Deteção de alterações nos padrões de consumo de energia em instalações de média tensão

por

Aline Pereira dos Santos

Dissertação de Mestrado em Modelação, Análise de Dados e Sistemas de Apoio à Decisão

Orientado por:

Professor João Gama, Faculdade de Economia do Porto

Professora Rita Ribeiro, Faculdade de Ciências do Porto

Dra. Susana Magalhães, EDP Distribuição - Energia, S.A.

Set/2016

Nota biográfica do autor

Aline Pereira dos Santos concluiu a sua licenciatura em Gestão, pela Faculdade de Economia da Universidade do Porto (FEP), em Julho de 2009. Em Dezembro do mesmo ano, começou o seu percurso profissional na EDP Valor - Gestão Integrada de Serviços, S.A., trabalhando na área de contabilidade e reporte financeiro desde então.

Com o objetivo de desenvolver novas competências e tendo observado uma crescente procura no mercado por profissionais capazes de trabalhar de forma eficiente com grandes bases de dados, em 2014, a autora deste trabalho decidiu retornar à FEP e inscrever-se no Mestrado em Modelação, Análise de Dados e Sistemas de Apoio à Decisão.

Agradecimentos

A conclusão deste projeto não seria possível sem o apoio de diversas pessoas que tornaram este percurso mais interessante e estimulante. Neste sentido, gostaria de começar por agradecer aos meus orientadores, Professor João Gama, Professora Rita Ribeiro, Dra. Susana Magalhães e Engenheiro Mário Lemos que se mostraram sempre disponíveis para esclarecer as dúvidas que foram surgindo, assim como me obrigaram a ir sempre mais além, apresentando novos desafios e questões.

Aproveito também para agradecer à Dra. Alice Jorge por todo o apoio e carinho ao longo destes anos em que trabalhamos juntas, por me incentivar a continuar a estudar e evoluir e por ter sempre um sorriso e uma palavra de apoio para dar.

Gostaria ainda de agradecer à minha família e aos meus amigos, que souberam compreender os momentos de ausência devido ao tempo necessário para concluir este projeto e por ajudarem a relembrar que a vida não é feita apenas de responsabilidades profissionais e académicas.

Agradeço em especial aos meus pais, Célia Márcia e Arsenio Santos, pelo apoio e amor incondicional e principalmente pelo exemplo de vida. Obrigada por me ensinarem a seguir os meus sonhos e a nunca desistir, mostrando que sou capaz de atingir todos os meus objetivos.

Também gostaria de agradecer ao meu irmão, Felipe Santos, pelo incentivo e pelas visitas para saber se eu continuava viva. Muito obrigada por acreditares em mim e por fazeres parte do meu dia-a-dia.

Por último mas de modo algum menos importante, gostaria de agradecer ao meu marido, Manuel Parente, por me apresentar o mundo das ferramentas de data mining e otimização e por estar sempre pronto para novas aventuras. Para além de todo o apoio, carinho e imensa paciência para ouvir sobre séries temporais e identificação de padrões, foste sempre o meu porto seguro e um dos motivos para ser tão feliz e realizada.

Resumo

O consumo energético apresenta-se como um tema de grande interesse para a sociedade atual devido a questões relacionadas com a sustentabilidade, por exemplo. Contudo, analisar e compreender os padrões de consumo assume uma importância ainda maior para empresas como a EDP Distribuição - Energia, S.A. (EDPD), visto esta informação poder auxiliar na tomada de decisões diária.

Neste sentido, o presente trabalho pretende desenvolver um estudo sobre o consumo de energia elétrica em instalações de média tensão, de modo a identificar os padrões frequentes e os padrões anómalos verificados num conjunto de dados, para um determinado horizonte temporal. A identificação dos padrões frequentes permite compreender o que é normal e, consequentemente, esperado para uma determinada instalação, enquanto os padrões anómalos permitem identificar situações extraordinárias que podem ocorrer devido a avarias de equipamentos, alterações da capacidade produtiva, erros da equipa técnica ou, no limite, fraudes.

De modo a atingir os objetivos definidos e contando com a colaboração da EDPD, serão analisadas algumas séries temporais disponibilizadas, que apresentam apenas informação sobre a potência média registada, em intervalos de 15 minutos, num determinado período de tempo. Com base nesta informação, serão aplicados os métodos considerados mais indicados a fim de identificar os padrões existentes. Os padrões identificados serão analisados e validados de modo a compreender a veracidade e importância do conhecimento extraído.

Assim sendo, apresenta-se uma análise exploratória dos dados, com enfoque no potencial existente nas bases de dados extraídas pela EDPD no decorrer da sua atividade, demonstrando as capacidades e vantagens das ferramentas de extração de conhecimento útil dos dados na matéria em estudo.

Palavras-chave: séries temporais, consumo de energia, identificação de padrões frequentes, identificação de padrões anómalos, análise exploratória dos dados

Abstract

Energy consumption is a relevant issue in contemporary society, especially but not exclusively from a sustainability point of view. However, for large companies such as EDP Distribuição - Energia, S.A. (EDPD), its importance is even greater in the context of analysing and understanding consumption patterns, as this kind of information is paramount to support the daily decision making process.

As such, the current project aims to study the consumption of electrical energy in medium voltage buildings, in order to detect frequent and anomalous patterns verified in the available data for a specified time period. On the one hand, the identification of frequent patterns allows for the understanding of regular situations, providing insight concerning the normal functioning of these buildings. On the other hand, the detection of anomalous patterns identifies extraordinary circumstances that may occur due to equipment malfunction, alterations in production capacity, technical team errors, or even fraud.

In order to realize the established goals, the analysis of some temporal series will be performed. These were attained through the collaboration with EDPD, and contain information exclusively concerning the average power consumption, divided into 15 minutes intervals, of some medium voltage buildings over a particular period of time. Leveraging on the availability of these temporal series, the most appropriate methods for identifying the current patterns will be considered and applied. In the final stages of the process, the detected patterns will be analysed and validated, in an attempt to assess the relevance of the obtained knowledge.

Therefore, this project comprises an exploratory data analysis, focused on the potential of the databases collected by EDPD in the normal course of its activity. In addition, it demonstrates the capabilities and advantages of using data mining tools in this type of studies.

Keywords: temporal series, energy consumption, frequent pattern identification, anomalous pattern identification, exploratory data analysis

Índice

Índice de figuras

Capítulo 1. Introdução

Ao longo do presente capítulo serão introduzidos alguns aspetos a ter em consideração relativamente a esta dissertação, nomeadamente, a motivação para a escolha do tema a estudar, os principais objetivos propostos e a descrição do problema em causa. Para além disso, faz-se referência à estrutura definida para este trabalho, realizando uma breve descrição do que será apresentado em cada capítulo.

1.1. Motivação

A energia é essencial para o desenvolvimento da sociedade ao representar um recurso básico em quase todos os processos produtivos (Amador 2010). Neste sentido, a análise do mercado energético, desde a sua produção até ao consumo, revela-se como um tema de elevado interesse para todos os intervenientes.

Para uma empresa como a EDP - Energias de Portugal, S.A., presente na produção, distribuição e comercialização de energia, obter e analisar informação associada ao mercado energético torna-se fundamental para o seu desenvolvimento. Tendo em atenção o negócio da distribuição de eletricidade, a cargo da EDP Distribuição - Energia, S.A. (EDPD), é possível afirmar que uma análise detalhada dos consumos permite um melhor planeamento e uma gestão mais eficaz da rede.

A EDPD, através dos dados recolhidos nos pontos de medição, consegue obter uma quantidade significativa de informação sobre os consumos dos seus clientes, da qual é possível extrair conhecimento útil para o desenvolvimento da sua atividade. A análise dos perfis de consumo dos seus clientes representa uma área de extrema importância para a empresa, sendo essencial compreender os padrões de consumo existentes e detetar possíveis alterações que possam ocorrer nos consumos ao longo do tempo.

Considerando a existência e o acesso a uma base de dados com potencial para ser utilizado para esse fim, esta tese foca-se, em termos gerais, na utilização de um método capaz de analisar os padrões de consumo para um conjunto de instalações e detetar as alterações ocorridas nestes mesmos padrões.

1.2. Objetivos

Como referido anteriormente, com base na informação disponível, o objetivo deste trabalho consiste em detetar alterações nos padrões de consumo em instalações de média tensão. Assim, pretende-se desenvolver um método capaz de analisar os dados de consumo de diferentes instalações de modo a identificar os padrões frequentes e anómalos existentes. Os padrões frequentes permitem compreender o consumo normal esperado de uma dada instalação, enquanto os padrões anómalos possibilitam a identificação de situações fora do normal como fraudes, desperdícios, alterações da capacidade produtiva, avaria de equipamentos ou erros da equipa técnica.

De modo a atingir este objetivo serão estudadas diferentes metodologias de extração de conhecimento de dados na área da deteção de padrões frequentes e anómalos. A EDPD apresenta um elevado número de clientes e, portanto, dispõe de uma quantidade de informação muito significativa. Por esta razão, os algoritmos a utilizar devem ser genéricos e funcionais, de modo a possibilitar a aplicação a diferentes instalações, exigindo um esforço computacional aceitável.

1.3. Definição do problema

Para a eficaz análise de qualquer problema é essencial começar por compreender as características do mesmo. Como a qualidade dos resultados obtidos em problemas relacionados com extração de conhecimento de dados é extremamente dependente da base de dados disponível, torna-se imperativo começar pela compreensão desta informação.

Os dados a utilizar neste trabalho foram disponibilizados pela EDPD e consistem em diagramas de carga que indicam a potência média de cada 15 minutos, registada ao longo do tempo, para diferentes instalações de média tensão. A necessária análise dos dados inclui a preparação da informação para posterior utilização, podendo implicar a estimativa de dados em falta ou a eliminação de informação considerada irrelevante ou duvidosa.

Dada a elevada quantidade de dados disponíveis através dos diagramas de carga, a análise gráfica destes, para além de apresentar uma componente subjetiva, apenas permite obter uma ideia geral sobre os padrões de consumo existente numa instalação, não permitindo obter informação detalhada sobre os mesmos. Por esta razão, foram desenvolvidos e apresentados na literatura diferentes métodos e algoritmos capazes de analisar os padrões

presentes em base de dados de elevada dimensão, como os diagramas de carga disponibilizados pela EDPD.

Neste contexto, o próximo capítulo faz o estudo e análise crítica da literatura existente afeta ao tema em estudo, sendo apresentados alguns métodos capazes de auxiliar na deteção de padrões frequentes e anómalos. Este estudo tem como objetivo selecionar os melhores métodos a aplicar à base de dados disponível, de modo a atingir os objetivos propostos para o presente trabalho.

1.4. Organização

O presente trabalho está dividido em cinco capítulos. No presente capítulo é realizada uma introdução onde são expostos o tema e a motivação para o desenvolvimento deste trabalho, assim como os objetivos definidos para o mesmo. De modo a enquadrar o trabalho a desenvolver nos estudos já realizados e publicados, o segundo capítulo pretende rever a literatura existente sobre o tema em estudo, apresentando em pormenor os métodos escolhidos para alcançar os objetivos propostos.

No terceiro capítulo será apresentado o problema em análise com maior detalhe, indicando algumas questões relacionadas com as bases de dados disponibilizadas para o efeito e a preparação necessária dos dados para a aplicação dos métodos escolhidos. Neste capítulo, serão ainda expostos em detalhe os processos realizados para a deteção de padrões frequentes e anómalos, descrevendo os diferentes passos de cada método.

Os resultados obtidos, nomeadamente os padrões frequentes e anómalos identificados, serão expostos no quarto capítulo. Em primeiro lugar serão apresentados os resultados considerando apenas os dados inicialmente disponibilizados, nomeadamente a potência média registada e o respetivo horizonte temporal. Contudo, no final do capítulo, será apresentada uma agregação da informação obtida para os padrões frequentes e anómalos com alguma informação adicional disponibilizada pela EDPD, de modo a auxiliar na interpretação dos resultados obtidos.

O quinto capítulo apresenta uma breve conclusão do trabalho desenvolvido, permitindo compreender o contributo deste trabalho. Adicionalmente, serão indicadas algumas limitações encontradas e potenciais trabalhos futuros a desenvolver.

Capítulo 2. Revisão de literatura

Tendo em vista os objetivos propostos no capítulo anterior, procedeu-se à pesquisa e ao estudo da literatura existente sobre o tema em análise. Assim, o presente capítulo começa por apresentar alguns conceitos e técnicas associados às séries temporais, referindo de seguida alguns métodos encontrados na literatura para a deteção de padrões frequentes e anómalos. Apesar da quantidade e variedade de métodos existentes, ao longo deste capítulo serão citados apenas aqueles que se considera com maior relevância para o problema em causa. Neste sentido, apenas os métodos utilizados nesta dissertação serão expostos de forma mais pormenorizada, como por exemplo a representação simbólica da série temporal através do SAX, conforme exposto de seguida. Dada a importância de avaliar a qualidade dos resultados obtidos através da aplicação dos algoritmos, serão apresentados ainda alguns métodos que poderão ser utilizados para suprir esta necessidade de validação dos modelos.

2.1. Séries temporais e técnicas associadas

O problema em estudo consiste na análise de diagramas de carga que registam a potência média observada em instalações de média tensão em intervalos consecutivos de 15 minutos, permitindo obter um elevado volume de dados ordenados cronologicamente. Por essa razão, estes devem ser tratados como séries temporais, uma vez que estas são definidas como uma sequência de observações cujos elementos apresentam uma determinada sequência cronológica (Li et al. 2013).

Considerando a notação usual, uma série temporal $T = t_1, \dots, t_n$ consiste num conjunto ordenado de *n* valores reais. Dado o elevado número de observações presentes numa série temporal, pode ser necessário analisar partes desta em detrimento de uma análise global, sendo assim utilizadas subsequências. Dada uma série temporal *T* de tamanho *n*, uma subsequência *S* pertencente a *T* consiste numa amostra de tamanho $m < n$ de elementos contíguos de *T*, tais que $S = t_p, \dots, t_{p+m-1}$ para $1 \leq p \leq n\text{-}m\text{+}1$.

As características associadas às séries temporais incluem a possível atualização constante da informação com o decorrer do tempo e a elevada quantidade de dados, que apresenta uma natureza numérica e contínua, devendo ser considerada como um todo e não apenas como um conjunto de elementos numéricos independentes (Fu 2011).

Devido à possível elevada dimensão das séries temporais, obter uma boa representação das mesmas pode ser referido como um problema fundamental a solucionar. De modo a extrair eficazmente conhecimento das bases de dados disponíveis, o método a utilizar para reduzir o número de observações de uma série temporal apresenta-se como uma decisão de extrema importância.

Neste sentido, a primeira técnica encontrada na literatura para solucionar esta questão foi proposta por Agrawal et al. (1993), ao propor a aplicação da *Discrete Fourier Transform* (DFT) a uma série temporal. Este método recorre a ferramentas matemáticas complexas para transformar uma sequência de uma série temporal num único ponto, com base na sua frequência. Como consequência deste agrupamento de dados, com o potencial de reduzir o número de leituras de uma série temporal extensa, é facilitada a gestão e análise da informação existente.

De facto, de acordo com Keogh, Chakrabarti, Pazzani, et al. (2001), reduzir o número de observações de uma série temporal definindo um método para indexar os valores obtidos na série temporal original consiste na abordagem mais promissora encontrada na literatura. Assim sendo, estes autores propõem uma técnica denominada *Piecewise Aggregate Approximation* (PAA), que pressupõe a segmentação da série temporal em frações de igual dimensão, sendo calculado e registado o valor médio de cada fração. Deste modo, uma série temporal C de tamanho n pode ser representada num espaço de dimensão w através do vetor $\bar{C} = \bar{c}_1, \ldots, \bar{c}_w$, sendo $1 \leq w \leq n$. Os elementos de $\bar{C}$ podem ser obtidos através da Equação 3.1:

$$\bar{c}_i = \frac{w}{n} \sum_{j=\frac{n}{w}(i-1)+1}^{\frac{n}{w}i} c_j \tag{3.1}$$

Gera-se assim um vetor com os valores médios dos w segmentos considerados, o que constitui a representação reduzida dos dados originais. A técnica de PAA tem como vantagem ser simples de compreender e de implementar e exigir menos esforço computacional do que outros métodos. Para além disso, permite a utilização de diferentes medidas para o cálculo das distâncias, sendo possível construir um gráfico linear para a representação dos dados reduzidos. A Figura 2.1 permite visualizar um exemplo de uma

série temporal (C) com 128 observações, que através da técnica de PAA é reduzida para 8 dimensões ($\bar{C}$):

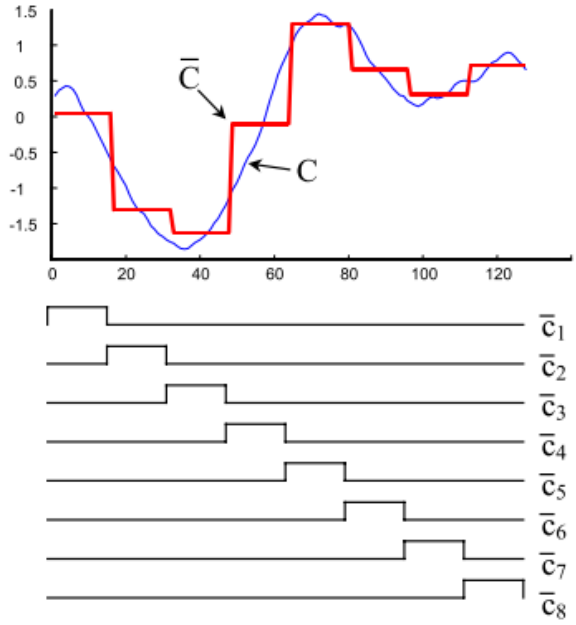


Figura 2.1: Representação gráfica de uma série temporal original C e reduzida $\bar{C}$ (extraída de Lin et al., 2002)

Apesar das vantagens apresentadas para a técnica de PAA, esta pressupõe que todos os segmentos apresentem a mesma dimensão. Neste sentido, Keogh, Chakrabarti, Mehrotra, et al. (2001) propõem uma nova técnica intitulada de *Adaptative Piecewise Constant Approximation* (APCA) que permite considerar segmentos de diferentes dimensões, sendo necessário registar não apenas o valor médio das observações pertencentes ao segmento em questão, mas também o último ponto deste segmento. Neste sentido, uma série temporal *C* de tamanho *n* será representada como:

$$C = \{< cv_1, cr_1 >, \dots, < cv_M, cr_M >\}, \quad cr_0 = 0 \tag{3.2}$$

onde *M* representa o número de segmentos, cv_i regista o valor médio das observações no segmento *i* e cr_i indica o último ponto deste segmento na série temporal.

O tamanho de cada segmento pode ser então calculado através da diferença $cr_i - cr_{i-1}$. Este método consegue aumentar a qualidade da aproximação comparativamente à técnica de PAA ao ser possível considerar segmentos de menor ou maior dimensão, conforme a estrutura dos dados em análise, ou seja, é possível considerar poucos segmentos para regiões com pouca atividade e muitos segmentos para regiões com muita atividade.

Independentemente do método escolhido para a redução do número de observações de uma série temporal, outro problema importante a ter em atenção consiste na capacidade de encontrar e representar subsequências de uma série temporal (Antunes & Oliveira 2001). Neste contexto, o método mais comum para encontrar subsequências envolve a utilização de uma janela deslizante (do inglês, *sliding window*) com uma dimensão *z*, que será posicionada em cada possível posição da série temporal. Deste modo, cada uma das janelas obtidas define uma subsequência composta pelos elementos presentes dentro desta mesma janela.

2.2. Padrões frequentes: *Motifs*

No decorrer da análise de uma série temporal, identificar e compreender os padrões frequentes existentes na base de dados pode auxiliar na compreensão da informação disponível. A identificação de padrões frequentes, desconhecidos à partida, indica que existe um conjunto de séries temporais ou subsequências de uma longa série temporal que são muito semelhantes entre si, recebendo estes conjuntos o nome de *motifs* (Lin et al. 2002).

Os *motifs*, ao sumarizar a informação disponível, facilitam a visualização da base de dados e possibilitam a utilização desta informação em tarefas posteriores. Na literatura foram encontrados diferentes métodos capazes de identificar *motifs* sendo possível citar alguns destes métodos, uma vez que existe potencialmente interesse em aprofundar ou explorar diferentes opções.

Existem diversas alternativas, tais como o método proposto por Vespier et al. (2013) para descobrir as características e possíveis sobreposições de *motifs* em múltiplas escalas de tempo com base na observação de alterações sistemáticas. A consideração das frequências observadas de um item numa determinada janela deslizante, cujo tamanho pode ser definido de forma dinâmica, é proposto por Thanh Lam & Calders (2010). Este método apresenta a vantagem de não exigir a definição de nenhum parâmetro específico, tal como o tamanho da janela deslizante. A utilização de um algoritmo exato para a descoberta de *motifs* em séries temporais é apresentada por Mueen et al. (2009), que afirmam que este algoritmo consegue ser muito competitivo em termos de esforço computacional. Por outro lado, Minnen et al. (2007) sugerem um método para encontrar *motifs* multivariados utilizando a estimativa da densidade das subsequências. Também é proposto um método

probabilístico para a descoberta de *motifs* em séries temporais (Chiu et al. 2003), cuja probabilidade de efetivamente detetar os *motifs* existentes apresenta-se bastante elevada perante a existência de ruído na base de dados.

O algoritmo proposto por Lin et al. (2002), intitulado de *Enumeration of Motifs through Matrix Approximation* (EMMA), apresenta um elevado interesse de aplicação para o problema em causa. Neste algoritmo e de modo a identificar os *motifs* existentes são consideradas as subsequências presentes numa série temporal, determinando o conjunto de subsequências que são semelhantes entre si. A semelhança entre subsequências é definida com base no cálculo das distâncias entre as subsequências em análise, sendo proposta a utilização da distância Euclidiana para este fim, por exigir uma menor capacidade de armazenamento de dados e menor esforço computacional para o seu cálculo. De referir que dadas duas subsequências *A* e *B* com a mesma dimensão *n*, a distância euclidiana pode ser calculada através da Equação 3.3:

$$D(A,B) = \sqrt{\sum_{i=1}^{n}(a_i - b_i)^2} \tag{3.3}$$

A principal vantagem apontada ao método em questão consiste na rapidez de obtenção de respostas. Contudo, o algoritmo EMMA obriga a definição de um parâmetro *R*, que representa o limite abaixo do qual duas subsequências serão consideradas compatíveis ou seja, se a distância calculada entre duas subsequências for igual ou inferior ao parâmetro *R,* estas subsequências serão consideradas compatíveis.

No que diz respeito às subsequências compatíveis, deve ser tida em atenção a existência de combinações triviais (do inglês, *trivial matches*), que consistem em subsequências pertencentes a janelas deslizantes adjacentes, diferindo o seu início ou fim poucas unidades para a direita ou para a esquerda, e que, por este motivo, tendem a ser compatíveis devido à proximidade observada. A Figura 2.2 pretende ilustrar o conceito de combinação trivial ao apresentar uma subsequência *C* e duas subsequências próximas desta que podem ser consideradas compatíveis com esta mas que, por estarem muito próximas da subsequência *C*, recebem o nome de combinação trivial, não devendo ser consideradas para a definição dos *motifs* de uma série temporal.

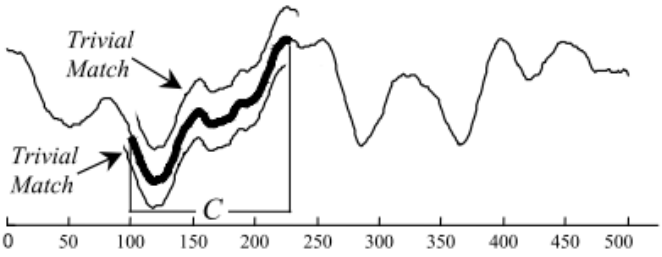


Figura 2.2: Representação gráfica de combinações triviais (extraída de Lin et al. 2002*)*

Apesar do problema em análise não apresentar o risco de admitir combinações triviais, pelo facto de serem realizadas análises diárias e, consequentemente, serem utilizadas janelas deslizantes sequenciais e não sobrepostas, a definição de um parâmetro R pode ser considerada uma desvantagem.

De modo a ultrapassar a questão da imposição deste parâmetro, optou-se por recorrer ao algoritmo *MrMotif (Multiresolution Motif Discovery in Time Series)*, apresentado por Castro & Azevedo (2010). Para a aplicação do algoritmo *MrMotif* a base de dados deve ser, em primeiro lugar, normalizada de modo a apresentar média zero e desvio padrão um. Este processo tem por base o estudo apresentado por Keogh & Kasetty (2003), que refere que comparar séries temporais com amplitudes muito diferentes não apresenta resultados relevantes, não sendo possível garantir as semelhanças ou dissemelhanças entre as séries temporais em análise. Uma vez normalizada a série temporal, devem ser realizados dois passos intermédios que consistem na redução da numerosidade, aplicando a técnica de PAA conforme apresentado na Secção 2.1., e na discretização, que implica a transformação da base de dados reduzida num conjunto de α símbolos com igual probabilidade entre si. De modo a realizar esta transformação, o algoritmo *MrMotif* recorre à representação simbólica conhecida na literatura por *indexable Symbolic Aggregate approXimation (iSAX)*, apresentada por Shieh & Keogh (2008). Esta representação consiste numa extensão da representação simbólica de séries temporais conhecida por SAX (*Symbolic Aggregate approXimation*) e apresentado por Lin et al. (2003).

Para a aplicação do SAX é necessário definir o número de símbolos a utilizar, representados por α, de modo a encontrar os pontos de corte (do inglês, *breakpoints*) na base de dados reduzida. Estes pontos consistem numa lista ordenada de números $B = \beta_1, \dots, \beta_{\alpha-1}$ tais que a área sob a curva da distribuição normal N(0,1) de β_i à β_{i+1} é igual

a 1/α. Assim sendo, os pontos de corte podem ser obtidos através de tabelas estatísticas, conforme apresentado na Figura 2.3:

β_i \ α	3	4	5	6	7	8	9	10
β_1	-0.43	-0.67	-0.84	-0.97	-1.07	-1.15	-1.22	-1.28
β_2	0.43	0.00	-0.25	-0.43	-0.57	-0.67	-0.76	-0.84
β_3		0.67	0.25	0.00	-0.18	-0.32	-0.43	-0.52
β_4			0.84	0.43	0.18	0.00	-0.14	-0.25
β_5				0.97	0.57	0.32	0.14	0.00
β_6					1.07	0.67	0.43	0.25
β_7						1.15	0.76	0.52
β_8							1.22	0.84
β_9								1.28

Figura 2.3: Tabela com os pontos de corte que dividem uma distribuição normal em um número arbitrário de regiões com igual probabilidade (adaptada de Lin et al., 2002)

Ao comparar os valores da base de dados reduzida com os pontos de corte a considerar, cada observação da base de dados reduzida passa a ser representada por um determinado símbolo. A concatenação destes símbolos faz com que cada subsequência passe a ser representada por uma palavra (*word*) que irá ser utilizada para comparar as semelhanças entre as subsequências da série temporal e, consequentemente, identificar os *motifs.*

O benefício da extensão conhecida por *iSAX* consiste no facto de não ser necessário definir o número de símbolos a utilizar, na medida em que este algoritmo considera a possibilidade de representar uma subsequência assumindo diferentes resoluções. As resoluções consistem no número de símbolos a utilizar, sendo possível considerar palavras com 2, 4, 8, 16, 32 ou 64 resoluções. Os respetivos pontos de corte são definidos seguindo o princípio apresentado acima para o algoritmo SAX, sendo necessário recorrer a tabelas estatísticas para obter os valores necessários.

A Figura 2.4 pretende demonstrar o processo de transformação de uma subsequência numa palavra assumindo duas resoluções diferentes. Assim, os gráficos representam uma subsequência com 128 observações, dividida em 8 segmentos de igual dimensão e onde cada segmento é representado por um determinado símbolo de acordo com a resolução e os respetivos pontos de corte a considerar. No gráfico a) é considerada uma resolução de 4 e a representação da palavra consiste em 1,2,3,2,1,0,1,1[4] enquanto no gráfico b) a

resolução escolhida é de 16 e a palavra é apresentada como 5,11,15,11,6,1,6,6[16]. Ambas as representações são consideradas pelo algoritmo como sendo uma mesma palavra, apenas apresentando resoluções diferentes.

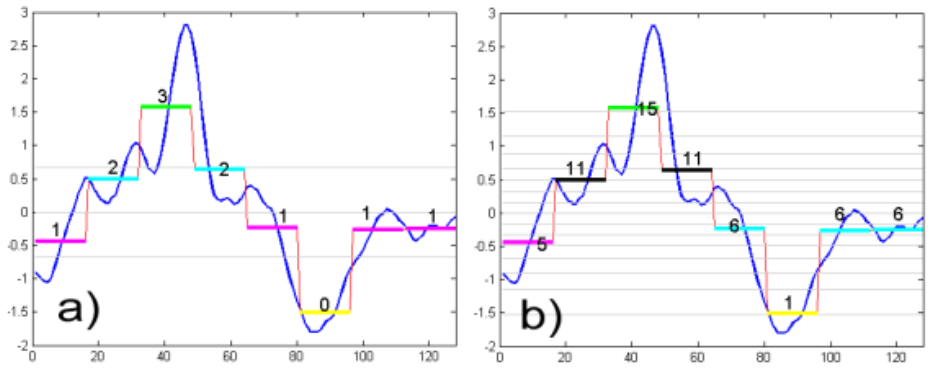


Figura 2.4: Exemplo da transformação de uma série temporal com 128 observações numa palavra com 8 símbolos para duas resoluções diferentes (extraído de Castro & Azevedo (2010)*)*

A possibilidade de representar a mesma palavra através de diferentes resoluções permite converter uma palavra com maior resolução numa palavra com menor resolução. Esta conversão pode ser facilmente realizada ao considerar a representação binária dos símbolos que compõem as palavras e ao dividir pela metade a resolução inicial. Contudo a conversão inversa, de uma palavra com menor resolução para uma palavra com maior resolução, não será tão linear pois existem diversas possibilidades ao aumentar a resolução considerada.

A deteção de *motifs* através do algoritmo *MrMotif* consiste na contagem do número de vezes que cada palavra ocorre e na consequente identificação daquelas que ocorrem mais vezes ao longo da base de dados. Estas serão aquelas que permitem compreender o normal comportamento da série temporal em análise, sendo identificadas como *motifs*. De referir que ao comparar duas séries temporais, quanto maior a resolução escolhida para representá-las, mais semelhantes estas deverão ser para que o algoritmo represente ambas através da mesma palavra. Assim sendo, uma palavra com maior resolução tende a apresentar um menor número de ocorrências. No entanto, a semelhança entre as séries temporais ou subsequências que a compõem tende a ser muito maior.

Deste modo, dada a existência de uma ferramenta gratuita e de fácil acesso para a aplicação do algoritmo apresentado e dadas as características da informação disponibilizada pela EDPD, considerou-se que a aplicação do algoritmo *MrMotif* seria uma mais-valia para os intervenientes no processo, tendo em atenção que o estudo

realizado pode vir a ser utilizado posteriormente pela empresa. De referir que uma das vantagens apontadas para este algoritmo é a possibilidade de utilizar este tipo de representação simbólica em bases de dados de fluxo contínuo (*streaming data*).

