

TRATAMENTO DE DADOS DE CURVAS DE CARGA VIA ANÁLISE DE  
AGRUPAMENTOS E TRANSFORMADA WAVELETS

Luiz Antonio Alves de Oliveira

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia de Sistemas e Computação.

Orientador: Adilson Elias Xavier

Rio de Janeiro  
Dezembro de 2013

TRATAMENTO DE DADOS DE CURVAS DE CARGA VIA ANÁLISE DE  
AGRUPAMENTOS E TRANSFORMADA WAVELETS

Luiz Antonio Alves de Oliveira

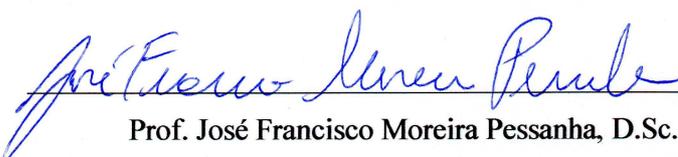
TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ  
COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA  
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS  
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM  
CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Examinada por:



---

Prof. Adilson Elias Xavier, D.Sc.



---

Prof. José Francisco Moreira Pessanha, D.Sc.



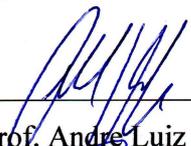
---

Prof. Mauro Antonio Rincon, D.Sc.



---

Prof.ª Marcia Helena Costa Fampa, D.Sc.



---

Prof. Andre Luiz Diniz Souto Lima, D.Sc.



---

Prof. João Lauro Dornelles Facó, D.Sc.

RIO DE JANEIRO, RJ – BRASIL

DEZEMBRO DE 2013

Oliveira, Luiz Antonio Alves de

Tratamento de Dados de Curvas de Carga via Análise de Agrupamentos e Transformada Wavelets/Luiz Antonio Alves de Oliveira. - Rio de Janeiro: UFRJ/COPPE, 2013.

X, 74p.:il.; 29,7cm.

Orientador: Adilson Elias Xavier

Tese (doutorado) - UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2013.

Referências Bibliograficas:p. 69-74.

1. Análise de Agrupamentos. 2. Transformada Wavelet. I. Xavier, Adilson Elias II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III Título.

# Dedicatórias

À minha querida família. Minha querida esposa Regina Maria Ribeiro de Carvalho e aos meus queridos filhos: Maria Luiza Carvalho Alves de Oliveira e Jorge Luiz Carvalho Alves de Oliveira.

À minha querida mãe Judith Alves.

Ao meu pai José Antonio de Oliveira.

À minha sogra Maria Rute Ribeiro de Carvalho (*in memoriam*).

Ao meu sogro João Batista Pereira de Carvalho.

Ao amigo e grande profissional Ricardo Diniz Rangel (*in memoriam*).

# Agradecimentos

A minha esposa e filhos que sempre apoiaram-me, neste projeto e compreendeu minha ausência na elaboração deste trabalho.

A minha mãe Judith Alves pelo apoio e carinho.

Ao meu orientador Adilson Elias Xavier, pelo apoio, pelo respeito, pelo carinho e pela paciência nesta longa jornada.

A Sra. Solange, Leonardo Xavier e Vinicius Xavier, pela hospitalidade, pelo carinho e pela confiança.

A José Francisco Moreira Pessanha, pela ajuda.

Ao CEPEL, por ter incentivado este trabalho.

A Flávio Rodrigo de Miranda Alves que confiou e apoiou esta tese.

Aos colegas de trabalho que sempre incentivaram e deram sugestões. Especificamente: Ricardo Penido Dutt Ross, Gilberto Pires de Azevedo, Juan Ignacio Patricio Rossi Gonzalez, Sergio Porto Romeno, Roberto Baitelli, Sergio Gomes Junior, Sergio Luiz Varricchio, Javier Ruben Ojuda Soto, Lucimar Gerhardt de Souza, Leonardo Carpi, Leonardo Pinto Almeida, Plutarcho Maravilha Lourenço, Wo Wei Ping, Fabricio Lucas Lírio, Luciano de Souza Moulin, Cristiano de Oliveira Costa, Fabiola Ferreira Clement Veliz, Luciano de Oliveira Daniel, Andrea de Mattos Rei Javaroni, Tiago Santana do Amaral e Amanda Martini Santana.

