

**UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO  
CENTRO TECNOLÓGICO  
DEPARTAMENTO DE ENGENHARIA ELÉTRICA  
PROJETO DE GRADUAÇÃO**



**AMANDA ALVARENGA DIAS**

**ANÁLISE DA TIPOLOGIA DE CURVAS DE CARGA  
BASEADA EM MÉTODOS DE CLASSIFICAÇÃO DE DADOS**

VITÓRIA – ES  
OUTUBRO/2016

AMANDA ALVARENGA DIAS

**ANÁLISE DA TIPOLOGIA DE CURVAS DE CARGA BASEADA EM MÉTODOS DE  
CLASSIFICAÇÃO DE DADOS**

Parte manuscrita do Projeto de Graduação da aluna **Amanda Alvarenga Dias**, apresentado ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Engenheira Eletricista.

Orientador: Dr. Lucas Frizera Encarnação  
Orientador: MSc. Rodrigo Fiorotti

VITÓRIA – ES  
OUTUBRO/2016

AMANDA ALVARENGA DIAS

**ANÁLISE DA TIPOLOGIA DE CURVAS DE CARGA BASEADA EM MÉTODOS DE CLASSIFICAÇÃO DE DADOS**

Parte manuscrita do Projeto de Graduação da aluna Amanda Alvarenga Dias, apresentado ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Engenheira Eletricista.

Aprovada em 11 de outubro de 2016.

**COMISSÃO EXAMINADORA:**

---

**Prof. Dr. Lucas Frizera Encarnação**  
**Universidade Federal do Espírito Santo**  
**Orientador**

---

**Prof. MSc. Rodrigo Fiorotti**  
**Instituto Federal do Espírito Santo**  
**Orientador**

---

**Profa. Dra. Jussara Farias Fardin**  
**Universidade Federal do Espírito Santo**  
**Examinadora**

---

**MSc. Roger Garcia Almeida**  
**EDP/ESCELSA**  
**Examinador**

Dedico esse meu trabalho para meus familiares, que me sempre me ajudaram e fizeram sacrifícios em favor dos meus sonhos. Dedico também a todos os estudantes de engenharia elétrica da UFES, que sabem como são difíceis os desafios até chegar aqui.

## **AGRADECIMENTOS**

Agradeço a Deus, por me iluminar sempre.

Aos meus pais, Iris e Waldir por todo apoio, conselhos e amor incondicional.

A disponibilidade e excelência dos meus orientadores, Prof. MSc. Rodrigo Fiorotti e Prof. Dr. Lucas Frizera Encarnação, que mesmo tendo muitos compromissos, sempre estiveram à disposição quando requisitados.

Agradeço imensamente aos meus colegas de faculdade, em especial ao Christian Alpoim, pelos estudos e apoio fornecidos ao longo de todo o curso, tornando essa caminhada muito mais leve e prazerosa.

Por fim, aos professores do Departamento de Engenharia Elétrica da Universidade Federal do Espírito Santo, e profissionais com quem tive a oportunidade de trabalhar, pelo crescimento pessoal e profissional.

## RESUMO

O comportamento dos consumidores de energia elétrica é categorizado em diversas classes de consumo ou atividades, integrando o planejamento de distribuição das concessionárias de energia elétrica e as revisões tarifárias do sistema elétrico brasileiro. Para uma operação técnica e economicamente viável do sistema, é importante que os perfis típicos de consumo sejam conhecidos, segundo a determinação da Agência Nacional de Energia Elétrica. Tal conhecimento pode ser obtido através da caracterização de carga de todos os clientes da empresa, no entanto, é tecnicamente inviável devido ao alto custo financeiro gerado e ao longo tempo necessário para a aquisição de dados dos inúmeros transformadores existentes. Tendo em vista a importância da determinação das tipologias de curvas de cargas para as concessionárias de energia elétrica, é apresentado neste trabalho um estudo baseado na teoria de amostragem associada às técnicas de análise de dados, o que torna possível caracterizar o universo de consumidores a partir de uma quantidade reduzida de amostras de medições da carga. Para atingir este objetivo, a classificação das amostras é feita por meio de métodos heurísticos tradicionais e práticas do setor, optando pelo método de partição - método das nuvens dinâmicas e pelo método de agregação - Ward. A comprovação da eficácia da classificação é realizada utilizando ferramentas estatísticas que validam as tipologias das curvas obtidas no processo. A caracterização da carga demonstra que os métodos adotados permitem uma análise mais precisa do mercado de energia elétrica, não só conhecendo o comportamento de suas redes e de seus consumidores mas acompanhando-os sistematicamente.

Palavras-chave: Caracterização de consumo. Tipologia de Curvas de carga. Classificação de dados.

## LISTA DE FIGURAS

Figura 1 – Perfis típicos de curvas de carga de diferentes classes de consumo .....	12
Figura 2 – Etapas da metodologia para a caracterização da carga .....	18
Figura 3 – Seleção das curvas características .....	22
Figura 4 – Sequência de agrupamento realizado pelo método K-means .....	28
Figura 5 – Sequência de agrupamento realizado pelo método Ward .....	29
Figura 6 – Dendrograma aplicando o método Ward .....	29
Figura 7 – Diagrama simplificado da metodologia empregada de caracterização da carga ..	32
Figura 8 – Diagrama simplificado da rede A3.....	32
Figura 9 – Exemplo do processo de exportação dos dados .....	34
Figura 10 – Processo de obtenção das tipologias de curvas de cargas .....	36
Figura 11 – Etapas para efetuar a análise de agrupamento.....	36
Figura 12 – Área de trabalho do software R .....	38
Figura 13 – Curvas de carga em forma de matriz inseridas no programa R .....	39
Figura 14 – Normalização das curvas de carga .....	40
Figura 15 – Dendrograma - dia útil .....	41
Figura 16 – Composição dos agrupamentos (K-means): (a) Cluster 1, (b) Cluster 2 e (c) Cluster 3.....	42
Figura 17 – Estatística Pseudo-F e participação da inércia entre clusters (%BSS) .....	45
Figura 18 – Comparação de tipologias em p.u. da média: (a) método Ward e K-Means e (b) método da distância euclidiana simples .....	46
Figura 19 – Estratificação do mercado de energia das redes A3 estudadas .....	54
Figura 20 – Planilha com tipologias ajustadas ao mercado .....	55

## LISTA DE GRÁFICOS

Gráfico 1 – Exemplo de gráfico para auxílio da escolha das curvas típicas.....	35
Gráfico 2 – Gráfico das médias para cada cluster - dia útil.....	43
Gráfico 3 – Curva de demanda da tipologia 1 .....	51
Gráfico 4 – Curva de demanda da tipologia 2 .....	51
Gráfico 5 – Curva de demanda da tipologia 3 .....	52
Gráfico 6 – Curva de demanda total das redes A3 .....	53
Gráfico 7 – Curvas de demanda sobrepostas das 3 tipologias obtidas .....	53



## **LISTA DE QUADROS**

Quadro 1 – Requisitos das informações para a caracterização da carga .....	18
Quadro 2 – Parâmetros de entrada para os cálculos do ajuste da tipologia ao mercado .....	30
Quadro 3 – Parâmetros complementares no ajuste da tipologia ao mercado .....	31
Quadro 4 – Características dos elementos amostrais escolhidos.....	33
Quadro 5 – Parâmetros de entrada para ajuste da tipologia ao mercado .....	47

## LISTA DE TABELAS

Tabela 1 – Comparação de resultados .....	45
Tabela 2 – Parâmetros de entrada para ajuste da tipologia ao mercado .....	47
Tabela 3 – Perfis de curvas típicas de cada cluster .....	48
Tabela 4 – Valores úteis para o ajuste de tipologia ao mercado .....	50
Tabela 5 – Perfil de demanda em p.u. das tipologias .....	50

## LISTA DE ABREVIATURAS E SIGLAS

ANEEL	Agência Nacional de Energia Elétrica
AT	Alta Tensão
BSS	<i>Between sum of squares</i>
BT	Baixa Tensão
MT	Média Tensão
PRODIST	Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional
SNACC	Sistema Nacional de Análise de Curvas de Carga - desenvolvido pelo Departamento Nacional de Águas e Energia Elétrica (DNAEE)
SRD	Superintendência de Regulação dos Serviços de Distribuição
TUSD	Tarifas de Uso do Sistema de Distribuição
UFES	Universidade Federal do Espírito Santo
WSS	<i>Within-group sum of squares</i>

## SUMÁRIO

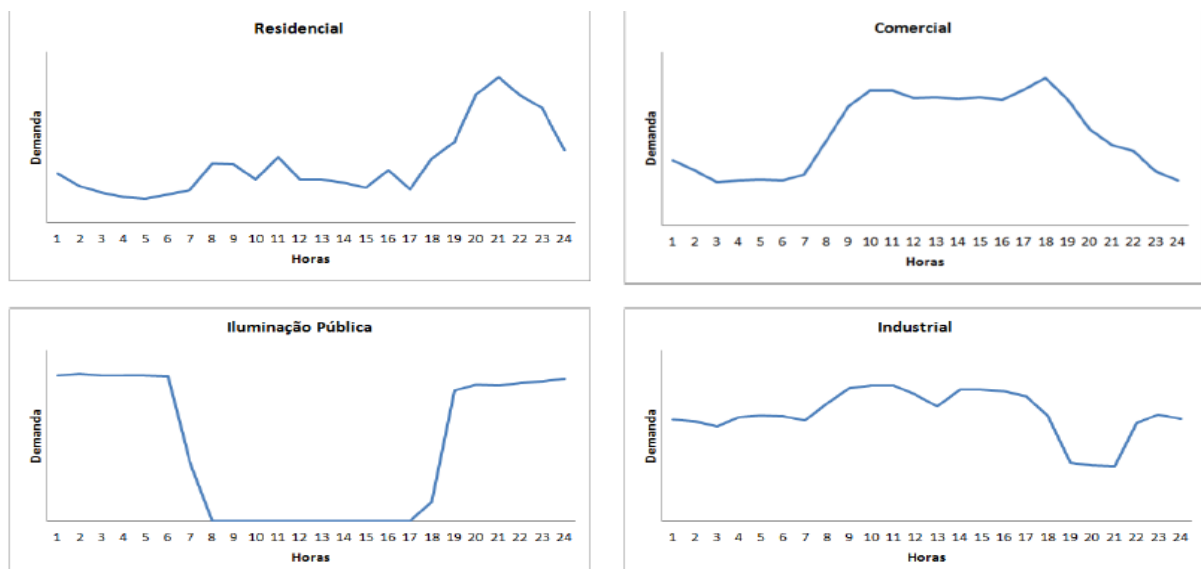
<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>12</b>
1.1	Apresentação .....	12
1.2	Objetivo .....	14
<b>2</b>	<b>ESTADO DA ARTE.....</b>	<b>16</b>
2.1	A origem.....	16
2.2	Caracterização da carga.....	16
2.2.1	Obtenção das amostras .....	18
2.2.1.1	Recuperação de medições .....	19
2.2.1.2	Campanhas de medidas .....	19
2.2.2	Seleção de curvas típicas.....	22
2.2.3	Classificação de dados.....	24
2.2.3.1	Método de nuvens dinâmicas .....	26
2.2.3.2	Método de Ward.....	28
2.2.4	Ajuste da tipologia ao mercado .....	30
<b>3</b>	<b>METODOLOGIA E RESULTADOS.....</b>	<b>32</b>
3.1	Dimensionamento e escolha das amostras .....	32
3.2	Seleção de curvas típicas .....	34
3.3	Metodologia para classificação de dados .....	35
3.3.1	Programa R.....	38
3.3.2	Agrupamento pelo método de Ward .....	39
3.3.3	Nuvem dinâmica - Método das K-Médias (K-Means).....	41
3.3.4	Resultados e avaliações .....	43
3.3.5	Comparação entre métodos de classificação .....	45
3.4	Ajuste da tipologia ao mercado anual de energia .....	46
<b>4</b>	<b>CONCLUSÕES .....</b>	<b>56</b>
	<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>58</b>
	<b>APÊNDICE A .....</b>	<b>61</b>

# 1 INTRODUÇÃO

## 1.1 Apresentação

As formas assumidas pelas curvas de carga retratam o uso da energia elétrica pelos consumidores e apresentam perfis discrepantes nas diversas classes de consumo. A maneira como a demanda de potência evolui ao longo do dia é uma informação fundamental na determinação das tarifas de eletricidade (PESSANHA et al, 2004; GUARDIA et al, 2010). Esta informação encontra-se na curva de carga diária das unidades consumidoras que descreve a trajetória da potência demandada ao longo das horas do dia, conforme ilustrado na Figura 1.

Figura 1 - Perfis típicos de curvas de carga de diferentes classes de consumo



Fonte: PESSANHA et al, 2015.

Uma curva de carga de um consumidor residencial se caracteriza por um consumo praticamente constante durante o dia todo, com aumento no fim da tarde e pico de demanda durante a noite. A curva de carga de consumidores comerciais é caracterizada por ter uma demanda durante o horário comercial (em geral, das 8h às 18h) com leve declínio no horário de almoço, fora do horário comercial a demanda desses consumidores é praticamente para iluminação e refrigeração. A curva de carga de iluminação pública caracteriza-se por um consumo constante durante a noite/madrugada e por não consumir energia durante o período diurno, quando há iluminação natural. Esta é bastante influenciada pela adoção do horário de verão, visto que, se desloca o momento em que a iluminação pública é acionada, estimulando

a redução na demanda de potência no período de pico. As curvas de cargas de consumidores industriais, por sua vez, apresentam enorme variação de atividades, e são estimadas por ramo de atividade e por localização (ALVES, 2011).

A formação de grupos tarifários com base no perfil de carga individual, em função da semelhança de comportamento, possibilita agregar e estratificar os consumidores que impõem custos similares ao sistema, representando os perfis de demanda de cada classe. Essas informações podem ser utilizadas para prever o crescimento das novas redes, melhorando o cálculo de dimensionamento do sistema e melhorando os estudos de planejamento de médio a longo prazo (PESSANHA, 2006).

A análise da curva de carga auxilia o planejamento do setor energético, tanto técnica quanto economicamente, propiciando a alocação otimizada de recursos na expansão da malha de distribuição, nos estudos de cálculos de níveis de tensão futuros, nos cálculos de perdas e na avaliação de investimentos, visando à qualidade da entrega de energia (QUEIROZ, 2011).

Além disso, para o planejamento estratégico de distribuição de energia elétrica é fundamental o mapeamento da curva de carga, a fim de produzir subsídios para a concessionária realizar uma contratação prévia de demanda, principalmente no horário de ponta, buscando maior eficiência do sistema e propor um plano de obras com o intuito de tornar o suprimento de energia elétrica confiável e seguro (QUEIROZ, 2011).

Devido ao grande número de curvas de carga dos consumidores, torna-se necessário a utilização de técnicas de agrupamento e classificação de dados, tanto para a seleção das curvas características de cada transformador quanto para o estudo e obtenção dos grupos ou *clusters* na definição das tipologias.

A tipologia ou perfil típico visa reduzir esse universo de consumidores através da análise dos comportamentos mais incidentes e distintos de uma população. A importância da tipologia na caracterização da curva de carga motiva o desenvolvimento de algoritmos associando a massa de dados das medições de modo a fornecer um diagnóstico mais real do sistema, assim como fornece elementos para controle e intervenção conforme necessário.

Embora haja muitas técnicas de *clusterização* de dados que podem ser aplicadas à construção

de tipologias de curvas de carga, a busca por métodos que realizem um processo de extração de conhecimento de uma ampla base de dados com informações de diferentes tipos de consumidores de energia elétrica é ainda um tópico de grande interesse das concessionárias (ALVES, 2011).

Vários trabalhos têm sido desenvolvidos para a determinação das tipologias das unidades consumidoras, entretanto, essas metodologias atuais não possuem padronização, nem pela ANEEL (Agência Nacional de Energia Elétrica) e nem pelos profissionais que atuam neste mercado. Fato que se torna evidente, principalmente nas seleções das curvas típicas, pois, na maioria das vezes, esta seleção é feita manualmente pelas equipes de estudos, havendo influência do modo de escolha dessas curvas pelo técnico responsável, e, conseqüentemente da classificação nos resultados.

## **1.2 Objetivo**

O objetivo da presente pesquisa é realizar um estudo sobre a caracterização da carga, no qual são utilizados métodos de agrupamentos para a obtenção das tipologias, a partir de amostras de medições de curvas de carga de clientes (transformadores) da rede de distribuição.

Na realização desse estudo, será analisada uma combinação de dois métodos estatísticos multivariados, tradicionalmente utilizada nos processos de classificação de dados, adequando-se aos propósitos da caracterização de carga para o plano das revisões tarifárias.

É importante apresentar de forma individual e incisiva, a classificação dos dados e desta forma, as tipologias de curvas de carga são construídas por meio das funções de análise de agrupamentos disponibilizadas no programa R, *software* livre e *open source*, altamente extensível, no qual são disponibilizadas várias funções para análise de dados e rotinas gráficas, nativas ou obtidas em pacotes (*packages*) distribuídos na internet. Especificamente, neste trabalho é apresentada uma implementação computacional em ambiente R dos métodos *K-Means* e *Ward*.

Após o desenvolvimento e a implementação da metodologia apresentada, a eficácia dos métodos é analisada por meio de ferramentas estatísticas de avaliação. Como parte do processo, as tipologias são ajustadas ao mercado de referência de energia do respectivo estrato.

Vale ressaltar que o principal ganho, possivelmente, seja o aumento da eficácia nos estudos da definição de tipologias de carga, no que diz respeito a confiabilidade, a qualidade dos resultados e ao tempo gasto na análise, que ainda é um tópico de grande interesse das concessionárias.



## **2 ESTADO DA ARTE**

### **2.1 A origem**

Entre as décadas de 1970 e 1980, a crise econômica vivida pelo Brasil causou dificuldades econômico-financeiras, levando a cortes drásticos nos investimentos de expansão do sistema elétrico e comprometendo a qualidade e a quantidade da eletricidade fornecida aos consumidores. Essa situação forçou o setor elétrico a buscar uma melhor utilização da capacidade instalada do sistema, onde se verificou que incentivos tarifários poderiam convencer os consumidores a reorganizarem suas demandas, minimizando os custos para o atendimento de um mesmo consumo através do uso mais racional das instalações existentes (GEMIGNANI, 2009).

Em 1985, a Eletrobras propôs a utilização dos custos marginais, anteriormente aplicada com sucesso em outros países. O custo marginal – custo de fornecimento para atender a um consumo unitário adicional ao existente – informa o custo que será incorrido pelo sistema elétrico para atender o crescimento do consumo. Obtido o custo de fornecimento, pode-se estabelecer uma base tarifária que induza o consumidor a buscar consumos onde a energia é mais barata para ser fornecida. Essa metodologia foi desenvolvida no Ministério de Minas e Energia nos anos anteriores à sua publicação (“Nova Tarifa de Energia Elétrica”, Eletrobras), onde são apresentados os princípios da tarifação pelo custo marginal, um sumário da sequência dos cálculos das tarifas e as modalidades de aplicação na Nova Tarifa (ELETROBRAS, 1985).

