Instituto Superior de Engenharia de Lisboa

Licenciatura em Matemática Aplicada à Tecnologia e à Empresa

Relatório de Estágio na empresa Águas do Tejo Atlântico

Eficiência Energética da Estação Elevatória 3 Através de Métodos Estatísticos

David Alexandre Victoriano Ferreira Lima

Orientado por:

Professora Doutora Ana Martins (ISEL)

Professora Doutora Iola Pinto (ISEL)

Engenheira Rita Alves (AdTA)

Engenheiro Nuno Pimentel (AdTA)

Engenheiro David Figueiredo (AdTA)

Lisboa

2020

Resumo

Este estudo teve como tema principal a otimização de estações elevatórias (EE) com base em indicadores de eficiência, através de métodos estatísticos. O objetivo deste trabalho foi a avaliação do funcionamento de uma estação elevatória, através da avaliação das diferentes combinações de grupos eletrobomba, comparando a sua eficiência energética.

Foi escolhido como caso de estudo a estação elevatória 3 localizada no subsistema de Alcântara. Trata-se da maior estação elevatória de águas residuais da Tejo Atlântico e é constituída por 5 grupos eletrobomba, havendo registos em tempo real para a o caudal elevado, o nível do poço e a potência fornecida pelos grupos eletrobomba, num período de tempo de Abril de 2017 a Setembro de 2018.

A metodologia utilizada consistiu numa análise descritiva e exploratória dos dados disponíveis. Mostrou-se essencial um pré-processamento dos dados originais dada a distinta granularidade dos mesmos e a existência de muitos valores omissos. Com o objetivo de comparar os diferentes grupos eletrobomba foi utilizado o teste não paramétrico de Kruskal-Wallis, bem como os métodos de comparação múltipla com a correção de Bonferroni. Adicionalmente estimaram-se funções de densidade para as variáveis de interesse através de métodos de Kernel.

Com base na análise verificou-se que havia grupos eletrobomba com indicadores de eficiência melhores que outros e inferiu-se que nos instantes em que os grupos estão a funcionar individualmente, o mais eficiente é utilizar com o grupo eletrobomba 5 . Quando estão a trabalhar dois grupos ao mesmo tempo, o estudo permitiu concluir as alternativas mais eficientes são combinar os grupos 4 e 2, ou, os grupos 4 e 5. Contando com 3 bombas em funcionamento as duas opções mais eficientes são combinar os grupos de bombas 1, 2 e 4, ou, 2, 4 e 5.

Palavras Chave

Estações elevatórias; Bombas; Caudal; Indicador de Eficiência; Métodos estatísticos, Kruskall-Wallis, Estimativa de densidade Kernel.

Índice

Lista de Figuras **4**

Lista de Tabelas **6**

Glossário **8**

1 Enquadramento **9**
1.1 Introdução . . . 9
1.2 Águas do Tejo Atlântico . . . 9
1.3 Sistema de drenagem de águas residuais . . . 11
1.3.1 Sistemas de drenagem . . . 11
1.3.2 Estações Elevatórias . . . 11
1.3.3 Fábricas de Água . . . 12
1.4 Objetivos do estágio . . . 13

2 Metodologia **14**
2.1 Tratamento de dados . . . 14
2.1.1 Sincronização da série temporal . . . 14
2.1.2 Outliers . . . 15
2.2 Teste de Kruskal-Wallis . . . 16
2.3 Testes de comparação múltipla . . . 17
2.4 Estimativa de densidade Kernel . . . 18

3 Caso de Estudo **19**
3.1 Subsistema de Alcântara . . . 19
3.2 Estação Elevatória 3 de Alcântara . . . 20

4 Análise e Discussão de Resultados **22**
4.1 Dados base . . . 22
4.2 Análise preliminar dos dados . . . 23
4.3 Análise descritiva dos dados . . . 24
4.3.1 Variável *Caudal* . . . 24
4.3.2 Variável *Nível* . . . 26
4.3.3 Variável *Potência* . . . 27
4.3.4 Variável *Indicador de eficiência* . . . 28
4.3.5 Variável *Código* . . . 29
4.4 Inferência Estatística . . . 46
4.4.1 Um grupo eletrobomba em funcionamento . . . 46
4.4.2 Dois grupos eletrobomba em funcionamento . . . 52
4.4.3 Três grupos eletrobomba em funcionamento . . . 57

Conclusão **62**

Referências **63**

Lista de Figuras

1.1 Empresas do Grupo AdP [2] . . . 10
1.2 Municípios abrangidos pela AdTA [3] . . . 10
1.3 Exemplo de uma estação elevatória [5] . . . 12
1.4 Fábrica de Água [2] . . . 13

3.1 Fábrica de Água de Alcântara . . . 19
3.2 Curvas características da instalação e curvas dos grupos eletrobomba FLYGT + 1 grupo ABS em paralelo [10] . . . 20

4.1 Base de dados original . . . 22
4.2 Base de dados sincronizada . . . 23
4.3 Boxplot e Histograma da variável caudal. . . . 25
4.4 Boxplot e Histograma da variável nível. . . . 26
4.5 Boxplot e Histograma da variável potência. . . . 28
4.6 Boxplot da variável indicador de eficiência. . . . 29
4.7 Boxplot da variável caudal para um grupo funcionamento. . . . 31
4.8 Boxplot da variável potência para um grupo em funcionamento. . . . 32
4.9 Gráfico de dispersão caudal vs potência para 1 grupo em funcionamento. . 33
4.10 Gráfico de dispersão caudal vs indicador para um grupo em funcionamento 34
4.11 Boxplot da variável caudal para dois grupos em funcionamento. . . . 36
4.12 Boxplot da variável potência para dois grupos em funcionamento . . . 37
4.13 Gráfico de dispersão caudal vs potência para dois grupos em funcionamento. . . . 38
4.14 Gráfico de dispersão caudal vs indicador de eficiência para dois grupos em funcionamento. . . . 39
4.15 Boxplot da variável caudal para três bombas em funcionamento. . . . 42
4.16 Boxplot da variável potência para três bombas em funcionamento. . . . 43
4.17 Boxplot da variável indicador de eficiência para três bombas em funcionamento. . . . 44
4.18 Gráfico de dispersão caudal vs potência para três grupos em funcionamento. 44
4.19 Gráfico de dispersão caudal vs indicador para três grupos em funcionamento . . . 45
4.20 Gráficos de densidade estimada da variável caudal, para um grupo em funcionamento. . . . 49
4.21 Gráficos de densidade estimada da variável potência, para 1 grupo em funcionamento. . . . 50
4.22 Gráfico de densidade estimada da variável indicador de eficiência, para 1 grupo em funcionamento. . . . 51
4.23 Gráficos de densidade estimada da variável caudal, para dois grupos em funcionamento. . . . 54

4.24 Gráficos de densidade estimada da variável potência, para dois grupos em funcionamento. 55
4.25 Gráficos de densidade estimada da variável indicador de eficiência, para dois grupos em funcionamento. 56
4.26 Gráficos de densidade estimada da variável caudal, para três grupos em funcionamento. 59
4.27 Gráficos de densidade estimada da variável potência, para três grupos em funcionamento. 60
4.28 Gráficos de densidade estimada da variável indicador de eficiência, para três grupos em funcionamento. 61

Lista de Tabelas

3.1 Poço de bombagem (Níveis de arranque). 21

4.1 Frequências de funcionamento das combinações dos grupos eletrobomba . . 30
4.2 Combinações de funcionamento dos grupos eletrobomba 30
4.3 Indicadores descritivos da variável caudal para um grupo em funcionamento 31
4.4 Indicadores descritivos da variável potência para um grupo em funcionamento 32
4.5 Indicadores descritivos da variável indicador de eficiência para um grupo em funcionamento. 33
4.6 Correlações entre o caudal e a potência para um grupo em funcionamento. 34
4.7 Correlações entre o caudal e o indicador de eficiência para um grupo em funcionamento. 35
4.8 Indicadores descritivos da a variável caudal para dois grupos em funcionamento. 35
4.9 Indicadores descritivos da a variável potência para dois grupos em funcionamento. 37
4.10 Indicadores descritivos da a variável indicador de eficiência para dois grupos em funcionamento. 38
4.11 Correlações entre o caudal e a potência para dois grupos em funcionamento 39
4.12 Correlações entre o caudal e o indicador de eficiência para dois grupos em funcionamento. 40
4.13 Indicadores descritivos da variável caudal para três bombas em funcionamento 41
4.14 Indicadores descritivos da variável potência para três bombas em funcionamento. 42
4.15 Indicadores descritivos da variável indicador de eficiência para 3 bombas em funcionamento 43
4.16 Correlações entre o caudal e a potência para três grupos em funcionamento 45
4.17 Correlações entre o caudal e o indicador para três grupos em funcionamento 46
4.18 Testes de comparação múltipla para a variável caudal com um grupo em funcionamento. 47
4.19 Testes de comparação múltipla para a variável potência com um grupo em funcionamento. 48
4.20 Testes de comparação múltipla para a variável indicador de eficiência com um grupo em funcionamento. 48
4.21 Testes de comparação múltipla para a variável caudal com dois grupos em funcionamento. 53
4.22 Testes de comparação múltipla para a variável potência com dois grupos em funcionamento. 53
4.23 Testes de comparação múltipla para a variável indicador de eficiência com dois grupos em funcionamento. 53

4.24 Testes de comparação múltipla com três grupos em funcionamento, para a variável caudal. 58
4.25 Testes de comparação múltipla com três grupos em funcionamento, para a variável potência. 58
4.26 Testes de comparação múltipla com três grupos em funcionamento, para a variável indicador de eficiência. 58

Glossário

AdTA - Águas do Tejo Atlântico
EE - Estação elevatória
ETAR - Estação de tratamento de águas residuais
FA - Fábrica de água
SCADA - Sistemas de supervisão e aquisição de dados
EDK - Estimativa de densidade por Kernel
CVR - Coeficiente de variação resistente

Capítulo 1

Enquadramento

1.1 Introdução

O presente relatório descreve o trabalho desenvolvido durante Estágio Curricular na empresa Águas do Tejo Atlântico (AdTA), com o objetivo de obter o grau de licenciado em Matemática Aplicada à Tecnologia e à Empresa, pelo Instituto Superior de Engenharia de Lisboa. O trabalho foi desenvolvido na fábrica de água (FA) de Alcântara e integrado na Área de Investigação, Desenvolvimento e Inovação da empresa.

O estágio realizado teve como tema principal a otimização do modo de funcionamento de estações elevatórias (EE) com base em indicadores de eficiência, através de métodos estatísticos.

Neste relatório será inicialmente apresentada a empresa Tejo Atlântico, a sua constituição e o sistema de drenagem de águas residuais. De seguida será apresentado o objetivo do estágio e o caso de estudo, a Estação Elevatória 3 do subsistema de Alcântara.

Posteriormente é apresentada a metodologia do trabalho que inclui os métodos estatísticos aplicados, todas as variáveis utilizadas e por fim os resultados provenientes da análise e inferência estatística. Nesta análise foram usados dados relativamente ao funcionamento de uma estação elevatória. Por último são apresentadas as conclusões e perspetivas futuras do trabalho realizado.

1.2 Águas do Tejo Atlântico

A AdTA é uma empresa pertencente ao Grupo Águas de Portugal, com sede na Fábrica de Água de Alcântara. A Tejo Atlântico é responsável pela gestão e exploração do sistema multimunicipal de saneamento de águas residuais da Grande Lisboa e Oeste por um prazo de 30 anos. [1]

A Figura 1.1 representa todas as empresas do Grupo Águas de Portugal.

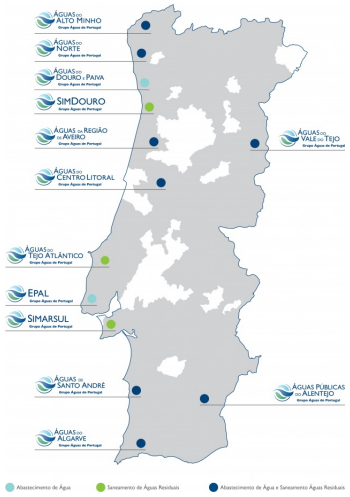

Figura 1.1: Empresas do Grupo AdP [2]

A AdTA tem por missão explorar e gerir o sistema multimunicipal de saneamento de águas residuais da Grande Lisboa e Oeste, garantindo a qualidade, a continuidade e a eficiência dos serviços públicos de águas, no sentido da proteção da saúde pública, do bem-estar das populações, da acessibilidade aos serviços públicos, da proteção do ambiente e da sustentabilidade económica e financeira do setor, num quadro de equidade e estabilidade tarifária, contribuindo ainda para o desenvolvimento regional e o ordenamento do território. [3]

Figura 1.2: Municípios abrangidos pela AdTA [3]

O sistema multimunicipal constituído por vários subsistemas de Saneamento de águas residuais compreende os respetivos intercetores, emissários, estações elevatórias e FA onde afluem as águas residuais para tratamento adequado.

Deste Sistema fazem parte:

- 103 FA
- 268 EE
- 5 emissários submarinos
- 1093.4 km de intercetores e emissários
- 2.3 milhões de habitantes servidos
- 4.145 km^2 de área coberta de 23 municípios [4]

1.3 Sistema de drenagem de águas residuais

1.3.1 Sistemas de drenagem

A água residual, resultante da utilização da água pelas populações e actividades produtivas, segue para a FA, através de rede em baixa e rede em alta. A rede em baixa contém emissários, coletores e EE da responsabilidade do município. A rede em alta contém EE, FA, emissários/intercetores da responsabilidade da empresa responsável pela tratamento das águas residuais, neste caso, da AdTA. [5]

A rede de coletores integra os sistemas públicos de drenagem de água residual e tem por finalidade assegurar a sua condução para o destino final adequado. Os emissários/ intercetores possibilitam a recolha dos efluentes que derivam das várias redes de coletores e a sua condução até às FA para tratamento adequado e destino final. [5]

Dentro da rede de drenagem que conduz a água residual para a FA, podem existir vários tipos de sistemas de drenagem da água residual, dos quais o sistema:

- **Unitário** : constituído por uma única rede de coletores, onde são admitidas conjuntamente as águas residuais domésticas, comerciais e industriais e águas pluviais; recolhem e drenam a totalidade das águas;

- **Separativo** : constituído por duas redes de coletores distintas, uma destinada à drenagem das águas residuais domésticas, comerciais e industriais, e uma outra à drenagem das águas pluviais ou similares.

1.3.2 Estações Elevatórias

As EE consistem em infraestruturas com equipamentos preparados para a elevação de águas pluviais e/ou residuais (figura 1.3). Quando não é possível efetuar a drenagem das águas somente por gravidade até à FA, através da utilização de bombas submersíveis, as EE permitem que a água residual seja elevada para um nível superior ao local de recolha.

Adicionalmente à elevação da água, as EE exercem também um papel fundamental na remoção de sólidos de maiores dimensões através de sistemas de gradagem, que por sua vez, facilitam o processo de pré-tratamento antes das águas residuais chegarem à FA para tratamento adequado. [5]

Uma EE pode ser composta por:

- Poço
- Grupos eletrobomba
- Dispositivos de descida e elevação
- Válvulas
- Quadro elétrico de comando das bombas
- Interruptores de nível
- Tubagens de interligação
- Sistemas de gradagem [6]

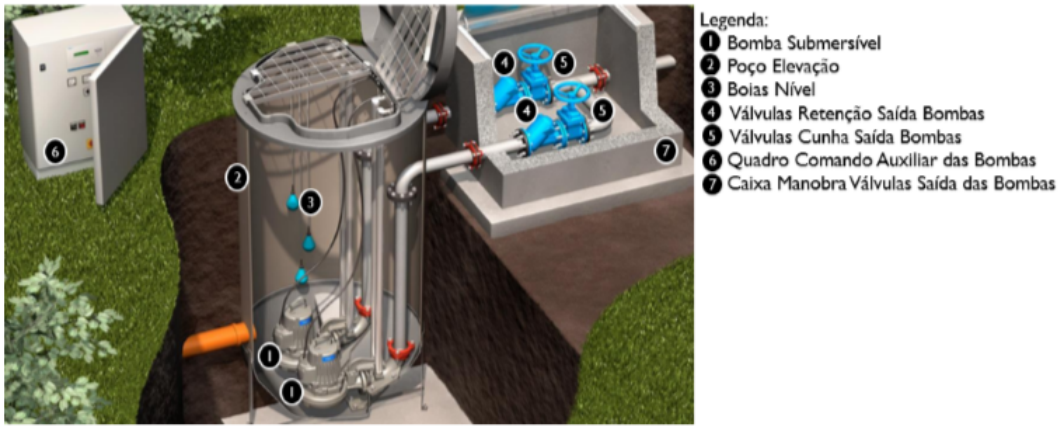

Figura 1.3: Exemplo de uma estação elevatória [5]

1.3.3 Fábricas de Água

A AdTA é pioneira na incorporação dos desafios da economia circular na gestão do ciclo urbano da água, fazendo evoluir o tratamento das águas residuais para um novo paradigma onde a valorização máxima dos recursos é a prioridade.[3]

É neste contexto que surgem as FA, um novo conceito que reforça o carácter industrial do trabalho realizado numa Estação de Tratamento de Águas Residuais (ETAR), onde a água é tratada como uma matéria-prima plena de recursos para usar, reutilizar, reciclar e valorizar.[3]

A água residual, resultante da utilização da água pelas populações e atividades produtivas, é recolhida e encaminhada para a FA (figura 1.4). Depois de tratada na FA, parte desta água é reutilizada para regas e lavagens e a restante é devolvida à natureza em condições ambientalmente seguras, permitindo assegurar a reposição de água nos meios hídricos sem comprometer a saúde pública e os ecossistemas, protegendo a natureza e a biodiversidade.[2]

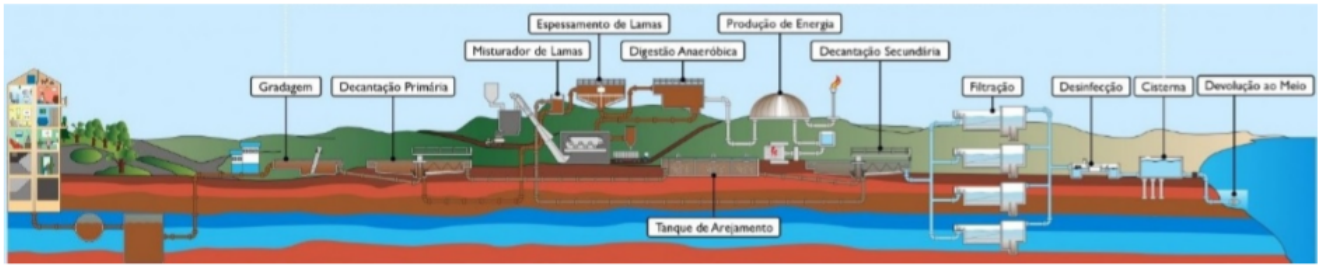

Figura 1.4: Fábrica de Água [2]

1.4 Objetivos do estágio

Como referido, a AdTA é responsável pela gestão de inúmeras EE. Estas infraestruturas representam um elevado consumo de energia, principalmente associado à elevação das águas residuais. Para além disso, nas EE poderão existir diferentes grupos eletrobomba, não só em termos de marcas, como também nos tempos de funcionamento e outras características dissemelhantes, logo a eficiência dos grupos eletrobomba poderá ser diferente.