Aos colegas de Juiz de Fora-MG, João Alberto Passos Filho e Ricardo Mota Henriques pela ajuda e incentivo.

A secretaria do PESC pela ajuda.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

TRATAMENTO DE DADOS DE CURVAS DE CARGA VIA ANÁLISE DE  
AGRUPAMENTOS E TRANSFORMADA WAVELETS

Luiz Antonio Alves de Oliveira

Dezembro/2013

Orientador: Adilson Elias Xavier

Programa: Engenharia de Sistemas e Computação

As curvas de carga representam o perfil do consumo de energia elétrica em função do tempo em uma determinada região geográfica. Tanto para a operação e como para o planejamento do fornecimento de energia elétrica é extremamente importante prever o consumo e antecipar ações para garantir o correto fornecimento de energia elétrica. As previsões são baseadas nos valores passados da própria carga. Em geral o histórico com os registros dos valores de carga apresentam problemas como lacunas de dados e observações aberrantes. A presença destes problemas compromete a qualidade da previsão e para isto é necessário dispor de métodos com a finalidade de corrigir os erros presentes nos registros da carga. Neste trabalho são propostas novas metodologias de tratamento de dados de carga utilizando análise de agrupamentos e *wavelet*.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

TREATMENT OF LOAD CURVE DATA VIA CLUSTERING ANALYSIS AND  
WAVELET TRANSFORM

Luiz Antonio Alves de Oliveira

December/2013

Advisor: Adilson Elias Xavier

Department: Systems Engineering and Computer Science

The load curves represent the profile of electricity consumption as a function of time in a particular geographic region. Both for the operation and how to plan the supply of electricity is extremely important to provide consumption and anticipate actions to ensure proper supply of electricity. The predictions are based on past values of the charge itself. In general the history of the recorded load values have problems with data gaps and aberrant. The presence of these problems compromise the quality of the forecast and it is necessary to have methods in order to correct the errors in the records of the load. This thesis proposed new methodologies for data processing load using cluster analysis and wavelet.

# Sumário

1.	Introdução.....	1
2.	Curvas de carga .....	4
3.	Filtragem de Dados.....	7
4.	Metodologias de Filtragem.....	11
4.1.	Transformada Wavelet .....	11
4.1.1.	Transformada <i>Wavelet</i> Contínua e Discreta .....	12
4.1.2.	Decomposição <i>Wavelet</i> .....	14
4.2.	Boxplot .....	16
4.3.	LOESS.....	17
4.4.	Aplicação Metodológica.....	18
4.4.1.	Identificação de Falhas .....	19
4.4.2.	Correção de Falhas por descontinuidade .....	22
5.	Análise de Agrupamento .....	25
5.1.	Definição dos dados de entrada .....	26
5.2.	Problema de análise de agrupamentos.....	26
5.2.1.	Função de distância .....	27
5.2.2.	Coefficientes de similaridade.....	27
5.2.2.1.	Coefficiente de Gower .....	28
5.2.3.	Medidas de dispersão interna .....	28
5.2.4.	Função objetivo .....	29
5.2.5.	Métodos de análise de agrupamentos .....	29
5.2.5.1.	Métodos hierárquicos .....	29
5.2.5.2.	Métodos de realocação iterativa .....	30
5.2.5.3.	Métodos de programação matemática .....	30
5.2.5.3.1.	Métodos da Soma dos Mínimos Quadrados .....	30
5.2.5.4.	Agrupamento pela soma dos mínimos quadrados .....	31
6.	HSCM.....	33
7.	Experimentos Computacionais .....	39
7.1.	Apresentação dos grupos .....	41
7.2.	Apresentação dos <i>Centroids</i> .....	51
7.3.	Comparação entre <i>centroids</i> .....	56
7.4.	Síntese.....	58
8.	Aplicações de tipologias.....	59
8.1.	Correção de Lacunas .....	59
8.2.	Previsão de carga .....	61
8.3.	Cálculo de Tarifas.....	63
9.	Conclusões.....	65