Para a determinação das tarifas de referência, alguns processos devem ser realizados: Caracterização da carga, cálculos dos custos marginais do sistema, cálculo das tarifas de referência e aplicação das tarifas, sendo a caracterização da carga, o processo considerado e detalhado ao longo do trabalho.

### **2.2 Caracterização da carga**

O conhecimento temporal do comportamento da carga elétrica dos consumidores é um dos principais fatores que auxiliam as concessionárias em seus planejamentos. Estas curvas de carga são informações importantes da forma como é consumida a energia elétrica.

Além do conhecimento de consumo da energia por parte dos usuários, a informação da potência demandada a cada instante também é de grande importância para os estudos de otimização técnica e econômica do setor elétrico. Tal informação é primordial para os estudos de investimentos, planejamento e tarifação em uma estrutura que pretenda utilizar o sistema elétrico de forma racional, pois não se pode otimizar um sistema se sua carga não for bem conhecida (GEMIGNANI, 2009).

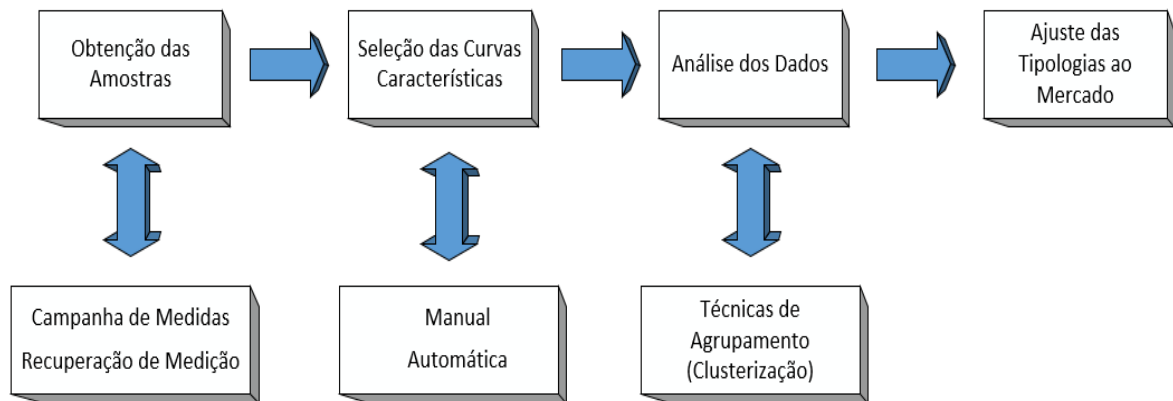
Neste sentido, caracterizar a carga implica:

- Em conhecer o perfil de consumo dos clientes;
- Acompanhar o carregamento horário das redes por nível de tensão;
- Responsabilidade do consumo no custo de expansão;
- Estudos de perdas;
- Planejamento considerando a decisão de investimentos e dimensionamento do sistema;
- Nos projetos de gerenciamento da carga e programas de conservação de energia;
- Na previsão da demanda de energia elétrica nos subsistemas.

Conforme a nota técnica nº63 – SRD/ANEEL (2014, p.6), uma parte expressiva das concessionárias de distribuição apresenta problemas para caracterizar a carga de seus consumidores e do sistema. As diferentes estruturas de mercado e o seu crescimento em cada área de concessão afetam as tipologias das redes, o que requer uma caracterização periódica.

A fim de se conhecer e detalhar o tipo de carga, são realizados os estudos de: obtenção de dados (amostra da população), seleção de curvas características dos elementos da amostra, classificação dos dados (agrupamentos, definição das curvas tipo) e ajuste da tipologia ao mercado. Assim, a Figura 2 apresenta o diagrama do processo da caracterização de carga.

Figura 2 – Etapas da metodologia para a caracterização da carga



Fonte: Produção do próprio autor.

A distribuidora deverá encaminhar à ANEEL, em formato específico definido, a caracterização da carga e de seu sistema elétrico conforme descrito no Quadro 1.

Quadro 1 – Requisitos das informações para a caracterização da carga

Informação	Especificação	Unidade	Periodicidade	Observação
Tipologia de rede	Representada por curvas de carga típicas para um dia útil, um sábado e um domingo, conforme estratificações definidas no Módulo 2	MW	De acordo com o cronograma da Revisão Tarifária Periódica – RTP	<ul style="list-style-type: none"> <li>- Considerar as premissas de medição definidas no Módulo 2</li> <li>- A construção das tipologias a partir das medições realizadas deverá basear-se em conceitos estatísticos consolidados</li> <li>- A integralização da curva deverá ser realizada em intervalos de 15 minutos</li> <li>- Devem ser indicados os postos tarifários (ponta e fora de ponta)</li> <li>- As tipologias deverão ser ajustadas ao mercado de energia da distribuidora</li> </ul>

Fonte: PRODIST, 6, rev. 12, seção 6.2.

### 2.2.1 Obtenção das amostras

A estrutura tarifária atualmente definida realiza apenas medições para o faturamento da energia consumida para a maioria dos consumidores, sendo necessária a obtenção dos dados de potência demandada por todos os segmentos do mercado. As curvas típicas que representam as formas mais importantes e distintas do comportamento das cargas dos consumidores e do sistema elétrico são definidas por amostragem.

Na alta tensão (AT), devido ao pequeno número de consumidores, este trabalho consiste apenas em sistematizar e analisar a coleta de informações disponíveis. Nas médias e baixas tensões (MT e BT), o conhecimento das curvas de carga de cada cliente é inviável, por consequência da grande quantidade de clientes no mercado, ocasionando custo extremamente elevado e uma demanda de trabalho que levaria tempo. A alternativa utilizada atualmente nas revisões tarifárias do sistema elétrico brasileiro emprega a teoria de amostragem associada à técnica de análise de dados, visando reduzir esse universo através da análise de comportamentos mais incidentes e distintos de uma população (GEMIGNANI, 2009).

Segundo a Eletrobras, existem duas maneiras principais para a obtenção das curvas de carga: a recuperação de medições e a campanha de medidas (ELETROBRAS, 1985).

#### **2.2.1.1 Recuperação de medições**

Na recuperação de medições, o trabalho se concentra em recuperar e organizar os arquivos das medições existentes, sendo normalmente utilizado em subestações e nos clientes de alta e média tensão (não se aplica para os clientes convencionais da média tensão), que usualmente possuem o medidor adequado instalado de forma definitiva e não é necessária a instalação de equipamentos exclusivamente para este fim, como nos clientes convencionais de média tensão e na totalidade dos consumidores de baixa tensão.

#### **2.2.1.2 Campanhas de medidas**

A campanha de medidas consiste em um conjunto de técnicas estatísticas e procedimentos organizacionais, que permitem, a partir de amostras, caracterizar o nível do sistema ou universo de consumidores que se deseja conhecer. Para isso, é necessária a instalação de equipamento de medição para o registro das curvas de carga dos consumidores escolhidos para a amostra (GEMIGNANI, 2009). O objetivo de uma campanha de medidas é a obtenção de medições de quantidade e qualidade necessárias para o conhecimento do segmento do mercado que está sendo analisado. Para a realização de uma campanha de medidas adequada, também é preciso planejar bem as atividades intermediárias, que vão desde a organização até a montagem do arquivo final de dados.

O Módulo 2 do PRODIST descreve a atual regulamentação sobre a campanha de medidas. O item 6 estabelece que apenas as concessionárias de distribuição devem realizar a campanha de

medidas para fins de caracterização da sua carga e do seu sistema. Essa obrigação não se aplica às concessionárias de serviço público de distribuição. Além disso, as concessionárias de distribuição que não operam redes em tensão superior a 44 kV podem enviar as tipologias na forma simplificada (PRODIST, 2016).

Segundo a Eletrobras (1985) o principal problema das campanhas de medidas “... é garantir que as amostras que são usadas [...] sejam obtidas por processo adequado e permitam conhecer a população estudada”. Para solucionar esse problema é necessário determinar as variáveis explicativas do comportamento da população, e, para que os dados obtidos sejam consistentes, principalmente ter o cuidado com os critérios determinados para escolha da amostra.

Uma amostra é composta por uma parte da população sobre a qual as variáveis de interesse serão observadas e analisadas, e transfere-se o conhecimento adquirido com a amostra para a população (processo inferencial). Na seleção da amostra, a representatividade e a precisão são importantes para garantir que o conhecimento da mesma possa ser transferido para a população. Não há um método preciso que garanta a representatividade de uma amostra, sendo uma boa tática a estratificação da população pelas variáveis que explicam o comportamento da potência demandada pelos consumidores.

As variáveis que mais influenciam a demanda por energia elétrica são: portes de consumo ou demanda, porte da cidade, atividade econômica ou processo produtivo, renda, equipamentos elétricos utilizados, dentre outros. Unir o processo aleatório (sorteio das medições que servirá como amostra) à estratificação, leva a uma amostra com probabilidade bastante reduzida de não ser representativa. A estratificação força que os consumidores de certos segmentos estejam presentes na amostra independente do processo aleatório de seleção.

O tamanho da amostra está relacionado a um segundo ponto importante, que é o erro amostral. Se toda a população fosse estudada, não haveria erro amostral. A precisão do trabalho é inversamente proporcional ao erro amostral. A estatística básica para medir o erro amostral é o desvio padrão da variável de interesse da amostra. Mesmo sabendo-se que a alta precisão não significa necessariamente representatividade (daí a importância da estratificação), o tamanho da amostra a ser selecionada torna-se fator relevante, desde que relacionado com os dois principais fatores levantados (GEMIGNANI, 2009). Ao se conhecer

a dispersão da potência demandada e admitindo-se uma determinada probabilidade de erro, pode-se calcular o tamanho da amostra. A Equação (1) apresenta esse cálculo.

$$n = \frac{t^2 * \left(\frac{CV}{R}\right)^2}{1 + \frac{t^2}{N} * \left(\frac{CV}{R}\right)^2} \quad (1)$$

Onde:

$n$ -tamanho da amostra;

$CV$ -coeficiente de variação médio (desvio padrão dividido pela média);

$R$ -erro amostral;

$N$ -tamanho da população;

$t$ -fator de abrangência, valor retirado das Tabelas de distribuição normal, sendo:

$t = 1,64$  para intervalo de confiança de 90%;

$t = 1,96$  para intervalo de confiança de 95%;

A distribuição normal é utilizada para populações onde o desvio padrão é conhecido e em casos onde não se conhece, utiliza-se a distribuição t-Student. Como a população dos estudos de caracterização de carga é normalmente superior a 100 elementos, o valor “ $t$ ” pode ser bem aproximado pela distribuição normal. Na Equação (1), os CVs são calculados para todas as demandas medidas em cada consumidor e transformador (GEMIGNANI, 2009).

De acordo com o PRODIST – Módulo 2 (rev. 7, sec. 2.2, 2016), “o nível de erro aceitável da amostra de cada estratificação deve ser no máximo 20% com o nível de confiança de 95%”. A SRD (Superintendência de Regulação dos Serviços de Distribuição) entende que esse percentual máximo está adequado visto que a sua redução causaria aumento no número de amostras de cada estrato, o que resultaria em mais custos na realização da campanha de medidas. Assim, propõe-se a manutenção do erro amostral máximo em 20%, e a fórmula que determina o número de amostras se resume na Equação (2).

$$n = \frac{1.96^2 * \left(\frac{CV}{0.2}\right)^2}{1 + \frac{1.96^2}{N} * \left(\frac{CV}{0.2}\right)^2} \quad (2)$$

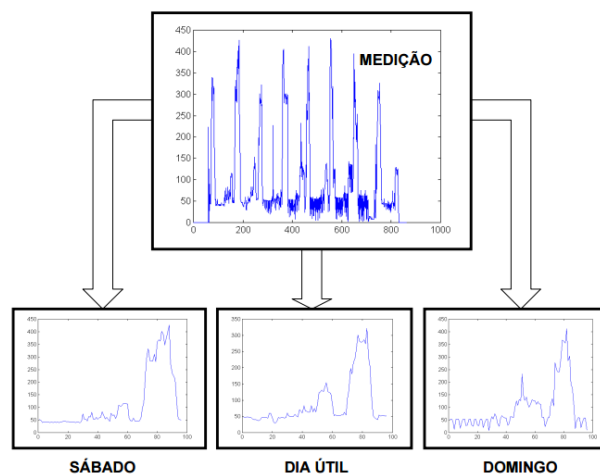
### 2.2.2 Seleção de curvas típicas

O objetivo desta etapa é a obtenção das curvas de carga características que melhor represente o perfil típico de cada cliente. Os clientes elegidos na amostra passam por um processo de seleção de curvas de carga, onde são escolhidas três curvas por cliente (transformador), uma curva característica para dias úteis (segunda-feira a sexta-feira), uma curva característica para sábado e uma curva característica para domingo. As curvas características do dia útil formam o conjunto de curvas a serem classificadas, portanto, são elas que definem os perfis das tipologias, já as curvas de sábado e domingo servem apenas para calcular fatores de ponderação usados no ajuste dos perfis típicos ao montante de energia anual (MWh) do segmento que eles representam.

A seleção das curvas típicas pode ser obtida conforme as seguintes etapas de procedimento de tratamento dos dados: leitura de dados, eliminação das curvas incompletas, eliminação das curvas com valores nulos e seleção da melhor curva.

Na leitura de dados, as curvas de cargas obtidas durante a campanha de medidas são agrupadas e separadas em dias úteis, dias de sábado e dias de domingo que caracterizam o ponto de medição. Essa separação expressa os reais hábitos de consumo de energia elétrica que variam em dias úteis (segunda-feira à sexta-feira), sábado e domingo, resultando em três curvas que representam o comportamento típico, conforme ilustrado na Figura 3.

Figura 3 – Seleção das Curvas Características



Fonte: ANEEL, 1999.

A próxima fase é a eliminação das curvas que não apresentem a mesma quantidade de pontos esperados para a mesma leitura. De uma forma geral, o conjunto de dados coletados na campanha de medidas durante, por exemplo, 15 dias com intervalos de discretização de 15 minutos resulta em vetor com 96 pontos por dia, para os 15 dias. Assim, toda curva de carga que apresente uma quantidade inferior de pontos, é eliminada. Geralmente são enquadradas neste critério de descarte, as primeiras e as últimas curvas, pois tal situação depende do horário de início e término da obtenção das medidas (GEMIGNANI, 2009).

Curvas de cargas que possuem valores nulos em percentual superior a um determinado valor de referência também são usualmente eliminadas. Por exemplo, pode-se admitir que curvas de cargas com valores zerados acima de 30% sejam eliminadas e desconsideradas do processo. A presença de valores nulos nos conjuntos de amostras ocorre principalmente devido a eventuais desligamentos ou erros de leituras causados por defeitos nos equipamentos.

Realizado os procedimentos de eliminação das curvas incompletas e com valores nulos dos dados de leituras, têm-se então os conjuntos das curvas que melhor representam o comportamento da carga.

Para a seleção das curvas características, existem duas metodologias utilizadas atualmente: manual e automática. A seleção manual é feita a partir da análise visual das curvas de carga de cada arquivo de medição. A identificação das curvas é realizada por meio de uma inspeção visual, porém, este procedimento consome muito tempo do analista conhecedor do mercado.

Existem programas disponíveis no mercado que permitem a visualização das curvas medidas facilitando a escolha das curvas características. Esta rotina não elimina o trabalho do analista, mas visa auxiliá-lo. A ideia é que no momento da inspeção visual o programa forneça a indicação de três curvas ao analista, deixando para ele a decisão final de aceitar a sugestão do programa ou a selecionar manualmente outras curvas. Quando as curvas são muito semelhantes, o resultado fica sujeito à precisão visual do usuário. A visualização pode ser realizada em diferentes intervalos de integração das curvas, porém, com intervalos maiores, as curvas se tornam mais suaves e a identificação pode se tornar mais fácil. Esta análise das curvas é feita com critérios característicos da equipe de estudos, com avaliações possivelmente subjetivas e sem muito rigor na padronização. Os resultados podem variar de acordo com o analista envolvido (GEMIGNANI, 2009).



A seleção automática é feita em *softwares* com o objetivo de agilizar o processo, aumentar a confiabilidade e permitir a reprodução de resultados. Esse tipo de seleção visa eliminar o caráter subjetivo da análise visual, permitindo que pequenas diferenças no perfil das curvas sejam consideradas no processo. Uma etapa que se adota no modo automático é o cálculo da curva média para cada conjunto. O critério implementado na rotina é bastante simples: dado um arquivo de medição, calculam-se as três curvas diárias médias, uma para sábado, uma para domingo e outra para o dia útil. Em seguida identificam-se para cada um destes dias quais as medições diárias registradas mais próximas de suas curvas médias.

Essas curvas características selecionadas compreendem o primeiro passo para a determinação das tipologias dos consumidores através das campanhas de medidas. Essas melhores curvas selecionadas serão utilizadas para determinação das tipologias de curvas de cargas dos consumidores.

### **2.2.3 Classificação de Dados**

Na análise de tipologia, quase todos os métodos usam os conceitos de semelhança e distância entre indivíduos para agrupá-los ou separá-los. O agrupamento dessas curvas semelhantes força que se tenham grupos com desvios padrões menores entre os seus respectivos elementos. Nesta etapa é utilizado um algoritmo de *cluster analysis* para agrupar as curvas do dia útil.

Segundo ANEEL (2000, p.15):

[...] Dado um conjunto de curvas características, as técnicas de “cluster analysis” segregam este conjunto em subconjuntos ou “clusters”, de tal forma que as curvas de um mesmo subconjunto apresentem uma grande semelhança entre si, mas as curvas de subconjuntos diferentes apresentem uma baixa similaridade. No final, cada subconjunto será representado por uma curva típica, ou melhor, uma tipologia.

Para se agregar as curvas típicas em formas típicas ou homogêneas, é necessário se utilizar técnicas de agrupamento estatístico. Existem dois grandes métodos de classificação: métodos hierárquicos e os métodos não hierárquicos.

- Hierárquico: consiste em uma série de sucessivos agrupamentos ou sucessivas divisões de elementos e envolve a construção de uma hierarquia de uma estrutura do tipo árvore, chamada de dendrograma. Basicamente existem dois tipos de

procedimentos hierárquicos de agrupamento - aglomerativo e divisivo (DONI, 2004).

- Não hierárquico: são técnicas de realocação iterativas que para melhorar a qualidade do agrupamento, os algoritmos são executados diversas vezes com diferentes partições iniciais e a melhor configuração obtida a partir desses resultados é selecionada como o agrupamento final (SOUZA, 2003).

Nos métodos hierárquicos aglomerativos, cada objeto começa como seu próprio agrupamento. Em passos seguintes, os dois objetos mais próximos são combinados em um novo agregado, reduzindo assim o número de agrupamentos em uma unidade em cada passo. Em alguns casos dois grupos de objetos formados em um estágio anterior podem se juntar em um novo agrupamento. Eventualmente, todos os objetos são reunidos em um único agregado. Existe uma variedade de métodos aglomerativos que são caracterizados de acordo com o critério utilizado para definir as distâncias entre grupos, tais como: métodos de ligação (*single linkage, complete linkage, average linkage, median linkage*); métodos de centróide; métodos de soma de erros quadráticos ou variância (método de Ward).

Uma característica importante dos procedimentos hierárquicos deve-se ao fato dos resultados de um estágio anterior ser sempre aninhado com os resultados de um estágio posterior, criando algo parecido com uma árvore.

