



Projecto

Mestrado em Engenharia Electrotécnica

***Diagnóstico de Consumos Anómalos de Energia:  
Abordagem por Classificação***

**Fábio Henrique Manso Dias**

Leiria, Setembro de 2014



Projecto

Mestrado em Engenharia Electrotécnica

***Diagnóstico de Consumos Anómalos de Energia:  
Abordagem por Classificação***

**Fábio Henrique Manso Dias**

Projecto de Mestrado realizada sob a orientação do Doutor Luís Neves, Professor da Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Leiria e co-orientação do Doutor João Sousa, Professor da Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Leiria.

Leiria, *Setembro* de 2014

## **AGRADECIMENTOS**

Uma vida académica cheia de projectos enriquecedores chega ao fim, sendo este o culminar de um longo e demorado processo de aprendizagem. Foram muitas as pessoas que foram fazendo parte da minha vida nos últimos anos a nível académico, profissional e pessoal. Não será possível agradecer a todas elas, contudo pretendo salientar as pessoas que me são mais próximas e que me foram acompanhando em todo este percurso.

Tenho que agradecer a minha namorada e aos meus pais que foram as grandes forças impulsionadoras e inspiradoras para o meu sucesso, que nunca me deixaram desistir mesmo nas horas em que o cansaço era superior à vontade. Aos meus companheiros de estudos e amigos Hugo Ferreira, Filipe Cordeiro, Hugo Moita, Nelson Seabra, João Fazendeiro, Marco Antunes, Inês Santos e Renato Gaspar. Um grande agradecimento aos meus professores que me ajudaram em todo este processo final de curso e que finalmente terão um pouco de descanso das dores de cabeça que lhes fui provocando principalmente os meus orientadores do projecto de mestrado Luís Neves, João Sousa e ao professor Pedro Marques a quem tenho que agradecer por ter cedido os dados que me permitiram o uso da aplicação em tempo real.



## RESUMO

Durante o período de funcionamento de uma instalação eléctrica podem ocorrer várias anomalias. Enquanto muitas delas apenas são identificadas tardiamente, outras acabam por nunca serem identificadas como um potencial problema. A identificação atempada dessas anomalias permite a realização de um diagnóstico que leve à correcção das suas causas evitando assim os desperdícios e prejuízos inerentes.

A identificação de um consumo anómalo pode ser realizada, de forma automática ou semi automática através de sistemas de apoio que permitam sinalizar falhas ou comportamentos anormais. O trabalho apresentado nesta dissertação pretende possibilitar esta sinalização apenas através da análise dos dados de consumo medidos em tempo real e comparados com dados históricos através de uma abordagem baseada em classificação, recorrendo a métodos de *clustering*.

Foram testadas diferentes abordagens em três casos distintos, dois relativos a consumidores residenciais para os quais existiam registos de consumo durante um período alargado, e um relativo a uma instalação desportiva, para a qual é possível aceder em tempo real ao sistema de gestão de consumos via web.

O sistema implementado proporciona vários tipos de informação ao utilizador, permitindo visualizar graficamente a existência de uma potencial anomalia quando a disparidade entre a classificação do consumo no instante e a classe do consumo de referência for significativa.

**Palavras-chave** – Análise de consumos energéticos, Detecção de anomalias, *Clustering*.



## **ABSTRACT**

There are several anomalies that can occur during the operation of an electrical installation. While many of these are lately identified, others may never be identified as a potential problem. The early identification of these anomalies allows the execution of a diagnosis that leads to the correction of the causes, avoiding waste and the inherent losses.

The identification of anomalous consumption can be taken, automatically or semi-automatically, through support systems that report failures or anomalous behaviors. The study presented in this dissertation is intended to allow such signaling only through the analysis of consumption data measured in real time and compared with historical data through a classification-based approach, using clustering.

Three separate cases were tested through different approaches, two cases related to residential consumers for which there were records of consumption over an extended period, and another case referent to a sports facility for which the consumption management system can be accessed in real time web.

The implemented system provides various types of information to the user, allowing to graphically see the presence of a potential anomaly when the disparity between the classification of the consumer at the moment and the class of the reference consumption is significant.

**Keywords** – Analysis of energy consumption, Anomaly detection, Clustering.





# ÍNDICE

1.	INTRODUÇÃO .....	1
2.	ESTADO DA ARTE .....	3
2.1.	MÉTODO DE <i>CLUSTERING FUZZY C-MEANS</i> .....	5
2.2.	MÉTODO DE <i>KOHONEN</i> .....	7
2.3.	MÉTODO DE <i>PERCENTIS</i> .....	9
3.	PROPOSTA DE IMPLEMENTAÇÃO.....	11
3.1.	DESCRIÇÃO DAS PROPOSTAS DE IMPLEMENTAÇÃO .....	11
3.1.1.	MÉTODO <i>FUZZY C-MEANS</i> .....	11
3.1.2.	MÉTODO <i>PERCENTIS</i> .....	13
3.1.3.	MÉTODO <i>FUZZY C-MEANS</i> COM REFERÊNCIA ESCOLHIDA PELO UTILIZADOR .....	14
3.1.4.	MÉTODO <i>FUZZY C-MEANS</i> EM TEMPO REAL.....	14
3.2.	DESCRIÇÃO RESUMIDA DO FUNCIONAMENTO DAS APLICAÇÕES .....	15
3.3.	LEITURA DE DADOS PELAS APLICAÇÕES DESENVOLVIDAS.....	15
3.4.	ESTRUTURA DA APRENDIZAGEM .....	16
3.5.	CARACTERIZAÇÃO DE UM OUTLIER NAS APLICAÇÕES.....	17
3.6.	FUNCIONAMENTO DO INTERFACE GRÁFICO .....	17
4.	ESTUDO DE CASOS.....	25
4.1.	APLICAÇÃO DO <i>FUZZY C-MEANS</i> .....	25
4.2.	APLICAÇÃO DOS <i>PERCENTIS</i> .....	33
4.3.	APLICAÇÃO DO <i>FUZZY C-MEANS</i> COM REFERÊNCIA ESCOLHIDA PELO UTILIZADOR .....	35
4.3.1.	CASO DB1 .....	35
4.3.2.	CASO DB2 .....	38
4.4.	APLICAÇÃO DO <i>FUZZY C-MEANS</i> EM TEMPO REAL.....	40
5.	CONCLUSÕES .....	47
6.	BIBLIOGRAFIA.....	49
7.	ANEXOS .....	51



## LISTA DE FIGURAS

Figura 1 – Exemplo gráfico de iterações.....	6
Figura 2 – Mapa de ligações do método <i>Kohonen</i> . ....	8
Figura 3 – Exemplo de um gráfico em <i>percentis</i> [8]. ....	10
Figura 4 – Painel de comando - GUI. ....	18
Figura 5 – Parâmetros iniciais para arranque do programa - GUI.....	18
Figura 6 – Exemplo de tabela de comparação de potência – GUI.....	19
Figura 7 – Exemplo de tabelas de alarmes – GUI. ....	19
Figura 8 – Exemplo de dados de consumo – GUI.....	21
Figura 9 – Exemplo de gráfico diferencial. ....	21
Figura 10 – Exemplo de gráfico de classe actual. ....	22
Figura 11 – Exemplo de gráfico de classe anterior.....	22
Figura 12 – Exemplo de gráfico de potências activas.....	22
Figura 13 – Exemplo de gráfico de potência activa do dia actual. ....	23
Figura 14 – Exemplo de gráfico de potência activa esperada. ....	23
Figura 15 – Dados com apresentação sob a forma de pontos – relativo à base de dados DB1. ....	25
Figura 16 – Dados com apresentação sob a forma de pontos – relativo à base de dados DB2. ....	26
Figura 17 – Dados com apresentação sob a forma de área – relativo à base de dados DB1. ....	26
Figura 18 – Dados com apresentação sob a forma de área – relativo à base de dados DB2. ....	27
Figura 19 – Dados com apresentação sob a forma de pontos separados por <i>cluster</i> segundo o método <i>Fuzzy C-Means</i> – DB1. ....	27
Figura 20 – Dados com apresentação sob a forma de pontos separados por <i>cluster</i> segundo o método <i>Fuzzy C-Means</i> – DB2. ....	28
Figura 21 – Gráficos de potência activa e da separação por classes com base nos valores actuais e estipulados – DB1. ....	28
Figura 22 – Gráficos de potência activa e da separação por classes com base nos valores actuais e estipulados – DB2. ....	29
Figura 23 – Semáforo de alertas – DB1. ....	30
Figura 24 – Semáforo de alertas – DB2. ....	30
Figura 25 – Diferencial de potência activa – DB1. ....	31
Figura 26 – Diferencial de potência activa – DB2. ....	31
Figura 27 – Tabela de classificações (Painel) – DB1. ....	32
Figura 28 – Tabela de classificações (Painel) – DB2. ....	32
Figura 29 – Dados com apresentação sob a forma de pontos separados para o primeiro e segundo método – DB1. ....	33
Figura 30 – Dados com apresentação sob a forma de pontos separados para o primeiro e segundo método – DB2. ....	33

Figura 31 – Comparativo de classificação entre o primeiro e segundo método – DB1.....	34
Figura 32 – Comparativo de classificação entre o primeiro e segundo método – DB2.....	34
Figura 33 – Comparativo de alertas entre o primeiro e segundo método – DB1.....	34
Figura 34 – Comparativo de alertas entre o primeiro e segundo método – DB2.....	35
Figura 35 – Resultado da comparação com dia anterior – DB1.....	36
Figura 36 – Comparação de resultados de potência activa e classificação entre a primeira e segunda abordagem do método <i>Fuzzy C-Means</i> para um atraso de um dia – DB1.....	36
Figura 37 – Comparação de resultados do semáforo entre a primeira e segunda abordagem do método <i>Fuzzy C-Means</i> para um atraso de um dia – DB1.....	37
Figura 38 – Comparação de diferencial de potência activa entre a primeira e segunda abordagem do método <i>Fuzzy C-Means</i> para um atraso de um dia – DB1. ....	37
Figura 39 – Comparação de resultados de potência activa e classificação da segunda abordagem do método <i>Fuzzy C-Means</i> para um atraso de um e sete dias – DB1.....	37
Figura 40 – Comparação do semáforo da segunda abordagem do método <i>Fuzzy C-Means</i> para um atraso de um e sete dias – DB1.....	38
Figura 41 – Comparação do diferencial de potência activa da segunda abordagem do método <i>Fuzzy C-Means</i> para um atraso de um e sete dias – DB1. ....	38
Figura 42 – Resultado da comparação com dia anterior – DB2.....	38
Figura 43 – Comparação de resultados de potência activa e classificação entre a primeira e segunda abordagem do método <i>Fuzzy C-Means</i> para um atraso de um dia – DB2.....	39
Figura 44 – Comparação de resultados de potência activa e classificação da segunda abordagem do método <i>Fuzzy C-Means</i> para um atraso de um e sete dias – DB2.....	39
Figura 45 – Comparação de resultados de potência activa e classificação da segunda abordagem do método <i>Fuzzy C-Means</i> para um atraso de um e catorze dias – DB2.....	39
Figura 46 – Dados com apresentação sob a forma de pontos – relativo a DB3. ....	41
Figura 47 – Dados com apresentação sob a forma de pontos separados por <i>cluster</i> para a aplicação <i>Fuzzy C-Means</i> em tempo real – DB3.....	41
Figura 48 – Resultado da comparação com um dia de atraso – <i>Fuzzy C-Means</i> em tempo real .....	42
Figura 49 – Resultado da comparação de potência activa e classificação com um dia de atraso – <i>Fuzzy C-Means</i> em tempo real .....	42
Figura 50 – Diferencial de potência activa com um dia de atraso – <i>Fuzzy C-Means</i> em tempo real.....	43
Figura 51 – Semáforo de alertas com um dia de atraso – <i>Fuzzy C-Means</i> em tempo real.....	43
Figura 52 – Comparação de resultados de potência activa e classificação do método <i>Fuzzy C-Means</i> em tempo real para um atraso de um e sete dias. ....	44
Figura 53 – Comparação de resultados de potência activa e classificação do método <i>Fuzzy C-Means</i> em tempo real para um atraso de um e catorze dias.....	44
Figura 54 – Comparação do diferencial de potência activa do método <i>Fuzzy C-Means</i> em tempo real para um atraso de um e catorze dias.....	44
Figura 55 – Comparação do semáforo do método <i>Fuzzy C-Means</i> em tempo real para um atraso de um e catorze dias.....	45

## **LISTA DE SIGLAS**

BD – Base de Datos

GUI - *Graphical User Interfaces*

MAPE - *Mean Absolute Percentage Error*



## 1. INTRODUÇÃO

Nos últimos anos tem surgido um largo interesse por parte dos consumidores, na obtenção de informação mais detalhada relativamente ao valor energético consumido nas diversas instalações, pretendendo saber de que forma se pode potenciar o uso racional de energia. Os chamados *smart meters* têm sido apresentados como dotados de variadíssimas utilidades, uma das quais reside na possibilidade de alertar o utilizador para anomalias verificadas no consumo, quer sejam consumos excessivos decorrentes de equipamentos que ficaram ligados indevidamente, quer seja por inexistência de consumos devido a equipamentos que não funcionam como deveriam, eventualmente devido à ocorrência de avarias. Contudo, a concretização dessa funcionalidade é normalmente proposta por recurso à colocação de sensores diversos, que podem na maioria das vezes impor custos que não podem ser desprezados. Com a utilização de sistemas que controlem e identifiquem as diversas variações energéticas ao longo do tempo, possibilita-se o ajuste da despesa operacional associada aos equipamentos de uso final e na verificação precoce de anomalias.

O objectivo do projecto consiste em proporcionar informação ao consumidor que lhe permita aferir a normalidade do consumo em tempo real, usando apenas tratamento dos dados recolhidos pelo normal contador de energia através de uma classificação em níveis que podem ser comparados com a expectativa do consumidor. A necessidade de identificar anomalias no consumo sem implicar o uso de um número significativo de sensores, com todas as respectivas implicações inerentes à sua utilização, levou à proposta de um processo de detecção baseado na classificação do consumo actual de acordo com o histórico de consumo energético existente. Ao incorporar esta visualização permite ao utilizador detectar eventuais não conformidades de acordo com a sua própria expectativa.

Existem na literatura diversas abordagens metodológicas, no entanto nem todos os métodos proporcionam o mesmo resultado para o mesmo problema. O método apresentado recorre ao reajuste dos dados recolhidos através de uma determinada estrutura, agrupando-os por categorias que permitam a obtenção de resultados, fornecendo uma resposta segundo um determinado padrão matemático. A separação por classes é obtida recorrendo às funcionalidades dos métodos de *clustering*, que ao agrupar os dados recolhidos por classes, ou *clusters*, permite ao utilizador verificar o estado actual ou estados anteriores de consumos, facilitando a identificação de anomalias resultantes da operacionalidade dos sistemas em análise.

A estrutura do texto apresentado encontra-se dividida em três partes principais. O segundo capítulo consiste no estudo das ferramentas utilizadas normalmente na segmentação de dados, referindo igualmente quais as principais funcionalidades dos métodos apresentados. O terceiro capítulo é referente ao estudo da implementação dos *softwares* que permitiu o tratamento de dados e a obtenção de valores resultantes desta análise. E por último o quarto capítulo consistiu na análise e na discussão de resultados, discutindo as diferenças encontradas nos diversos casos de estudo e consoante as implementações dos algoritmos desenvolvidos.



## 2. ESTADO DA ARTE

O trabalho descrito neste relatório partiu de uma hipótese de definição de potenciais alertas por classificação do consumo medido numa determinada janela temporal deslizante de acordo com classes previamente definidas em função do histórico de consumo. Desta forma, um consumidor poderá reagir se confrontado com uma medição classificada como correspondendo a um nível elevado de consumo, entender que esse nível não se coaduna com a sua expectativa porque não seria suposto ter algum equipamento com peso considerável no consumo em funcionamento. O mesmo poderá acontecer em sentido inverso, ou seja, se um nível de consumo baixo for considerado não concordante com uma expectativa associada ao esperado funcionamento de equipamentos de uso final. Neste sentido, uma etapa prévia consiste na definição dos níveis ou perfis de consumo que servirão de padrão de comparação e que deverão ser determinados a partir do histórico. Embora uma possibilidade de concretização desta definição pudesse partir de um tratamento estatístico dos dados históricos, a hipótese formulada para este trabalho consistiu no uso de métodos de *clustering*, possibilitando assim uma definição automática e dinâmica desses padrões de consumo, que pode ir sendo actualizada no tempo e que se espera ter uma maior imunidade a valores extremos dos dados de partida, que podem inclusive ser causados por erros de leitura. Este capítulo incidirá assim especialmente sobre os métodos de *clustering* que serão utilizados na concepção do *software* que serve de produto final do projecto em causa.

Em diversas situações é necessário saber identificar e classificar padrões. O método de processamento de padrões poderá ser identificado como uma procura por estruturas de dados e a sua classificação definida por categorias. Numa estrutura de dados à medida que vão sendo analisados mais valores, estes vão correspondendo às diversas categorias consoante o seu grau de relevância, ou seja, cada valor terá uma relação de pertença aos vários *clusters*, sendo o valor de pertença baixo para as classes as quais o dado não se enquadra e alto para a classe à qual o valor será associado.

Na utilização de um método de *clustering*, uma das primeiras tarefas consiste na separação dos dados por grupos que contenham parâmetros semelhantes, dividindo-os por classes sendo assim mais fácil adoptar técnicas que identifiquem um padrão num conjunto global de dados. Os processos que consistem na exploração de padrões existentes em bases de dados podem ser diversos, consistindo no conceito genérico de *data mining*. Chen & Du citado em [1], dividem o processo de *data mining* nos seguintes passos:

- 1) Conhecimento e compreensão do problema em análise, tal como a identificação dos objectivos resultantes dos dados existentes;
- 2) Particionar o conjunto de dados fornecidos e seleccionar o conjunto que seja relevante aos objectivos pretendidos;
- 3) Eliminação e limpeza de dados excedentários, nomeadamente a exclusão de ruídos e a correcção de dados em falta;
- 4) Identificação de padrões no conjunto de dados e consequente visualização dos mesmos;
- 5) Aplicar os conhecimentos aprendidos ao problema em estudo;
- 6) Validação dos *clusters* e interpretação dos resultados.

A validação dos resultados pode ser feita através da utilização de um indicador métrico de erro que permita verificar até que ponto os valores de referência apresentados estão perto dos resultados reais. Um possível indicador de erro relativo é denominado MAPE (*Mean Absolute Percentage Error*), como está indicado na equação 1.

$$MAPE_{(\%)} = \frac{1}{N} \sum_{i=1}^N \left| 1 - \frac{Y_{pi}}{Y_{ri}} \right| \times 100$$

**Equação 1-** Análise de erro [2].

A definição dos parâmetros é a seguinte:

- $N$  – Número de elementos do conjunto;
- $Y_{pi}$  – O valor de referência com base no histórico;
- $Y_{ri}$  – O valor real.

A possibilidade de verificação do erro percentual entre o valor real e o valor de referência para um determinado período do dia com base no histórico, pode contribuir numa melhoria significativa na análise dos resultados obtidos.

Existe uma grande diversidade de algoritmos de *clustering*. O método a escolher deverá depender do tipo de aplicação pretendida e dos dados disponíveis. Podem aplicar-se ao mesmo tempo vários tipos de algoritmos sobre os mesmos dados, facilitando a comparação de resultados e a identificação do algoritmo mais eficaz para o tipo de dados em causa. Os algoritmos podem ser classificados de várias formas, sendo a mais usual a distinção entre métodos hierárquicos e métodos de partição.

Os algoritmos que utilizam o método hierárquico definem-se pela organização dos dados, estruturando-os por semelhança ou diferença. Cada elemento de uma matriz organiza-se de acordo com a sua relevância, dispondo-se de acordo com as variáveis seleccionadas, resultando na obtenção de uma hierarquia em partições ( $p_1, p_2, p_3, \dots, p_n$ ), integrantes num conjunto total de  $n$  objectos em  $k$  grupos [3].

Relativamente aos algoritmos que utilizem os métodos de partição, estes consistem na associação de um conjunto de indivíduos a grupos sem criar uma estrutura hierárquica. Cada partição deverá depois ser avaliada utilizando um critério específico, ou seja, para um determinado número de objectos ( $n$ ) e um número de partições ( $k$ ), deverão ser organizados os  $n$  objectos em  $k$  partições, sendo que cada partição representa um *cluster*. Os *clusters* são definidos de modo a otimizar um critério objectivo de partição, desta forma os objectos da mesma partição são similares entre si, no entanto objectos de *clusters* diferentes são diferentes nos seus atributos [3].

## **2.1. MÉTODO DE *CLUSTERING FUZZY C-MEANS***

Os *clusters* consistem num conjunto de dados agrupados segundo determinadas características ou atributos. O objectivo do agrupamento *Fuzzy* é o reconhecimento e quantificação desses conjuntos. Este tipo de atribuição consiste em dois processos, a identificação da importância de qualquer um dos dados em qualquer um dos grupos e a realização de um processo de iteração onde se pretende especificar a localização dos centros dos *clusters* [4].

Em aplicações reais os dados dificilmente se encontram organizados em conjuntos bem definidos, sendo possível uma sobreposição de alguns dos conjuntos. No entanto, é possível que o mesmo dado se encontre em *clusters* diferentes com diferentes graus de pertença. Existem dois tipos de métodos:

- O *fuzzy c-means* que consiste num parcelamento de *clusters*;
- O *fuzzy equivalence relation based hierarchical clustering* que tem como especificação relações de equivalência entre determinados *clusters*.

O método *fuzzy c-means* é um dos algoritmos mais usados de agrupamento *fuzzy*, consistindo numa modificação apresentada por Jim Bezdek de um método não difuso. O método tem como base o algoritmo Isodata [4] que consiste no cálculo do ponto central, implicando assim uma reorganização dos elementos aos respectivos *clusters*. Este método especifica o intervalo  $[1, n]$ , como o parâmetro de lógica difusa que determina a relevância ( $m$ ), também

denominado por grau de pertença. Por exemplo para  $m=1$  o resultado é uma atribuição não difusa dos elementos aos *clusters*, no entanto para  $m > 1$  a relevância entre os pontos no espaço de decisão aumenta [5].

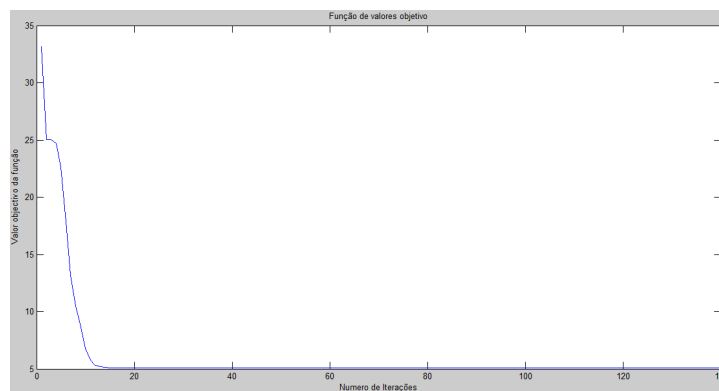
O algoritmo possui duas etapas, a primeira consistindo no cálculo dos centros dos *clusters*, e a segunda consistindo na atribuição dos pontos aos centros dos *clusters* com base na distância euclidiana, sendo esta tarefa iterativa até à obtenção de um centro nos *clusters* que seja estável. Um *cluster* estável consiste na obtenção de um grau de pertença único para cada elemento de um conjunto de dados, ou seja, o mesmo elemento não pode pertencer a dois *clusters* com um grau de pertença elevado, tendo sempre em consideração que a distância entre os limites do *cluster* em relação ao valor central será igual.

$$J = \sum_{i=1}^n \sum_{j=1}^p \mu_{ij}^m d(x_i, c_j)^2 \quad (2)$$

**Equação 2** – Função objectivo – *Fuzzy C-Means* [6].

Pretende-se com este algoritmo a minimização da função objectivo apresentada na equação 2, sendo a definição dos parâmetros a seguinte:

- $n$  – O número de dados;
- $p$  – O número de *clusters* considerados no algoritmo;
- $m > 1$  – Parâmetro de lógica difusa usando somente valores racionais, normalmente no intervalo [1,25;2];
- $x_i$  - Vector de dados de teste onde ( $i= 1,2,\dots, n$ ), sendo cada posição do vector um atributo do conjunto de dados;
- $c_j$  – Ponto central de um agrupamento *fuzzy* ( $j = 1,2,\dots, p$ );
- $d(x_i, c_j)^2$  – Distância Euclidiana ao quadrado entre  $x_i$  e  $c_j$ .