Nesse sentido existe a necessidade de otimizar o funcionamento de uma EE, de modo a encontrar uma solução de funcionamento que apresente menores consumos energéticos.

Tendo em conta esta necessidade, este estágio teve como principal objetivo a avaliação do funcionamento de uma EE, através da avaliação das diferentes combinações de grupos eletrobomba, comparando a sua eficiência, através de indicadores de eficiência.

Capítulo 2

Metodologia

Neste capítulo é apresentada a metodologia proposta para a concretização dos objetivos deste trabalho.

Primeiramente, é importante fazer um tratamento inicial dos dados, sincronizando a série temporal, identificando e eliminando valores anómalos. De seguida, após se fazer uma análise exploratória, com o intuito de comparar o comportamento das variáveis de interesse dos grupos eletrobomba distintos, e dada a presença de desvios à normalidade nas variáveis, utilizou-se o teste não paramétrico de Kruskal- Wallis.

Para os casos em que o teste de Kruskal-Wallis deu resultados significativos, procedeu-se à realização dos testes de comparação múltipla de Dunn.

Por fim, avalia-se quais as combinações de grupos eletrobomba mais eficientes analisando as estimativas de densidade Kernel de cada um dos grupos e os respetivos indicadores descritivos.

Durante o estágio foi utilizado o programa EXCEL para construir a base de dados necessária e o software R para efetuar todos os cálculos estatísticos essenciais.

2.1 Tratamento de dados

Como referido, antes de ser feita qualquer análise é necessário que se faça um tratamento inicial dos dados para que se possa construir a base de dados.

2.1.1 Sincronização da série temporal

De modo a proceder-se como um todo, é preciso preparar a base de dados sincronizando todas as variáveis com o mesmo espaço temporal. Para efetuar esta sincronização será necessário efetuar uma interpolação linear dos dados.

Esta interpolação consiste em aproximar um conjunto de dados entre 2 pontos (x_0, y_0) e (x_1, y_1) , através de uma função linear do tipo **ax + b**, podendo ser deduzida por:

$$y = y_0 + (y_1 - y_0)\frac{x - x_0}{x_1 - x_0} \tag{2.1}$$

Por exemplo, numa base de dados com uma variável com espaços temporais de 10 minutos e outra variável com 5, fazendo uma interpolação linear na variável com observações de 10 em 10 minutos, é possível obter uma aproximação das observações, de 5 em 5 minutos, sincronizando assim as variáveis com o mesmo espaço temporal.

2.1.2 Outliers

Seguidamente é fundamental identificar e eliminar valores anómalos que possam influenciar uma futura análise. O gráfico "caixa de bigodes"permite evidenciar a existência de outliers, caso estes existam no conjunto de dados.

Para construir o gráfico "caixa de bigodes"é necessário calcular [7]:

- Mínimo - $x_{(1)}$
- Máximo - $x_{(n)}$
- Quantil de ordem 0.25 - $Q_{0.25}$
- Quantil de ordem 0.5 (mediana) - $Q_{0.50}$
- Quantil de ordem 0.75 - $Q_{0.75}$
- Amplitude inter-quartis - IQR $= Q_{0.75} - Q_{0.25}$
- Barreira interna inferior - $b_I = Q_{0.25} - 1.5IQR$
- Barreira interna superior - $b_S = Q_{0.75} + 1.5IQR$
- Valores adjacentes (limites dos "bigodes")
  - Inferior: $a_I =$ menor x_i tal que $x_i \geq b_I$
  - Superior: $a_S =$ maior x_i tal que $x_i \leq b_I$
- No caso de existirem outliers devem então calcular-se as barreiras externas:
- Barreira externa inferior - $B_l = Q_{0.25} - 3IQR$
- Barreira externa superior - $B_S = Q_{0.75} + 3IQR$

A observação x_i diz ser um outlier se $x_i < b_I$ ou $x_i > b_S$. Caso hajam outliers, então estes podem ser moderados ou severos.

- A observação x_i é um outlier moderado se $B_I \leq x_i < b_I$ ou $b_S < x_i \leq B_S$
- A observação x_i é um outlier severo $x_i < B_I$ ou $x_i > B_S$

2.2 Teste de Kruskal-Wallis

O teste de *Kruskal-Wallis* é um teste adequado para **comparar funções de distribuição de uma variável medida em k amostras independentes**.

O teste pode ser aplicado quando:[8]

- as k amostras **não** são provenientes de populações com distribuição normal (ou pelo menos uma delas não é);
- variâncias das k populações não são idênticas;
- as k amostras independentes e os dados são de natureza quantitativa;

Este procedimento testa a **hipótese** de que **as k amostras são provenientes de populações com a mesma distribuição** contra a hipótese alternativa de que pelo menos duas das k amostras são provenientes de populações com distribuições diferentes entre si.[8]

No entanto, este teste é particularmente sensível a diferenças de localização, pelo que as hipóteses são usualmente escritas em função das medianas populacionais.

1. **Hipóteses:**

$$H_0 : Me_{X_1} = Me_{X_2} = ... = Me_{X_k}$$

vs

H_1: Pelo menos duas das k medianas populacionais diferem entre si.

ou, no caso de se admitir a simetria das k populações:

$$H_0 : \mu_{X_1} = \mu_{X_2} = ... = \mu_{X_k}$$

vs

H_1: Pelo menos dois dos k valores médios populacionais diferem entre si.

A **estatística de teste** de *Kruskal-Wallis* assenta na comparação das ordens que as observações das diferentes amostras em teste recebem quando consideradas juntamente com as observações das restantes amostras.

É calculada uma medida que evidencia a discrepância existente entre as ordens que os valores de cada amostra recebem na amostra conjunta e aquelas que se esperaria que recebessem caso as amostras fossem homogéneas, ou seja, provenientes da mesma população.

Um valor reduzido da estatística de teste revela que as ordens das observações das diversas amostras estão próximas do que delas se espera, sob a validade da hipótese nula.

Um valor elevado para a estatística de teste revela precisamente o contrário, levando à rejeição da hipótese nula.

2. **Nível de significância**: α

3. **Estatística de teste**:

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(n+1) \frown \chi^2_{k-1} \tag{2.2}$$

- $R_j = \sum_{j=1}^{n_j} R_{ij}$ soma dos números de ordem (ranks) da j-ésima amostra, j= 1, 2, ..., k;
- n_j é a dimensão da amostra j, com j= 1, 2, ..., k;
- **n** é a dimensão da amostra conjunta, ou seja, $\sum_{j=1}^{k} n_j$.

No caso de existirem empates, a estatística de teste vem:

$$H_E = H/(1 - \frac{\sum_{j=1}^{g}(u_i^3 - u_i)}{n^{3-n}}) \frown \chi^2_{k-1} \tag{2.3}$$

- **g** é o número de grupos de empates;
- u_i representa o número de empates no i-ésimo grupo de observações iguais

4. **Decisão**:

Rejeitar H_0 se:

- $H_{obs}\,(ou\, H_{E_{obs}}) \geqslant \chi^2_{k-1,1-\alpha}$ (quantil de probabilidade 1 - α da distribuição do qui-quadrado com k-1 graus de liberdade);
- ou de modo equivalente quando o valor de prova (valor-p) é inferior ou igual ao nível de significância α.

2.3 Testes de comparação múltipla

Quando a hipótese nula no teste de Kruskal-Wallis é rejeitada, ao nível de significância α, significa que pelo menos um dos k grupos é diferente de pelo menos um dos outros. No entanto, não indica quais é que são diferentes entre si.

Os **testes de comparação múltipla** entre as médias dos ranks permitem responder à questão. Os testes que se apresentam de seguida, são essencialmente uma aplicação do método de Bonferroni-Dunn ao modelo de teste de Kruskal-Wallis. [8]

- **Hipóteses:**

$$H_0 : Me_{X_i} = Me_{X_j} ; i \neq j$$

vs

$$H_1 : Me_{X_i} \neq Me_{X_j}, i, j = 1, 2, ..., k;\ i \neq j$$

- **Decisão:**

  Rejeita-se H_0, ao nível de significância α, se :

$$\left|\bar{R}_i - \bar{R}_j\right| \geqslant z_{1-\frac{\alpha}{k(k-1)}} \sqrt{\frac{n(n+1)}{12}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \tag{2.4}$$

  - $\bar{R}_i$ e $\bar{R}_j$ são as médias dos ranks das amostras i e j envolvidas na comparação;
  - n_i e n_j são as dimensões das amostras i e j; $n = \sum_{j=1}^{k} n_j$
  - $z_1 - \frac{\alpha}{k(k-1)}$ é o quantil de probabilidade $1 - \frac{\alpha}{k(k-1)}$ da distribuição normal padrão.

2.4 Estimativa de densidade Kernel

A estimativa de densidade por Kernel é uma forma não-paramétrica para estimar a função de densidade de probabilidade de uma variável aleatória. A EDK é uma metodologia fundamental quando se tratar o tema do ajustamento através de funções suavizadoras. O conhecimento destas funções suavizadoras permitem realizar inferências sobre a população, com base em uma amostra finita de dados. [9]

Seja $(x_1, x_2, ..., x_n)$ uma amostra de uma variável aleatória independente e identicamente distribuída com uma densidade f desconhecida. O objetivo é estimar a forma dessa função f. A estimativa de densidade por kernel pode ser feita através da seguinte fórmula [9]:

$$\hat{f}_b(x) = \frac{1}{n}\sum_{i=1}^{n} K_b(x - x_i) = \frac{1}{nb}\sum_{n}^{i=1} K(\frac{x - x_i}{b}) \tag{2.5}$$

em que K é um Kernel fixo e $b > 0$ é um parâmetro de distância chamado de largura de banda, que controla a influência dos vizinhos na estimativa final com base no método de Kernel. A largura de banda por *default* no software R é dada por :

$$\hat{b} = 0.9min(\hat{\sigma}, R/1.34)n^{-1/5} \tag{2.6}$$

onde a variável R representa a amplitude interquartil (IQR).

O comando do software R *geom_density*() calcula e desenha a estimativa de densidade Kernel (EDK) , que é uma versão suavizada do histograma.

Capítulo 3

Caso de Estudo

Neste estágio, de modo a avaliar a eficiência energética de uma estação elevatória, foi escolhido como caso de estudo a EE3 localizada no subsistema de Alcântara, que é composto pela FA de Alcântara e 11 estações elevatórias.

3.1 Subsistema de Alcântara

O subsistema de Alcântara, que serve os municípios de Amadora, Lisboa e Oeiras, transporta as águas residuais da zona alta de Lisboa, por gravidade através do caneiro de Alcântara, até à FA de Alcântara. As águas residuais da zona ribeirinha da cidade de Lisboa são transportadas por um sistema intercetor de 23 km e 11 EE, até à EE3 de Alcântara que posteriormente transporta as águas residuais para a FA de Alcântara.

Esta FA (figura 3.1), que foi dimensionada para tratar o equivalente populacional a 756000 habitantes, tem uma capacidade de tratamento de 3.3 m^3/s em tempo seco e pode aumentar para 6.6 m^3/s em tempo húmido. Posteriormente, o efluente tratado é descarregado no estuário do Tejo.

Figura 3.1: Fábrica de Água de Alcântara

3.2 Estação Elevatória 3 de Alcântara

A EE3 de Alcântara é a maior estação elevatória de águas residuais do país e, como referido anteriormente, reúne as águas residuais provenientes da zona ribeirinha da cidade de Lisboa para posterior transporte para a FA de Alcântara. Esta EE está equipada com 5 grupos eletrobomba do tipo submersível para a instalação em poço seco (4 grupos eletrobomba FLYGT e 1 grupo ABS). [10]

Os grupos eletrobomba funcionam, normalmente, em modo automático controlado através de um autómato. Neste modo de funcionamento da instalação, será a variação contínua da velocidade de rotação dos grupos em serviço, em função do nível no poço e do caudal, que se quer elevar. [10]

Na figura 3.2 encontram-se as curvas características da instalação que retratam o comportamento das várias combinações de bombas relacionando grandezas que caracterizam o seu funcionamento tais como o caudal (litros por segundo), a velocidade nominal das bombas (hertz) e o nível do poço (mm).

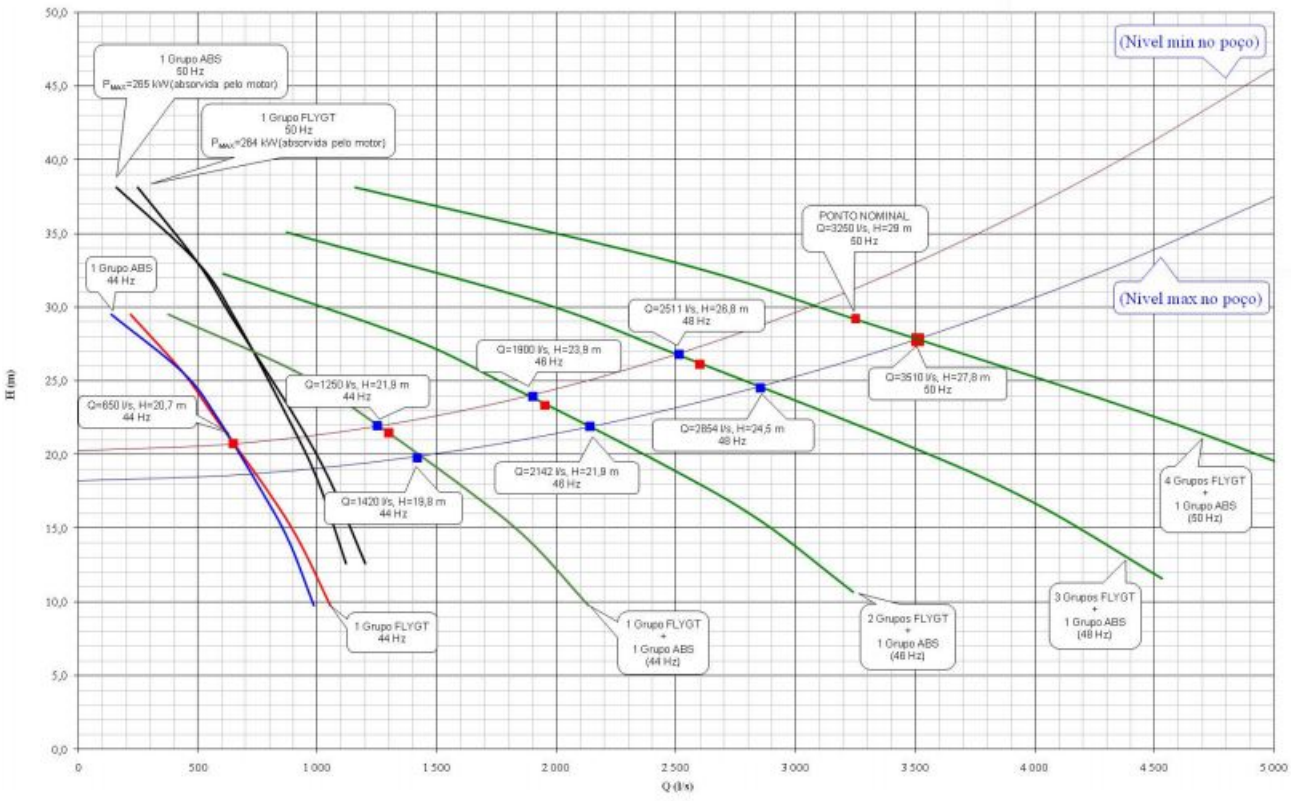

Figura 3.2: Curvas características da instalação e curvas dos grupos eletrobomba FLYGT + 1 grupo ABS em paralelo [10]

Foram definidos os volumes de água do poço para as várias alturas conforme a tabela 3.1:

Tabela 3.1: Poço de bombagem (Níveis de arranque).

Nível	**h (mm)**	**V (m^3)**
LSAHH.01	**5150**	**341.87**
LSH.05	4950	324.89
LSH.04	4650	299.42
LSH.03	4350	273.95
LSH.02	4050	248.48
LSH.01	3750	223.01
LSL.05	3450	197.54
LSL.04	3150	173.30
LSL.03	2850	149.87
LSL.02	2550	127.58
LSL.01	2250	106.54
LSALL.01	**2000**	**90.26**

Em que:

- LSAHH.01 - Nível alto de alarme - Inundação;
- LSH.05 - Nível de arranque do 5^{o} grupo;
- LSH.04 - Nível de arranque do 4^{o} grupo;
- LSH.03 - Nível de arranque do 3^{o} grupo;
- LSH.02 - Nível de arranque do 2^{o} grupo;
- LSH.01 - Nível de arranque do 1^{o} grupo;
- LSL.05 - Paragem do 5^{o} grupo;
- LSL.04 - Paragem do 4^{o} grupo;
- LSL.03 - Paragem do 3^{o} grupo;
- LSL.02 - Paragem do 2^{o} grupo;
- LSL.01 - Paragem do 1^{o} grupo;
- LSLLA.01 - Nível baixo de alarme - Nível Mínimo.

Os dados relativos ao consumo de energia da EE3, mais concretamente a potência, foram fornecidos pela EDP distribuição e os dados relativos ao funcionamento das bombas da EE3, nomeadamente o caudal elevado, o nível do poço e o estado de funcionamento das bombas, foram fornecidos pelos sistemas de supervisão e aquisição de dados (SCADA).

Capítulo 4

Análise e Discussão de Resultados

No presente capítulo são apresentados os dados base utilizados para a realização deste trabalho, assim como os resultados e a sua discussão depois de aplicada a metodologia descrita no Capítulo 3.

4.1 Dados base

Neste estudo foram utilizados dados relativos ao funcionamento EE3, nomeadamente do nível da água no poço, em mm, o caudal elevado, em m^3/h, o estado de funcionamento de cada grupo eletrobomba, que foram retirados dos registos do sistema de supervisão e aquisição de dados (SCADA). Relativamente aos dados do consumo de energia da EE3, mais concretamente a potência ativa da estação elevatória, em kW, foram retirados do portal da EDP distribuição. Adicionalmente foi calculado o indicador de eficiência energética, em kWh/m^3, para cada instante.

Foram considerados registos das variáveis referidas para o período de Abril de 2017 a 24 de Setembro de 2018. As variáveis apresentavam intervalos de tempo diferentes, conforme a figura 4.1, e com inúmeras observações repetidas.