# Lista de Figuras

Figura 1. 1 Exemplo de Curva de Carga .....	1
Figura 3. 1 Observações Aberrantes.....	7
Figura 3. 2 Descontinuidades na curva de carga .....	7
Figura 3. 3 Lacunas de dados .....	8
Figura 3. 4 Segmentação da curva de carga .....	9
Figura 4. 1 Wavelets mãe: (a) Haar, (b) Daubechies, (c) Coiflet, (d) Symmlet.....	13
Figura 4. 2 Decomposição com Wavelet (Reis & Silva, 2004).....	14
Figura 4. 3 Decomposição de curva de carga com TDW (GUIRELLI, 2006).....	15
Figura 4. 4 Exemplo de BoxPlot .....	16
Figura 4. 5 Fluxograma da filtragem de dados.....	18
Figura 4. 6 Decomposição de curva de carga.....	20
Figura 4. 7 Decomposição de curva de carga com lacuna.....	22
Figura 4. 8 Correção de descontinuidade .....	23
Figura 5. 1 Cálculo de $Z_j$ ao <i>centroid</i> mais próximo .....	32
Figura 7. 1 Grupos e seus <i>centroids</i> .....	40
Figura 7. 2 (a) Grupo 1 .....	41
Figura 7. 2 (b) Grupo 2.....	42
Figura 7. 2 (c) Grupo 3 .....	42
Figura 7. 2 (d) Grupo 4.....	43
Figura 7. 2 (e) Grupo 5 .....	43
Figura 7. 2 (f) Grupo 6 .....	44
Figura 7. 2 (g) Grupo 7.....	45
Figura 7. 2 (h) Grupo 8.....	45
Figura 7. 2 (i) Grupo 9.....	46
Figura 7. 2 (j) Grupo 10.....	47
Figura 7. 3 Mapa Percentual.....	48
Figura 7. 4 Mapa Percentual – Ampliação .....	49
Figura 7. 5 (a) <i>Centroid</i> referente ao grupo 1.....	51
Figura 7. 5 (b) <i>Centroid</i> referente ao grupo 2.....	51
Figura 7. 5 (c) <i>Centroid</i> referente ao grupo 3.....	52
Figura 7. 5 (d) <i>Centroid</i> referente ao grupo 4.....	52
Figura 7. 5 (e) <i>Centroid</i> referente ao grupo 5.....	52
Figura 7. 5 (f) <i>Centroid</i> referente ao grupo 6 .....	53
Figura 7. 5 (g) <i>Centroid</i> referente ao grupo 7.....	53
Figura 7. 5 (h) <i>Centroid</i> referente ao grupo 8.....	54
Figura 7. 5 (i) <i>Centroid</i> referente ao grupo 9 .....	54
Figura 7. 5 (j) <i>Centroid</i> referente ao grupo 10 .....	55
Figura 7. 6 – Comparativo entre o <i>centroids</i> dos grupos 1 e 5 .....	56
Figura 7. 7 – Comparativo entre o <i>centroids</i> dos grupos 1 e 3 .....	57
Figura 7. 8 – Comparativo entre o <i>centroids</i> dos grupos 3 e 4 .....	58
Figura 8. 1 Tipologia para preenchimento de lacuna .....	50
Figura 8. 2 Curva de carga com lacuna preenchida .....	50
Figura 8. 1 Padrões semanais típicos da carga .....	63
Figura 8. 2 Níveis de tensão e Perfis típicos de clientes e redes .....	64

# Lista de Tabelas

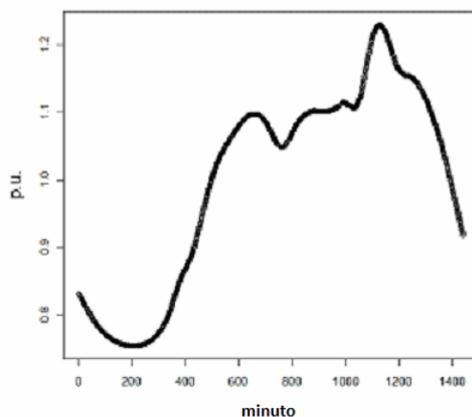
Tabela 7. 1 Medidas de validação e tempo de processamento .....	40
Tabela 7. 2 Participação dos dias de semana nos grupos .....	47
Tabela 7. 3 Participação das estações do ano nos grupos.....	50