O outro tipo de agrupamento hierárquico é o método divisivo. Este método começa com um grande agregado que contém todos os objetos. Em passos sucessivos, as observações mais diferentes entre si são separadas e transformadas em agrupamentos menores. Esse processo continua até que cada objeto seja um agrupamento por si mesmo. Os métodos divisivos são pouco mencionados na literatura, pois exigem uma maior capacidade computacional que os métodos aglomerativos (KAUFMAN, 1990).

Os métodos não hierárquicos caracterizam-se por admitir que um indivíduo seja alocado em um grupo e posteriormente em outro. Os métodos mais comuns deste tipo são o agrupamento em torno de centros móveis (k-medóides) e o método das nuvens dinâmicas (k-médias). A ideia central da maioria dos métodos não hierárquicos é escolher uma partição inicial dos elementos e, em seguida, alterar os membros dos grupos para obter-se a melhor partição. Quando comparado com o método hierárquico, o método por particionamento é mais rápido porque não é necessário calcular e armazenar, durante o processamento, a matriz de

similaridade. Em geral, os métodos por particionamento diferem entre si pela maneira que constituem a melhor partição (DONI, 2004).

No setor elétrico brasileiro, o SNACC (Sistema Nacional de Análise de Curvas de Carga) foi o primeiro sistema computacional para construção de tipologias de curvas de carga. Neste sistema, desenvolvido em meados da década de 1980, os *clusters* são obtidos pelo uso combinado de dois métodos estatísticos para análise de agrupamentos: nuvens dinâmicas e método de descrição e classificação ascendente hierárquico. Os dois métodos são executados em sequência: primeiro o método das nuvens dinâmicas identifica os *clusters* iniciais (formas fortes) e em seguida estes *clusters* são agrupados por um método hierárquico semelhante ao de Ward (PESSANHA, 2006).

### **2.2.3.1 Método de nuvens dinâmicas**

Os algoritmos do tipo nuvens dinâmicas são uma família de métodos de *cluster* não hierárquicos que têm por objetivo obter uma partição de um conjunto de elementos em um número predefinido de classes e identificar um conjunto de indivíduos que apresenta atributos semelhantes entre si minimizando um critério que mede a adequação entre as classes e os protótipos.

O algoritmo inicia com um conjunto de representantes ou uma partição aleatória e aplica iterativamente uma etapa de alocação para afetar cada elemento a classe onde a proximidade entre o elemento e o protótipo é mínima, e uma etapa de representação onde os protótipos são atualizados de acordo com o resultado da etapa de alocação. Estas duas etapas são realizadas até a convergência do algoritmo, quando o critério de ajustamento alcança um valor estacionário. Para melhorar a qualidade do agrupamento, o algoritmo é executado diferentes vezes com diferentes partições iniciais, e a melhor configuração de pontos é usada como resultado do agrupamento (SOUZA, 2003).

O algoritmo utilizado neste método tem dois passos de realocação iterativos envolvendo a construção das classes e a identificação de representantes ou protótipos das classes tais como media, mediana, distribuição de probabilidade, etc. otimizando um critério que mede a adequação entre as classes e seus representantes. O algoritmo de *k-means* é um caso particular dos algoritmos do tipo nuvem dinâmica onde os protótipos são os centroides e a função critério é baseada na distância euclidiana medindo as diferenças entre os centroides e as

classes (SOUZA, 2003). A Equação (3) expressa que para uma dada partição dos  $n$  objetos, o centroide do  $j$ -ésimo *cluster*,  $j=1,k$  é a média dos objetos classificados no *cluster*.

$$C_j = \frac{1}{n_j} \sum_{cluster\_k} X_i \quad (3)$$

O critério utilizado pelo método *k-means* na classificação dos objetos em  $k$  *clusters* consiste em minimizar a variabilidade dentro dos agrupamentos, mas, o algoritmo não garante a convergência para uma solução ótima. Conforme a Equação (4), ele é expresso pela soma dos quadrados dos desvios entre o centroide do *cluster* e os objetos que nele foram classificados (*within-group sum of squares* - WSS):

$$WSS = \sum_{k=1}^K \sum_{i \in cluster\_k} \sum_{j=1}^P (X_{ij} - C_{kj})^2 \quad (4)$$

Onde:

$X_{ij}$ - é a demanda da  $i$ -ésima curva de carga na  $j$ -ésima variável ( $j=1,p$ );

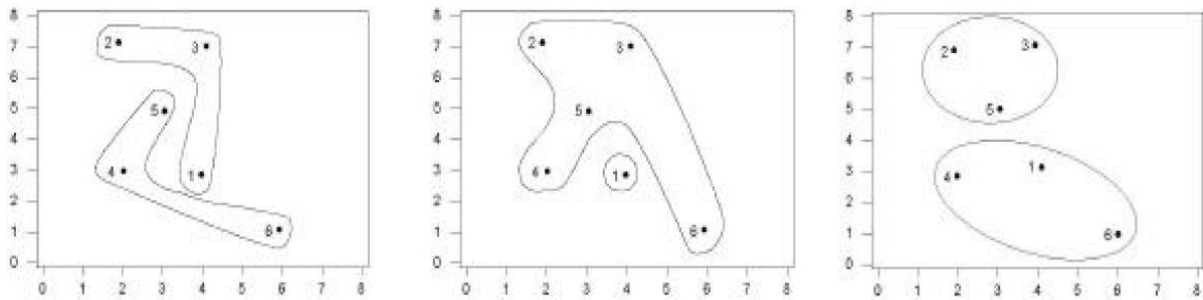
$C_{kj}$ - é a  $j$ -ésima coordenada do centroide do  $k$ -ésimo agrupamento e  $i \in cluster\_k$  denota os índices de todas as curvas de carga classificadas no  $k$ -ésimo agrupamento.

A inércia entre os agrupamentos (BSS – *between sum of squares*), por sua vez, expressa a variabilidade entre os centroides dos agrupamentos com base na Equação (5) e também é utilizada como ferramenta de validação dos métodos de agrupamento.

$$BSS = \sum_{k=1}^k n_k \sum_{j=1}^p (c_{kj} - \bar{x}_j)^2 \quad (5)$$

A fim de exemplificar o método, a Figura 4 particiona um conjunto de 6 elementos, assumindo como parâmetro de entrada 2 grupos iniciais tais como (1,2,3) e (4,5,6), onde são feitas as iterações verificando-se a necessidade de realocação dos objetos. Observa-se que cada elemento está associado ao grupo com o centroide mais próximo, ou seja, cada elemento possui menor distância em relação ao grupo no qual faz parte do que em relação ao outro grupo, conforme estabelece o método apresentado (DONI, 2004).

Figura 4 – Sequência de agrupamento realizado pelo método K-means



Fonte: DONI, 2004.

### 2.2.3.2 Método de Ward

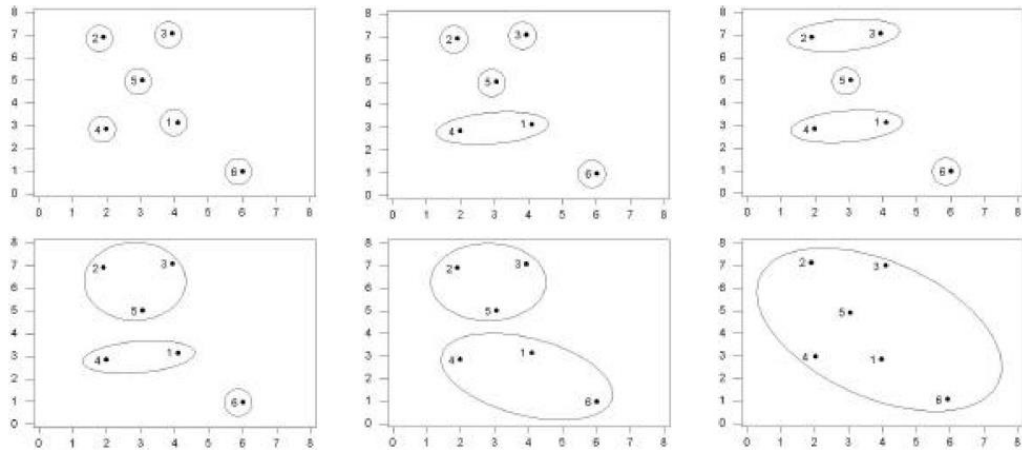
O método de Ward é um método hierárquico aglomerativo altamente eficiente na formação de grupos e é o mais comumente utilizado. Como explicado anteriormente, no início cada agrupamento (*cluster*) tem-se apenas uma curva de carga, e em cada iteração de execução do algoritmo os *clusters* são agregados dois a dois até que reste apenas um grupo com todos os elementos. No método Ward, a distância entre dois agrupamentos é a soma dos quadrados entre ambos, feita sobre todas as variáveis. Em cada estágio do procedimento de agrupamento, a soma interna de quadrados é minimizada sobre todas as partições que podem ser obtidas pela combinação de dois agregados do estágio anterior.

O método baseia-se em uma matriz simétrica de ordem  $n$ , onde o elemento  $ij$  guarda a distância entre os *clusters*  $i$  e  $j$ . Inicialmente as distâncias correspondem aos quadrados das distâncias euclidianas entre os próprios objetos, pois cada *cluster* tem apenas um elemento. Os *clusters* mais próximos são os mais semelhantes e, portanto, são os primeiros a serem agrupados. À medida que os *clusters* vão sendo agrupados (Figura 5), a ordem da matriz de distâncias diminui e as distâncias são recalculadas com base na Equação (6):

$$d_{ij} = \frac{p_i p_j}{p_i + p_j} d^2(c_i, c_j) \quad (6)$$

em que  $p_i$  e  $p_j$  denotam as quantidades de objetos nos *clusters*  $i$  e  $j$  respectivamente e  $d^2(c_i, c_j)$  representa o quadrado da distância euclidiana entre os centroides dos agrupamentos  $i$  e  $j$  (PESSANHA, 2015).

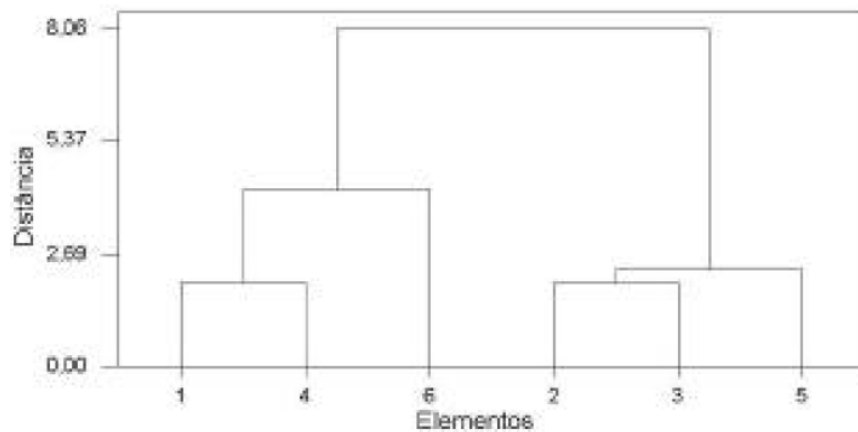
Figura 5 – Sequência de agrupamento realizado pelo método Ward



Fonte: DONI, 2004.

A Figura 6 mostra o dendrograma gerado pelo método de ligação de Ward.

Figura 6 – Dendrograma aplicando o método Ward



Fonte: DONI, 2004.

O dendrograma proporciona soluções para diferentes níveis de agregação dos elementos. Quanto maior o comprimento dos segmentos verticais mais distintos são os ramos que convergem para um mesmo ponto no dendrograma, ou seja, mais distintos são os *clusters* agrupados nessa etapa. A princípio são agrupados os elementos ou *clusters* mais semelhantes, assim, na base do dendrograma os segmentos verticais são curtos. À medida que o processo de aglomeração se desenvolve, *clusters* cada vez mais distintos são agrupados e os segmentos verticais tornam-se cada vez mais longos. Logo, uma boa estratégia para a formação de agrupamentos homogêneos é observar o momento em que os ramos ficam longos e então classificar todos os elementos conectados ao ramo em um mesmo *cluster*.

### 2.2.4 Ajuste da tipologia ao mercado

Devido à inviabilidade prática da construção de tarifas a partir da análise do comportamento individual da curva de carga dos consumidores e das instalações de transformação de tensão, torna-se necessária a definição de um número conveniente de curvas de carga típicas. Estas curvas de carga devem representar a totalidade dos consumidores e das instalações de transformação de tensão da concessionária. Os agregados das tipologias de carga já ajustados ao mercado do ano teste para cada subgrupo tarifário da distribuidora são estimados considerando as curvas típicas encaminhadas à ANEEL nos períodos de revisão tarifária.

O ajuste consiste na extrapolação da energia das tipologias ao mercado anual do nível de tensão ou classe de consumo que elas representam. Para se realizar os cálculos necessários para fazer o ajuste da tipologia ao mercado, é essencial possuir as informações descritas no Quadro 2.

Quadro 2 – Parâmetros de entrada para os cálculos do ajuste da tipologia ao mercado

U(h)	Curva de carga das tipologias no dia útil
S(h)	Curva de carga das tipologias no sábado
D(h)	Curva de carga das tipologias no domingo
M	Mercado anual em MWh
Nu	Total de dias úteis
Ns	Total de sábados
Nd	Total de domingos

Fonte: Produção do próprio autor.

As curvas são obtidas do estudo de caracterização da carga; o mercado anual e a quantidade de dias são dados conhecidos pelas empresas. Com todas essas informações é possível determinar os parâmetros definidos no Quadro 3 através das suas respectivas fórmulas.

Quadro 3 – Parâmetros complementares no ajuste da tipologia ao mercado

Parâmetro	Fórmula
$FC_i$ = fator de carga da tipologia i	$\frac{\sum_{h=1}^{24} U(h)_i / 24}{\max(U_h)_i} \cdot 100$
$EU_i$ = Energia diária do dia útil da tipologia i	$\sum_{h=1}^{24} U(h)_i$
$ES_i$ = Energia diária do sábado da tipologia i	$\sum_{h=1}^{24} S(h)_i$
$ED_i$ = Energia diária do domingo da tipologia i	$\sum_{h=1}^{24} D(h)_i$
$PS_i$ = Ponderação do sábado da tipologia i	$ES_i / EU_i$
$PD_i$ = Ponderação do domingo da tipologia i	$ED_i / EU_i$
$CA_i$ = consumo anual da tipologia i	$EU_i \cdot 24 \cdot (U_i + S_i \cdot PS_i + D_i \cdot PD_i)$
$\alpha_i$ = Participação da tipologia i	$CA_i / \sum_{i=1}^N CA_i$
$ET_i$ = Energia anual da tipologia i	$\alpha_i \cdot M$
$DM_i$ = Demanda média da tipologia i	$ET_i / 24 \cdot (U_i + S_i \cdot PS_i + D_i \cdot PD_i)$
$U(h)_i^{pu}$ = demanda da tipologia i na hora h em p.u. da média	$U(h)_i / \left( \frac{\sum_{h=1}^{24} U(h)_i}{24} \right)$

Fonte: GEMIGNANI, 2009.

Para completar o ajuste é preciso calcular o fator de carga geral da tipologia, o fator de carga fora ponta, o fator de carga de ponta e a energia consumida, pois são importantes nos cálculos dos custos marginais da TUSD (Tarifa de Uso do Sistema de Distribuição) e outros estudos técnicos e de mercado das concessionárias de energia elétrica.

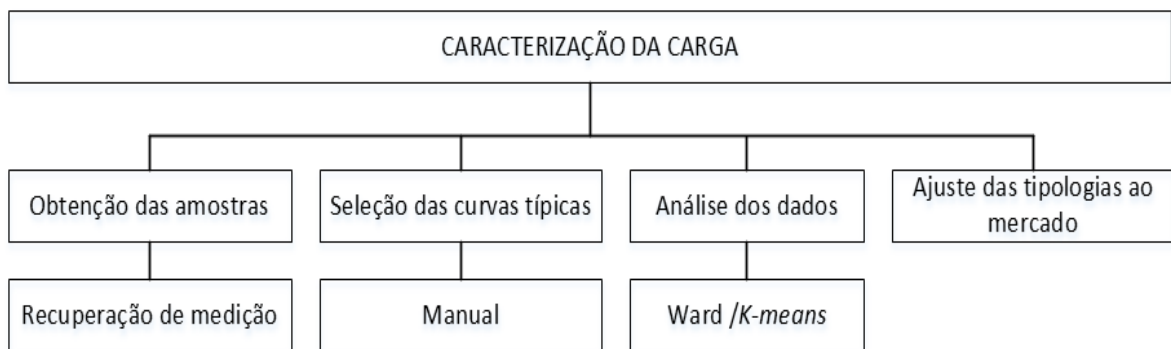
Com todos os resultados adquiridos corretamente, conclui-se a etapa de caracterização da carga.



### 3 METODOLOGIA E RESULTADOS

Este capítulo descreve a metodologia empregada para a caracterização das cargas e redes, além dos resultados obtidos. As etapas utilizadas na caracterização da carga serão descritas segundo a prática atual do mercado. A Figura 7 apresenta o diagrama simplificado do processo segundo a metodologia adotada neste trabalho.

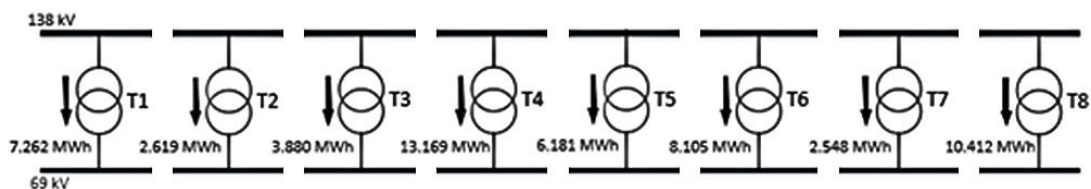
Figura 7 – Diagrama simplificado da metodologia empregada para caracterização da carga



Fonte: Produção do próprio autor.

Os dados amostrais foram obtidos a partir do levantamento realizado no trabalho de (FIOROTTI, 2013). Eles são provenientes de um segmento do mercado de transformadores de rede classe A3 de uma determinada localidade e possuem valores de grandezas reais, tais como: as medições, o mercado anual de energia, a quantidade de dias úteis, de sábados e de domingos do ano em que as medições foram realizadas, entre outras. A fim de preservar a identidade dos dados técnicos do mercado e dos transformadores, os elementos serão denominados conforme mostrado na Figura 8.

Figura 8 – Diagrama simplificado da rede A3



Fonte: Produção do próprio autor.

#### 3.1 Dimensionamento das amostras

Todas as medições foram obtidas através do procedimento da recuperação das medições, visto

que esses dados são periodicamente utilizados nos setores técnicos e comerciais da concessionária local, e, todos os elementos deste segmento do mercado serão utilizados para a elaboração da caracterização da carga. Esta etapa serve apenas como dado de entrada para os estudos seguintes. O Quadro 4 mostra os nomes fictícios dos transformadores, o número de elementos utilizados na amostra e o tamanho da amostra.

Quadro 4 – Características dos elementos amostrais escolhidos

Transformadores	T1 – T8
Número total de elementos	8
Tamanho da amostra	8
Número de pontos da curva diária	96

Fonte: Produção do próprio autor.