**Figura 1** – Exemplo gráfico de iterações.

A figura 1 ilustra a evolução da função objectivo em função das iterações de *clustering*. Para a realização do cálculo deverão ainda ser seguidos alguns passos. Deverá ser iniciada a variável  $\mu$  com um valor entre zero e um, sendo calculado o centro do *cluster* usando a equação seguinte.

$$c_j = \frac{\sum_{i=1}^n \mu_{ij}^m x_i}{\sum_{i=1}^n \mu_{ij}^m} \quad (3)$$

**Equação 3** - Função de centro de *cluster* J - *Fuzzy C-Means* [6].

Após calcular o valor inicial para J usando a equação 2, deverá ser calculada a tabela da função de relevância *fuzzy*  $\mu$  conforme a equação seguinte.

$$u_{ij} = \frac{\left(\frac{1}{d(x_i; c_j)}\right)^{\frac{2}{m-1}}}{\sum_{k=1}^p \left(\frac{1}{d(x_i; c_k)}\right)^{\frac{2}{m-1}}} \quad (4)$$

**Equação 4** - Função de relevância - *Fuzzy C-Means* [6].

Mesmo que se utilize uma boa parametrização do algoritmo, existem sempre alguns pontos que não se identificam em nenhum dos grupos, sendo estes denominados por *outliers*. Algumas das técnicas de manipulação de dados têm dificuldades em trabalhar os *outliers* mais difíceis no entanto, alguns métodos de *fuzzy* atribuem a estes pontos um grau de relevância muito superior ou inferior relativamente aos *clusters* definidos, consoante este seja um valor anómalo inferior ou superior relativamente ao valor padrão do *cluster* [5].

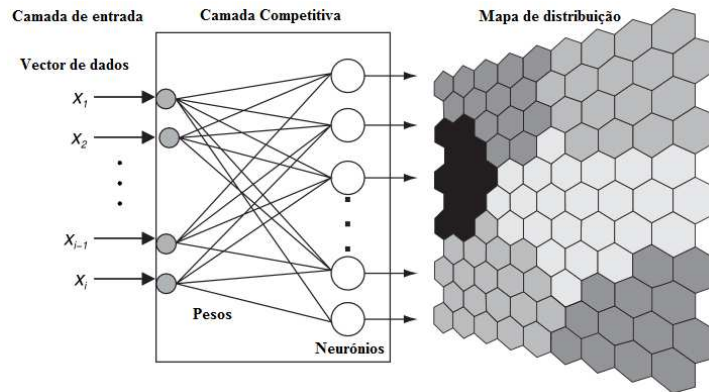
## 2.2. MÉTODO DE KOHONEN

O método de *Kohonen* consiste na implementação de um modelo que simula uma actividade cerebral. Esta teoria baseia-se na premissa que os nós de saída procuram representar uma resposta adequada à informação presente à entrada. Este comportamento tem como base a simulação de uma actividade cerebral, daí a designação deste método como rede neuronal.

O elemento base de uma rede neuronal é o neurónio, aqui visto como um elemento que processa informações que lhe chegam via ligações de entrada. As ligações entre os neurónios são designadas sinapses.

Existem dois tipos de camadas interligadas entre si a camada de entrada e a camada competitiva [7]. Na camada de entrada os neurónios transmitem somente as interacções que

recebem, no caso da camada competitiva os neurónios processam a informação resultante dos resultantes neurónios, quer sejam da camada de entrada quer da própria camada competitiva. A resposta de uma rede neuronal deste tipo consiste numa interacção de sinapses entre as duas camadas [7]. O peso das diferentes sinapses vai evoluindo ao longo do processo de aprendizagem, resultando numa auto-organização da rede que leva à identificação dos diferentes *clusters* (Figura 2).



**Figura 2** – Mapa de ligações do método *Kohonen*.

Podemos resumir-se o algoritmo em cinco principais etapas:

- 1) Inicialização – Selecção de valores aleatórios para os vectores iniciais de peso  $w_j(0)$ , sendo ‘z’ o número de neurónios na base de dados. Dependendo da aplicação é aconselhável o uso de um valor baixo nos pesos.
- 2) Amostragem – Escolha de um valor de amostra  $x$  do espaço de entrada com uma determinada probabilidade. O vector  $x$  consiste na representação do padrão de estímulos do conjunto.
- 3) Assimilar estímulo – Ao utilizar o critério de distância mínima deverá obter-se o neurónio vencedor.

$$i(x) = \operatorname{argmin}_j \|x(n) - w_j\|, j = 1, 2, \dots, z \quad (5)$$

**Equação 5** – Critério da mínima distância euclidiana [7].

Definição dos parâmetros:

- $i(x)$  – Identifica o neurónio que conjuga melhor com o vector de entrada  $x$  (neurónio vencedor);
- $x(n)$  – Vector de entrada de dados;
- $w_j$  – Vector dos pesos sinápticos;
- $z$  – Número total de neurónios.

4) Actualização – usando a fórmula de actualização deverá ajustar-se os pesos de todos os neurónios.

$$w_i(n+1) = w_j(n) + \eta(n)h_{j,i(x)}(n)(x(n) - w_j(n)) \quad (6)$$

**Equação 6** – Actualização dos neurónios das redes neuronais [7].

Definição dos parâmetros:

- $\eta(n)$  - Taxa de aprendizagem;
- $h_{j,i(x)}(n)$  – Função de vizinhança central no neurónio vencedor  $i(x)$ .

A vizinhança do neurónio vencedor refere-se ao nível de interligação entre o neurónio  $i$  e o neurónio  $j$  (neurónio na sua vizinhança). Esse valor é simétrico relativamente ao neurónio  $j$ , decrescendo gradualmente consoante o aumento da distância euclidiana entre os neurónios. O distanciamento da vizinhança diminui com o tempo, implicando assim valores  $h_{j,i(x)}$  menores ao longo do tempo definindo deste modo uma vizinhança mais estreita.

5) Finalização – As etapas anteriores irão decorrer até não se verificarem alterações significativas.

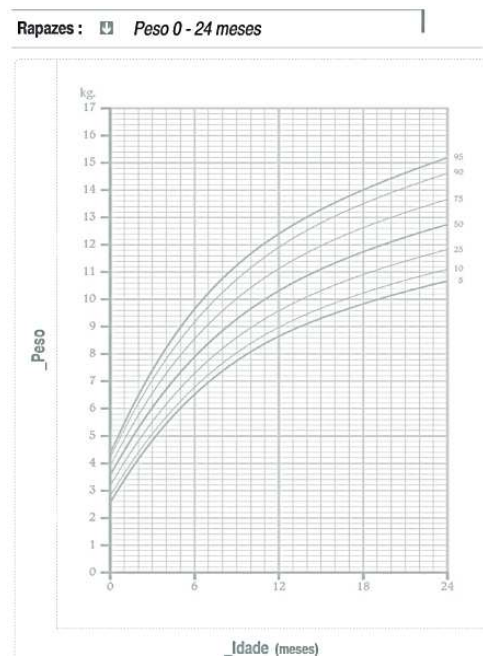
### 2.3. MÉTODO DE PERCENTIS

Apesar do método de *Percentis*, não se tratar de um método clássico de classificação, este consiste na utilização de ferramentas estatísticas que permitem a definição de uma separação de dados bem estruturados, podendo esta ser uma ferramenta alternativa aos métodos referidos.

A separação de classes com base em *percentis* consiste na separação ordenada por ordem crescente de uma base de dados em cem partes iguais. Para um conjunto de dados  $B_r$  (sendo 'B' o conjunto da base de dados e 'r' a razão em *percentis*), a distribuição dos elementos implica a correspondência de um conjunto de dados a uma dada percentagem, sendo que nesse conjunto os valores são semelhantes. Como se trata de valores crescentes o percentil um é um conjunto sempre com um valor inferior a um percentil superior, como por exemplo com o valor cinco.

Sendo os percentis um caso particular dos quartis, estes são constituídos pela separação de um conjunto de dados em quatro (ou quartil), ou seja, separa a amostra em quatro partes com uma percentagem de dados idêntica. O primeiro e terceiro quartil são conhecidos como percentil 25 e 75, sendo a mediana considerado o percentil 50.

Uma forma fácil de entender no que consiste um percentil é na comparação feita pelos pediatras, quando recorrem a comparação da evolução do bebé ou criança relativamente a parâmetros já estabelecidos. Por exemplo no caso de um bebé se encontrar com um percentil 86 em termos de peso, refere-se que 86% dos indivíduos para a idade idêntica do bebé em causa, dispõem de um peso igual ou inferior ao apresentado. Para o caso do percentil ser 50, refere-se que o bebé encontra-se dentro de um valor intermédio para a idade, ou seja 50% dos bebés têm um peso igual ou inferior ao apresentado (figura 3).



**Figura 3** – Exemplo de um gráfico em *percentis* [8].



### **3. PROPOSTA DE IMPLEMENTAÇÃO**

A aplicação desenvolvida utiliza uma ferramenta matemática que estuda e implementa um programa de análise de consumos, estabelecendo critérios essenciais no seu processamento. No estudo comparativo, após a obtenção de valores de consumos eléctricos, são definidas categorias consoante o tipo de consumo em cada instante. Posteriormente, estes resultados serão comparados com os valores actualmente em análise, dando origem a informações referentes às diferenças encontradas entre os dois padrões de consumo. O funcionamento da aplicação elaborada processa-se em tempo real ou em ambiente de simulação.

#### **3.1. DESCRIÇÃO DAS PROPOSTAS DE IMPLEMENTAÇÃO**

As propostas de implementação incidiram na aplicação de dois métodos: um dependente da utilização de ferramentas de *clustering* através do método *Fuzzy C-Means* (tendo sido para este método desenvolvidas três aplicações consoante as funcionalidades pretendidas em cada uma) e outro método recorrendo ao uso de uma ferramenta de divisão com base em valores estatísticos (método *Percentis*). Relativamente ao primeiro método, a primeira opção serviu como comparação com o método de classificação estatística, sendo esta mais simples que as restantes, não dispondo de algumas funcionalidades opcionais, todavia têm as mesmas características de implementação que o método *Percentis*. A segunda opção do método *Fuzzy C-Means* consiste na aplicação de uma abordagem com base no histórico de consumo e a terceira corresponde a uma abordagem em tempo real.

##### **3.1.1. MÉTODO FUZZY C-MEANS**

Com o método *Fuzzy C-Means* e recorrendo a primeira opção, a classificação das classes vai recorrer ao histórico de consumos já existente. Neste caso perante toda a base de dados irá classificá-la por inteiro, dividindo a mesma por patamares limite não dando ênfase se o valor de consumo é referente ao dia anterior ou há um ano atrás, dividindo somente os valores consoante a sua semelhança. De seguida, como nem todas as bases de dados têm o mesmo tamanho, foi seguido um critério que irá executar um varrimento da base de dados procurando valores que sejam semelhantes para o mesmo período temporal, sendo a definição do que a aplicação considera como valores de consumo do mesmo período temporal dependente do tamanho da base de dados.

Ao se mencionar a divisão da análise das bases de dados por etapas consoante a sua dimensão, pretende-se que seja adoptada a opção que reúna um maior conjunto de dados de comparação sem perder fiabilidade temporal. Isto é, pretende-se que num conjunto de dados estes sejam o mais semelhante possível com o valor actual. Caso existam dados inferiores a um dia, os valores de comparação de resultados serão segmentados pelos intervalos temporais de quinze minutos. Para uma base de dados com tamanho até uma semana, os períodos de segmentação serão referentes à hora e minutos específicos da potência activa actual. No caso de uma base de dados com o tamanho de uma semana a um ano, o período será referente ao dia da semana, ou seja, se é uma segunda-feira, terça-feira ou outro dia da semana, considerando ainda o referido no caso anterior. Entre um ano e sete anos é idêntico ao anterior, todavia vai limitar o intervalo ao mês pretendido para confrontação. E, por último, para uma base de dados superior a sete anos, irá confrontar o dia da semana com o número da semana do calendário e as condições existentes no segundo caso.

Exemplo para o período temporal (04/08/2014 – 16h30m):

- 1) Menor que um dia: Equipara com todos os intervalos de trinta minutos;
- 2) Entre um dia e uma semana: Compara com todos os períodos horários das 16h30m;
- 3) Intervalo entre uma semana e um ano: Compara todas as segundas-feiras com o horário 16h30m;
- 4) Entre um ano e sete anos: Confronta com todas as segundas-feiras às 16h30m existentes no mês de Agosto;
- 5) Maior que sete anos: Confronta a semana trinta e dois, que sejam igualmente segundas-feiras e com valores referentes às 16h30m.

Esta implementação pretende verificar de que forma este padrão de consumos se mantém ou varia consoante o valor residual do diferencial de consumos. Uma das particularidades deste exemplo é o ajuste automático da comparação de resultados consoante a dimensão da base de dados, ou seja, a padronização do tipo de consumos para comparação vai sendo alterada progressivamente consoante maior ou menor for o volume de dados existentes para comparação.

### 3.1.2. MÉTODO *PERCENTIS*

Neste caso pretendeu-se a aplicação da segmentação de *clusters* sem recorrer às ferramentas do método *Fuzzy C-Means*. A actual aplicação consiste na separação do padrão de consumos, por um limite matemático relativo a valores intermédios de consumos. Com esta opção pretende-se verificar que ao dividir os consumos por um valor rígido, não é possível categorizar os valores mediante a dimensão de cada *cluster*, mas sim por um valor intermédio que poderá até nunca ser atingido. Por exemplo se existir uma grande diferença entre os consumos máximos e mínimos e nestes se encontrarem uma grande percentagem dos valores totais de consumos, as restantes divisões encontram-se com grau elevado de separação de classes. No entanto recorrendo às ferramentas de *clustering* é possível reorganizar os dados, permitindo assim um estudo de consumos de forma mais ponderada e objectiva.

Não se considerando a aplicação do método *Fuzzy C-Means* para a separação por classes, a categorização de resultados só é possível após a obtenção do valor médio de potência activa da base de dados, do valor mínimo, do valor máximo e do desvio padrão. A partir desses resultados é elaborada uma fórmula matemática que define a classificação em cinco classes. Caso se pretenda aumentar ou diminuir o número de classes para este método é necessário proceder à alteração do código da aplicação, para que este seja parametrizável.

Sabendo que ao considerar o desvio padrão o valor de consumo estará com 95% de certeza nesse intervalo seguindo uma distribuição normal [9], pretende-se excluir os valores mais antagónicos ao valor médio, relegando-os para as classes limites inferior e superior. Sendo considerado o desvio padrão para a constituição de classes, a classificação encontra-se definida da seguinte forma:

- 1) A primeira classe é constituída para valores inferiores a 33% do valor negativo do desvio padrão relativamente ao valor médio;
- 2) A segunda classe é constituída para valores inferiores entre 33% e 50% do valor negativo do desvio padrão relativamente ao valor médio;
- 3) A terceira classe é constituída para valores inferiores entre 50% do valor negativo do desvio padrão e valores superiores a 33% relativamente ao valor médio;
- 4) A quarta classe é constituída para valores superiores entre 33% e 50% do valor positivo do desvio padrão relativamente ao valor médio;

5) A quinta classe é constituída para valores superiores a 50% do valor positivo do desvio padrão relativamente ao valor médio.

### **3.1.3. MÉTODO *FUZZY C-MEANS* COM REFERÊNCIA ESCOLHIDA PELO UTILIZADOR**

Na utilização do método *Fuzzy C-Means* nesta segunda opção recorre-se aos *clusters*, tal como no primeiro caso, sendo os mesmos obtidos recorrendo à ferramenta FCM do *software Matlab*. Esta ferramenta implementa o algoritmo *fuzzy* para a obtenção dos resultados separados por classes, separando toda a base de dados em todo o seu eixo temporal em *clusters*. A diferença comparativamente com a primeira abordagem é que os valores que serão comparados com os consumos actuais em análise, não serão mediante uma média de valores para um intervalo estipulado consoante a dimensão da base de dados, mas sim consoante um valor atribuído pelo utilizador. Ou seja, neste caso o utilizador tem a liberdade de escolher qual será o atraso relativamente ao dia actual com o qual os valores na base de dados serão comparados. Será dada a informação ao utilizador do tamanho da base de dados e da estimativa de dias anteriores ao actual que poderão existir para comparação. A vantagem desta solução é que numa base de dados com muitos valores, existe a possibilidade de comparar para um dia anterior próximo do actual. Permite ainda comparar com um dia específico que o utilizador considere como semelhante ao actual. Uma das vantagens desta aplicação consiste na possibilidade de qualquer alteração de consumos de uma instalação eléctrica, poder dar valores mais ajustados à nova realidade. Esta alteração de consumos pode ser devida à utilização de novos equipamentos que impliquem a alteração do valor de potência activa consumida pelo utilizador ou mesmo devido a alteração dos padrões horários actuais.

São ainda apresentadas como informações adicionais a comparação entre o valor actual de potência activa e o valor de potência activa nas horas anteriores. Neste caso a comparação de valores é referente ao atraso seleccionado para comparação pelo utilizador e não um valor fixo definido automaticamente pelo programa, consoante o tamanho da base de dados usado nos casos anteriores.

### **3.1.4. MÉTODO *FUZZY C-MEANS* EM TEMPO REAL**

A última aplicação do método *Fuzzy C-Means* é idêntica ao referido na opção anterior, a diferença consiste na obtenção de resultados e alarmes em tempo real, possibilitando uma

abordagem a um caso que decorra em tempo real. Esta opção permite verificar a dinâmica do sistema, facilitando ao utilizador a detecção de anomalias em contexto real e em tempo real, permitindo a resolução das mesmas, limitando assim possíveis danos posteriores resultantes da ocorrência de consumos, que se encontrem fora do padrão normal.

### **3.2. DESCRIÇÃO RESUMIDA DO FUNCIONAMENTO DAS APLICAÇÕES**

A programação desenvolvida para os dois métodos e diferentes abordagens é semelhante. Conforme poderá ser observado nos fluxogramas que seguem em anexo, a primeira etapa antes da obtenção de dados ou da análise dos mesmos é a verificação da restrição temporal actual (as aplicações só são executadas de quinze em quinze minutos).

Caso se encontre activa a funcionalidade de separação da base de dados por *clusters* ou por *percentis*, a todos os valores da base de dados será atribuída uma classificação que permitirá a análise dos resultados. Durante o período temporal em que não se procede a classificação ou reclassificação da base de dados, o valor de potência activa referente a um novo instante temporal será classificado de acordo com o método em análise em uso.

Após efectuada as classificações são definidos alarmes caso existam, alertando assim o utilizador para eventuais anomalias que possam existir no sistema eléctrico. Por último são elaborados gráficos que possibilitem facilmente verificar todas as informações relevantes para o utilizador.

### **3.3. LEITURA DE DADOS PELAS APLICAÇÕES DESENVOLVIDAS**

Antes de explicar o funcionamento do interface gráfico é necessário salientar como se encontram implementadas as bases de dados e como estas se encontram interligadas com as aplicações. Consoante a aplicação em análise, os dados poderão ser introduzidos manualmente através de uma base dados fornecida ao utilizador ou poderá a mesma ser construída com base na obtenção de dados, através de um método externo de aquisição de dados.

Para a aplicação com o Método *Fuzzy C-Means* com execução em tempo real, deverá existir a interligação com uma base de dados externa que será lida por esta aplicação. A configuração da base de dados dever estar de acordo com o descrito em anexo nas “Configurações dos Softwares Matlab e HeidiSql”. Sendo possível fazer a interligação da base de dados com o Matlab este irá ler o último elemento da base de dados e irá gravar numa base de dados

própria com todas as informações necessárias, para a execução da análise de consumos. Salienta-se que caso a base de dados contenha dados inferiores a um período de um dia, a aplicação ficará em “*Loop*” contínuo até à obtenção de valores que permitam fazer uma análise de consumos correcta.

Para as restantes aplicações, as bases de dados deverão estar configuradas de forma a que o programa consiga interpretá-las correctamente. Para a sua execução, os dados devem ser introduzidos através da implementação de uma tabela no *Matlab*, com todos os valores referentes aos valores de potência activa conhecidos. A descrição e o método de implementação dessas bases de dados estão explicados com mais detalhe nos anexos deste relatório. Neste subcapítulo não se encontram descritas em pormenor as configurações necessárias das bases de dados pois tal situação não se justifica, no contexto em que se pretende expor somente a definição das características das aplicações.

### **3.4. ESTRUTURA DA APRENDIZAGEM**

Com excepção da classificação de consumos com base no método de *Percentis*, os restantes métodos irão ser actualizados relativamente a classificação de dados, perante duas situações distintas no caso de alteração do número de *clusters* e perante uma transição horária para um novo dia.

As aplicações permitem que o utilizador implemente a classificação da base de dados consoante o número de *clusters* que pretender, no entanto sempre que esta classificação é alterada é necessário executar novamente a classificação dos dados consoante os novos parâmetros, sendo possível executar essa alteração em qualquer período horário.

As aplicações têm também implementado na sua rotina, sempre que exista uma transição horária para um novo dia (00h00min), a execução de uma nova classificação da base de dados. Por diversos motivos pode existir num curto espaço de tempo alterações no padrão de consumos da instalação, tal como a existência de dias atípicos que justifiquem a classificação dos dados de outra forma. Por esse motivo, a reclassificação da base de dados após a transição horária para um novo dia é necessária, permitindo assim uma reorganização dos dados e uma apresentação dos valores com uma tipologia mais adequada aos resultados actuais.

### 3.5. CARACTERIZAÇÃO DE UM OUTLIER NAS APLICAÇÕES

Foi introduzido nestas aplicações um alerta referente à detecção de *outliers*. A implementação de um sistema de detecção de *outliers* consiste na identificação de um valor de potência eléctrica fora do normal referente a um pico de consumo eléctrico. Normalmente um valor tem que ter um elevado grau de pertença para o *cluster* em que foi identificado para o método *Fuzzy C-Means*. Por exemplo no último *cluster*, definido como superior, se os valores próximos do centro do cluster forem semelhantes, ou seja, com uma pequena variação, e surgir um valor extremo que seja em muito superior aos intervalos que haviam sido definidos para o *cluster*, este valor pode vir a ser considerado como *outlier*.

Na implementação do método *Fuzzy C-Means*, foi considerado que o grau de pertença na classe superior com o valor inferior a 0,4 (40%) é um evento anómalo, devendo esta situação ser um forte alerta para uma análise profunda por parte do utilizador, na identificação da causa do seu aparecimento. Na aplicação de análise de resultados com base em percentagens (Método *Percentis*), pode igualmente referir-se que os *outliers* serão valores fora do conjunto do desvio padrão. Considerando uma distribuição normal dos valores da base de dados na utilização da ferramenta estatística, os valores que estarão fora da gama de média serão aproximadamente referentes a duas vezes o valor do desvio padrão. Para o caso das aplicações desenvolvidas os valores considerados no último nível, próximo de zero, não foram considerados como *outliers*, apesar do sistema devolver alarmes nas aplicações não serão assinalados como *outliers*, pois estes podem ser resultantes de um sistema que normalmente se encontra desligado em alguns períodos do dia.