# Capítulo 1

## Introdução

A questão energética, em particular, a questão de energia elétrica é de fundamental importância no desenvolvimento sócio-econômico de qualquer país. No Brasil, a demanda por energia elétrica tem aumentado nos últimos anos, esse aumento é reflexo do desenvolvimento econômico (EPE, 2013). Podemos entender como demanda, o consumo de todos os consumidores em uma determinada região geográfica. A representação desse fenômeno como função do tempo é denominada curva de carga.

As curvas de carga demonstram nesse caso, o comportamento temporal do consumo de energia elétrica em uma dada região geográfica. Um exemplo de curva de carga é ilustrado na figura 1.1. Esta curva ilustra um comportamento típico. O dia começa com baixo consumo e, depois das 06:00h, o consumo aumenta até o primeiro máximo (aproximadamente ao meio dia). Seguido por ligeira diminuição e, então, o consumo aumenta até o segundo máximo, aproximadamente às 20:00hs. Durante o resto da noite, o consumo decresce.



**Figura 1. 1 Exemplo de Curva de Carga**

É importante salientar que toda energia gerada deve ser consumida imediatamente, isto porque as possibilidades de armazenamento de energia elétrica, no momento, são economicamente inviáveis. Os perfis típicos da demanda diária por energia elétrica são informações fundamentais em diversas etapas do planejamento e da operação de sistemas elétricos de potência. Por exemplo, no cálculo das tarifas que remuneram o serviço de distribuição de eletricidade (PESSANHA, 2001), na avaliação das perdas de

energia, no despacho das unidades geradoras, na previsão de demanda e no dimensionamento do sistema, entre outras aplicações importantes para a operação econômica e segura de um sistema elétrico. Do ponto de vista de planejamento é importante prever o consumo ou gerar cenários confiáveis de consumo futuro e então antecipar ações que permitam o fornecimento de energia de forma estável.

As curvas de carga são determinadas pelos registros de consumo de energia elétrica mensurados na operação do sistema. Esses dados coletados estão sujeitos a eventos aleatórios provocando uma série de falhas nas curvas de carga. Em síntese, podem ocorrer: (i) erros de medição provocando pontos de descontinuidade na curva; ou (ii) perda de dados medidos provocando grandes lacunas na curva (PESSANHA, 2011).

Desta forma, é necessário aplicar uma metodologia para filtrar os dados iniciais de carga. Existem várias metodologias fundamentadas em diferentes ferramentas matemáticas e estatísticas para efetuar esses procedimentos de filtragem de dados, por exemplo, as metodologias descritas por CHEN (2010), TORRES (2006), XIAOXING (2008) e PESSANHA et al (2011).

O presente trabalho tem por finalidade, primeiramente descrever metodologias para tratamento de dados de carga para corrigir aos diversos tipos de erros e falhas. Para isso, serão utilizados: a transformada discreta *wavelet* DAUBECHIES (1992) denominada TDW e o método de regressão não paramétrico LOESS - *Locally Weighted Regression and Smoothing Scatterplots* (MARTINEZ & MARTINEZ, 2002; HASTIE et al, 2009).

A segunda finalidade é descrever uma metodologia para identificação de perfis típicos de demanda (tipologias). Normalmente esses perfis correspondem aos dias da semana, aos feriados e dias extraordinários. Nessa finalidade, será utilizado o método de análise de agrupamento baseado em suavização hiperbólica XAVIER (2010,2011) denominado HSCM - *Hyperbolic Smoothing Clustering Method*.

Resumidamente, os métodos de análise de agrupamentos têm por finalidade a classificação de objetos em grupos, de tal forma que os objetos semelhantes sejam classificados no mesmo grupo, enquanto objetos distintos sejam classificados em grupos diferentes.

