montgomery, D. C., & Runger, G. C. (2010). *Applied statistics and probability for engineers*. John Wiley & sons.
- Morais, R., S. (2017) *Monitorização de deslocamentos em grandes barragens utilizando GNSS: aplicação à barragem do Cabril*. de Mestrado em Engenharia Civil, Instituto Superior de Engenharia de Lisboa.
- Moura, G. M. G. (2005). *Determinação de tensões a partir de extensões observadas em barragens de betão*. Dissertação de Mestrado de Engenharia Civil, Universidade do Porto.
- R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna.
- Rdocumentation.org, [Home - RDocumentation](https://www.rdocumentation.org/), consultado em 20 de maio de 2023
- RSB. (2018). *Regulamento de Segurança de Barragens*. Decreto-Lei n.º 21/2018 de 28 de março, Diário da República, Lisboa.



- Rijo, M. (2007). *ESTRUTURAS HIDRÁULICAS. Barragens*. Documento de apoio às Estruturas Hidráulicas dos Cursos de Licenciatura de Engenharia Civil, de Engenharia dos Recursos Hídricos da Universidade de Évora e das disciplinas de opção de Estruturas Hidráulicas dos segundos ciclos de Engenharia Agrónoma, Engenharia Civil e Engenharia dos Recursos, também da Universidade de Évora.
- Russell, S., and Norvig, P. (2002). *Artificial Intelligence: A Modern Approach*. Second. Prentice Hall Upper Saddle River, Nova Jersey.
- Schnitter, N.J. (1994). *A history of dams. The useful pyramids*. A.A. Balkema, Rotterdam.
- Schubert, E., Sander, J., Ester, M., Kriegel, H. P., & Xu, X. (2017). *DBSCAN revisited, revisited: why and how you should (still) use DBSCAN*. ACM Transactions on Database Systems, Volume 42, Edição 3, Artigo No.: 19, págs.1–21, Nova Iorque.
- Siguel A., Faggion P., Veiga L., Nadal, C., Mattos, M., Soares, M. (2013). *Aplicação do método de irradiação tridimensional no monitoramento de barragens*. Boletim de Ciências Geodésicas 19, setembro de 2013, págs. 391-406, Brasil.
- Silva, J. A. (2022a). *Desenvolvimento de Modelos (de Machine Learning) Baseados em Dados de Monitorização Contínua de Barragens de Betão para Interpretação do Comportamento Estrutural Observado*. Dissertação de Mestrado em Engenharia Civil, Universidade do Porto.
- Silva, J. A, Mata, J., Pereira, S., & Cunha, A. (2022b). *Desenvolvimento de modelos de interpretação quantitativa do tipo HTT com recurso a medições da temperatura do ar. Aplicação aos deslocamentos horizontais observados na Barragem do Baixo Sabor*. JPÉE2022-6, 6ª Jornadas Portuguesas de Engenharia de Estruturas, 2022.
- Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). *Integration k-means clustering method and elbow method for identification of the best customer profile cluster*. IOP Conference Series: Materials Science and Engineering, 2nd International Conference on Vocational Education and Electrical Engineering (ICVEE), Volume 336, 9 de novembro de 2017, Surabaya.
- Tanchev, L. (2014). *Dams and Appurtenant Hydraulic Structures (2nd edition)*. CRC Press/Balkema, Leiden.
- Xu, H., Zhang, L., Li, P., & Zhu, F. (2022). *Outlier detection algorithm based on k-nearest neighbors-local outlier factor*. Journal of Algorithms & Computational Technology, Volume 16, 1 de janeiro de 2022, págs 1-12, Sage Publications, Reino Unido.
- Zipper, M., Zulauf, N. & Bendels, M.H.K. (2020). *The Banqiao Dam Catastrophe*. Zentralblatt für Arbeitsmedizin, Arbeitsschutz und Ergonomie, Volume 70, setembro de 2019, págs. 230–235.