Tendo conhecimento prévio dos valores do erro amostral de campanha de medidas anteriores, é possível prognosticar o número de amostras necessárias para um determinado intervalo de confiança e um erro amostral através da Equação (1).

De acordo com o que foi explicado anteriormente, como base de cálculo das amostras será aplicado um intervalo de confiança de 95%, fator de abrangência  $t=1,96$  e um erro amostral máximo de 20%, e com estes parâmetros qualitativos definidos com base nas medidas do PRODIST, o número de amostras necessárias é determinado pela Equação (2). Como todos os elementos desta amostra são obtidos pelo processo da recuperação das medições, não foi necessário se conhecer o coeficiente de variação para calcular o número de amostras.

No cálculo para determinar o erro amostral incorrido no processo inferencial, isolando  $R$  e para  $t = 1,96$  na Equação (1), de acordo com a Equação (7), temos que:

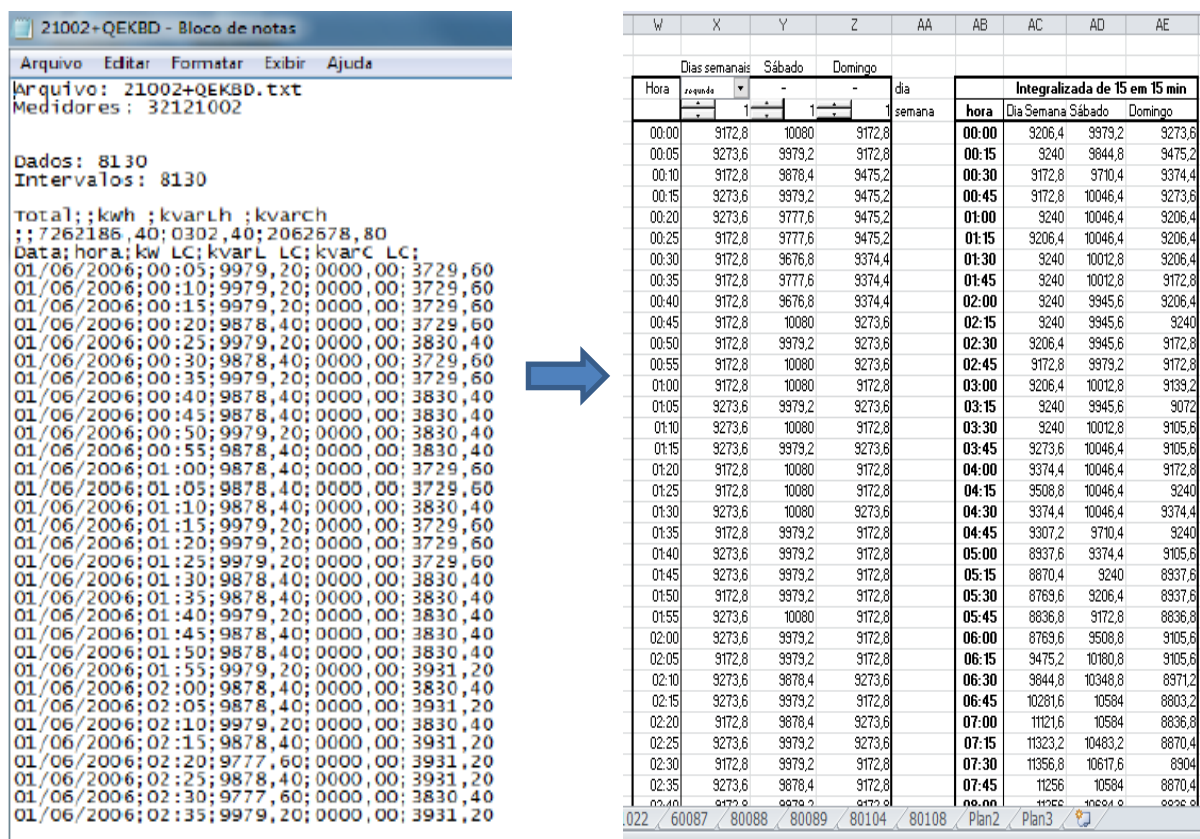
$$R = 1,96 * CV \sqrt{\frac{1}{n} - \frac{1}{N}} \quad (7)$$

Como  $n = N = 8$ , temos que para qualquer valor de  $CV$  aplicado na equação anterior, o erro amostral será zero. Desta forma, sempre que se utilizam todos os elementos da amostra, esta etapa da caracterização da carga é desnecessária.

### 3.2 Seleção de Curvas Típicas

O primeiro estudo realizado se refere à seleção de curvas características. Utilizando as curvas dos clientes da amostra, foi realizado o processo de seleção. As medições da curva de carga dos oito transformadores estão compactadas em um tipo de arquivo usualmente chamado de arquivo@. Para visualização deste arquivo utiliza-se o *software* ELO50. O *software* proporciona a exportação das medições no formato .txt. O arquivo é importado para o Excel, onde é programado uma macro para formatar o arquivo e separar as medições em dias das semanas automaticamente (Figura 9).

Figura 9 – Exemplo do processo de exportação dos dados



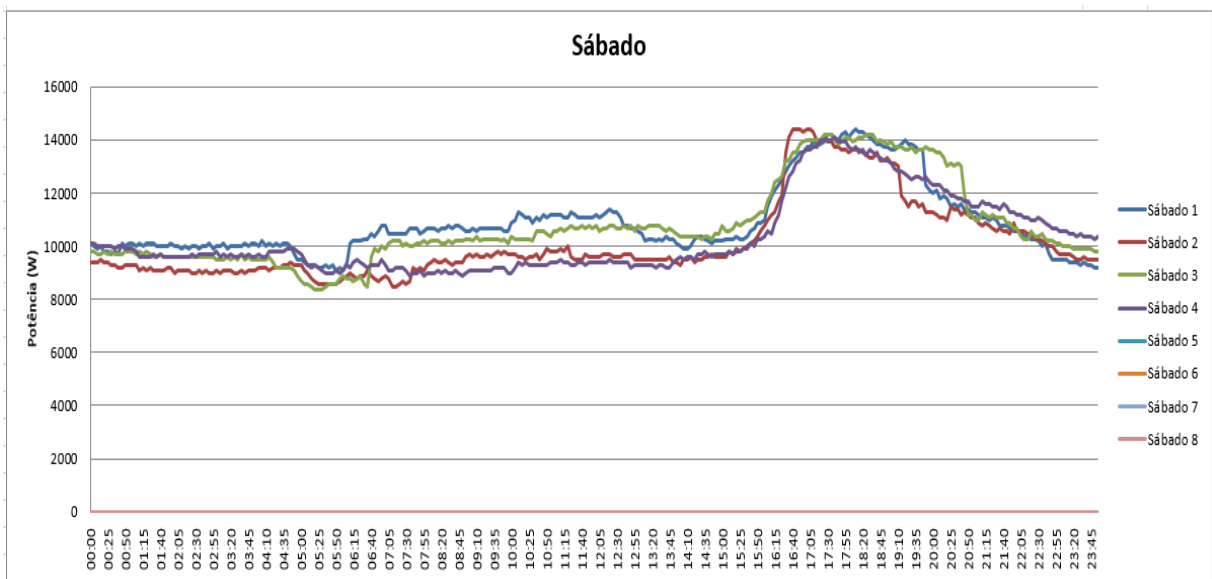
Fonte: Produção do próprio autor.

Com as medições separadas em dias da semana, para cada dia da semana foi plotado um gráfico com todas as suas respectivas medições diárias com o intuito de auxiliar visualmente o analista na seleção das curvas que irão representar o perfil da potência demandada de cada transformador. O resultado final para cada cliente (ou transformação) é sempre a indicação de três curvas características: para sábado, domingo e dia útil, conforme foi ilustrado na Figura 3. Em seguida é necessário integralizar essas medições a cada 15 minutos, se estiverem integralizadas em outro intervalo de tempo.

Finaliza-se esta etapa da caracterização da carga com todos os elementos da amostra possuindo suas três curvas de cargas características com intervalo de integralização de 15 minutos.

O Gráfico 1 mostra o formato dos gráficos realizados para facilitar a seleção da curva característica que representa os sábados de uma unidade consumidora.

Gráfico 1 – Exemplo de gráfico para auxílio da escolha das curvas típicas

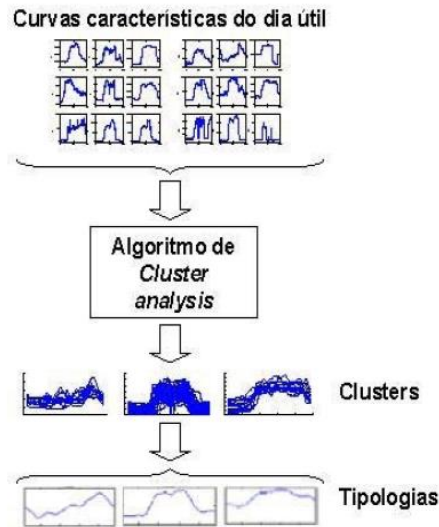


Fonte: Produção do próprio autor.

### 3.3 Metodologia para Classificação de Dados

A identificação das tipologias de um determinado segmento (nível de tensão, classe de consumo ou tipo de rede) envolve a aplicação de alguma técnica de análise de agrupamentos, com a finalidade de dividir o conjunto de curvas características do dia útil em agrupamentos internamente homogêneos (*clusters*), a partir dos quais são extraídas as tipologias, por exemplo, a média das curvas em cada cluster. Este processo é resumido na Figura 10.

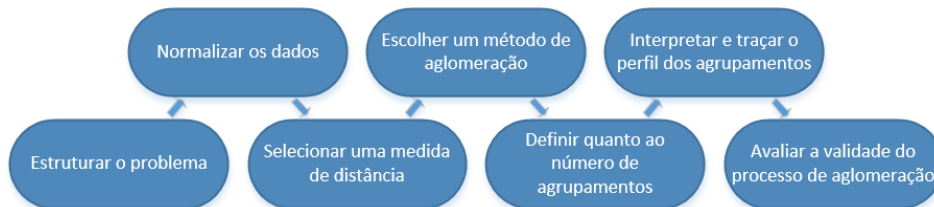
Figura 10 – Processo de Obtenção das Tipologias de Curvas de Cargas



Fonte: ALVES, 2011.

Três questões fundamentais devem ser consideradas na aplicação da análise de agrupamento: primeira, como será medida a similaridade dos dados; segunda, como formar os agrupamentos; e por fim como decidir quantos grupos formar. Um procedimento para se efetuar a análise de agrupamentos é mostrado na Figura 11.

Figura 11 – Etapas para efetuar a análise de agrupamento



Fonte: Produção do próprio autor.

As características de cada objeto são combinadas em uma medida de semelhança, que pode ser de similaridade ou dissimilaridade, calculada para todos os pares de elementos, possibilitando a comparação de qualquer elemento com outro pela medida de similaridade e a associação dos elementos semelhantes por meio da análise de agrupamento. As medidas de distância representam a similaridade, que é representada pela proximidade entre as observações ao longo das variáveis. Normalmente nos métodos de análise de dados são utilizadas três fórmulas para o cálculo da distância entre dois indivíduos: distância euclidiana simples, distância euclidiana ponderada e distância euclidiana quadrática, sendo a distância euclidiana simples a medida de distância mais utilizada (VICINI, 2005).

Considerando-se um conjunto de  $n$  indivíduos (consumidores de energia elétrica), sobre os quais  $p$  variáveis foram observadas (demandas de energia), a distância euclidiana simples é calculada conforme a Equação (8) (GEMIGNANI, 2009).

$$D(i, j) = \sqrt{\sum_{h=1}^{p=N} (X_{ih} - X_{jh})^2} \quad (8)$$

Onde:

$D(i, j)$  - distância entre os indivíduos  $i$  e  $j$ ;

$X_{ih}$  - valor da  $h$ -ésima variável do  $i$ -ésimo indivíduo;

$N$  - número de pontos da curva (24 para intervalos de medição de 1 hora, 96 para intervalos de 15 minutos, etc.).

( $h = 1, 2, \dots, p$  e  $i = 1, 2, \dots, n$ )

Um conceito importante utilizado quando o objetivo é agrupar um indivíduo a um dos grupos definidos na análise de tipologia é o da distância entre um indivíduo e um grupo de indivíduos. Considerando-se grupo  $A$  com  $N_A$  elementos, a medida de distância entre este grupo e um indivíduo  $Y$  mais utilizada nos métodos de análise de tipologia é definida de acordo com a Equação (9) (GEMIGNANI, 2009).

$$d(Y, A) = \sum_{h=1}^{p=96} (X_{yh} - X_h)^2 \quad (9)$$

Onde:

$X_h$  - ( $X_1, X_2, \dots, X_p$ ), chamado de centroide do grupo  $A$ , é a média da variável  $h$ ;

$X_{yh}$  - valor da  $h$ -ésima variável do  $y$ -ésimo indivíduo.

Com base nas medidas de distância, para os procedimentos de aglomeração deve-se optar por um método específico. Os vários métodos de análise de agrupamentos visam resolver o problema de como dividir um conjunto com  $n$  objetos em  $k$  *clusters* mutuamente exclusivos, de tal forma, que os objetos em um mesmo cluster sejam semelhantes entre si, mas diferentes dos objetos pertencentes aos outros *clusters*. Neste trabalho é utilizado o método Ward para a aglomeração hierárquica e o método K-médias para a aglomeração não hierárquica. Detalha-

se, a partir de agora, os procedimentos para realização da análise.

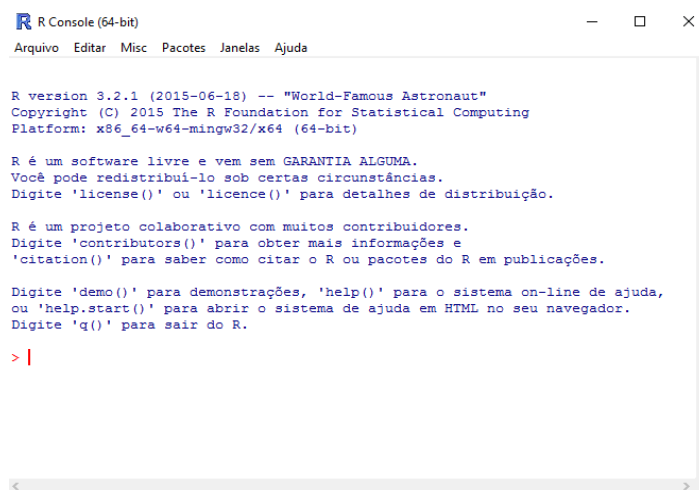
### 3.3.1 Programa R

O uso de *softwares* dedicados à construção de tipologias de curvas de carga ou programas comerciais para a análise de agrupamentos é de grande importância no que se refere a interpretação de resultados. Contudo observa-se que estes apresentam um custo de aquisição relativamente elevado, ou a criação de programas alternativos.

O programa R é um ambiente de *software* livre capaz de realizar cálculos e gerar gráficos, além de interagir com outros programas estatísticos e banco de dados. A tela inicial do R contém o *prompt* em vermelho e algumas informações sobre o sistema e comandos básicos na cor azul (Figura 12).

O R pode ser obtido em <http://www.r-project.org/> onde é disponibilizado em versões de acordo com o sistema operacional UNIX, Windows ou Macintosh. Ele é usado neste trabalho na versão 3.2.1 como ferramenta auxiliar na classificação de dados e na validação do método apresentado.

Figura 12 – Área de trabalho do programa R



```
R Console (64-bit)
Arquivo Editar Misc Pacotes Janelas Ajuda

R version 3.2.1 (2015-06-18) -- "World-Famous Astronaut"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R é um software livre e vem sem GARANTIA ALGUMA.
Você pode redistribuí-lo sob certas circunstâncias.
Digite 'license()' ou 'licence()' para detalhes de distribuição.

R é um projeto colaborativo com muitos contribuidores.
Digite 'contributors()' para obter mais informações e
'citation()' para saber como citar o R ou pacotes do R em publicações.

Digite 'demo()' para demonstrações, 'help()' para o sistema on-line de ajuda,
ou 'help.start()' para abrir o sistema de ajuda em HTML no seu navegador.
Digite 'q()' para sair do R.

> |
```

Fonte: Produção do próprio autor.

Os resultados gerados pelo programa foram obtidos por meio de uma rotina de cálculo desenvolvida pela autora do trabalho e disponível no Apêndice A. É importante salientar que o R não é um programa estatístico, mas que devido a suas rotinas permite a manipulação,

avaliação e interpretação de procedimentos estatísticos aplicados a dados. Ele permite ao usuário acrescentar funcionalidades, tornando-o flexível para gerar análises e receber contribuições de diversos pesquisadores através de pacotes específicos como a inclusão de outros métodos para análise de agrupamentos não considerados neste trabalho.

### 3.3.2 Agrupamento pelo Método de Ward

Para realização da análise de agrupamentos utilizando-se o método de Ward, serão salientados alguns princípios gerais de interpretação dos resultados numéricos e gráficos.