As curvas de carga são um tipo de série temporal. Em um contexto mais amplo, a análise de agrupamento em séries temporais apresenta grande aplicabilidade em diversas áreas: Economia (identificação de padrões de oferta e procura), Engenharia (reconhecimento de voz), Ciência (identificação de padrões sísmicos) e Medicina (identificação de padrões em eletrocardiograma).

Na literatura técnica, são encontrados diferentes métodos de análise de agrupamentos, esses são empregados na identificação de perfis de carga. Entre os vários métodos disponíveis, foi feita a escolha do método HSCM de XAVIER (2010, 2011) pela capacidade de tratar problemas com grande número de componentes fornecendo resultados com consistência, precisão e robustez.

A tese está organizada em oito capítulos. A seguir, no capítulo 2, é feita uma descrição das curvas de carga, bem como, de suas aplicações. No capítulo 3, é descrita a filtragem de dados. No capítulo 4, são descritas as metodologias de filtragem de dados. No capítulo 5 é descrita a análise de agrupamento. No capítulo 6 são apresentados os procedimentos de construções de agrupamentos. No capítulo 7, são apresentados os experimentos computacionais. No capítulo 8, são apresentados os usos das tipologias e no capítulo 9, são descritas as conclusões.

# Capítulo 2

## Curvas de carga

Neste capítulo são apresentadas as curvas de carga, tema principal deste trabalho. As curvas de carga são séries temporais que representam a demanda de energia elétrica ao longo do dia.

No contexto dos sistemas elétricos de potência, estes são divididos em três subsistemas: geração, transmissão e distribuição.

Segundo a EPE – Empresa de Pesquisa de Energética (2013), a geração no Brasil é realizada por usinas: hidrelétrica, térmicas (gás natural, carvão e óleo combustível), biomassa, termonuclear e eólica. Na matriz elétrica brasileira de 2012, de acordo com a EPE (2013), as usinas hidrelétricas representaram 76,90% da geração de energia elétrica.

O local adequado para construção destas usinas é majoritariamente determinado pelo maior aproveitamento do potencial hidrelétrico. Desta forma, geralmente as usinas hidrelétricas são instaladas distantes do mercado consumidor.

Por esta razão é necessário construir uma grande rede de transmissão, feita em alta tensão para minimizar as perdas intrínsecas à mesma. O subsistema de transmissão é formado por linhas de transmissão de alta tensão (230kv até 765kv). O objetivo é transmitir energia elétrica a grandes distâncias.

A distribuição é realizada por linhas de transmissão de baixa tensão (13,8kv até 138kv). Nesse caso, objetiva-se fornecer energia elétrica para o consumidor final. O comportamento do conjunto de consumidores finais, seja uma pequena família ou uma grande indústria, determina a definição das curvas de carga.

Estas partes que compõem o sistema elétrico de potência possuem características distintas de operação, mas possuem o objetivo comum de suprir o mercado de energia da melhor forma possível. Isto é conseguido minimizando as interrupções (*blackout*) e também mantendo os níveis de tensão e de frequência em limites pré-estabelecidos.

No contexto de operação do sistema elétrico, existem vários aspectos que devem ser observados: econômico (energia a baixo custo), segurança (minimização das interrupções), recomposição (restauração do sistema após uma falha), qualidade (respeito aos limites de tensão e frequência).

Nesta dinâmica, como dito anteriormente, existe a figura do consumidor final e cada um com o seu consumo (demanda), determina a quantidade de energia elementar que deve ser gerada (produzida) para o seu suprimento individual. Ademais, toda energia gerada deve ser consumida imediatamente, desde que, as possibilidades tecnológicas de armazenamento de energia elétricas são inviáveis economicamente.

O comportamento do conjunto de consumidores finais em um dado espaço geográfico é determinado por um conjunto de fatores: temperatura, período de chuva, tarifas de energia, período de férias escolares e etc. Em suma o comportamento é probabilístico.