2.3. Padrões anómalos: *Discords*

Como referido anteriormente, na análise de uma série temporal torna-se relevante observar a existência de padrões frequentes de modo a verificar o normal comportamento da base de dados em estudo. Contudo, encontrar os padrões anómalos existentes numa série temporal permite identificar situações dissemelhantes que podem ser de extrema importância na análise de um conjunto de dados. Neste contexto os padrões anómalos em séries temporais recebem o nome de *discords* e consistem nas subsequências menos semelhantes face às restantes subsequências da série temporal (Fu et al. 2006).

Para a deteção de *discords* são propostos diversos métodos na literatura, sendo possível citar alguns destes, facilitando pesquisas futuras, caso haja o interesse em conhecer e explorar outros métodos. Deste modo, Li et al. (2013) propõe um algoritmo que recorre à representação binária da informação, comparando e registando a variação observada entre os valores médios, calculados através da técnica de PAA, para cada segmento (1 ou 0, caso o segmento apresente um valor médio acima ou abaixo, respetivamente, do seu antecessor). Com base nesta informação, pode-se visualizar as tendências existentes na série temporal, sendo então aplicadas técnicas de agrupamento (*clustering*) a fim de reunir as subsequências que apresentem os mesmos padrões, evidenciando assim os *discords* existentes na base de dados em análise.

Devido à elevada quantidade de informação que algumas séries temporais apresentam, Yankov et al. (2008) apresentam um algoritmo exato, capaz de lidar com estas séries temporais com um menor esforço computacional ao exigir apenas duas buscas lineares de informação. Este algoritmo seleciona os candidatos a *discord*, identificando de seguida os valores anómalos presentes neste conjunto de dados.

Outro algoritmo encontrado para a deteção de *discords* consiste no algoritmo WAT (*wavelet and augmented trie*) proposto por Bu et al. (2007). Este método pretende descobrir não apenas as subsequências mais incomuns, mas principalmente o *top-k discords* existentes na série temporal em análise, recorrendo para tal à transformada de

Haar Wavelet para aproximar a série temporal e representar, de forma geral, a sequência temporal das observações.

Apesar da existência de diversos métodos para detetar eficazmente *discords*, para o desenvolvimento do presente trabalho optou-se por utilizar o método proposto por Keogh et al. (2005) e denominado por HOT SAX. A escolha deste método teve por base a evidência de que este método já foi testado em séries temporais relativas ao consumo de energia, apresentando bons resultados. Além disso, este método recorre a algumas tarefas semelhantes às utilizadas no método escolhido para a deteção de *motifs*, nomeadamente nos processos de redução do número de observações e na representação simbólica dos dados, permitindo assim um melhor aproveitamento do tempo e do trabalho a executar.

Tal como na deteção de *motifs*, de modo a encontrar os *discords* numa série temporal devem ser excluídas as combinações triviais, logo torna-se relevante definir o conceito de *non-self match*. Assim sendo, Fu et al. (2006) referem que dada uma série temporal C, que contém uma subsequência T de tamanho n, com início na posição p e uma subsequência compatível M com início em q, pode-se afirmar que M é uma *non-self match* de T se $|p - q| \geq n$. Com base neste conceito, pode-se afirmar que dada uma série temporal C, a subsequência D de tamanho n com início na posição l é considerada *discord* de C se D apresentar a maior distância face ao seu *non-self match* mais próximo.

De acordo com a definição de *discord* apresentada, verifica-se que uma subsequência não pode ser candidata a *discord* se for possível encontrar uma qualquer subsequência na série temporal em análise cuja distância à atual candidata seja inferior à distância desta ao seu *non-self match* mais próximo. De referir que o cálculo das distâncias deve considerar a distância Euclidiana, apresentada anteriormente na Equação 3.3. Contudo, de acordo com Keogh et al. (2005), a série temporal deve ser normalizada, apresentando média zero e desvio padrão um, antes de serem calculadas estas distâncias, na medida em que este processo permite reduzir a redundância dos dados e a possibilidade destes serem inconsistentes.

Como referido, para a aplicação do algoritmo HOT SAX deve-se começar por reduzir o número de observações da série temporal, recorrendo à técnica de PAA, transformando de seguida os coeficientes obtidos em símbolos, através do algoritmo SAX. Dado que estes processos já foram apresentados nas secções anteriores, torna-se relevante

apresentar nesta fase a continuação do processo para a deteção de *discords* através deste método.

Deste modo, as subsequências já representadas por palavras são inicialmente registadas e ordenadas de acordo com a sua ocorrência temporal numa tabela (*array*). A informação assim organizada irá permitir a criação de uma árvore denominada por *augmented trie*. A Figura 2.5 permite visualizar o desenvolvimento deste método ao apresentar o exemplo de uma série temporal de tamanho *m* (quadro *raw time series*), da qual são extraídas subsequências de tamanho *n* e representadas por c_i, que são convertidas em palavras, $\hat{c}_i$. Estas palavras são então organizadas cronologicamente num *array*, conforme apresentado no lado esquerdo da Figura 2.5, cujo índice irá de 1 a *(m-n)+1*. De referir que a última coluna do *array* deve conter o número de vezes que uma determinada palavra é observada na série temporal. O lado direito da Figura 2.5 ilustra parte da árvore gerada, cujas folhas contêm a lista de todos os índices do *array* construído, que concorrem para a formação do respetivo nó terminal.

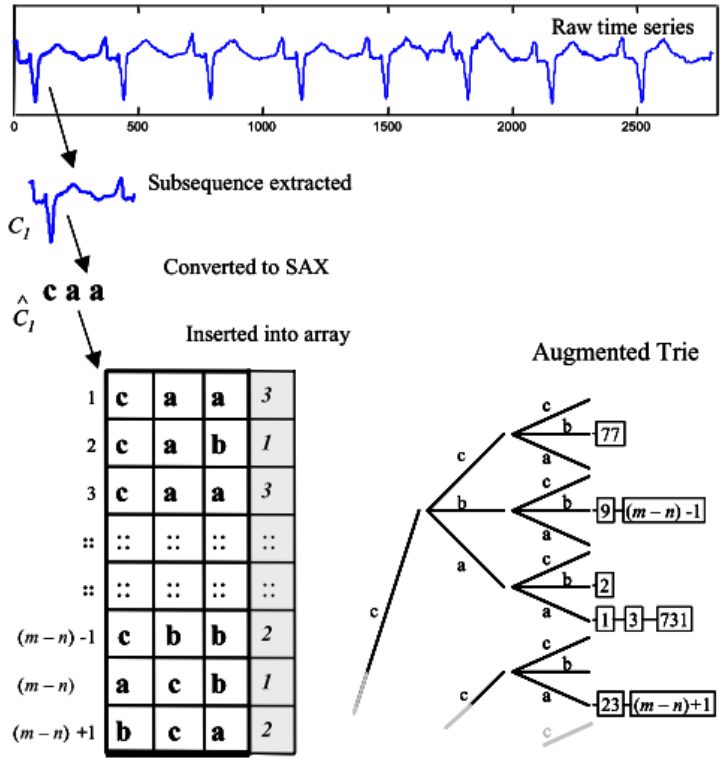


Figura 2.5: Ilustração do processo de outer loop (extraído de Keogh et al., 2005)

Para facilitar a compreensão da Figura 2.5 e, consequentemente, do método em estudo, é possível ilustrar alguns exemplos. Assim sendo, para o caso de existir interesse na palavra

caa, é possível verificar através da árvore que esta palavra ocorre três vezes, nos índices 1, 3 e 731. Contudo, também é possível verificar qual a palavra existente numa determinada posição e quantas vezes esta palavra ocorre. De modo a ilustrar esta situação, ao considerar, por exemplo, o índice *(m-n)-1*, verifica-se que a palavra correspondente ao referido índice é *cbb* e que a mesma ocorre duas vezes.

Com base na informação obtida, dá-se início ao processo denominado por *outer loop* que consiste na análise da última coluna do *array*, procurando pelo menor valor apresentado. A busca pelo menor número de ocorrências tem por base a noção de que uma subsequência anómala tende a apresentar uma representação simbólica praticamente única. Assim, os índices das palavras com menor número de ocorrências são registados, indicando a existência de potenciais candidatos a *discords*. Estes candidatos são então analisados através de um processo denominado por *inner loop*, onde são calculadas as distâncias entre os candidatos e as restantes subsequências de modo a eleger aquele que apresente a maior distância face às restantes subsequências, sendo então atingido o objetivo proposto e detetado o *discord* da série temporal em análise.

2.4. Avaliação dos resultados

De acordo com Gama et al. (2012), "a avaliação experimental de um algoritmo de extração de conhecimento de dados pode ser realizada segundo diferentes aspetos, tais como taxa de acerto do modelo gerado, compreensibilidade do conhecimento extraído, tempo de aprendizagem, requisitos de armazenamento do modelo, entre outros".

No caso em estudo, ao estarem disponíveis apenas os diagramas de carga de diferentes instalações sem qualquer informação adicional sobre situações normais e/ou anómalas possivelmente já identificadas pela EDPD, não se torna possível recorrer às métricas de avaliação que têm por base o cálculo da taxa de erro dos modelos. Neste sentido, a avaliação exata dos resultados obtidos será realizada em conjunto com a EDPD que, ao dispor de informação adicional, conseguirá validar se uma determinada subsequência identificada pelos modelos desenvolvidos como sendo *motif* ou *discord*, efetivamente representa uma situação regular ou excecional, respetivamente.

Apesar da necessidade de recorrer a terceiros para obter uma efetiva validação dos resultados obtidos, considera-se relevante encontrar uma forma de validar estes resultados antes mesmo de os apresentar à EDPD. Assim, de acordo com Keogh et al. (2005) pode

ser realizada uma validação visual dos *discords* identificados. Para tal, após conhecer quais as subsequências identificadas como *discords* numa determinada série temporal, deve-se observar a representação gráfica destas subsequências na série temporal original, validando visualmente se as mesmas demonstram um comportamento anómalo.

Na literatura estudada, é ainda usual encontrar comparações da performance de diferentes algoritmos capazes de identificar *motifs* e *discords.* Estas tendem a ser realizadas com base no esforço computacional exigido por cada algoritmo (Li et al. 2013; Yankov et al. 2008), sendo dada uma melhor avaliação aos que conseguem obter melhores resultados num menor espaço de tempo. De relembrar que a escolha dos algoritmos para o desenvolvimento deste trabalho teve em atenção este aspeto, dada a elevada quantidade de dados que a EDPD dispõe e pretende analisar.

Adicionalmente, ao serem utilizados dois métodos distintos para a identificação de *motifs* e de *discords*, torna-se possível avaliar a qualidade dos modelos ao comparar os resultados obtidos por cada algoritmo. Esta avaliação consiste em verificar que as subsequências identificadas como *motifs* por um algoritmo não coincidem com as subsequências indicadas como *discords* pelo outro algoritmo.

2.5. Comentários finais

A análise da literatura demonstra a existência e eficácia de alguns métodos para a deteção de alterações nos padrões de consumo. A utilização de algoritmos para a deteção de *motifs* permitem compreender o normal comportamento de uma série temporal, enquanto os algoritmos para deteção de *discords* evidenciam as anomalias verificadas. Este estudo permitiu uma compreensão aprofundada dos métodos existentes e do potencial associado a cada um deles, tornando possível um planeamento mais informado do processo necessário à resolução do problema tratado neste projeto.

Capítulo 3. Estudo de caso

O presente projeto tem como objetivo a deteção de alterações nos padrões de consumo de energia em instalações de média tensão. Para tal, recorreu-se a dados coletados pela EDPD no decorrer da sua atividade e disponibilizados para o referido efeito. Estes dados consistem em diagramas de carga que registam a potência média observada nas respetivas instalações ao longo do tempo. Não sendo utilizada qualquer informação adicional que permita identificar a instalação ou a atividade desenvolvida pela mesma. Ao longo deste capítulo serão apresentadas em detalhe as bases de dados utilizadas, bem como as operações realizadas a fim de preparar a informação para aplicar corretamente os algoritmos escolhidos para a deteção de *motifs* e de *discords*. Os resultados obtidos com a aplicação dos algoritmos serão descritos no capítulo seguinte, juntamente com a avaliação dos métodos utilizados.

3.1. Descrição das bases de dados

A EDPD, no decurso da sua atividade, procedeu à instalação de sistemas de medição com leitura remota (telecontagem) em várias empresas (instalações de média tensão), tornando possível a obtenção de informação detalhada sobre o consumo de energia elétrica destes locais. Estes equipamentos registam a potência média observada em kW, em intervalos de 15 em 15 minutos (perfazendo um total de 96 registos diários), possibilitando a criação de um diagrama de carga para cada instalação. Através dos dados recolhidos nos pontos de medição, a EDPD consegue obter uma quantidade significativa de informação da qual é possível extrair conhecimento útil para o desenvolvimento da sua atividade.

Ao indicar a potência média ao longo do tempo, os diagramas de carga contêm informação relevante sobre os acontecimentos dentro de uma instalação, tornando possível analisar os padrões de consumo dos clientes e, consequentemente, a existência de padrões frequentes e anómalos. De notar que os diagramas de carga disponibilizados pela EDPD contêm apenas dados sobre a potência média consumida nas instalações ao longo do tempo, não sendo possível identificar a instalação em questão nem as características da mesma.

Deste modo, uma vez que o estudo foi elaborado apenas com base nos valores observados para a potência média registada num determinado horizonte temporal, não subsiste o

problema da análise a efetuar sofrer influência de outras informações, como características da instalação ou dados comerciais, para a deteção de padrões. Por outro lado, tal como referido anteriormente, a ausência de informação adicional impossibilita a validação definitiva dos resultados finais. Neste sentido, de modo a ultrapassar este obstáculo, a validação final será efetuada em conjunto com a EDPD, após realizar a análise inicial dos padrões identificados.

Os diagramas de carga permitem obter um elevado volume de dados ordenados cronologicamente, sendo possível reunir a informação por dia (Figura 3.1), por semana (Figura 3.2), por mês (Figura 3.3) ou por ano (Figura 3.4), conforme o objetivo da análise a desenvolver.

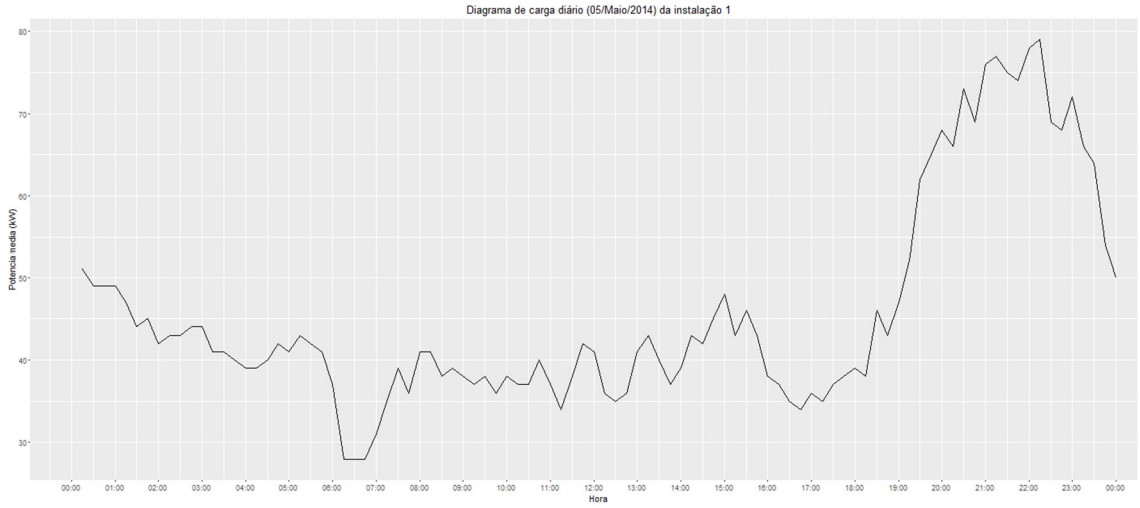


Figura 3.1: Representação gráfica do diagrama de carga diário de 05 de maio de 2014 da instalação 1

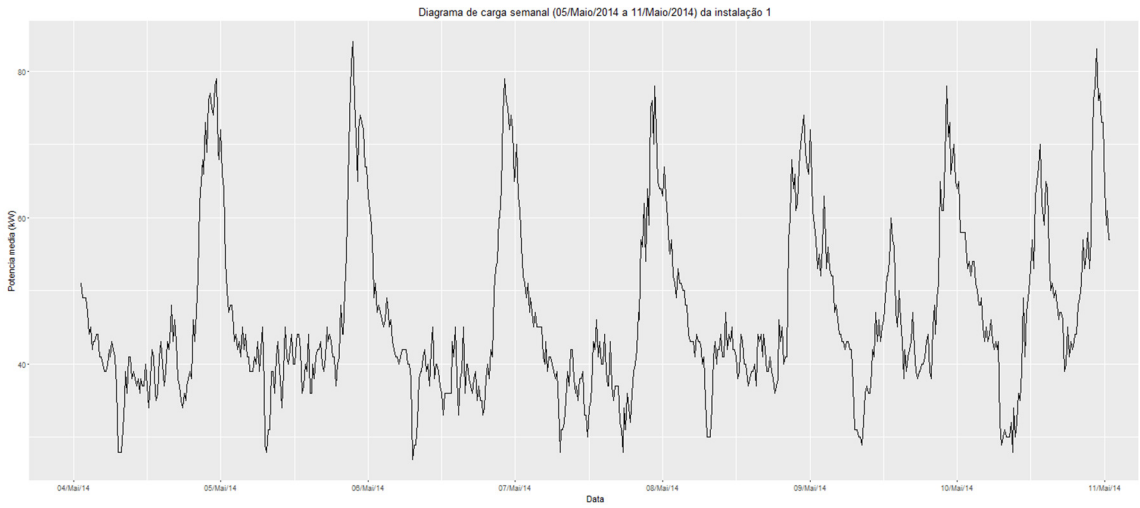


Figura 3.2: Representação gráfica do diagrama de carga semanal de 05 de maio de 2014 a 11 de maio de 2014 da instalação 1

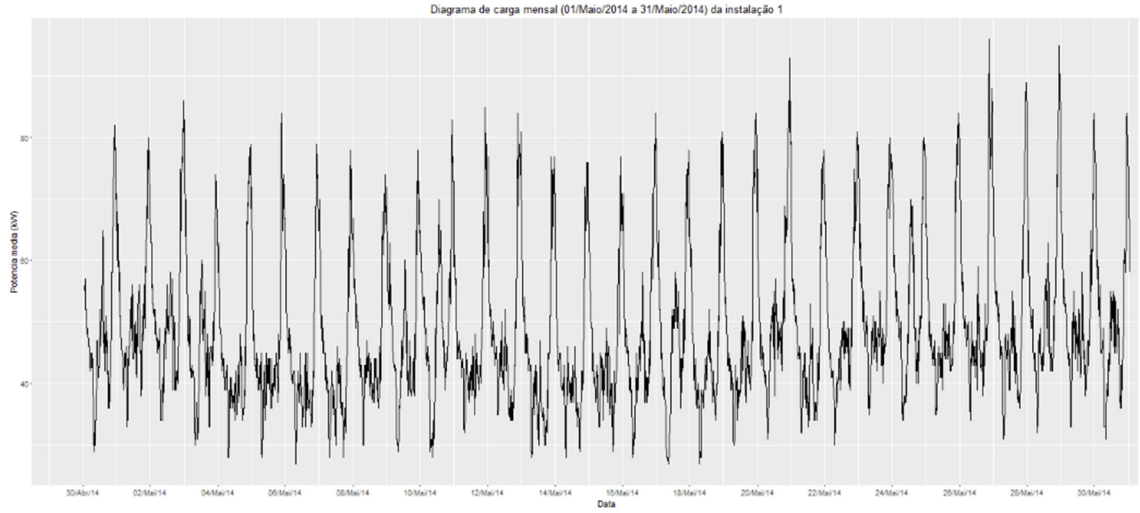


Figura 3.3: Representação gráfica do diagrama de carga mensal de maio de 2014 da instalação 1

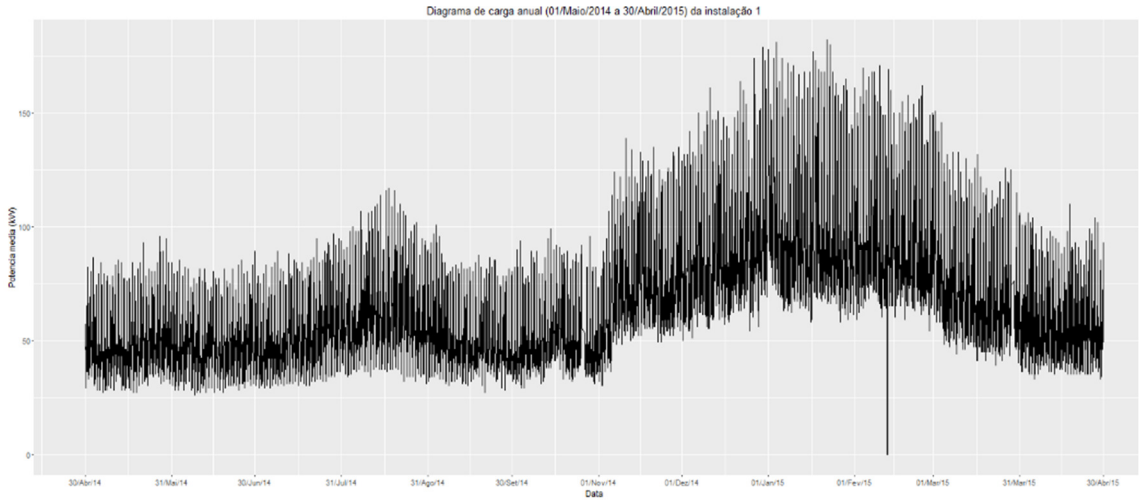


Figura 3.4: Representação gráfica do diagrama de carga anual de maio de 2014 a 30 de abril de 2015 da instalação 1

Através da visualização dos diagramas de carga apresentados é possível verificar que existem padrões nos consumos de energia das instalações. Contudo, é percetível que o alargamento do período de análise acarreta um aumento significativo no número de observações. Neste sentido, quanto maior o horizonte temporal a considerar, mais difícil se torna a realização de uma análise apenas visual da informação, para além deste tipo de análise poder ser considerada demasiado simples e subjetiva.

De referir que o consumo de energia em instalações de média tensão pode apresentar uma maior estabilidade em comparação ao consumo doméstico, dadas as características inerentes ao processo produtivo de uma instalação. Contudo, as instalações de média tensão podem apresentar comportamentos muito diferentes entre si, no que diz respeito ao consumo energético, com base na atividade que cada uma desenvolve. Tal afirmação

pode ser comprovada ao comparar os diagramas de carga de diferentes instalações para um mesmo horizonte temporal.

Neste sentido, ao comparar as Figuras 3.2 e 3.5, que representam os diagramas de carga para a semana de 05 a 11 de maio de 2014 das instalações 1 e 4, respetivamente, verifica-se que existe uma diferença significativa nos seus padrões de consumo. Admitindo que a semana em causa é uma semana normal para ambas as instalações, pode-se afirmar que a instalação 1 aparenta utilizar uma potência média entre 20 e 50kW, apresentando picos de maior consumo (entre 70 e 80kW) entre as 19:00 e as 22:00. Entretanto, a instalação 4 aparenta ter maiores variações ao longo do dia, com uma potência média que pode variar entre os 70 e os 130kW e apresentando picos de menor consumo (entre 40 e 60kW) entre as 06:00 e as 09:00.

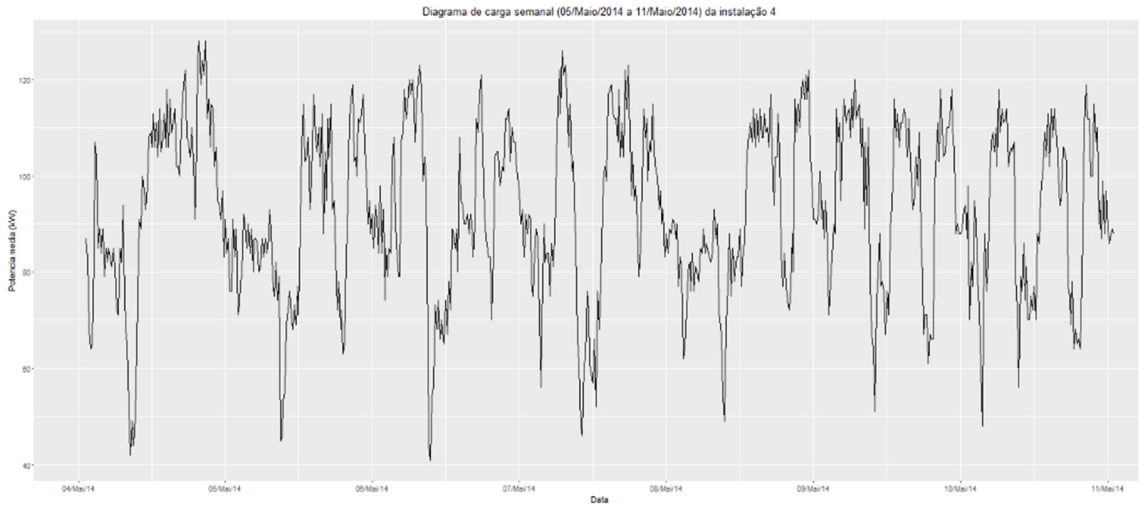


Figura 3.5: Representação gráfica do diagrama de carga semanal de 05 de maio de 2014 a 11 de maio de 2014 da instalação 4

Para além da subjetividade presente na análise visual dos dados, foi referido pela EDPD que identificar padrões nos consumos considerando apenas a simples análise de valores médios e desvios padrão pode ocultar informação relevante. Esta limitação consiste no facto das instalações poderem apresentar diferentes padrões ao longo do tempo, em consequência da normal evolução das suas atividades. Neste sentido, a simples comparação dos valores médios observados não permite obter um conhecimento detalhado da situação, não evidenciando as alterações que podem ocorrer nos padrões com o decorrer do tempo. Deste modo, demonstra-se a necessidade de recorrer a métodos mais sofisticados para a deteção dos padrões existentes.

3.2. Preparação da base de dados

Conhecida a natureza da informação disponível, torna-se necessário preparar as bases de dados de modo a ser possível aplicar os algoritmos escolhidos para a deteção de padrões frequentes e anómalos. A elevada quantidade de informação disponível exige a utilização de ferramentas capazes de trabalhar com grandes quantidades de dados de forma eficiente, permitindo a realização de diferentes estudos e a extração de conhecimento útil dos dados. Neste sentido, existem diversas ferramentas capazes de auxiliar na análise de dados, tais como *MS Excel* (Microsoft Corporation 2013), *R* (R Core Team 2014), *Matlab* (The MathWorks Inc. 2016)¸ *RapidMiner* (Hofmann & Klinkenberg 2013), *Weka* (Hall et al. 2009), entre outros, sendo que a escolha da ferramenta depende do objetivo em causa. Relativamente ao problema em estudo, o *software R* demonstra ser a ferramenta com maior potencial para alcançar os objetivos propostos ao apresentar uma interface gráfica para o usuário (*Graphical User Interface - GUI*). Esta apresenta as vantagens de ser gratuita (*freeware*) e compatível com múltiplas plataformas como *Windows*, *Linux* ou *Mac OS*, para além de recorrer à linguagem de programação *R* direcionada para a análise estatística de dados. Adicionalmente, a ferramenta possibilita o incremento das suas capacidades através da instalação de *packages* específicos, de acordo com os objetivos definidos, facilmente disponíveis e acessíveis. Assim, para a realização deste projeto foi utilizado maioritariamente o *software R*, versão 3.3.0, através do ambiente de desenvolvimento integrado (*Integrated Development Environment – IDE*) denominado por *RStudio*, versão 0.99.489 (RStudio Team 2015).

Relativamente às bases de dados disponíveis, sabe-se que os diagramas de carga, ao registarem a potência média observada nos últimos 15 minutos, apresentam como primeiro registo do dia a observação das 00:15, sendo a última observação deste mesmo dia registada às 00:00 do dia seguinte. Deste modo, a primeira etapa do processo de preparação dos dados consistiu na harmonização dos dados de forma a considerar apenas dias completos, com 96 observações cada, tendo-se optado por considerar apenas meses completos. Neste sentido, através do *MS Excel* foram eliminadas as primeiras e últimas observações das bases de dados, correspondentes a períodos incompletos, em conformidade com o exposto.

Numa segunda etapa, recorrendo à linguagem *R*, foi analisada a existência de dados incompletos nas bases de dados, nomeadamente, períodos sem indicação da potência

média observada. Neste sentido, através da criação de funções desenvolvidas para o efeito foram identificados os períodos com observações em falta, tendo sido definido estimar os respetivos valores ausentes. Para a estimativa dos valores em falta, foram consideradas as observações localizadas exatamente antes e depois deste período, sendo estimado para os períodos em falta o valor médio das observações consideradas.

De modo a concluir a preparação dos dados foi necessário realizar mais uma etapa por ter sido verificado que os algoritmos a utilizar consideram janelas deslizantes de dimensão fixa. De facto, ao realizar análises diárias, as janelas deslizantes devem apresentar uma dimensão de 96 observações; contudo, nos dias em que ocorre mudança de hora existem quatro observações a mais ou a menos devido a passagem para o horário de inverno ou verão, respetivamente. Neste sentido, optou-se por eliminar a informação relativa aos dias de mudança de hora de forma a ultrapassar o problema de existirem efetivamente dias com 23 e 25 horas em termos de potência média registada.

A Figura 3.6 apresenta uma tabela com o resumo da informação das cinco bases de dados disponíveis, sendo possível comparar os dados inicialmente disponibilizados com os dados a considerar, após as diferentes etapas de preparação das bases de dados:

		DC1	**DC2**	**DC3**	**DC4**	**DC5**
ORIGINAL	**Número de observações**	64.343	76.069	100.727	100.943	73.285
	Primeira observação	18/01/2014 00:15	17/09/2013 12:30	01/01/2013 00:15	01/01/2013 00:15	16/10/2013 15:00
	Última observação	19/11/2015 09:15	19/11/2015 02:15	19/11/2015 00:30	19/11/2015 00:45	19/11/2015 02:15
FINAL	**Número de observações**	60.864	72.576	98.688	98.688	69.696
	Primeira observação	01/02/2014 00:15	01/10/2013 00:15	01/01/2013 00:15	01/01/2013 00:15	01/11/2013 00:15
	Última observação	01/11/2015 00:00	01/11/2015 00:00	01/11/2015 00:00	01/11/2015 00:00	01/11/2015 00:00

Figura 3.6: Tabela com informação das cinco bases de dados disponíveis antes e depois da preparação dos dados

3.3. Identificação de *motifs*

Concluída a fase de preparação da base de dados e tendo em consideração a opção pela utilização do algoritmo *MrMotif* para a identificação de *motifs*, optou-se por utilizar a

ferramenta desenvolvida pelos autores deste algoritmo. Esta ferramenta recebeu o nome de *iMotifs* (*Interactive Time Series Motif Discovery and Visualization Tool*) e foi desenvolvida em *JAVA*, tendo demonstrado ser de fácil acesso e compreensão e apresentando resultados de forma rápida e intuitiva. A única desvantagem a apontar consiste no facto de não permitir a exportação dos resultados obtidos, sendo necessário realizar a extração manual e individual dos resultados.