Para encontrar os grupos de variáveis com as mesmas características, que constituem o dendrograma na análise, deve-se importar as amostras de curva de carga do banco de dados do Excel para o programa R, armazenando as informações em amostras [i,j], onde o objeto ‘amostras’ guarda o número da amostra *j* na curva de carga do transformador *i*, conforme Figura 13:

Figura 13 – Curvas de carga em forma de matriz inseridas no programa R

```
> amostras
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15] [,16] [,17] [,18] [,19] [,20]
[1,] 9844.8 9811.2 9710.4 9643.2 9542.4 9676.8 9643.2 9475.2 9408.0 9408.0 9408.0 9475.2 9508.8 9475.2 9475.2 9508.8 9542.4 9643.2 9710.4 9844.8
[2,] 2234.4 2184.0 2167.2 2167.2 2150.4 2133.6 2150.4 2116.8 2116.8 2100.0 2100.0 2083.2 2116.8 2167.2 2167.2 2217.6 2234.4 2284.8 2352.0 2452.8
[3,] 2923.2 2822.4 2856.0 2889.6 2856.0 2788.8 2755.2 2822.4 2856.0 2889.6 2822.4 2755.2 2688.0 2856.0 2856.0 2822.4 2856.0 2956.8 2923.2 3057.6
[4,] 14806.4 14716.8 14515.2 14179.2 14022.4 13708.8 13619.2 13529.6 13507.2 13440.0 13462.4 13552.0 13664.0 13708.8 13596.8 13552.0 13731.2 13731.2 14022.4 14336.0
[5,] 4569.6 7862.4 7694.4 7526.4 7425.6 7190.4 7156.8 7022.4 7123.2 7089.6 6988.8 6820.8 6880.0 6753.6 6753.6 6787.2 6753.6 6921.6 7123.2
[6,] 8841.6 9043.2 9331.2 9792.0 10166.4 10454.4 10195.2 10137.6 10022.4 9504.0 9705.6 10166.4 10080.0 10339.2 10166.4 10108.8 10339.2 10800.0 11001.6 11174.4
[7,] 2671.2 2637.6 2604.0 2536.8 2520.0 2520.0 2536.8 2520.0 2520.0 2536.8 2520.0 2503.2 2452.8 2452.8 2486.4 2486.4 2469.6 2503.2 2587.2 2620.8
[8,] 9441.6 9408.0 9240.0 8971.2 9038.4 8904.0 8870.4 8836.8 8736.0 8702.4 8803.2 8635.2 8635.2 8635.2 8736.0 8769.6 8971.2 9105.6 9307.2 9576.0
      [,21] [,22] [,23] [,24] [,25] [,26] [,27] [,28] [,29] [,30] [,31] [,32] [,33] [,34] [,35] [,36] [,37] [,38] [,39] [,40]
[1,] 9844.8 9508.8 9105.6 9004.8 9004.8 8937.6 9105.6 9172.8 9609.6 10348.8 10281.6 10416.0 10382.4 10449.6 10550.4 10684.8 10684.8 10718.4 10785.6 10785.6
[2,] 2553.6 2671.2 2772.0 2872.8 2973.6 2923.2 2940.0 2990.4 3091.2 3292.8 3427.2 3511.2 3494.4 3544.8 3578.4 3544.8 3561.6 3544.8 3460.8 3360.0
[3,] 3292.8 3360.0 3494.4 3796.8 4065.6 3998.4 4200.0 4435.2 4704.0 5443.2 5880.0 6182.4 6216.0 6350.4 6652.8 6652.8 6652.8 6652.8 6552.0 6595.6
[4,] 15097.6 15523.2 16060.8 16755.2 17360.0 17628.8 18278.4 19174.4 20339.2 20899.2 20966.4 21100.8 21347.2 21772.8 22131.2 22131.2 22377.6 22780.8 22601.6 22848.0
[5,] 7392.0 7761.6 7828.8 8299.2 8668.8 8937.6 8769.6 8232.0 7896.0 8668.8 9878.4 10617.6 11524.8 12364.8 13104.0 13406.4 13944.0 14078.4 14044.8 14078.4
[6,] 11462.4 11664.0 11318.4 11232.0 11664.0 11404.8 10944.0 10857.6 11059.2 12844.8 14112.0 14457.6 13737.6 13276.8 13449.6 13766.4 13651.2 13766.4 13478.4 13536.0
[7,] 2721.6 2839.2 2906.4 2973.6 3024.0 2973.6 3024.0 2956.8 3057.6 3410.4 3528.0 3595.2 3544.8 3628.8 3679.2 3662.4 3544.8 3444.0 3561.6 3595.2
[8,] 9676.8 10214.4 10416.0 10953.6 11390.4 11726.4 11995.2 11961.6 12969.6 14011.2 15019.2 15388.8 15993.6 16800.0 16968.0 17035.2 17136.0 17068.8 17136.0 17169.6
      [,41] [,42] [,43] [,44] [,45] [,46] [,47] [,48] [,49] [,50] [,51] [,52] [,53] [,54] [,55] [,56] [,57] [,58] [,59] [,60]
[1,] 10281.6 9878.4 9576.0 9609.6 9979.2 9744.0 10584.0 10819.2 10886.4 11088.0 10785.6 10852.8 10819.2 11356.8 11424.0 11390.4 11524.8 11760.0 11726.4 11625.6
[2,] 3427.2 3628.8 3729.6 3679.2 3393.6 3158.4 3225.6 3225.6 3376.8 3595.2 3796.8 3763.2 3746.4 3746.4 3712.8 3712.8 3780.0 3729.6 3712.8 3645.6
[3,] 6619.2 6585.6 6686.4 6417.6 5577.6 5376.0 5342.4 5241.6 5443.2 5812.8 6182.4 6316.8 6552.0 6585.6 6552.0 6484.8 6350.4 6686.4 6619.2 6619.2
[4,] 23049.6 23004.8 22825.6 21324.8 19667.2 19286.4 19196.8 20048.0 21750.4 22041.6 22377.6 22803.2 23116.8 23139.2 22892.8 22937.6 23072.0 23430.4 23385.6 23273.6
[5,] 14280.0 14347.2 14448.0 14313.6 13540.8 11659.2 11155.2 10550.4 10684.8 12297.6 12667.2 13372.8 13540.8 13843.2 13910.4 14246.4 14448.0 14448.0 14313.6 14851.2
[6,] 13305.6 12787.2 12556.8 12960.0 11664.0 10339.2 10051.2 9878.4 10512.0 11577.6 12182.4 12616.0 13075.2 13622.4 13564.8 13017.6 12844.8 12787.2 13276.8 13219.2
[7,] 3578.4 3612.0 3511.2 3393.6 3276.0 3124.8 3057.6 3158.4 3326.4 3528.0 3578.4 3578.4 3528.0 3544.8 3612.0 3595.2 3696.0 3746.4 3645.6 3595.2
[8,] 16800.0 16900.8 16967.2 16598.4 16060.8 16027.2 15825.6 15993.6 16128.0 16699.2 16667.2 17035.2 17035.2 17035.2 17035.2 17035.2 17035.2 17035.2 17035.2 17035.2
      [,61] [,62] [,63] [,64] [,65] [,66] [,67] [,68] [,69] [,70] [,71] [,72] [,73] [,74] [,75] [,76] [,77] [,78] [,79] [,80]
[1,] 11894.4 12096.0 12432.0 12465.6 13204.8 14112.0 14448.0 14985.6 15052.8 14380.8 13540.8 13372.8 13104.0 13440.0 13171.2 13171.2 13104.0 12969.6 12969.6 12835.2
[2,] 3494.4 3561.6 3645.6 3628.8 3645.6 3780.0 3763.2 3864.0 3914.4 4183.2 4569.6 4838.4 4989.6 4704.0 4502.4 4384.8 4149.6 4048.8 3914.4 3780.0
[3,] 6552.0 6686.4 6652.8 6787.2 7257.6 7660.8 8097.6 8500.8 8702.4 9744.0 10651.2 11088.0 11188.8 10416.0 9475.2 9072.0 8400.0 7929.6 7862.4 7123.2
[4,] 23497.6 23699.2 23721.6 23632.0 23923.2 24304.0 24147.2 23699.2 24550.4 25939.2 26857.6 27552.0 28246.4 27552.0 26499.2 25356.8 24841.6 24640.0 23497.6 23161.6
[5,] 14380.8 14179.2 14481.6 14582.4 14918.4 15254.4 15724.8 15724.8 15657.6 15993.6 16665.6 17707.2 12969.6 2049.6 1680.0 672.0 0.0 0.0 0.0 0.0
[6,] 13132.8 13248.0 13651.2 14544.0 14745.6 14688.0 15350.4 15753.6 16617.6 17856.0 19929.6 20793.6 19152.0 16617.6 15638.4 15062.4 13939.2 13852.8 13651.2 12528.0
[7,] 3578.4 3511.2 3612.0 3712.8 3662.4 3712.8 3679.2 3712.8 3830.4 3964.8 4267.2 4653.6 4485.6 4216.8 4132.8 4065.6 3948.0 3897.6 3864.0 3763.2
[8,] 17068.8 17539.2 17841.6 17976.0 17908.8 18211.2 18412.8 18782.4 19857.6 21369.6 23822.4 26174.4 27249.6 25972.8 24528.0 23721.6 22579.2 21705.6 21537.6 20160.0
      [,81] [,82] [,83] [,84] [,85] [,86] [,87] [,88] [,89] [,90] [,91] [,92] [,93] [,94] [,95] [,96]
[1,] 12667.2 12432.0 12297.6 12129.6 11961.6 11692.8 11256.0 10953.6 10684.8 10584.0 10382.4 10248.0 10113.6 9979.2 9945.6 9878.4
[2,] 3612.0 3494.4 3360.0 3208.8 3074.4 2990.4 2906.4 2805.6 2805.6 2654.4 2604.0 2503.2 2452.8 2385.6 2352.0 2318.4
```

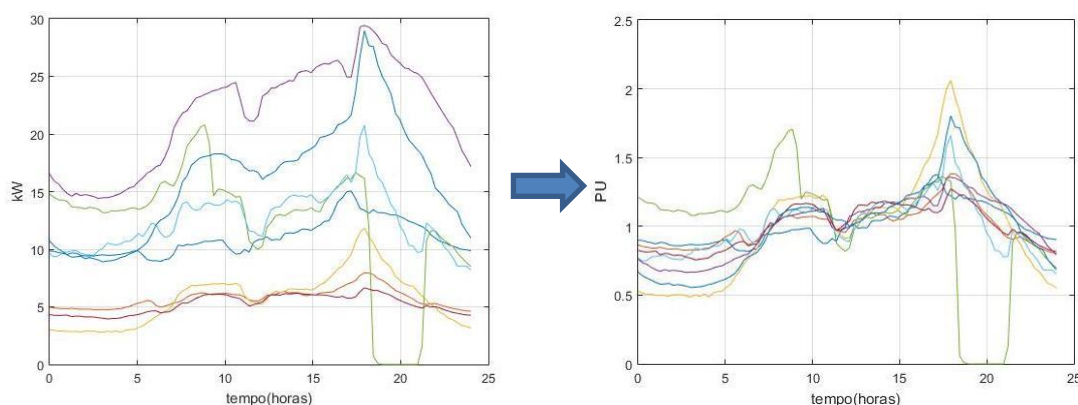
Fonte: Produção do próprio autor.

Dado que o objetivo consiste em identificar grupos de curvas com formas semelhantes, as medições devem ser normalizadas pelos respectivos valores médios, antes de serem processadas pela análise de agrupamentos (PESSANHA et al, 2004). Este procedimento é utilizado a fim de facilitar a comparação e identificação de padrões semelhantes.



A normalização da curva de carga é obtida em p.u. da demanda média. Para tal, calcula-se a demanda média em cada curva de carga e divide-se cada curva pela respectiva média. A partir dessa etapa, um consumidor qualquer pertencente a amostra passa a ser definido pelas suas curvas típicas normalizadas, como é mostrado na Figura 14.

Figura 14 – Normalização das curvas de carga

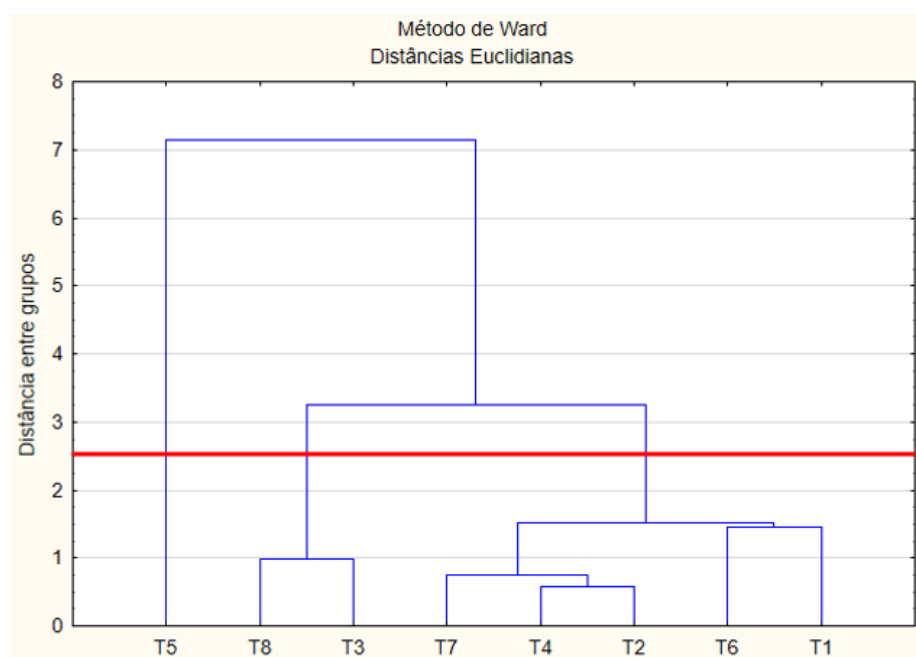


Fonte: Produção do próprio autor.

Em seguida, o método de Ward é executado gerando gráficos úteis na avaliação do número de agrupamentos em um conjunto de dados. No dendrograma, a escala vertical indica o nível de similaridade, e no eixo horizontal são marcados os indivíduos, na ordem em que são agrupados. A sequência de agregações é ilustrada nos dendrogramas das curvas do dia útil, pois são essas curvas que definem os perfis das tipologias.

Três grandes ramos ou agrupamentos emergem do dendrograma ilustrado na Figura 15, sendo que eles são bem distintos, conforme sugerido pelos segmentos verticais que expressam o grau de dissimilaridade entre os agrupamentos. Os grupos foram definidos pelo traçado de uma linha paralela ao eixo horizontal, denominada “Linha Fenon”. O primeiro grupo é formado por T1, T2, T4, T6 e T7; o segundo grupo é formado por T3 e T8, e, o terceiro grupo é formado por T5. Pelo fato de T5 ter formado um grupo isolado, isso significa dizer que esta variável é heterogênea em relação às outras. Considerou-se que o corte deve ser realizado entre as distâncias 1.52 e 3.25. É evidente a presença de três ramos distintos apesar deste não ser o maior segmento vertical.

Figura 15 – Dendograma – dia útil



Fonte: Produção do próprio autor.

Antes de concluir a análise sobre o dendrograma, é pertinente lembrar que o corte no gráfico que determina o número de grupos, geralmente, é realizado em relação às maiores distâncias em que os grupos foram formados, levando-se sempre em consideração os critérios adotados por cada analista.

### 3.3.3 Nuvens dinâmicas – Método das K-Médias (K-Means)

A análise de agrupamentos é efetuada em sequência por meio do método das nuvens dinâmicas (algoritmo *K-Means*), um método não hierárquico. O *K-Means* classifica o conjunto de objetos em um número de categorias (*clusters*) especificado previamente. O número de *clusters* não é conhecido a priori, mas, com base no dendrograma ilustrado na Figura 15, pode-se considerar que três *clusters* é uma boa solução inicial.

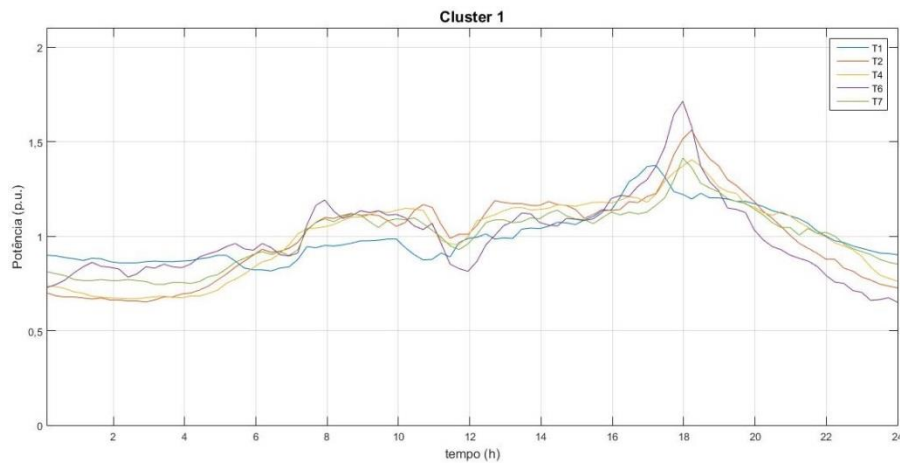
No R, a classificação das curvas normalizadas em três grupos, é realizada com inicializações dos centroides iniciais e 99 iterações do algoritmo *K-Means*.

O método *K-Means* pode ser implementado segundo o algoritmo a seguir, onde os centroides dos *clusters* são obtidos de forma iterativa:

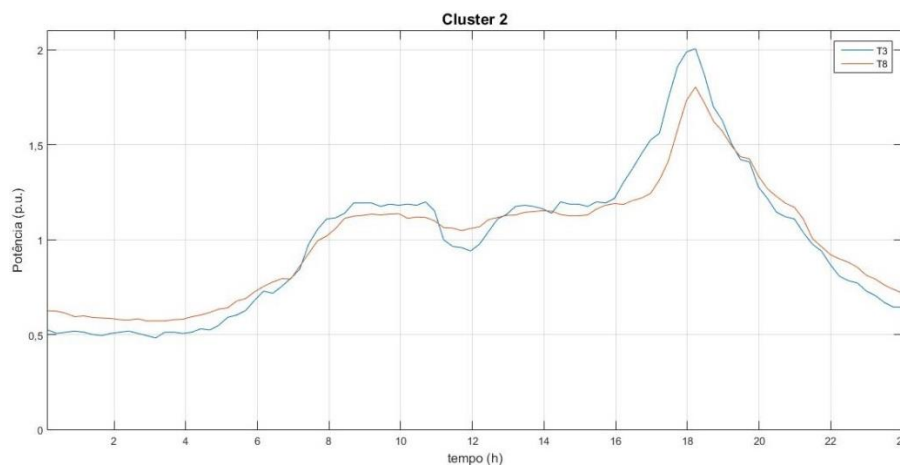
- I. Inicializar os centroides dos *clusters* com  $K$  objetos entre os  $n$  a serem classificados.
- II. Alocar cada objeto ao *cluster* com o centroide mais próximo.
- III. Calcular o valor da soma dos quadrados dos desvios entre as observações e o centroide do cluster. Pare o algoritmo se o valor da função estiver abaixo de uma tolerância pré-especificada, se a melhoria em relação à iteração anterior for desprezível ou se o número máximo de iterações for alcançado.
- IV. Atualizar os centroides dos *clusters* e voltar para o passo 2.

A visualização das curvas classificadas em cada agrupamento é mostrada pela Figura 16.

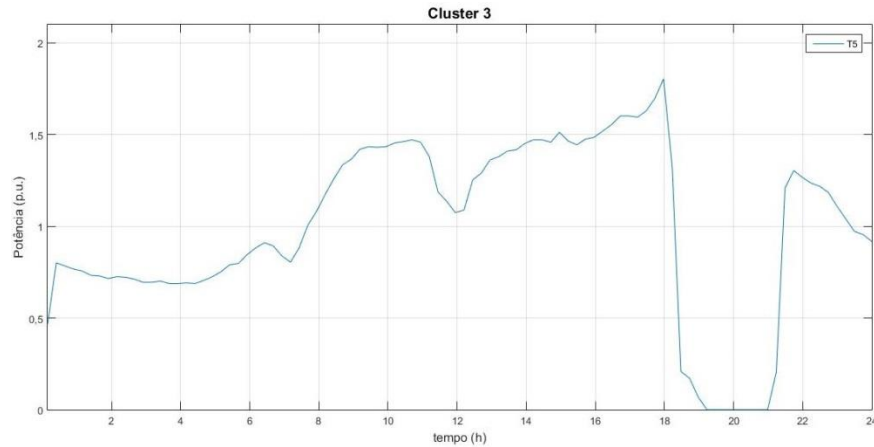
Figura 16 – Composição dos agrupamentos (K-means): (a) Cluster 1, (b) Cluster 2 e (c) Cluster 3.