Data	Hora	Activa kW		BD SQL Alcântara		SCADA EE Alcânta
01/04/17	00:00:00	334	Data	Gradagem [Medidor EE3]	Data	EE 3 [Nível]
01/04/17	00:15:00	326	[dd/mm/aaaa hh:mm:ss]	[m3/h]	[dd/mm/aaaa hh:mm:ss]	[m]
01/04/17	00:30:00	321	01/01/2017 00:01:11	gradagem [Medidor EE3]	01/01/2017 00:00:05	3146
01/04/17	00:45:00	319	01/01/2017 00:02:54	2265,625	01/01/2017 00:00:06	3146
01/04/17	01:00:00	321	01/01/2017 00:04:37	2276,042	01/01/2017 00:05:10	3180
01/04/17	01:15:00	313	01/01/2017 00:06:21	2291,667	01/01/2017 00:10:01	3229
01/04/17	01:30:00	301	01/01/2017 00:09:07	2348,958	01/01/2017 00:15:00	3191
01/04/17	01:45:00	275	01/01/2017 00:10:51	2421,875	01/01/2017 00:20:02	3160
01/04/17	02:00:00	276	01/01/2017 00:12:34	2484,375	01/01/2017 00:25:01	3178
01/04/17	02:15:00	289	01/01/2017 00:14:19	2432,292	01/01/2017 00:30:03	3164
01/04/17	02:30:00	286	01/01/2017 00:16:01	2395,833	01/01/2017 00:35:04	3154
01/04/17	02:45:00	284	01/01/2017 00:17:45	2380,208	01/01/2017 00:40:03	3163
01/04/17	03:00:00	280	01/01/2017 00:19:29	2375	01/01/2017 00:45:06	3136

Figura 4.1: Base de dados original

Importa referir que durante o período de analise o grupo eletrobomba 3 não esteve em funcionamento, não tendo sido incluído na posterior análise de dados.

A base de dados foi construída com recurso ao programa EXCEL e os restantes cálculos estatísticos deste trabalho foram realizados no software R.

4.2 Análise preliminar dos dados

De modo a proceder-se como um todo, foi necessário preparar a base de dados sincronizando todas as variáveis com o mesmo espaço temporal, identificar e eliminar valores com instantes repetidos.

Para as variáveis *caudal e nível*, que se encontravam maioritariamente com intervalos de 5 minutos, fez-se uma interpolação linear dos dados de forma a que todas as variáveis tenham um intervalo de 2 minutos e meio entre as observações. No caso da *potência*, como o intervalo entre as observações era maior (15 minutos), optou-se por considerar que os valores foram iguais durante o intervalo de tempo entre as observações.

Após o procedimento anterior obteve-se uma base de dados cujas observações têm uma granulidade de 2 minutos e meio, tal como se observa na figura 4.2.

Data	Hora	Nível[mm]	Caudal[m3/h]	Potencia[kW]
01/04/2017	00:00:00	3460	3243	334
01/04/2017	00:02:30	3432	3162	334
01/04/2017	00:05:00	3404	3081	334
01/04/2017	00:07:30	3395	3078	334
01/04/2017	00:10:00	3386	3074	334
01/04/2017	00:12:30	3412	3079	334
01/04/2017	00:15:00	3438	3083	326
01/04/2017	00:17:30	3418	3066	326

Figura 4.2: Base de dados sincronizada

Seguidamente averiguou-se a presença de outliers que possam influenciar a análise.

Para o **nível** constatou-se que existem 452 outliers. Destes outliers eliminaram-se os que se tratavam de situações onde existiam erros nos dados, nomeadamente:

1. Valores com nível zero - trata-se de uma situação impossível de acontecer pois está sempre a entrar água por ser uma das principais estações elevatórias.

2. Valores superiores a 5830 mm - pois a 5300 mm temos o alarme de inundação, onde são fechadas as comportas, o que impossibilita a entrada de água. Considerou-se valores entre 5300 e 5830 mm, pois as comportas vão fechando gradualmente e também se teve em conta a existência de algum possível erro de medição.

Os restantes valores mantêm-se pois não se encontrou razão para a sua remoção. Em suma, dos 452 outliers, eliminaram-se 353 destes ficando apenas 99 outliers.

Para o **caudal** primeiramente encontraram-se 6336 outliers. Destes eliminaram-se:

1. Valores com caudal a zero, correspondentes a momentos em que estava em funcionamento pelo menos um grupo eletrobomba.

2. Valores com caudal a zero que aparentam ser um erro de medição devido ao tempo e horário em que estes estão registados.

3. Valores acima do caudal máximo.

   Considerando a figura 3.2 e que uma dos grupos eletrobomba do tipo FLYGT (grupo 3), nunca esteve em funcionamento neste período de tempo, iremos assumir que o caudal máximo, trabalhando em simultâneo os restantes grupos é de 2854 l/s.

   Assim sendo eliminaram-se as observações com o caudal superior a 10300 m^3/h (2854 l/s).

4. Eliminaram-se também valores observados em dias em que ocorreu chuva, de forma a evitar que dados anómalos pudessem influenciar a análise dos dados.

Os restantes valores mantêm-se pois não se encontrou razão para a sua remoção. Em suma, dos 6336 outliers eliminaram-se 6046, tendo-se mantido na base de dados a analisar apenas 290 observações consideradas como outliers.

Devido à quantidade de valores omissos causados pela remoção das observações anteriores e por falhas no sistema, decidiu-se não utilizar nenhum método de preenchimento de dados.

Por fim criou-se uma variável *código* que informa quais são os grupos eletrobomba que se encontram em funcionamento em cada instante.

No total, depois das alterações referidas anteriormente, conta-se para análise com:

- 312191 observações na série temporal;
- 5 variáveis das quais 4 quantitativas e 1 qualitativa.

4.3 Análise descritiva dos dados

Já com a base de dados completa prosseguiu-se para uma análise exploratória dos dados.

4.3.1 Variável *Caudal*

A variável *Caudal* representa a quantidade de água elevada (em m^3/h) pelos grupos eletrobomba em funcionamento da EE3 num determinado instante, sendo uma **variável quantitativa** com uma **escala absoluta contínua**. Foram encontrados os seguintes indicadores descritivos:

- Mínimo: 0 m^3/h
- Máximo: 9552 m^3/h
- Média: 2939 m^3/h
- Mediana: 3094 m^3/h
- 1º Quartil: 2245 m^3/h

- 3^{o} Quartil: 3624 m^3/h
- IQR: 1379 m^3/h
- Desvio-padrão: 978.0526
- CVR: 44,5701 %
- Curtose: -0.4950
- Assimetria: -0.3087
- Valores omissos: 94228

Com base nos resultados anteriores, observou-se que do todal das 312191 observações, 94228 são valores omissos, tendo informação apenas de 217963 observações.

Verificou-se que a mediana é de 3094 m^3/h, ou seja, 50% dos valores são menores ou iguais a esse caudal, que o 1^{o} quartil é de 2245 m^3/h, ou seja, 25% dos dados são menores ou iguais a esse valor e que o 3^{o} quartil é igual a 3624 m^3/h, ou seja, 25% dos dados são maiores ou iguais a esse valor, o que leva com que a amplitude inter-quartis(IQR) seja de 1379 m^3/h.

Os dados apresentam uma variabilidade alta, pois o coeficiente de variação resistente (CVR) é superior a 30%.

De seguida fez-se uma **análise gráfica**, apresentada na figura 4.3, onde se pode confirmar alguns valores referidos anteriormente:

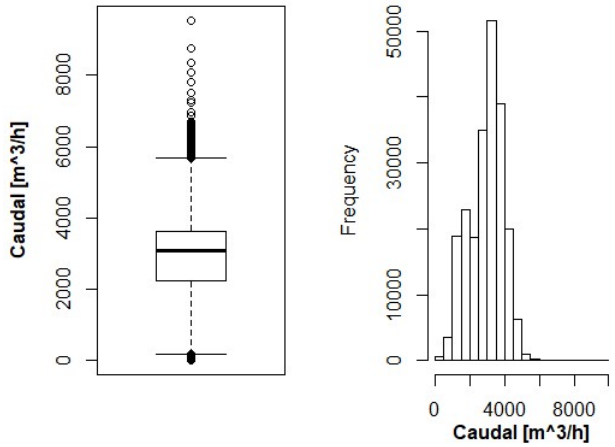

Figura 4.3: Boxplot e Histograma da variável caudal.

Pode-se ver que há uma assimetria negativa pois o coeficiente de assimetria é menor que zero. No bloxpot nota-se a presença de inúmeros outliers.

Dos 290 outliers encontrados, 285 são outliers moderados e 5 são severos. É importante salientar que os outliers abaixo da barreira inferior do boxplot, foram ocorridos em instantes em que a frequência dos grupos eletrobomba diminuía, até à sua paragem, para se poder alternar o funcionamento dos mesmos.

4.3.2 Variável *Nível*

A variável *Nível* representa o nível da água no poço (em mm) da EE3 num determinado instante, sendo uma **variável quantitativa** com uma **escala absoluta contínua**. Encontrou-se os seguintes indicadores descritivos:

- Mínimo: 1480 mm
- Máximo: 5810 mm
- Média: 3637.1333 mm
- Mediana: 3559 mm
- $1^{\underline{o}}$ Quartil: 3207 mm
- $3^{\underline{o}}$ Quartil: 4071 mm
- IQR: 864 mm
- Desvio-padrão: 576.7753
- CVR: 24.2765%
- Curtose: -0.5771
- Assimetria: 0.3480
- Valores omissos: 84032

Com base no output anterior, viu-se que das 312191 observações, 84032 são valores omissos, tendo informação apenas de 228159 observações.

Verificou-se também que a mediana é de 3559 mm, o $1^{\underline{o}}$ quartil é de 3207 mm e que o $3^{\underline{o}}$ quartil é igual a 4071 mm, tendo uma amplitude inter-quartis seja de 864 mm.

Os dados apresentam uma variabilidade média, pois o coeficiente de variação resistente está entre os 15% e os 30%.

Posteriormente fez-se uma **análise gráfica**, apresentada na figura 4.4, onde se pode confirmar alguns valores referidos anteriormente:

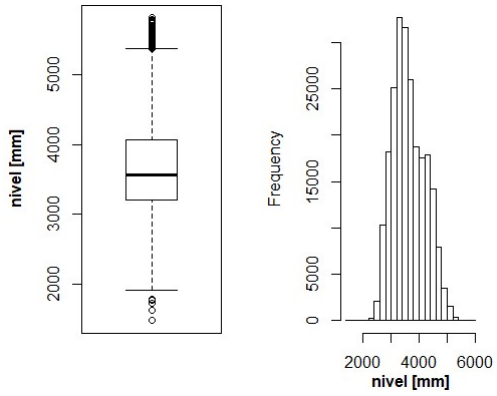

Figura 4.4: Boxplot e Histograma da variável nível.

É possível notar que há uma assimetria positiva, pois o valor de assimetria é ligeiramente maior que zero e no histograma da figura 4.4 vê-se uma pequena "cauda" de pontos de distribuição à direita, ou seja, a frequência de haver um nível entre 1480 e 3559 é maior que os restantes valores.

No boxplot consegue-se ver a presença de alguns outliers, em que todos os 99 outliers encontrados são moderados.

4.3.3 Variável *Potência*

A variável *Potência* representa a energia consumida pelos grupos em funcionamento numa determinada unidade de tempo (em kW), sendo uma **variável quantitativa** com uma **escala absoluta contínua**. Foram encontrados os seguintes indicadores descritivos:

- Mínimo: 5 kW
- Máximo: 962 kW
- Média: 326.0581 kW
- Mediana: 322 kW
- $1^{\underline{o}}$ Quartil: 280 kW
- $3^{\underline{o}}$ Quartil: 365 kW
- IQR: 85 kW
- Desvio-padrão: 99.3144 kW
- CVR: 26.3975%
- Curtose: 3.5564
- Assimetria: 1.0408
- Valores omissos: 18462

Com base nos resultados anteriores, verificou-se que das 312191 observações, 18462 são valores omissos, tendo informação apenas de 293729 observações.

Contatou-se que a mediana é de 322 kW, que o $1^{\underline{o}}$ quartil é de 280 kW e que o $3^{\underline{o}}$ quartil é igual a 365 kW, tendo amplitude inter-quartis seja de 85 kW. Os dados apresentam uma variabilidade média, pois o coeficiente de variação resistente está entre 15% e 30%.

De seguida fez-se uma **análise gráfica**, apresentada na figura 4.5, onde se pode confirmar alguns valores referidos anteriormente:

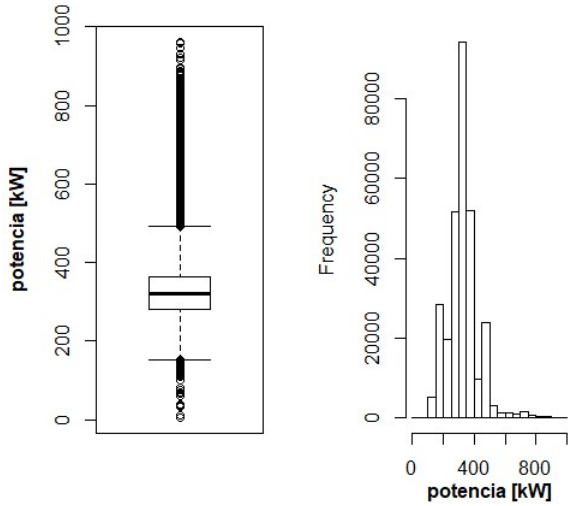

Figura 4.5: Boxplot e Histograma da variável potência.

Pode-se observar que há uma assimetria positiva e no histograma da figura 4.5 vê-se uma uma "cauda" de pontos de distribuição à direita, ou seja, a frequência de haver um nível entre 0 e 322 é maior que os restantes valores.

Constatou-se a presença de 16182 outliers, sendo 11574 são moderados e 4608 graves.

4.3.4 Variável *Indicador de eficiência*

A variável *Indicador de eficiência* representa o consumo de energia (em kWh/m^3), segundo o caudal elevado e a potência fornecida por cada grupo. É uma **variável quantitativa** com uma **escala absoluta contínua**. Encontrou-se os seguintes indicadores descritivos:

- Mínimo: 0.0036 kWh/m^3
- Máximo: 16.0769 kWh/m^3
- Média: 0.1177 kWh/m^3
- Mediana: 0.1080 kWh/m^3
- 1º Quartil: 0.1001 kWh/m^3
- 3º Quartil: 0.1223 kWh/m^3
- IQR: 0.0222 kWh/m^3
- Desvio-padrão: 0.0698 kWh/m^3
- CVR: 20.5556%
- Curtose: 16488.4854
- Assimetria: 93.0560
- Valores omissos: 111631

Com base nos resultados anteriores, nota-se que mais de um terço das observações são valores omissos.

Verificou-se também que a mediana é de 0.1080 kWh/m^3, o $1^{\underline{o}}$ quartil é de 0.1001 kWh/m^3 e que o $3^{\underline{o}}$ quartil é igual a 0.1223 kWh/m^3, tendo uma amplitude inter-quartis seja de 0.0222 kWh/m^3 . Os dados apresentam uma variabilidade média.

Por fim fez-se uma **análise gráfica**, apresentada na figura 4.6, onde se pode confirmar alguns resultados referidos anteriormente:

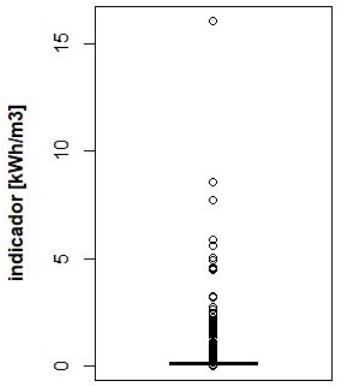

Figura 4.6: Boxplot da variável indicador de eficiência.

Há uma grande assimetria entre os dados e vê-se que quase todos os dados estão entre 0 e 1 kWh/m^3, tendo depois vários outliers acima desse valor .

Foram identificados 19103 outliers, sendo que 11502 são moderados e 7601 severos.

4.3.5 Variável *Código*

A variável *Código* representa os grupos eletrobomba que se encontram em funcionamento em cada instante, sendo uma **variável qualitativa** com uma **escala nominal**.

Na tabela 4.1 tem-se a frequência com que foram utilizados os conjuntos de grupos eletrobomba referidos anteriormente, sendo que "1" é quando o grupo está em funcionamento e "0" quando está desligado. Por exemplo, o código "01010"significa que no instante em que ocorre a observação, encontrava-se em funcionamento os grupos 2 e 4.

Também se considerou o código "99999" quando a informação sobre o conjunto grupos é desconhecida.

Tabela 4.1: Frequências de funcionamento das combinações dos grupos eletrobomba

Nº de Grupos	Código	Frequência absoluta	Frequência relativa
0	00000	4562	1.46%
1	10000	19609	6.28%
	01000	5464	1.75%
	00010	23189	7.43%
	00001	742	0.24%
2	11000	30964	9.92%
	10010	51912	16.63%
	10001	1632	0.52%
	01010	52704	16.88%
	01001	951	0.30%
	00011	975	0.31%
3	11010	18849	6.04%
	11001	1379	0.44%
	10011	2896	0.93%
	01011	1054	0.34%
4	11011	2824	0.90%
NA	99999	92485	29.62%
Total		312191	100%

Pelos dados da tabela 4.1, vê-se que quando os grupos funcionam individualmente, o que trabalha a maior parte do tempo é o grupo 4 e a seguir o grupo 1. Quando estão agrupados 2 a 2, o tempo de funcionamento da combinação composta pelos grupos 1 e 4 e a combinação formada pelos grupos 2 e 4, é consideravelmente maior que os restantes. Nos instantes em que estão em funcionamento 3 grupos simultaneamente, das 4 combinações possíveis, a mais usada é a formada pelos grupos 1, 2 e 4. O que nos permite concluir que, no período de tempo considerado, o grupo que trabalha mais é o 4 e o menos utilizado é o 5.

É igualmente importante notar que, no período a que respeitam os dados analisados, em aproximadamente 30% do tempo, não temos informação do conjunto em funcionamento.

Para o funcionamento da estação elevatória é necessário que pelo menos um grupo eletrobomba funcione. A quantidade de grupos a funcionar depende do nível do poço e do caudal que queremos elevar [10]. Sendo que o grupo 3 não funcionou durante estes 2 anos, tem-se os seguintes conjuntos/padrões relativos ao número de grupos em funcionamento.

Tabela 4.2: Combinações de funcionamento dos grupos eletrobomba

Nº de grupos eletrobomba em funcionamento			
1	2	3	4
Grupo 1	Grupos 1 e 2	Grupos 1, 2 e 4	Grupos 1, 2, 4 e 5
Grupo 2	Grupos 1 e 4	Grupos 1, 2 e 5	
Grupo 4	Grupos 1 e 5	Grupos 1, 4 e 5	
Grupo 5	Grupos 2 e 4	Grupos 2, 4 e 5	
	Grupos 2 e 5		
	Grupos 4 e 5		

De seguida fez-se um estudo descritivo do caudal, da potência e do indicador de eficiência para as **diferentes combinações de funcionamento**:

- **Um grupo em funcionamento**

  a) Caudal

  A partir de um output gerado no R, pode-se ter em consideração alguns **indicadores descritivos** dos dados :

Tabela 4.3: Indicadores descritivos da variável caudal para um grupo em funcionamento

Código	**10000**	**01000**	**00010**	**00001**	**Global**
Mínimo	213	37	31	396	31
Máximo	5636	5361	4782	4075	5636
Média	2414.38	1398.48	2518.96	2272.64	2348
Mediana	2547.50	1323	2810.50	1905	2375
1^{o} Quartil	1507	1107	1459	1582	1393
3^{o} Quartil	3237	1661	3415	3101	3261
IQR	1730	554	1956	1519	1868
Desvio-padrão	967.33	476.38	1012.40	84.86	1005.53
CV	40.07	34.06	40.19	37.09	42.81
CVR	67	417	69.60	79.74	78.65
Curtose	-12	53	-1.41	-1.15	-1.36
Assimetria	-0.09	1.32	-0.23	0.52	0.03

Quando está um grupo em funcionamento, o mínimo de água que se elevou foi de 31 m^3/h, com o grupo 4 em funcionamento e o máximo foi 5636, com o grupo 1, contudo, o que elevou mais água, em média, foi o grupo 4, tendo também o maior desvio padrão.