A Figura 3.7 apresenta a interface da referida ferramenta, sendo possível visualizar os parâmetros de entrada necessários e os resultados obtido:

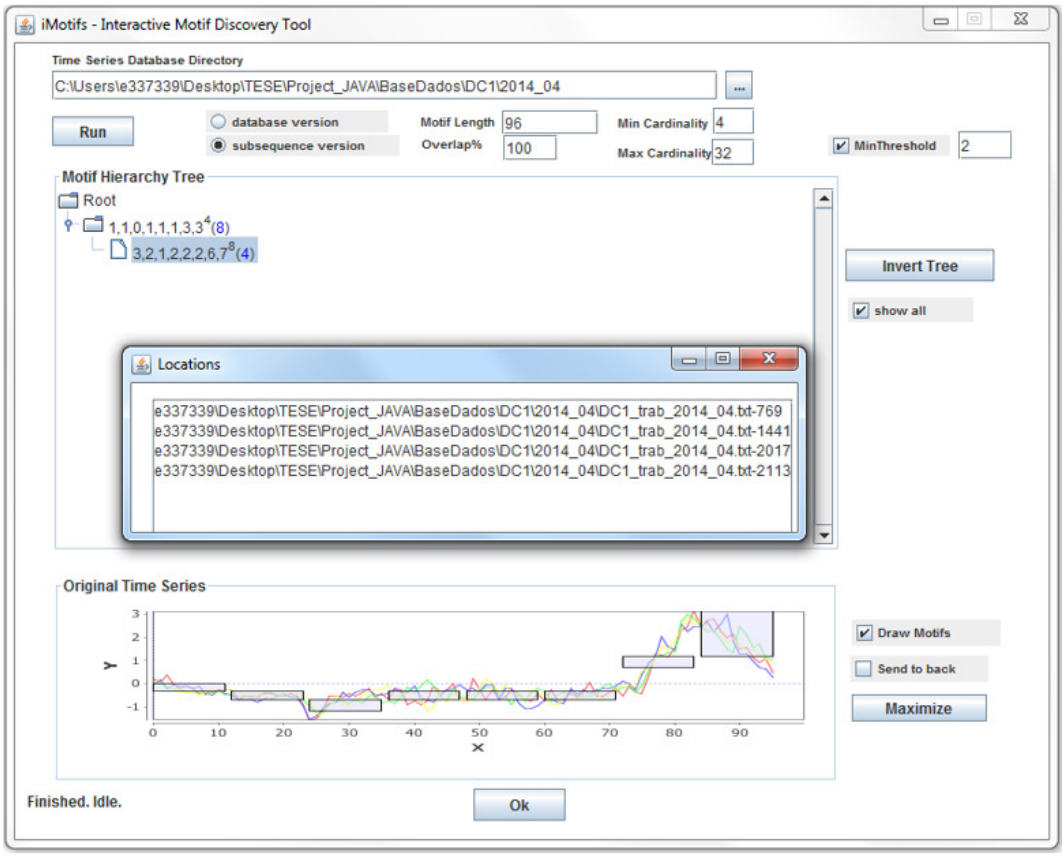


Figura 3.7: Interface da ferramenta iMotifs

Para a deteção de *motifs* diários, após indicar a localização da base de dados a considerar, deve ser indicado que o estudo será realizado com base em subsequências, escolhendo para tal a opção "*subsequence version*" e indicando no campo "*Motif Length*" a dimensão de cada subsequência que, para a análise em causa, devem ser as 96 observações diárias. Dado que a análise deve ser realizada considerando os dias completos, não deve ser permitido que o algoritmo compare subsequências sobrepostas, devendo assim ser indicado no campo "*Overlap%*" o valor de 100. De referir que foi testada a hipótese de considerar um valor de 0 para o referido campo, contudo os resultados obtidos

demonstraram que o algoritmo neste caso considera que as subsequências podem estar sobrepostas ao gerar mais palavras do que o número de dias existentes. Em termos de resolução pode ser mantido o intervalo de 4 a 32 símbolos para que o algoritmo trabalhe com todas as possíveis resoluções. Para finalizar é possível definir que uma palavra apenas será considerada pelo algoritmo como um potencial *motif* se ocorrer pelo menos três vezes na base de dados em estudo, preenchendo para tal o campo "*MinThreshold*" com o valor 2.

Introduzida a informação necessária, conforme exposto, é possível correr o algoritmo e analisar os resultados obtidos. A ferramenta *iMotifs* apresenta os *motifs* detetados através de uma árvore organizada hierarquicamente com indicação das palavras e respetivas resoluções, assim como o número de vezes que cada palavra ocorre. Ao aceder aos respetivos *motifs*, representados pelas palavras concebidas, é possível obter informação adicional que poderá auxiliar na análise dos resultados. Esta informação consiste num gráfico com a representação de todos os dias que compõem este *motif* e a indicação da localização dos mesmos na base de dados original, permitindo identificar os dias correspondentes.

Recorrendo à ferramenta referida e com base nos valores indicados para os parâmetros, foram extraídos os potenciais *motifs* para as cinco instalações, considerando, de forma individualizada, os dados por mês e por dia da semana. A informação obtida foi compilada no *MS Excel* e recorrendo à ferramenta *PivotTable* foi possível definir diversos critérios de seleção, possibilitando a realização de diferentes análises.

Os *motifs* mais relevantes foram aqueles cuja palavra ocorreu um maior número de vezes, sendo de referir que, ao observar os resultados da análise por dias da semana, os *motifs* identificados apresentam um maior número de ocorrências do que na análise por mês para todas as bases de dados. Esta variação no número de ocorrências tem origem no facto de ser considerado, na extração dos resultados, um mínimo de três palavras iguais na base de dados em causa para que uma palavra seja indicada nos resultados como potencial *motif*. Dado que na análise mensal são considerados apenas os dados do mês em estudo, pode suceder de uma palavra ocorrer apenas uma ou duas vezes neste mês, fazendo com que a mesma não será refletida nos resultados mensais por não ultrapassar a barreira das três palavras. Entretanto, ao realizar o estudo por dias da semana são considerados todos dias da base de dados que correspondem a um determinado dia da semana, podendo assim

uma palavra apresentar um maior número de ocorrências pois a base de dados utilizada é consideravelmente maior. De notar ainda que, em algumas situações pontuais, foi necessário reduzir o número mínimo de ocorrência das palavras a serem consideradas nos resultados como *motifs* para dois, apesar de ser um número muito reduzido para admitir a efetiva existência de um padrão nos dados.

Por fim, foi criada uma função no *software R* com o objetivo de obter a lista completa de palavras para uma determinada base de dados. Dado que a ferramenta *iMotifs* não permite a extração automática das palavras geradas, a função criada pretende obter a mesma lista de palavras mas permitindo a extração da informação. Esta relação de palavras possibilita a realização de análises adicionais para além da identificação dos *motifs* obtidos através do *iMotifs*. A Figura 3.8 apresenta a função desenvolvida em linguagem R, que recorre ao *package seewave* para criar a representação simbólica das palavras através do algoritmo SAX, apresentando também um pequeno exemplo da lista de palavras gerada:

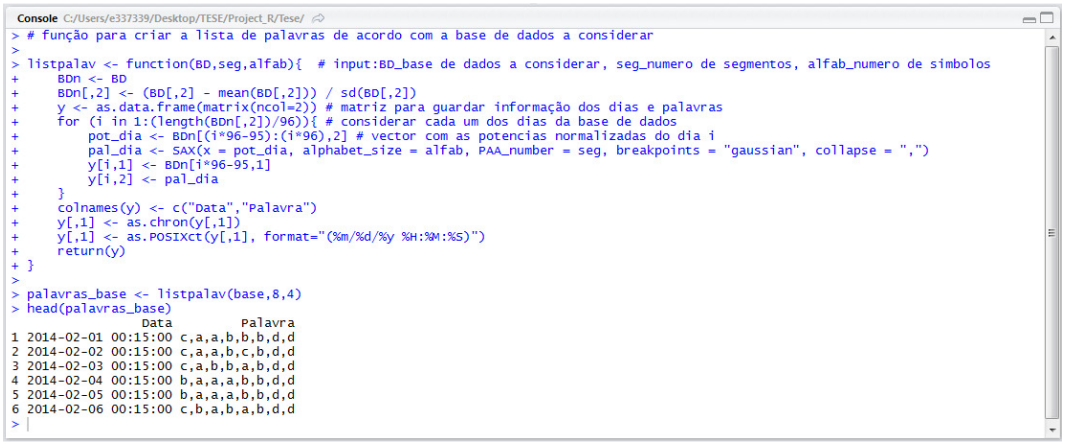


Figura 3.8: Interface do R com indicação da função criada a fim de obter as palavras por dia

3.4. Identificação de *discords*

Identificados os padrões habituais numa instalação e de modo a alcançar todos os objetivos definidos para o presente trabalho, procedeu-se à deteção de alterações nos padrões de consumo nestas mesmas instalações. Neste sentido, para a identificação dos *discords* optou-se por recorrer ao *package jmotif*, versão 1.0.2 (Senin 2016), disponível em linguagem *R*. Na medida em que o objetivo do presente estudo consiste na análise diária dos padrões, todos os dias das bases de dados devem ser considerados de forma consistente, garantido que todos têm início a mesma hora, apresentando como primeira observação o período 00:15, e contendo as 96 observações diárias. De forma a garantir que tal se verifica, foi necessário ajustar o código do *package* em questão de acordo com

os objetivos definidos. A alteração referida esteve relacionada com a forma como o código original do *package* executa a deslocação das janelas deslizantes, visto este não considerar uma deslocação igual à dimensão das subsequências, permitindo assim a sobreposição das janelas.

A Figura 3.9 permite visualizar a aplicação do referido *package* através da utilização da função *find_discords_hot_sax*, que tem como resultado a distância e a posição dos *discords* identificados. Para além da indicação da base de dados a considerar, esta função exige a definição de diferentes parâmetros que podem alterar o resultado final, nomeadamente a dimensão da janela deslizante (*w_size*), a dimensão dos segmentos a considerar na redução do número de observações (*paa_size*), a dimensão do alfabeto (*a_size*) e o limite a partir do qual a base de dados deve ser normalizada (*n_threshold*).

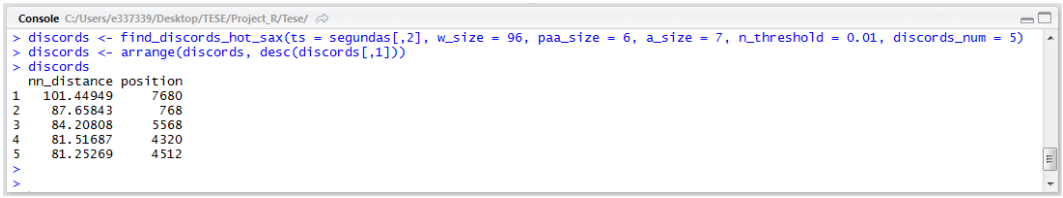


Figura 3.9: Interface do R com a aplicação da função find_discords_hot_sax e exemplo de resultados

De modo a definir os melhores valores a considerar para os parâmetros, foram definidos diferentes cenários, alterando os valores dos parâmetros da função referida e analisando os resultados obtidos para as cinco bases de dados disponíveis. Tendo em consideração uma análise diária dos *discords*, a dimensão da janela deslizante manteve-se igual pelo facto de existirem 96 observações diárias. Relativamente ao número de segmentos a considerar dentro de cada janela deslizante, foram testados os valores de 24, 12, 6 e 4, agrupando as observações diárias em segmentos de 1, 2, 4 e 6 horas, respetivamente. No que diz respeito ao número de letras do alfabeto a considerar, analisou-se a possibilidade de utilizar 3, 5, 7, 9, 11 e 15 letras.

Foram realizados vários testes com diferentes combinações de valores para os parâmetros referidos e, com base no estudo realizado, concluiu-se que os resultados obtidos para os diferentes cenários são consistentes, não tendo sido observadas alterações significativas com a variação dos valores dos parâmetros em causa. Neste sentido, e de modo a reduzir o esforço computacional, decidiu-se considerar os seguintes valores para os parâmetros em causa (Figura 3.10):

Parâmetro	Valor	Descrição
w_size	96	96 observações por dia
paa_size	6	6 segmentos por dia (4 horas por segmento)
a_size	7	7 letras possíveis (a, b, c, d, e, f, g)

Figura 3.10: Tabela com valores dos parâmetros a considerar para a função find_discords_hot_sax

Recorrendo à função referida anteriormente e com base nos valores indicados na tabela da Figura 3.10 para os parâmetros, foram extraídos os cinco *discords* mais relevantes por mês para as cinco bases de dados disponíveis. De referir que a opção pela extração de cinco *discords* teve em vista não limitar demasiado o algoritmo na pesquisa a efetuar por padrões anómalos. Os resultados obtidos foram posteriormente ordenados de acordo com a distância mínima calculada entre o dia candidato a *discord* e os restantes dias do mês em análise. De acordo com a definição de *discord*, quanto maior a distância observada entre uma subsequência e as restantes subsequências da base de dados em causa, mais dissemelhantes serão estas observações.

De modo a facilitar a análise dos resultados obtidos, optou-se por extrair a representação gráfica dos diagramas de carga mensais, evidenciando os cinco *discords* identificados por mês. Estes gráficos permitem validar, visualmente, se o método utilizado conseguiu efetivamente identificar os *discords*, assim como permite analisar a existência de algum tipo de padrão ao longo do tempo que justifique as situações anómalas identificadas, visto as análises terem sido efetuadas considerando os meses individualmente.

3.5. Considerações finais

Ao compreender o problema em estudo, as técnicas e ferramentas existentes e a informação disponível, torna-se possível organizar as tarefas a executar de modo a atingir os objetivos definidos. Deste modo, o presente capítulo teve como propósito descrever as tarefas e processos realizados de modo a obter os resultados esperados, nomeadamente, a identificação de padrões frequentes e anómalos. Assim sendo, no capítulo seguinte serão apresentados em detalhe os resultados obtidos para as cinco bases de dados disponíveis. Visto as instalações em estudo poderem apresentar comportamentos diferentes, as conclusões serão apresentadas de forma individualizada de modo a evidenciar as especificidades de cada instalação.

Capítulo 4. Resultados

Ao longo do presente capítulo são expostos os resultados obtidos nos processos de identificação de padrões frequentes e anómalos. De modo a facilitar a apresentação dos resultados, os *motifs* e os *discords* serão expostos inicialmente de forma separada, sendo de seguida realizada uma análise de forma agregada. Esta agregação permite realizar uma primeira validação da informação obtida, visto os padrões frequentes identificados não poderem ser iguais aos padrões anómalos. A validação efetiva é apresentada no fim deste capítulo e foi realizada em conjunto com a EDPD por ser necessário recorrer a informação adicional para validar os resultados. De referir ainda que, dadas as possíveis especificidades de cada instalação de média tensão, a apresentação dos resultados é individualizada por instalação.

4.1. Resultados da identificação de *motifs*

O processo detalhado na Secção 3.3. foi executado para as cinco bases de dados em estudo, tendo sido extraídos os *motifs* identificados para cada instalação. Os resultados obtidos foram compilados e através das tabelas apresentadas nos Anexos 1 e 2 é possível observar os *motifs* mais relevantes para cada instalação. Os *motifs* mais relevantes consistem naqueles cuja respetiva palavra ocorreu um maior número de vezes. A tabela apresentada no Anexo 1 reúne os resultados da análise mensal, enquanto a tabela presente no Anexo 2 contém os resultados da análise por dias da semana. Nos anexos 3 a 7 é possível observar a representação simbólica de todos os dias que compõem as cinco bases de dados, facilitando a visualização dos padrões que podem existir. Em alguns casos verifica-se que são formadas demasiadas palavras, sugerindo que os dias apresentam elevada variabilidade entre si, o que poderá dificultar a identificação dos *motifs*.

De seguida é apresentada uma breve análise dos principais resultados obtidos para cada uma das bases de dados de forma individual, visto as instalações em estudo poderem apresentar comportamentos diferentes e, consequentemente, as conclusões diferirem de acordo com as especificidades de cada instalação. No entanto, é necessário ter em consideração que, para além do intervalo temporal e da potência média registada, nesta fase não existe qualquer tipo de informação adicional que auxilie na explicação dos *motifs* encontrados.

Instalação 1

A base de dados relativa à instalação 1 contém informação sobre as potências médias registadas entre Fevereiro/2014 e Outubro/2015, sendo assim composta por 634 dias divididos em 92 sábados, 87 domingos e 91 dias para cada dia útil da semana.

De acordo com os resultados obtidos, esta instalação apresenta como *motifs* mais relevantes os dias representados pelas palavras "1,0,0,1,1,1,3,3", "1,0,0,1,2,1,3,3" e "1,0,1,1,1,1,3,3", todas com resolução 4. Estas palavras são relevantes para ambas as análises, ocupando a mesma posição tanto na análise mensal como na análise por dias da semana.

O primeiro *motif* "1,0,0,1,1,1,3,3" (Figura 4.1) ocorre essencialmente durante os meses de inverno, entre novembro e março, sendo um padrão observado maioritariamente à sexta-feira. O terceiro *motif* "1,0,1,1,1,1,3,3" (Figura 4.3) é muito semelhante ao primeiro, variando apenas o símbolo do terceiro segmento, e também apresenta uma maior incidência nos meses de inverno, entre dezembro e janeiro, contudo, este padrão é observado maioritariamente à segunda-feira. Por sua vez, o segundo *motif* "1,0,0,1,2,1,3,3" (Figura 4.2) ocorre maioritariamente nos meses de verão, julho e agosto, mas a sua distribuição pelos dias da semana aparenta ser mais constante e sem grandes variações.

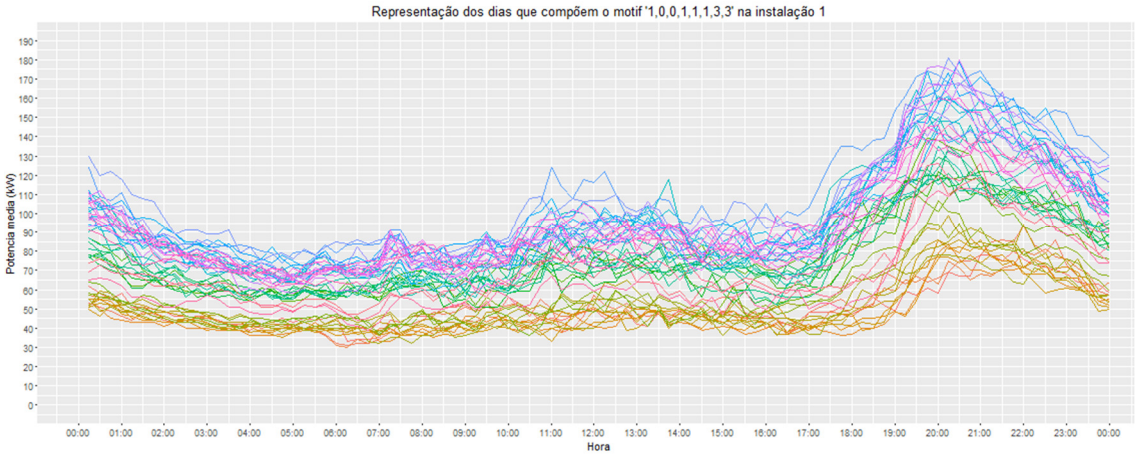


Figura 4.1: Representação gráfica do primeiro motif identificados para a instalação 1 ("1,0,0,1,1,1,3,3")

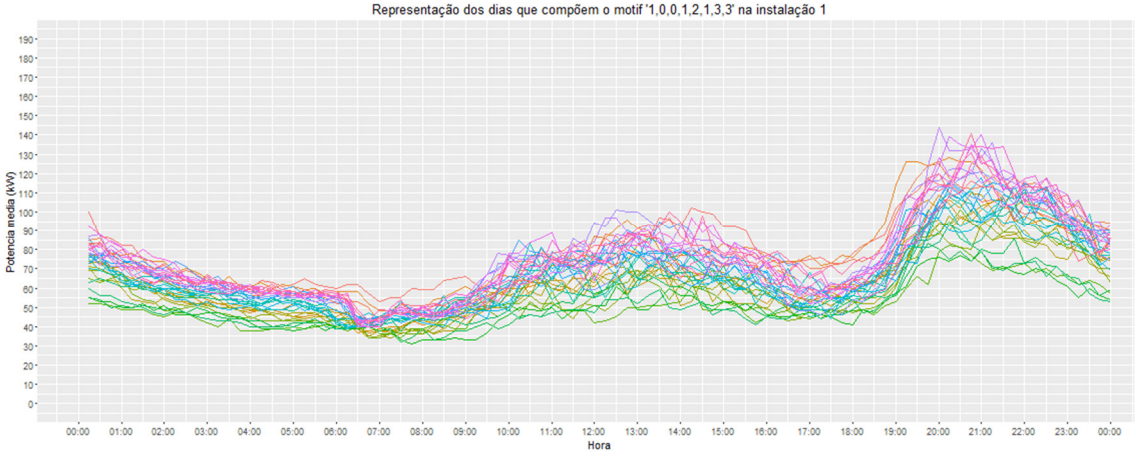


Figura 4.2: Representação gráfica do segundo motif identificados para a instalação 1 ("1,0,0,1,2,1,3,3")

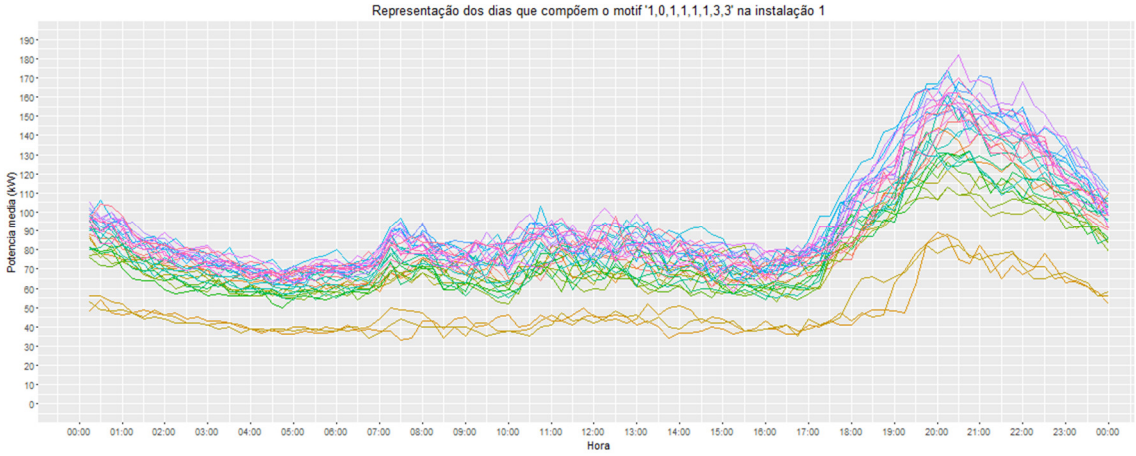


Figura 4.3: Representação gráfica do terceiro motif identificados para a instalação 1 ("1,0,1,1,1,1,3,3")

As representações gráficas apresentadas nas figuras acima auxiliam a compreensão dos *motifs* identificados, na medida em que os gráficos são compostos pela representação de todos os dias que compõem o *motif* em causa. Através destes gráficos é possível visualizar as diferenças e semelhanças entre os padrões identificados. Em termos de semelhanças, é visível que a potência média observada nunca é inferior a 30kW e que o valor máximo é registado às 21:00, existindo um aumento da potência média a partir das 17:00. Contudo, no inverno os valores máximos atingem os 180kW e no verão a potência média registada fica no limite dos 140kW.

Em termos dos valores observados para a potência média é possível referir que o primeiro *motif* apresenta uma maior amplitude de valores (entre 40 e 100kW), essencialmente das 6:00 às 17:00. No mesmo intervalo temporal, o terceiro padrão apresenta valores médios entre 60 e 90kW, sendo notório o aumento da potência a partir das 7:00. Analisando esta

informação em conjunto com os dados apresentados nas tabelas é possível supor que esta instalação incorre em maiores consumos no início da semana devido ao arranque dos seus equipamentos, dado este terceiro *motif* ocorrer essencialmente à segunda-feira.

De modo a completar a análise, pode ser mencionado que durante o verão a potência média registada apresenta valores entre 50 e 80kW, maioritariamente das 9:00 as 18:00. Adicionalmente, é possível apurar que, independentemente do período do ano em causa, a potência média diária apresenta os valores mais elevados entre as 18:00 e as 00:00, atingindo os valores máximos às 21:00, como referido anteriormente.

Instalação 2

Relativamente à base de dados da instalação 2 foram considerados 756 dias, entre Outubro/2013 e Outubro/2015, dos quais 103 dias são domingos, 108 dias são segundas e os restantes dias da semana apresentam 109 dias cada.

Com base nos resultados obtidos, a instalação 2 apresenta como *motifs* mais relevantes, com uma resolução de 4, os dias representados pelas palavras "0,0,2,3,3,3,0,0", "2,2,1,1,1,1,1,3" e "0,0,1,3,3,3,0,0". De notar que estas palavras ocupam a mesma posição tanto na análise por meses como na análise por dias da semana.

O segundo *motif* "2,2,1,1,1,1,1,3" (Figura 4.5) ocorre essencialmente durante o verão ao apresentar um elevado número de ocorrências no mês de agosto (15 dias em 2014 e 18 dias em 2015) e também aos finais de semana (34 sábados e 40 domingos). O primeiro e o terceiro *motifs* (Figuras 4.4 e 4.6) são muito semelhantes entre si, ao variar apenas o terceiro segmento da palavra, demonstrando ser este o padrão habitual nos dias úteis e nos restantes meses do ano. Neste sentido, é possível supor que esta instalação opera habitualmente entre as 8:00 e as 18:00 nos dias úteis, estando potencialmente encerrada no mês de agosto e nos finais de semana.

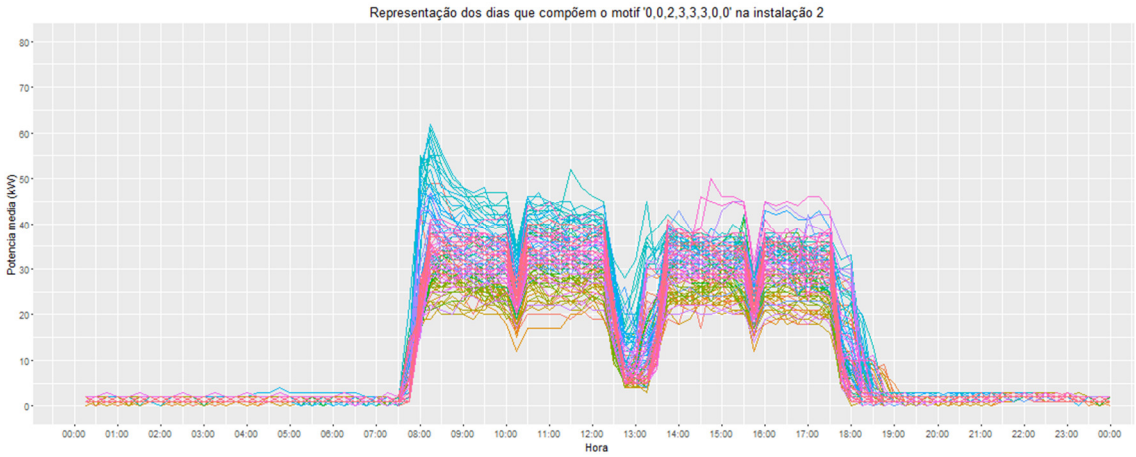


Figura 4.4: Representação gráfica do primeiro motif identificado para a instalação 2 ("0,0,2,3,3,3,0,0")

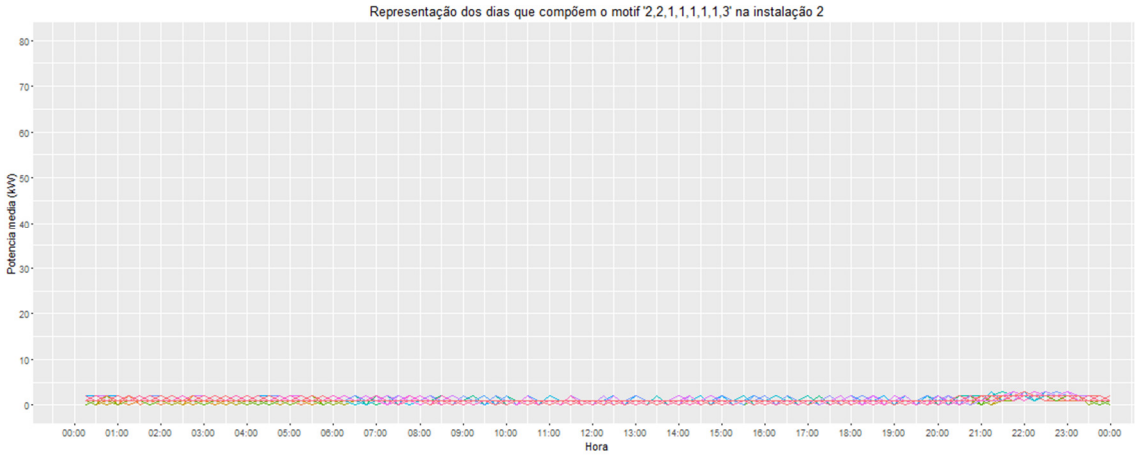


Figura 4.5: Representação gráfica do segundo motif identificado para a instalação 2 ("2,2,1,1,1,1,1,3")

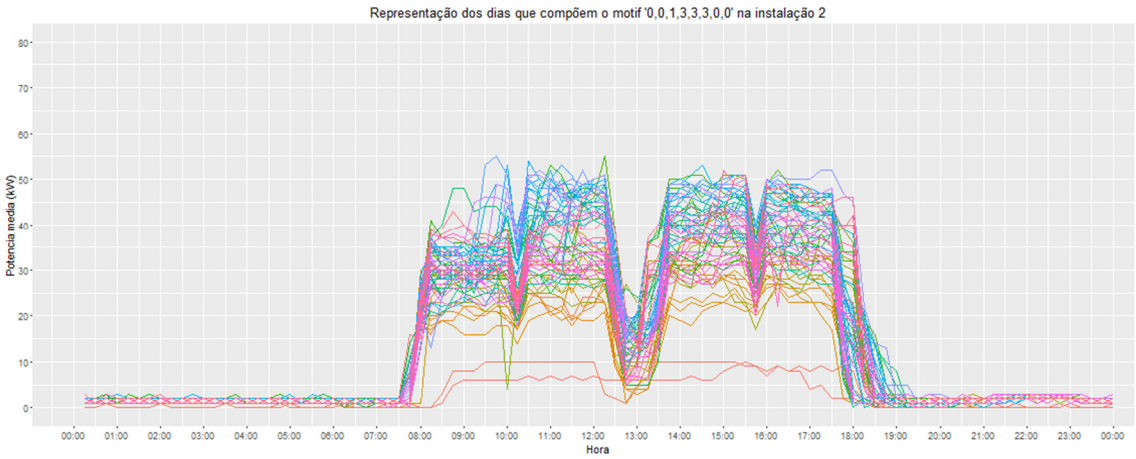


Figura 4.6: Representação gráfica do terceiro motif identificado para a instalação 2 ("0,0,1,3,3,3,0,0")

As figuras acima contêm as representações gráficas de todos os dias que compõem cada um dos *motifs* identificados, permitindo validar que o segundo *motif* diz respeito ao período em que a instalação se encontra encerrada, visto a potência média registada ser

residual (muito próxima de zero). Os restantes padrões demonstram que a instalação opera habitualmente entre as 8:00 e as 18:00, parando para o período de almoço entre as 12:00 e as 14:00 e realizando pausas pontuais às 10:00 e às 16:00. Estas pausas podem estar relacionadas com a atividade desenvolvida pela instalação visto ocorrerem exatamente a meio dos períodos da manhã e da tarde. De modo a concluir esta análise é possível referir que a potência média registada para o primeiro *motif* atinge valores máximos pouco acima dos 60kW, por volta das 8:00. Por sua vez, o terceiro *motif* não apresenta valores acima dos 55kW, sendo que estes valores podem ocorrer ao longo do dia e não apenas no início da atividade da instalação.