### 3.6. FUNCIONAMENTO DO INTERFACE GRÁFICO

Nas aplicações desenvolvidas é necessário dispor das versões de Matlab 2012, MySql 1.1 e HeidiSql 8.3.0.4694 ou superior. Os *softwares* referidos, ao trabalharem em simultâneo, irão utilizar uma quantidade de recursos de sistema elevados, pelo que os sistemas deverão ser instalados em unidades de processamento com elevada capacidade computacional. O *software Matlab* permitiu o desenvolvimento das aplicações referidas, enquanto os restantes *softwares* permitiram a interligação das aplicações com uma base de dados.

Para existir uma boa interligação entre os vários sistemas, as aplicações deverão ser executadas numa só pasta. A pasta seleccionada para a execução das aplicações deverá

permitir escrita. Caso se encontrem em pastas diferentes ou não permita escrita, o sistema ao ser iniciado não será executado. Durante o processamento de dados serão guardadas algumas variáveis que serão utilizadas no painel de comando *Graphical User Interfaces* (GUI). A Figura 4 é um exemplo das informações apresentadas ao utilizador.



Figura 4 – Painel de comando - GUI.

Para o arranque da aplicação é necessária a existência de quatro variáveis, que são atribuídas pelo utilizador e guardadas através do painel de comando. As variáveis requeridas são “actualização\_2”, “clusters”, “arranque” e “num\_comparar”. A primeira variável verifica a existência de alteração de parâmetros iniciais que implique o reajuste dos parâmetros de análise. A segunda variável é referente ao número de “clusters”, com o qual o programa executa a separação de dados para análise. A variável “arranque” estipula se o programa está preparado para a sua inicialização, ou se o programa se encontra a executar as tarefas finais antes do seu encerramento. E por último, a variável “num\_comparar” especifica qual o atraso pretendido a ser definido, como comparação de resultados (Figura 5).

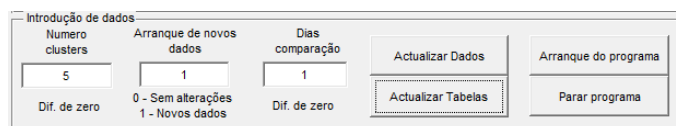


Figura 5 – Parâmetros iniciais para arranque do programa - GUI.

As restantes opções do painel servem somente como visualização, proporcionando a visualização de tabelas com valores relevantes da execução da aplicação, tal como gráficos que permitem ao utilizador ter uma ideia da potência activa consumida na instalação eléctrica



em causa. Enquanto o programa não for executado pela primeira vez não serão apresentados quaisquer resultados.

A tabela de comparação de potência activa permite não só a visualização do diferencial da potência activa entre a data actual e a data estipulada para comparação, mas permite também comparar o tipo de classes existentes nas diversas situações, tal como o nível de alerta para cada caso (Figura 6).

Hora [H]	Minutos [M]	Diferencial Pot. Act. [kW]	Pot. Act. dia actual [kW]	Pot. Act. esperado [kW]	Classes Actuais	Classes Anteriores	Tipo de alerta
0	0	0.0840	0.1040	0.0200	4	1	3
0	15	-0.0760	0.0200	0.0960	1	4	1
0	30	0.0800	0.1040	0.0240	4	2	3
0	45	-0.0760	0.0160	0.0920	1	3	1
1	0	0.0800	0.1080	0.0280	4	2	3
1	15	-0.0720	0.0160	0.0880	1	3	1
1	30	0.0520	0.1000	0.0480	4	2	3
1	45	-0.1000	0.0240	0.1240	2	4	1
2	0	0.0680	0.0960	0.0280	4	2	3

Figura 6 – Exemplo de tabela de comparação de potência – GUI.

Estão implementadas duas tabelas informativas dos estados funcionais do programa **Alerta** e **Histórico** (Figura 7), facilitando a identificação dos estados do programa desde a sua inicialização. A tabela Alerta é referente aos alertas identificados pela aplicação desde a sua primeira utilização, contudo não menciona os estados considerados como normais pelo sistema. No caso da tabela Histórico, são considerados todos os estados funcionais com ou sem anomalias desde o início do programa. Consoante o tipo de alerta é atribuída uma cor específica a cada um, estabelecendo assim um efeito visual que permite ao utilizador verificar com uma maior rapidez o tipo de alerta existente.

Alarmes		N.º de alerta actual
Alertas		3
19-Sep-2014 23:15:00 - Possível Anomalia potencia activa media superior ao normal		3
Historico		
19-Sep-2014 23:15:00 - Possível Anomalia potencia activa media superior ao normal		3
		0 - Sem dados
		1 - Consumo inferior (AZ)
		2 - Consumo Normal (VD)
		3 - Consumo Superior (VR)
		4 - Outlier (AM)

Figura 7 – Exemplo de tabelas de alarmes – GUI.

Existe ainda a indicação do tipo de alerta actual, sendo que cada alerta é identificado com um número. A legenda indicativa do tipo de alerta encontra-se referida numa janela junto às tabelas. Os alertas existentes podem ser dos seguintes tipos:

- Tipo '0' – Não existem dados actuais para emitir qualquer tipo de alerta;
- Tipo '1' – A potência activa actual encontra-se numa categoria inferior ao estipulado para comparação;

- Tipo ‘2’ – A potência activa actual encontra-se na mesma categoria do valor estipulado para comparação;
- Tipo ‘3’ – A potência activa actual encontra-se numa categoria superior ao estipulado para comparação;
- Tipo ‘4’ – A potência activa actual encontra-se numa categoria superior ao estipulado para comparação. O grau de pertença da classe em causa encontra-se com um valor bastante distanciado do centro do *cluster*, ou seja, o valor pode ser resultante de um pico de consumo que diverge dos parâmetros normais.

No decorrer da análise da aplicação são facultados ao utilizador parâmetros referentes à evolução da potência activa nas últimas horas. Esta informação é relevante para a decisão, por parte do utilizador, da eventual necessidade de se proceder a alguma alteração dos parâmetros introduzidos inicialmente, sendo útil igualmente para a observação da evolução do diagrama de carga nos instantes mais imediatos (Figura 8). As informações disponibilizadas são as seguintes:

- Diferença de potência activa uma hora atrás – Consiste na comparação do valor actual de potência activa com a média dos valores de potência activa registada na hora anterior à actual;
- Diferença de potência activa duas horas atrás – Consiste na comparação do valor actual de potência activa com a média dos valores de potência activa registada duas horas anteriores à actual;
- Diferença de potência activa três horas atrás – Consiste na comparação do valor actual de potência activa com a média dos valores de potência activa registada três horas anteriores à actual;
- Diferença de potência activa quatro horas atrás – Consiste na comparação do valor actual de potência activa com a média dos valores de potência activa registada quatro horas anteriores à actual;
- Diferença de potência entre uma e duas horas atrás – Consiste na comparação média dos valores de potência activa registadas entre uma e duas horas anteriores à actual;
- Diferença de potência activa entre duas e três horas atrás – Consiste na comparação média dos valores de potência activa registadas entre duas e três horas anteriores à actual;
- Diferença de potência activa entre três e quatro horas atrás – Consiste na comparação média dos valores de potência activa registadas entre três e quatro horas anteriores à actual;
- Classe actual – Indica a classe actual da potência activa relativamente à hora do sistema;

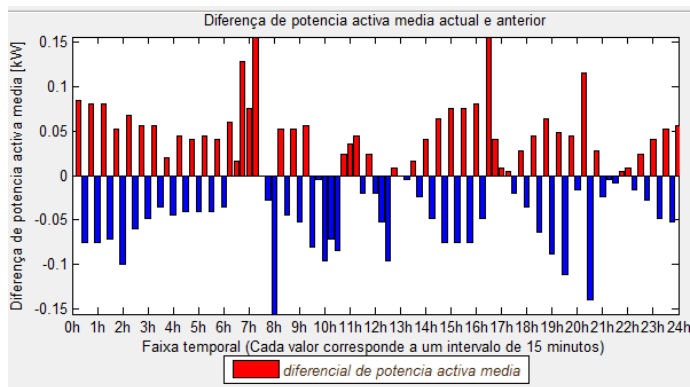
- Classe anterior – Indica a classe estipulada da potência activa relativamente à hora do sistema;
- Potência activa actual – Indica a potência activa actual relativamente à hora do sistema;
- Potência activa anterior – Indica a potência activa considerada como comparação relativamente à hora actual do sistema;
- Número de *clusters* actuais – Refere o número de *clusters* que o sistema considera actualmente na organização dos valores de potência activa;
- Número de dias para comparação de resultados – Menciona o número máximo de dias que podem ser considerados para comparação de resultados.

Visualização de dados de consumo				
diferença pot. act. 1 hora atras	diferença pot. act. 2 hora atras	diferença pot. act. 3 hora atras	diferença pot. act. 4 hora atras	Numero clusters actuais
0.0040 [kW]	0.0010 [kW]	-0.0920 [kW]	-0.0310 [kW]	5
diferença pot. act. entre 1 e 2 horas atras	diferença pot. act. entre 2 e 3 horas atras	diferença pot. act. entre 3 e 4 horas atras	Classe actual	Pot. Act. actual [kW]
-0.0030 [kW]	-0.0930 [kW]	0.0610 [kW]	3	0.0680 [kW]
			Classe anterior	Pot. Act. anterior [kW]
			1	0.0160 [kW]
Numero de dias para comparação de resultados				421 [dias]

**Figura 8** – Exemplo de dados de consumo – GUI.

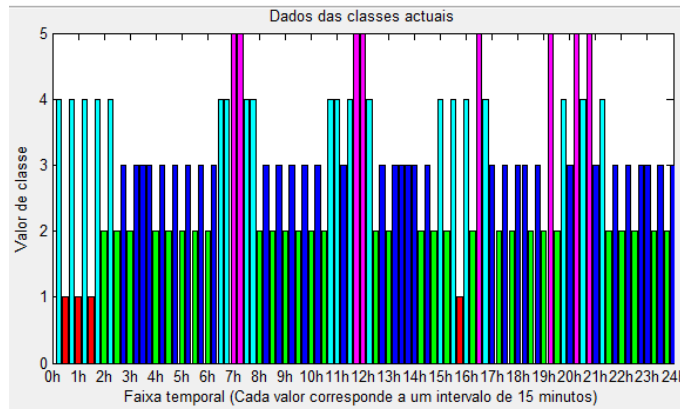
Existem dois ambientes gráficos que asseguram a comparação de resultados, ou seja, é possível seleccionar individualmente qual o tipo de gráfico a visualizar pelo utilizador, sendo que ambas as janelas apresentam os mesmos gráficos. Perante a possibilidade de comparação gráfica de resultados é possível verificar com maior detalhe quais as diferenças entre os parâmetros actuais e os estipulados para comparação. Os gráficos existentes para comparação são os seguintes:

- 1) Diferencial de potência activa – Mostra graficamente o valor diferencial de potência activa entre o dia actual e o estipulado para comparação (Figura 9);



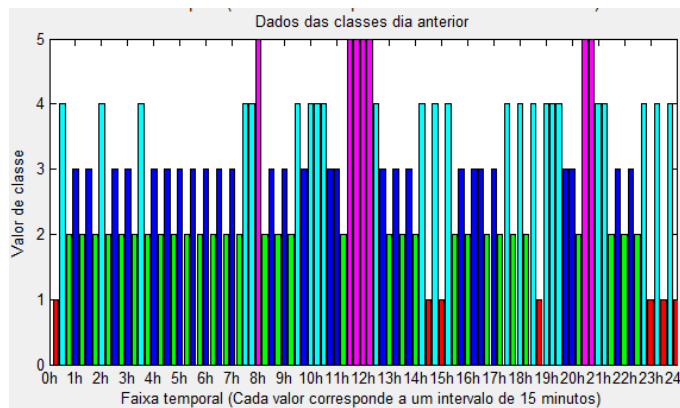
**Figura 9** – Exemplo de gráfico diferencial.

2) Classe actual – Exibe o valor das classes para o dia actual em análise (Figura 10);



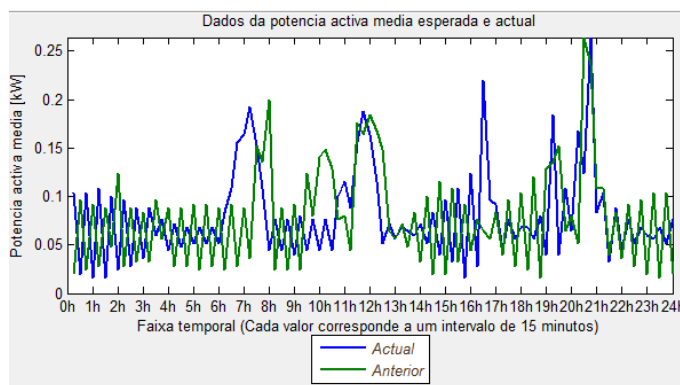
**Figura 10** – Exemplo de gráfico de classe actual.

3) Classe anterior – Apresenta o valor das classes para o dia estipulado como comparação (Figura 11);



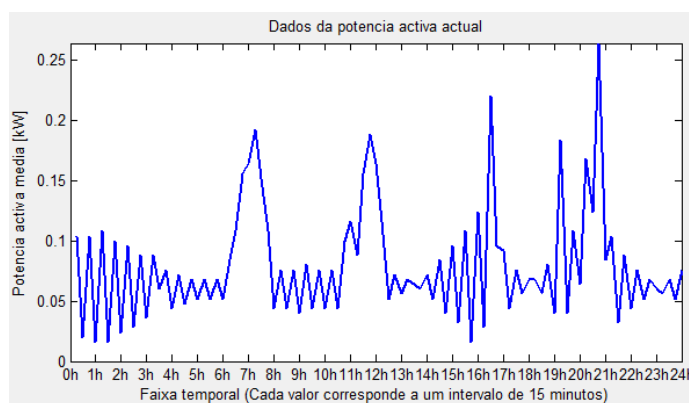
**Figura 11** – Exemplo de gráfico de classe anterior.

4) Ambas potências activas – Indica graficamente a potência activa para o dia actual e o dia estipulado como comparação (Figura 12);



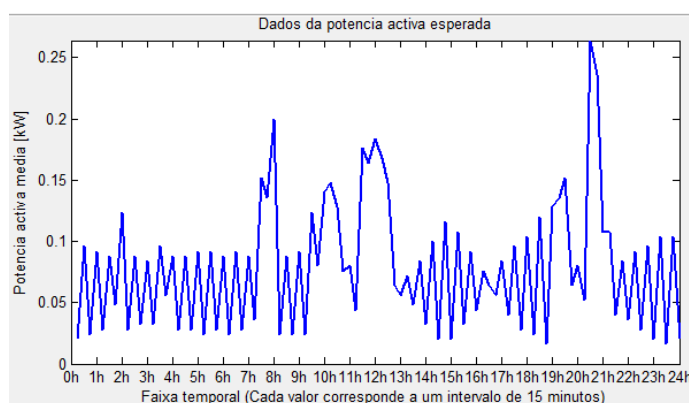
**Figura 12** – Exemplo de gráfico de potências activas.

5) Potência activa actual – Exibe a potência activa para o dia actual em análise (Figura 13);



**Figura 13** – Exemplo de gráfico de potência activa do dia actual.

6) Potência activa esperada – Apresenta a potência activa para o dia estipulado como comparação (Figura 14).



**Figura 14** – Exemplo de gráfico de potência activa esperada.

O caso especificado consiste no interface para o programa de análise em tempo real, apesar dos valores diferenciais apresentados nos gráficos das figuras anteriores serem elevados as imagens apresentadas são somente ilustrativas (Figuras 9, 10, 11, 12, 13 e 14). Os gráficos referidos neste subcapítulo foram elaborados utilizando valores arbitrários, não se tratando de valores num espaço temporal bem definido. Como neste subcapítulo apenas se pretende expor o funcionamento genérico das aplicações, não se enquadrou uma análise profunda dos dados introduzidos. Para as restantes aplicações desenvolvidas algumas das funcionalidades referidas não se encontram disponíveis por não serem necessárias para o seu funcionamento, sendo utilizados nesses casos metodologias de análise diferentes.



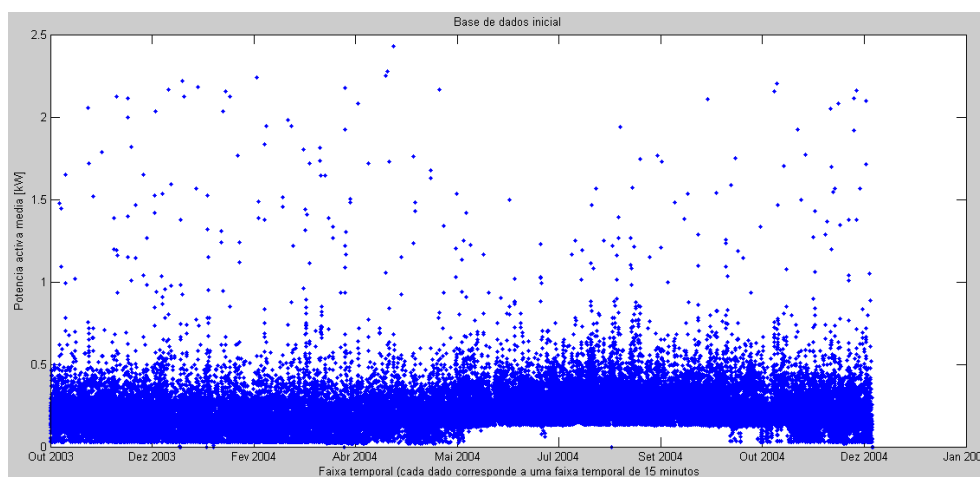
## 4. ESTUDO DE CASOS

Os casos estudados recorreram a dois tipos de bases de dados. Dois casos são residenciais (DB1 e DB2) e usam dados históricos de consumos de consumidores reais disponibilizados pelo INESC - Coimbra. O terceiro caso (DB3) corresponde às instalações da Juventude Desportiva do Lis, para a qual foi possível aceder em tempo real à base de dados disponibilizada pelo seu sistema de gestão de consumos.

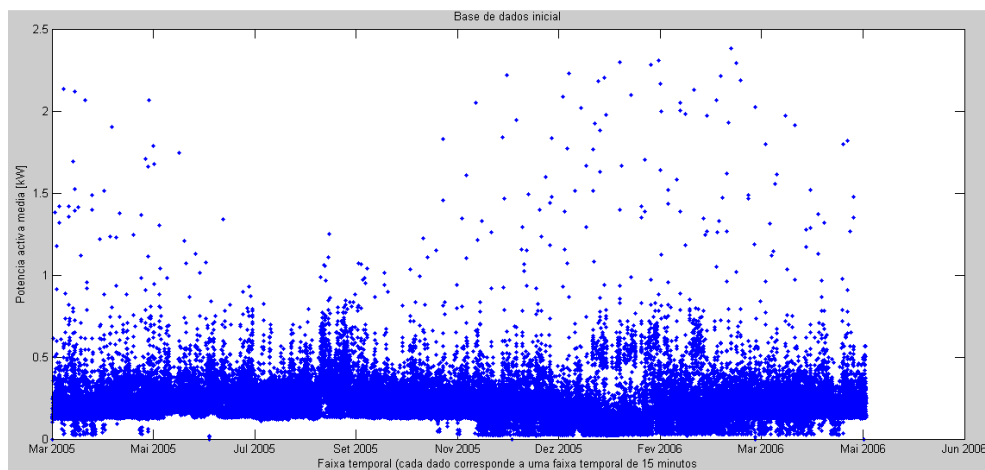
Os três casos foram estudados com as duas abordagens já descritas no capítulo anterior, o método de *Percentis* e o método *Fuzzy C-Means*, este último com três funcionalidades diferentes.

### 4.1. APLICAÇÃO DO *FUZZY C-MEANS*

Existem várias formas de apresentar os resultados da base de dados, possibilitando em alguns casos a verificação de resultados com uma maior ou menor rapidez e facilidade. As figuras seguintes apresentam graficamente todos os valores da base de dados, sendo cada ponto referente a um valor de potência activa. Esta apresentação é referente ao conjunto total de valores num espaço temporal contínuo (eixo dos xx's) e à potência de activa (eixo dos yy's).



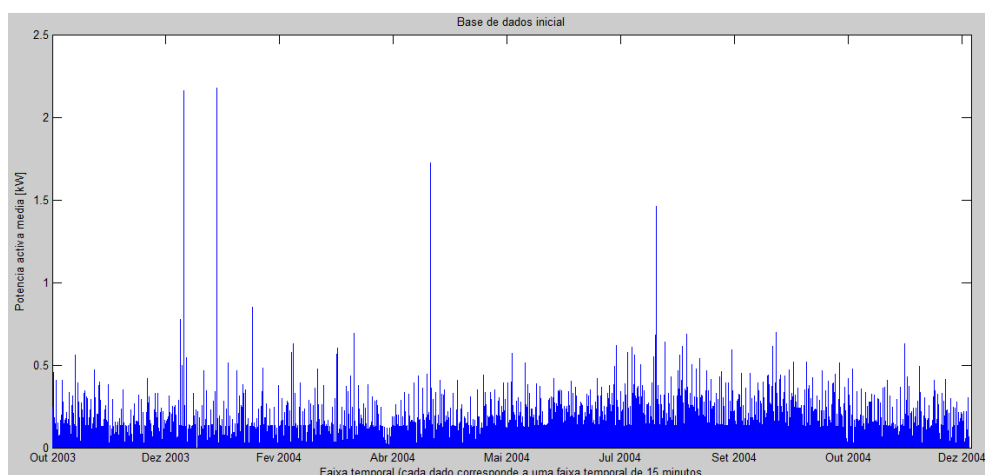
**Figura 15** – Dados com apresentação sob a forma de pontos – relativo à base de dados DB1.



**Figura 16** – Dados com apresentação sob a forma de pontos – relativo à base de dados DB2.

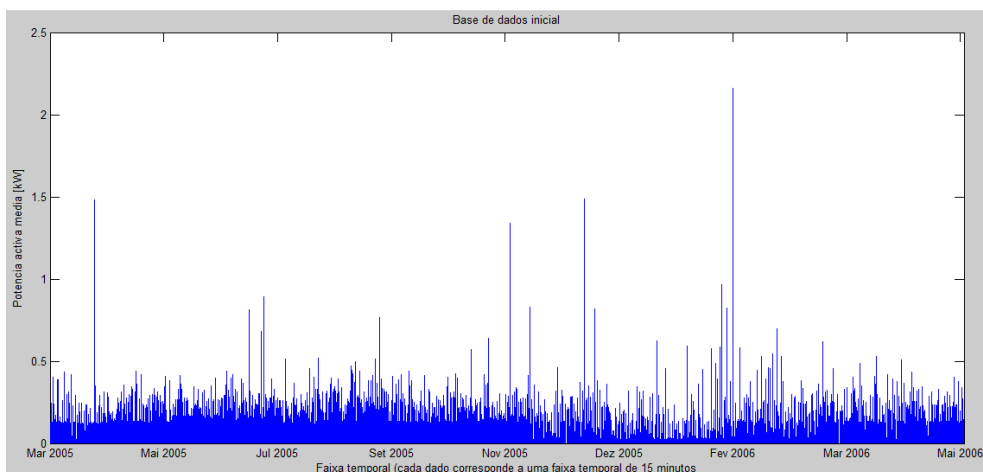
Nas Figuras 15 e 16 são apresentados dois consumidores diferentes, com valores de potência activa referentes aos períodos entre Outubro de 2003 a Dezembro de 2004 para o primeiro caso, e Março de 2005 a Maio de 2006 para o segundo caso. Para ambos os casos entre Maio e Novembro o valor de potência activa raramente é inferior a 200 W, situação que poderá ser devida a uma superior utilização de energia nos meses mais quentes.

As figuras 17 e 18 expõem a base de dados de outro modo, permitindo uma visualização da variação de carga com maior percepção ao longo do tempo. A desvantagem da utilização desta representação gráfica relativamente à anterior, é que não permite a visualização imediata dos valores praticamente nulos, nos meses normalmente com temperaturas mais baixas.