De seguida fez-se uma **análise gráfica**, apresentada na figura 4.7, onde se pode confirmar alguns resultados referidos anteriormente:

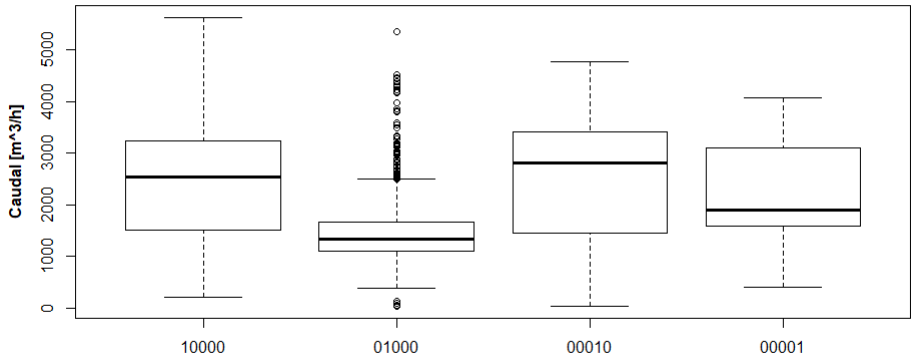

Figura 4.7: Boxplot da variável caudal para um grupo funcionamento.

Pela mediana, vê-se que há uma clara diferença entre o caudal do grupo 2 com os restantes, os seus valores encontram-se constantemente em torno do seu valor central, mas por outro lado apresenta inúmeros outliers.

b) Potência

Considerando alguns **indicadores descritivos** dos dados, tem-se :

Tabela 4.4: Indicadores descritivos da variável potência para um grupo em funcionamento

Código	**10000**	**01000**	**00010**	**00001**	**Global**
Mínimo	59	5	111	123	5
Máximo	804	481	662	545	804
Média	267.94	183.26	286.45	251.22	267.08
Mediana	287	169	304	252	282
1^{o} Quartil	184	152	211	160	185
3^{o} Quartil	324	201	343	305	330
IQR	140	49	132	145	145
Desvio-padrão	78.3	43.17	77.87	92.60	81.37
CV	29.22	51.85	27.18	36.86	30.47
CVR	48.78	28.99	43.42	57.54	51.42
Curtose	-0.59	2.17	-0.55	-0.06	-0.71
Assimetria	0.06	1.34	-0.01	0.69	0.15

Em média, a potência fornecida pelas bombas foi de 267 kW, no entanto, a mediana da potência fornecida pela bomba 2 é bastante inferior que as restantes e é a que tem menor variabilidade dos dados.

De seguida fez-se uma **análise gráfica**, apresentada na figura 4.8, onde se pode confirmar alguns valores referidos anteriormente:

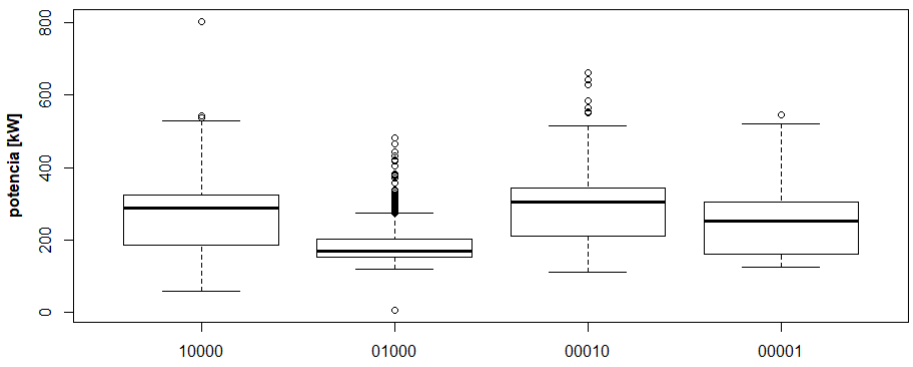

Figura 4.8: Boxplot da variável potência para um grupo em funcionamento.

Pelo valor da mediana, nota-se com clareza diferenças entre a potência do grupo 2 com os restantes, os seus valores encontram-se constantemente em torno do seu valor central, contudo apresenta inúmeros outliers.

c) Indicador de eficiência

Tendo-se em conta alguns **indicadores descritivos** dos dados, temos :

Tabela 4.5: Indicadores descritivos da variável indicador de eficiência para um grupo em funcionamento.

Código	**10000**	**01000**	**00010**	**00001**	**Global**
Mínimo	0.0286	0.0036	0.0475	0.0536	0.0036
Máximo	1.1300	5.0505	20.9444	0.5556	20.9444
Média	0.1243	0.1468	0.1255	0.1168	0.1268
Mediana	0.1083	0.1216	0.1092	0.0997	0.1101
1^{o} Quartil	0.0981	0.1046	0.0995	0.0899	0.0990
3^{o} Quartil	0.1287	0.1516	0.1293	0.1334	0.1309
IQR	0.0306	0.0470	0.0298	0.0435	0.0319
Desvio-padrão	0.0701	0.1063	0.1473	0.0445	0.0676
CV	56.3958	72.4114	117.3705	38.0993	53.3061
CVR	28.2548	38.6513	27.2894	43.6309	28.9665
Curtose	51.2043	912.5200	17533.6200	15.6266	1059.1152
Assimetria	6.3533	22.2200	125.1645	2.8089	19.1229

O grupo que teve o menor indicador de eficiência foi o grupo 2 e o que teve maior foi o grupo 4, contudo, é o grupo 5 que em média, tem os menores indicadores, tendo 50% dos seus valores menores que 0.0997 kWh/m^3, sendo que foi o grupo menos utilizado.

Na figura 4.9 está representado o gráfico de dispersão entre as variáveis caudal e potência:

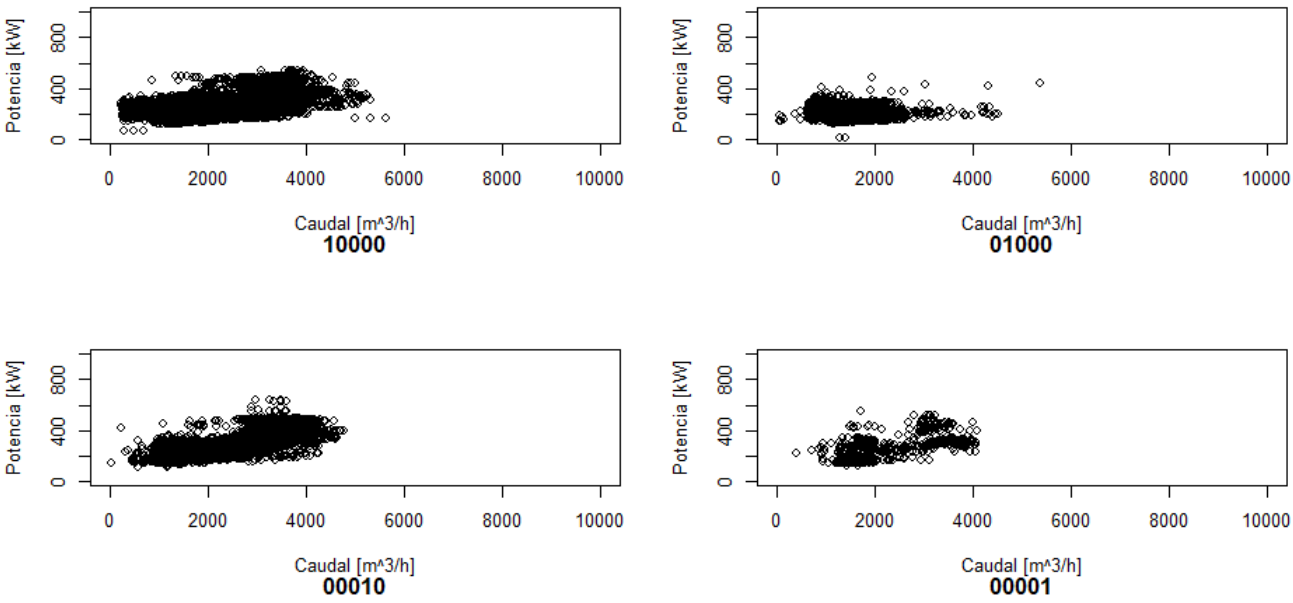

Figura 4.9: Gráfico de dispersão caudal vs potência para 1 grupo em funcionamento.

No gráfico de dispersão dos momentos em que está em funcionamento unicamente o grupo 2, não é evidente que haja algum tipo de correlação entre a potência e o caudal, ao contrário dos outros três gráficos, onde vê-se que quando o caudal aumenta, a potência também aumenta. Esta observação é mais evidente quando se encontra em funcionamento o grupo 1 ou o 4.

De modo a estudar a existência de associação linear significativa entre as duas variáveis, realizaram-se os testes à nulidade do coeficientes de correlação populacional.

Apresenta-se na tabela 4.6 os resultados das estimativas do coeficiente de correlação Pearson e de Spearman e os correspondentes valores-p.

Através do teste de associação entre as variáveis calculou-se os seguintes **coeficientes de correlação** e os respetivos **p-value**:

Tabela 4.6: Correlações entre o caudal e a potência para um grupo em funcionamento.

Código	**Coeficiente de Pearson**	**valores-p**	Coeficiente de Spearman	valores-p
10000	0.8090	< 0.001	0.8503	< 0.001
01000	0.0229	0.0948	0.0741	< 0.001
00010	0.8667	< 0.001	0.8833	< 0.001
00001	0.6280	< 0.001	0.6676	< 0.001

Enquanto a correlação de Pearson avalia associações lineares, a correlação de Spearman é uma medida de correlação não paramétrica que avalia associações monótonas, sejam elas lineares ou não.

Analisando os dois coeficientes e os respectivos valores-p, conclui-se que há uma correlação positiva entre o caudal e a potência, excepto quando está em funcionamento o grupo 2.

Representou-se também o gráfico de dispersão entre as variáveis caudal e o indicador de eficiência:

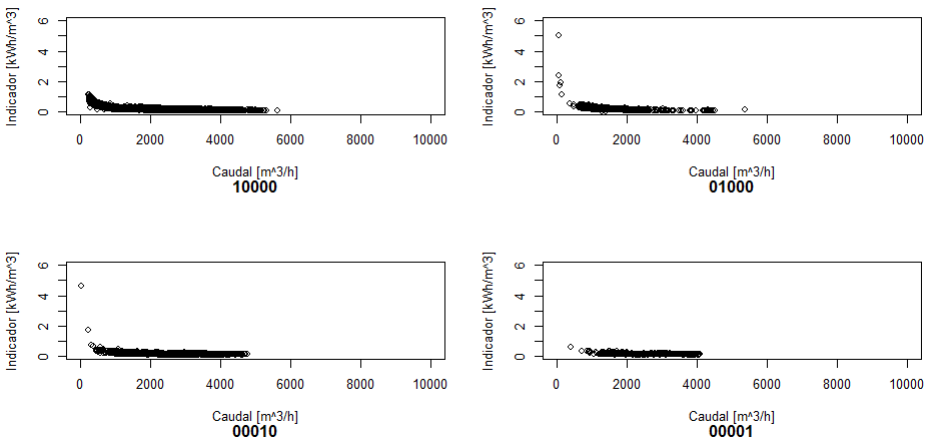

Figura 4.10: Gráfico de dispersão caudal vs indicador para um grupo em funcionamento

Observando as quatro combinações, constata-se alguma correlação negativa entre as duas variáveis, quando o caudal aumenta, o indicador de eficiência tende a diminuir.

De seguida calculou-se os seguintes **coeficientes de correlação** e os respetivos **valores-p** :

Tabela 4.7: Correlações entre o caudal e o indicador de eficiência para um grupo em funcionamento.

Código	**Coeficiente de Pearson**	**valores-p**	**Coeficiente de Spearman**	**valores-p**
10000	-0.5048	< 0.001	-0.7547	< 0.001
01000	-0.4891	< 0.001	-0.8225	< 0.001
00010	-0.5837	< 0.001	-0.8083	< 0.001
00001	-0.3803	< 0.001	-0.4311	< 0.001

Embora os coeficientes de correlação e os respectivos valores-p, confirmarem que há uma correlação negativa entre o caudal e o indicador de eficiência para os quatro grupos, no caso do grupo 5, o seu coeficiente de correlação Spearman é baixo sendo quase metade dos restantes, por isso não se pode concluir que há uma associação linear com significado estatístico entre o caudal e o indicador de eficiência quando está em funcionamento o grupo 5.

Em resumo, o grupo 2 apresenta um funcionamento "diferente"não sendo um dos grupos mais utilizados. Os grupos 1 e 4 são os mais utilizados e apresentam um comportamento semelhante.

- **2 Grupos em funcionamento**

  a) Caudal

  A partir de um output gerado no R, pode-se ter em consideração alguns **indicadores descritivos** dos dados :

Tabela 4.8: Indicadores descritivos da a variável caudal para dois grupos em funcionamento.

Código	**11000**	**10010**	**10001**	**01010**	**01001**	**00011**	**Global**
Mínimo	169	116	503	167	178	962	116
Máximo	5476	7801	6476	4891	4199	5323	7801
Média	2564.32	3191.28	2633.68	3226.64	2839.47	3522.37	3057.30
Mediana	2613	3197	2463	3289	2882	3663	3142
$1^{\underline{o}}$ Quartil	1846	2735	1655	2914	2347	2786	2633
$3^{\underline{o}}$ Quartil	3186	3665	3563	3601	3355	4209	3555
IQR	1340	930	1908	687	1008	1423	921
Desvio-padrão	759.43	713.25	1094.57	504.67	613.69	861.17	713.59
CV	29.62	22.35	41.56	15.64	21.61	24.45	23.34
CVR	51.28	29.09	77.47	20.89	34.98	38.85	29.33
Curtose	-0.91	-0.02	-0.92	0.60	-0.55	-0.9	-0.01
Assimetria	0.09	-0.02	0.46	-0.65	-0.32	-0.26	-0.34

Nos instantes em que funcionam dois grupos simultaneamente, a combinação que elevou menos água, em média, foi o formado pelos grupos 1 e 2. O que elevou mais água, em média, foi a combinação feita pelos grupos 4 e 5.

Na figura 4.11 apresenta os boxplot da variável caudal para dois grupos em funcionamento.

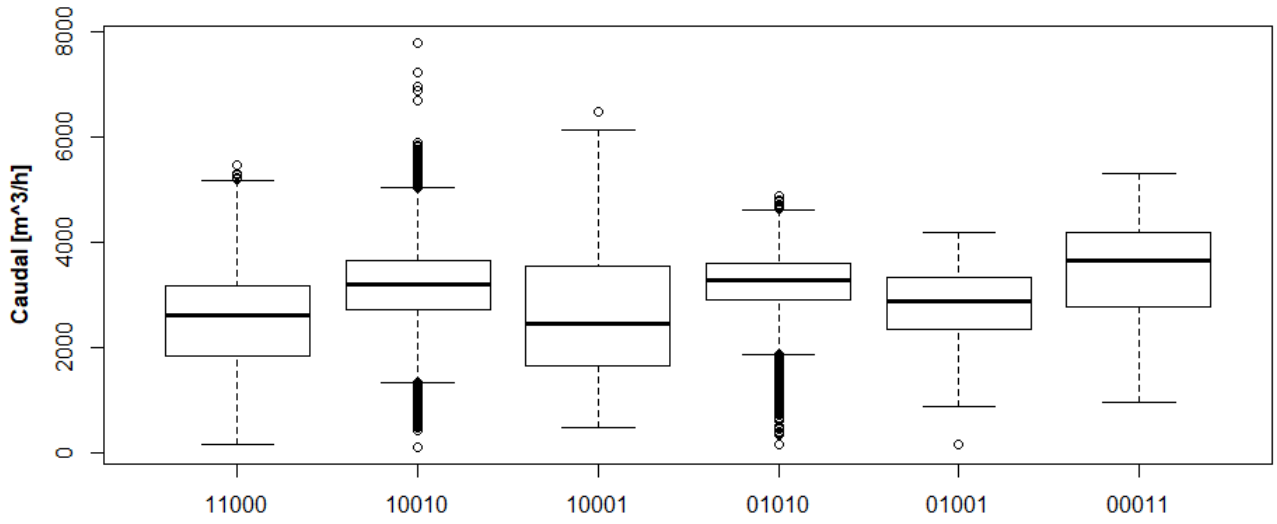

Figura 4.11: Boxplot da variável caudal para dois grupos em funcionamento.

O caudal do conjunto formado pelos 1 e 4 e o formado pelos grupos 2 e 4 apresentam inúmeros outliers, comparando com os restantes. Também é importante salientar, observando as barreiras dos boxplot, a grande dispersão da combinação composta pelos grupos 1 e 5.

b) Potência

Considerando alguns **indicadores descritivos** dos dados, temos :

Tabela 4.9: Indicadores descritivos da a variável potência para dois grupos em funcionamento.

Código	**11000**	**10010**	**10001**	**01010**	**01001**	**00011**	**Global**
Mínimo	124	136	59	136	123	161	59
Máximo	704	799	765	817	724	748	817
Média	298.96	354.36	343.66	340.75	322.04	443.59	337.32
Mediana	298	342	280	339	308	449	331
1^{o} Quartil	272	310	237	313	271	353	300
3^{o} Quartil	331	383	437	364	336	521	366
IQR	59	73	200	51	65	168	66
Desvio-padrão	55.87	73.16	141.03	45.58	83.64	123.16	67.06
CV	18.69	20.65	41.04	13.38	25.97	27.76	19.88
CVR	19.80	21.35	71.43	15.04	21.10	37.42	19.94
Curtose	2.03	2.12	0.24	3.91	5.68	0.01	3.50
Assimetria	0.39	0.85	1.06	0.67	1.86	0.12	0.95

O conjunto formado pelos grupos 1 e 5, tem a menor mediana, seguindo pelo conjunto com os grupos 1 e 2. Este por sua vez tem o menor 1^{o} e 3^{o} quartil, enquanto que a combinação dos grupos 4 e 5 tem a maior mediana e o maior 1^{o} e 3^{o} quartil.