Instalação 3

A análise da base de dados da instalação 3 considera o período entre Janeiro/2013 e Outubro/2015, admitindo 1028 dias divididos por 141 domingos, 147 segundas e 148 dias para cada um dos restantes dias da semana.

De acordo com os resultados obtidos, as três palavras que apresentam um maior número de ocorrências e que, consequentemente, são denominadas por *motifs* para a instalação 3 são "1,1,2,3,2,2,1,1", "1,1,1,3,3,2,1,1" e "1,1,1,3,2,3,1,1". De referir que ao observar as palavras em causa, é possível verificar que as três são muito semelhantes entre si ao apresentar na representação simbólica cinco segmentos exatamente iguais, nomeadamente o primeiro, o segundo, o quarto, o sétimo e o oitavo segmentos. As Figuras 4.7, 4.8 e 4.9 apresentam as representações gráficas do primeiro, segundo e terceiro *motifs* identificados, respetivamente.

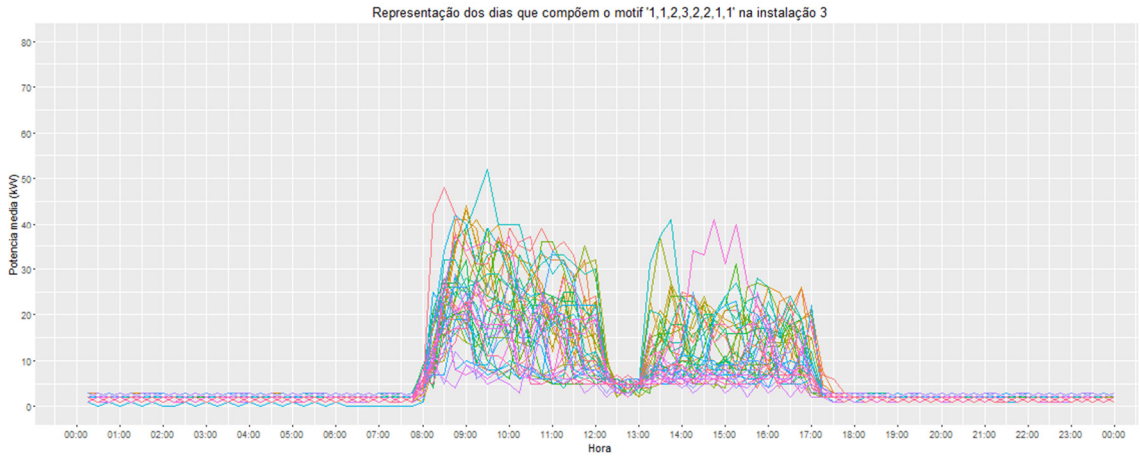


Figura 4.7: Representação gráfica do primeiro motif identificado para a instalação 3 ("1,1,2,3,2,2,1,1")

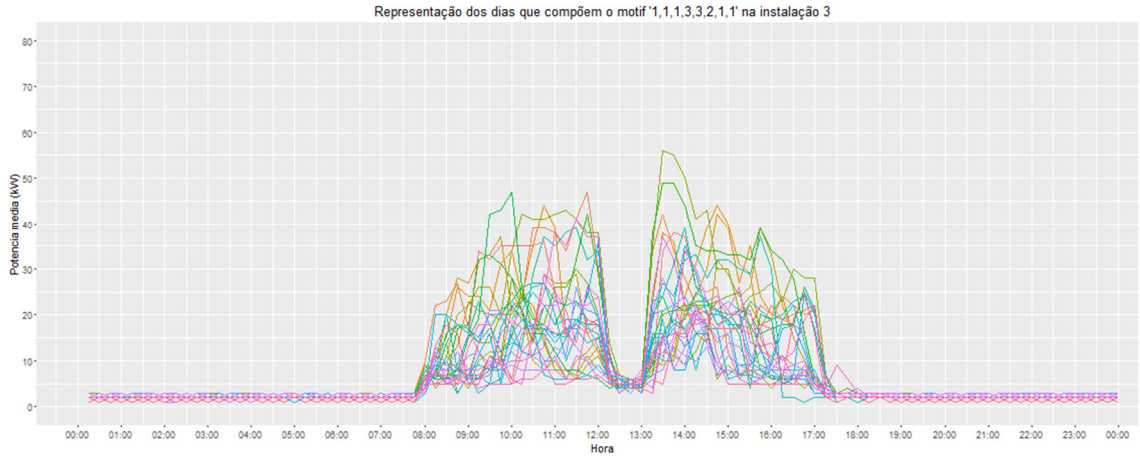


Figura 4.8: Representação gráfica do segundo motif identificado para a instalação 3 ("1,1,1,3,3,2,1,1")

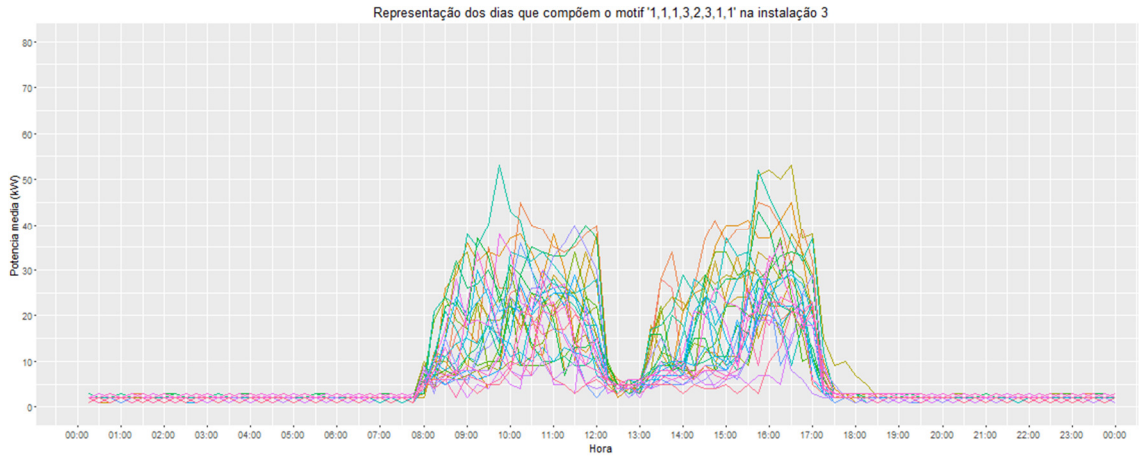


Figura 4.9: Representação gráfica do terceiro motif identificado para a instalação 3 ("1,1,1,3,2,3,1,1")

Com base nos resultados da análise mensal não é possível relacionar diretamente um *motif* a um período específico do ano mas os resultados da análise por dias da semana indica que os padrões identificados dizem respeito a dias úteis, visto que todos ocorrem de segunda a sexta-feira. Neste sentido, os resultados podem indicar que a instalação 3 realiza a sua atividade habitualmente nos dias úteis.

Através das representações gráficas presentes nas figuras acima é possível comprovar que esta instalação inicia a sua atividade habitualmente as 8:00 e encerra as 17:30, realizando uma pausa para o almoço entre as 12:00 e as 13:00. Em termos de potência média registada nos períodos normais de atividade, verifica-se que para os três padrões identificados os valores permanecem entre os 5 e os 40kW, atingindo valores máximos de cerca de 55kW pontualmente.

Instalação 4

No que diz respeito à base de dados da instalação 4, à semelhança da instalação 3, é considerado o período entre Janeiro/2013 e Outubro/2015, admitindo 1.028 dias divididos por 141 domingos, 147 segundas e 148 dias para cada um dos restantes dias da semana.

Os *motifs* detetados para esta instalação são representados pelas palavras "1,3,2,1,1,0,3,2", "1,0,0,2,2,3,3,1" e "2,0,2,2,0,2,2,1". Contudo, o processo para deteção destes *motifs* apresentou ligeiras diferenças face às restantes instalações na medida em que esta instalação aparenta ter um comportamento muito instável ao longo do tempo, dificultando a definição de padrões. De referir que para alguns meses foram identificados apenas dois dias representados pela mesma palavra e em outros meses não foi possível atingir esta barreira dos dois dias.

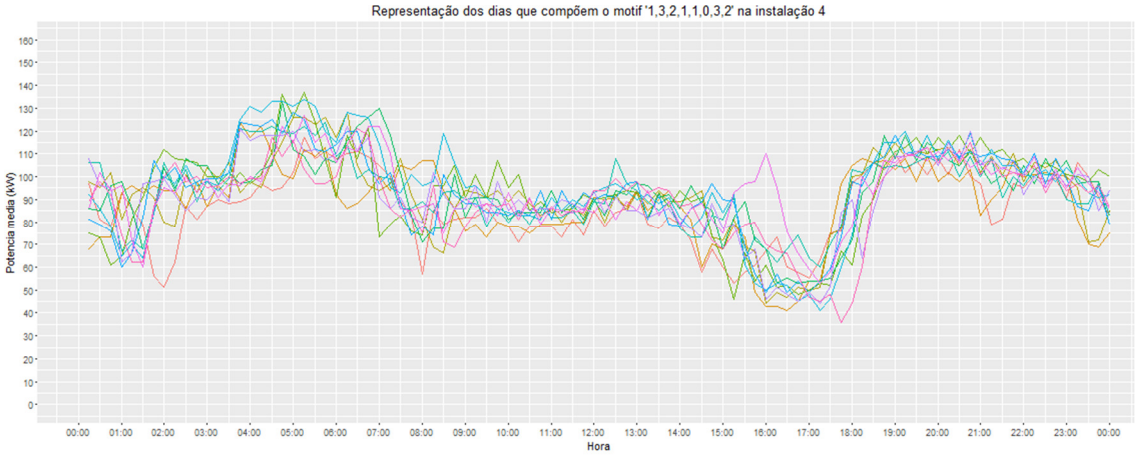


Figura 4.10: Representação gráfica do primeiro motif identificado para a instalação 4 ("1,3,2,1,1,0,3,2")

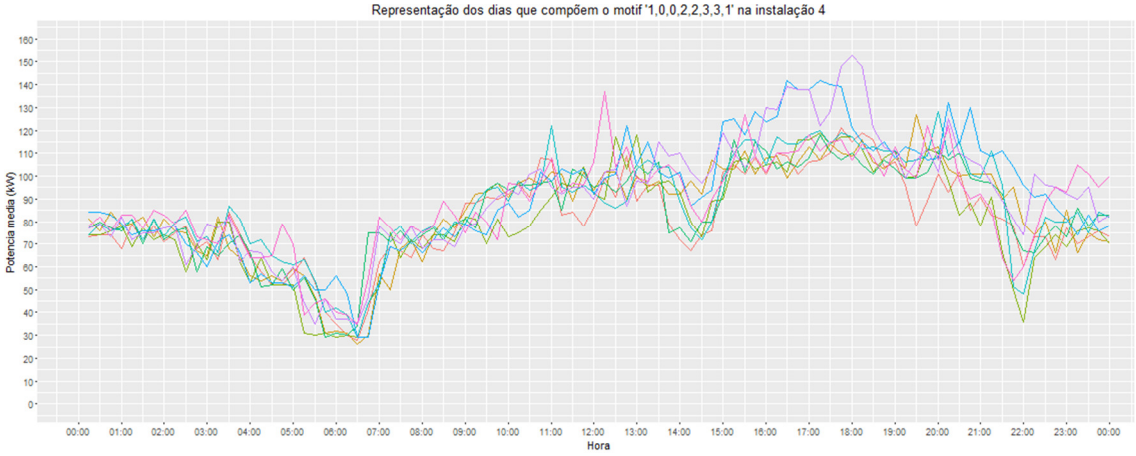


Figura 4.11: Representação gráfica do segundo motif identificado para a instalação 4 ("1,0,0,2,2,3,3,1")

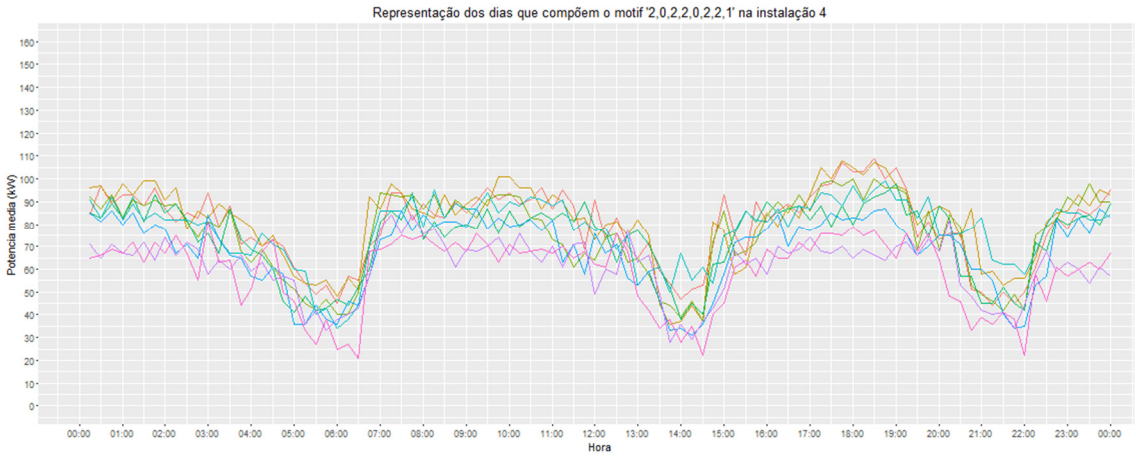


Figura 4.12: Representação gráfica do terceiro motif identificado para a instalação 4 ("2,0,2,2,0,2,2,1")

As Figuras 4.10, 4.11 e 4.12 apresentam as representações gráficas dos dias que compõem os três *motifs* referidos, sendo possível verificar que mesmo se tratando dos três principais *motifs*, o número de dias representados em cada um é muito reduzido.

Ao analisar os *motifs* identificados, verifica-se que o primeiro apresenta valores mínimos a rondar os 40kW, entre as 16:00 e as 17:30, contrariamente ao segundo *motif* que apresenta no mesmo período os valores mais elevados ao atingir os 140kW para a potência média observada. Em relação ao terceiro *motif* é possível verificar que este apresenta valores mais reduzidos em comparação aos restantes. Este último apresenta valores máximos a rondar os 100kW e três períodos de potência média observada muito reduzida, entre 20 e 40kW, das 5:00 as 6:30, das 13:30 as 14:30 e das 20:30 as 22:00.

As grandes variações observadas na potência média registada ao longo do mesmo dia fazem com que, apesar de serem considerados muitos dias na análise, seja raro encontrar dias semelhantes. Neste sentido, considera-se pouco fiável admitir uma determinada palavra como padrão.

Instalação 5

A base de dados da instalação 5 contém informação sobre a potência média registada entre Novembro/2013 e Outubro/2015, num total de 726 dias, dos quais 100 são domingos, 105 são sextas, sábados apresentam o mesmo número de dias de sexta e os restantes dia da semana apresentam 104 dias cada.

Com base nos resultados obtidos é possível referir que a instalação 5 apresenta quatro *motifs* principais em que, por oposição às restantes instalações, o terceiro *motif*

identificado pela análise mensal difere do identificado pela análise por dias da semana. Assim sendo, as palavras que apresentam maior número de ocorrências são "1,1,1,2,2,2,1,1", "1,1,1,2,2,2,2,1", "1,1,1,3,2,2,1,1" (pela análise por mês) e "1,1,1,2,2,2,2,2" (pela análise por dias da semana), tal como ilustrado nas Figuras 4.13, 4.14, 4.15 e 4.16, respetivamente.

Apesar da análise dos resultados por mês não permitir identificar um padrão para um determinado período do ano, a análise por dias da semana demonstra que esta instalação deve operar essencialmente nos dias úteis. Ao analisar os dois primeiros *motifs* é possível verificar que estes são muito semelhantes entre si pelo facto de variar apenas o símbolo do sétimo segmento e por ambos apresentarem um maior número de ocorrências nos dias úteis, apesar do segundo ocorrer essencialmente às segundas, quartas e sextas.

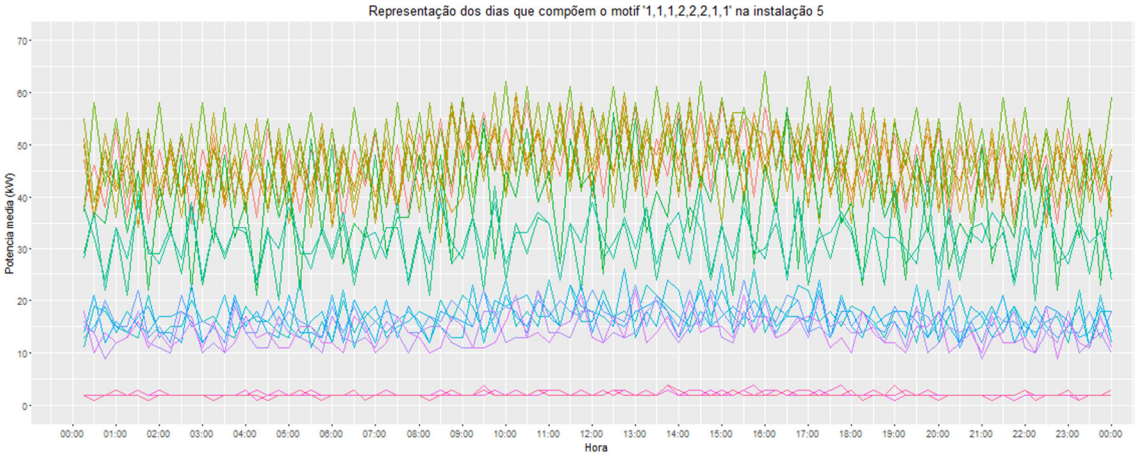


Figura 4.13: Representação gráfica do primeiro motif identificado para a instalação 5 ("1,1,1,2,2,2,1,1")

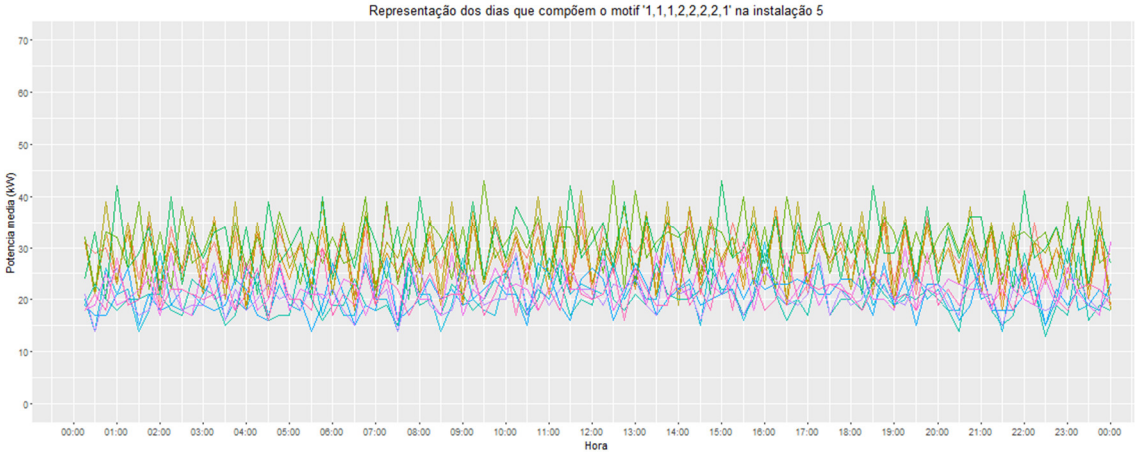


Figura 4.14: Representação gráfica do segundo motif identificado para a instalação 5 ("1,1,1,2,2,2,2,1")

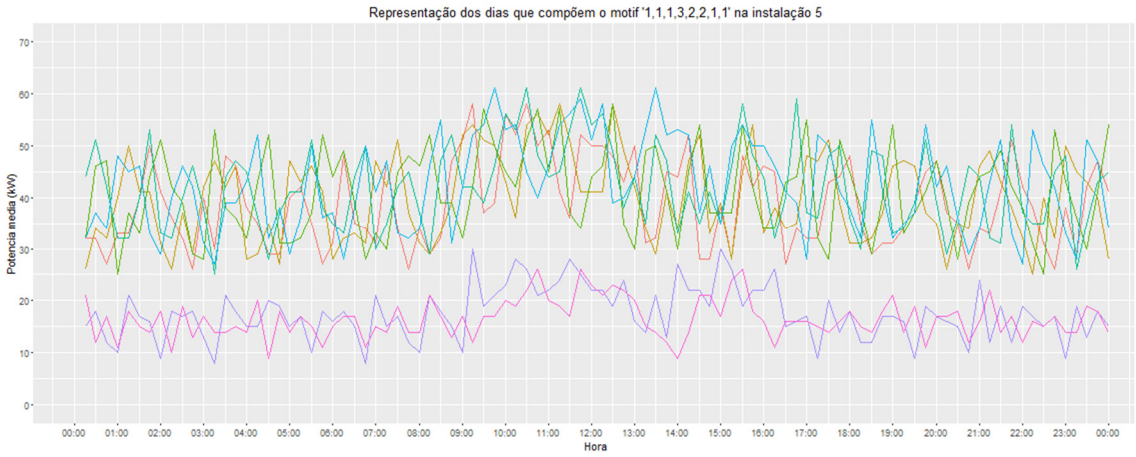


Figura 4.15: Representação gráfica do terceiro motif identificado pela análise mensal para a instalação 5 ("1,1,1,3,2,2,1,1")

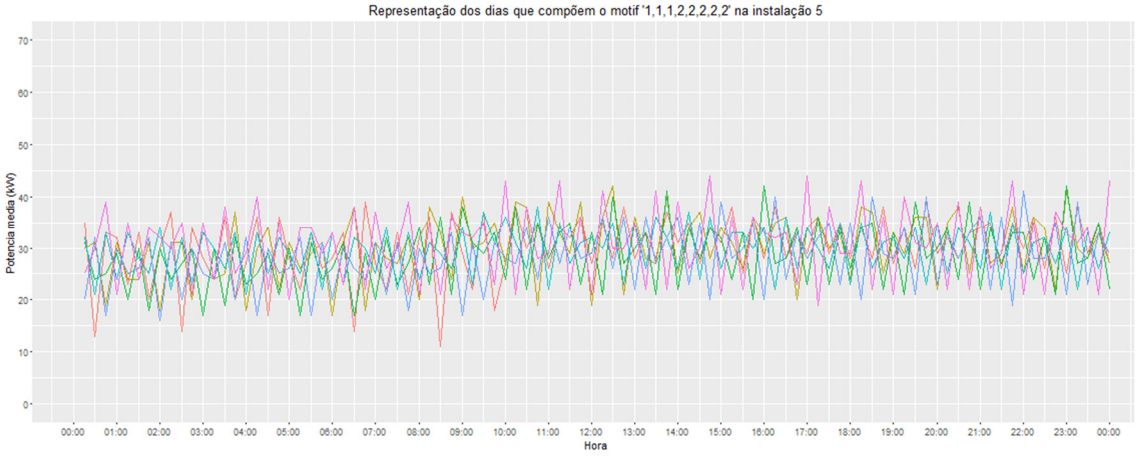


Figura 4.16: Representação gráfica do terceiro motif identificado pela análise semanal para a instalação 5 ("1,1,1,2,2,2,2,2")

Através das representações gráficas dos *motifs* identificados, é possível supor que esta instalação opera continuamente, sem pausas, e que a sua atividade exige variações significativas na potência necessária, visto serem registados valores muito elevados seguidos de valores muito reduzidos. Esta elevada variabilidade na potência média registada ao longo do dia pode exigir uma análise mais detalhada a fim de encontrar os padrões efetivos existentes nesta instalação. Neste sentido, o método utilizado para detetar os *motifs* neste trabalho pode não ser o mais indicado para este tipo de instalações, na medida em que, após normalizar os dados, o algoritmo *MrMotif* ao dividir um dia em oito segmentos, considera os valores médios observados em intervalos de três horas e, como já referido, os valores médios podem ocultar informação relevante.

4.2. Resultados da identificação de *discords*

O objetivo principal do presente trabalho consiste em detetar alterações nos padrões de consumo em instalações de média tensão. Neste sentido, analisar os resultados obtidos com a aplicação do processo detalhado na Secção 3.4. demonstra-se essencial para a concretização deste objetivo. Assim, os resultados obtidos foram compilados e através das tabelas apresentadas nos Anexos 8 e 9 é possível observar os *discords* mais relevantes para cada instalação considerando a análise mensal e semanal, respetivamente. Estes *discords* consistem nos dias que apresentam a maior distância mínima calculada face aos restantes dias em estudo, indicando assim os mais dissemelhantes.

De modo a auxiliar a análise e atestar a existência dos *discords* detetados pode-se recorrer aos Anexos 10 a 14, onde são apresentados as representações gráficas dos diagramas de carga mensais das cinco instalações em estudo, com evidência dos cinco principais *discords* identificados por mês. Nestes gráficos os *discords* são evidenciados através de uma escala hierárquica de cores, onde a cor identificada por "1" representa o primeiro *discord* mais relevante e a cor identificada por "5" representa o quinto *discord* mais relevante identificados no mês em causa. A análise conjunta destes gráficos também permite observar a existência de padrões ao longo do tempo, facilitando a análise de períodos homólogos, por exemplo.

De forma distinta do processo de deteção de *motifs*, no processo de identificação de *discords* cada dia que compõe a base de dados é dividido em seis segmentos de igual dimensão, que correspondem a quatro horas cada. Estes segmentos são transformados em símbolos de acordo com o valor médio observado nestes períodos de quatro horas, podendo assumir um dos sete símbolos existentes conforme o valor médio calculado.

De seguida é apresentada uma breve análise dos resultados obtidos para cada uma das instalações de forma individual, sendo necessário ter em consideração que, para além do intervalo temporal e da potência média registada, numa primeira fase não será considerada qualquer informação adicional que auxilie na explicação dos *discords* encontrados.

Instalação 1

A base de dados relativa à instalação 1 apresenta como *discord* mais relevante o dia 12 de fevereiro de 2015, quinta-feira, que corresponde exatamente à quinta-feira mais

dissemelhante da base de dados. Através da Figura 4.17 é possível comprovar que neste dia houve uma redução muito significativa na potência média, dado que entre as 9:45 e as 17:45 o valor registado foi nulo (igual a zero). Esta redução pode indicar uma avaria em algum equipamento ou o encerramento extraordinário da instalação, visto que é possível deduzir que esta instalação opera de forma contínua e sem interrupções.

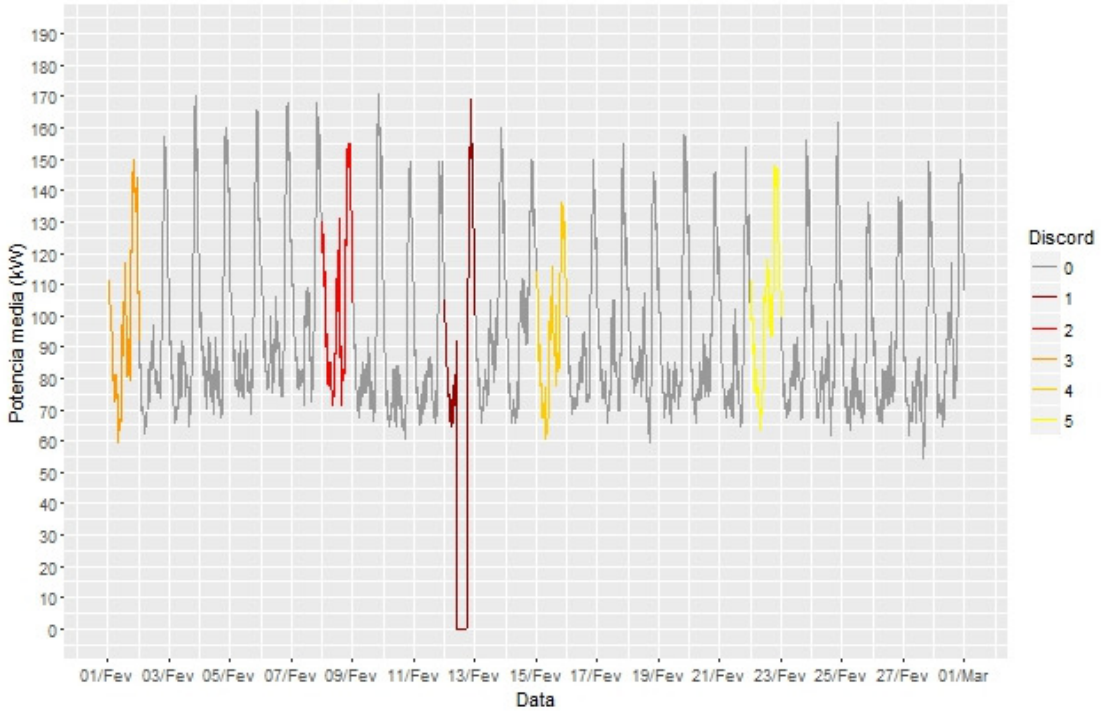


Figura 4.17: Representação gráfica do diagrama de carga de fevereiro de 2015 para a instalação 1

O segundo *discord* mais relevante ocorreu no dia 25 de setembro de 2015, sexta-feira, correspondendo também à sexta-feira mais dissemelhante da base de dados. Neste dia ocorreu uma redução considerável na potência média observada, tendo sido verificado que isso surge logo após um período com três registos em falta, nomeadamente entre as 15:15 e as 15:45. Através da Figura 4.18 é possível notar que a potência média registada nesta instalação passa a apresentar valores praticamente residuais depois desta redução, mantendo o comportamento no mês seguinte. Assim sendo, esta situação pode indicar uma alteração súbita na atividade da instalação ou, no limite, uma falha nos equipamentos de contagem.