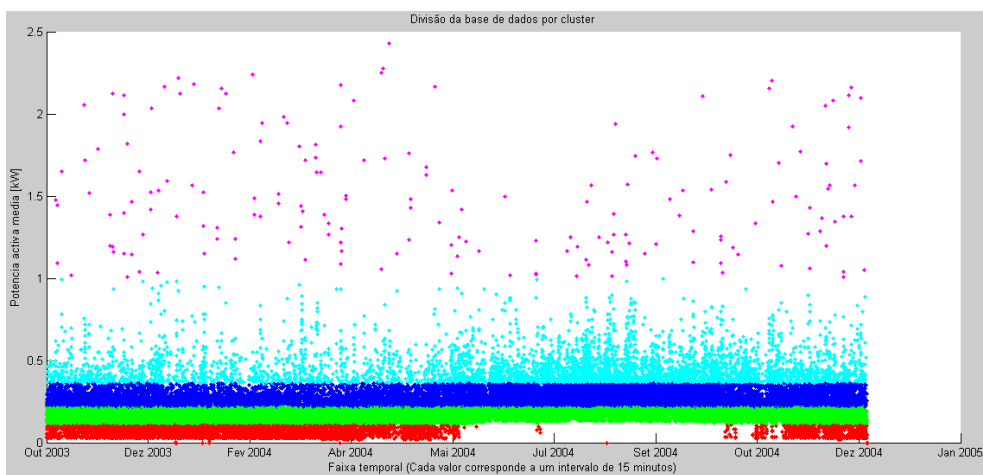
**Figura 17** – Dados com apresentação sob a forma de área – relativo à base de dados DB1.



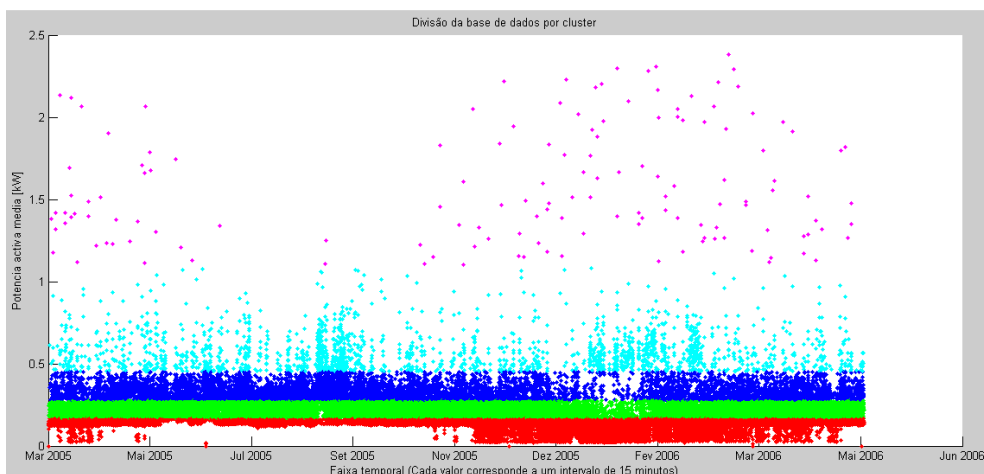


**Figura 18** – Dados com apresentação sob a forma de área – relativo à base de dados DB2.

No decorrer da execução da aplicação, os valores das bases de dados são categorizados por classes. Os gráficos que inicialmente foram apresentados e que englobam todos os valores das bases de dados (Figuras 15 e 16), após um processo de *clustering* resultam em gráficos apresentados com cores diferentes, consoante a sua atribuição por classes segundo o método *Fuzzy C-Means*. Em ambos os exemplos, na separação por classes foi considerado que existiria uma separação por cinco classes (Figuras 19 e 20).

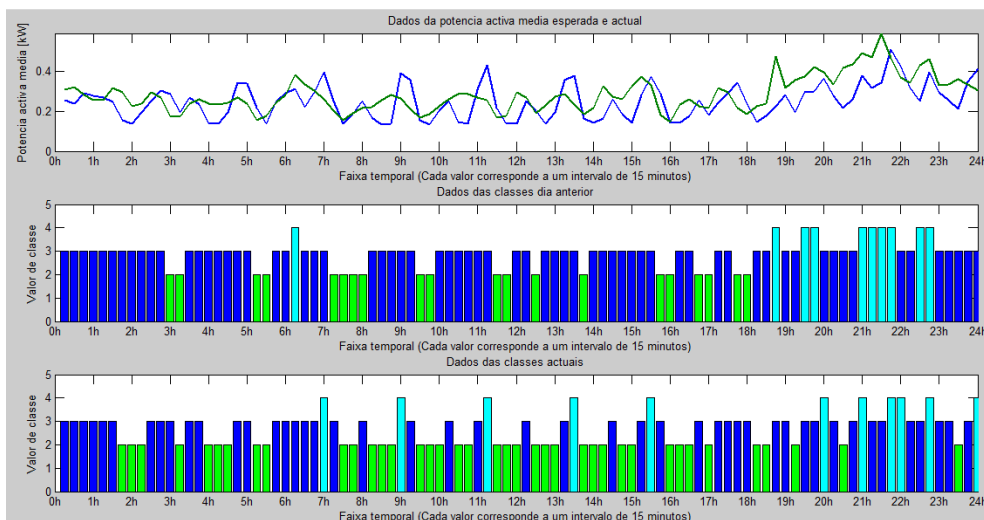


**Figura 19** – Dados com apresentação sob a forma de pontos separados por *cluster* segundo o método *Fuzzy C-Means* – DB1.

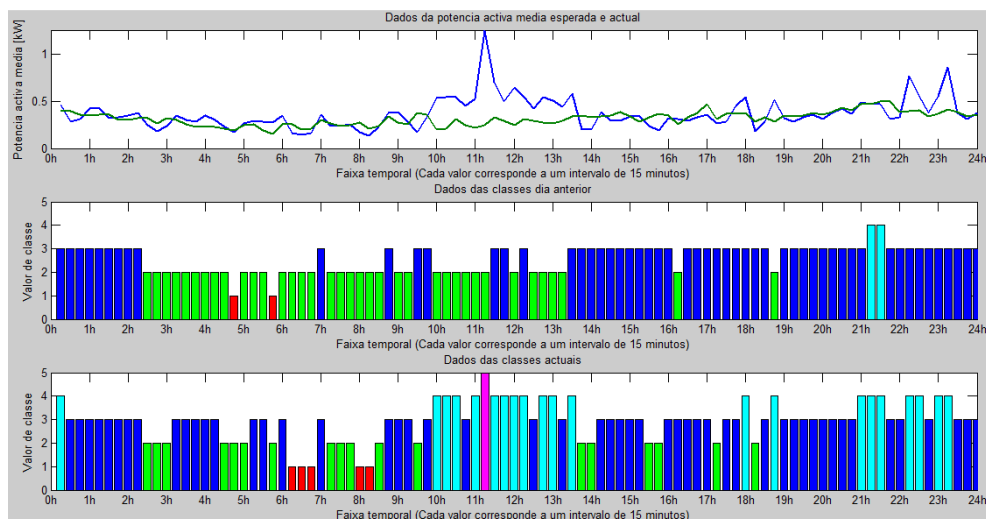


**Figura 20** – Dados com apresentação sob a forma de pontos separados por *cluster* segundo o método *Fuzzy C-Means* – DB2.

Na classe um encontram-se os valores de potência activa mais baixos. A classe três é referente aos valores de potência activa considerados como comuns e a classe dois como valores ligeiramente abaixo do comum. As classes quatro e cinco separam valores de potência activa altos em dois níveis, sendo que na classe cinco alguns dos resultados podem até ter sido causados por valores anormalmente elevados ou *outliers*.



**Figura 21** – Gráficos de potência activa e da separação por classes com base nos valores actuais e estipulados – DB1.

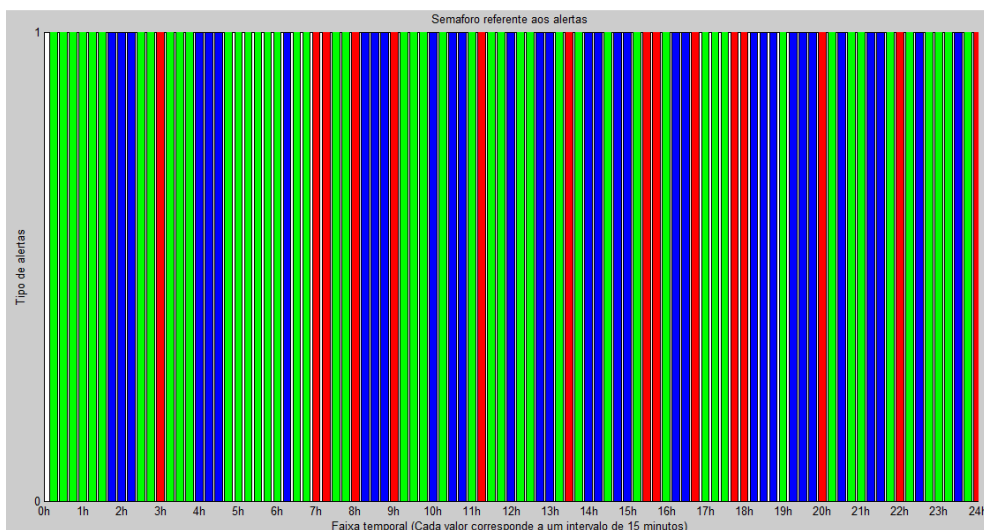


**Figura 22** – Gráficos de potência activa e da separação por classes com base nos valores actuais e estipulados – DB2.

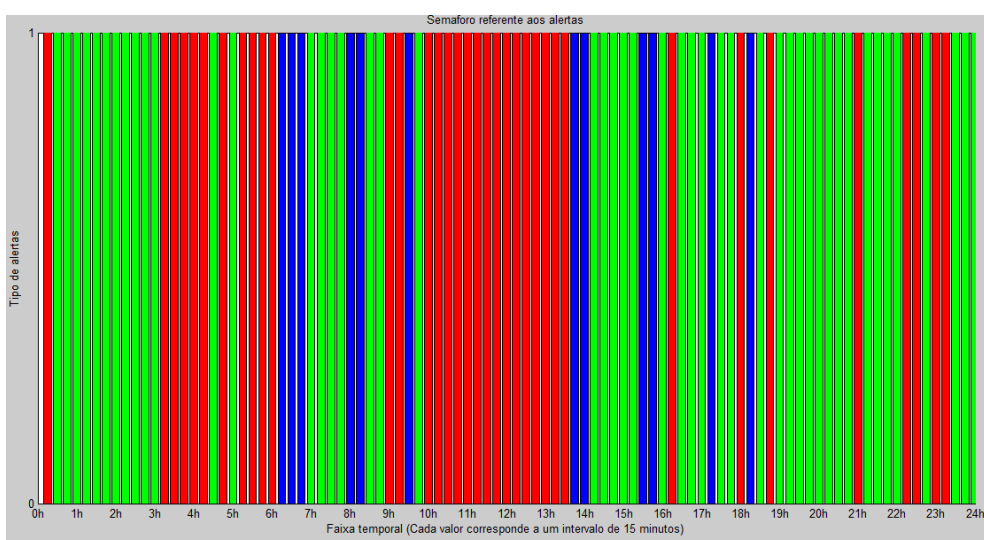
No exemplo ilustrado, que corresponde a simulação com base em dados históricos, o gráfico apresentado como actual é referente ao dia que foi considerado para teste. A curva no primeiro gráfico apresentado a azul nas Figuras 21 e 22 é referente aos valores médios da potência activa ao longo do dia. A segunda curva apresentada a verde é referente ao diagrama de referência que será comparado com o dia em análise. O segundo e terceiro gráfico mostram a variação das classes que foram atribuídas ao longo dos períodos temporais considerados.

No caso da Figura 21 os dois diagramas são muito próximos, apesar de existirem pequenas variações nas classes atribuídas, não existe nenhuma variação brusca de resultados. Para a situação da Figura 22 já é diferente, existindo variações com significado entre as 10h e a 13h30 e entre as 22h e as 24h. Estas diferenças poderão ser facilmente explicáveis com base em variações na actividade dos espaços monitorizados, ou poderão corresponder a anomalias, algo que apenas o utilizador/gestor desses espaços poderá identificar quando confrontado com a informação.

A determinação do diagrama de referência ocorre após a definição do que se considera como resultados temporais semelhantes para comparação com o dia e período actual. Assim, procede-se à obtenção de um diagrama médio com valores para 96 períodos num dia (cada período corresponde a um intervalo de quinze minutos). A comparação das classes determinadas para o consumo do dia em análise com as classes atribuídas ao diagrama de referência permitirá emitir alertas, assinalando que poderá estar a decorrer uma anomalia do sistema, nomeadamente se a classificação for muito diferente.

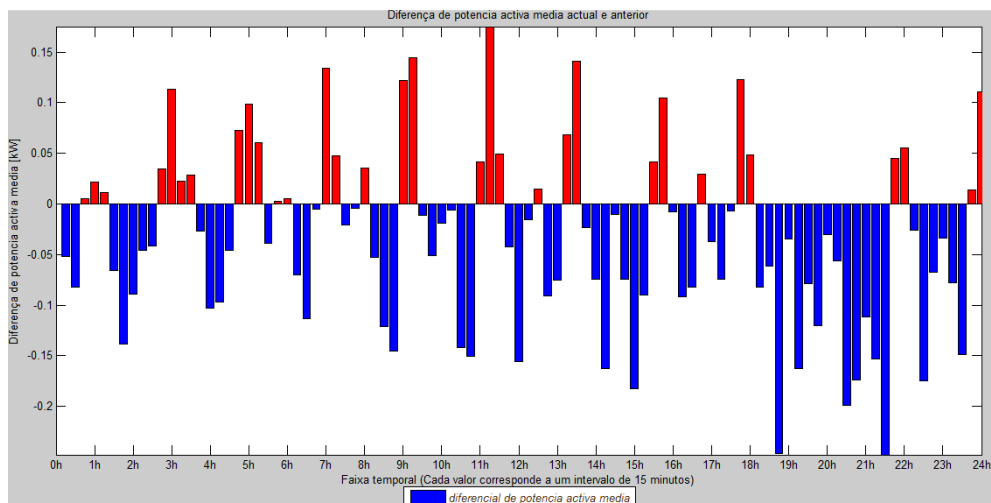


**Figura 23** – Semáforo de alertas – DB1.

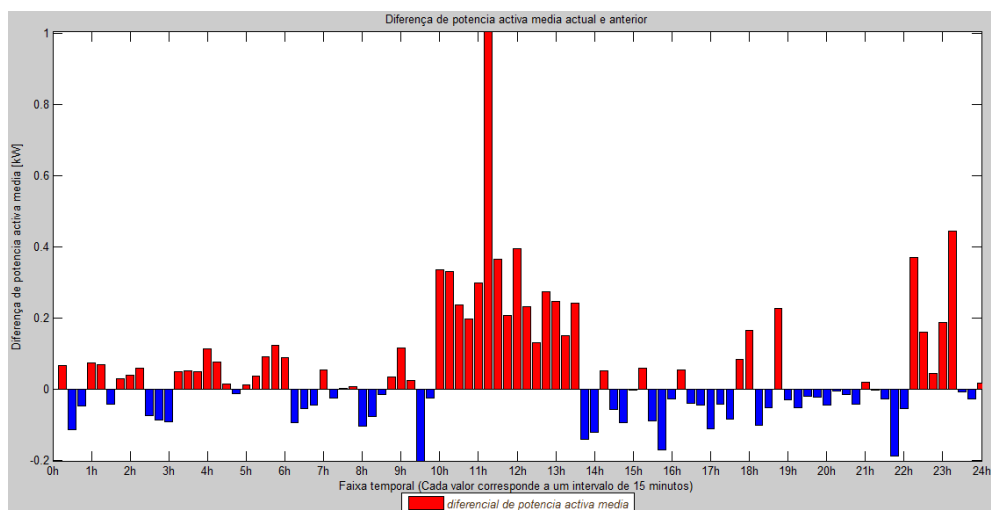


**Figura 24** – Semáforo de alertas – DB2.

O semáforo de alertas foi criado com o intuito de registar e apresentar graficamente os vários níveis de alertas que vão surgindo ao longo do dia. Nas Figuras 23 e 24 é possível identificar facilmente as variações que foram existindo ao longo do dia. Na base de dados 1 não se verificam diferenças significativas de classes ao longo do dia, apesar da existência de alguns alertas vermelhos (classe superior ao normal) e alertas azuis (classe inferior ao normal). Na base de dados 2 é visível a existência de consumos significativamente superiores ao normal, resultando num alerta de potência activa com nível excessivo nos períodos temporais já enunciados anteriormente (10h-13h30 e 22h-24h).



**Figura 25** – Diferencial de potência activa – DB1.



**Figura 26** – Diferencial de potência activa – DB2.

Como complemento é apresentado um gráfico de diferencial de potência activa que ajuda a ilustrar a diferença dos valores de potência activa nos dois casos, ajudando igualmente a justificar os vários alertas que foram sendo implementados ao longo do dia e a dar uma avaliação quantitativa dos desvios verificados face ao padrão esperado. No primeiro caso Figura 25, as classificações não sofrem alterações porque as variações de potência activa são de valor reduzido, no entanto, visualiza-se uma diferença de potência activa expressiva que justifica alertas emitidos, sendo visível na Figura 27 um excerto desses alertas.

Hora [H]	Minutos [M]	Diferencial Pot. Act. [kW]	Pot. Act. dia actual [kW]	Pot. Act. esperada [kW]	Classes Actuais	Classes Anteriores	Tipo de alerta
20	0	-0.0560	0.2760	0.3320	3	3	2
20	15	-0.1992	0.2160	0.4152	2	3	1
20	30	-0.1736	0.2600	0.4336	3	3	2
20	45	-0.1120	0.3760	0.4880	4	4	2
21	0	-0.1528	0.3160	0.4688	3	4	1
21	15	-0.2480	0.3400	0.5880	3	4	1
21	30	0.0448	0.5080	0.4632	4	4	2
21	45	0.0552	0.4240	0.3688	4	3	3
22	0	-0.0256	0.3160	0.3416	3	3	2
22	15	-0.1752	0.2520	0.4272	3	4	1
22	30	-0.0672	0.3920	0.4592	4	4	2
22	45	-0.0336	0.2960	0.3296	3	3	2
23	0	-0.0776	0.2560	0.3336	3	3	2

Alarmes		N.º de alerta actual	
Alertas	sem alertas	2	2
Historico	19-Aug-2014 22:30:00 - Potencia activa media normal	2	
	19-Aug-2014 22:30:00 - Potencia activa media normal	2	
	19-Aug-2014 22:30:00 - Potencia activa media normal	2	

Figura 27 – Tabela de classificações (Painel) – DB1.

O alerta apresentado como actual na Figura 27 é referente às 22h30 do dia 19/08/2004. O alerta indica consumo normal para o período especificado, sendo o valor de potência activa actual de 390 W para uma classificação de nível quatro. O valor a ser comparado estipula que para este período temporal o valor da potência activa deveria ser aproximadamente de 450 W para uma classe igualmente de nível quatro. Como se trata apenas de um valor residual de diferencial de potência activa, não é emitido nenhum alarme. Os valores de potência activa esperados correspondem ao diagrama de referência.

Hora [H]	Minutos [M]	Diferencial Pot. Act. [kW]	Pot. Act. dia actual [kW]	Pot. Act. esperada [kW]	Classes Actuais	Classes Anteriores	Tipo de alerta
20	45	0.0200	0.4640	0.4640	4	3	3
21	0	-0.0024	0.4680	0.4704	4	4	2
21	15	-0.0256	0.4680	0.4936	4	4	2
21	30	-0.1872	0.3080	0.4952	3	3	2
21	45	-0.0536	0.3280	0.3816	3	3	2
22	0	0.3712	0.7680	0.3968	4	3	3
22	15	0.1608	0.5600	0.3992	4	3	3
22	30	0.0448	0.3800	0.3352	3	3	2
22	45	0.1888	0.5520	0.3632	4	3	3
23	0	0.4456	0.8600	0.4144	4	3	3
23	15	-0.0072	0.3760	0.3832	3	3	2
23	30	-0.0272	0.3080	0.3352	3	3	2
23	45	0.0184	0.3760	0.3576	3	3	2

Alarmes		N.º de alerta actual	
Alertas	sem alertas	2	2
Historico	19-Aug-2014 22:30:00 - Potencia activa media normal	2	

Figura 28 – Tabela de classificações (Painel) – DB2.

Para a Figura 28 relativa ao caso DB2, o alerta de nível dois apresentado como actual é referente às 22h30 do dia 19/08/2005, sendo este referente igualmente a um valor de potência activa dentro dos parâmetros esperados. Apesar de se verificar que a partir das 22h deste dia, foram surgindo alertas de níveis superiores ao normal, para o registo horário das 22h30, em

que se encontra o alerta actual, os valores estão dentro do normal com uma classe de categoria três e um diferencial de potência activa de 44,8 W. Nos períodos horários imediatamente antes e depois, ou seja, 22h15 e 22h45 a tabela regista alertas de nível três, contendo em ambos os casos uma diferença de potência activa superior a 160,8 W e 188,8 W.

## 4.2. APLICAÇÃO DOS PERCENTIS

Esta solução pretende ser um método alternativo à utilização da classificação de resultados com o recurso ao método *Fuzzy C- Means*. Ao estabelecer-se um novo método, é possível comparar se uma abordagem mais simples poderá ser ou não uma solução com uma maior versatilidade que a anterior. Nesta abordagem, ao invés de utilizar as ferramentas de *clustering*, a atribuição de classes é realizada através da separação de valores por categorias, definindo cada categoria por um intervalo percentual relativamente ao valor médio de potência activa.

Nos testes realizados verificam-se diferenças na definição das classes, o que já seria de esperar, visto que se tratam de duas tipologias de classificação diferentes. Com a utilização da separação por *Percentis*, as classes médias e superiores divergem largamente das apresentadas pelo outro método. Ao ter uma abertura maior de classes podem incluir valores que já não se encontram tão próximos do valor médio do *cluster*, ou seja, podem incluir valores que já diferem bastante da média, o que dará uma informação errada relativamente ao pretendido.

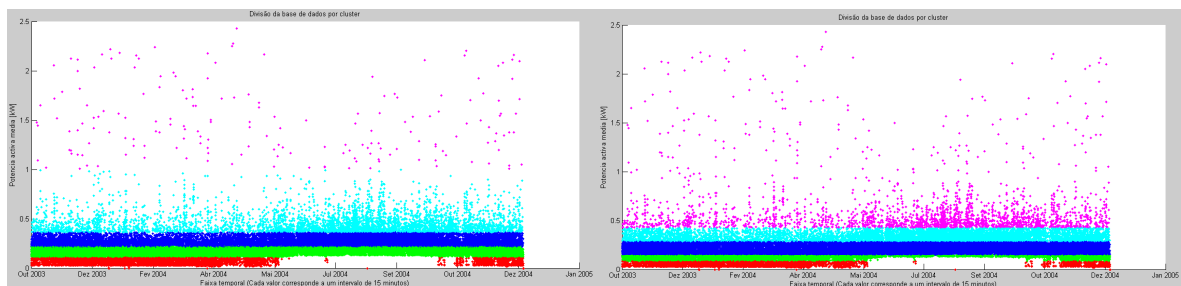


Figura 29 – Dados com apresentação sob a forma de pontos separados para o primeiro e segundo método – DB1.