De seguida fez-se uma **análise gráfica**, apresentada na figura 4.12, onde podemos confirmar alguns valores referidos anteriormente:

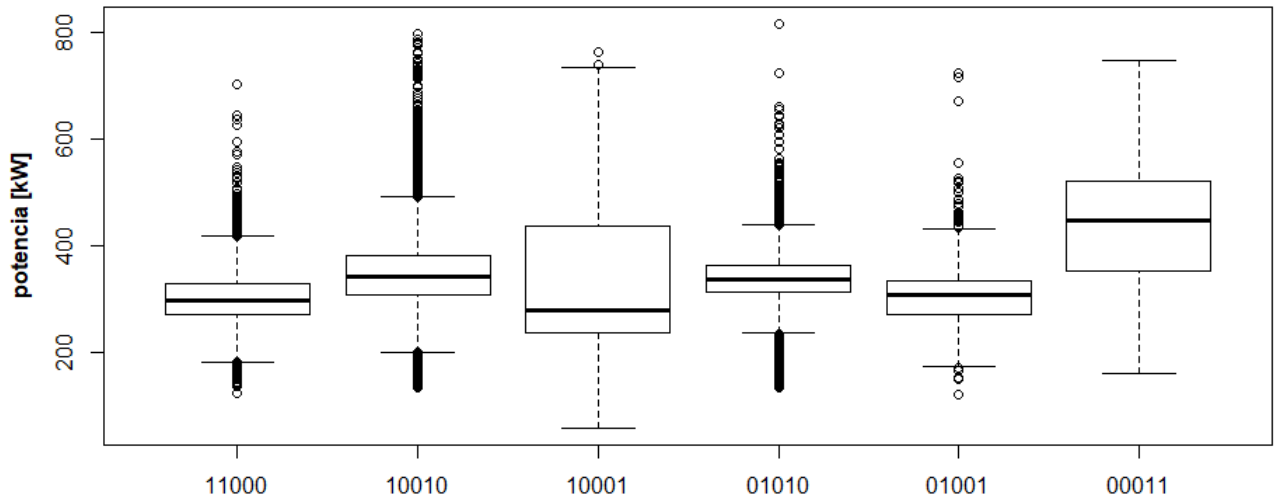

Figura 4.12: Boxplot da variável potência para dois grupos em funcionamento

A potência fornecida pelos conjuntos com os grupos 1 e 5 e com os grupos 4 e 5 parecem diferir das restantes. A potência dos outros conjuntos andam em volta do seu valor central, embora tenham muitos mais outliers que os outros dois conjuntos.

c) Indicador de eficiência

Tendo em conta alguns **indicadores descritivos** dos dados, temos :

Tabela 4.10: Indicadores descritivos da a variável indicador de eficiência para dois grupos em funcionamento.

Código	**11000**	**10010**	**10001**	**01010**	**01001**	**00011**	**Global**
Mínimo	0.0521	0.0581	0.0153	0.0735	0.0506	0.0504	0.0153
Máximo	1.5680	1.4397	0.5821	1.8922	1.5225	0.2983	1.8922
Média	0.1217	0.1117	0.1292	0.1069	0.1187	0.1155	0.1126
Mediana	0.1104	0.1078	0.1252	0.1020	0.1087	0,1106	0.1070
1^{o} Quartil	0.0996	0.1029	0.0995	0.0991	0.0946	0.1018	0.004
3^{o} Quartil	0.1380	0.1173	0.1555	0.1112	0.1347	0.1260	0.1185
IQR	0.0384	0.0144	0.0560	0.0121	0.0401	0.0242	0.0181
Desvio-padrão	0.0475	0.0331	0.0416	0.0414	0.0575	0.0256	0.0255
CV	39.0304	29.6329	32.1981	38.7278	48.4414	22.1645	22.6172
CVR	34.7826	13.3581	44.7284	11.8627	36.8905	21.8807	16.9196
Curtose	34.1920	64.2292	19.5636	70.8877	373.8164	5.5640	446.4206
Assimetria	0.9610	0.5393	2.2088	0.4145	15.7622	1.3533	9.8678

Observa-se que o conjunto com a menor mediana é o conjunto formado pelos grupos 2 e 4, contudo a que tem menor 1^{o} quartil é o formado pelos grupos 2 e 5, sendo que destes 2 conjuntos a que tem menor coeficiente de variação resistente é o formado pelos grupos 2 e 4.

Na figura 4.13 está representado o gráfico de dispersão entre as variáveis caudal e potência:

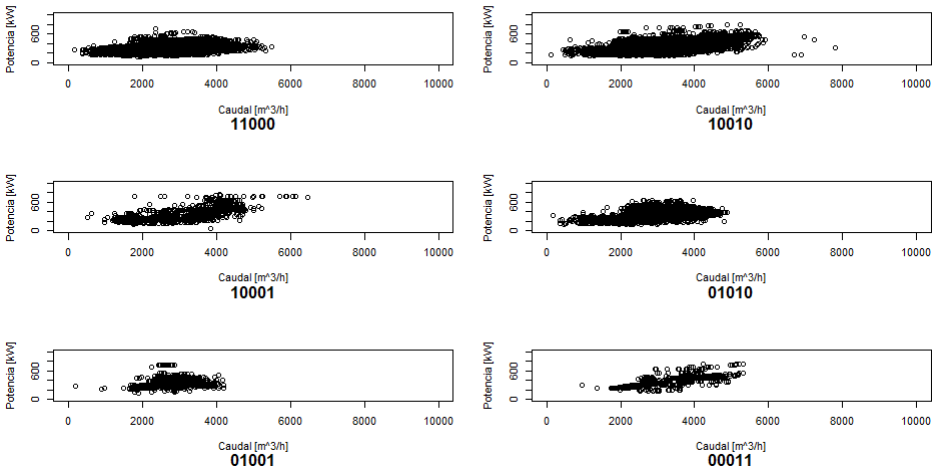

Figura 4.13: Gráfico de dispersão caudal vs potência para dois grupos em funcionamento.

Nos gráficos de dispersão da figura 4.13, com exceção do conjunto formado pelos grupos 2 e 5, observa-se alguma correlação positiva entre as variáveis, ou seja, quando o caudal aumenta, a potencia igualmente aumenta.

Seguidamente realizaram-se os testes à nulidade do coeficiente de correlação.

Na tabela 4.11 apresentam-se os coeficientes de correlação de Pearson e Spearman e os respetivos valores-p.

Tabela 4.11: Correlações entre o caudal e a potência para dois grupos em funcionamento

Código	**Coeficiente de Pearson**	**p-value**	**Coeficiente de Spearman**	**p-value**
11000	0.6179	< 0.001	0.7171	< 0.001
10010	0.7557	< 0.001	0.8161	< 0.001
10001	0.7300	< 0.001	0.7087	< 0.001
01010	0.6874	< 0.001	0.8150	< 0.001
01001	0.2426	< 0.001	0.4397	< 0.001
00011	0.6893	< 0.001	0.7184	< 0.001

Apesar de os coeficientes de correlação e os respetivos p-value comprovarem a existência de uma correlação positiva entre o caudal e a potência nos instantes em que funcionam dois grupos ao mesmo tempo, observa-se que os coeficientes de correlação do conjunto formado pelos grupos 2 e 5 é bastante inferior aos restantes, por isso não se pode assumir que há uma associação linear entre as duas variáveis nos instantes em que estão em funcionamento os grupos 2 e 5.

Representou-se na figura 4.14 o gráfico de dispersão entre as variáveis caudal e o indicador de eficiência:

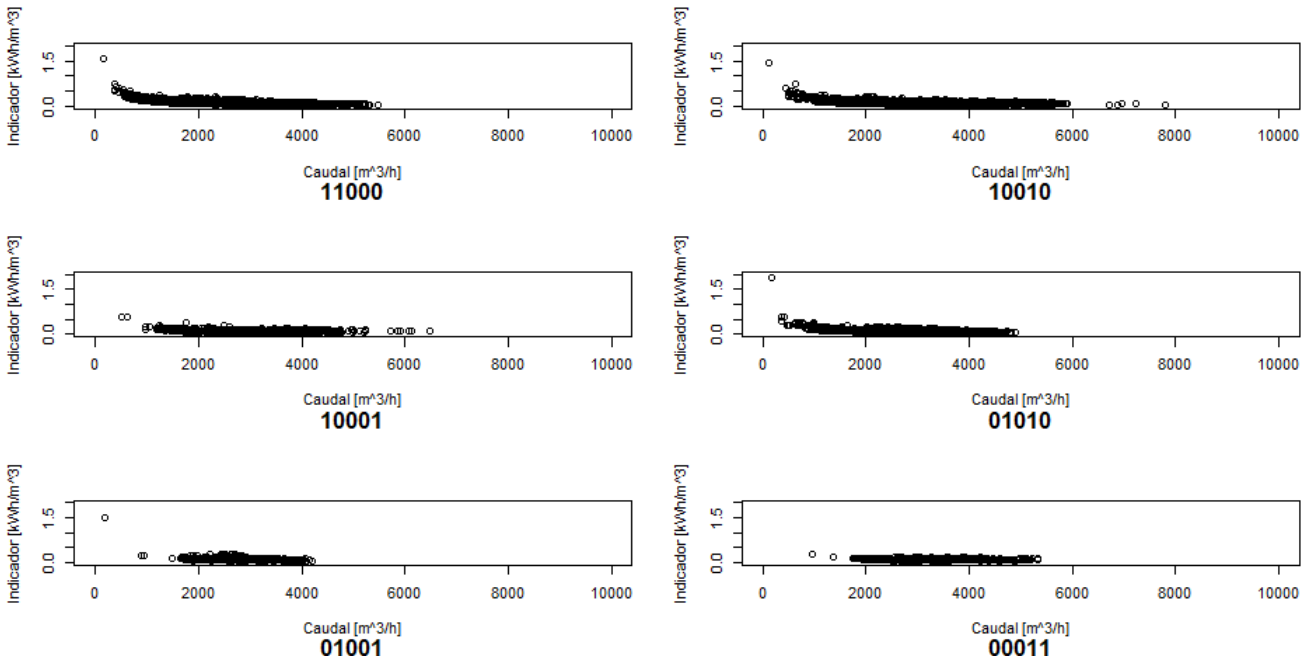

Figura 4.14: Gráfico de dispersão caudal vs indicador de eficiência para dois grupos em funcionamento.

É possível observar no gráfico dos instantes em que os grupos 1 e 2 trabalham ao mesmo tempo, que quando o caudal aumenta, o indicador de eficiência propende a diminuir enquanto que nos restantes gráficos não é possível fazer uma observação clara.

Seguidamente calculou-se os seguintes **coeficientes de correlação** e os respetivos **p-value** :

Tabela 4.12: Correlações entre o caudal e o indicador de eficiência para dois grupos em funcionamento.

Código	**Coeficiente de Pearson**	**p-value**	**Coeficiente de Spearman**	**p-value**
11000	-0.7135	< 0.001	-0.7857	< 0.001
10010	-0.5232	< 0.001	-0.5681	< 0.001
10001	-0.5319	< 0.001	-0.6914	< 0.001
01010	-0.5344	< 0.001	-0.6720	< 0.001
01001	-0.4250	< 0.001	-0.6716	< 0.001
00011	-0.2827	< 0.001	-0.3711	< 0.001

Embora os coeficientes de correlação e os respetivos valores-p demonstrarem que há uma correlação negativa entre o caudal e o indicador de eficiência, os coeficientes de correlação do conjunto dos grupos 4 e 5 é baixo, sendo metade dos restantes, deste modo não se pode assumir que há uma associação com significado estatístico entre as duas variáveis quando funcionam os grupos eletrobomba 4 e 5.

Em suma, as combinações formadas pelos grupos 2 e 5 e pelos grupos 4 e 5 apresentam um funcionamento "dissemelhante"aos demais, sendo das combinações com menor utilização.

- **Três grupos em funcionamento**

  a) Caudal

  Na tabela 4.13 apresentam-se alguns **indicadores descritivos** dos dados :

Tabela 4.13: Indicadores descritivos da variável caudal para três bombas em funcionamento

Código	**11010**	**11001**	**10011**	**01011**	**Global**
Mínimo	666	1556	1879	1861	666
Máximo	9552	6259	5691	5491	9552
Média	4365.23	4075.21	4261.22	4194.28	4336.42
Mediana	4351.5	4089	4297	4181.75	4340
1^{o} Quartil	4188	3784	4041	3791	4157
3^{o} Quartil	4520	4406	4545	4521	4516
IQR	332	622	504	730	359
Desvio-padrão	380.72	515.12	463.73	472.76	405.42
CV	8.72	12.64	10.88	11.27	9.35
CVR	7.63	15.21	11.73	17.46	8.28
Curtose	14.86	1.99	1.52	0.36	10.55
Assimetria	-0.66	-0.35	-0.49	0.06	-0.68

Com três grupos eletrobomba em funcionamento, elevou-se em média 4336 m^3/h. A quantidade de água elevada, em média, por cada uma das 4 combinações, não diferia muito, contudo, a combinação formada pelos grupos 1, 2 e 5, que eleva menos, é a que presenta maior desvio padrão.

De seguida fez-se uma **análise gráfica**,apresentada na figura 4.15, onde se pode confirmar alguns resultados referidos anteriormente:

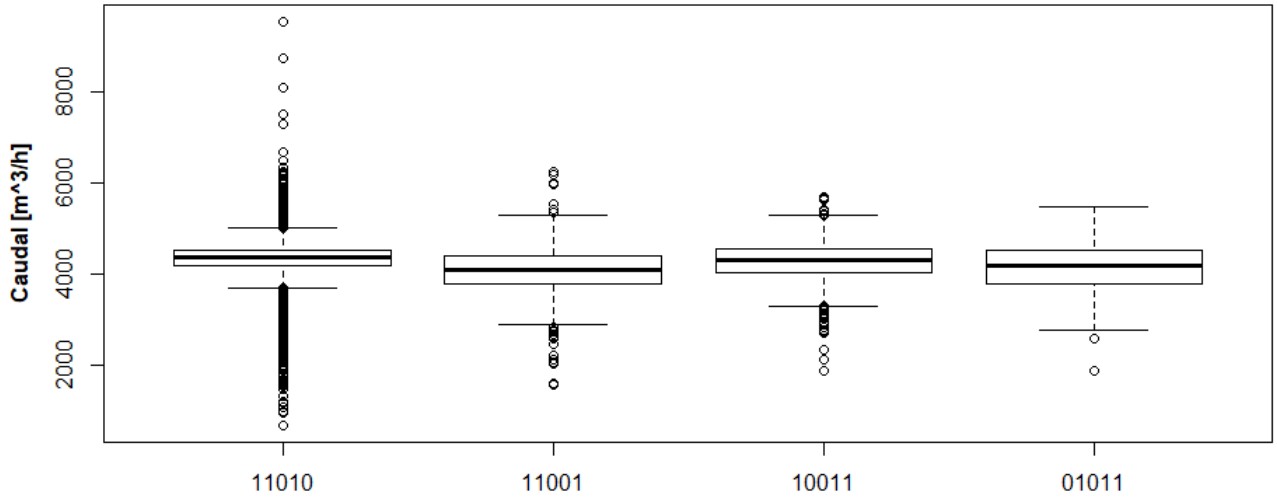

Figura 4.15: Boxplot da variável caudal para três bombas em funcionamento.

Apesar de a primeira combinação (grupos 1, 2 e 4) apresentarem muito mais outliers que as restantes 3 combinações, o caudal elevado pelas quatro combinações não parece diferir significativamente, as suas médias e medianas são semelhantes e os valores das variáveis situam-se constantemente em torno dos seus valores centrais.

b) Potência

Considerando alguns **indicadores descritivos** dos dados, tem-se :

Tabela 4.14: Indicadores descritivos da variável potência para três bombas em funcionamento.

Código	**11010**	**11001**	**10011**	**01011**	**Global**
Mínimo	32	212	69	179	32
Máximo	810	778	896	774	896
Média	459.59	451.86	561.89	491.47	474.23
Mediana	477	444	500	454	477
$1^{\underline{o}}$ Quartil	432	411	449	435	435
$3^{\underline{o}}$ Quartil	483	478	719	537	486
IQR	51	67	270	102	51
Desvio-padrão	72.69	104.39	152.60	101.58	97.46
CV	15.82	23.10	27.16	20.67	20.55
CVR	10.69	15.09	54.00	22.47	10.69
Curtose	3.16	1.11	-0.93	1.02	2.99
Assimetria	-0.06	0.69	0.32	0.78	1.00

A potência fornecida, em média, pelos grupos quando trabalham três ao mesmo tempo é de 474 kW. O conjunto de grupos eletrobomba que fornece mais potência

é o formado pelos grupos 1, 4 e 5 e também é o que tem maior desvio padrão.

Na figura 4.16 apresenta-se o boxplot da variável potência para três grupos em funcionamento, onde se pode confirmar resultados referidos anteriormente.

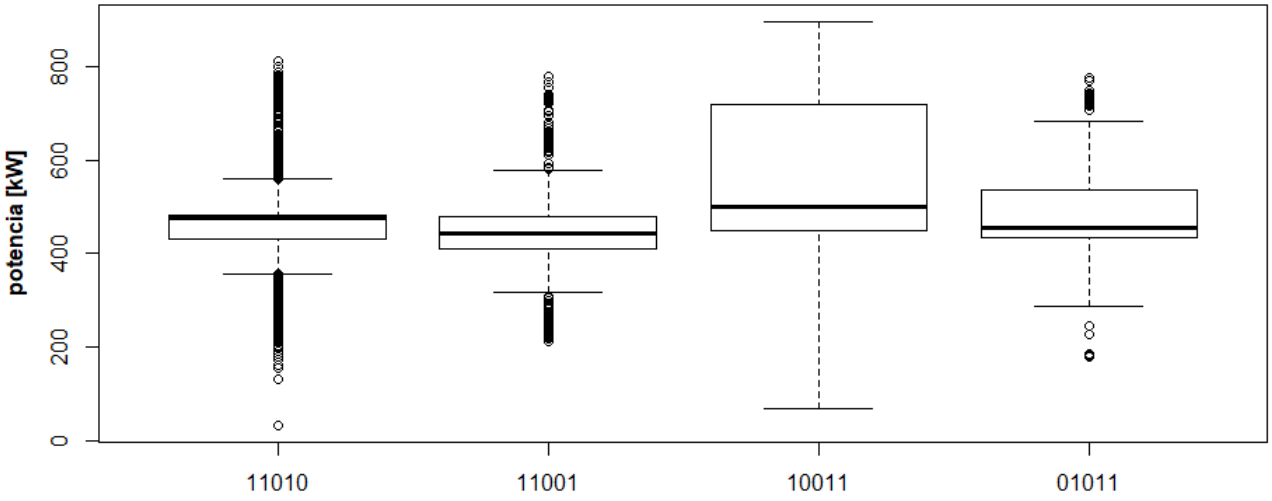

Figura 4.16: Boxplot da variável potência para três bombas em funcionamento.

A potência do conjunto dos grupos 1, 4 e 5 destaca-se dos demais conjuntos. Tem uma maior variabilidade dos dados, o seu valor central encontra-se mais acima e ao contrário dos outros grupos, não apresenta nenhum outlier.

c) Indicador de eficiência

Tendo-se em conta alguns **indicadores descritivos** dos dados, tem-se :

Tabela 4.15: Indicadores descritivos da variável indicador de eficiência para 3 bombas em funcionamento

Código	**11010**	**11001**	**10011**	**01011**	**Global**
Mínimo	0.0161	0.0538	0.0617	0.0496	0.0161
Máximo	0.3677	0.1924	0.2363	0.1810	0.3677
Média	0.1048	0.1076	0.1106	0.1109	0.1055
Mediana	0.1076	0.1069	0.1072	0.1076	0.1076
1^{o} Quartil	0.0965	0.0965	0.1013	0.1006	0.0972
3^{o} Quartil	0.1135	0.1177	0.1188	0.1155	0.1138
IQR	0.0170	0.0212	0.0175	0.0149	0.0166
Desvio-padrão	0.0175	0.0223	0.0236	0.0219	0.01854
CV	16.6985	20.7249	21.3382	19.7475	17.5695
CVR	15.7993	19.8316	16.3246	13.8476	15.4198
Curtose	12.1795	2.0072	2.4668	1.5222	9.0609
Assimetria	1.0562	0.7000	0.9940	0.7935	1.0601

Os conjuntos de grupos eletrobomba com maior indicador de eficiência, em média, é o formado pelos grupos 1, 4 e 5 e os grupos 2, 4 e 5. A combinação com menor indicador de eficiência é a feita pelos grupos 1, 2 e 4 e é o que tem menor desvio padrão.