(a)



(b)



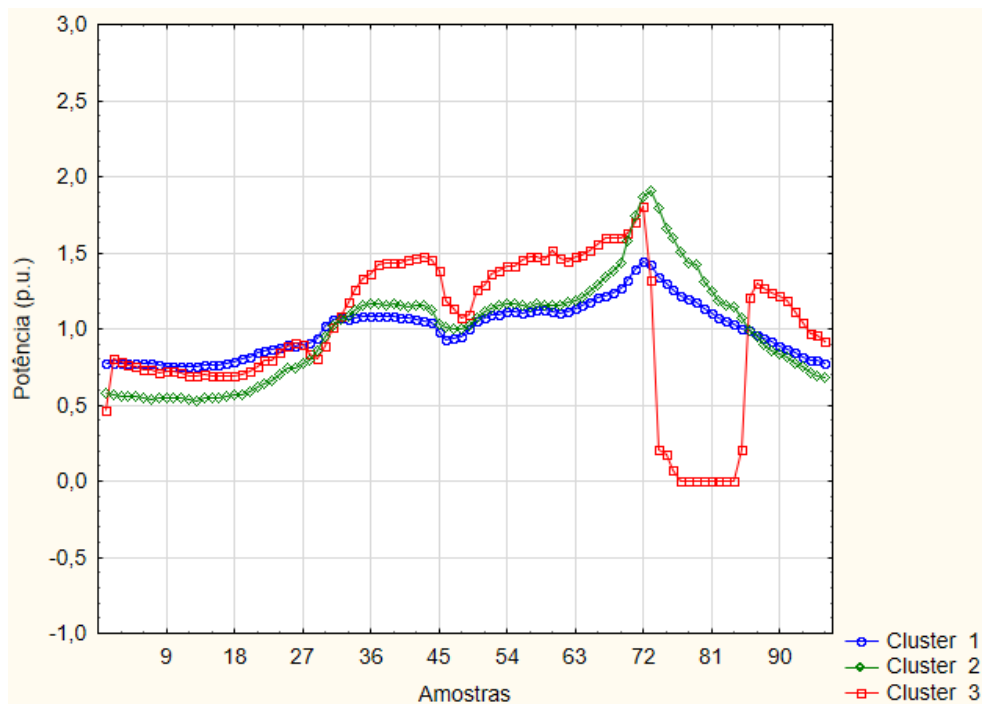
(c)

Fonte: Produção do próprio autor.

### 3.3.4 Resultados e avaliações

As tipologias de redes representam os perfis típicos das solicitações de potência em pontos selecionados da rede de distribuição (transformadores). Os perfis típicos de cada cluster foram definidos conjuntamente na utilização sequencial dos métodos Ward e *K-means*, conforme o Gráfico 2.

Gráfico 2 – Gráfico das médias para cada cluster - dia útil



Fonte: Produção do próprio autor.

É possível avaliar a qualidade dos agrupamentos identificados a partir das medidas do grau de compacidade dos agrupamentos e da separação entre eles, por exemplo. A variabilidade total do conjunto de dados analisados ou inércia total pode ser decomposta em duas parcelas: a inércia entre os agrupamentos (*Between sum of squares* - BSS) e a inércia dentro dos agrupamentos (*within-group sum of squares* - WSS).

Cada parcela da WSS corresponde a um *cluster* e é definida na Equação (4) pela soma dos quadrados dos desvios entre o centroide do *cluster* e os objetos que nele foram classificados (PESSANHA, 2015). O resultado da soma obtida foi de 2,99024.

A Equação (5) apresenta BSS como a soma dos quadrados dos desvios entre os centroides dos agrupamentos e a curva média, ponderados pelos totais de elementos classificados em cada *cluster*. A variabilidade entre os centroides e os agrupamentos foi de 22,42916.

A razão entre BSS e a variabilidade total do conjunto, mostra que com apenas três agrupamentos a maior parte da variabilidade (88,23%) reside entre os agrupamentos conforme o desejado. Ao agrupar as curvas, busca-se uma classificação que tenha grande homogeneidade interna e que seja diferente dos demais agrupamentos. Por outro lado, a inércia dentro dos agrupamentos (WSS) é uma medida problemática, pois diminui inexoravelmente à medida que o número de agrupamentos aumenta (LATTIN et al, 2011).

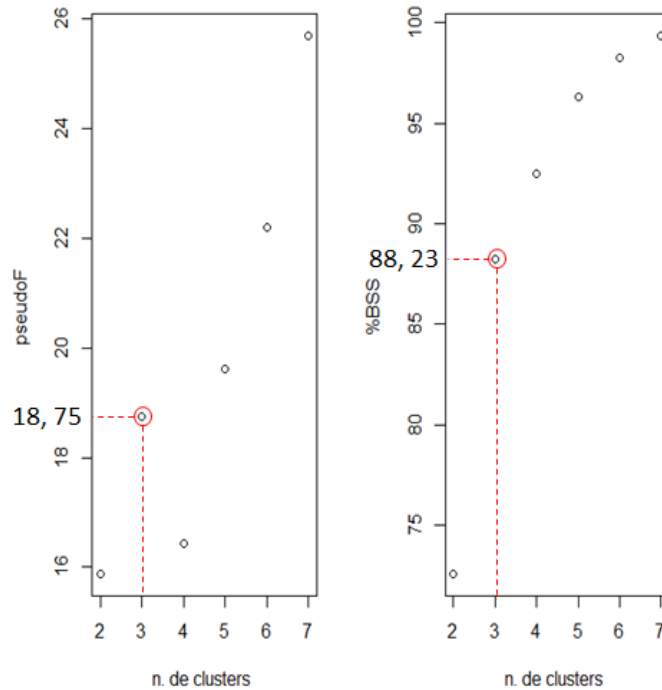
Uma estatística complementar a este fim é o pseudo-F, definido pela razão do quadrado médio entre os clusters  $BSS/(k-1)$  e o quadrado médio dentro dos clusters  $WSS/(N-k)$  conforme a Equação (10). Por meio da pseudo-F pode-se levar em conta os *tradeoffs* entre o aumento da homogeneidade interna dos agrupamentos e o número de *clusters*. Grandes valores de pseudo-F indicam união entre os elementos e aglomerados separados. Em particular, os picos no pseudo-F são indicadores de maior separação dos *clusters* (PESSANHA, 2015).

$$pseudo - F = \frac{BSS/(k - 1)}{WSS/(N - k)} \quad (10)$$

Na Figura 17, é evidente que o máximo do pseudo-F e do BSS se encontra na solução com o maior número de agrupamentos simulados, porém, o número de clusters ainda seria elevado, visto que o conjunto possui 8 elementos. O pico na curva do pseudo-F é observado para a

solução com três agrupamentos e sua inércia entre os agrupamentos mantém-se em 88,23% da inércia total, sendo assim, uma solução satisfatória.

Figura 17 - Estatística Pseudo-F e participação da inércia entre *clusters* (%BSS)



Fonte: Produção do próprio autor.

### 3.3.5 Comparação entre métodos de classificação

Nessa etapa, o procedimento utilizado para classificar as curvas é comparado à distância euclidiana simples, proposta por (FIOROTTI, 2013), do qual as amostras foram obtidas.

Na execução das técnicas, obteve-se 3 formas fortes a partir dos grupos formados. A seguir, a Tabela 1 resume os valores dos somatórios dos quadrados intra-classe (WSS) e inter-classe (BSS), obtida por cada método de classificação, comparando o desempenho das técnicas.

Tabela 1 – Comparação de resultados

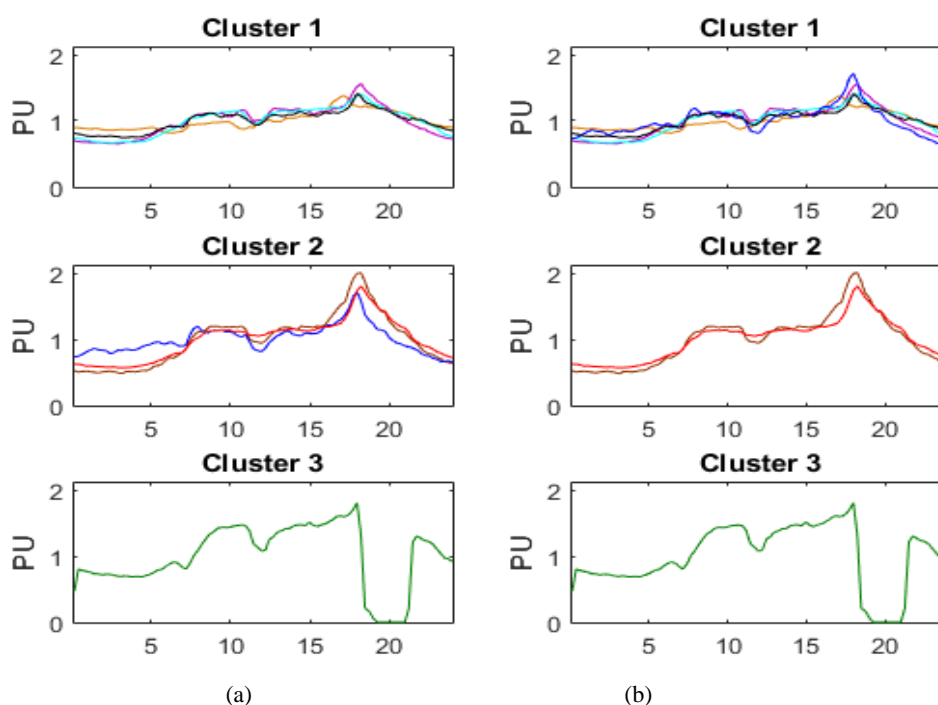
Métodos de classificação	BSS	WSS
Distância Euclidiana Simples	20.95755	4.46186
Ward e K-means	22.42916	2.99024

Fonte: Produção do próprio autor.

Neste contexto, uma boa aplicação de *cluster* requer que o método usado forneça um

agrupamento com alta similaridade dentro dos grupos e alta dissimilaridade entre os grupos. Portanto, os resultados da análise de validação apresentados na Tabela 1 indicam que o melhor desempenho foi alcançado pelo método utilizado neste trabalho, a saber, método de Ward e *K-means*. A proximidade nos valores de ambos os métodos é traduzida pela grande semelhança das tipologias obtidas, conforme apresenta a figura 18.

Figura 18 – Comparação de tipologias em p.u. da média: (a) método da distância euclidiana simples e (b) método Ward e *K-means*



Fonte: Produção do próprio autor.

Pode-se considerar que os dois métodos identificam corretamente as tipologias, apesar das composições dos *clusters* serem ligeiramente distintas. Eles apresentam um grau satisfatório de compacidade nos agrupamentos e de separação entre eles, observados na avaliação estatística. Os resultados evidenciam que outras técnicas podem apresentar resultados superiores ou compatíveis aos obtidos pelos tradicionais.

### 3.4 Ajuste da Tipologia ao Mercado Anual de Energia

Assim como as tipologias de clientes são definidas para cada classe de consumo e representam a diversidade do comportamento dos consumidores da classe, as tipologias de redes representam os perfis típicos das solicitações de potência em pontos selecionados da rede de distribuição, alimentadores e transformadores (PESSANHA, 2006).

Nesta última etapa do processo de caracterização da carga, o perfil de consumo anual do estrato é determinado a partir do perfil de consumo das amostras. Cada *cluster* obtido na etapa de classificação é transformado em uma única forma chamada forma forte por meio da soma de cada uma das suas curvas de carga, enquanto a junção das formas fortes irá determinar o perfil de consumo do estrato. Essas formas representam a tendência de um cliente, que está inserido em um estrato do mercado, possuir perfil de demanda similar a outro cliente nas mesmas condições.

Para realizar o ajuste da tipologia ao mercado, é necessário obter alguns parâmetros de entrada, tais como: número de dias úteis, sábados e domingos do ano em que as medições foram realizadas, o mercado anual de energia e os horários adotados de ponta e fora ponta da concessionária de energia em questão. O Quadro 5 mostra os parâmetros de entrada citados, necessários para realizar os cálculos de ajuste da tipologia ao mercado anual de energia.

Quadro 5 – Parâmetros de entrada para ajuste da tipologia ao mercado

Parâmetros	
Ano de análise das amostras	2016
Mercado anual de energia (MWh)	8.725.935,84
Horário de ponta (horas por dia)	3
Horário fora de ponta (horas por dia)	21

Fonte: Produção do próprio autor.

O valor do mercado anual de energia utilizado é proveniente da Resolução Homologatória n° 2.018/16 (ANEEL, 2016), e o posto tarifário ponta, composto por três horas diárias consecutivas, é definido pela Resolução Normativa n° 414/10 com exceção feita aos sábados, domingos e alguns feriados nacionais (ANEEL, 2010). A Tabela 2, por sua vez, apresenta todos os dias do ano de análise das amostras divididas em segmentos específicos.

Tabela 2 – Parâmetros de entrada para ajuste da tipologia ao mercado

Parâmetros	Total
Número de dias úteis (Nu)	250
Número de sábados (Ns)	52
Número de domingos + feriados (Nd)	63
<b>Total</b>	<b>365</b>

Fonte: Produção do próprio autor.

Inicialmente, as curvas do dia útil (U), do sábado (S) e do domingo (D) são somadas separadamente em cada *cluster* obtendo três curvas agregadas (perfis típicos) para cada *cluster* *i*:



$$U_h^i, S_h^i, D_h^i, \forall h = 1, 24 \text{ horas.} \quad (11)$$

Totalizamos as curvas horárias do dia útil, sábado e domingo em cada *cluster i*, conforme demonstra a Equação (12).

$$E_u^i = \sum_{h=1}^{24} U_h^i \quad E_s^i = \sum_{h=1}^{24} S_h^i \quad E_d^i = \sum_{h=1}^{24} D_h^i \quad (12)$$

A Tabela 3 mostra os resultados dos somatórios das curvas típicas de cada cluster.

Tabela 3 – Perfis de curvas típicas de cada cluster

HORA	Tipologias em MW								
	TIPO 1			TIPO 2			TIPO 3		
	Dia útil	Sábado	Domingo	Dia útil	Sábado	Domingo	Dia útil	Sábado	Domingo
1	38.359,4	38.412,6	36.754,8	12.138,0	13.910,4	13.398,0	6.913,2	8.257,2	6.888,0
2	38.204,8	37.055,6	35.768,8	11.718,0	13.154,4	12.583,2	7.198,8	7.761,6	6.526,8
3	37.384,8	37.006,4	35.074,4	11.550,0	12.734,4	12.238,8	7.005,6	7.408,8	6.090,0
4	37.932,8	36.695,6	34.331,0	11.499,6	12.616,8	11.919,6	6.804,0	6.938,4	5.737,2
5	39.345,4	38.217,4	35.089,4	12.188,4	12.986,4	12.138,0	6.896,4	6.745,2	5.686,8
6	42.222,0	39.798,0	35.564,6	13.801,2	13.910,4	12.566,4	7.820,4	6.930,0	5.586,0
7	44.334,6	40.817,2	34.961,0	15.943,2	15.598,8	14.246,4	8.652,0	6.426,0	4.964,4
8	50.837,2	45.709,4	35.682,4	19.899,6	18.698,4	19.656,0	9.265,2	6.375,6	3.645,6
9	53.089,4	48.226,4	37.082,2	23.167,2	21.478,8	21.495,6	12.600,0	8.534,4	4.939,2
10	54.021,8	48.809,0	37.579,2	23.730,0	21.882,0	21.487,2	14.036,4	9.290,4	5.149,2
11	52.430,0	48.088,4	37.494,0	23.368,8	21.159,6	21.042,0	14.347,2	9.634,8	5.644,8
12	46.719,4	45.116,6	37.026,4	21.361,2	20.269,2	20.244,0	11.726,4	8.979,6	5.577,6
13	52.054,2	45.590,0	35.929,4	22.621,2	20.470,8	17.304,0	12.255,6	8.853,6	5.451,6
14	54.888,8	44.990,8	35.866,4	23.814,0	20.487,6	14.985,6	13.885,2	9.122,4	5.468,4
15	55.369,4	46.393,2	35.599,8	23.688,0	20.748,0	14.691,6	14.515,2	9.391,2	5.636,4
16	56.689,8	46.285,0	36.500,2	24.276,0	21.898,8	15.136,8	14.406,0	9.584,4	5.938,8
17	60.795,4	49.360,6	40.392,8	26.208,0	23.839,2	16.371,6	15.405,6	9.828,0	6.854,4
18	67.666,2	62.395,6	50.777,6	32.852,4	29.744,4	23.528,4	16.506,0	13.616,4	11.281,2
19	65.623,2	63.424,6	53.341,8	35.406,0	31.852,8	27.619,2	4.342,8	8.408,4	6.207,6
20	58.339,0	56.473,8	47.888,4	29.324,4	26.720,4	23.679,6	0,0	4.141,2	1.822,8
21	53.170,4	51.545,6	44.572,6	24.746,4	23.410,8	20.781,6	0,0	2.805,6	529,2
22	48.807,8	47.344,6	43.108,6	20.420,4	20.319,6	16.749,6	9.777,6	6.098,4	8.257,2
23	43.882,8	44.043,4	38.971,4	17.320,8	17.824,8	14.288,4	11.650,8	10.088,4	8.652,0
24	39.195,0	39.825,2	35.895,4	15.078,0	15.498,0	12.742,8	9.534,0	9.198,0	6.980,4
<b>Total</b>	<b>1.191.363,6</b>	<b>1.101.625,0</b>	<b>931.252,6</b>	<b>496.120,8</b>	<b>471.214,8</b>	<b>410.894,4</b>	<b>235.544,4</b>	<b>194.418,0</b>	<b>139.515,6</b>

Fonte: Produção do próprio autor.

As curvas do dia útil definem os perfis das tipologias, enquanto as curvas de sábado e domingo são usadas apenas no ajuste dos perfis típicos ao montante de energia anual do segmento que eles representam. Essas curvas servem para calcular fatores de ponderação do sábado ( $P_s^i$ ) e do domingo ( $P_d^i$ ) de cada *cluster i*, de acordo com a Equação (13):

$$P_s^i = \frac{E_s^i}{E_u^i} \quad P_d^i = \frac{E_d^i}{E_u^i} \quad (13)$$

Em seguida, é feito o cálculo da energia anual da curva agregada do *cluster i*, descrito na Equação (14).

$$C_i^{anual} = E_u^i \cdot 24(N_u + N_s \cdot P_s^i + N_d \cdot P_d^i) \quad (14)$$

Com o resultado anterior, é possível calcular a participação do *cluster i* no mercado total do segmento que ele representa, em conformidade com a Equação (15).

$$\alpha_i = \frac{C_i^{anual}}{\sum_{i=1}^k C_i^{anual}} \quad (15)$$

Multiplicando a participação do *cluster i* pelo valor do mercado anual de energia, obtemos a parcela de consumo do mercado correspondente ao *cluster i*, expressa pela Equação (16).

$$ET_i = \alpha_i \cdot M \quad (16)$$

Onde  $M$  é o mercado anual em MWh do segmento que ele representa.

Consequentemente, o cálculo da demanda média da curva agregada do *cluster i* pode ser realizado conforme a Equação (17) (PESSANHA, 2006).

$$D_i^{média} = \frac{ET_i}{24(N_u + N_s P_s^i + N_d P_d^i)} \quad (17)$$

Os resultados dos passos anteriores estão relacionados na Tabela 4. Os resultados obtidos são extrapolados no estudo das amostras para o mercado anual de energia elétrica do mesmo segmento de mercado com base nesses valores.

Tabela 4 – Valores úteis para o ajuste de tipologia ao mercado

	TIPO 1	TIPO 2	TIPO 3
ponderação sábado	0,9247	0,94980	0,8254
ponderação domingo	0,7817	0,8282	0,5923
energia anual (MWh)	9.931.063.531	4.186.073.203	1.866.847.651
mercado (%)	62,13	26,19	11,68
mercado (MWh)	5.421.541	2.285.250	1.019.145
demanda média (MW)	650,39	270,84	128,59

Fonte: Produção do próprio autor.