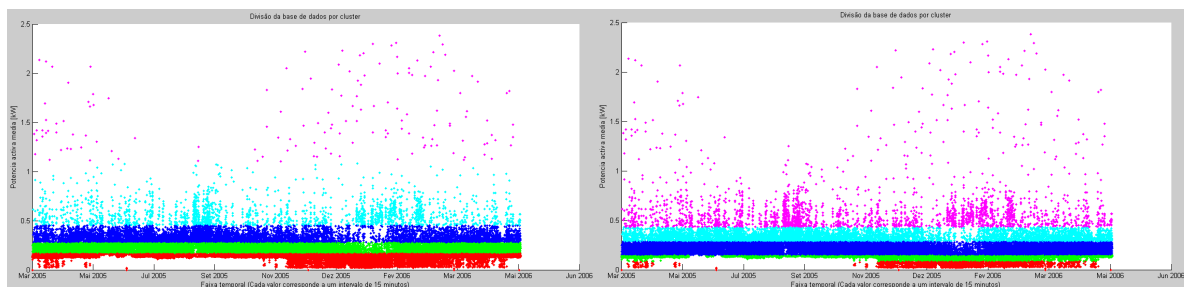


Figura 30 – Dados com apresentação sob a forma de pontos separados para o primeiro e segundo método – DB2.

Ao aplicar o método *Percentis*, as classes são ajustadas e reorganizadas sendo a classe superior mais abrangente do que nos casos que utilizaram o método *Fuzzy C-Means*, logo menos discriminativa (Figuras 29 e 30).

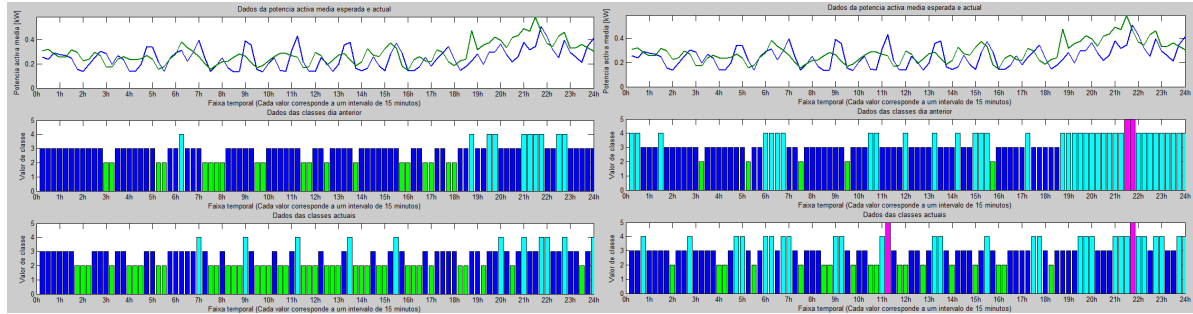


Figura 31 – Comparativo de classificação entre o primeiro e segundo método – DB1.

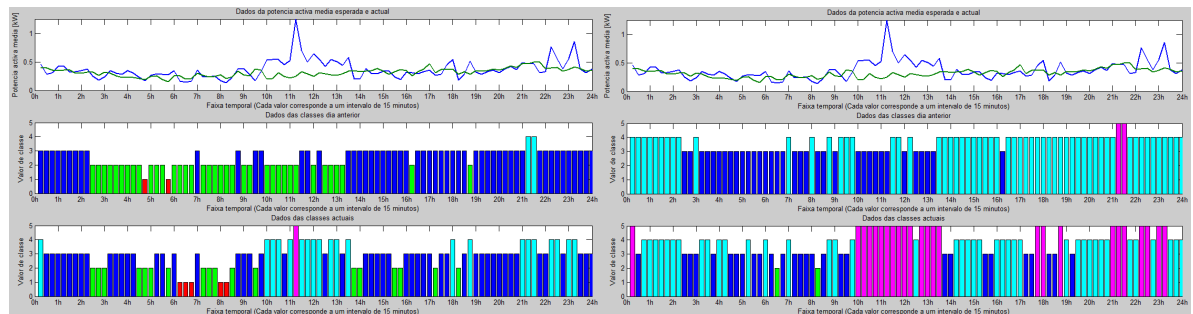


Figura 32 – Comparativo de classificação entre o primeiro e segundo método – DB2.

Comparando os métodos verifica-se a existência de uma elevada disparidade na classificação de alguns dos valores de potência activa (Figuras 31 e 32), tendendo o método de *Percentis* a atribuir classes superiores. Essa situação poderá ou não influenciar a emissão de níveis de alertas, como se pode ver nas Figuras 33 e 34, em que apesar de tudo não existiram diferenças significativas. Contudo, o método *Fuzzy C-Means* permite aparentemente detectar melhor diferenças com significado quando os consumos são mais baixos (caso DB2 entre as 3h00 e as 6h00).

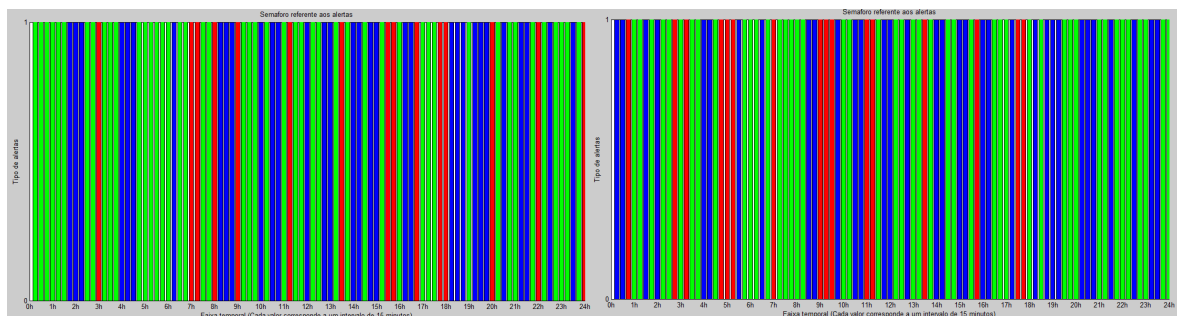
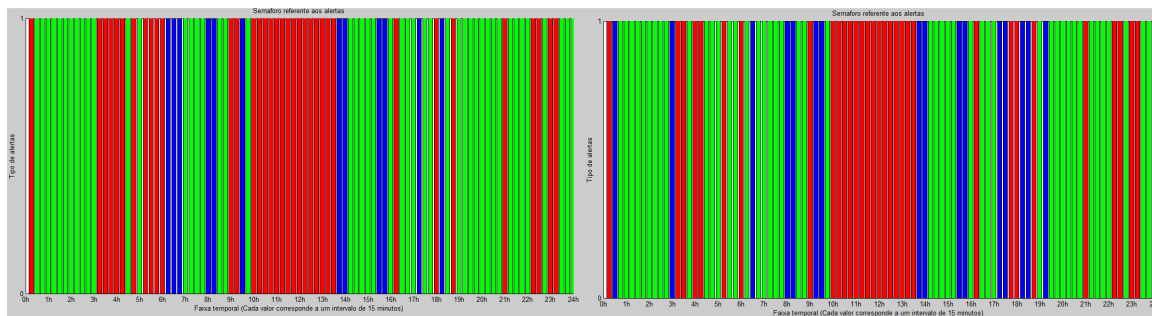


Figura 33 – Comparativo de alertas entre o primeiro e segundo método – DB1.





**Figura 34** – Comparativo de alertas entre o primeiro e segundo método – DB2.

### **4.3. APLICAÇÃO DO *FUZZY C-MEANS* COM REFERÊNCIA ESCOLHIDA PELO UTILIZADOR**

Na implementação deste método pretendeu-se deixar ao utilizador a opção sobre a formação da referência com a qual os consumos actuais serão comparados, permitindo escolher entre o dia anterior, da semana anterior, ou qualquer outro dia que se pretenda analisar. É dada ao utilizador a informação de quantos dias existentes na base de dados são possíveis para comparação. Todavia, este sistema só irá funcionar correctamente caso a base de dados disponha de valores superiores a um dia, caso contrário o sistema vai aguardar até esse horizonte temporal ser atingido.

Uma das funcionalidades existentes nestas aplicações é a possibilidade dada ao utilizador de comparação do valor de potência activa actual com outros valores nas últimas horas. Esta opção permite que o utilizador disponha de informações relativamente à evolução de carga nos períodos temporais mais próximos da hora actual.

#### **4.3.1. CASO DB1**

Os resultados são referentes ao primeiro consumidor usado para demonstrar resultados nas aplicações anteriores. Para o instante considerado como actual referido na Figura 35 a potência activa actual é de 392W e a classificação de nível quatro. Ajustando os valores de comparação para um dia anterior ao actual, é devolvido ao utilizador a informação de que o alerta actual é de nível três (valor superior ao normal), já que o valor de potência activa de referência de 284 W, classificado na classe três.

Alarmes		N.º de alerta actual	
Alertas		3	
Historico		3	
Introdução de dados		Arranque de novos dados	
Numero clusters	5	Dias comparação	1
Dif. de zero	0 - Sem alterações 1 - Novos dados	Dif. de zero	
Atualizar Dados		Arranque do programa	
Atualizar Tabelas		Parar programa	
Visualização de dados de consumo			
diferença pot. act. 1 hora atras	diferença pot. act. 2 hora atras	diferença pot. act. 3 hora atras	diferença pot. act. 4 hora atras
-0.0050 [kW]	0.1100 [kW]	0.1040 [kW]	0.1860 [kW]
diferença pot. act. entre 1 e 2 horas atras	diferença pot. act. entre 2 e 3 horas atras	diferença pot. act. entre 3 e 4 horas atras	Numero clusters actuais
0.1150 [kW]	-0.0060 [kW]	0.0820 [kW]	5
Classe actual			Pot. Act. actual
4			0.3920 [kW]
Classe anterior			Pot. Act. anterior
3			0.2840 [kW]
Numero de dias para comparação de resultados			420 [dias]

Figura 35 – Resultado da comparação com dia anterior – DB1.

A vantagem do uso desta aplicação comparativamente com a primeira opção deste método é a possibilidade de análise de diferentes dias escolhidos como referência. Assim, verifica-se na Figura 36 que a curva a azul nos dois gráficos é a mesma, correspondendo aos consumos do dia em análise, e que são identicamente classificados. A diferença estabelece-se apenas pela diferente definição da referência que originará diferentes classes para comparação, alterando assim a identificação dos alertas. Deste modo é possível ver que a calibração da referência pode ser um processo importante para que os alertas automáticos possam ter significado.

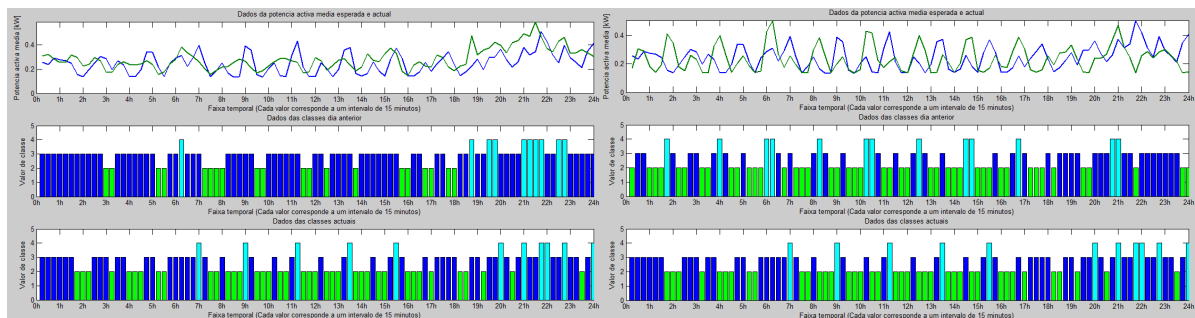
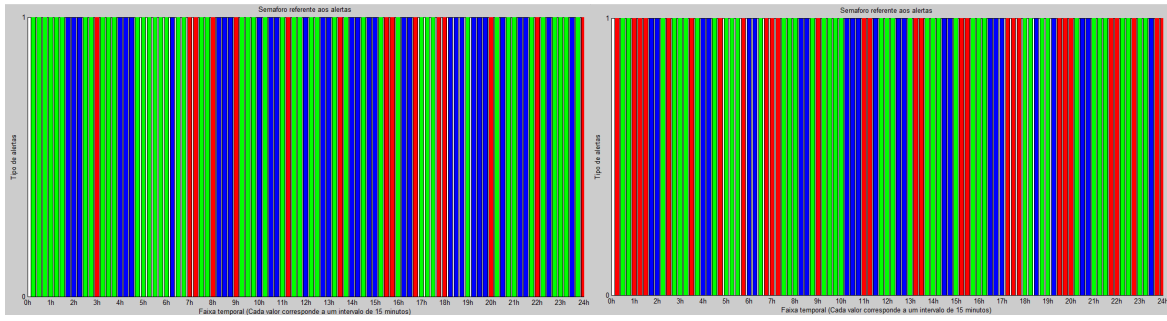


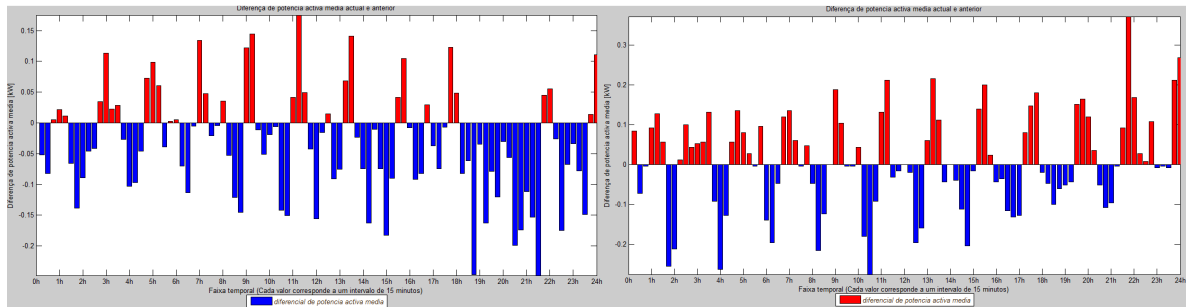
Figura 36 – Comparação de resultados de potência activa e classificação entre a primeira e segunda abordagem do método Fuzzy C-Means para um atraso de um dia – DB1.

Nesta segunda abordagem são apresentados para este dia um número superior de valores normais e superiores ao normal do que o verificado na primeira abordagem. Os resultados inferiores ao normal foram bastante inferiores (Figura 37).



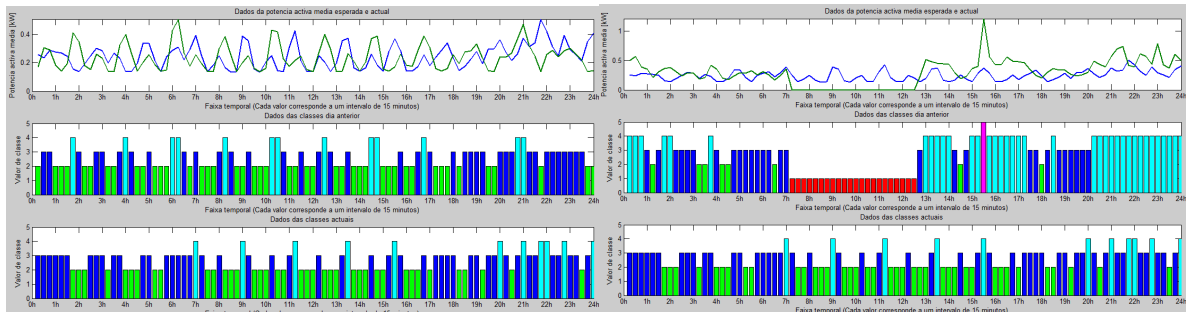
**Figura 37** – Comparação de resultados do semáforo entre a primeira e segunda abordagem do método *Fuzzy C-Means* para um atraso de um dia – DB1.

Na variação de potência activa verifica-se que existe uma variação até duas vezes superior ao primeiro caso. Com a primeira aplicação o máximo de diferença que existe entre dois valores consecutivos de potência activa é de 150 W, enquanto com esta abordagem a diferença passa para o dobro, ou seja 300 W (Figura 38).

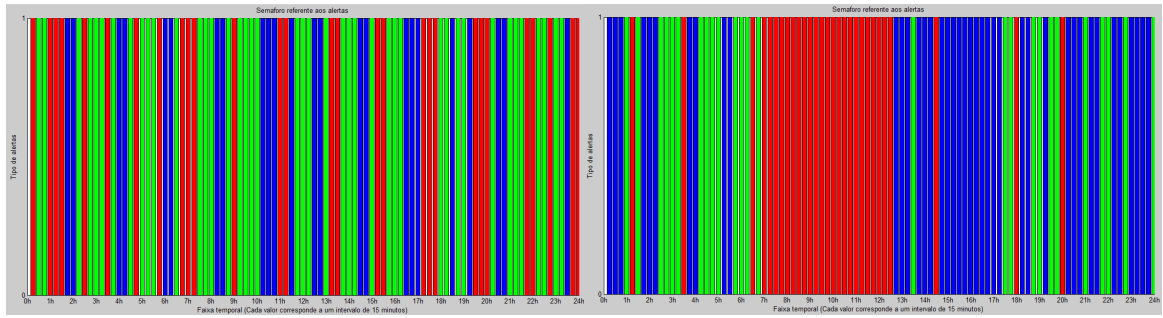


**Figura 38** – Comparação de diferencial de potência activa entre a primeira e segunda abordagem do método *Fuzzy C-Means* para um atraso de um dia – DB1.

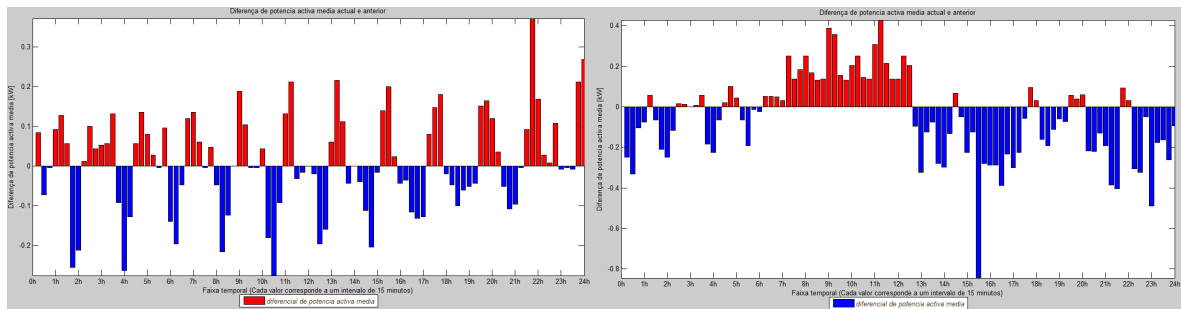
Usando como referência o dia correspondente da semana anterior provoca um maior número de alertas, demonstrando assim eventualmente uma menor correspondência entre as duas situações, o que é particularmente visível nas figuras 39, 40, 41.



**Figura 39** – Comparação de resultados de potência activa e classificação da segunda abordagem do método *Fuzzy C-Means* para um atraso de um e sete dias – DB1.



**Figura 40** – Comparação do semáforo da segunda abordagem do método *Fuzzy C-Means* para um atraso de um e sete dias – DB1.



**Figura 41** – Comparação do diferencial de potência activa da segunda abordagem do método *Fuzzy C-Means* para um atraso de um e sete dias – DB1.

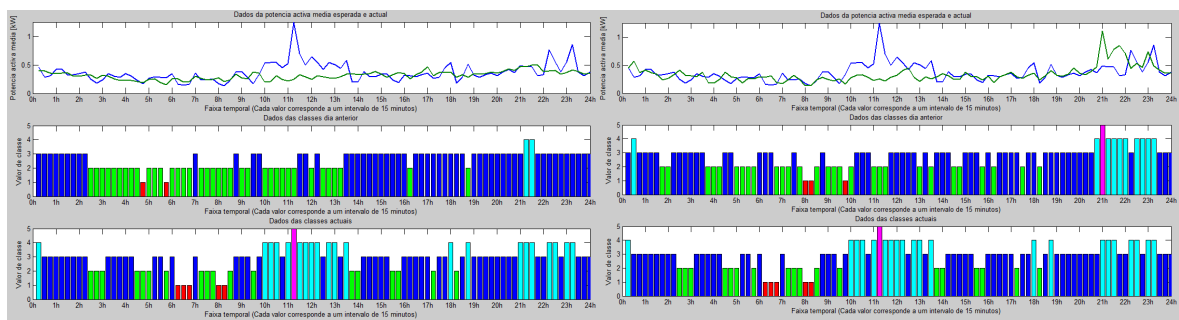
### 4.3.2. CASO DB2

No estudo com a DB2 obtêm-se situações similares às já referidas no estudo com a DB1. Para o atraso de um dia visualiza-se um alerta de nível um, apesar de os valores de potência activa não dispõem de uma diferença de valores elevada, foram classificados em *clusters* diferentes originando o alarme referido (Figura 42).

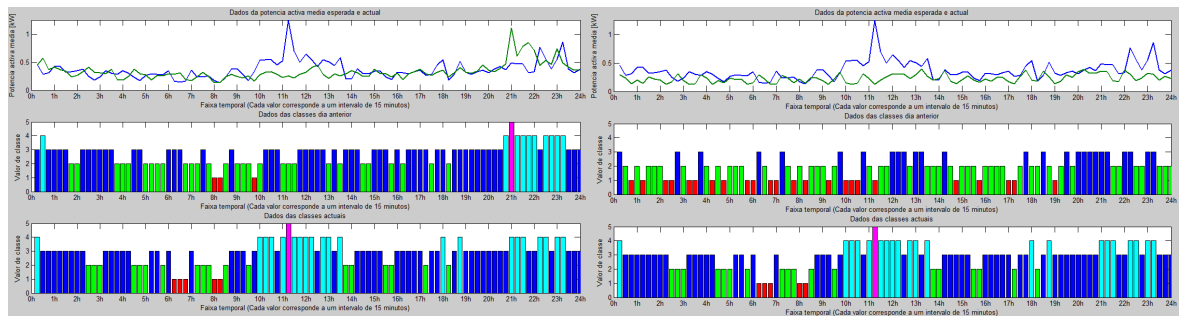
Alarmes		N.º de alerta actual		
Alertas		1		
19-Aug-2014 22:30:00 - Possível Anomalia potencia activa media inferior ao normal		1		
19-Aug-2014 22:30:00 - Possível Anomalia potencia activa media inferior ao normal		1		
Historico		0 - Sem dados		
19-Aug-2014 22:30:00 - Possível Anomalia potencia activa media inferior ao normal		1 - Consumo Inferior (AZ)		
19-Aug-2014 22:30:00 - Possível Anomalia potencia activa media inferior ao normal		2 - Consumo Normal (VD)		
		3 - Consumo Superior (VR)		
		4 - Outlier (AM)		
Introdução de dados				
Numero clusters	Arranque de novos dados	Dias comparação		
5	1	1	Actualizar Dados	
Dif. de zero	0 - Sem alterações	Dif. de zero	Actualizar Tabelas	
	1 - Novos dados		Arranque do programa	
			Parar programa	
Visualização de dados de consumo				
diferença pot. act. 1 hora atras	diferença pot. act. 2 hora atras	diferença pot. act. 3 hora atras	diferença pot. act. 4 hora atras	Numero clusters actuais
-0.0130 [kW]	-0.0320 [kW]	0.0600 [kW]	0.0570 [kW]	5
diferença pot. act. entre 1 e 2 horas atras	diferença pot. act. entre 2 e 3 horas atras	diferença pot. act. entre 3 e 4 horas atras	Classe actual	Pot. Act. actual
-0.0190 [kW]	0.0920 [kW]	-0.0030 [kW]	3	0.3800 [kW]
			Classe anterior	Pot. Act. anterior
			4	0.4560 [kW]
Numero de dias para comparação de resultados				417 [dias]

**Figura 42** – Resultado da comparação com dia anterior – DB2.

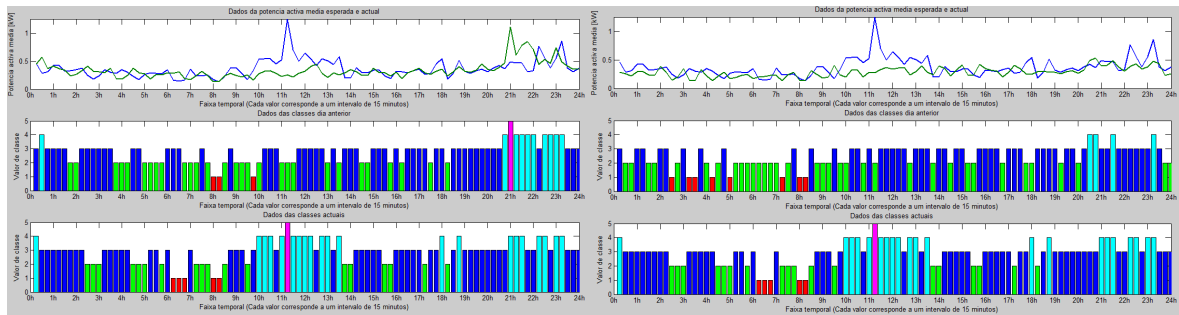
Comparativamente com a primeira e a segunda abordagem, o padrão que foi obtido para um dia semelhante é bastante idêntico, somente nos períodos pós laboral entre as 21h e as 24h é que foram surgindo pequenas alterações, sendo para o atraso de sete e catorze dias apresentados curvas similares, apesar dos valores de potência activa serem mais baixos no primeiro caso e superiores no segundo, o que originou a alteração das classificações nesses períodos. Em todos os casos para o período de almoço independente do atraso que lhes foi dado não surgiu numa curva similar ao apresentado para o dia considerado como actual, reafirmando assim que realmente poderá ter sido um dia atípico em termos do período entre as 10h e as 13h30 (Figuras 43, 44 e 45).