Na figura 4.17 representa-se o boxplot da variável indicador de eficiência para três bombas em funcionamento.

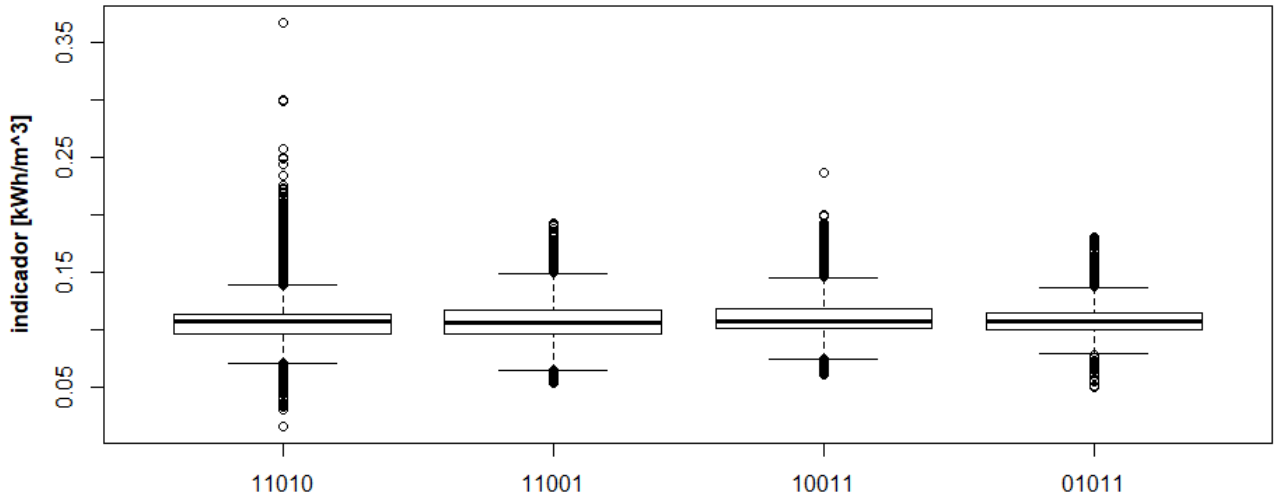

Figura 4.17: Boxplot da variável indicador de eficiência para três bombas em funcionamento.

Na figura 4.18 está representado o gráfico de dispersão entre as variáveis caudal e potência:

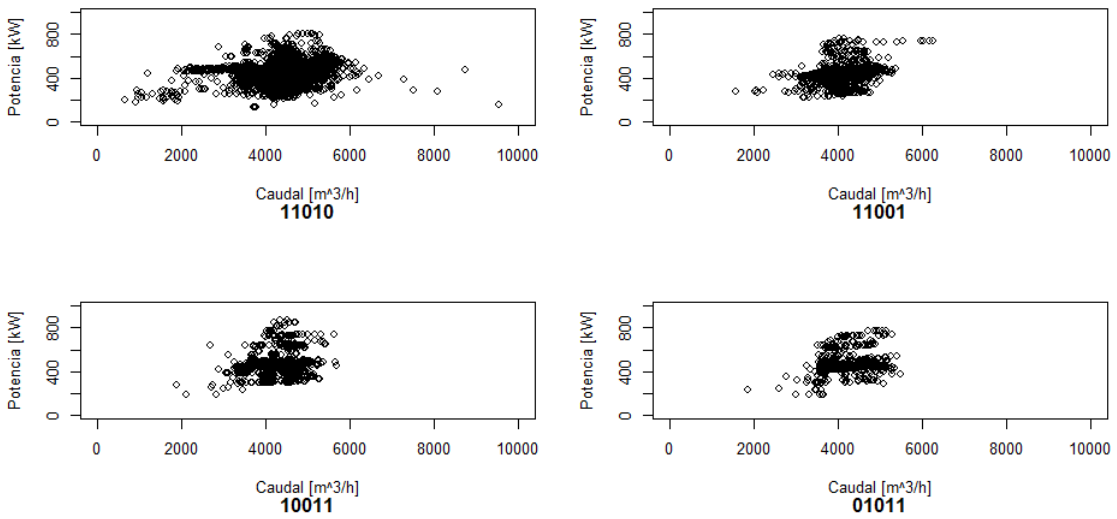

Figura 4.18: Gráfico de dispersão caudal vs potência para três grupos em funcionamento.

Analisando os gráficos da figura 4.18, não é possível ter uma ideia clara do comportamento entre o caudal e a potência.

Na tabela 4.16 apresenta-se os coeficientes de correlação entre o caudal e a potência e os respectivos valores-p.

Tabela 4.16: Correlações entre o caudal e a potência para três grupos em funcionamento

Código	**Coeficiente de Pearson**	**valores-p**	**Coeficiente de Spearman**	**valores-p**
11010	0.1458	< 0.001	0.2203	< 0.001
11001	0.3083	< 0.001	0.4315	< 0.001
10011	0.2514	< 0.001	0.3680	< 0.001
01011	0.2649	< 0.001	0.3349	< 0.001

A partir dos coeficientes de correlação da tabela 4.10, não se pode assumir que existe uma associação linear entre as variáveis com significado estatístico, pois estes valores são muito baixos, portanto a correlação é fraca.

Representou-se também o gráfico de dispersão entre as variáveis caudal e o indicador de eficiência:

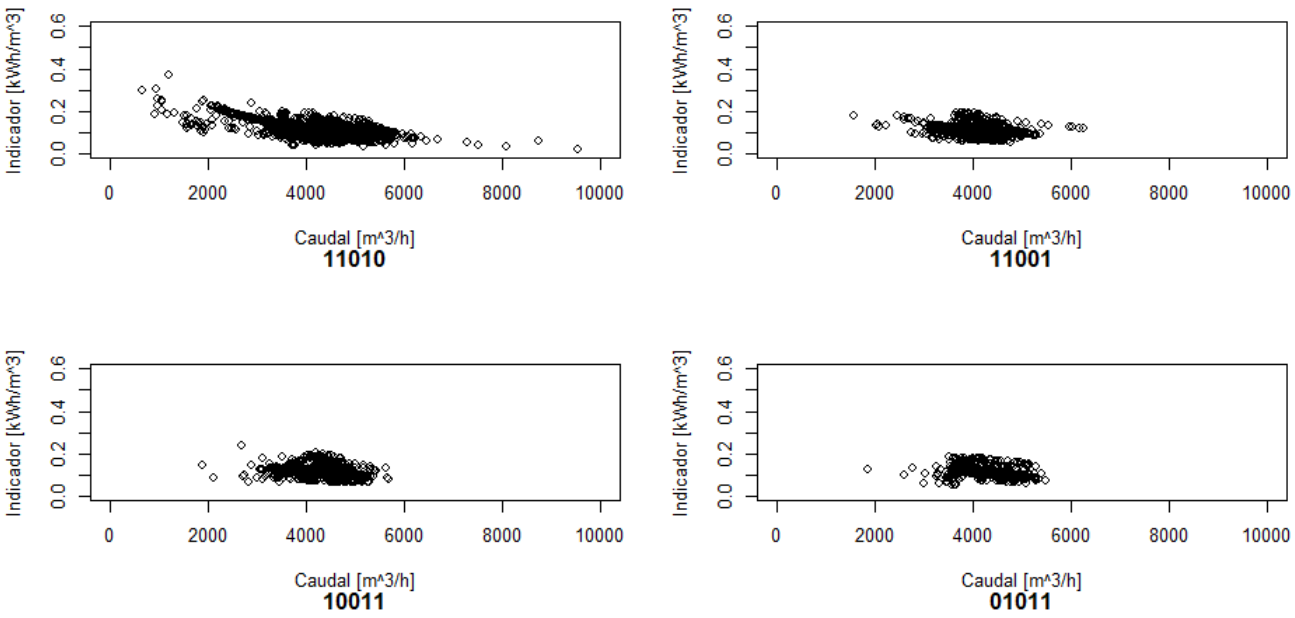

Figura 4.19: Gráfico de dispersão caudal vs indicador para três grupos em funcionamento

No gráfico dos instantes em que estão em funcionamento simultaneamente os grupos 1, 2 e 4, nota-se que, à medida que o caudal aumenta, o indicador de eficiência também diminui, enquanto que nos restantes gráficos não é tão claro.

Representa-se na tabela 4.17 as estimativas dos coeficientes de correlação de Pearson e de Spearman e os valores-p dos testes à nulidade dos coeficientes de correlação.

Tabela 4.17: Correlações entre o caudal e o indicador para três grupos em funcionamento

Código	**Coeficiente de Pearson**	**valores-p**	**Coeficiente de Spearman**	**valores-p**
11010	-0.5138	< 0.001	-0.5364	< 0.001
11001	-0.3273	< 0.001	-0.4387	< 0.001
10011	-0.2499	< 0.001	-0.3762	< 0.001
01011	-0.2834	< 0.001	-0.4994	< 0.001

Na generalidade dos casos, a associação entre as variáveis nas diferentes combinações é fraca.

Em conclusão tem-se que as quatro combinações têm funcionamentos semelhantes entre si, sendo que a combinação dos grupos eletrobomba 1, 2 e 4 foi mais utilizada, por outro lado os grupos eletrobomba têm funcionamentos "diferentes"do que quando funcionam individualmente ou em combinações de dois.

4.4 Inferência Estatística

De forma a se poder tirar conclusões sobre as combinações de bombas, é necessário utilizar métodos de inferência estatística, mais concretamente o teste de Kruskal-Wallis e os testes de comparação múltipla e de seguida analisar gráficos de densidade kernel, de maneira a inferir a combinação de grupos eletrobomba mais eficiente.

4.4.1 Um grupo eletrobomba em funcionamento

- Caudal

De modo a comparar se a quantidade de água elevada pelos grupos eletrobomba, quando funcionam individualmente, é idêntica ou não, fez-se um teste de Kruskal-Wallis.

Obtiveram-se os seguintes resultados:

- **Valor observado da estatística de teste e valor-p**

  Valor observado = 5263.4 e valores-p < 0.001

- **Decisão e conclusão**

  Como o valores-p $< \alpha = 0.05$, rejeita-se a hipótese nula a este nível de significância, ou seja, conclui-se que a distribuição da quantidade de água elevada não é idêntica para pelo menos dois pares de grupos.

- Potência

De seguida, comparou-se se a distribuição da potência fornecida pelos grupos eletrobomba, quando trabalham individualmente, é idêntica para todos os grupos.

Obtiveram-se os seguintes resultados:

- **Valor observado da estatística de teste e valor-p**

  Valor observado = 7709.5 e valores-p < 0.001

- **Decisão e conclusão**

  Como o valores-p $< \alpha = 0.05$, rejeita-se a hipótese nula a este nível de significância, ou seja, conclui-se que a distribuição da potência fornecida pelos grupos não é idêntica para pelo menos dois grupos eletrobomba.

- Indicador de eficiência

Seguidamente, verificou-se se a distribuição dos indicadores de eficiência é idêntica para os quatro grupos.

Obtiveram-se os seguintes resultados:

- **Valor observado da estatística de teste e valor-p**

  Valor observado = 683.46 e valores-p < 0.001

- **Decisão e conclusão**

  Como o valores-p $< \alpha = 0.05$, rejeita-se a hipótese nula a este nível de significância, ou seja,conclui-se que a eficiência não é idêntica para pelo menos 2 grupos eletrobomba.

Com os resultados observados concluiu-se que o caudal elevado, a potência fornecida e os indicadores de eficiência **não são idênticos** em pelo menos dois grupos, então fez-se um **teste de comparação múltipla de Bonferroni**, para averiguar quais os grupos que diferem entre si, tendo-se obtido os seguintes resultados:

Tabela 4.18: Testes de comparação múltipla para a variável caudal com um grupo em funcionamento.

Caudal			
Bombas	**10000**	**01000**	**00010**
01000	< 0.001	-	-
00010	< 0.001	< 0.001	-
00001	0.2	< 0.001	< 0.001

Tabela 4.19: Testes de comparação múltipla para a variável potência com um grupo em funcionamento.

Potência			
Bombas	**10000**	**01000**	**00010**
01000	< 0.001	-	-
00010	< 0.001	< 0.001	-
00001	< 0.001	< 0.001	< 0.001

Tabela 4.20: Testes de comparação múltipla para a variável indicador de eficiência com um grupo em funcionamento.

Indicador de eficiência			
Bombas	**10000**	**01000**	**00010**
01000	< 0.001	-	-
00010	< 0.001	< 0.001	-
00001	< 0.001	< 0.001	< 0.001

Com base nos valores-p das tabelas 4.18, 4.19 e 4.20 conclui-se que os vários grupos eletrobomba são diferentes entre si para todas as variáveis, com exceção dos grupos 1 e 5, que elevam, em média, caudais idênticos como se pode averiguar na tabela 4.18.

De modo a inferir sobre o grupo eletrobomba mais eficiente, é necessário comparar as variáveis de cada grupo. Para tal utilizou-se o método de Kernel para estimar as diferentes funções de densidade.

Na figura 4.20 apresentam-se os gráficos de densidade estimada da variável caudal, para um grupo em funcionamento:

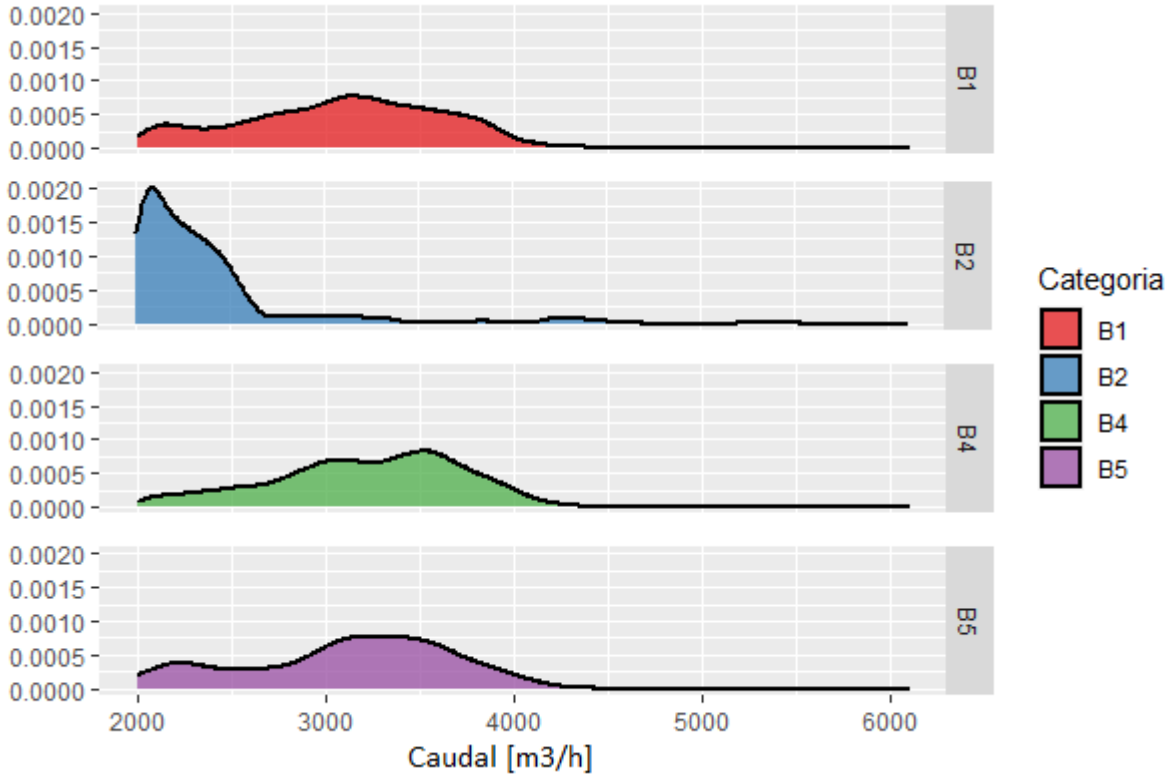

Figura 4.20: Gráficos de densidade estimada da variável caudal, para um grupo em funcionamento.

Pelos gráficos da figura 4.20 observa-se que o grupo que eleva caudais menores com mais frequência é o grupo 2. Por exemplo, a probabilidade estimada de obter caudais menores que 2500 m_3/h é superior no grupo 2. Os restantes grupos elevam caudais superiores, sensivelmente com a mesma frequência.

Na figura 4.21 apresentam-se os gráficos de densidade estimada da variável potência, para um grupo em funcionamento.

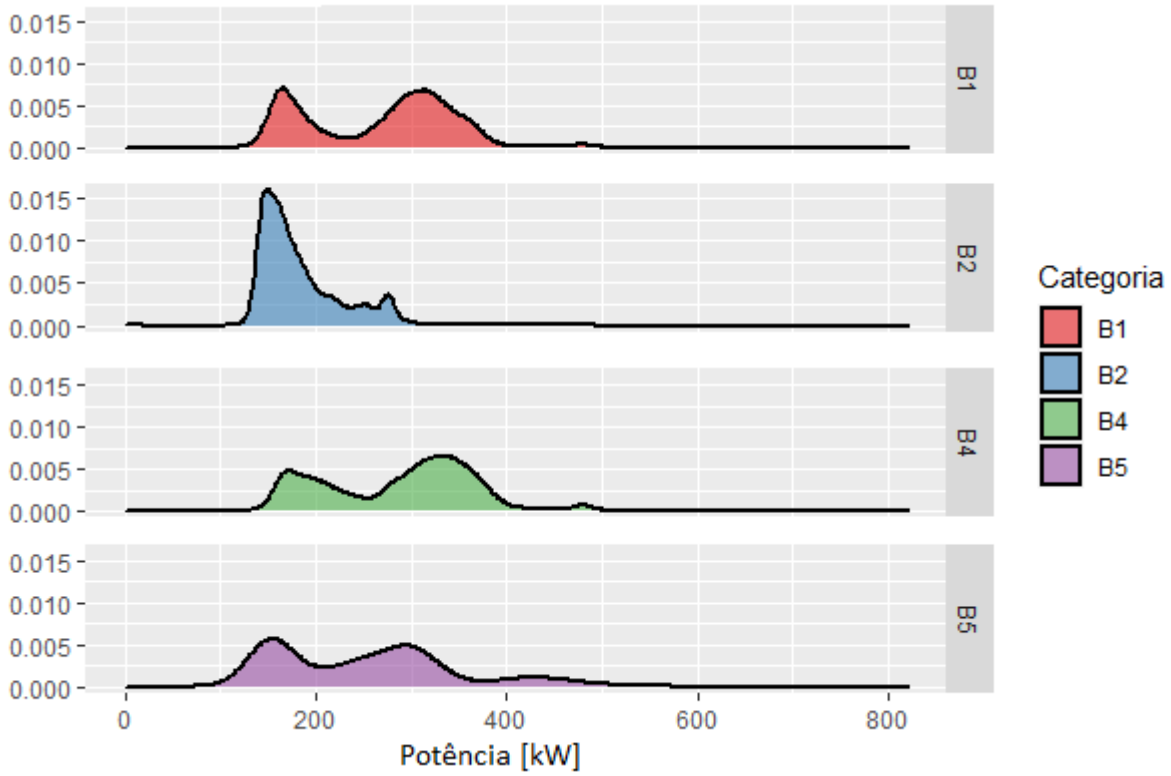

Figura 4.21: Gráficos de densidade estimada da variável potência, para 1 grupo em funcionamento.