O consumo agregado dos consumidores brasileiros é quantizado por medições temporais feitas por equipamento específico: os medidores. Infelizmente, ocorrem erros de medição e assim os dados de consumo apresentam falhas que podem ser lacunas ou simplesmente dados aberrantes (descontínuos).

A demanda agregada do conjunto de consumidores finais é refletida em gráficos, especificamente em curvas de carga. As curvas de cargas demonstram o consumo agregado (de carga elétrica) ao longo do tempo. Estas curvas podem ser discretizadas em pequenos intervalos (ordem de minutos) e também em grandes intervalos (ordem de horas).

Como dito anteriormente, a demanda de energia determina a geração de energia no tempo. Assim, do ponto de vista do planejamento e da operação é fundamental ter informações futuras do consumo para antecipar medidas a serem tomadas para garantir o fornecimento de energia de forma adequada e com qualidade. Isto é feito através de métodos de previsão que se fundamentam nessas curvas de carga.

Previsões de carga precisas são fundamentais para a operação econômica e segura de um sistema elétrico de potência. Se por um lado previsões superestimadas tendem a elevar os custos operacionais do sistema com a necessidade de uma maior reserva operativa, por outro lado previsões subestimadas contribuem para reduzir a reserva operativa comprometendo a segurança do sistema.

A necessidade de previsões precisas tem conduzido ao desenvolvimento de um amplo conjunto de metodologias para previsão de carga, sobretudo de metodologias para previsão de curto prazo, em horizontes que abrangem desde uma hora até uma semana à frente, em bases horárias. Também fazem parte desse escopo às previsões de curtíssimo prazo para horizontes de 10 até 30 minutos à frente.

Atualmente há uma variedade de metodologias para previsão de carga, desde os tradicionais métodos estatísticos até métodos de inteligência computacional, entre os quais se destacam as redes neurais artificiais, a lógica fuzzy e as máquinas de vetor de suporte.

Independentemente da metodologia de previsão utilizada e da forma como esta é empregada, a construção de um modelo de previsão baseia-se no comportamento passado da carga e das suas relações com outras variáveis explicativas, por exemplo, a temperatura. Portanto, para que o modelo identificado tenha uma boa capacidade preditiva é imprescindível que os dados tenham a maior qualidade possível. No caso ideal, os dados devem estar livres de falhas e perturbações provocadas por problemas no sistema de medição (falhas de medidores e na transmissão de dados) ou provocadas por eventos como curtos-circuitos, falha de equipamentos e demais causas não naturais que afetam a trajetória da carga.

Os dados com falhas não são apropriados para serem utilizados em qualquer método de previsão. Para contornar esse problema é necessária a utilização de um processo de tratamento de dados para filtragem (limpeza).

Dado uma série histórica de curvas de carga livre de observações deficientes, com a análise de agrupamentos procura-se identificar perfis típicos de cada dia da semana e também dos feriados. Esses perfis típicos, denominado tipologias, apresentam grande aplicabilidade no planejamento do sistema elétrico, como por exemplo, subsidiar os métodos de previsão de carga. Estas tipologias, adicionalmente, são utilizadas para completar as eventuais lacunas de dados nos registros da carga.

# Capítulo 3

## Filtragem de Dados

Neste trabalho, a filtragem de dados consiste na identificação e correção de falhas. Estas falhas ocorrem em dados resultantes dos registros de consumo.

As falhas identificáveis são classificadas da seguinte forma: observações aberrantes (Figura 3.1), descontinuidades (Figura 3.2) e lacunas de dados (Figura 3.3). Conforme PESSANHA et al. (2011)

As observações aberrantes são destacadas na figura 3.1. Estas ocorrem devido a erros no registro de medições.