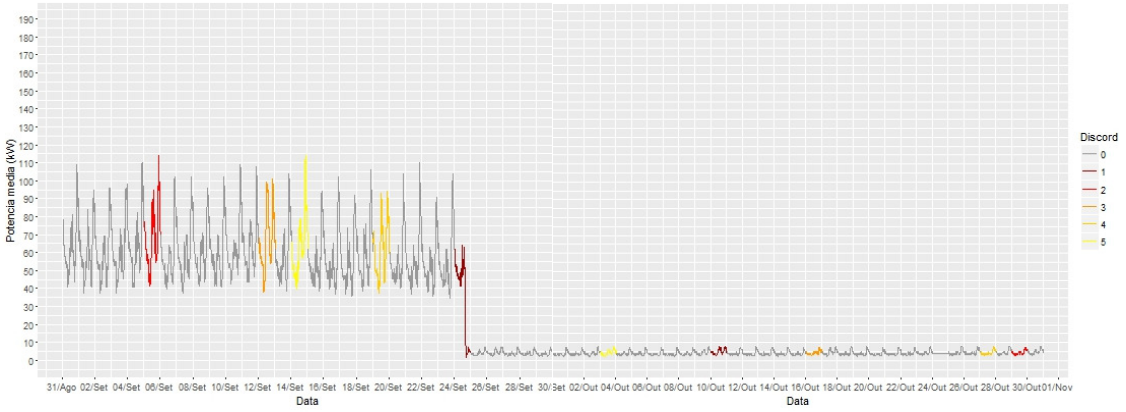


Figura 4.18: Representação gráfica do diagrama de carga de setembro e outubro de 2015 para a instalação 1

Os restantes *discords* identificados, que ocorrem maioritariamente aos domingos e feriados, não apresentam grande expressão, principalmente ao proceder a análise visual das representações gráficas mensais. Estes *discords* foram identificados pelo facto dos dias em causa apresentarem comportamentos diferentes dos restantes dias da base de dados considerada, seja por mês ou por dias da semana.

Instalação 2

Relativamente à base de dados da instalação 2, a análise mensal indica como *discord* mais significativo o dia 26 de fevereiro de 2014, quarta-feira, constando o dia 21 do mesmo mês, sexta-feira, como terceiro *discord* mais importante. Através da representação gráfica presente na Figura 4.19, é possível observar claramente uma alteração no comportamento da instalação na segunda quinzena deste mês. Na segunda parte do mês existe um aumento significativo da potência média registada, deixando de apresentar valores inferiores a 10kW para atingir valores acima dos 30kW. Recorrendo à base de dados original, verifica-se que estes aumentos ocorreram principalmente entre as 9:00 e as 13:45, no dia 26 e entre as 10:45 e as 19:15, no dia 21. Ao observar os dados dos meses anteriores (de outubro de 2013 a janeiro de 2014) verifica-se que a potência média registada é residual, havendo assim indícios de que a instalação não se encontrava em funcionamento, dando início à sua atividade em meados de fevereiro de 2014.

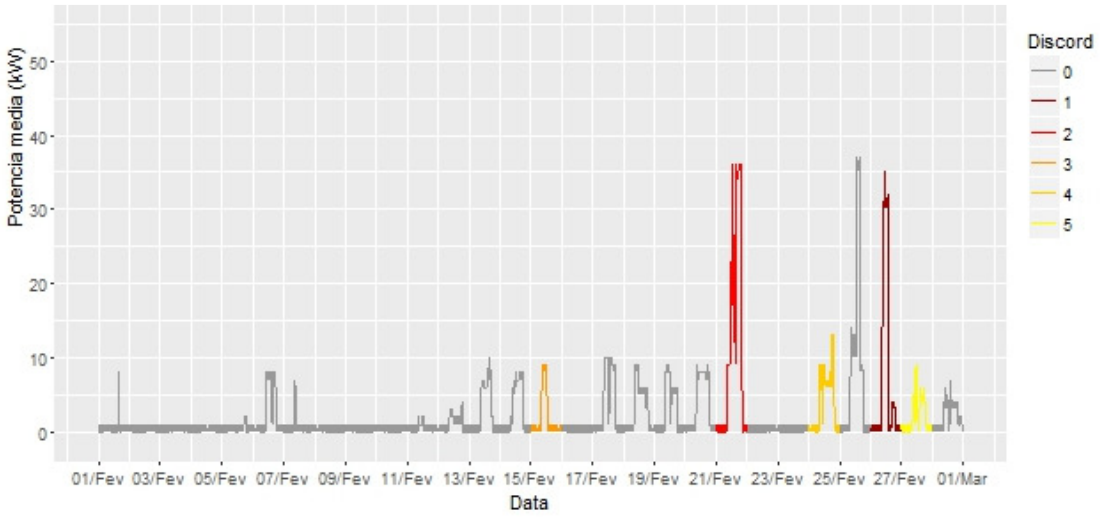


Figura 4.19: Representação gráfica do diagrama de carga de fevereiro de 2014 para a instalação 2

A partir do mês de março de 2014 e conforme referido na análise dos *motifs* identificados, esta instalação passa a demonstrar que desenvolve as suas atividades apenas nos dias úteis, encerrando aos finais de semana, feriados e segunda quinzena de agosto. Esta suposição pode justificar o facto dos restantes *discords* ocorrerem maioritariamente às sextas, devido ao potencial acréscimo de trabalho antes do fim de semana.

Das restantes anomalias identificadas pode-se ainda destacar o quarto e o sexto *discords*, conforme representados nas Figuras 4.20 e 4.21, respetivamente. O quarto *discord* ocorreu no dia 17 de abril de 2014, quinta-feira Santa (Páscoa), sendo verificado um aumento da potência média registada ao serem observados valores acima dos 30kW entre as 14:00 e as 20:00 deste dia. Como já referido, este aumento pode ser justificado por trabalhos adicionais antes do período de encerramento da instalação durante a Páscoa.

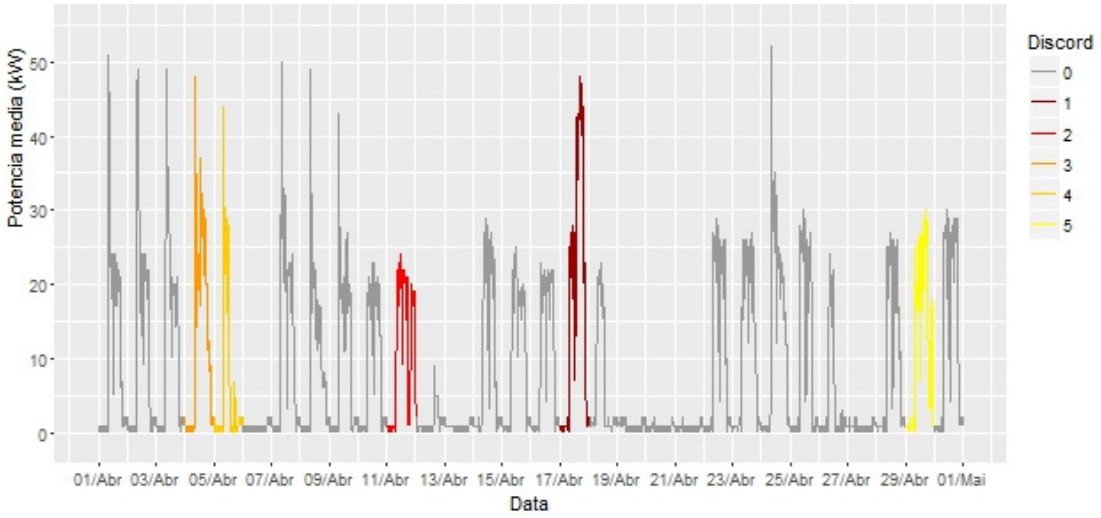


Figura 4.20: Representação gráfica do diagrama de carga de abril de 2014 para a instalação 2

Relativamente ao sexto *discord*, observa-se uma redução na potência média registada, comparativamente aos restantes dias do mês, que pode ser explicada pelo facto do respetivo dia, 17 de fevereiro de 2015, ter sido terça-feira de Carnaval. Com base nos dados presentes no diagrama de carga é possível observar que neste dia, a instalação apenas opera durante o período da manhã, nomeadamente entre as 8:30 e as 12:30, apresentando valores residuais no resto do dia.

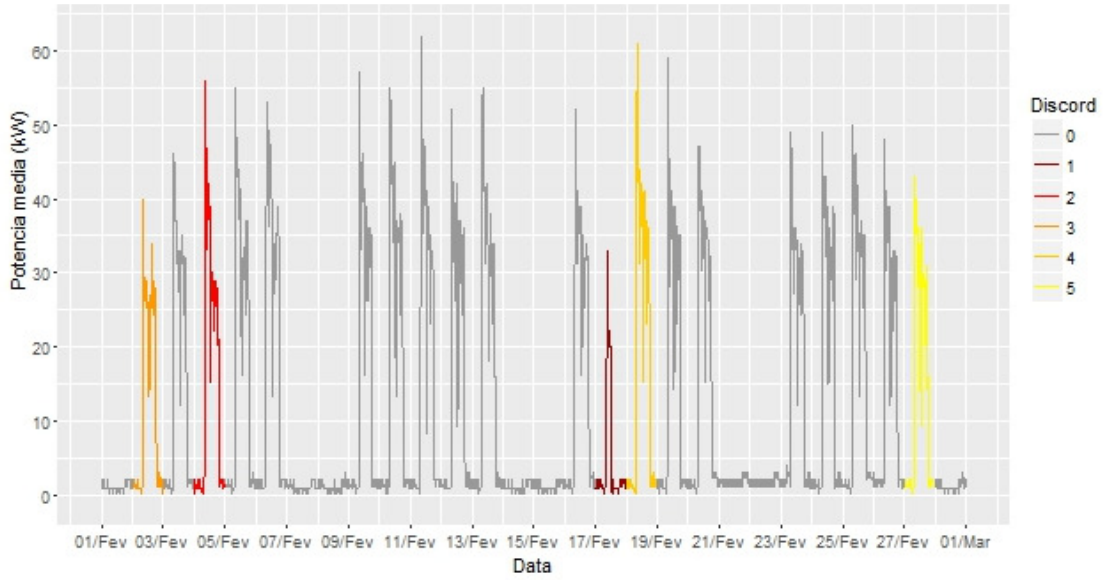


Figura 4.21: Representação gráfica do diagrama de carga de fevereiro de 2015 para a instalação 2

Instalação 3

Com base na análise dos resultados mensais para a base de dados da instalação 3, o *discord* mais significativo ocorre no dia 4 de fevereiro de 2013, segunda-feira. Contudo, ao considerar os resultados da análise por dias da semana verifica-se que a segunda-feira mais dissemelhante ocorre quinze dias depois, no dia 18 de fevereiro de 2013. Através da representação gráfica do diagrama de carga deste mês, presente na Figura 4.22, verifica-se que o dia 18 aparenta maiores diferenças face aos restantes dias do mês do que o dia 4 ao apresentar uma potência média registada muito elevada. Neste sentido, não é possível garantir que o dia 4 deva ser efetivamente considerado como *discord*, apesar do algoritmo o ter identificado como tal.

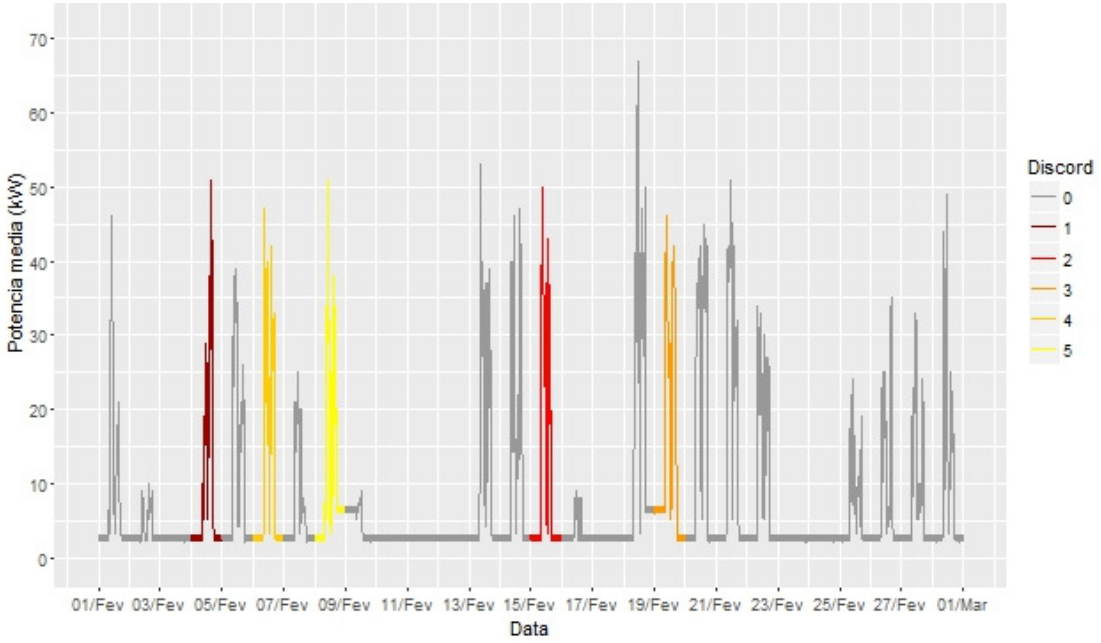


Figura 4.22: Representação gráfica do diagrama de carga de fevereiro de 2013 para a instalação 3

O segundo e o terceiro *discords* identificados são os dias 22 de abril de 2015 e 16 de maio de 2013, respetivamente. Estes demonstram uma maior probabilidade de serem dias anómalos ao serem observadas as representações gráficas dos diagramas de carga dos meses em questão. Relativamente ao mês de abril de 2015 (Figura 4.23), verifica-se que tanto o dia 22, quarta-feira, como o dia 23, quinta-feira, estão evidenciados como *discords* ao apresentarem períodos com valores acima dos 40kW para a potência média em ambos os dias. Este aumento pode ser justificado por trabalhos adicionais antes do feriado do Dia da Liberdade (25 de abril) ou como forma de compensar o período da Páscoa, que teve lugar no início do mês, quando a instalação aparenta ter encerrado.

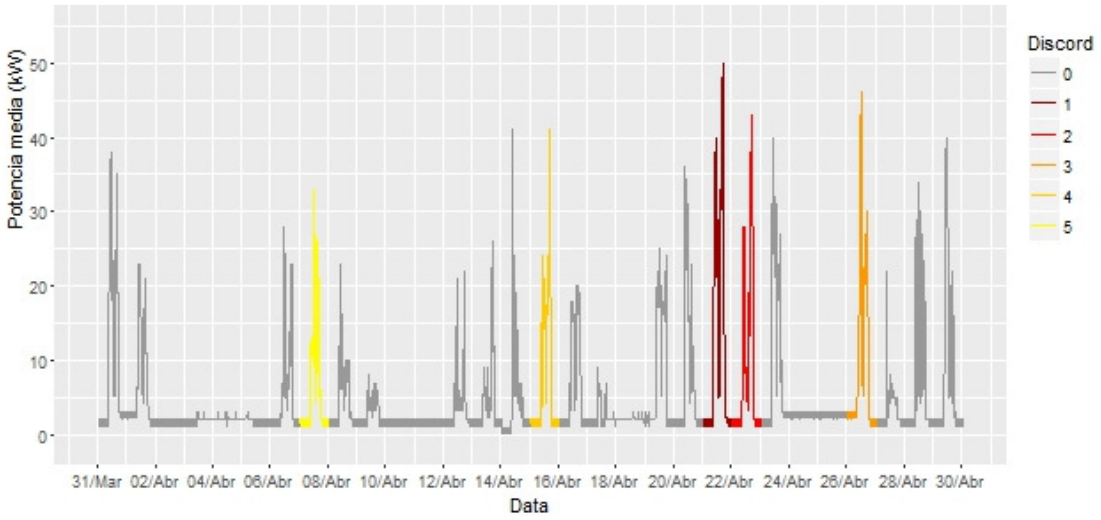


Figura 4.23: Representação gráfica do diagrama de carga de abril de 2015 para a instalação 3

No que diz respeito ao terceiro *discord*, representado na Figura 4.24, o dia 16 de maio de 2013 apresenta-se também como a quinta-feira mais dissemelhante da base de dados. Neste dia, a potência média registada chega a ultrapassar os 50kW entre as 15:45 e as 16:30, o que pode ser explicado pelo facto do dia anterior apresentar valores muito reduzidos, com uma média agregada diária inferior a 4kW.

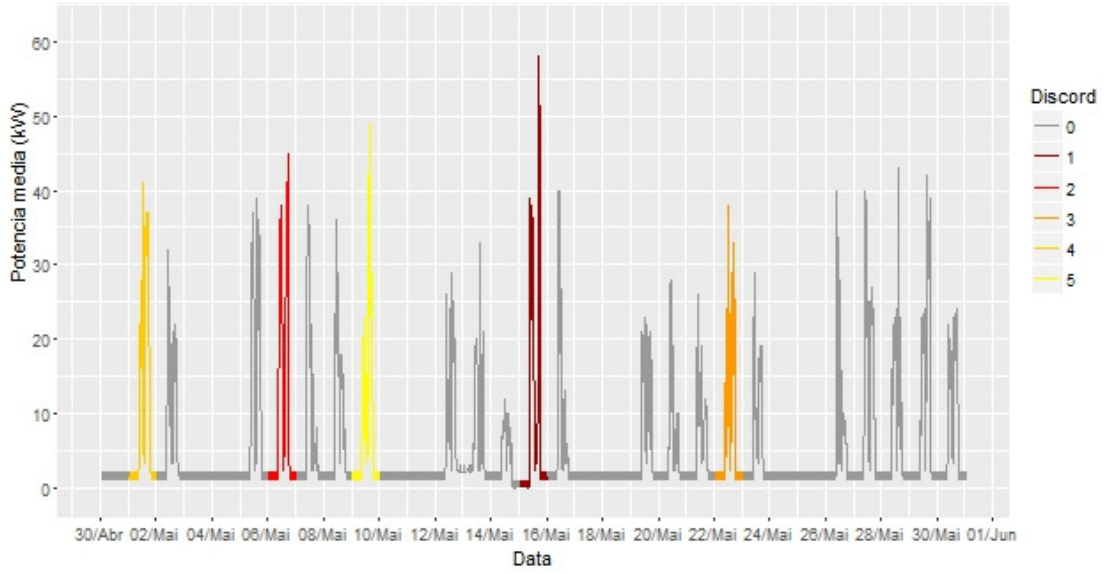


Figura 4.24: Representação gráfica do diagrama de carga de maio de 2013 para a instalação 3

Instalação 4

No que diz respeito à base de dados da instalação 4, o *discord* mais significativo ocorre no dia 17 de setembro de 2014, identificado também como a quarta-feira mais dissemelhante da base de dados. A Figura 4.25 apresenta a representação gráfica do diagrama de carga de setembro de 2014 sendo possível validar que no dia 17 houve uma redução significativa na potência média registada. Recorrendo aos dados inicialmente disponibilizados, é possível verificar que no intervalo entre as 14:30 e as 14:45 deste dia a potência média foi nula. No intervalo seguinte, entre as 14:45 e as 15:00, existe uma observação em falta que foi corrigida no processo de preparação da base de dados, passando a apresentar um valor estimado para a potência deste intervalo. Esta situação pode indiciar uma falha momentânea na rede elétrica ao dar origem a uma redução tão brusca na potência média registada.

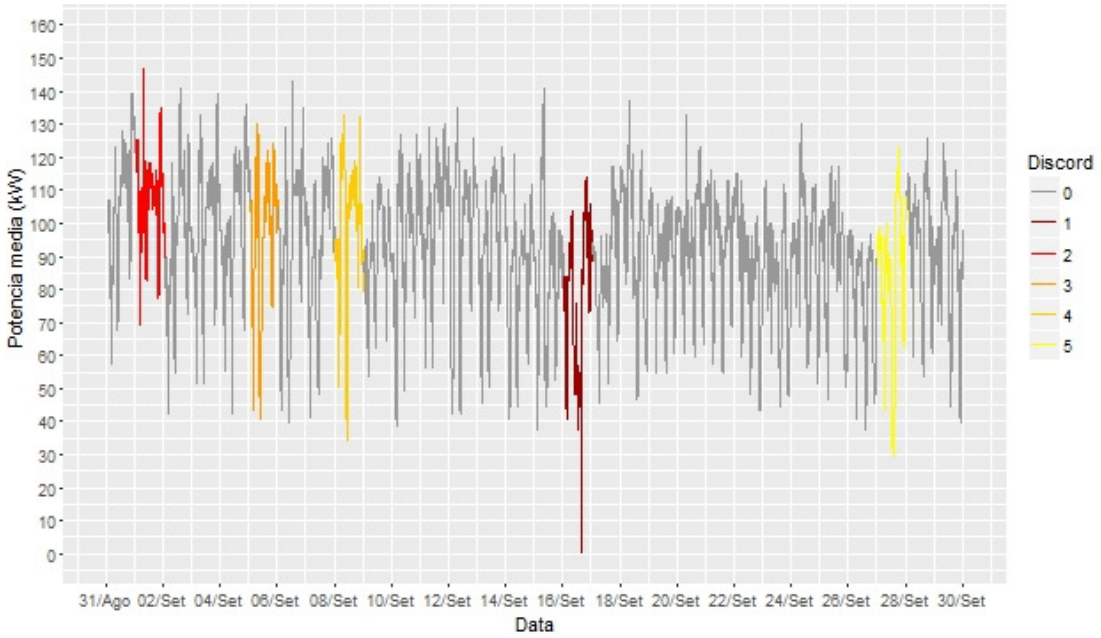


Figura 4.25: Representação gráfica do diagrama de carga de setembro de 2014 para a instalação 4

Relativamente ao segundo *discord* identificado através da análise mensal, verifica-se que o dia 26 de junho de 2013 apresenta 31 observações em falta. Neste sentido, 32% dos dados considerados para este dia foram estimados para o desenvolvimento deste trabalho. Dado que esta instalação apresenta uma grande variação na potência média registada ao longo do dia, estimar um valor constante para todo o intervalo com valores em falta faz com que o algoritmo detete uma irregularidade no padrão diário. Através da Figura 4.26 é possível comprovar a elevada variabilidade observada ao longo dos dias, bem como a utilização de um valor constante de 46kW para o intervalo entre as 9:45 e as 17:15. Assim sendo, é provável que tenha ocorrido uma avaria no equipamento de telecontagem neste intervalo ou uma interrupção extraordinária na atividade da instalação.

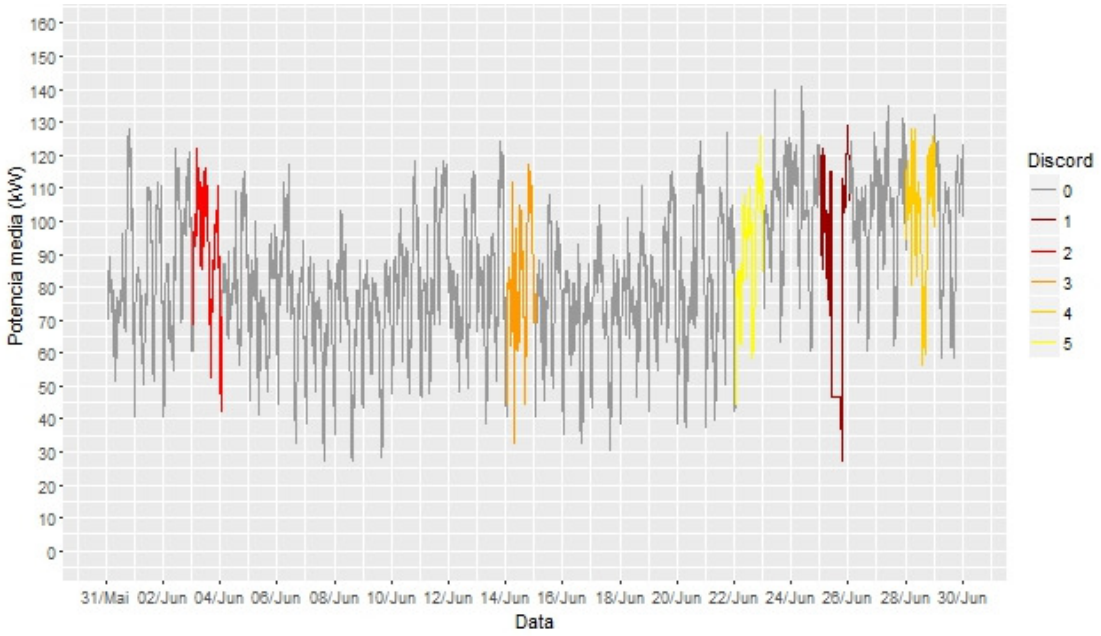


Figura 4.26: Representação gráfica do diagrama de carga de junho de 2013 para a instalação 4

Apesar de identificados como nono e décimo *discords*, os dias 22 e 9 de junho de 2015, respetivamente, motivam uma segunda análise ao conjugar a informação presente na tabela do Anexo 8 com a visualização da representação gráfica do diagrama de carga do referido mês (Figura 4.27). Em ambos os dias, muitos intervalos apresentam valores registados para a potência média acima de 100kW. Contudo, a seguir ao dia 9, terça-feira e véspera de feriado (10 de junho - Dia de Portugal, de Camões e das Comunidades Portuguesas), verifica-se uma alteração no padrão verificado até ao momento. Entre os dias 10 e 15 de junho de 2015, a potência média registada agregada do período apresenta valores na ordem dos 65kW, com um valor máximo de 115kW entre as 16:45 e as 17:00 do dia 10. Assim sendo, esta alteração no padrão de consumo durante praticamente uma semana pode indicar que, apesar desta instalação operar continuamente (não encerrando em feriados ou finais de semanas), pode ter ocorrido uma diminuição da atividade da instalação que deu origem a menores consumos de energia elétrica neste período.

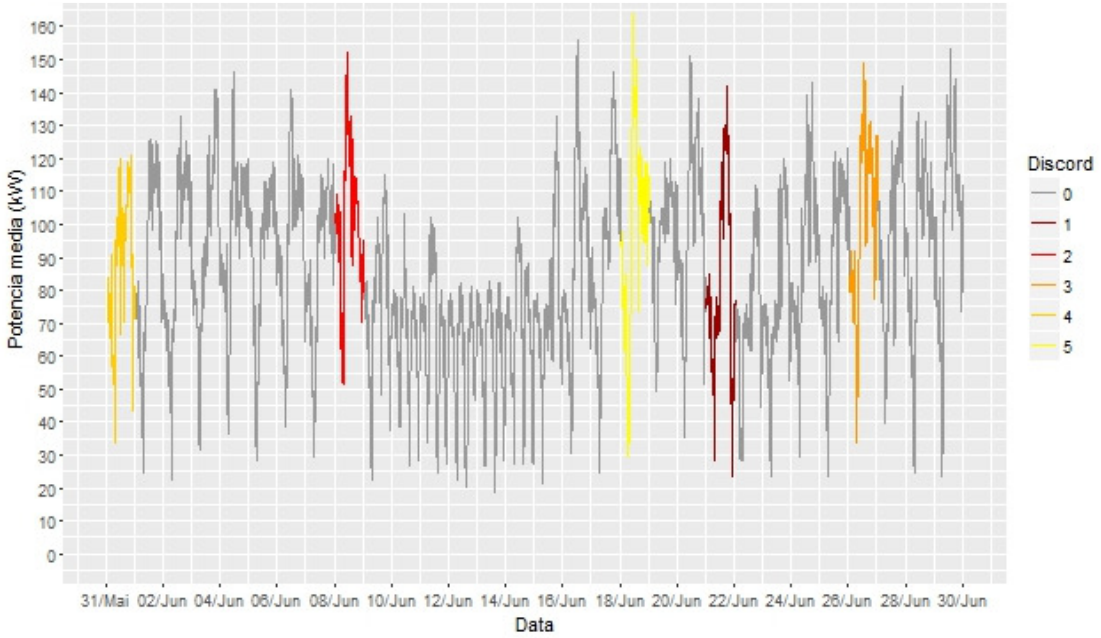


Figura 4.27: Representação gráfica do diagrama de carga de junho de 2015 para a instalação 4

Instalação 5

A base de dados da instalação 5 apresenta como *discord* mais significativo o dia 15 de novembro de 2013, sexta-feira, identificado tanto na análise mensal como na análise por dias da semana. A Figura 4.28 permite observar a representação gráfica do diagrama de carga do referido mês, sendo de notar que o dia seguinte, 16 de novembro, também consta na tabela do Anexo 8 como quarto *discord* mais relevante nesta base de dados.

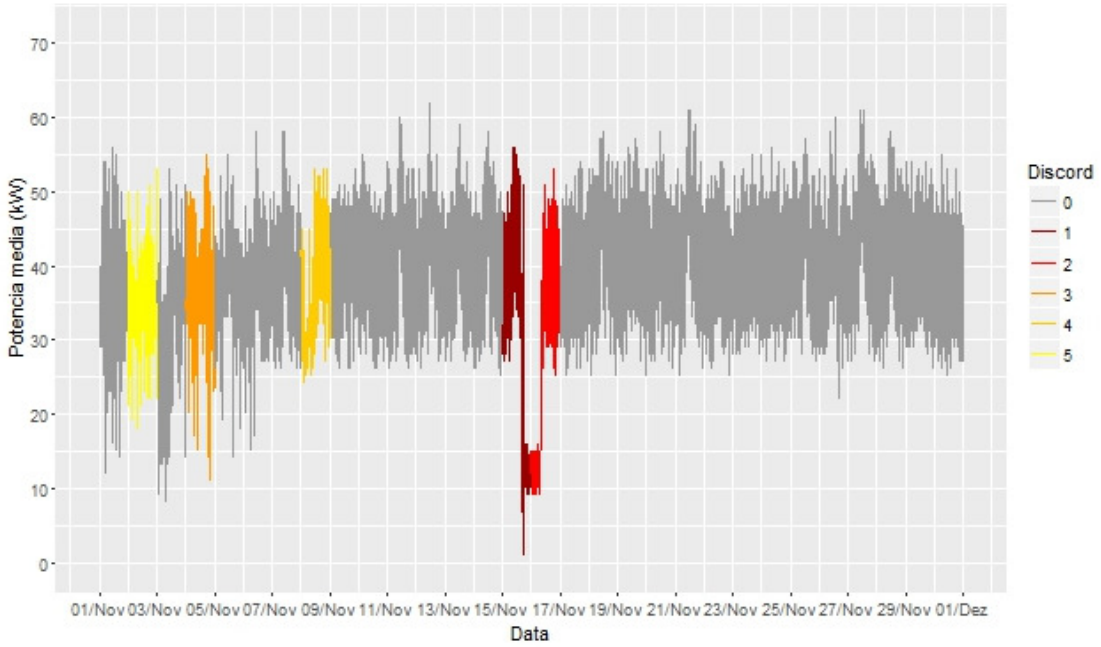


Figura 4.28: Representação gráfica do diagrama de carga de novembro de 2013 para a instalação 5

Através da visualização deste diagrama de carga é possível observar que os dois dias apresentam um período de reduzida potência média registada. Com o auxílio da base de dados original, é possível comprovar que entre as 18:00 do dia 15 e as 7:30 do dia 16, a potência média apresenta valores muito inferiores aos registados nos restantes dias do mês, com uma média agregada de aproximadamente 13kW neste intervalo. Dado que estavam em falta cinco observações no intervalo entre as 16:00 e as 17:00 no dia 15, é possível pressupor que ocorreu uma falha no fornecimento de energia neste período. Esta falha pode ter originado uma paragem extraordinária na atividade, exigindo a redefinição de processos, sendo esta uma possível justificação para o padrão anómalo identificado.