**Figura 43** – Comparação de resultados de potência activa e classificação entre a primeira e segunda abordagem do método *Fuzzy C-Means* para um atraso de um dia – DB2.



**Figura 44** – Comparação de resultados de potência activa e classificação da segunda abordagem do método *Fuzzy C-Means* para um atraso de um e sete dias – DB2.



**Figura 45** – Comparação de resultados de potência activa e classificação da segunda abordagem do método *Fuzzy C-Means* para um atraso de um e catorze dias – DB2.

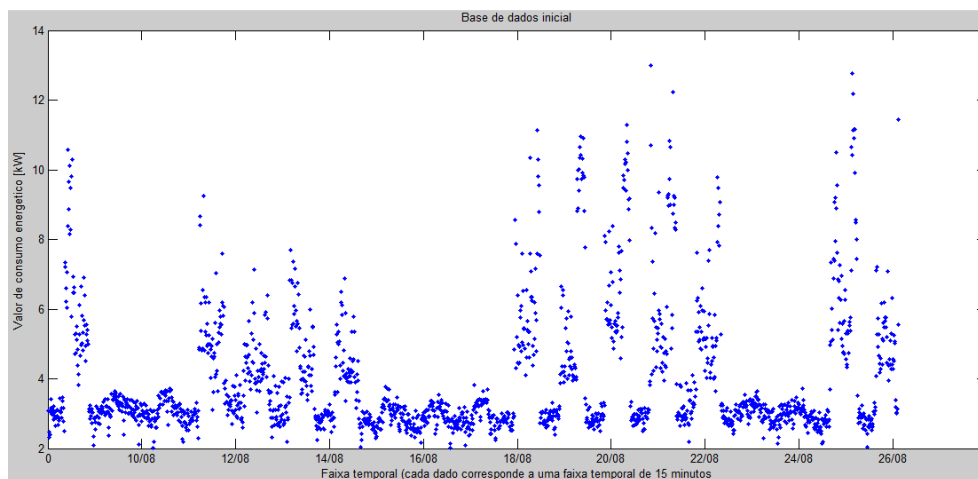
#### 4.4. APLICAÇÃO DO *FUZZY C-MEANS* EM TEMPO REAL

Os estudos de casos desenvolvidos anteriormente consistiram no uso de dados históricos, apenas sendo possível simular o que poderia ser o resultado de uma monitorização em tempo real, e não existindo qualquer informação que pudesse identificar anomalias que pudessem validar o processo. Neste caso o processo é aplicado a uma instalação desportiva que possui um sistema de gestão de consumos que pode ser consultado via web, permitindo por isso a monitorização em tempo real.

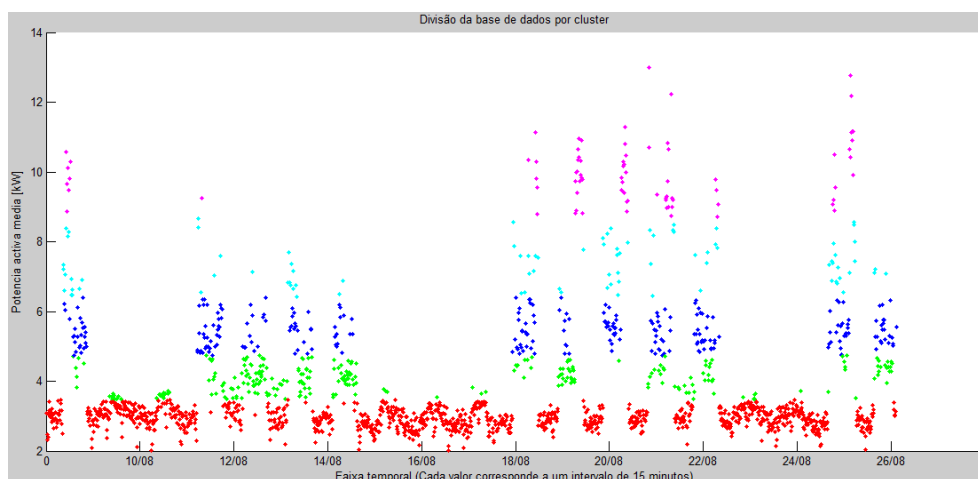
A consulta à base de dados do sistema de gestão de consumos recorre aos *softwares* Mysql e HeidiSql. Neste caso o Matlab vai ler sempre o último valor existente na base de dados e guarda-o numa base de dados própria, evitando provocar inadvertidamente alterações na base de dados original. De notar que nas aplicações com base em dados históricos, os gráficos são visualizados na sua totalidade, com todo o intervalo de tempo preenchido, enquanto na aquisição de dados em tempo real os gráficos vão sendo completados à medida que o dia vai avançando, verificando-se assim a evolução actual de cargas, gerando alertas em tempo real.

A aplicação desenvolvida implementou o método *Fuzzy C-Means*, mas deixando ao utilizador a escolha da referência, estratégia identificada como *Fuzzy C-Means* com referência escolhida pelo utilizador. Para os três diferentes períodos temporais que foram considerados para comparação de resultados (um, sete e catorze dias), os resultados obtidos são diferentes.

Uma vez que só foi possível aproveitar a oportunidade de aceder a um sistema de gestão de consumos em tempo real já no final do ano lectivo, os resultados são apenas referentes ao mês de Agosto, um mês de férias, resultando num índice menor de utilização do espaço desportivo em causa. Os dados apresentados são respeitantes ao dia 26/08/2014 para o período horário 20h45, sendo que o histórico de resultados dispõe de valores até dezassete dias antes da data actual (Figura 46). Na análise foi igualmente considerada uma separação de dados em cinco *clusters* (Figura 47).

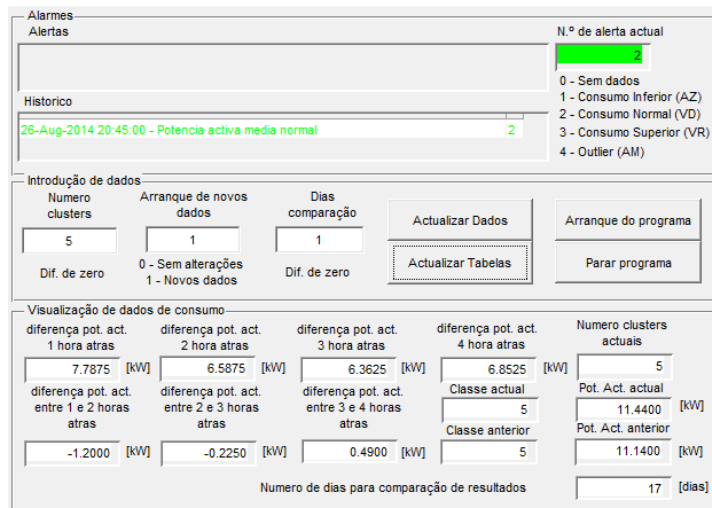


**Figura 46** – Dados com apresentação sob a forma de pontos – relativo a DB3.



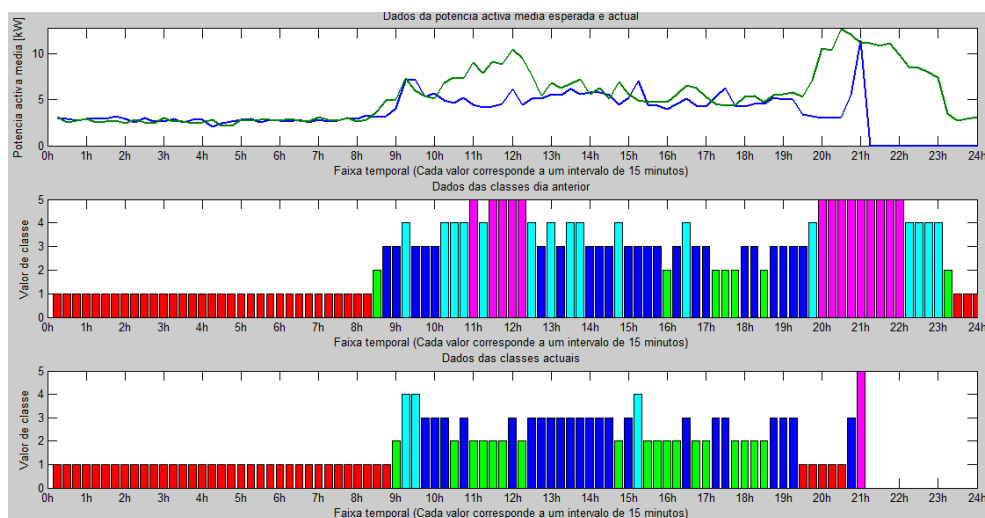
**Figura 47** – Dados com apresentação sob a forma de pontos separados por *cluster* para a aplicação *Fuzzy C-Means* em tempo real – DB3.

Verifica-se que a diferença na potência activa comparativamente às quatro horas anteriores é elevada (Figura 48). Todavia, tratando-se de um período temporal em que poderá existir um maior consumo devido, por exemplo, ao uso da iluminação artificial do espaço de jogo, essa diferença pode ser justificada. Já usando como referência o dia anterior, os valores monitorizados resultantes são bastante próximos dos valores de referência, gerando menos alarmes.



**Figura 48** – Resultado da comparação com um dia de atraso – *Fuzzy C-Means* em tempo real

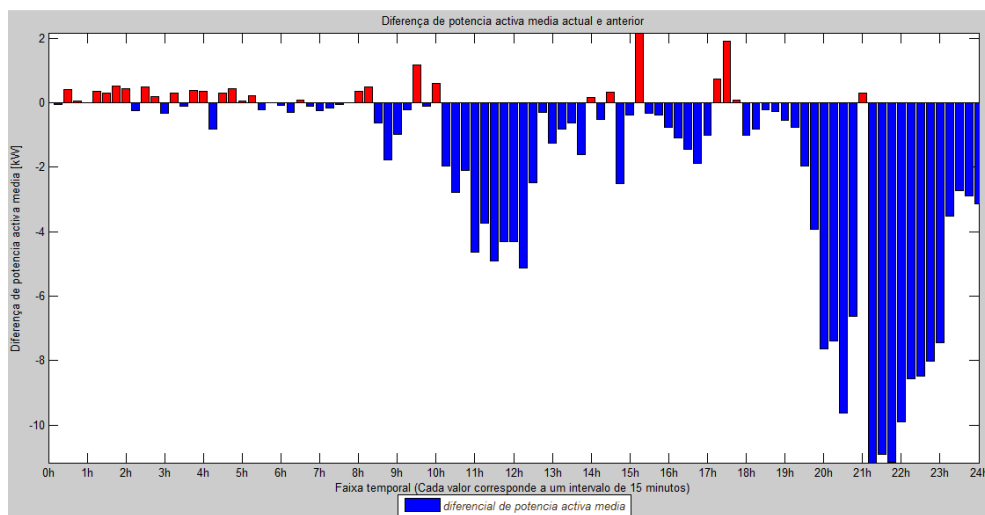
Ao comparar os diagramas de carga monitorizado e de referência na Figura 49 verifica-se alguma semelhança entre a meia-noite e as 9h da manhã do período em que o espaço desportivo deverá estar fechado ou com ocupação mais reduzida, assim como entre as 13h e as 19h divergindo apenas no final da manhã e no início da noite. Como não existem valores após as 20h45m, a curva parece decair abruptamente. No entanto, tal facto deve-se a não existirem ainda valores para os períodos temporais seguintes.



**Figura 49** – Resultado da comparação de potência activa e classificação com um dia de atraso – *Fuzzy C-Means* em tempo real

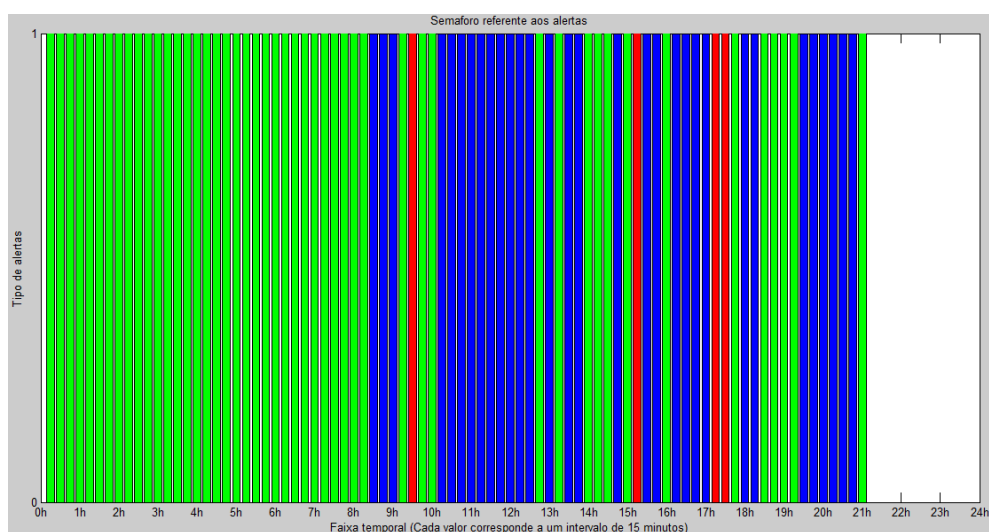
A comparação anteriormente descrita pode ser melhor visualizada na Figura 50. Contudo, deve de novo chamar-se a atenção de que o período temporal entre as 20h45 e as 24h00 não deve ser considerado por ausência de dados monitorizados.





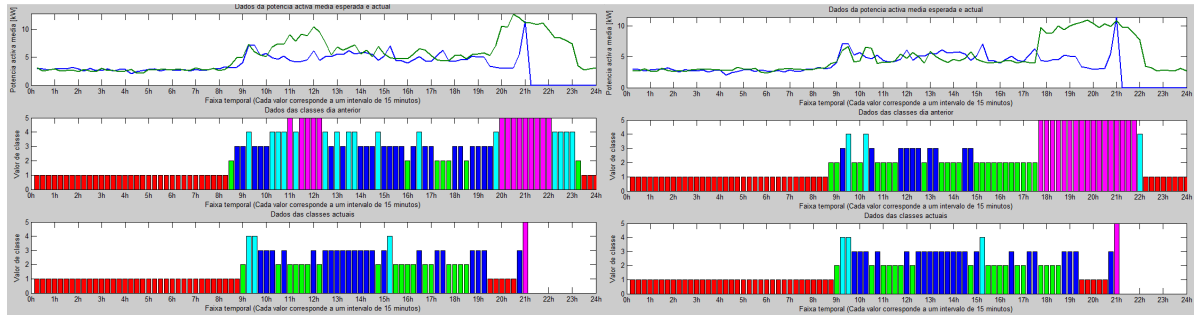
**Figura 50** – Diferencial de potência activa com um dia de atraso – *Fuzzy C-Means* em tempo real.

Em termos de alertas verifica-se que acompanha as variações da curva da potência activa e da evolução do diferencial de potência activa, verificando-se um período temporal praticamente constante da meia-noite as 9h da manhã (Figura 51).

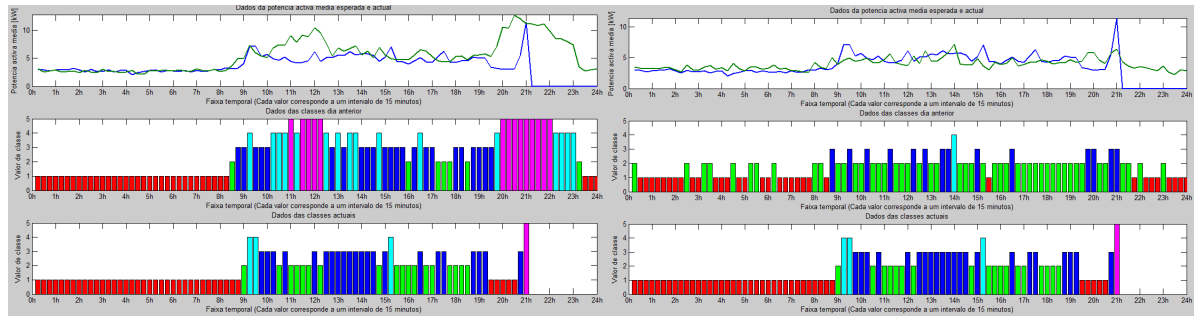


**Figura 51** – Semáforo de alertas com um dia de atraso – *Fuzzy C-Means* em tempo real

Ao verificar-se o resultado para os atrasos de sete e catorze dias, na primeira situação os valores de referência a partir das 17h30 são em larga escala diferentes dos valores apresentados para o dia actual (Figura 52). No entanto para uma referência a um atraso de catorze dias resultou num diagrama razoavelmente semelhante, mas isso não nos garante que noutra qualquer dia do ano o mesmo viesse a acontecer (Figura 53).

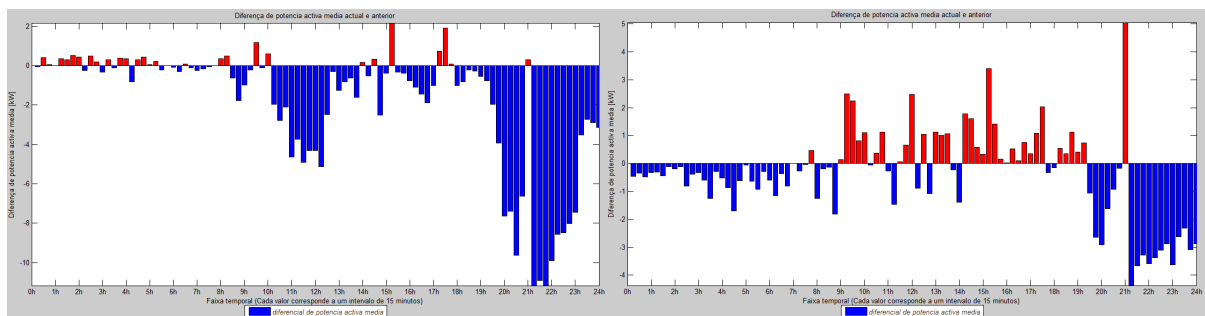


**Figura 52** – Comparação de resultados de potência activa e classificação do método *Fuzzy C-Means* em tempo real para um atraso de um e sete dias.

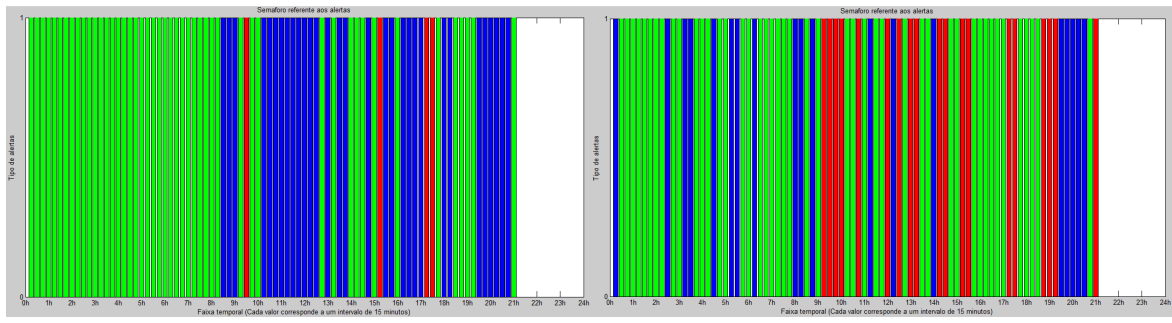


**Figura 53** – Comparação de resultados de potência activa e classificação do método *Fuzzy C-Means* em tempo real para um atraso de um e catorze dias.

Na Figura 54 observa-se uma variação positiva significativa, comparativamente entre o atraso de um e catorze dias. Por esse motivo existem alguns alertas referentes a valores excessivos, existindo uma predominância de valores normais, sendo reduzidos o número de alarmes (Figura 55).



**Figura 54** – Comparação do diferencial de potência activa do método *Fuzzy C-Means* em tempo real para um atraso de um e catorze dias.



**Figura 55** – Comparação do semáforo do método *Fuzzy C-Means* em tempo real para um atraso de um e catorze dias.



## 5. CONCLUSÕES

O trabalho realizado consistiu no desenvolvimento de uma aplicação que permita a análise das medições de consumos eléctricos numa determinada instalação com vista à detecção de eventuais anomalias correspondentes a consumos excessivos, insuficientes ou inexistentes face ao esperado. Com a análise e verificação do estado de funcionamento da instalação em causa, é possível minimizar as consequências de tais situações, evitando prejuízos e aumentando assim o valor acrescentado de soluções mais inteligentes de medida.

A proposta efectuada baseia-se na classificação dos consumos medidos de acordo com agrupamentos dos dados históricos efectuados com recurso a ferramentas de *clustering*. Essa classificação pode assim, ser usada directamente por um utilizador/gestor que usará a sua própria compreensão sobre o estado de funcionamento de equipamentos de uso final, para interpretar se corresponderá ou não ao expectável. Pode ainda ser comparada com a classificação dada a um consumo de referência, determinado pelo utilizador, para gerar automaticamente alertas que permitam a um gestor prestar a atenção necessária para eventualmente detectar a necessidade de intervir.

O trabalho foi implementado em Matlab, mas futuramente espera-se que a implementação das aplicações possa ser desenvolvida noutro tipo de plataformas permitindo a sua inclusão em sistemas de gestão de energia ou *smart-meters*. Desta forma será possível aumentar o valor acrescentado dessas soluções, tornando-as mais atractivas aos utilizadores e potenciando assim a sua generalização, com vantagens para a gestão do sistema.