Finalmente, ajustamos a tipologia  $i$  ao mercado, multiplicando a demanda média pela demanda horária em p.u. da média, de acordo com a Equação (18) (PESSANHA, 2006).

$$T_h^i = D_i^{média} \cdot \frac{U_h^i}{\sum_{h=1}^{24} U_h^i / 24} \quad (18)$$

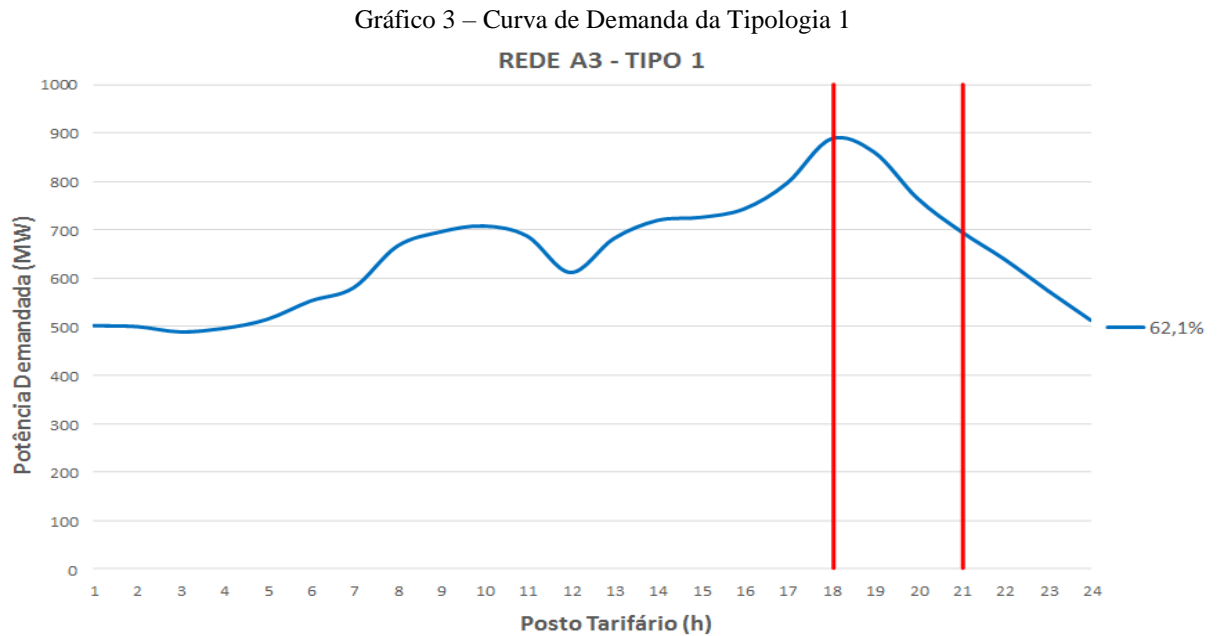
A Tabela 5 mostra os valores da potência demandada integralizada de hora em hora e normalizadas em p.u. da média das tipologias determinadas a partir do ajuste da tipologia ao mercado.

Tabela 5 – Perfil de demanda em p.u. das tipologias

HORA	TIPO 1	TIPO 2	TIPO 3	AGREGADO
1	0,77	0,59	0,70	0,72
2	0,77	0,57	0,73	0,71
3	0,75	0,56	0,71	0,70
4	0,76	0,56	0,69	0,70
5	0,79	0,59	0,70	0,73
6	0,85	0,67	0,80	0,80
7	0,89	0,77	0,88	0,86
8	1,02	0,96	0,94	1,00
9	1,07	1,12	1,28	1,11
10	1,09	1,15	1,43	1,15
11	1,06	1,13	1,46	1,13
12	0,94	1,03	1,19	1,00
13	1,05	1,09	1,25	1,08
14	1,11	1,15	1,41	1,16
15	1,12	1,15	1,48	1,17
16	1,14	1,17	1,47	1,19
17	1,22	1,27	1,57	1,28
18	1,36	1,59	1,68	1,46
19	1,32	1,71	0,44	1,32
20	1,18	1,42	0,00	1,09
21	1,07	1,20	0,00	0,97
22	0,98	0,99	1,00	0,99
23	0,88	0,84	1,19	0,91
24	0,79	0,73	0,97	0,80
<b>Consumo (%)</b>	<b>62,13</b>	<b>26,19</b>	<b>11,68</b>	<b>100,00</b>

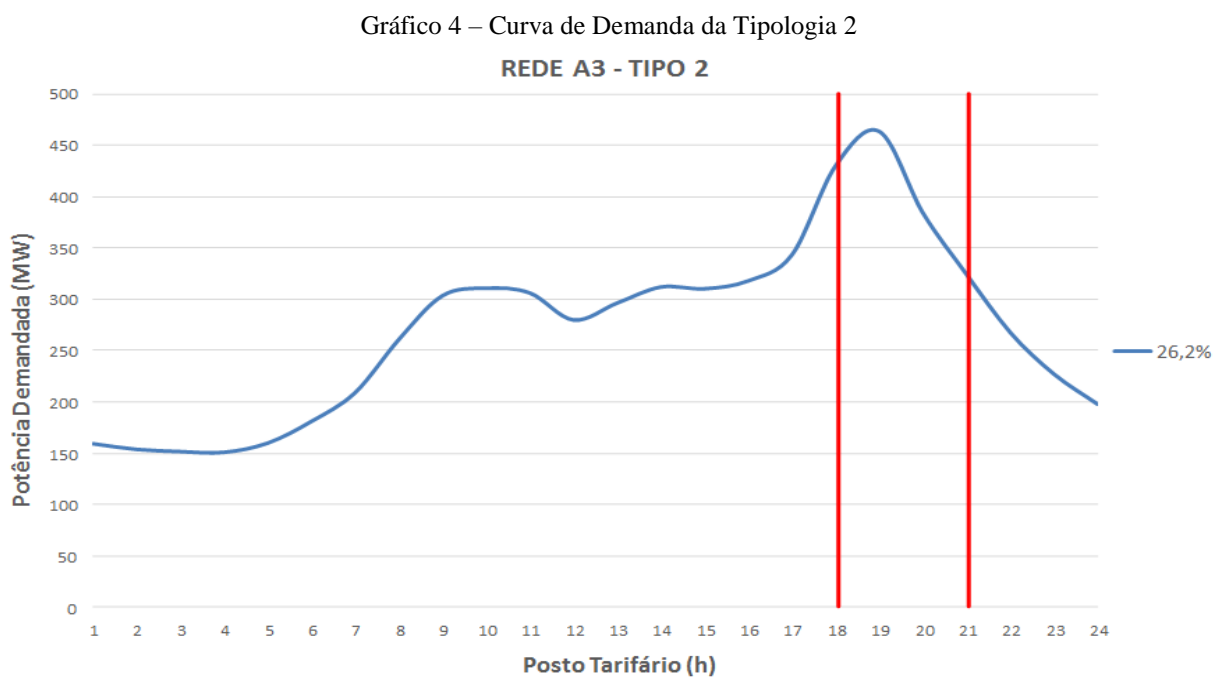
Fonte: Produção do próprio autor.

Os agregados das tipologias de carga já ajustados ao mercado das três formas fortes obtidas dos transformadores de 138 kV / 69 kV (Redes A3) da região estudada são apresentados nos Gráficos 3, 4 e 5.



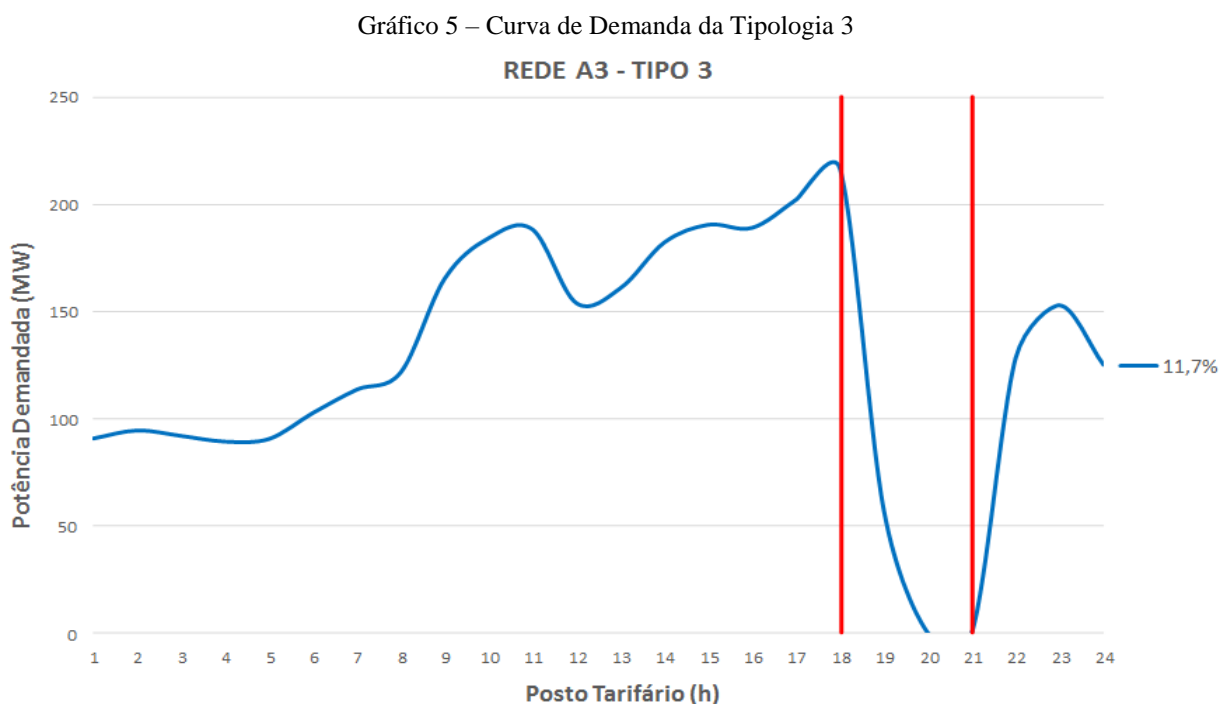
Fonte: Produção do próprio autor.

A Tipologia 1 é formada pelos transformadores 1, 2, 4, 6 e 7, responsáveis pela maior parte (62.1%) do consumo de energia que são transferidos das Redes A2 (nível de tensão de 138 kV). Além disso, possui fator de carga de 73,36%, o maior das tipologias estudadas.



Fonte: Produção do próprio autor.

Um decréscimo quase linear é verificado das 19h às 24h na tipologia 2, formada pelos transformadores 3 e 8. Presume-se que essa queda na demanda de potência seja causada pelo fim do horário comercial (por volta das 18h). Esse perfil é responsável por 26,2% do consumo de energia com de fator de carga de 58,38%.

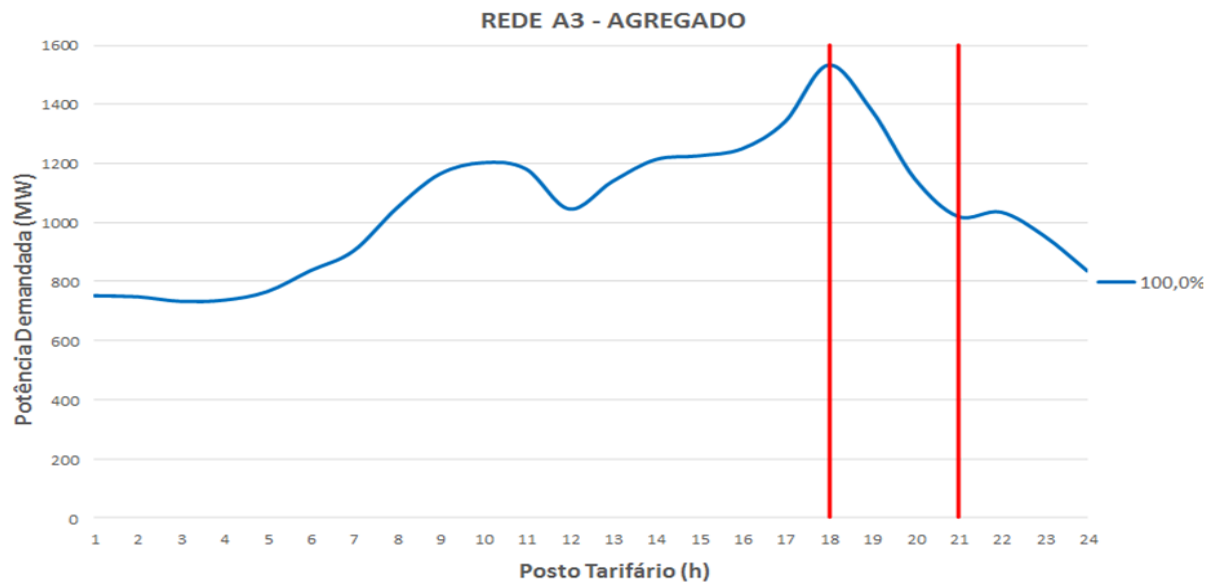


Fonte: Produção do próprio autor.

A forma mais distinta dentre as tipologias é a do perfil 3, composta apenas pelo transformador 5. A potência demandada dessa tipologia no período de ponta diminui drasticamente, sendo responsável em sua totalidade por 11,7% do mercado anual de energia com fator de carga de 59,46%. Provavelmente, isto se deve a algum grande cliente que não consome energia no horário de ponta para minimizar seus gastos com energia elétrica ou que alguma geração é despachada neste horário.

A tipologia total dos transformadores de 138 kV / 69 kV (Redes A3) da região estudada é obtida através da composição de todas as três tipologias mostradas anteriormente, levando-se em conta o peso que cada uma exerce sobre a demanda de potência que é solicitada das redes A2. O Gráfico 6 mostra a curva de demanda total das redes A3 da região estudada.

Gráfico 6 – Curva de demanda total das redes A3

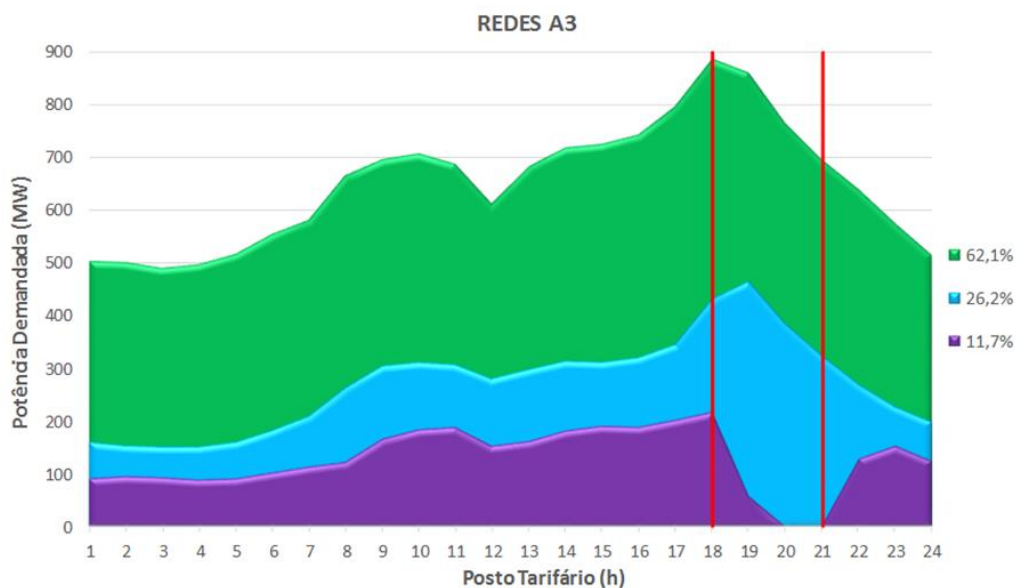


Fonte: Produção do próprio autor.

A curva de demanda total se assemelha à curva da tipologia 2 e à curva de carga de consumidores comerciais que é caracterizada por uma demanda durante o horário comercial (em geral, das 8h às 18h) com um leve declínio no horário de almoço e um pico de demanda às 18h.

No Gráfico 7, nota-se a sobreposição das curvas geradas e a influência de cada uma delas no mercado de energia.

Gráfico 7 – Curvas de demanda sobrepostas das 3 tipologias obtidas



Fonte: Produção do próprio autor.

Para finalizar o processo da caracterização da carga, foi elaborada uma tabela com a estratificação do mercado de energia das Redes A3 estudadas, apresentada na Figura 19. A estratificação do mercado de energia é necessária em alguns estudos técnicos e comerciais das empresas concessionárias de energia, em especial, para servir como base aos cálculos de custos marginais.

Figura 19 – Estratificação do mercado de energia das redes A3 estudadas

MERCADO ESTRATIFICADO - REDE A3						
Consumidor Tipo	Consumo (MWh)			Demanda (MW)		
	Ponta	Fora de Ponta	Total	Ponta	Fora de Ponta	Máxima
1	627.679,71	4.793.861,07	5.421.540,78	886,56	796,54	886,56
2	319.632,91	1.965.617,45	2.285.250,37	463,89	343,38	463,89
3	68.290,34	950.854,35	1.019.144,69	216,26	201,84	216,26
4						
5						
6						
<b>Total</b>	<b>1.015.602,97</b>	<b>7.710.332,87</b>	<b>8.725.935,84</b>	<b>1.566,72</b>	<b>1.341,76</b>	<b>1.566,72</b>

Fonte: Produção do próprio autor.

Estas informações devem refletir a condição de operação do sistema. As demandas máximas identificadas, por sua vez, irão definir as demandas teóricas responsáveis pelas ampliações e custos dos sistemas de distribuição.

De forma sintetizada, pode-se gerar uma planilha eletrônica (Figura 20) com as informações mais relevantes e com as tipologias para os dias úteis, sábado e domingo, ajustadas ao mercado e prontas para serem anexadas às propostas tarifárias da distribuidora.