A Figura 4.29 permite visualizar a representação gráfica do diagrama de carga onde constam o segundo *discord* identificado, que ocorre no dia 6 de janeiro de 2014, correspondente também à segunda-feira mais dissemelhante da base de dados. Neste dia verifica-se uma situação semelhante ao exposto no primeiro *discord*, existindo uma redução da potência média registada entre dois dias consecutivos, o que pode estar relacionado com a própria atividade da instalação visto repetir-se ao longo do tempo. O princípio do dia 6 apresenta valores médios agregados de aproximadamente 15kW. Contudo, a seguir às 8:00 apura-se um aumento considerável da potência, atingindo valores acima de 60kW para a potência média.

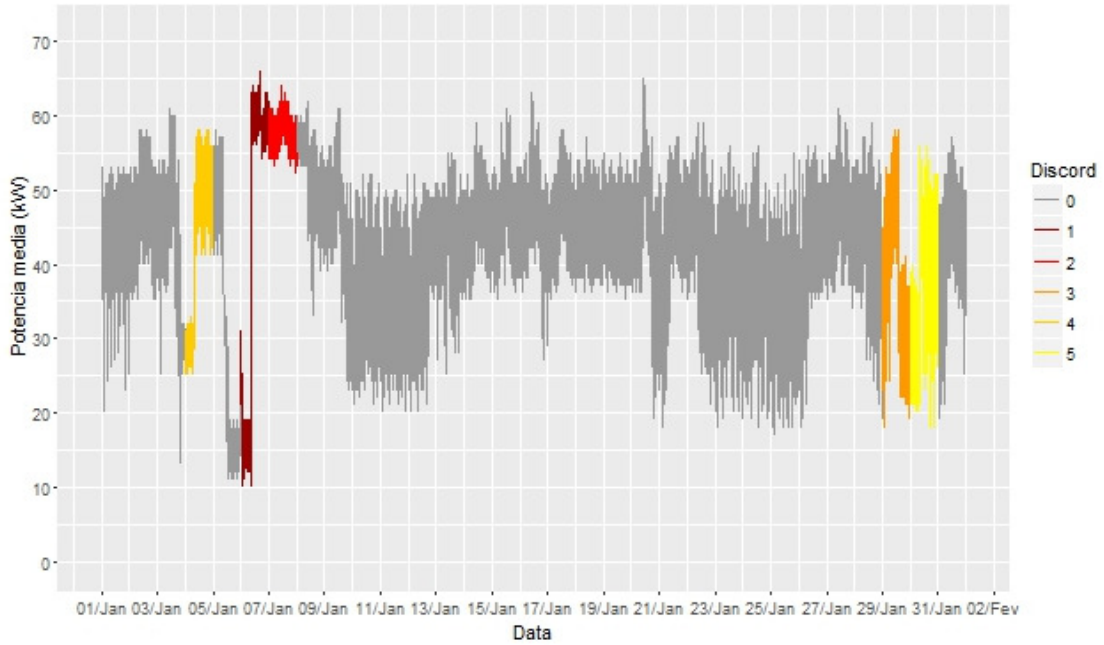


Figura 4.29: Representação gráfica do diagrama de carga de janeiro de 2014 para a instalação 5

A Figura 4.30 apresenta o oitavo e o décimo *discords* detetados relativos aos dias 25 e 9 de julho de 2014, respetivamente. No dia 25, sexta-feira, a potência média registada passa de valores médios agregados de aproximadamente 6kW, entre as 0:00 e as 15:15, para cerca de 30kW para o resto do dia. Contudo, no dia 9, quarta-feira, a variação acontece no sentido oposto ao iniciar o dia com potências mais elevadas, a rondar os 30kW, que vão diminuindo com o passar das horas, atingindo um valor médio agregado de aproximadamente 9kW depois das 15:00. Conjugando esta informação com a visualização do diagrama de carga é possível supor que a instalação sofreu uma redução nos consumos devido a um período de férias entre estes dois dias, não ocorrendo o encerramento da instalação mas uma redução na atividade geral.

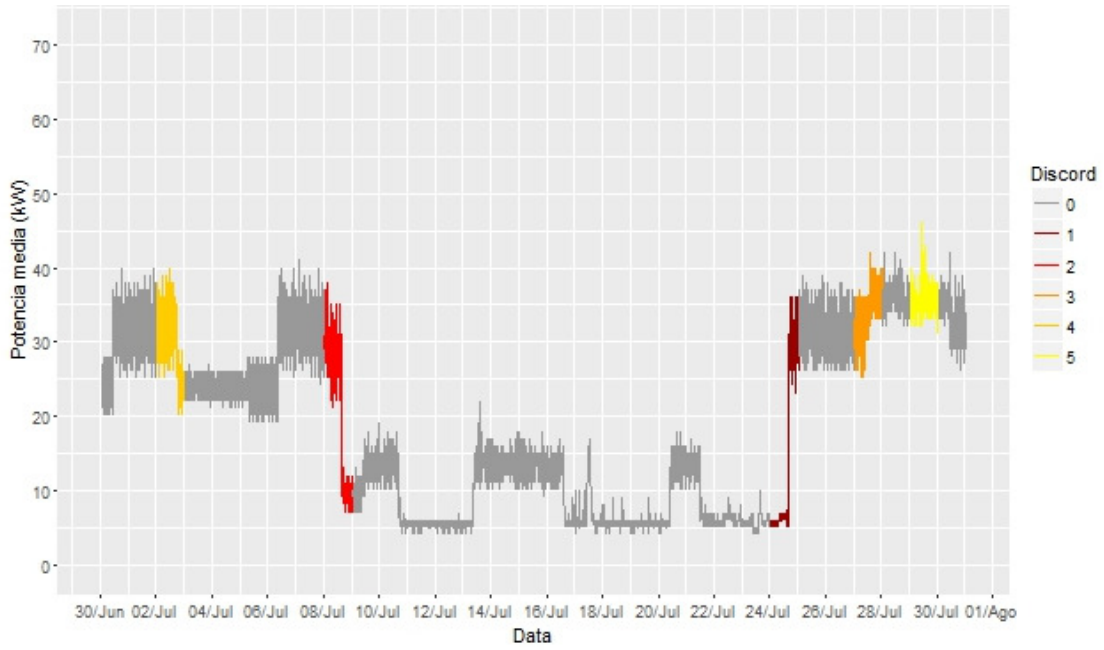


Figura 4.30: Representação gráfica do diagrama de carga de julho de 2014 para a instalação 5

4.3. Resultados agregados e avaliação

Após analisar individualmente os resultados obtidos durante os processos de identificação de *motifs* e de *discords* para cada uma das instalações, revela-se interessante agregar as conclusões extraídas anteriormente. Esta agregação permite não apenas validar que os *discords* e os *motifs* não se sobrepõem, como possibilita uma melhor compreensão dos diferentes padrões verificados nestas instalações.

As análises realizadas nas secções anteriores consideraram apenas a informação disponível nos diagramas de carga, ou seja, o horizonte temporal e a potência média registada neste período. Contudo, para proceder a uma interpretação dos resultados, a

EDPD auxiliou na análise final, procurando validar os resultados obtidos e encontrar justificações para as conclusões extraídas com base em alguns dados adicionais que foram disponibilizados. Neste sentido, nesta secção são apresentadas as observações finais para cada uma das instalações, com a agregação da informação extraída através da identificação de *motifs* e de *discords*, conjugada com informação adicional sobre a realidade de cada instalação.

Instalação 1

A instalação 1 não consiste numa instalação de média tensão, como seria de esperar, tratando-se afinal de um posto de transformação de distribuição (PTD). Ao representar um PTD, a informação presente no diagrama de carga desta instalação contém informação agregada de um conjunto de 260 clientes de baixa tensão, que são maioritariamente residências particulares.

Esta informação adicional permite compreender o motivo dos *motifs* indicarem que os consumos aumentam a seguir as 17:00, atingindo valores máximos por volta das 21:00, visto ser este o padrão habitual nas instalações residenciais. Neste sentido, apesar de não se tratar de uma instalação de média tensão, é interessante verificar que os algoritmos utilizados conseguem lidar corretamente com os dados, identificando verdadeiros padrões frequentes neste género de instalação.

Tendo em conta a informação disponível nesta fase, não foi possível encontrar justificação para os *discords* identificados nesta instalação, nomeadamente nos dias 12 de fevereiro de 2015 e 25 de setembro de 2015. De fato, a informação adicional disponibilizada pela EDPD indicou que não houve qualquer tipo de intervenção técnica no período referido, assim como foi comprovado que o PTD em causa continua ativo. Contudo, a EDPD referiu que as situações identificadas deverão ser posteriormente analisadas em pormenor, de modo a compreender o motivo de tais anomalias.

Instalação 2

A informação disponibilizada para a instalação 2 refere que houve um período de transição entre dois clientes diferentes, ou seja, em 11 de março de 2014 observa-se uma alteração na titularidade do contrato de fornecimento de energia. Neste sentido, a potência média registada no início do diagrama de carga diz respeito a uma fábrica de têxteis mas,

após a referida transição, os valores indicados passam a ser relativos a uma sociedade de investimentos. Pode-se assim concluir que houve uma alteração significativa na atividade desenvolvida nesta instalação ao passar de uma unidade fabril para um escritório com atividades maioritariamente administrativas.

A análise dos *discords* identificados indica que entre os meses de outubro de 2013 e janeiro de 2014 a potência média registada apresentou valores residuais, o que pode ser justificado pelo encerramento da fábrica têxtil. Apesar da titularidade apenas ter ocorrido oficialmente em 11 de março, durante o mês de fevereiro podem ter sido desenvolvidos trabalhos no sentido de efetuar as alterações necessárias na estrutura da instalação para dar início à atividade da sociedade de investimentos. Assim sendo, esta transição entre clientes pode explicar os *discords* identificados em fevereiro de 2014 pois ocorreu efetivamente uma alteração no padrão de consumo da instalação.

Com base na análise de *motifs* e no conhecimento da atividade desenvolvida, depreende-se que esta instalação desenvolve a sua atividade nos dias úteis, encerrando aos feriados, fins de semana e na segunda quinzena de agosto para férias. O horário de trabalho habitual é das 8:00 as 18:00, nos dias úteis, com intervalo para almoço entre as 12:00 e as 14:00 e realizando pausas pontuais às 10:00 e às 16:00. Este pode ser considerado um padrão habitual para uma atividade administrativa deste género, sendo os *discords* maioritariamente observados às sextas-feiras devido a um potencial acréscimo de trabalho para concluir o trabalho da semana.

Instalação 3

A instalação 3 diz respeito a uma serração de madeira e cortiça que, de acordo com a análise dos *motifs* identificados anteriormente, opera habitualmente nos dias úteis entre as 8:00 e as 17:30, encerrando para o período de almoço entre as 12:00 e as 13:00. Este padrão pode ser considerado habitual numa unidade fabril que não opera por turnos de forma contínua.

Relativamente aos *discords*, os dias identificados como anómalos correspondem maioritariamente a dias de maior potência média registada. Este aumento de potência pode indicar trabalhos adicionais em determinados períodos para compensar alguns dias de paragem (por exemplo, feriados) ou para garantir o cumprimento de prazos de entrega de material. Neste sentido, o *discord* identificado no dia 22 de abril de 2015 pode ser

justificado pelo facto desta instalação aparentemente ter reduzido a produção ou mesmo encerrado durante o período da Páscoa, que ocorreu no início do referido mês. No que diz respeito ao dia 16 de maio de 2013, também identificado como *discord*, o aumento do consumo pode ser justificado por questões técnicas internas. Por exemplo, dado que nos dias anteriores a potência média registada apresenta valores inferiores, e existindo uma diminuição progressiva ao longo da semana, é possível supor que houve uma quebra no *stock* de matérias-primas nesta semana. Neste caso, ao repor o *stock* de matérias-primas no dia 16, quinta-feira, pode ser retomada a produção, aumentando a atividade para compensar o período em que não havia matéria-prima suficiente.

Apesar das justificações acima mencionadas poderem ser apenas suposições, a identificação de padrões anómalos pode, em alguns casos, indiciar efetivas situações de fraude que são de extrema importância para a EDPD. Os padrões frequentes também auxiliam nesta análise pois permitem compreender o que é normal e consequentemente esperado numa dada instalação. Assim, a presença de alterações significativas e sem causa aparente nos *motifs* ao longo do tempo podem indicar a existência de uma situação fraudulenta, sendo do interesse da EDPD identificar e analisar estas mesmas alterações.

Neste sentido, pode-se referir que em termos de padrões frequentes esta instalação encerra em determinados períodos do ano, nomeadamente, para férias de verão (agosto ou setembro) e na altura da Páscoa.

Instalação 4

A instalação 4 corresponde a uma instalação agropecuária, não existindo informação complementar sobre a atividade exata desta instalação. No entanto, de acordo com as análises realizadas anteriormente, a potência média registada ao longo de um dia de trabalho apresenta grande variabilidade, dificultando a identificação de dias semelhantes. A atividade desenvolvida por esta instalação, ao poder estar sujeita a condições naturais pouco controláveis, pode exigir um maior consumo de energia de modo a combater situações imprevistas. Do mesmo modo, a definição do processo produtivo diário pode exigir diferentes potências consoante as fases do processo produtivo e os equipamentos utilizados. Apesar dos dias serem habitualmente atípicos, observa-se a existência de um padrão ao longo do tempo ao verificar que a potência média registada apresenta habitualmente valores acima dos 50kW, independentemente do período em análise.

Relativamente ao *discord* identificado no dia 17 de setembro de 2014, este dia apresentou uma redução inesperada na potência média registada entre as 14:30 e as 15:00. Uma vez que a EDPD não dispõe de informações adicionais que indiquem a efetiva ocorrência de situações anómalas nesta instalação, a redução brusca verificada no consumo pode estar relacionada com uma situação interna da própria instalação.

Instalação 5

Com base na informação adicional disponibilizada, sabe-se que a instalação 5 diz respeito a uma empresa de tratamento de resíduos sólidos urbanos. Esta instalação realiza o tratamento do lixo da região Centro do país, cuidando da recolha, triagem e valorização ou eliminação dos resíduos.

Na análise dos resultados obtidos para a identificação de *motifs* foi referido que esta instalação apresenta uma grande variabilidade nos valores da potência média registada ao longo do mesmo dia. Esta variação pode estar relacionada com a atividade da empresa, que pode exigir alterações no consumo de energia de acordo com a tarefa a decorrer.

A análise de *discords* identifica períodos, como o primeiro *discord* detetado em 15 de novembro de 2013, onde um período de menor consumo é seguido de um aumento significativo nos valores da potência média registada. Esta situação pode ser justificada por um período de recolha dos resíduos, que pode exigir menor consumo de energia, para de seguida serem realizados a triagem e o tratamento final destes, que pode consistir num trabalho mais intensivo em termos de consumo energético.

Tendo em consideração a atividade desenvolvida por esta instalação, é compreensível que a mesma opere de forma contínua, não ocorrendo períodos de encerramento. Contudo, verificam-se períodos de menor atividade, com consumos praticamente residuais, entre final de julho e início de agosto de 2015. Dado que o encerramento da instalação para férias de verão não ocorre no ano anterior, com base na informação histórica disponível, esta situação não pode ser considerada como habitual. Assim sendo, esta redução no consumo pode estar relacionada com alguma situação extraordinária que ocorreu na instalação.

4.4. Considerações finais

Ao comparar os dias identificados como *discords* com os dias que compõem os *motifs* para cada uma das instalações foi possível validar que não existem situações em que o mesmo dia é apontado como *motif* e como *discord*, estando assim parcialmente validados os métodos utilizados. Após as análises acima indicadas pode-se concluir que os processos para deteção de *motifs* e de *discords* conseguem efetivamente identificar os padrões verificados nos consumos, considerando a representação simbólica dos dias em palavras e os valores registados na base de dados original.

Agregando a análise dos resultados obtidos inicialmente com alguma informação adicional disponibilizada pela EDPD, a fim de compreender as situações identificadas, foi possível validar a veracidade dos padrões identificados. Apesar de não ser possível compreender ao pormenor a atividade desenvolvida pelas diferentes instalações e justificar detalhadamente os padrões identificados (isso exigiria contactar diretamente com os responsáveis pelas instalações e obter informação interna), foi possível retirar conclusões admissíveis pelo senso comum e pela própria EDPD. Deste modo, considera-se que os objetivos propostos para este projeto foram devidamente alcançados ao serem identificados os padrões frequentes e anómalos em diferentes instalações de média tensão, apresentando-se como uma boa análise exploratória dos dados disponibilizados.

Capítulo 5. Conclusão

Dada a relevância do tema em análise, conforme apresentado no Capítulo 1, e a existência de diversas técnicas capazes de analisar os padrões existentes em série temporais, como demonstrado no Capítulo 2, considera-se que o presente trabalho representa uma mais-valia tanto para o aluno como para a EDPD que cedeu os dados para análise. A identificação de alterações nos padrões de consumo de energia em instalações de média tensão permite que a EDPD consiga analisar os dados de consumo dos seus clientes, identificando situações anómalas e possibilitando uma tomada de decisão que pode basear-se em factos analíticos.

A aplicação prática dos conceitos estudados, conforme descrito no Capítulo 3, permitiu obter os resultados esperados, sendo efetivamente identificados os padrões frequentes e anómalos existentes em cada instalação. A justificação detalhada dos resultados, apresentados no Capítulo 4, pode exigir informação adicional do domínio das próprias instalações em estudo. Contudo, a deteção da existência de tais padrões pode alertar a EDPD para a ocorrência de situações que podem indiciar a ocorrência de fraudes ou anomalias.

Neste sentido, o trabalho desenvolvido ao longo deste projeto apresentou-se como um desafio aliciante e motivador, permitindo uma evolução positiva no que diz respeito aos conhecimentos teórico-práticos do aluno em matérias de extração de conhecimento de dados, nomeadamente em séries temporais. De modo semelhante, este trabalho deu a conhecer à EDPD o potencial existente nos dados que reúne, demonstrando que através da aplicação de diferentes técnicas de extração de conhecimento de dados, é possível obter informação útil para o desenvolvimento do seu negócio.

5.1. Discussão

Dada a elevada quantidade de informação recolhida diariamente pela EDPD no decorrer da sua atividade, encontrar métodos capazes de analisar os dados e extrair conhecimento útil dos mesmos demonstra-se de extrema importância. Assim sendo, com o desenvolvimento deste projeto foi possível compreender os diferentes métodos existentes e capazes de identificar padrões em séries temporais, assim como verificar que os mesmos podem ser aplicados aos dados existentes com sucesso.

Neste sentido, o presente trabalho permitiu identificar um conjunto de padrões frequentes e anómalos para as cinco instalações em estudo, tendo por base apenas a informação da potência média registada num determinado horizonte temporal. De referir que o horizonte temporal em análise varia de acordo com a base de dados em estudo, dado que cada uma contém informação sobre um determinado período, conforme referido anteriormente. Os resultados obtidos apresentam-se como uma boa análise exploratória dos dados, demonstrando o potencial existente nos dados e nas técnicas aplicadas para o reconhecimento de padrões.

5.2. Limitações e projetos futuros

A utilização da ferramenta *iMotifs*, para a identificação dos *motifs*, apresenta uma limitação na medida em que esta ferramenta não permite exportar os resultados obtidos de forma automática. Esta exportação da informação poderia auxiliar na análise posterior dos resultados, permitindo utilizar diferentes ferramentas e realizar diversas operações a fim de encontrar relações e justificações para os padrões identificados. Pelo facto de não permitir uma extração automática dos resultados, acaba por dificultar o processo de análise dos padrões encontrados ao exigir demasiado tempo para executar a tarefa mecânica de extração manual dos resultados para outro formato (por exemplo, *MS Excel*).
O método utilizado pela ferramenta *iMotifs* demonstra apresentar bons resultados ao detetar efetivamente os padrões frequentes existentes, apesar do tempo e trabalho exigidos para a extração da informação. Neste sentido, um potencial projeto futuro consiste em desenvolver uma ferramenta única que consiga detetar os *motifs* e os *discords* presentes numa determinada série temporal, recorrendo aos métodos expostos por cada um dos algoritmos utilizados (*MrMotif* e *HOT SAX*).
Assim sendo, poderiam ser poupados esforços ao realizar os mesmos processos até à fase de representação simbólica dos dados, optando para tal pelo *SAX* ou pelo *iSAX*, por exemplo, para de seguida permitir a identificação de padrões frequentes ou anómalos, conforme o interesse. Dada a natureza da informação disponível através dos diagramas de carga, apresenta-se também interessante a hipótese desta ferramenta conseguir trabalhar com dados obtidos de forma contínua (*streaming data*), avaliando os padrões em tempo real.

Referências bibliográficas

Agrawal, R., Faloutsos, C. & Swami, A., 1993. Efficient similarity search in sequence databases. In *Proceedings of the 4th International Conference on Foundations of Data Organization and Algorithms*. pp. 69–84.

Amador, J., 2010. Produção e consumo de energia em Portugal: Factos Estilizados. *Boletim Económico | Banco de Portugal*, pp.71–86.

Antunes, C.M. & Oliveira, A.L., 2001. Temporal Data Mining: an overview. In *Proceedings of the KDD Workshop on Temporal Data Mining*. pp. 1–13.

Bu, Y. et al., 2007. WAT: Finding top-k discords in time series database. In *Proceedings of the 2007 SIAM International Conference on Data Mining (SDM'07)*. pp. 449–454.

Castro, N. & Azevedo, P., 2010. Multiresolution motif discovery in time series. In *Proceedings of the 2010 SIAM International Conference on Data Mining*. pp. 665–676.

Chiu, B., Keogh, E. & Lonardi, S., 2003. Probabilistic discovery of time series motifs. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD 03*. pp. 493–498.

Fu, A.W. et al., 2006. Finding time series discords based on haar transform. In *Proceedings of the 2nd International Conference Advanced Data Mining and Applications*. pp. 31–41.

Fu, T., 2011. A review on time series data mining. *Engineering Applications of Artificial Intelligence*, 24(1), pp.164–181.

Gama, J. et al., 2012. *Extração de Conhecimento Dados - Data Mining* 1ª Edição. E. Sílabo, ed., Lisboa.

Hall, M. et al., 2009. The WEKA Data Mining Software: An Update. *SIGKDD Explorations*, 11(1).

Hofmann, M. & Klinkenberg, R., 2013. *RapidMiner: Data mining use cases and business analytics applications.*, Chapman & Hall/CRC.

Keogh, E., Chakrabarti, K., Pazzani, M., et al., 2001. Dimensionality reduction for fast similarity search in large time series databases. *Knowledge and Information Systems*, 3(3), pp.263–286.

Keogh, E., Chakrabarti, K., Mehrotra, S., et al., 2001. Locally adaptive dimensionality reduction for indexing large time series databases. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*. pp. 151–162.

Keogh, E. & Kasetty, S., 2003. On the need for time series data mining benchmarks: A survey and empirical demonstration. *Data Mining and Knowledge Discovery*, 7(4), pp.349–371.

Keogh, E., Lin, J. & Fu, A., 2005. HOT SAX: Finding the most unusual time series subsequence: Algorithms and applications. In *Proceedings of the 5th IEEE International Conference on Data Mining*. pp. 226–233.

Li, G. et al., 2013. Finding time series discord based on bit representation clustering. *Knowledge-Based Systems*, 54(C), pp.243–254.

Lin, J. et al., 2003. A symbolic representation of time series, with implications for streaming algorithms. In *Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*. pp. 2–11.

Lin, J. et al., 2002. Finding motifs in time series. In *Proceedings of the 2nd Workshop on Temporal Data Mining*. pp. 53–68.

Microsoft Corporation, 2013. MS Excel. , p.URL http://office. microsoft.com/en-us/excel/. 14.

Minnen, D. et al., 2007. Discovering multivariate motifs using subsequence density estimation and greedy mixture learning. In *Proceedings of the 22nd AAAI Conference on Artificial Intelligence*. pp. 615–620.

Mueen, A. et al., 2009. Exact discovery of time series motifs. In *Proceedings of the 2009 SIAM International Conference on Data Mining*. pp. 473–484.

R Core Team, 2014. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, 0, p.{ISBN} 3-900051-07-0.

RStudio Team, 2015. RStudio: Integrated development for R.

Senin, P., 2016. Package “jmotif” - Time series analysis based on symbolic aggregate discretization.

Shieh, J. & Keogh, E., 2008. iSAX: Indexing and mining terabyte sized time series. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. pp. 623–631.

Thanh Lam, H. & Calders, T., 2010. Mining top-k frequent items in a data stream with flexible sliding windows. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. pp. 283–292.

The MathWorks Inc., 2016. MATLAB.

Vespier, U., Nijssen, S. & Knobbe, A., 2013. Mining characteristic multi-scale motifs in sensor-based time series. In *Proceedings of the 22nd ACM International Conference on Information and Knowledge Management*. pp. 2393–2398.

Yankov, D., Keogh, E. & Rebbapragada, U., 2008. Disk aware discord discovery: Finding unusual time series in terabyte sized datasets. *Knowledge and Information Systems*, 17(2), pp.241–262.

ANEXOS

Anexo 1:

Top 3 dos *motifs* identificados para todas as bases de dados, admitindo uma resolução de 4 e considerando apenas aqueles que apresentam o maior número de ocorrências (análise mensal)

Base de dados	***Rank***	**Palavra**	**Nº ocorrências**	**Distribuição por período**		
DC1	1	1,0,0,1,1,1,3,3	50	(3) Mai/14, (14) Nov/14, (3) Fev/15,	(3) Set/14, (9) Dez/14, (4) Mar/15	(4) Out/14, (10) Jan/15,
	2	1,0,0,1,2,1,3,3	35	(3) Mar/14, (3) Out/14,	(5) Ago/14, (7) Jul/15,	(3) Set/14, (14) Ago/15
	3	1,0,1,1,1,1,3,3	32	(3) Fev/14, (8) Dez/14,	(3) Out/14, (10) Jan/15,	(4) Nov/14, (4) Fev/15
DC2	1	0,0,2,3,3,3,0,0	118	(3) Abr/14, (11) Set/14, (3) Dez/14, (5) Mar/15, (4) Jun/15,	(5) Jun/14, (13) Out/14, (7) Jan/15, (9) Abr/15, (3) Ago/15, (16) Out/15	(8) Jul/14, (10) Nov/14, (11) Fev/15, (6) Mai/15, (4) Set/15,
	2	2,2,1,1,1,1,1,3	93	(4) Abr/14, (7) Jul/14, (8) Mai/15,	(7) Mai/14, (15) Ago/14, (9) Jun/15, (18) Ago/15	(9) Jun/14, (8) Abr/15, (8) Jul/15,
	3	0,0,1,3,3,3,0,0	65	(2) Fev/14, (10) Mai/15, (6) Ago/15,	(7) Set/14, (9) Jun/15, (10) Set/10	(3) Out/14, (18) Jul/15,
DC3	1	1,1,2,3,2,2,1,1	40	(3) Fev/13, (2) Set/13, (3) Abr/14, (3) Nov/14,	(2) Mar/13, (3) Out/13, (3) Mai/14, (4) Dez/14, (4) Jul/15	(3) Jun/13, (4) Mar/14, (2) Jul/14, (4) Mai/15,
	2	1,1,1,3,3,2,1,1	29	(3) Nov/13, (4) Mar/14, (3) Dez/14,	(3) Jan/14, (2) Jul/14, (2) Fev/15, (3) Jul/15	(4) Fev/14, (3) Nov/14, (2) Jun/15,

	3	1,1,1,3,2,3,1,1	27	(2) Mar/13, (4) Jan/14, (5) Out/14,	(2) Mai/13, (3) Abr/14, (4) Nov/14,	(2) Dez/13, (3) Set/14, (2) Abr/15
DC4	1	1,3,2,1,1,0,3,2	11	(2) Nov/13,	(3) Jan/14, (3) Mar/14	(3) Fev/14,
	2	1,0,0,2,2,3,3,1	8	(3) Mai/15,	(5) Jul/15	
	3	2,0,2,2,0,2,2,1	8	(3) Dez/14,	(3) Jan/15,	(2) Out/15
DC5	1	1,1,1,2,2,2,1,1	18	(3) Jan/14, (4) Dez/14,	(4) Fev/14, (2) Jan/15,	(2) Out/14, (3) Ago/15
	2	1,1,1,2,2,2,2,1	11	(3) Abr/14,	(2) Out/14, (3) Abr/15	(3) Mar/15,
	3	1,1,1,3,2,2,1,1	7	(5) Nov/13,	(2) Jul/15	

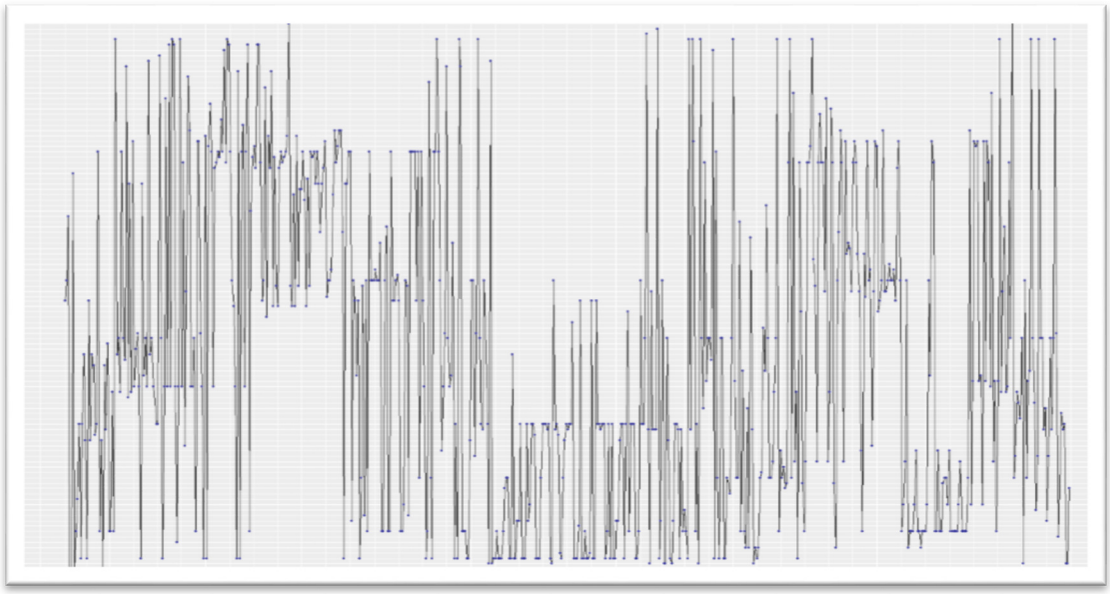
Anexo 2:

Top 3 dos *motifs* identificados para todas as bases de dados, admitindo uma resolução de 4 e considerando apenas aqueles que apresentam o maior número de ocorrências (análise por dias da semana)