## 6. BIBLIOGRAFIA

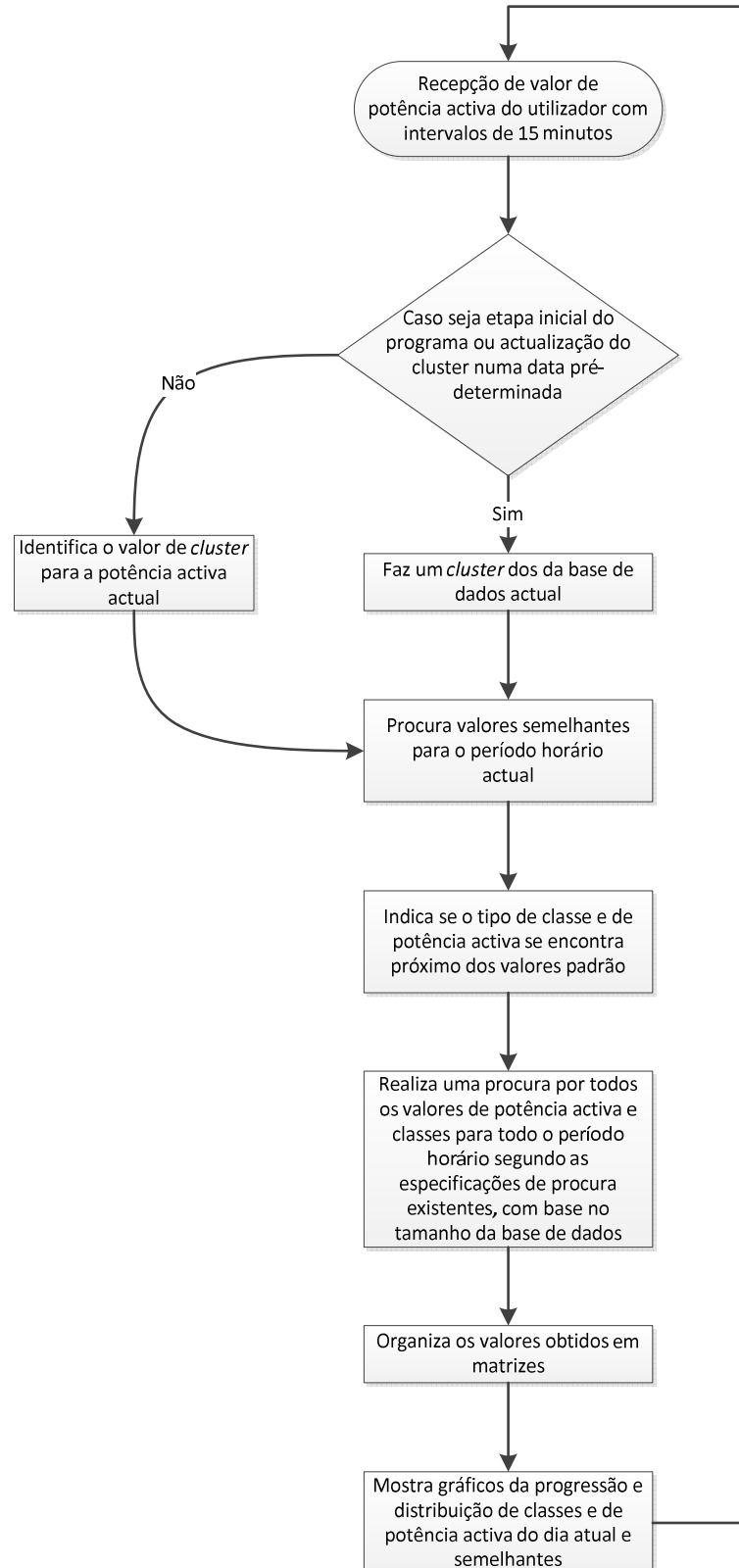
- [1] J. T. Almeida, "Utilização do algoritmo "Self-Organizing" na caracterização de organizações públicas portuguesas usando informação financeira," 2011.
- [2] A. L. Costa, C. Martin, J. P. J. Carvalho e M. K. Neto, "Agrupamento de padrões de curva de carga utilizando algoritmos e técnicas de agrupamento como alternativa tarifária.," 2010.
- [3] H.-J. L. Reimhold Decker, *Advances in Data Analysis*, Springer-Verlag Berlin Heidelberg, 2007.
- [4] M. C. A. C. A. G. E. Marta V. Modenesi, "Parallel Fuzzy c-Means Cluster Analysis," 2006.
- [5] R. B. A. M. N. S. Ahmad Shahi, "Detecting Effectiveness of Outliers and Noisy Data on Fuzzy System Using FCM," 2009.
- [6] K.-S. Chuang, H.-L. Tzeng, S. Chen, J. Wu e T.-J. Chen, "Fuzzy c-means clustering with spatial information for image segmentation," *computerized medical imaging and graphics*, vol. 30, n.º 1, pp. 9-15, 2006.
- [7] S. Haykin, *Redes neurais Princípios e prática*, 2ª Edição ed., Bookman, 2001, pp. 483-518.
- [8] J. B. I. d. C. Ana Rito, *Guia de Avaliação do estado Nutricional Infantil e Juvenil*, Lisboa: Instituto Nacional de Saúde Doutor Ricardo Jorge, 2010.
- [9] J. R. Taylor, *Introdução à Análise de Erros: O Estudo de Incertezas em Medições Físicas*, 2ª Edição ed., Bookman, 1984, pp. 145-153.
- [10] A. K. Jain, M. N. Murty e P. J. Flynn, "Data clustering: a review," *ACM computing surveys (CSUR)*, vol. 31, n.º 3, pp. 264-323, 1999.
- [11] D. F. Fonte, "Sistemas de detecção de eventos anómalos no padrão do consumo," 2012.
- [12] R. M. P. Bidarra e others, "Determinação de perfis de consumo baseada em mapas de Kohonen Modificados," 2013.
- [13] J. N. Fidalgo, M. A. Matos e L. Ribeiro, "A new clustering algorithm for load profiling based on billing data," *Electric Power Systems Research*, vol. 82, 2012.
- [14] N. L. C. Junior, "Clusterização baseada em algoritmos fuzzy," 2006.
- [15] F. Rodrigues, J. Duarte, V. Figueiredo, Z. Vale e M. Cordeiro, "A comparative analysis of clustering algorithms applied to load profiling," em *Machine Learning and Data Mining in Pattern Recognition*, Springer, 2003.
- [16] M. Sforza, "Data mining in a power company customer database," *Electric Power Systems Research*, vol. 55, n.º 3, 2000.

- [17] H. R. V. d. Rosário, “Excesso de peso e obesidade em crianças: implementação e avaliação de um programa de intervenção na escola,” 2011.
- [18] R. M. P. M. B. P. G. F. P. W. T. d. A. S. M. C. d. M. G. Daniel Perego, “Agrupamento fuzzy pelo algoritmo Gath-Geva na Shell Orion Data Mining Engine,” 2013.
- [19] R. C. D. Anil K. Jain, Algorithms for Clustering Data, Englewood Cliffs, New Jersey: Prentice Hall, 1988.
- [20] M. R. S. H. R. S. Eyke H ullermeier, “Comparing Fuzzy Partitions: A Generalization of the Rand Index and Related Measures,” 2012.
- [21] J. C. Bezdek, Fuzzy Models for Pattern Recognition: Methods That Search for Structures in Data, Sankar K. Pal, 1992.
- [22] S. K. Erkki Oja, Kohonen Maps, 1<sup>o</sup> ed., Elsevier Science B.V., 1999.

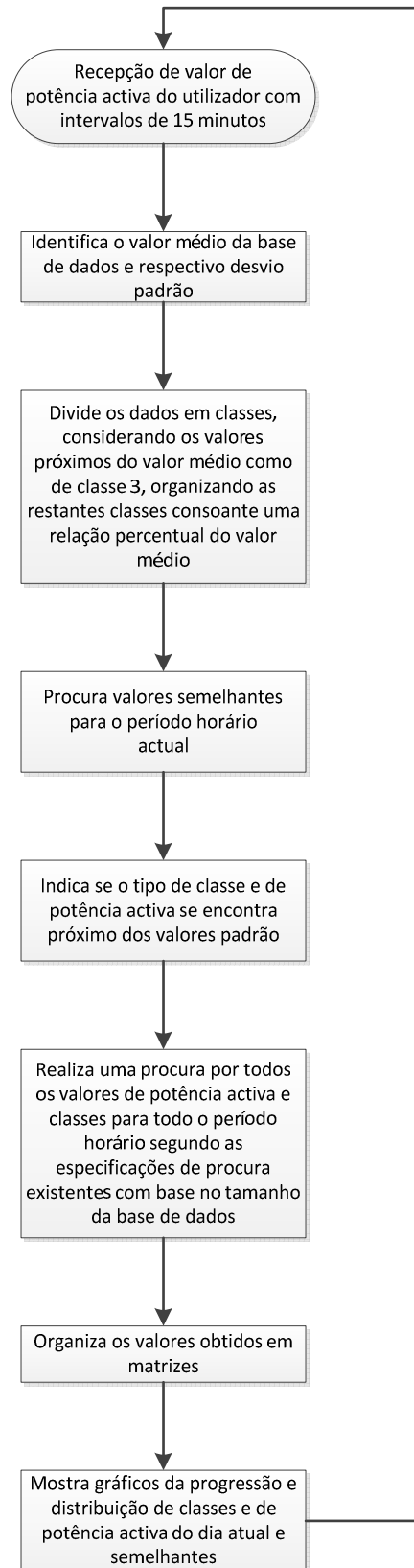


## 7. ANEXOS

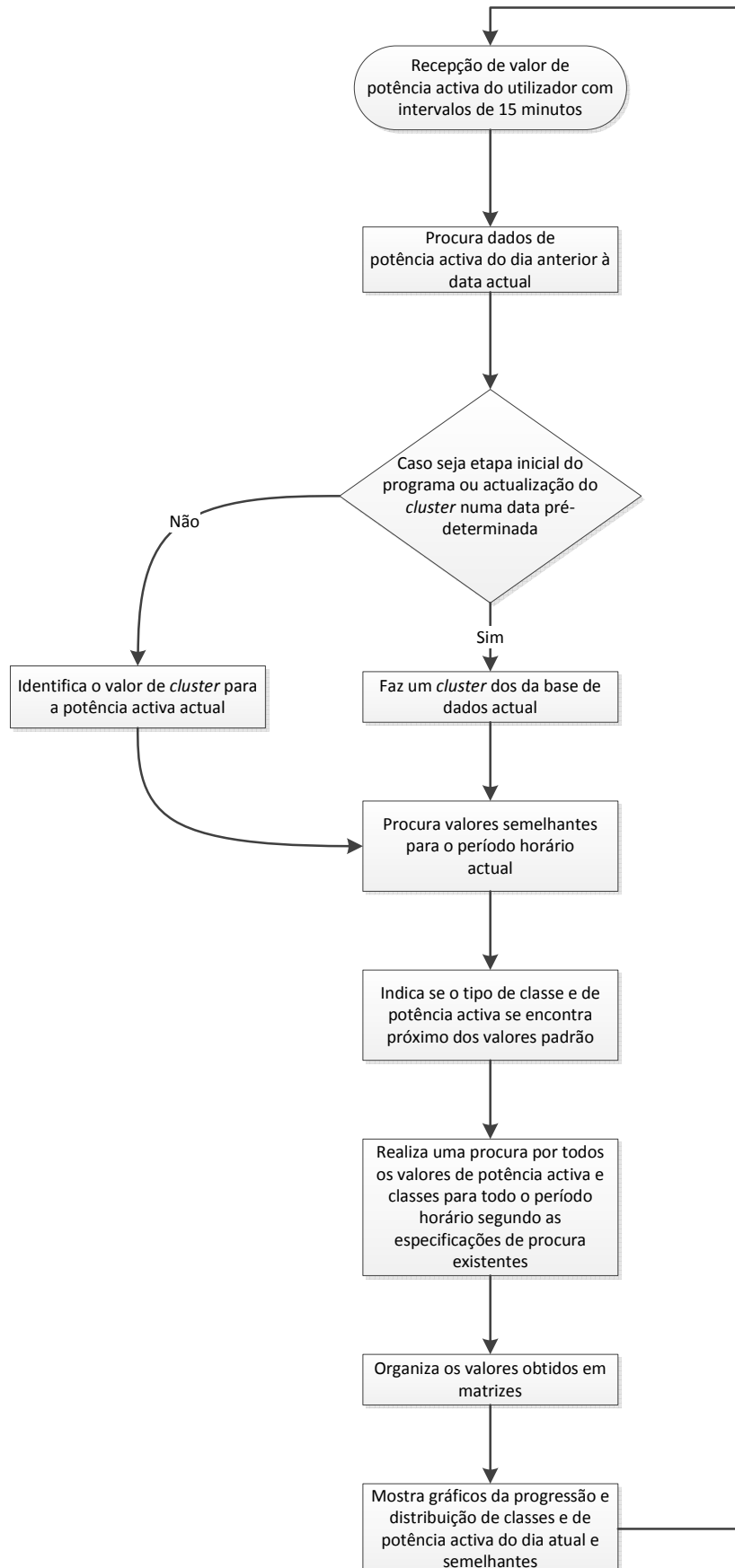
### Aplicação Tese Tempo Total



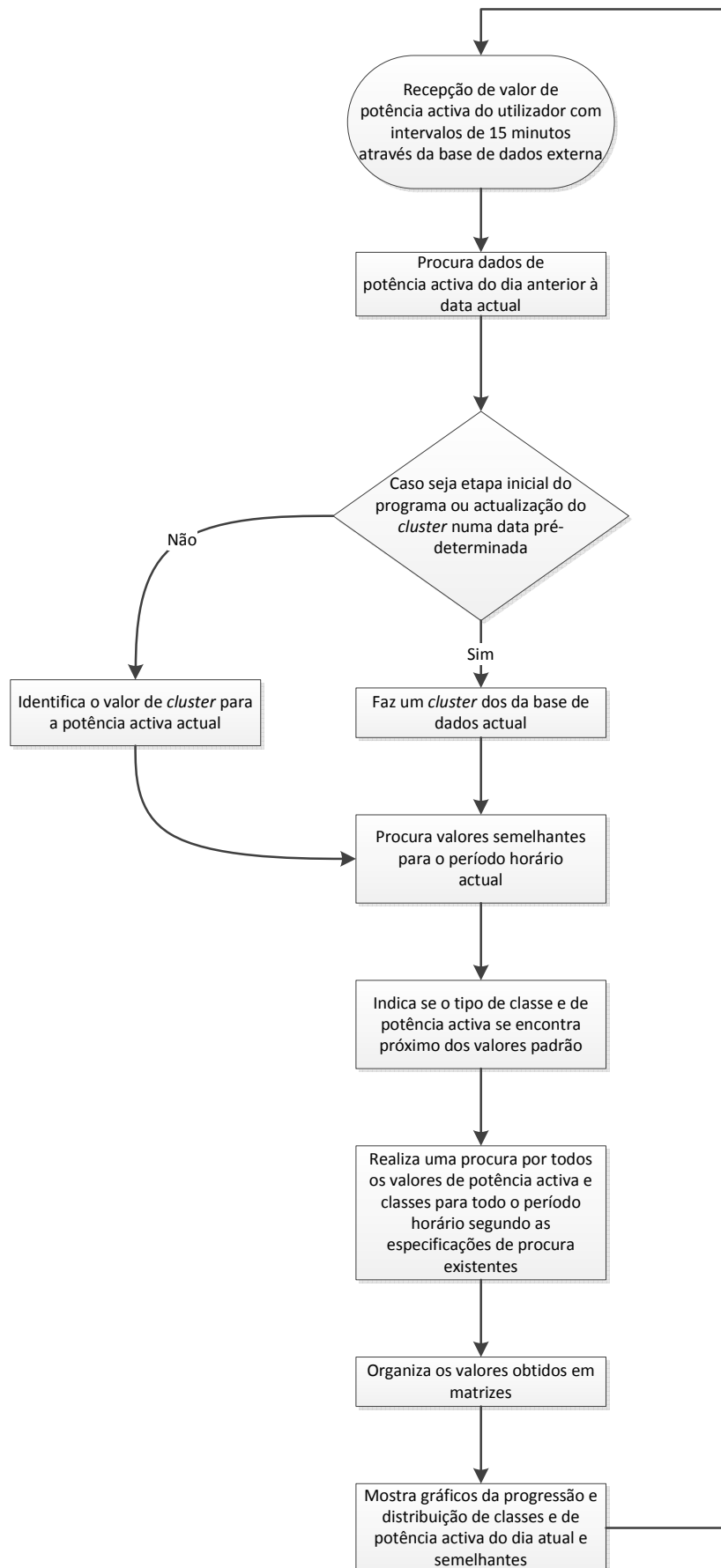
## Aplicação Tese Percentis



## Aplicação Tese 4H Simulação



## Aplicação Tese 4H Real



# MANUAL DO UTILIZADOR

## APLICAÇÕES DIAGNÓSTICOS DE CONSUMOS ANÓMALOS DE ENERGIA

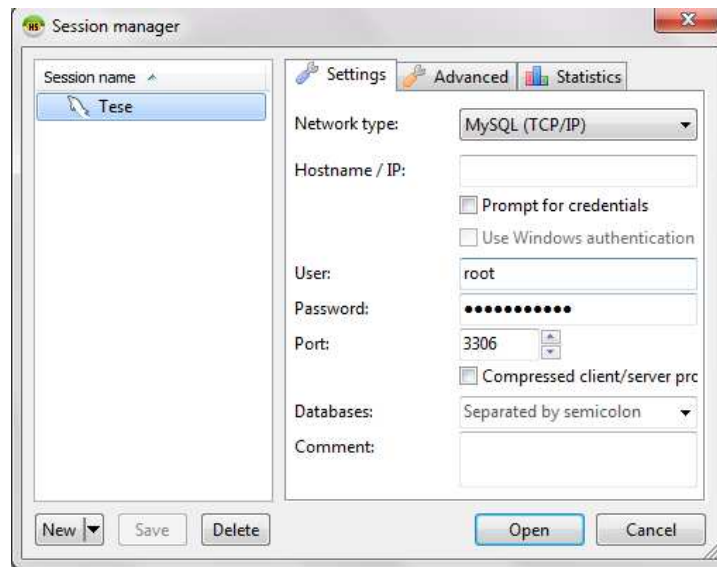


SETEMBRO DE 2014

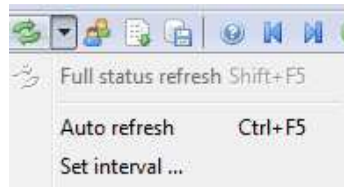
## Configurações dos *softwares* Matlab e HeidiSql

A parametrização da base de dados no programa HeidiSql só é necessária, caso a aplicação em análise for referente ao estudo de uma instalação em tempo real. Para as aplicações em que o sistema será executado em ambiente de simulação os dados devem ser inseridos manualmente.

Para a implementação da base de dados em tempo real usando o gestor de base de dados, deve-se antes de mais confirmar a ligação ao servidor, onde se encontra alojada a base de dados. Ao seleccionar uma nova sessão recorrendo à opção “*New*” (figura seguinte), deve-se estipular o tipo de comunicação por MySQL(TCP/IP), caso não se encontre instalado no computador o programa MySQL Workbench o gestor de base de dados não irá funcionar. No parâmetro de endereçamento *IP* ou *Hostname*, deverá ser colocado o endereço *web* ou endereço IP onde se encontra alojada a base de dados. Para executar o acesso à base de dados deverá ser igualmente introduzido a *password* e o *username* do utilizador, sendo a porta de comunicação dependente das configurações de acesso ao servidor.



Para o caso em análise é necessário só uma parte da base de dados, contudo para a correcta actualização da base de dados, deverá seleccionar-se o modo automático de actualização, recorrendo à opção *auto-refresh* do programa (figura seguinte). O tempo de intervalo de aquisição de dados deve ser ajustado para o intervalo mínimo, sendo este dependente do diferencial temporal de obtenção de um novo valor. Para o caso que se encontra em estudo foi definido um intervalo de 5 minutos (300 segundos).



No acesso às tabelas onde se encontram guardados os valores relevantes para análise da aplicação, estas devem ser expandidas, permitindo assim visualizar integralmente todos os elementos constituintes da base de dados. A base de dados ao ser expandida permite visualizar em tempo real a aquisição dos dados do servidor. Para o caso que se encontra parametrizado para o accionamento da aplicação em tempo real, a tabela de dados a ser utilizada é denominada **electricidade**, sendo os valores de potência activa acedidos na barra de tarefas com o nome **data** (figura seguinte).

Year	Month	Day	Hour	Minute	Second	Power (W)	Power (VA)	Power (kVA)	Power (kWh)	Power (kVAr)	Power (kVh)	Power (kVArh)	Power (kVArh)	Power (kVArh)	Power (kVArh)
2013	7	9	19	10	0.00	0.00	8.03	4.80	9.37	0.00	0.00	0.00	0.00	421.21	4
2013	7	9	19	20	0.00	0.00	7.87	4.88	9.30	0.00	0.00	0.00	0.00	422.02	4
2013	7	9	19	30	0.00	0.00	6.37	4.32	7.72	0.00	0.00	0.00	0.00	422.74	4
2013	7	9	19	40	0.00	0.00	4.83	3.62	6.06	0.00	0.00	0.00	0.00	420.63	4
2013	7	9	19	50	1.00	0.00	3.85	2.74	4.78	0.09	0.08	0.12	0.12	418.97	4
2013	7	9	20	0	1.00	0.00	4.03	2.90	5.00	1.58	1.35	2.10	2.10	419.53	4
2013	7	9	20	10	1.00	0.00	5.05	3.19	6.01	1.41	1.20	1.88	1.88	420.51	4
2013	7	9	20	20	1.00	0.00	4.99	3.66	6.21	1.51	1.11	2.05	2.05	420.82	4
2013	7	9	20	30	1.00	0.00	6.05	4.55	7.61	1.74	1.55	2.37	2.37	421.42	4
2013	7	9	20	40	1.00	0.00	4.09	3.18	5.22	2.06	1.71	2.74	2.74	421.98	4
2013	7	9	20	50	1.00	0.00	4.69	3.70	6.00	1.74	1.53	2.36	2.36	421.16	4
2013	7	9	21	0	1.00	0.00	5.06	4.16	6.61	1.68	1.48	2.38	2.38	420.37	4
2013	7	9	21	10	0.00	0.00	4.76	4.25	6.41	2.03	1.61	2.64	2.64	420.00	4
2013	7	9	21	20	1.00	0.00	4.58	3.51	5.83	1.95	1.74	2.62	2.62	418.89	4
2013	7	9	21	30	2.00	0.00	8.39	6.04	10.56	1.89	1.46	2.45	2.45	416.01	4
2013	7	9	21	40	1.00	0.00	9.67	5.51	11.14	4.91	3.24	6.18	6.18	414.74	4
2013	7	9	21	50	2.00	0.00	9.80	5.54	11.29	5.45	2.83	6.15	6.15	416.01	4
2013	7	9	22	0	1.00	0.00	8.69	4.81	9.94	5.54	2.77	6.22	6.22	416.27	4
2013	7	9	22	10	2.00	0.00	9.30	5.34	10.75	5.33	2.72	5.99	5.99	416.18	4
2013	7	9	22	20	1.00	0.00	9.32	5.78	10.99	5.57	2.85	6.44	6.44	416.15	4
2013	7	9	22	30	2.00	0.00	9.48	5.57	11.01	5.55	3.01	6.33	6.33	416.77	4
2013	7	9	22	40	1.00	0.00	9.32	5.37	10.78	5.56	2.84	6.25	6.25	417.88	4
2013	7	9	22	50	2.00	0.00	9.35	5.86	11.05	5.40	2.78	6.09	6.09	417.85	4
2013	7	9	23	0	2.00	0.00	9.72	6.01	11.46	5.50	2.95	6.26	6.26	418.37	4
2013	7	9	23	10	1.00	0.00	9.28	5.81	10.97	5.70	2.96	6.45	6.45	419.17	4

Nas aplicações em que a aquisição de dados é realizada com base em simulação basta ter os dados no formato Excel, formatá-los de forma que só exista um valor de potência activa por cada intervalo de quinze minutos e copiar os dados para um ficheiro do tipo tabela do Matlab (ficheiro com a extensão “.mat”), com o nome dados\_datos\_horas. O ficheiro deverá ficar com a distribuição semelhante ao identificado nas figuras seguintes.