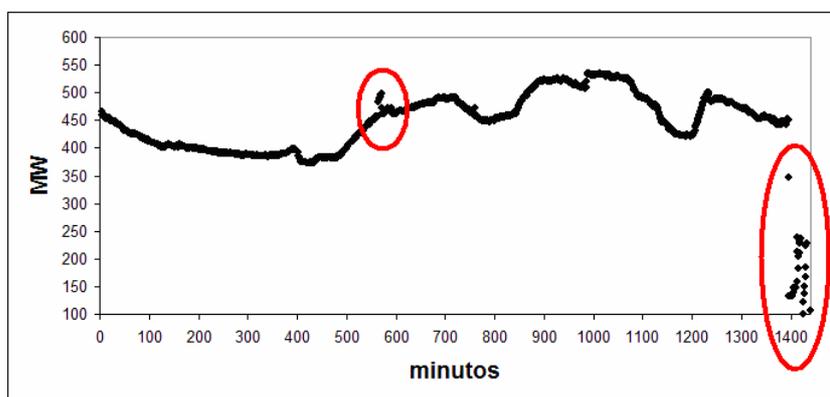


Figura 3. 1 Observações Aberrantes (PESSANHA, 2011)

Eventualmente ocorrem falhas pontuais nos registros de medição. Estas falhas são denominadas de descontinuidades e são destacadas na figura 3.2.

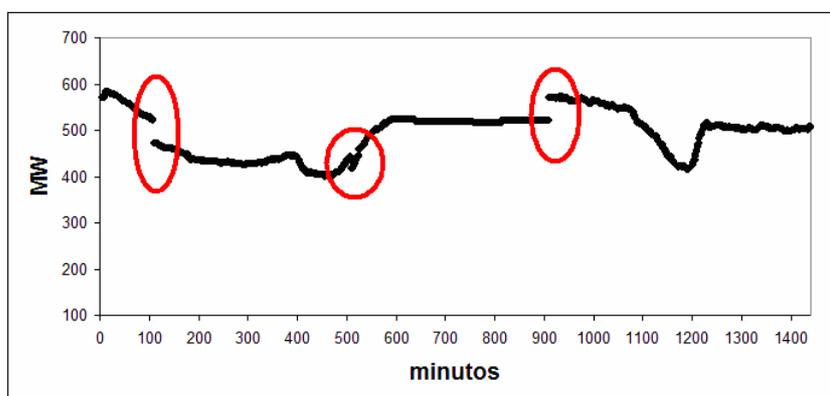
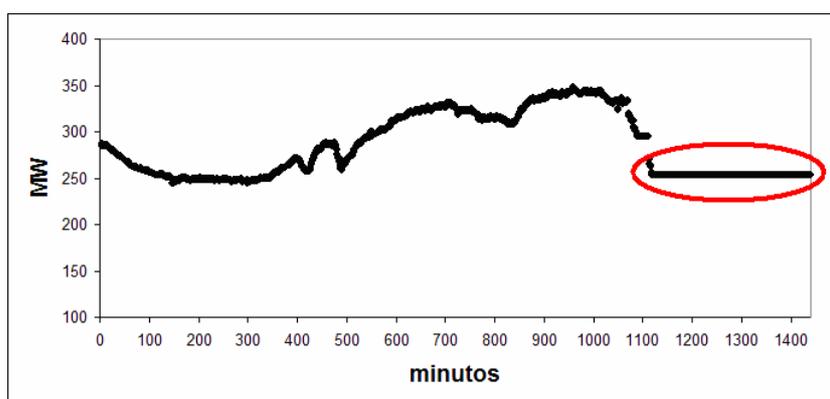


Figura 3. 2 Descontinuidades na curva de carga (PESSANHA, 2011)

As lacunas são caracterizadas pelo registro constante de consumo. Normalmente esse fenômeno apresenta-se quando ocorre perda de informações. Conseqüentemente o último valor registrado é repetido. Na figura 3.3 está destacada uma lacuna com valor constante repetido da última medição.

Infelizmente não é possível preencher uma lacuna, utilizando-se de informação da própria curva. Neste trabalho, para tratar este problema, são utilizados os perfis típicos de consumo gerados pela análise de agrupamentos aplicada as curvas de carga. Esta técnica será vista nos capítulos 5 e 6.



**Figura 3. 3 Lacunas de dados (PESSANHA, 2011)**

A presença destas falhas compromete a estimação dos parâmetros do modelo de previsão o que contribui para a especificação errônea do modelo e conseqüentemente para a sua perda de precisão.