Base de dados	***Rank***	**Palavra**	**Nº ocorrências**	**Distribuição por período**		
DC1	1	1,0,0,1,1,1,3,3	61	(8) Segunda, (11) Quinta,	(10) Terça, (16) Sexta,	(7) Quarta, (9) Sábado
	2	1,0,0,1,2,1,3,3	55	(9) Segunda, (8) Quinta,	(9) Terça, (8) Sexta, (8) Domingo	(5) Quarta, (8) Sábado,
	3	1,0,1,1,1,1,3,3	35	(12) Segunda, (3) Quinta,	(10) Terça, (5) Sexta	(5) Quarta,
DC2	1	0,0,2,3,3,3,0,0	122	(22) Segunda, (21) Quinta,	(28) Terça, (26) Sexta	(25) Quarta,
	2	2,2,1,1,1,1,1,3	100	(5) Segunda, (5) Quinta,	(5) Terça, (8) Sexta, (40) Domingo	(3) Quarta, (34) Sábado,
	3	0,0,1,3,3,3,0,0	75	(18) Segunda, (19) Quinta,	(18) Terça, (7) Sexta	(13) Quarta,
DC3	1	1,1,2,3,2,2,1,1	57	(5) Segunda, (17) Quinta,	(11) Terça, (12) Sexta	(12) Quarta,
	2	1,1,1,3,3,2,1,1	52	(9) Segunda, (5) Quinta,	(12) Terça, (14) Sexta	(12) Quarta,
	3	1,1,1,3,2,3,1,1	47	(12) Segunda, (7) Quinta,	(10) Terça, (8) Sexta	(10) Quarta,
DC4	1	1,3,2,1,1,0,3,2	7	(3) Segunda,	(4) Quarta	
	2	1,0,0,2,2,3,3,1	5	(2) Sexta,	(3) Sábado	
	3	2,0,2,2,0,2,2,1	3	(3) Domingo		
DC5	1	1,1,1,2,2,2,1,1	30	(3) Segunda, (4) Quinta,	(11) Terça, (6) Sexta	(6) Quarta,
	2	1,1,1,2,2,2,2,1	13	(4) Segunda,	(5) Quarta,	(4) Sexta
	3	1,1,1,2,2,2,2,2	12	(4) Terça,	(5) Quinta,	(3) Sexta

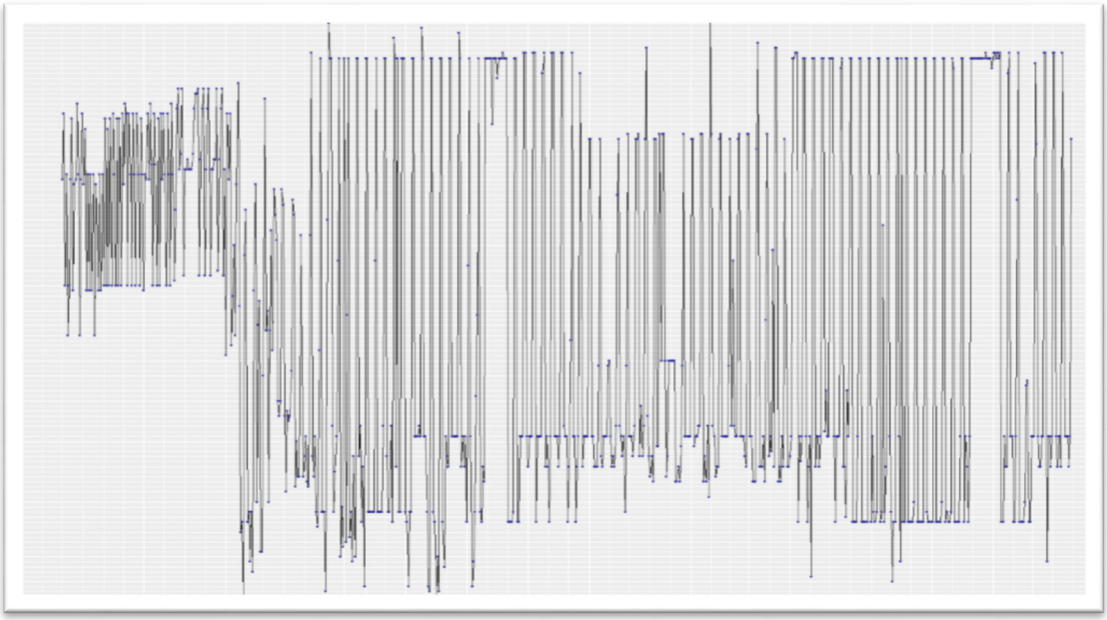
Anexo 3:

Representação gráfica das palavras geradas durante a análise de *motifs*, para todos os dias da base de dados da instalação 1, estando representados no eixo horizontal os dias da base de dados e no eixo vertical as respetivas palavras



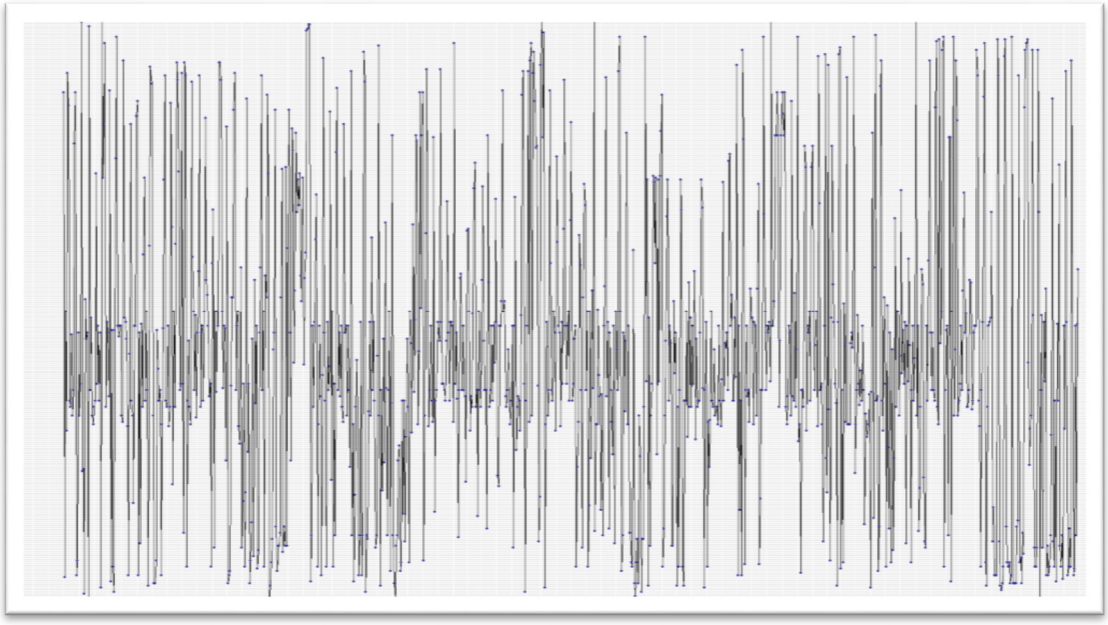
Anexo 4:

Representação gráfica das palavras geradas durante a análise de *motifs*, para todos os dias da base de dados da instalação 2, estando representados no eixo horizontal os dias da base de dados e no eixo vertical as respetivas palavras



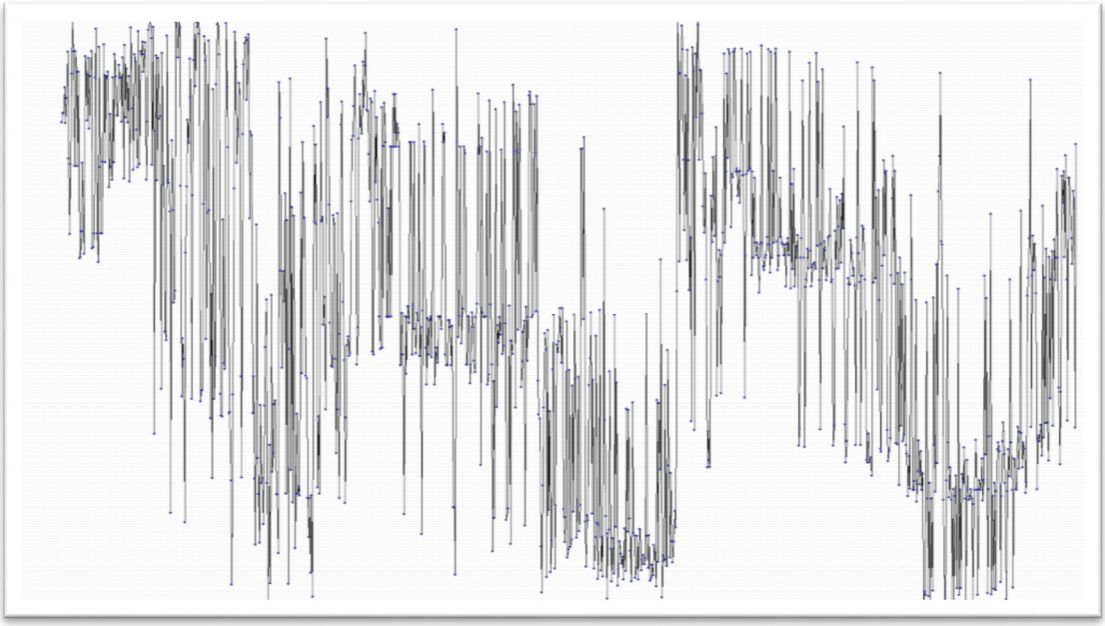
Anexo 5:

Representação gráfica das palavras geradas durante a análise de *motifs*, para todos os dias da base de dados da instalação 3, estando representados no eixo horizontal os dias da base de dados e no eixo vertical as respetivas palavras



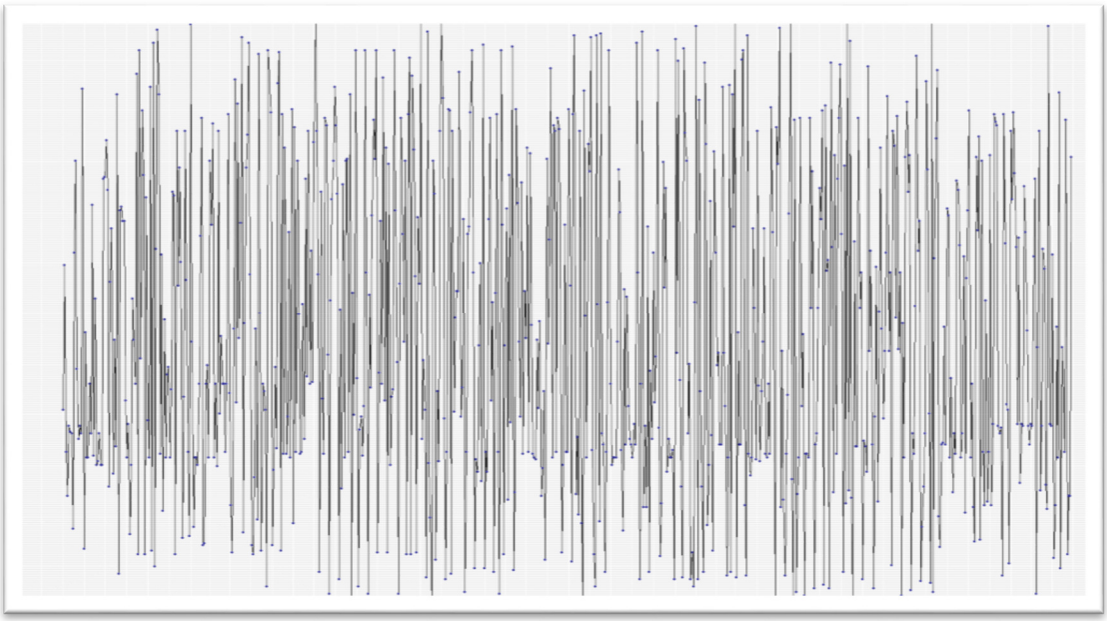
Anexo 6:

Representação gráfica das palavras geradas durante a análise de *motifs*, para todos os dias da base de dados da instalação 4, estando representados no eixo horizontal os dias da base de dados e no eixo vertical as respetivas palavras



Anexo 7:

Representação gráfica das palavras geradas durante a análise de *motifs*, para todos os dias da base de dados da instalação 5, estando representados no eixo horizontal os dias da base de dados e no eixo vertical as respetivas palavras



Anexo 8:

Top 10 dos *discords* identificados para todas as bases de dados considerando apenas aqueles que apresentam a maior distância face aos restantes (análise mensal)

Base de dados	*Rank*	Dia	Dia da semana	Distância	Observações
DC1	1	12-02-2015	Quinta	409,32872	
	2	25-09-2015	Sexta	283,01723	Faltam 3 obs.: 15:15 – 15:45
	3	01-03-2015	Domingo	185,02162	
	4	01-01-2015	Quinta	123,85475	Feriado: Ano Novo
	5	06-04-2014	Domingo	109,16959	
	6	18-01-2015	Domingo	109,02293	
	7	14-12-2014	Domingo	106,44247	
	8	24-12-2014	Quarta	106,44247	Véspera de feriado: Natal
	9	09-02-2014	Domingo	105,13800	
	10	08-02-2015	Domingo	99,37303	
DC2	1	26-02-2014	Quarta	97,70875	
	2	25-09-2015	Sexta	92,79009	
	3	21-02-2014	Sexta	90,88454	
	4	17-04-2014	Quinta	88,61151	Quinta-feira Santa (Páscoa)
	5	04-07-2014	Sexta	87,64131	
	6	17-02-2015	Terça	83,18053	Feriado: Terça-feira Carnaval
	7	16-05-2014	Sexta	77,31106	
	8	09-07-2014	Quarta	76,52451	
	9	11-07-2014	Sexta	76,52451	
	10	25-03-2014	Terça	74,66592	
DC3	1	04-02-2013	Segunda	81,57205	
	2	22-04-2015	Quarta	78,39005	
	3	16-05-2013	Quinta	76,31514	
	4	20-01-2015	Terça	75,96052	
	5	02-08-2013	Sexta	75,17313	
	6	15-02-2013	Sexta	73,73602	
	7	05-04-2013	Sexta	73,57989	
	8	12-08-2015	Quarta	73,52551	
	9	10-02-2014	Segunda	73,49150	
	10	25-02-2014	Terça	73,49150	

DC4	1	17-09-2014	Quarta	246,87900	Falta 1 obs.: 15:00
	2	26-06-2013	Quarta	237,61050	Faltam 31 obs.: 09:45 – 17:15
	3	28-05-2015	Quinta	182,71290	
	4	05-02-2014	Quarta	174,28710	
	5	27-09-2013	Sexta	173,21660	
	6	30-07-2013	Terça	171,96510	
	7	13-01-2015	Terça	171,70910	Faltam 3 obs.: 11:30 – 12:00
	8	09-06-2014	Segunda	169,22470	Véspera de feriado: 10/Junho
	9	22-06-2015	Segunda	168,28550	
	10	09-06-2015	Terça	167,33500	
DC5	1	15-11-2013	Sexta	138,43410	Faltam 5 obs.: 16:00 – 17:00
	2	06-01-2014	Segunda	121,30130	
	3	15-06-2014	Domingo	120,96690	
	4	16-11-2013	Sábado	120,61510	
	5	01-04-2014	Terça	119,46970	
	6	11-12-2013	Quarta	111,40920	
	7	30-12-2013	Segunda	105,57940	
	8	25-07-2014	Sexta	102,77650	
	9	17-03-2014	Segunda	95,47775	
	10	09-07-2014	Quarta	95,39392	

Anexo 9:

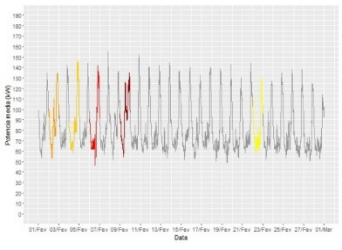
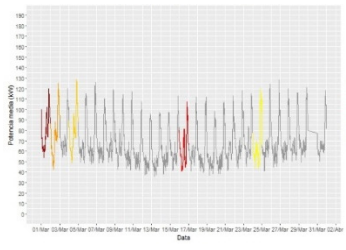
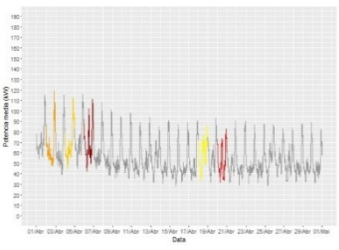
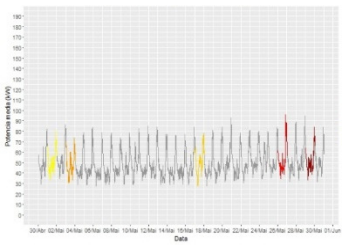
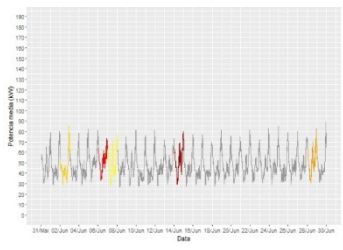
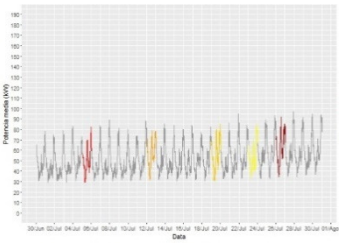
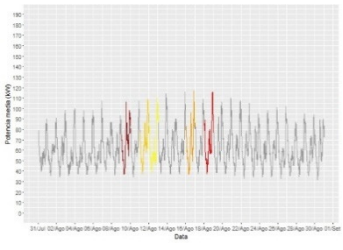
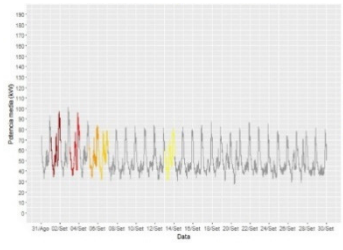
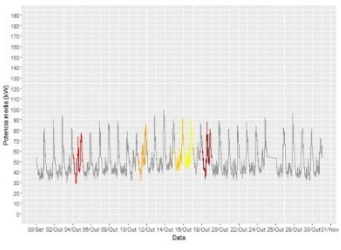
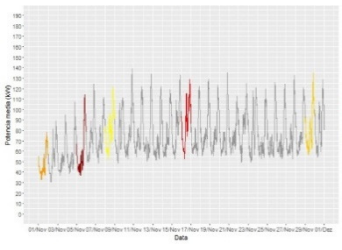
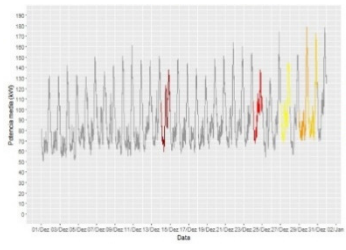
Discords identificados por dias da semana, para todas as bases de dados, considerando apenas aqueles que apresentam a maior distância face aos restantes dias da semana (análise semanal)

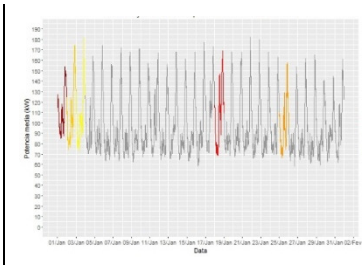
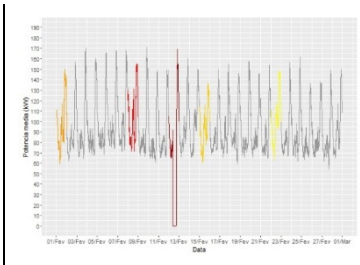
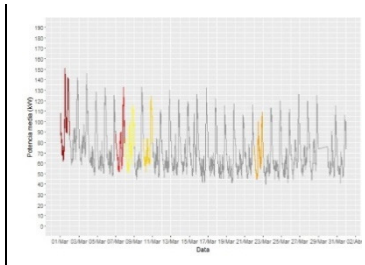
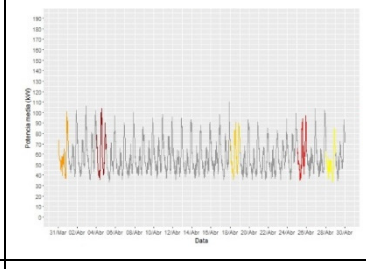
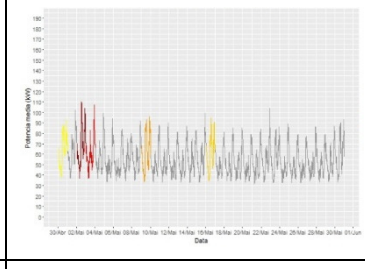
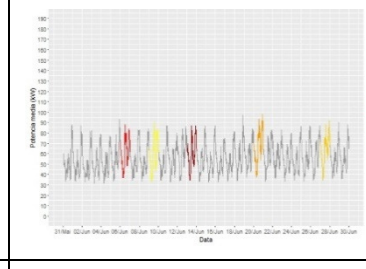
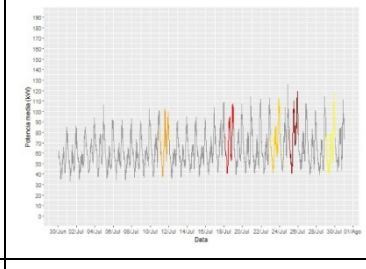
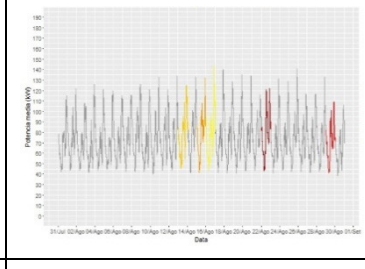
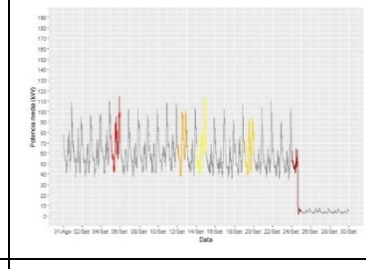
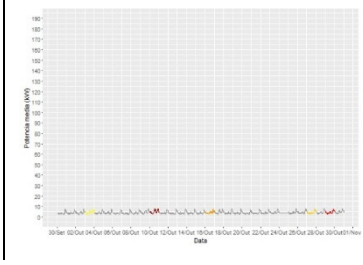
Base de dados	Dia da semana	Dia	Distância	Observações
	Segunda	17-08-2015	101,44949	
	Terça	17-02-2015	101,22253	Feriado: Terça-feira Carnaval
	Quarta	24-12-2014	167,83921	Véspera de Natal
DC1	Quinta	12-02-2015	364,49966	
	Sexta	25-09-2015	216,99712	Faltam 3 obs.: 15:15 – 15:45
	Sábado	17-01-2015	122,22111	Sábado e Domingo - sequência
	Domingo	18-01-2015	132,18548	Sábado e Domingo - sequência
	Segunda	05-05-2014	83,19255	
	Terça	25-02-2014	66,96268	
	Quarta	09-07-2014	93,56281	
DC2	Quinta	17-04-2014	85,69131	Quinta-feira Santa (Páscoa)
	Sexta	11-07-2014	83,06022	
	Sábado	31-01-2015	51,54610	
	Domingo	27-09-2015	14,59452	
	Segunda	18-02-2013	74,58552	
	Terça	19-02-2013	62,37788	
	Quarta	12-08-2015	77,67883	
DC3	Quinta	16-05-2013	71,65891	
	Sexta	08-08-2014	70,79548	
	Sábado	06-04-2013	29,34280	
	Domingo	08-12-2013	9,43398	Feriado: Dia da Imaculada Conceição
	Segunda	08-07-2013	179,99440	
	Terça	30-06-2015	182,15100	
	Quarta	17-09-2014	187,54530	
DC4	Quinta	23-05-2013	174,45630	
	Sexta	19-06-2015	176,31220	
	Sábado	03-08-2013	169,12720	
	Domingo	11-08-2013	167,27820	

	Segunda	06-01-2014	149,04362	Domingo e Segunda - sequência
	Terça	01-04-2014	97,84682	
	Quarta	25-12-2013	87,52143	Feriado: Natal
DC5	Quinta	23-01-2014	88,97191	
	Sexta	15-11-2013	118,81919	
	Sábado	30-11-2013	88,53813	
	Domingo	05-01-2014	106,59737	Domingo e Segunda - sequência

Anexo 10:

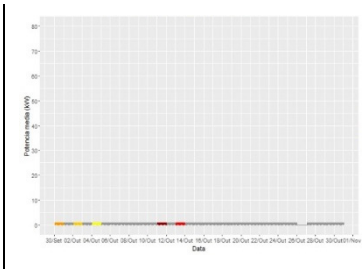
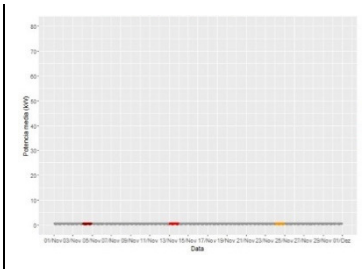
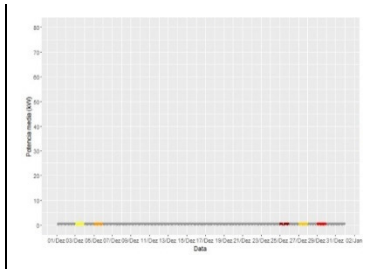
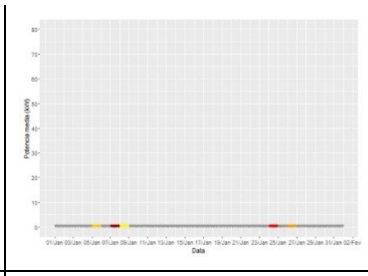
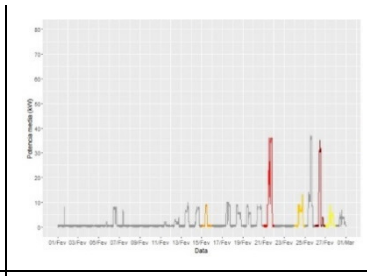
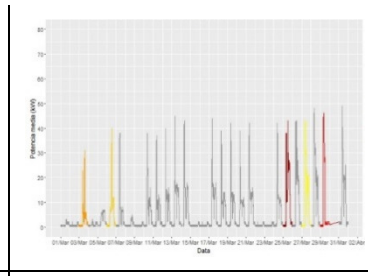
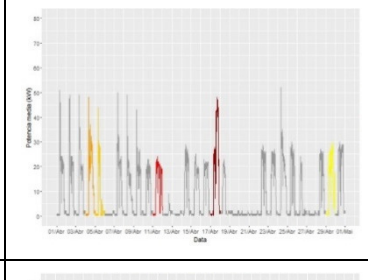
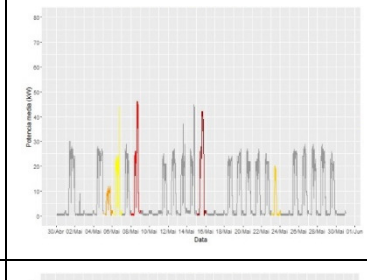
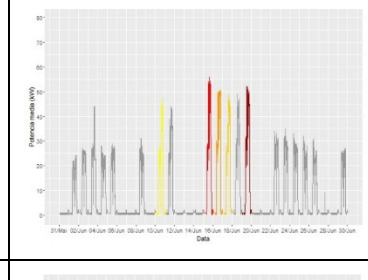
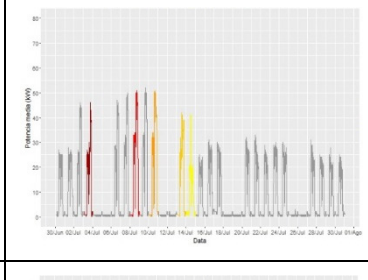
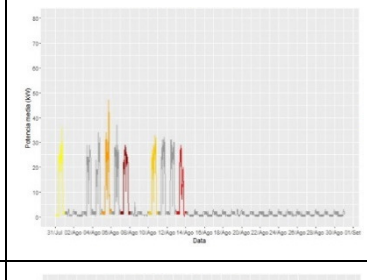
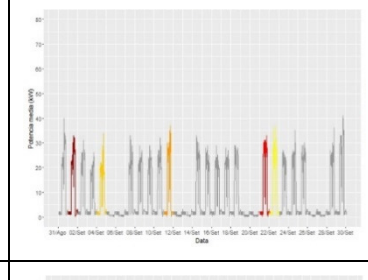
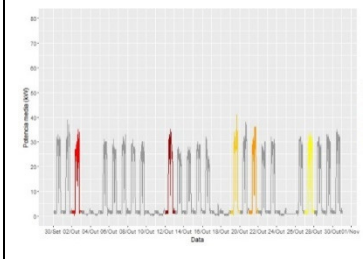
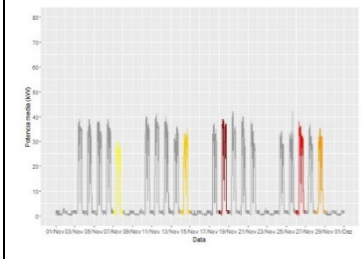
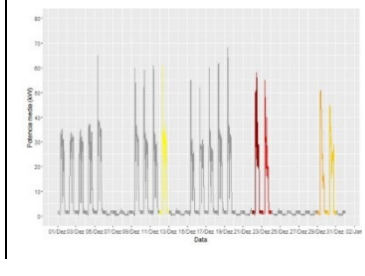
Representação gráfica dos diagramas de carga mensais da instalação 1 com evidência dos cinco *discords* identificados através de uma hierarquia de cores

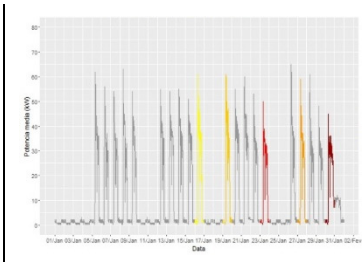
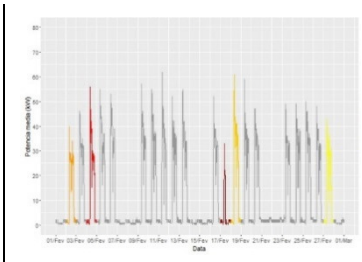
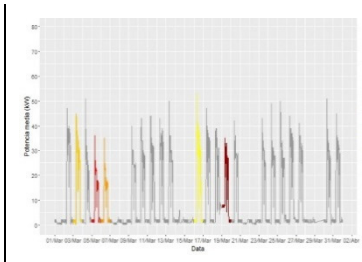
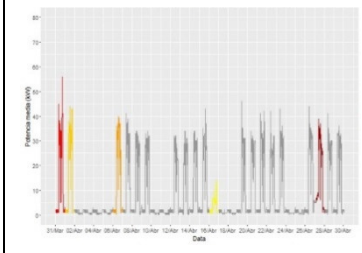
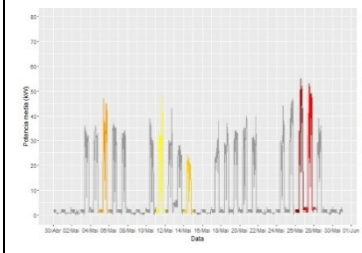
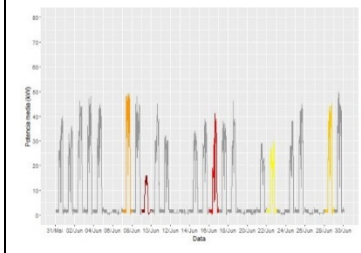
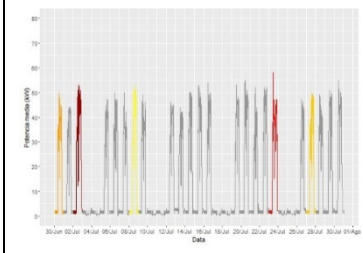
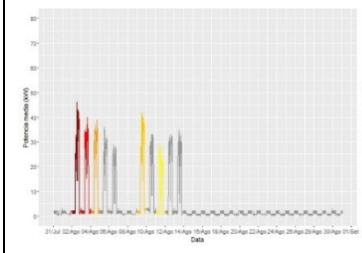
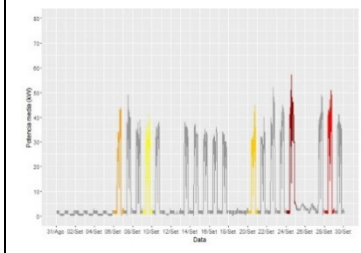
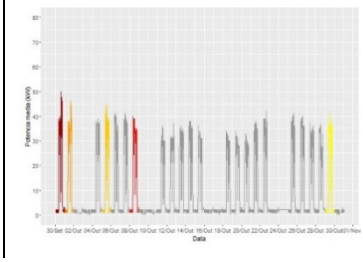
2014	1º Trimestre			
	2º Trimestre			
	3º Trimestre			
	4º Trimestre			