Pelos gráficos da figura 4.21 nota-se que a probabilidade estimada de fornecer potência menores, por exemplo entre 100 e 200 kW, são maiores no grupo 2.

Na figura 4.22 mostram-se os gráficos de densidade estimada da variável indicador de eficiência, para um grupo em funcionamento.

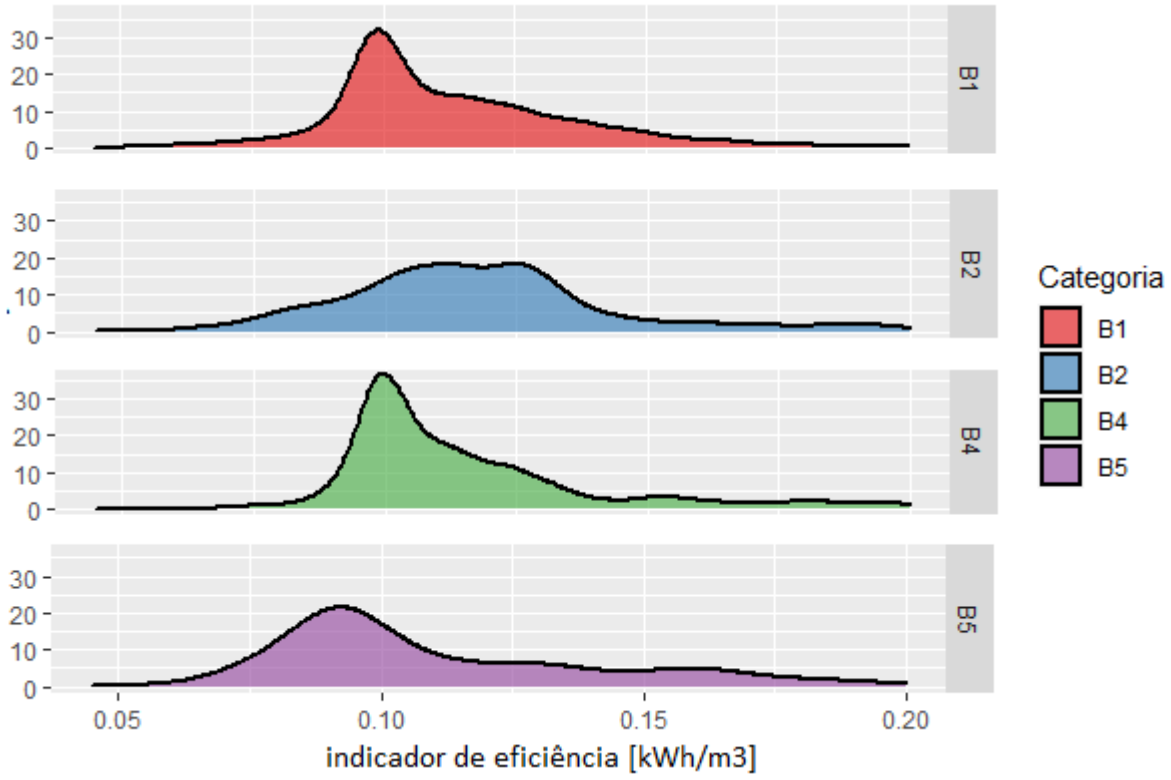

Figura 4.22: Gráfico de densidade estimada da variável indicador de eficiência, para 1 grupo em funcionamento.

Pelos gráficos da figura 4.22 verifica-se que a probabilidade de se ter indicadores de eficiência inferiores, é maior no grupo 5, seguindo o grupo 2. Contudo, o grupo 2 apresenta indicadores de eficiência mais elevados com alguma frequência.

Em suma, uma vez que na tabela 4.5, relativa às medidas descritivas do indicador de eficiência, verificou-se que o grupo 5 tem a menor mediana e o menor 1^{o} quartil e sendo que quanto menor o indicador de eficiência, mais eficiente é o grupo eletrobomba, conclui-se que quando os grupos estiverem em funcionamento individualmente, o mais eficiente seria utilizar o grupo eletrobomba 5, seguindo pelo grupo 1, uma vez que apresenta a segunda menor mediana, o segundo menor 1^{o} quartil e ainda o menor 3^{o} quartil.

4.4.2 Dois grupos eletrobomba em funcionamento

- Caudal

De forma a comparar se a quantidade de água elevada pelos conjuntos de grupos eletrobomba, quando funcionar em grupos de 2, é idêntica ou não, fez-se um teste de Kruskal-Wallis:

Obteve-se os seguintes valores:

- **Valor observado da estatística de teste e valor-p**

  H_{obs}= 16721 e valores-p < 0.001

- **Decisão e conclusão**

  Como o valores-p $< \alpha = 0.05$, rejeita-se a hipótese nula a este nível de significância, ou seja,conclui-se que a quantidade de água elevada, em média, não é idêntica para os 6 conjuntos.

- Potência

De seguida, averiguou-se se a potência fornecida pelos conjuntos, nos momentos em que funcionam em grupos de 2, é idêntica para todos os conjuntos.

Obteve-se os seguintes valores:

- **Valor observado da estatística de teste e valor-p**

  H_{obs}=16489 e valores-p < 0.001

- **Decisão e conclusão**

  Como o valores-p $< \alpha = 0.05$, rejeita-se a hipótese nula a este nível de significância, ou seja, conclui-se que a potência fornecida, em média, não é idêntica todos os conjuntos.

- Indicador de eficiência

Depois, verificou-se se a eficiência dos conjuntos, nos instantes em que encontram-se em funcionamento combinações de dois grupos, é idêntica para os 6 conjuntos.

Obteve-se os seguintes valores:

- **Valor observado da estatística de teste e valor-p**

  H_{obs}= 7087.3 e valores-p < 0.001

- **Decisão e conclusão**

  Como o valores-p $< \alpha = 0.05$, rejeita-se a hipótese nula a este nível de significância, ou seja, conclui-se que a eficiência não é idêntica para os 6 conjuntos de grupos eletrobomba.

Uma vez que se concluiu que os 6 grupos **não são idênticos** para todas as variáveis, fez-se um **teste de comparação múltipla de Bonferroni**, para averiguar quais grupos diferem entre si, obtendo os seguintes resultados:

Tabela 4.21: Testes de comparação múltipla para a variável caudal com dois grupos em funcionamento.

Caudal					
Bombas	**11000**	**10010**	**10001**	**01010**	**01001**
10010	< 0.001	-	-	-	-
10001	< 0.001	< 0.001	-	-	-
01010	< 0.001	< 0.001	< 0.001	-	-
01001	< 0.001	< 0.001	0.075	< 0.001	-
00011	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

Tabela 4.22: Testes de comparação múltipla para a variável potência com dois grupos em funcionamento.

Potência					
Bombas	**11000**	**10010**	**10001**	**01010**	**01001**
10010	< 0.001	-	-	-	-
10001	< 0.001	< 0.001	-	-	-
01010	< 0.001	< 0.001	< 0.001	-	-
01001	< 0.001	< 0.001	0.16	< 0.001	-
00011	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

Tabela 4.23: Testes de comparação múltipla para a variável indicador de eficiência com dois grupos em funcionamento.

Indicador de eficiência					
Bombas	**11000**	**10010**	**10001**	**01010**	**01001**
10010	< 0.001	-	-	-	-
10001	0.976	< 0.001	-	-	-
01010	< 0.001	< 0.001	< 0.001	-	-
01001	0.692	1.000	0.087	< 0.001	-
00011	< 0.001	< 0.001	< 0.001	0.660	< 0.001

Pelos valores-p da tabela 4.21 conclui-se que todas as combinações elevam, em média, caudais diferentes, com exceção das combinações formadas pelos grupos 1 e 5 e pelos grupos 2 e 5.

Observando os valores-p da tabela 4.22, conclui-se que todos os conjuntos fornecem, em média, potências diferentes excepto os conjuntos formados pelos grupos 1 e 5 e pelos grupos 2 e 5.

Por outro lado, com base na tabela 4.23, conclui-se que :

- O indicador de eficiência do conjunto formado pelos grupos 1 e 5 é idêntico ao conjunto formado pelos grupos 1 e 2;
- O indicador de eficiência da combinação formada pelos grupos 2 e 5 é idêntico às combinações formadas pelos grupos 1 e 2, pelas bombas 1 e 4 e pelas bombas 1 e 5;
- O indicador de eficiência do conjunto formado pelos grupos 1 e 5 é idêntico ao conjunto formado pelos grupos 2 e 4.

Com a finalidade de inferir o conjunto de bombas mais eficiente, comparou-se os gráficos de densidade de cada conjunto.

Na figura 4.23 apresenta-se os gráficos de densidade estimada para a variável caudal, para 2 grupos em funcionamento.

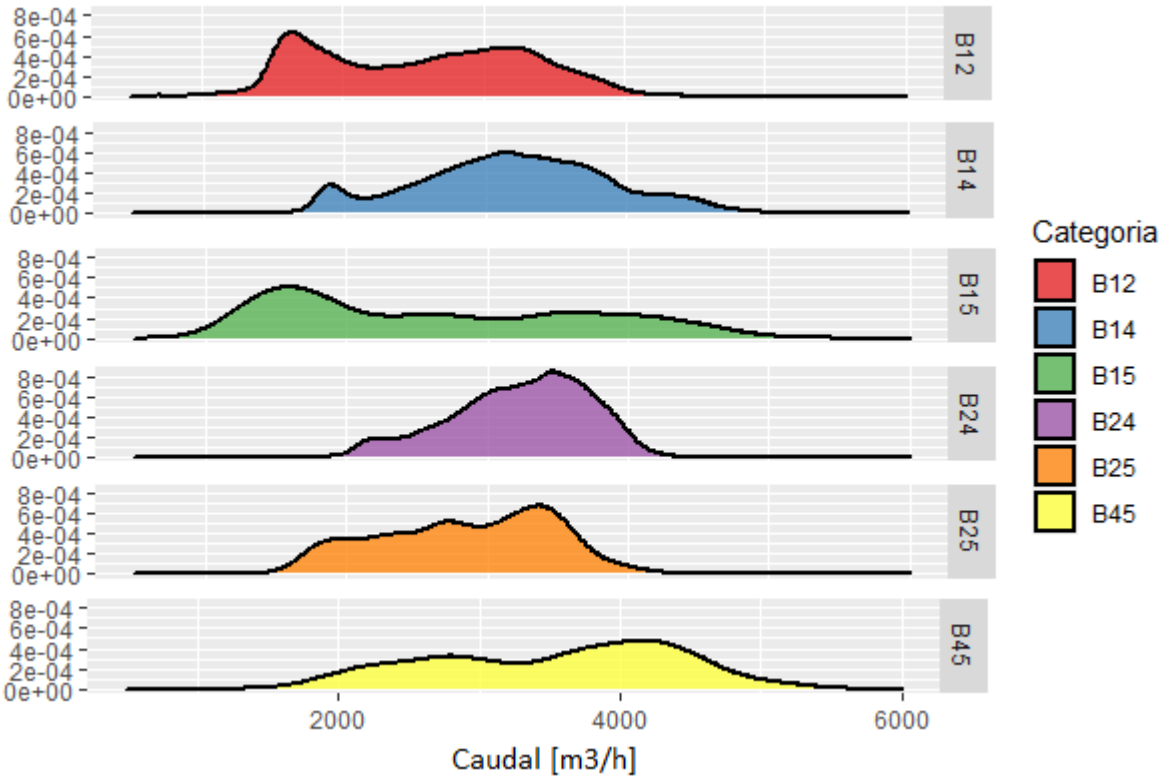

Figura 4.23: Gráficos de densidade estimada da variável caudal, para dois grupos em funcionamento.

Pelo gráfico da figura 4.23 observa-se que a probabilidade de ser elevado maiores valores de caudal é maior no conjunto formado pelos grupos 4 e 5, enquanto que o conjunto formado pelos grupos 1 e 5 é o que tem maior probabilidade de elevar caudais inferiores.

Na figura 4.24 representa-se os gráficos de densidade da variável potência para 2 grupos em funcionamento.

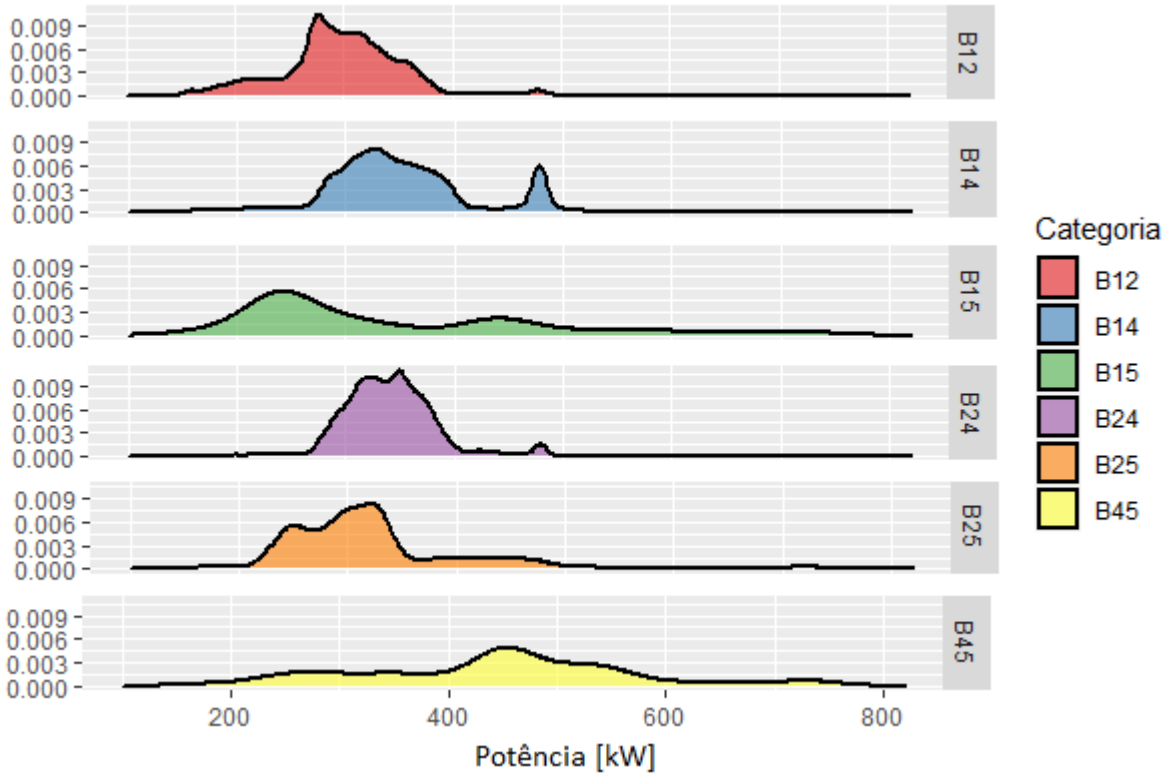

Figura 4.24: Gráficos de densidade estimada da variável potência, para dois grupos em funcionamento.

Com base no gráfico da figura 4.24 nota-se que a probabilidade estimada de fornecer potências inferiores é maior no conjunto com os grupos 1 e 5, contudo, juntamente com o conjunto formado pelos grupos 4 e 5, são os que fornecem superiores valores de potência. Os restantes conjuntos forneceram regularmente valores médios de potência.

Na figura 4.25 mostra-se os gráficos de densidade estimada da variável indicador de eficiência, para dois grupos em funcionamento.

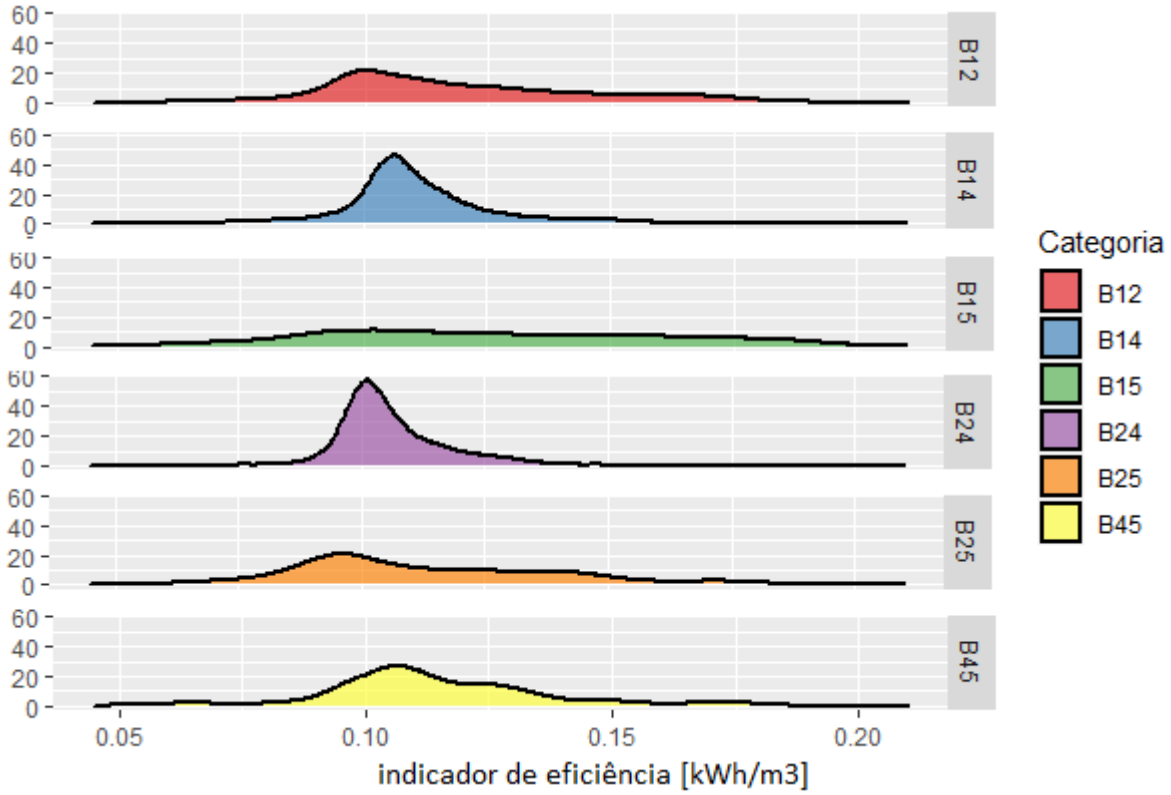

Figura 4.25: Gráficos de densidade estimada da variável indicador de eficiência, para dois grupos em funcionamento.