	A	B	C	D	E	F	G	H	I
1	Potência Activa Média (kW)	Dia	Mês	Ano	Hora	Minutos	Segundos	Dia da Semana	Número da Semana
2	2,71	30	7	2014	0	0	0	4	31
3	2,47	30	7	2014	0	15	0	4	31
4	2,72	30	7	2014	0	30	0	4	31
5	3,04	30	7	2014	0	45	0	4	31
6	2,17	30	7	2014	1	0	0	4	31
7	1,95	30	7	2014	1	15	0	4	31
8	1,84	30	7	2014	1	30	0	4	31
9	1,75	30	7	2014	1	45	0	4	31
10	2,08	30	7	2014	2	0	0	4	31

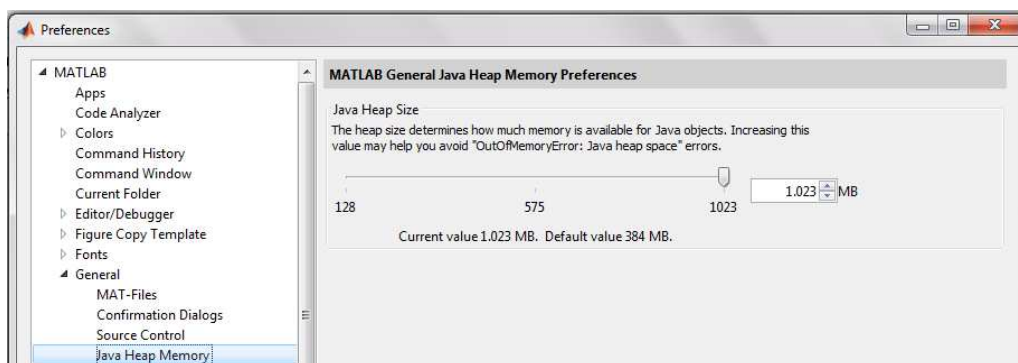
Editor - Tese\_4H\_real.m Variables - dados\_datos\_horas

dados\_datos\_horas 151x9 double

	1	2	3	4	5	6	7	8	9	10	11
1	2.7100	30	7	2014	0	0	0	4	31		
2	2.4700	30	7	2014	0	15	0	4	31		
3	2.7200	30	7	2014	0	30	0	4	31		
4	3.0400	30	7	2014	0	45	0	4	31		
5	2.1700	30	7	2014	1	0	0	4	31		
6	1.9500	30	7	2014	1	15	0	4	31		
7	1.8400	30	7	2014	1	30	0	4	31		
8	1.7500	30	7	2014	1	45	0	4	31		
9	2.0800	30	7	2014	2	0	0	4	31		
10	2.0700	30	7	2014	2	15	0	4	31		
11	2.0400	30	7	2014	2	30	0	4	31		
12	1.7700	30	7	2014	2	45	0	4	31		

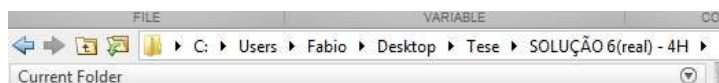
Como já foi referido anteriormente, a execução do programa desenvolvido recorre às funcionalidades do *software* Matlab com uma versão igual ou superior a 2012, caso contrário algumas das funcionalidades implementadas não serão executadas, o que resultará numa falha de execução do sistema. No entanto a implementação das aplicações foi realizada com a versão do Matlab 2014, podendo esta ser executada perfeitamente nas versões anteriores sem ser necessário qualquer conversão.

Devidos às restrições de memória estipuladas inicialmente no Matlab as capacidades de cálculo em memória têm que ser alteradas e estendidas para o máximo. Caso nunca tenham sido alteradas, deverá aceder-se às preferências do programa, seleccionar configurações gerais e a opção “*Java Heap Memory*” (figura seguinte).





Todos os ficheiros deverão encontrar-se numa só pasta em local que permita escrita, caso contrário o programa irá indicar que não é possível gravar as variáveis na pasta seleccionada, terminando a execução da aplicação. O local a ser alterado deve ser o indicado na figura seguinte.



Para correr o painel de comando e o programa de análise em simultâneo deverão ser abertos dois programas de cálculo Matlab, executando num deles a aplicação referente ao painel e no segundo a aplicação referente à aplicação de análise e tratamento dos dados.

Algumas das funcionalidades deverão estar activadas caso não se encontrem de origem no Matlab inicialmente. Duas das principais funcionalidades que terão que estar presentes são a *Database Toolbox* e *Fuzzy Logic Toolbox*. A *toolbox* “*database*” permite a ligação a uma base de dados possibilitando a interacção de dados com a ferramenta de cálculo, permitindo ler os dados existentes ou guardando os valores resultantes numa base de dados externa. A *toolbox* “*Fuzzy Logic*” permite fazer operações de classificação de dados, recorrendo as funcionalidades do algoritmo *fuzzy*, sem esta *toolbox* não será possível fazer a classificação por classes recorrendo as aplicações desenvolvidas.

### **Exemplos dos resultados gráficos das aplicações**

O sistema precisa de algumas variáveis específicas a serem introduzidas pelo utilizador para funcionar adequadamente. Neste caso é necessário atribuir pelo utilizador o número de *clusters* com o qual será programado o padrão de consumo, o sinal de arranque do sistema e seleccionar a opção de carregar dados. No fim de se proceder a essa operação, serão guardadas três variáveis que serão utilizadas no programa de processamento e análise de dados, que neste caso é denominado “Tese\_Tempo\_Total”.

Dado o início do programa este ficará a ser executado em *loop* contínuo, até que lhe seja dada ordem para terminar a operação. No entanto, este poderá parecer que por vezes se encontra parado. Essa situação deve-se ao formato de dados existentes na base de dados, que se encontram espaçados por um intervalo de quinze minutos. Somente quando o ciclo horário for referente ao minuto zero, quinze, trinta ou quarenta e cinco é que serão executadas as operações de análise.

A operação de *cluster* só se realiza se tiverem sido introduzidos novos dados, ou seja, caso tenha sido alterado o número de *clusters* pelo utilizador e tiver sido dado a ordem de arranque ou então caso tenha sido realizada a transição horária para um novo dia. Consoante os parâmetros horários actuais, a base de dados será confrontada com os valores existentes e irá definir quais os valores que serão representantes dos intervalos de dados em análise, para isso será construída uma matriz que irá incluir as várias comparações entre o valor horário actual e o histórico existente. Após serem analisados os valores resultantes, estes serão guardados em variáveis específicas.

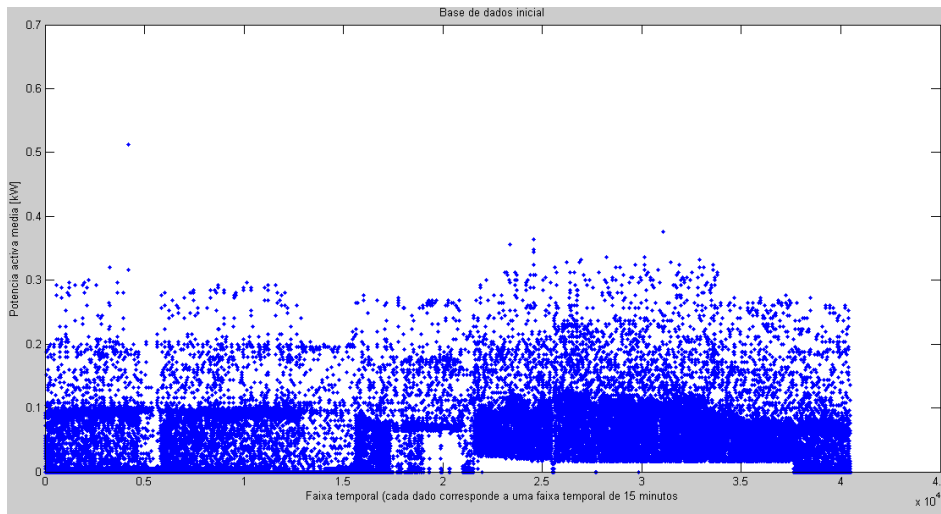
Se por acaso na base de dados tiverem sido introduzidos novos dados e ainda não tenham sido atribuídos graus de pertença, estes serão analisados através da aproximação do seu valor ao grau de pertença de cada grupo e será atribuído um valor referente à sua classe. Isto só acontecerá nas situações em que esteja a ser executado uma organização de valores por *cluster*, recorrendo às ferramentas *Fuzzy C-means clustering*.

Após a classificação do que se considera como resultados temporais semelhantes para comparação para o dia e período actual, procede-se à realização de um valor médio destes valores escalonados para 96 períodos horários de um dia (cada período horário corresponde a um intervalo de quinze minutos). Estes valores médios serão considerados como valores padrão para comparação com os valores actuais, permitindo assim ao utilizador comparar se os valores actuais se encontram de acordo com o valor médio esperado. Caso se encontrem em classificações diferentes, serão emitidos alertas para verificação do estado dos parâmetros actuais, assinalando que poderá estar a decorrer uma anomalia do sistema.

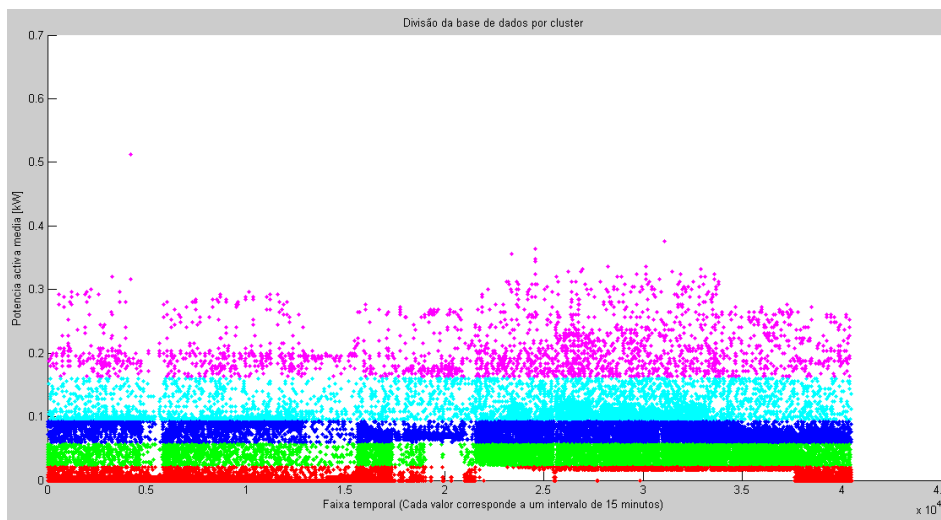
Com base nessas informações são elaborados vários gráficos, sendo mostrados no painel de comando apenas os que serão mais relevantes para o utilizador. Alguns dos gráficos apresentados pelo sistema servem somente para a comparação de resultados, sendo estes necessários para verificação da programação que foi desenvolvida. Caso fossem apresentados todos os gráficos resultantes, poderia tornar-se confuso para o utilizador na comparação de resultados. Os gráficos exportados pelo programa de análise de dados são referidos de seguida.

A figura seguinte é um exemplo da apresentação gráfica de todos valores da base de dados, sendo cada ponto referente a um valor de potência activa. Esta apresentação é referente ao

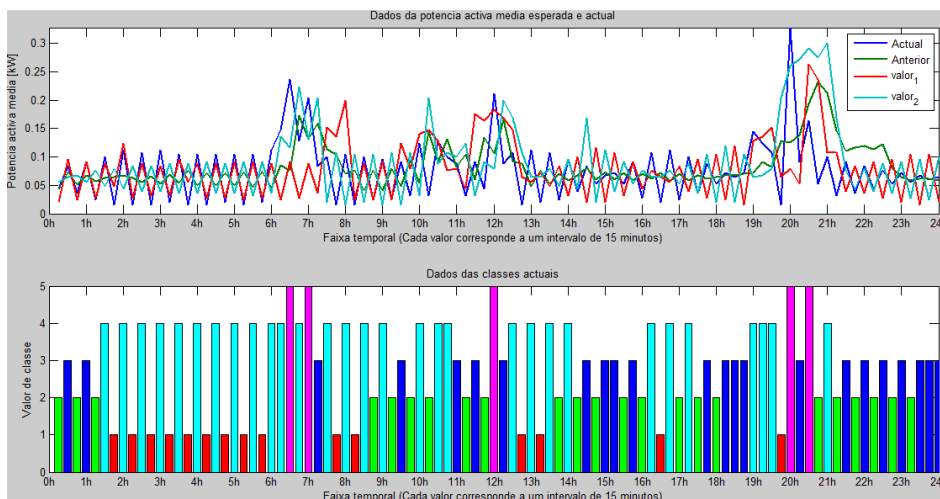
conjunto total de valores num espaço temporal contínuo (eixo dos xx's) e à potência de activa (eixo dos yy's).



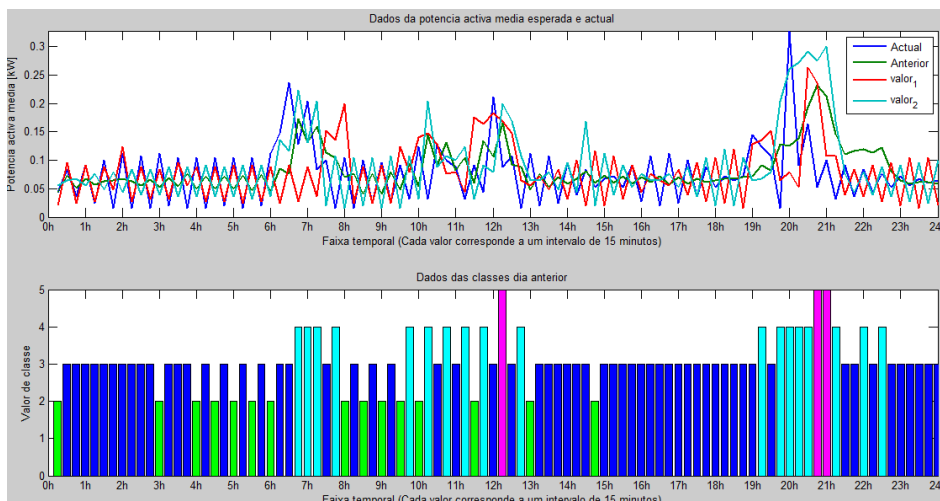
Após todos os valores de potência activa terem sido categorizados por classes, é visualizada uma imagem semelhante à anterior. Contudo, esta é separada por cores onde cada cor corresponde a uma classe diferente.



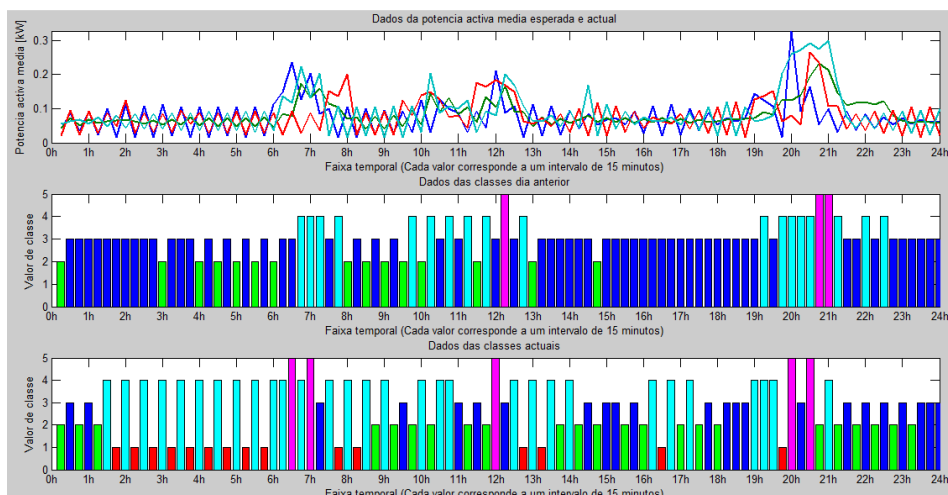
Na figura seguinte são incluídos dois gráficos, no primeiro é representado o gráfico de potência activa e no segundo o valor das classes actuais. O primeiro gráfico apresenta os valores de potência activa para o período actual, período médio e restantes períodos considerados para análise. O gráfico de classes estabelece o tipo de classe dividido por cores, facilitando assim a compreensão da alteração de classes nos vários períodos horários.



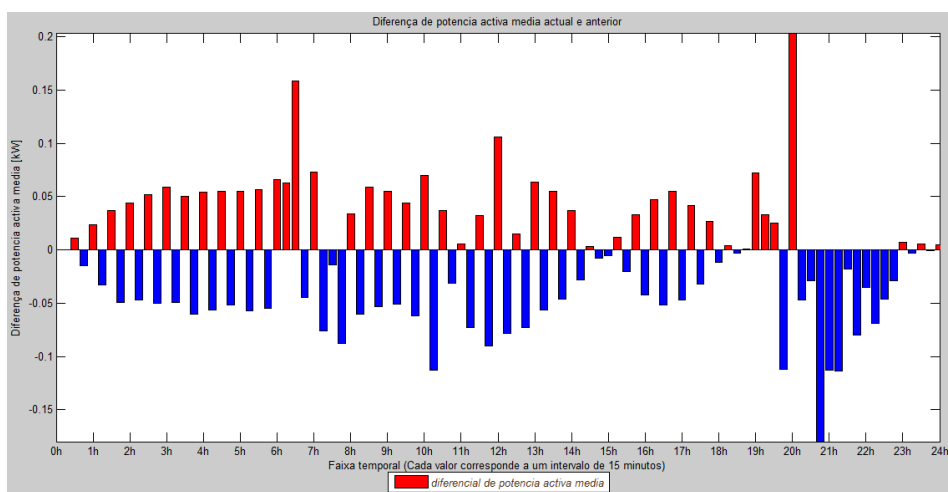
O próximo exemplo é idêntico ao anterior, a única diferença consiste no gráfico de classes que ao invés de ser referente às classes actuais é referente à classe de comparação de valores. Ao encontrar-se uma mistura de gráficos por vezes repetida, acaba por se tornar desinteressante e confuso para o utilizador. A razão da criação destas figuras deveu-se à necessidade de ao longo do desenvolvimento da aplicação de demonstrar graficamente a evolução dos valores de potência, tal como a identificação de anomalias de cálculo resultantes da programação da aplicação.



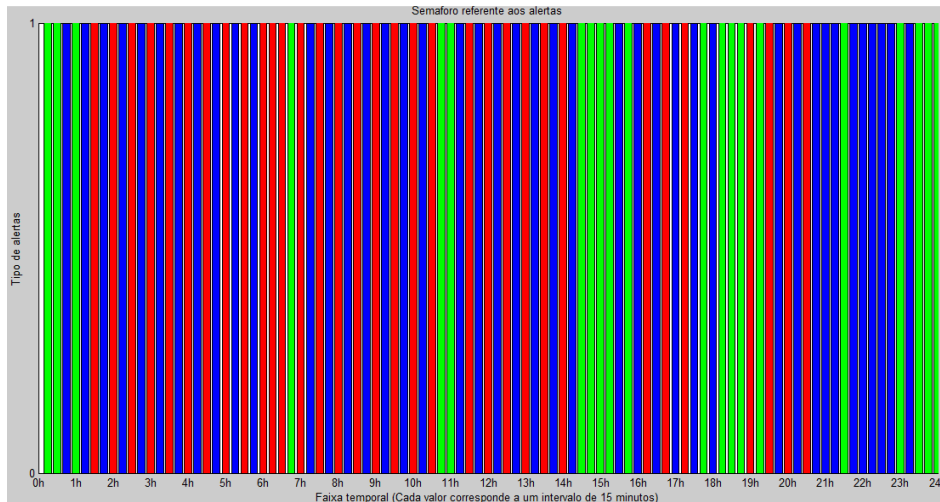
Ao se juntar o gráfico de potências já referido com os dois gráficos de classes (tanto as classes actuais como o gráfico com os valores de classes comparativas), possibilita a verificação da diferença do tipo de classes esperado com o valor das classes obtidas. É possível comparar os valores de potência com as classes, verificando assim a exactidão da sua atribuição e permitindo uma abordagem crítica sobre a razão de existência de um valor diferencial.



O gráfico diferencial de potência activa consiste na diferença dos valores de potência do dia actual com os valores do dia de comparação seleccionado. Por vezes em alguns gráficos, devido ao elevado número de dados existentes, não é permitida uma visualização intuitiva da diferença de potência entre os dois estados. Com este gráfico é possível interpretar com maior facilidade a diferença de valores.



É igualmente apresentado um gráfico que representa para o dia actual diferentes classes correspondendo a diferentes cores, sendo cada classe referente ao nível de alerta existe ao longo do dia.



Os restantes gráficos devolvidos pelo *software*, consistem numa figura única onde são apresentados somente uma das opções referidas anteriores de potência activa e classes (actuais ou esperados).

## Resumo de Instruções para o uso das aplicações

### Arranque das variáveis na aplicação Painel

O valor da variável “*clusters*” é associado ao primeiro campo de escrita identificado na Figura seguinte, sendo que este terá que ser sempre diferente de zero, caso contrário não será assumido qualquer *cluster* na aplicação.

Introdução de dados				
Numero clusters	Arranque de novos dados	Dias comparação	Actualizar Dados	Arranque do programa
5	1	1	Actualizar Tabelas	Parar programa
Dif. de zero	0 - Sem alterações 1 - Novos dados	Dif. de zero		

O segundo campo de escrita é associado à variável “*actualização\_2*”, neste caso o valor a ser introduzido deverá ser de ‘1’ (para um novo número de *clusters*, ou para a etapa inicial caso ainda não tenha sido iniciado o *software*), e ser ‘0’ caso não se pretenda realizar alterações. O terceiro campo de escrita está associado à variável “*num\_comparar*” que, como já foi referido, indica os dias de atraso relativamente ao actual com o qual a aplicação compara resultados. Na etapa inicial do programa o valor de “*num\_comparar*” deverá ser ‘1’.

Para carregar os valores dos campos de escrita para as variáveis deverá ser seleccionado o botão **Actualizar Dados**, após introduzir os dados referidos anteriormente. Existe ainda o botão **Actualizar Tabelas** que irá actualizar as tabelas. Se o programa ainda não tiver sido executado nenhuma vez as tabelas não serão actualizadas. Os botões **Arranque do programa**

e **Parar programa** atribuem o valor de '1' e '0' à variável arranque, que irá permitir o início ou o bloqueio da aplicação.

#### Instruções de uso para a aplicação em tempo real, sem a aplicação painel

Deverão ser definidas as seguintes variáveis:

clusters = 5; %para 5 clusters

actualizacao\_2 = 1; %para implementar a actualização de dados

**arranque = 1; %para dar arranque de dados**

**num\_comparar = 1; % se a base de dados for inferior a um dia e para não esperar um dia ate começar a introduzir resultados**

**save clusters.mat clusters;**

**save actualizacao\_2.mat actualizacao\_2;**

**save num\_comparar.mat num\_comparar;**

**save arranque.mat arranque;**

Para o caso de não existir um base de dados e não se queira esperar um dia para começar a testar a base de dados, deve utilizar o ficheiro Excel, alterar o dia para o actual e o anterior, e alterar igualmente o dia da semana (domingo =1, segunda = 2, ...). Deve-se criar uma nova base de dados, no *Matlab* e colar os valores somente até à hora actual. Se a hora actual for 15H58 colocar valores até às 15H45 e aguardar até à transição horaria. De seguida guardar a base de dados com o nome "dados\_datos\_horas".

**save dados\_datos\_horas.mat dados\_datos\_horas;**

De seguida pode fazer *Run* da aplicação Tese\_4H\_real

#### Instruções de uso para a aplicação em tempo real, com a aplicação painel

Abrir uma nova aplicação *Matlab*. Irão existir duas a trabalhar em simultâneo

Ao introduzir as variáveis nos campos correctos de *clusters*, dias de comparação que inicialmente deverá ser 1, e actualização que também deverá ser 1. Carregar em actualizar dados. Carregar igualmente em arranque de programa. de seguida caso não exista base de dados e não queira esperar igualmente para um dia de resultados introduzir manualmente a base de dados "dados\_datos\_horas.mat" como referido anteriormente