2015				
	1º Trimestre			
	2º Trimestre			
	3º Trimestre			
	4º Trimestre			

Anexo 11:

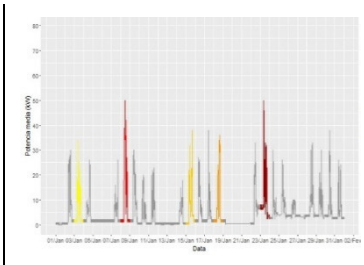
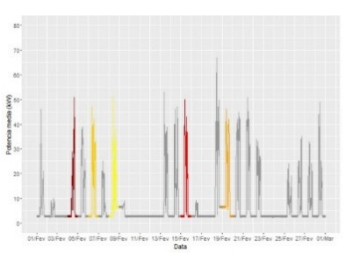
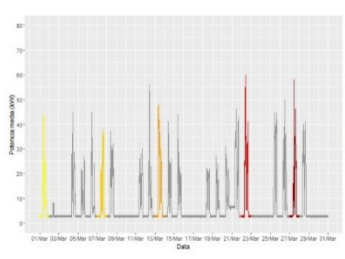
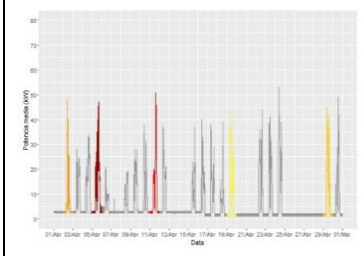
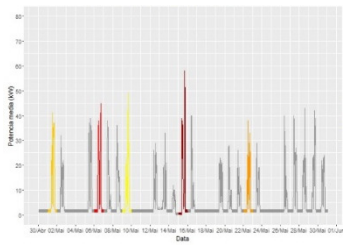
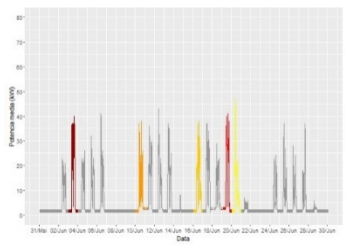
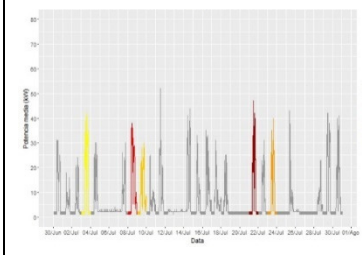
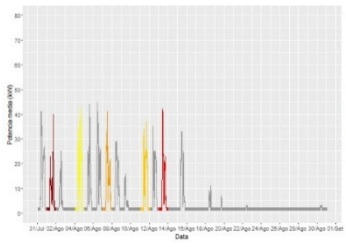
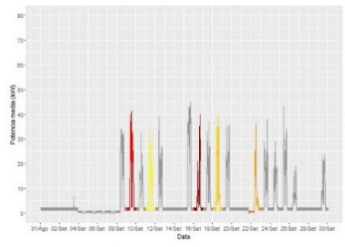
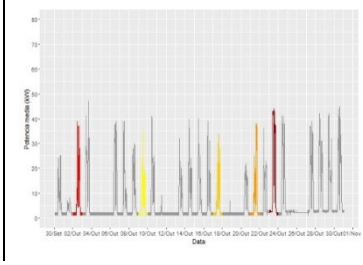
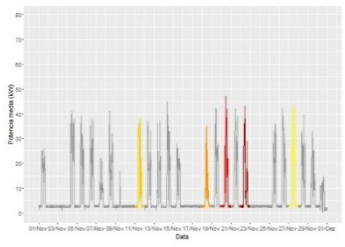
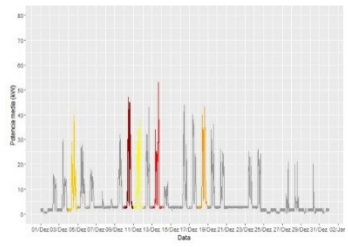
Representação gráfica dos diagramas de carga mensais da instalação 2 com evidência dos cinco *discords* identificados através de uma hierarquia de cores

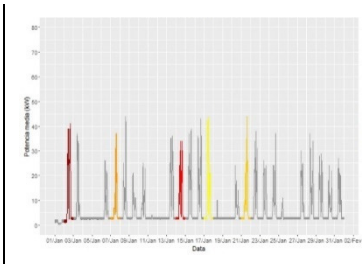
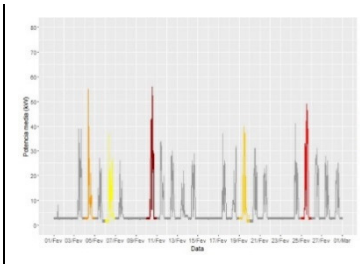
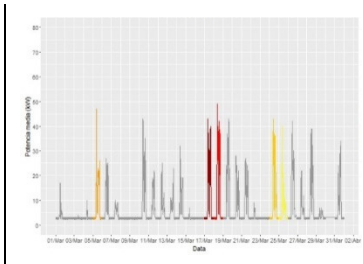
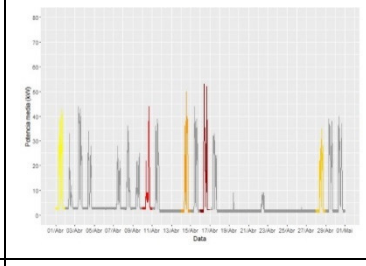
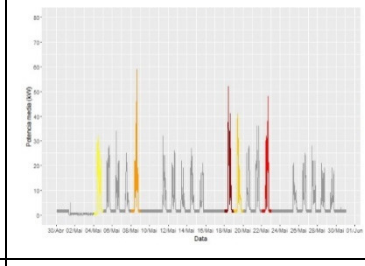
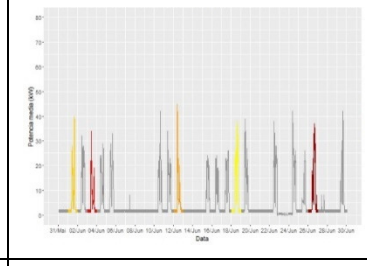
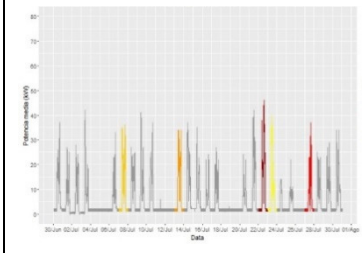
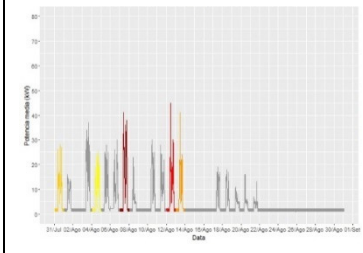
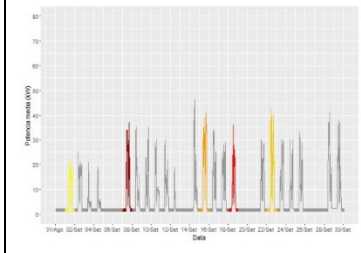
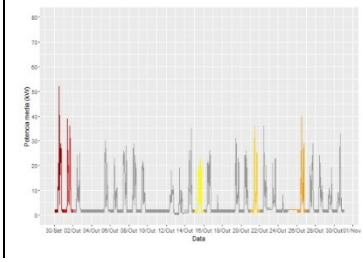
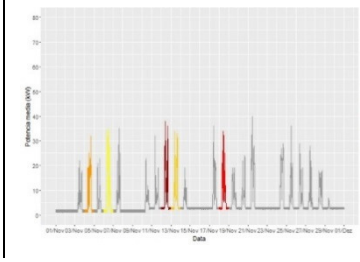
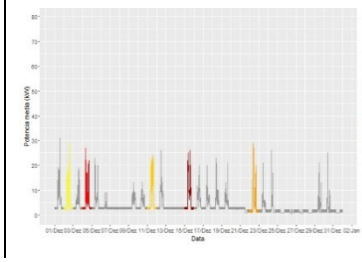
2013	4º Trimestre			
2014	1º Trimestre			
	2º Trimestre			
	3º Trimestre			
	4º Trimestre			

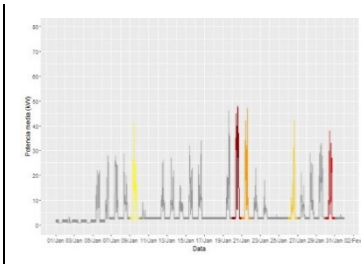
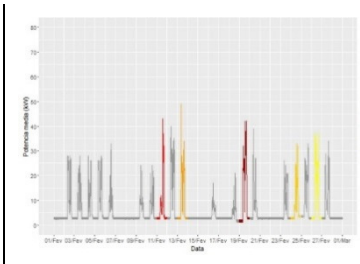
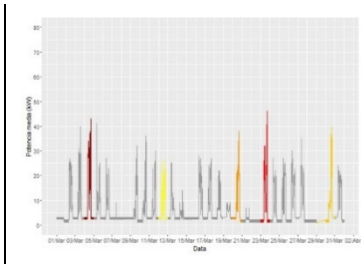
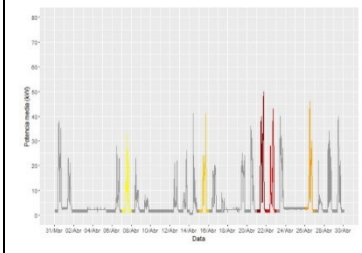
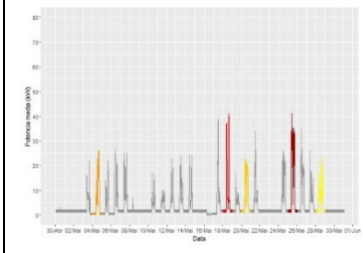
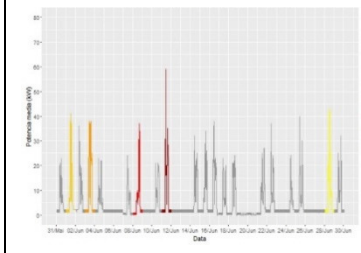
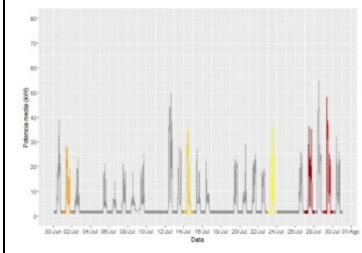
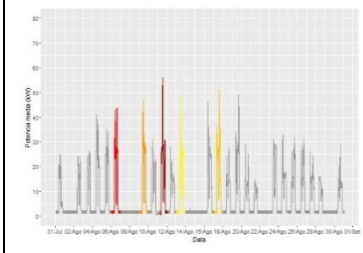
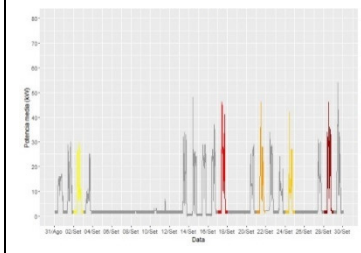
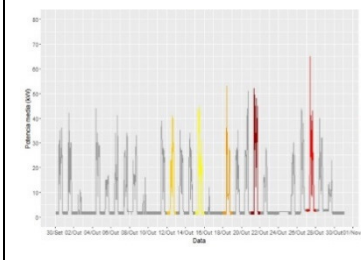
2015				
	1º Trimestre			
	2º Trimestre			
	3º Trimestre			
	4º Trimestre			

Anexo 12:

Representação gráfica dos diagramas de carga mensais da instalação 3 com evidência dos cinco *discords* identificados através de uma hierarquia de cores

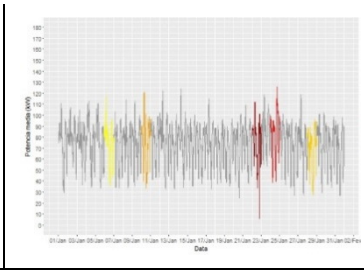
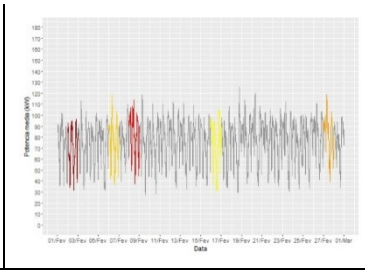
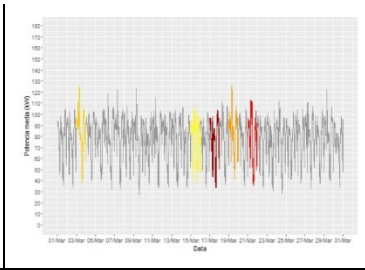
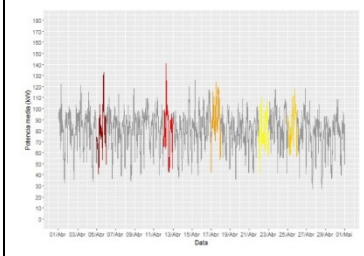
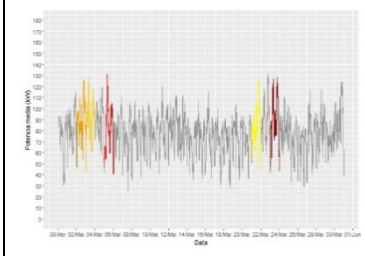
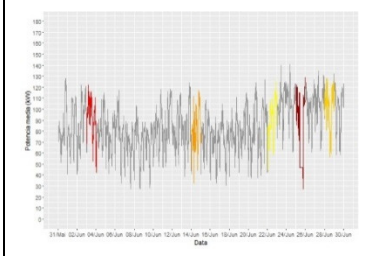
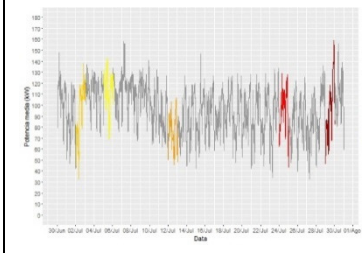
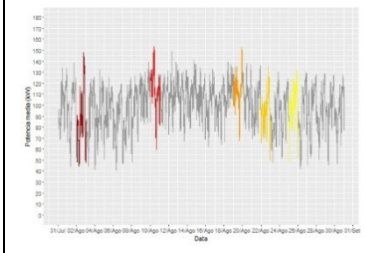
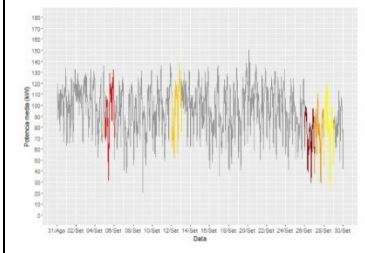
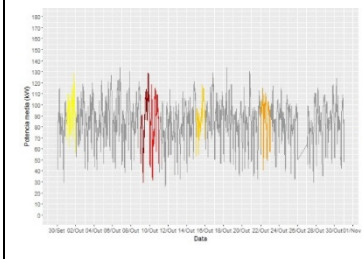
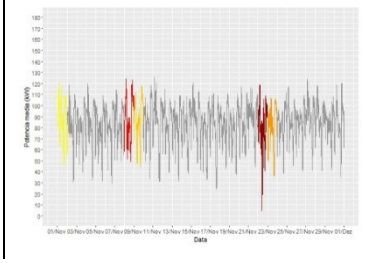
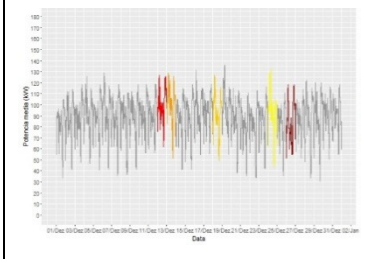
2013	1º Trimestre			
	2º Trimestre			
	3º Trimestre			
	4º Trimestre			

2014				
	1º Trimestre			
	2º Trimestre			
	3º Trimestre			
	4º Trimestre			

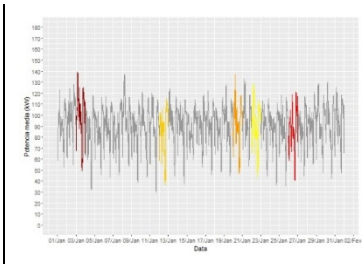
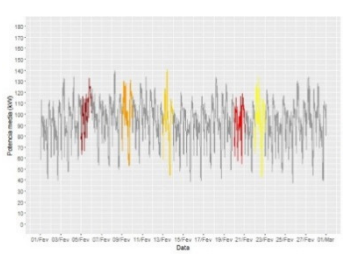
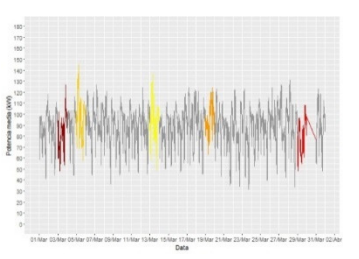
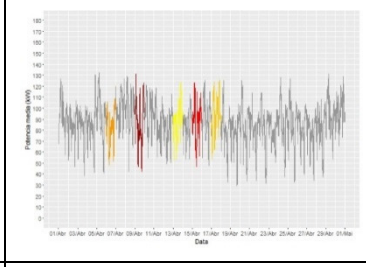
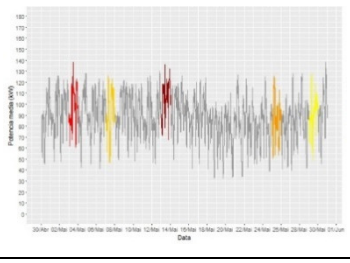
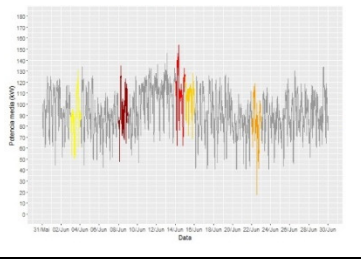
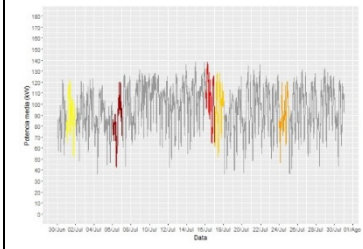
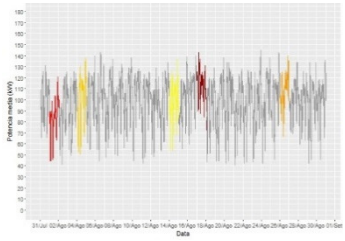
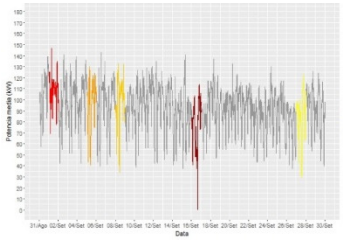
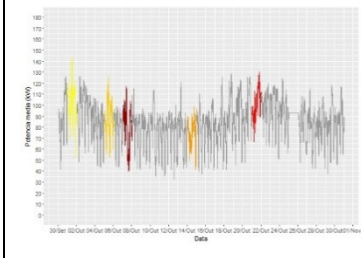
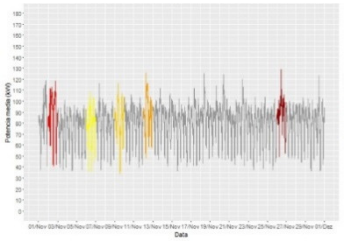
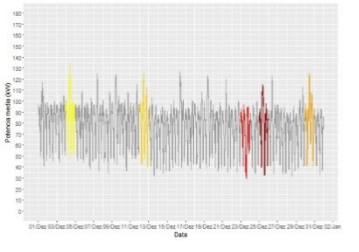
2015	1º Trimestre			
	2º Trimestre			
	3º Trimestre			
	4º Trimestre			

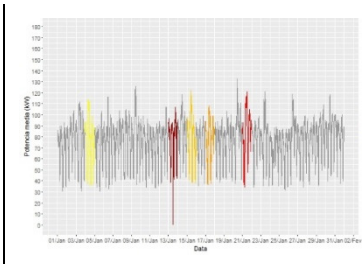
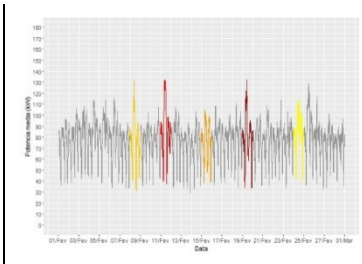
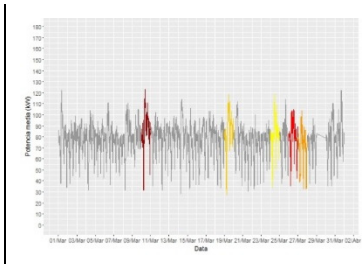
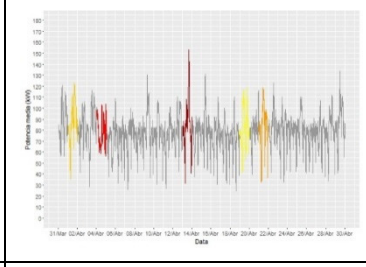
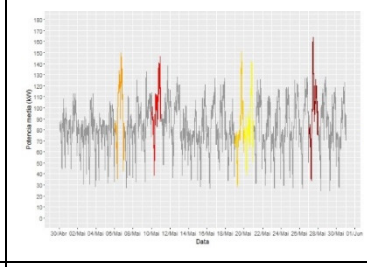
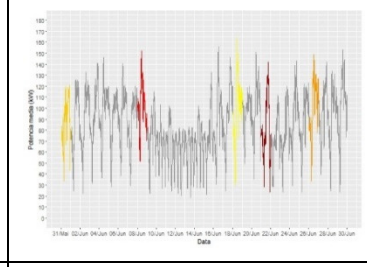
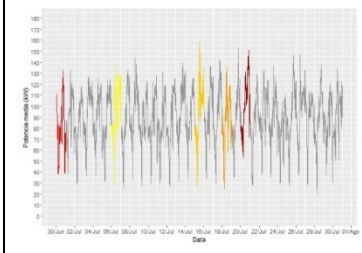
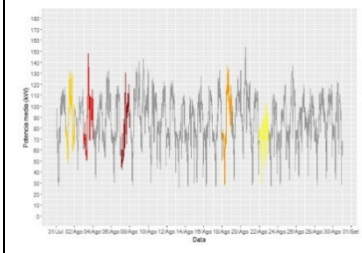
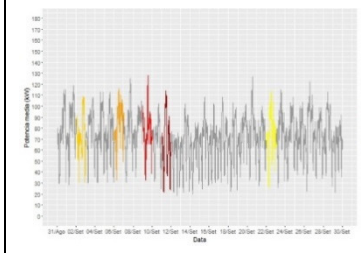
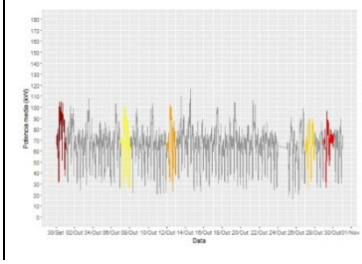
Anexo 13:

Representação gráfica dos diagramas de carga mensais da instalação 4 com evidência dos cinco *discords* identificados através de uma hierarquia de cores

2013	**1º Trimestre**			
	2º Trimestre			
	3º Trimestre			
	4º Trimestre			

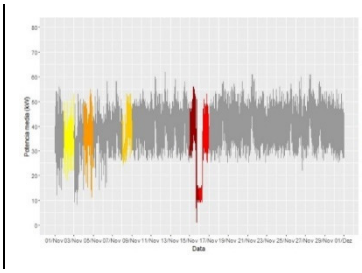
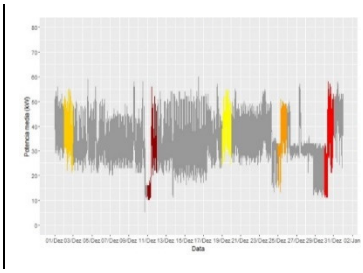
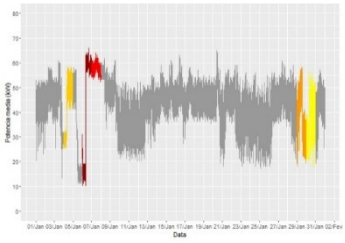
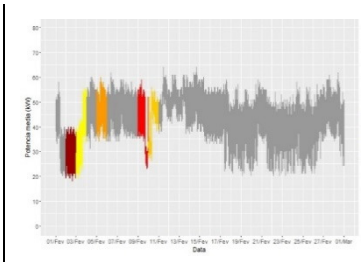
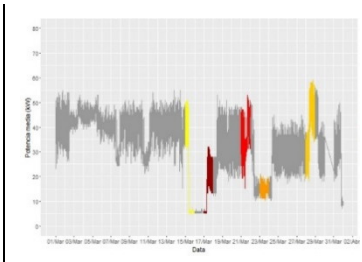
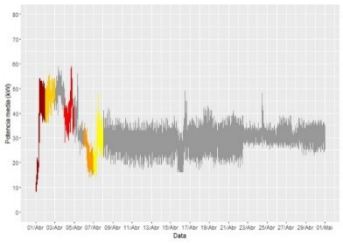
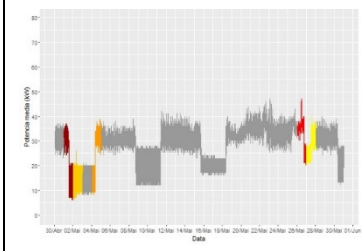
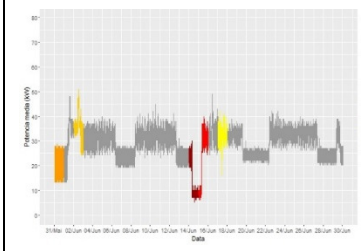
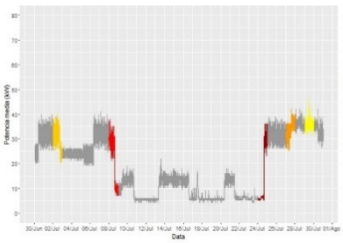
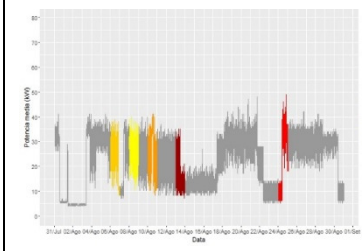
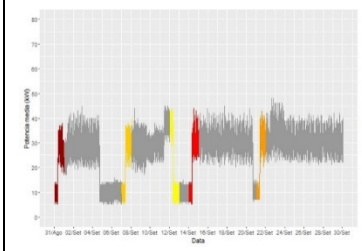
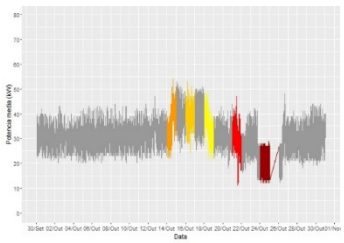
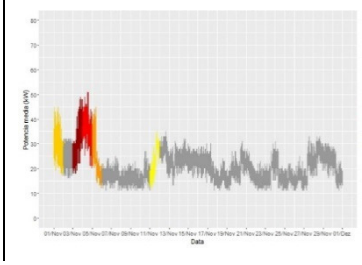
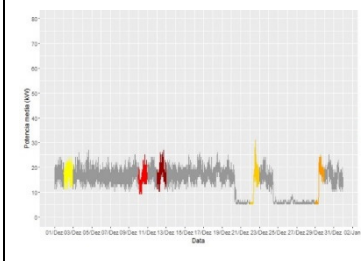
2014

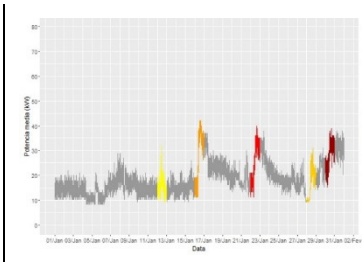
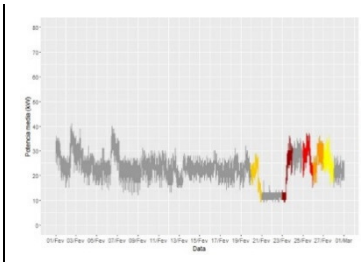
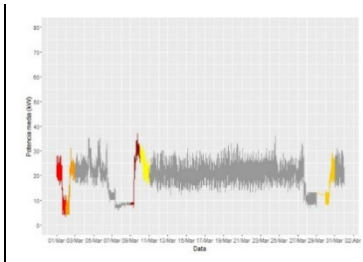
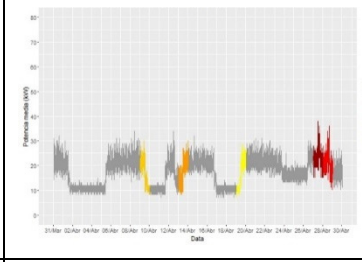
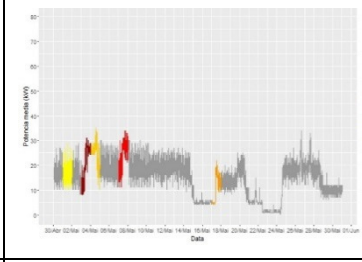
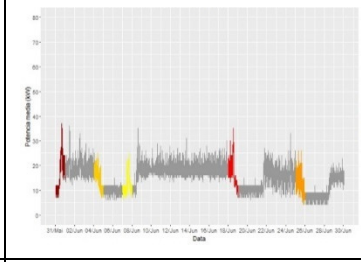
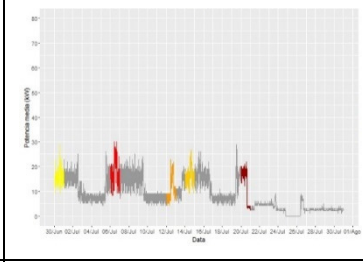
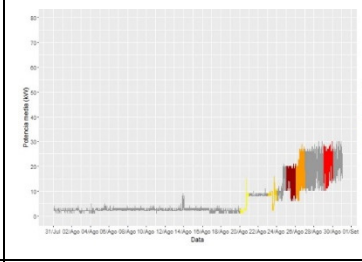
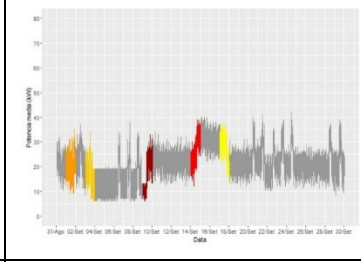
1º Trimestre			
2º Trimestre			
3º Trimestre			
4º Trimestre			

2015				
	1º Trimestre			
	2º Trimestre			
	3º Trimestre			
	4º Trimestre			

Anexo 14:

Representação gráfica dos diagramas de carga mensais da instalação 5 com evidência dos cinco *discords* identificados através de uma hierarquia de cores

2013	4º Trimestre			
2014	1º Trimestre			
	2º Trimestre			
	3º Trimestre			
	4º Trimestre			

2015				
	1º Trimestre			
	2º Trimestre			
	3º Trimestre			
	4º Trimestre