Através do gráfico da figura 4.25, observa-se que têm maior probabilidade de apresentar menores indicadores de eficiência, inferiores a 0.10 kWh/m^3, são as combinações formadas pelos grupos 2 e 4 e pelos grupos 2 e 5. Por outro lado, à semelhança das combinações formadas pelos grupos 1 e 2 e pelos grupos 1 e 5, a combinação dos grupos 2 e 5 estende com alguma regularidade os seus valores à direita para maiores indicadores de eficiência.

Com base na mediana e no 3^{o} quartil dos indicadores descritivos da tabela 4.8 para o caudal e na figura 4.23, constata-se que a combinação formada pelos grupos 2 e 4 é a que eleva caudais superiores, no entanto, com base na mediana e no 1^{o} quartil da tabela 4.10 e no gráfico da figura 4.25 é a que tem os menores indicadores de eficiência.

Dito isto e considerando que através do teste de comparação múltipla inferiu-se que os indicadores de eficiência do conjunto com as bombas 2 e 4 e do conjunto com as bombas 4 e 5 são idênticos, nos instantes em que estão em funcionamento duas bombas, o mais eficiente seria utilizar a combinação com o grupos eletrobomba 4 e 5 ou com os grupos 2 e 4.

4.4.3 Três grupos eletrobomba em funcionamento

- Caudal

De forma a comparar se a quantidade de água elevada pelos grupos eletrobomba, quando trabalham em grupos de 3, é idêntica ou não, fez-se um teste de Kruskal-Wallis:

Obteve-se os seguintes valores:

- **Valor observado da estatística de teste e valor-p**

  H_{obs}= 591.76 e valores-p < 0.001

- **Decisão e conclusão**

  Como o valores-p $< \alpha = 0.05$, rejeita-se a hipótese nula a este nível de significância, ou seja,conclui-se que a quantidade de água elevada, em média, não é idêntica para os 4 conjuntos de bombas.

- Potência

De seguida, verificou-se se a potência fornecida pelos conjuntos de grupos eletrobomba, nos momentos em que funcionam em grupos de 3, é idêntica para os quatro conjuntos.

Obteve-se os seguintes valores:

- **Valor observado da estatística de teste e valor-p**

  H_{obs}=814.92 e valores-p < 0.001

- **Decisão e conclusão**

  Como o valores-p $< \alpha = 0.05$, rejeita-se a hipótese nula a este nível de significância, ou seja, conclui-se que a potência fornecida pelas bombas, em média, não é idêntica para os quatro conjuntos.

- Indicador de eficiência

Depois, averiguou-se se a eficiência das bombas, nos instantes em que estão em fucuncionamento conjuntos de 3 grupos, é idêntica para todos os conjuntos.

Obteve-se os seguintes valores:

- **Valor observado da estatística de teste e valor-p**

  H_{obs}= 1133.5 e valores-p < 0.001

- **Decisão e conclusão**

  Como o valores-p $< \alpha = 0.05$, rejeita-se a hipótese nula a este nível de significância, ou seja,conclui-se que a eficiência não é idêntica para os 4 conjuntos de bombas.

Uma vez que se concluiu que os quatro grupos **não são idênticos** para todas as variáveis, fez-se um **teste de comparação múltipla de Bonferroni**, para averiguar quais grupos diferem entre si, obtendo os seguintes resultados:

Tabela 4.24: Testes de comparação múltipla com três grupos em funcionamento, para a variável caudal.

Caudal			
Bombas	**11010**	**11001**	**10011**
11001	< 0.001	-	-
10011	< 0.001	< 0.001	-
01011	< 0.001	< 0.001	< 0.001

Tabela 4.25: Testes de comparação múltipla com três grupos em funcionamento, para a variável potência.

Potência			
Bombas	**11010**	**11001**	**10011**
11001	< 0.001	-	-
10011	< 0.001	< 0.001	-
01011	0.052	< 0.001	< 0.001

Tabela 4.26: Testes de comparação múltipla com três grupos em funcionamento, para a variável indicador de eficiência.

Indicador de eficiência			
Bombas	**11010**	**11001**	**10011**
11001	< 0.001	-	-
10011	< 0.001	< 0.001	-
01011	0.0017	< 0.001	< 0.001

Com base nas tabelas 4.24, 4.26, a um nível de significância de 0.05 conclui-se que todas as combinações diferem entre si, enquanto que pela tabela 4.25, ao mesmo nível de significância, conclui-se que as combinações formadas pelos grupos 1, 2 e 4 e pelos grupos 2, 4 e 5 não apresentam diferenças com significado estatístico na potência fornecida.

Com o objetivo de inferir o conjunto de bombas mais eficiente, é necessário comparar as variáveis de cada conjunto. Para tal representou-se os seguintes gráficos de densidade de cada um dos conjuntos para as três variáveis.

Na figura 4.26 representa-se os os gráficos de densidade estimada da variável caudal, para três grupos em funcionamento.

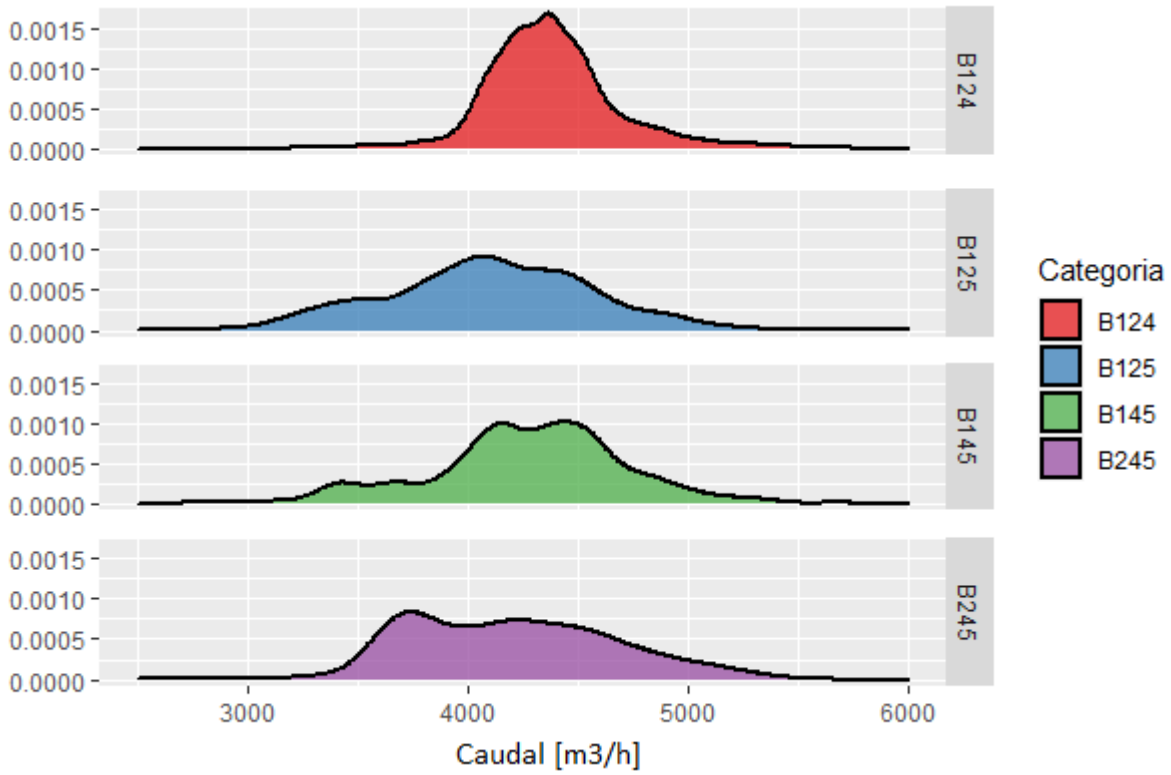

Figura 4.26: Gráficos de densidade estimada da variável caudal, para três grupos em funcionamento.

Com base no gráfico da figura 4.26 é possível observar que a probabilidade de elevar caudais superiores é maior na combinação formada pelos grupos 1, 2 e 4, enquanto que na combinação com os grupos 1, 2 e 5 a probabilidade de elevar caudais inferiores é maior.

Na figura 4.27 mostra-se os gráficos de densidade estimada da variável potência, para três grupos em funcionamento.

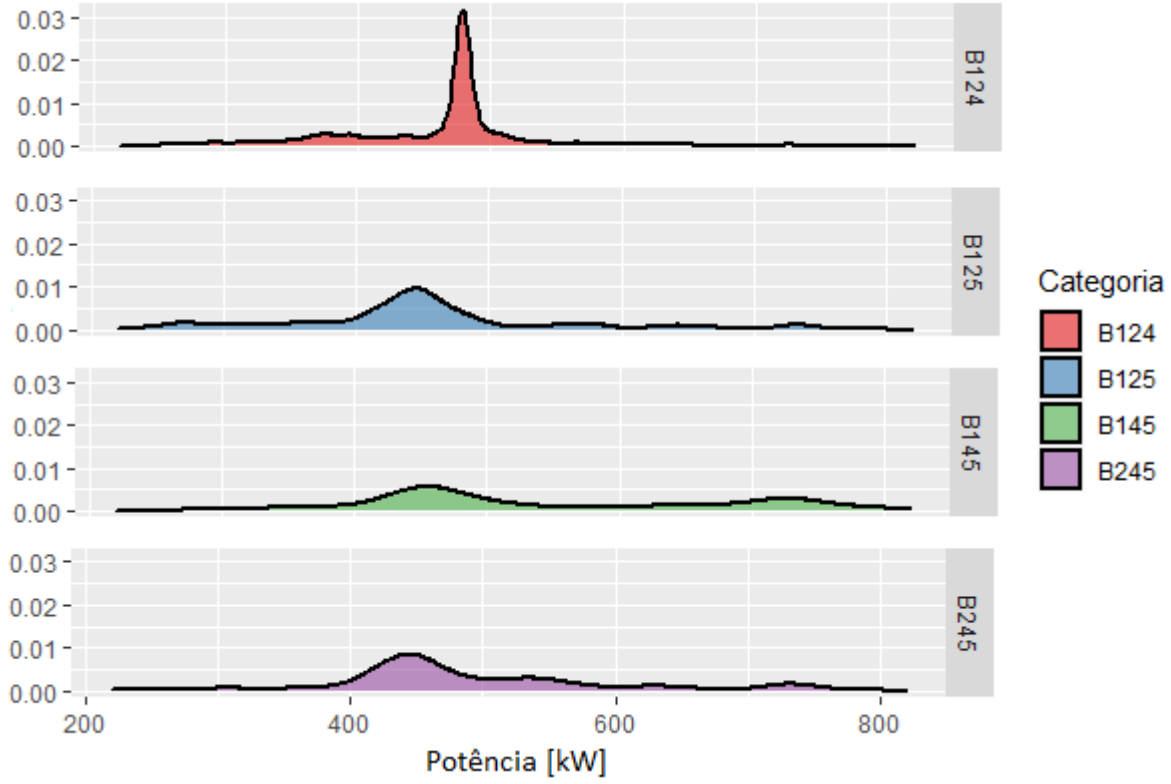

Figura 4.27: Gráficos de densidade estimada da variável potência, para três grupos em funcionamento.

Pelo gráfico da figura 4.27, observa-se que a combinação que fornece potências elevadas com maior frequência é a combinação composta pelos grupos 1, 4 e 5 e a que fornece potências mais baixas com maior frequência é a formada pelos grupos 1, 2 e 5.

Na figura 4.28 representa-se os gráficos de densidade estimada da variável indicador de eficiência, para três grupos em funcionamento.

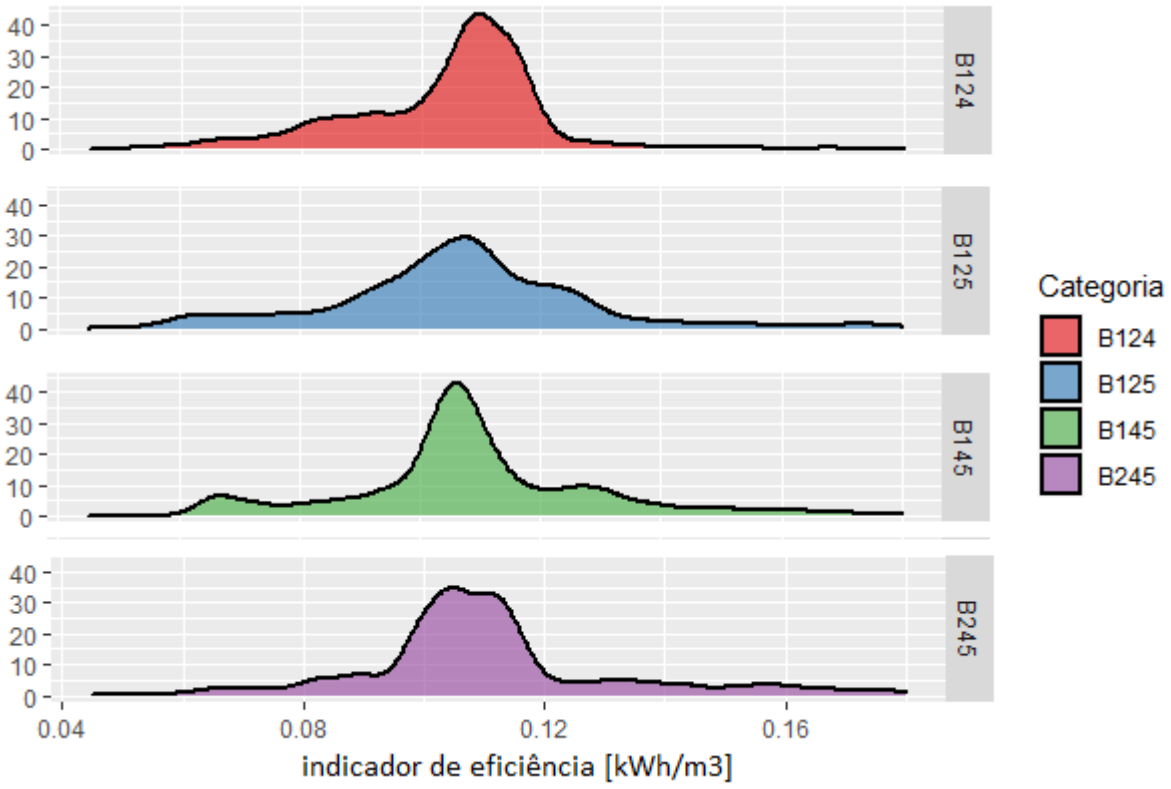

Figura 4.28: Gráficos de densidade estimada da variável indicador de eficiência, para três grupos em funcionamento.

Observando o gráfico da figura 4.28, nota-se que a probabilidade de se ter menores indicadores de eficiência é maior nos conjuntos formados pelos grupos 1, 2 e 5 e pelos 1, 4 e 5. Por outro lado, os conjuntos com os grupos 1, 4 e 5 e os grupos 2, 4 e 5 são os que têm maior probabilidade de se ter maiores indicadores de eficiência, enquanto que essa probabilidade é menor no conjunto com os grupos 1, 2 e 4.

Na tabela 4.15, relativa aos indicadores descritivos do indicador, observou-se que a combinação com os grupos 1, 2 e 5 tem a menor mediana e à semelhança da combinação com os grupos 1, 2 e 4, tem o menor 1^{o} quartil, sendo que a combinação dos grupos 1, 2 e 4 tem também o menor 3^{o} quartil.

Assim sendo, conclui-se que quando estão três grupos em funcionamento simultaneamente, o mais eficiente seria utilizar a combinação com os grupos eletrobomba 1, 2 e 5 ou com os grupos 1, 2 e 4.

Conclusões e Perspetivas futuras

Este estágio surgiu da necessidade de otimizar o funcionamento da EE 3 do subsistema de Alcântara, de modo a encontrar uma solução com menores consumos energéticos, neste contexto foram avaliadas as diferentes combinações de grupos eletrobomba, comparando a sua eficiência, com base em indicadores descritivos, no teste de Kruskal-Wallis e na estimativa de densidade Kernel.

Com base nos dados das variáveis, o caudal elevado, a potência fornecida e os indicadores de eficiência, para o período de tempo de Abril de 2017 a Setembro de 2018, foi efetuado um pre-processamento dos dados o qual permitiu obter uma base de dados com todas as variáveis sincronizadas no mesmo espaço temporal, e onde valores anómalos foram identificados, e mesmo eliminados, quando correspondiam a erros do sistema de recolha de dados.

Os indicadores descritivos, o teste de K-W e as estimativas kernel da densidade para as variáveis em estudo, permitiram concluir que quando os grupos eletrobomba funcionam individualmente, o mais eficiente é usar o grupo 5 seguido do grupo eletrobomba 1. Quando os grupos funcionam em conjuntos de dois, o mais eficiente é utilizar as combinações do grupo 4 com o grupo 2 ou a do grupo 4 com o grupo 5. Nos instantes em que funcionam 3 grupos ao mesmo tempo, as combinações mais eficientes são os conjuntos dos grupos 1, 2 e 4 e o conjunto dos grupos 1, 2 e 5. Nos momentos em que funcionam 4 grupos eletrobomba simultaneamente, só existe uma combinação.

Face aos resultados obtidos, conclui-se que se deve privilegiar combinações de grupos eletrobomba que incluam o grupo 5. No entanto, e apesar de o grupo eletrobomba 5 ter sido considerado o mais eficiente, foi o que teve menos tempo de funcionamento durante o período de análise. Este resultado pode estar associado ao facto deste grupo ter menos desgaste relativamente aos restantes grupos. De modo a obter maior fiabilidade neste estudo, seria relevante que as combinações de funcionamento tivessem uma frequência de funcionamento semelhante.

Para investigações futuras recomenda-se uma análise estatística com base na mesma metodologia, mas tendo em conta outros fatores que possam influenciar o caudal elevado, nomeadamente a precipitação e o nível de maré.

Referências

[1] Águas do Tejo Atlântico. (2017, 28 de abril). Contrato de Concessão. Água do Tejo Atlântico. Lisboa.

[2] Águas de Portugal. Águas de Portugal. Disponível a partir de https://www.adp.pt/en/

[3] Águas do Tejo Atlântico. Águas do Tejo Atlântico. Disponível a partir de https: /www.aguasdotejoatlantico.adp.pt/

[4] Águas do Tejo Atlântico. (2019, 29 de março). Relatório e Contas 2018. Água do Tejo Atlântico. Lisboa.

[5] Ferreira, C. (2020). *Análise de Clusters de interrupções de Funcionamento de Estações de Água Residual*. Relatório de Estágio, Águas do Tejo Atlântico - Instituto Superior de Engenharia de Lisboa. 97pp.

[6] H2TEC - Soluções Ambientais. H2TEC. Disponível a partir de https: /www.h2tec.pt/produtos/estacao-elevatoria-compacta/

[7] Aleixo, S. *Estatística Descritiva e Análise Exploratória Dados*. Instituto Superior de Engenharia de Lisboa.

[8] Aleixo, S. *Testes de hipóteses não paramétricos de localização central*. Instituto Superior de Engenharia de Lisboa.

[9] Venables, Ripley (2002). *Modern Applied Statistics with S*.

[10] SimTejo. *Empreitada de Concepção, Fornecimento e Montagem do Equipamento Electromecânico, Metalo-Mecânico e Instalações Eléctricas para a Estação Elevatória de Águas Residuais EE3 do Sistema Interceptor Algés-Alcântara.*