Figura 20 – Planilha com tipologias ajustadas ao mercado

AMANDA DIAS - Tabela Resumo														
INTERVALO	HORA	TIPO 1			TIPO 2			TIPO 3			AGREGADO			
		útil	sábado	domingo	útil	sábado	domingo	útil	sábado	domingo	útil	sábado	domingo	
00:00-01:00	1	38.359,4	38.412,6	36.754,8	12.138,0	13.910,4	13.398,0	6.913,2	8.257,2	6.888,0	57.410,6	60.580,2	57.040,8	
01:00-02:00	2	38.204,8	37.055,6	35.768,8	11.718,0	13.154,4	12.583,2	7.198,8	7.761,6	6.526,8	57.121,6	57.971,6	54.878,8	
02:00-03:00	3	37.384,8	37.006,4	35.074,4	11.550,0	12.734,4	12.238,8	7.005,6	7.408,8	6.090,0	55.940,4	57.149,6	53.403,2	
03:00-04:00	4	37.932,8	36.695,6	34.331,0	11.499,6	12.616,8	11.919,6	6.804,0	6.938,4	5.737,2	56.236,4	56.250,8	51.987,8	
04:00-05:00	5	39.345,4	38.217,4	35.089,4	12.188,4	12.986,4	12.138,0	6.896,4	6.745,2	5.686,8	58.430,2	57.949,0	52.914,2	
05:00-06:00	6	42.222,0	39.798,0	35.564,6	13.801,2	13.910,4	12.566,4	7.820,4	6.930,0	5.586,0	63.843,6	60.638,4	53.717,0	
06:00-07:00	7	44.334,6	40.817,2	34.961,0	15.943,2	15.598,8	14.246,4	8.652,0	6.426,0	4.964,4	68.929,8	62.842,0	54.171,8	
07:00-08:00	8	50.837,2	45.709,4	35.682,4	19.899,6	18.698,4	19.656,0	9.265,2	6.375,6	3.645,6	80.002,0	70.783,4	58.984,0	
08:00-09:00	9	53.089,4	48.226,4	37.082,2	23.167,2	21.478,8	21.495,6	12.600,0	8.534,4	4.939,2	88.856,6	78.239,6	63.517,0	
09:00-10:00	10	54.021,8	48.809,0	37.579,2	23.730,0	21.882,0	21.487,2	14.036,4	9.290,4	5.149,2	91.788,2	79.981,4	64.215,6	
10:00-11:00	11	52.430,0	48.088,4	37.494,0	23.368,8	21.159,6	21.042,0	14.347,2	9.634,8	5.644,8	90.146,0	78.882,8	64.180,8	
11:00-12:00	12	46.719,4	45.116,6	37.026,4	21.361,2	20.269,2	20.244,0	11.726,4	8.979,6	5.577,6	79.807,0	74.365,4	62.848,0	
12:00-13:00	13	52.054,2	45.590,0	35.929,4	22.621,2	20.470,8	17.304,0	12.255,6	8.853,6	5.451,6	86.931,0	74.914,4	58.685,0	
13:00-14:00	14	54.888,8	44.990,8	35.866,4	23.814,0	20.487,6	14.985,6	13.885,2	9.122,4	5.468,4	92.588,0	74.600,8	56.320,4	
14:00-15:00	15	55.369,4	46.393,2	35.599,8	23.688,0	20.748,0	14.691,6	14.515,2	9.391,2	5.636,4	93.572,6	76.532,4	55.927,8	
15:00-16:00	16	56.689,8	46.285,0	36.500,2	24.276,0	21.898,8	15.136,8	14.406,0	9.584,4	5.938,8	95.371,8	77.768,2	57.575,8	
16:00-17:00	17	60.795,4	49.360,6	40.392,8	26.208,0	23.839,2	16.371,6	15.405,6	9.828,0	6.854,4	102.409,0	83.027,8	63.618,8	
17:00-18:00	18	67.666,2	62.395,6	50.777,6	32.852,4	29.744,4	23.528,4	16.506,0	13.616,4	11.281,2	117.024,6	105.756,4	85.587,2	
18:00-19:00	19	65.623,2	63.424,6	53.341,8	35.406,0	31.852,8	27.619,2	4.342,8	8.408,4	6.207,6	105.372,0	103.685,8	87.168,6	
19:00-20:00	20	58.339,0	56.473,8	47.888,4	29.324,4	26.720,4	23.679,6	0,0	4.141,2	1.822,8	87.663,4	87.335,4	73.390,8	
20:00-21:00	21	53.170,4	51.545,6	44.572,6	24.746,4	23.410,8	20.781,6	0,0	2.805,6	529,2	77.916,8	77.762,0	65.883,4	
21:00-22:00	22	48.807,8	47.344,6	43.108,6	20.420,4	20.319,6	16.749,6	9.777,6	6.098,4	8.257,2	79.005,8	73.762,6	68.115,4	
22:00-23:00	23	43.882,8	44.043,4	38.971,4	17.320,8	17.824,8	14.288,4	11.650,8	10.088,4	8.652,0	72.854,4	71.956,6	61.911,8	
23:00-24:00	24	39.195,0	39.825,2	35.895,4	15.078,0	15.498,0	12.742,8	9.534,0	9.198,0	6.980,4	63.807,0	64.521,2	55.618,6	
<b>% mercado</b>			62,13		26,19				11,68		100			
<b>mercado MWh</b>			5.421.541		2.285.250			1.019.145			8.725.936			
<b>d demanda média MW</b>			650,39		270,84			128,59			1.049,815			
<b>ponderação sábado</b>			0,9247		0,94980			0,8254			0,9190			
<b>ponderação domingo</b>			0,7817		0,8282			0,5923			0,7705			
<b>fator de carga %</b>			73,36		58,38			59,46			68,47			
<b>d demanda máx MW Ponta</b>			886,56		463,89			216,26			1.533,26			
<b>d demanda máx MW Fora_Ponta</b>			796,54		343,38			201,84			1.341,76			

Fonte: Produção do próprio autor.



## 4 CONCLUSÕES

A partir dos registros contidos na recuperação de medições, uma base de dados foi obtida contendo os perfis horários de demanda de todos os transformadores amostrados, cujas formas descrevem o comportamento típico da demanda por eletricidade ao longo do dia.

Na seleção manual de curvas típicas, nota-se que bons resultados dependem dos critérios adotados por cada profissional para selecionar as curvas mais representativas. As ferramentas de criação e exibição de gráficos a partir de amostras são de grande utilidade nessa etapa da caracterização.

A construção de tipologias de curvas de carga, por sua vez, é uma etapa crítica nos processos de revisão das tarifas de uso dos sistemas de distribuição de energia elétrica (TUSD), pois envolve desde a obtenção de curvas selecionadas à identificação dos perfis latentes ou tipologias, que foi realizada por meio de técnicas de análise de agrupamentos. Neste trabalho são descritas as técnicas de *cluster analysis* e apresentadas formas de implementá-las computacionalmente em ambiente R. Adicionalmente são discutidas as medidas de validação que permitem determinar o número adequado de agrupamentos.

Na aplicação do método Ward foram encontrados três agrupamentos, assim como no método *K-Means*. A clusterização teve alta concordância, pois 100% dos transformadores foram agrupados nos mesmos agrupamentos em ambos os métodos, evidenciando a eficiência da robustez dos agrupamentos formados pelos dois métodos.

No método de Ward a qualidade dos agrupamentos pode ser percebida por meio do dendrograma, enquanto no método *K-Means* as inércias intra (WSS) e inter (BSS) *cluster* fornecem medidas do grau de compacidade dos agrupamentos e da separação entre eles respectivamente.

Independentemente do método de análise de agrupamentos utilizado, os *clusters* resultantes devem ser compactos, bem separados e estáveis. Na utilização dos dois métodos conjuntamente para a formação dos grupos, o método Ward define com mais eficiência a quantidade de agrupamentos que devem ser utilizados, enquanto que o método *K-Means* classifica de forma mais adequada os transformadores dentro dos agrupamentos.

Além dos métodos de classificação tradicionais abordados, novos métodos podem ser utilizados na agregação de curvas de carga. Baseada em inteligência artificial, a metodologia de classificação por redes neurais, por exemplo, vem apresentando-se como uma metodologia bastante promissora, estimulando futuras implementações.

Outro tópico passível de pesquisa é a medida de validação. Medidas mais gerais permitem avaliar a qualidade dos agrupamentos identificados, tais como: medidas de conectividade (relacionadas com o grau de vizinhança entre objetos classificados em um mesmo *cluster*), silhueta (homogeneidade interna), índice Dunn (separação entre os agrupamentos) e medidas da estabilidade dos agrupamentos.

Estes fatos sugerem a continuidade da aplicação de clusterização de dados nas concessionárias de energia, utilizando por exemplo, uma amostra maior de curvas de carga e diferentes números de classes.

## REFERÊNCIAS BIBLIOGRÁFICAS

ALVES, E. R. **Mineração de dados baseada em inteligência computacional: uma aplicação à determinação da tipologia de curvas de cargas.** 2011. Dissertação (Mestrado em Engenharia Elétrica) – Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal do Pará, Belém, 2011. Disponível em: <[http://repositorio.ufpa.br/jspui/bitstream/2011/2834/1/Dissertacao\\_MineraoDadosBaseada.pdf](http://repositorio.ufpa.br/jspui/bitstream/2011/2834/1/Dissertacao_MineraoDadosBaseada.pdf)>. Acesso em: 14 jun. 2016.

AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA. **Metodologia do Cálculo das Tarifas de Uso dos Sistemas de Distribuição:** Revisão das Tarifas Publicadas na Resolução ANEEL nº 286/99. Brasília, 2000. Disponível em: <[http://www2.aneel.gov.br/aplicacoes/Audiencia\\_Publica/audiencia\\_proton/ap003/Metodologia%20TUSD.pdf](http://www2.aneel.gov.br/aplicacoes/Audiencia_Publica/audiencia_proton/ap003/Metodologia%20TUSD.pdf)>. Acesso em: 11 mai. 2016.

\_\_\_\_\_. **PRODIST: Módulo 2 - Planejamento da Expansão do Sistema de Distribuição.** Disponível em:

<[http://www.aneel.gov.br/documents/656827/14866914/M%C3%B3dulo2\\_Revis%C3%A3o7.pdf/c1cf5bd8-b2bc-4d57-9b42-285a7fd8c2a5](http://www.aneel.gov.br/documents/656827/14866914/M%C3%B3dulo2_Revis%C3%A3o7.pdf/c1cf5bd8-b2bc-4d57-9b42-285a7fd8c2a5)>. Acesso em: 14 jun. 2016.

\_\_\_\_\_. **PRODIST: Módulo 6 - Informações Requeridas e Obrigações.** Disponível em:

<[http://www.aneel.gov.br/documents/656827/14866914/M%C3%B3dulo6\\_Revis%C3%A3o12.pdf/21341314-de91-44cb-b091-1e768ca3be8d](http://www.aneel.gov.br/documents/656827/14866914/M%C3%B3dulo6_Revis%C3%A3o12.pdf/21341314-de91-44cb-b091-1e768ca3be8d)>. Acesso em: 14 jun. 2016.

\_\_\_\_\_. **Resolução Homologatória nº 2.018 de 2 de fevereiro de 2016.** Homologa as cotas anuais da Conta de Desenvolvimento Energético – CDE para o ano de 2016 e dá outras providências. Disponível em: <

<http://www2.aneel.gov.br/aplicacoes/audiencia/arquivo/2015/083/resultado/reh20162018ti.pdf>>. Acesso em: 20 jul. 2016.

\_\_\_\_\_. **Resolução Normativa nº 414 de 9 de setembro de 2010.** Estabelece as Condições Gerais de Fornecimento de Energia Elétrica de forma atualizada e consolidada. Disponível em: < <http://www2.aneel.gov.br/cedoc/ren2010414comp.pdf> >. Acesso em: 20 jul. 2016.

DONI, M. V. **Análise de cluster:** métodos hierárquicos e de particionamento. 2004. Trabalho de Conclusão de Curso (Graduação em Sistemas de Informação) – Faculdade de Computação e Informática, Universidade Presbiteriana Mackenzie, São Paulo, 2004. Disponível em: <<http://meusite.mackenzie.com.br/rogerio/tgi/2004Cluster.PDF>>. Acesso em: 08 mai. 2016.

ELETROBRAS. **Nova Tarifa de Energia Elétrica:** metodologia e aplicação. Brasília, 1985. 444 p.

FIOROTTI, R. **Tipologia de cargas e redes: aplicação de uma metodologia para definir o perfil da curva de carga de unidades consumidoras e de redes elétricas.** 2013. Projeto de Graduação (Graduação em Engenharia Elétrica) – Colegiado de Engenharia Elétrica, Universidade Federal do Espírito Santo, Vitória, 2013.

GEMIGNANI, M. M. F. **Proposição e Análise Comparativa de Métodos Alternativos de Seleção e Classificação de Curvas de Carga para a Definição de Tipologias para Estudos**

**Tarifários.** 2009. Dissertação (Mestrado em Engenharia) – Universidade de São Paulo, São Paulo, 2009. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/3/3143/tde-29062009-151951/pt-br.php>>. Acesso em: 10 mai. 2016.

GUARDIA, E. C.; QUEIROZ, A.R.; MARANGON LIMA, J.W. Estimation of Electricity Elasticity for Demand Rates and Load Curve in Brazil. In: IEEE POWER AND ENERGY SOCIETY GENERAL MEETING, Minneapolis. 2010.

KAUFMAN, L.; ROUSSEEUW, P. J. Finding groups in data: an introduction to cluster analysis. New York: Wiley, 1990.

LATTIN, J.; CARROL, J. D.; GREEN, P. E. **Análise de Dados Multivariados.** São Paulo: Cengage Learning, 2011.

PESSANHA, J. F. M.; HUANG, J. L. C.; PEREIRA, L. A. C.; PASSOS JUNIOR, R.; CASTELLANI, V. L. O. Metodologia e sistema computacional para cálculo das tarifas de uso dos sistemas de distribuição. In: XXXVI SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL, São João del-Rei. 2004. Disponível em: <<http://www.din.uem.br/sbpo/sbpo2004/pdf/arg0065.pdf>>. Acesso em: 10 mai. 2016.

PESSANHA, J.F.M.; CASTELLANI, V.L.O.; ARAÚJO, A. L. A. Uma nova ferramenta computacional para construção de tipologias de curvas de carga. In: X SIMPÓSIO DE ESPECIALISTAS EM PLANEJAMENTO DA OPERAÇÃO E EXPANSÃO ELÉTRICA, Florianópolis. 2006. Disponível em: <[https://www.researchgate.net/publication/279940107\\_UMA\\_NOVA\\_FERRAMENTA\\_COMPUTACIONAL\\_PARA\\_CONSTRUCAO\\_DE\\_TIPOLOGIAS\\_DE\\_CURVAS\\_DE\\_CARGA](https://www.researchgate.net/publication/279940107_UMA_NOVA_FERRAMENTA_COMPUTACIONAL_PARA_CONSTRUCAO_DE_TIPOLOGIAS_DE_CURVAS_DE_CARGA)>. Acesso em: 10 mai. 2016.

PESSANHA, J.F.M.; XAVIER, V. L.; AMARAL, M. R. S.; LAURENCEL, L. C. Construindo tipologias de curvas de carga com o programa R. **Pesquisa Operacional para o Desenvolvimento**, Rio de Janeiro, v.7, n.1, p. 29-54, jan. a abr. 2015.

QUEIROZ, P. V. S. **Mensuração do consumo de energia elétrica:** algoritmo para detecção de potenciais usuários da termoacumulação como alternativa para deslocamento de carga 2011. Dissertação (Mestrado em Metrologia) – Programa de Pós-Graduação em Metrologia da PUC-Rio, Rio de Janeiro, 2011. Disponível em: <[http://www2.dbd.puc-rio.br/pergamum/tesesabertas/0913478\\_2011\\_pretextual.pdf](http://www2.dbd.puc-rio.br/pergamum/tesesabertas/0913478_2011_pretextual.pdf)>. Acesso em: 14 jun. 2016.

SOUZA, R. M. C. R. **Métodos de cluster para intervalos usando algoritmos do tipo nuvens dinâmicas.** 2003. Tese (Doutorado em Ciência da Computação) - Programa de Pós Graduação em Ciência da Computação do Centro de Informática, Universidade Federal do Pernambuco, Recife, 2003. Disponível em: <[http://repositorio.ufpe.br/bitstream/handle/123456789/1904/arquivo4841\\_1.pdf?sequence=1&isAllowed=y](http://repositorio.ufpe.br/bitstream/handle/123456789/1904/arquivo4841_1.pdf?sequence=1&isAllowed=y)>. Acesso em: 11 mai. 2016.

SUPERINTENDÊNCIA DE REGULAÇÃO DOS SERVIÇOS DE DISTRIBUIÇÃO; AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA. **NOTA TÉCNICA N°0063/2014-SRD/ANEEL.** Proposta de aprimoramento para a regulamentação sobre a caracterização da carga e do sistema elétrico das distribuidoras. Brasília, 2014. Disponível em:

<[http://www2.aneel.gov.br/aplicacoes/audiencia/arquivo/2014/028/documento/nota\\_tecnica\\_0063\\_2014\\_srd\\_aneel.pdf](http://www2.aneel.gov.br/aplicacoes/audiencia/arquivo/2014/028/documento/nota_tecnica_0063_2014_srd_aneel.pdf)>. Acesso em: 25 jun. 2016.

VICINI, L. **Análise multivariada da teoria a prática**. 2005. Monografia (Especialização em Estatística e Modelagem Quantitativa) – Programa de Pós Graduação em Estatística e Modelagem Quantitativa, Universidade Federal de Santa Maria, Santa Maria, 2005.

Disponível em:

< <http://w3.ufsm.br/adriano/livro/Caderno%20dedatico%20multivariada%20-%20LIVRO%20FINAL%201.pdf>>. Acesso em: 18 mai. 2016.

## APÊNDICE A – CÓDIGO DO PROGRAMA R

```

amostras <- matrix(nrow=8,ncol=96)      # matriz de 8 linhas e 96 colunas

amostras[1,] <- c(9844.8, 9811.2, 9710.4, ...) # curva 1 com 96 medições
amostras[2,] <- c(4972.8, 4905.6, 4872, ...) # curva 2 com 96 medições
amostras[3,] <- c(3024, 2956.8, 2889.6, ...) # curva 3 com 96 medições
amostras[4,] <- c(16531.2, 15926.4, 15680, ...) # curva 4 com 96 medições
amostras[5,] <- c(14784, 14515.2, 14280, ...) # curva 5 com 96 medições
amostras[6,] <- c(9619.2, 9331.2, 9417.6, ...) # curva 6 com 96 medições
amostras[7,] <- c(4317.6, 4267.2, 4250.4, ...) # curva 7 com 96 medições
amostras[8,] <- c(10718.4, 10281.6, 10012.8, ...) # curva 8 com 96 medições

d_media = apply(amostras,1,mean) # demanda média em cada curva de carga
curvas_pu = sweep(amostras,1,d_media,FUN='/') # curva normalizada pelo PU da média

result_ward = hclust(dist(curvas_pu),method='ward.D',members=NULL) # execução do
método de Ward
plot(result_ward) # exibição do dendograma
result_kmeans = kmeans(curvas_pu,centers=3,iter.max=100,nstart=10) # execução do método
K-means com 3 grupos, no máximo 100 iterações do algoritmo e 10 inicializações aleatórias
dos centroides iniciais
result_kmeans$cluster # agrupamentos das curvas de carga classificadas

result_kmeans$tot.withinss #WSS
result_kmeans$betweenss #BSS

# Pseudo-F e BSS para níveis de agregação entre 2 e 7 clusters
pseudoF = numeric(0)
bss = numeric(0)
n_curvas=dim(curvas_pu)[1] # número de curvas de carga

for (i in 2:7) {
result_kmeans = kmeans(curvas_pu,i,nstart=10)

```

```
auxpseudoF = (result_kmeans$betweenss/(i-1))/(result_kmeans$tot.withinss/( n_curvas-i))
auxbss = result_kmeans$betweenss/result_kmeans$totss
pseudoF = c(pseudoF,auxpseudoF)
bss = c(bss,auxbss)
}
par(mfrow=c(1,2))
plot(seq(2,7,1),pseudoF,xlab= 'n. de clusters ',ylab= 'pseudoF ',pch=1)
plot(seq(2,7,1),bss*100,xlab= 'n. de clusters ',ylab= '%BSS ',pch=1)
```