Desta forma, para mitigar os efeitos das falhas do sistema de medição e demais causas não naturais sobre a identificação dos modelos de previsão, a construção de qualquer modelo deve ser precedida pelo tratamento dos dados históricos da carga com a finalidade de corrigir ou atenuar as observações aberrantes, as discontinuidades e as lacunas de dados.

Na literatura técnica encontram-se algumas propostas para tratamento de dados de carga, por exemplo, SUAREZ-FARINAS et al (2004), YANG & STENZEL (2005), GUIRELLI (2006), XIAOXING & CAIXIN (2008), GRIGORAS et al. (2009), GUAN et al, 2009, CHEN et al (2010), CHUNXIA (2010) e PESSANHA et. al. (2011)

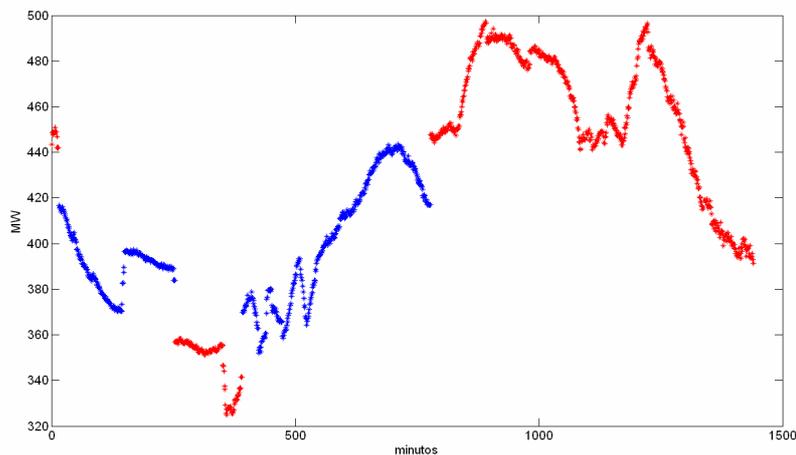
Para corrigir os dados errados deve-se primeiro identificá-los na curva de carga diária. Por exemplo, YANG & STENZEL (2005) mostram que a diferença de segunda ordem da carga é um meio eficaz na detecção dos dados errados. Seja  $L(t)$  a carga no

instante  $t$  da curva de carga diária, então a diferença de segunda ordem da carga é definida como:

$$\Delta^2 L(t) = [L(t) - L(t-1)] - [L(t-1) - L(t-2)] = L(t) - L(t-2) \quad (3.1)$$

Em sistemas físicos que variam continuamente no tempo, como a carga, a diferença de segunda ordem deve ser próxima de zero quando o sistema opera em regime normal. Porém, quando há uma mudança abrupta na tendência da carga causada por algum evento anormal ou erro de medida, a diferença de segunda ordem distancia-se do zero.

A classificação das diferenças de segunda ordem entre próximo e distante de zero baseia-se em um intervalo de confiança  $(-3, +3)$ , onde  $\bar{\Delta^2 L}$  e  $\sigma_{\Delta^2 L}$  são, respectivamente, a média e o desvio padrão das diferenças de segunda ordem de uma curva de carga (YANG & STENZEL, 2005).



**Figura 3. 4 Segmentação da curva de carga**

O uso da diferença de segunda ordem pode-se identificar os instantes das ocorrências de erros nos dados da carga e separar uma curva de carga diária em vários segmentos contínuos, conforme ilustrado na Figura 3.4. Os segmentos contínuos podem ser formados por dados válidos ou por dados errados. YANG & STENZEL (2005) propõem o uso de um modelo quadrático de regressão linear para avaliar a validade de cada segmento e fornecer as estimativas da carga caso o segmento não seja válido. Os autores advertem que esse procedimento pode gerar erros inaceitáveis em segmentos com longa duração e restringem essa proposta apenas aos segmentos com menos de 75 minutos.

Após a identificação dos segmentos e correção dos dados errados, YANG & STENZEL (2005) propõem a suavização da curva de carga resultante antes de introduzi-la no modelo de previsão de carga. Na suavização da curva carga os autores empregam o método de regressão não paramétrica LOESS - Locally Weighted Regression and Smoothing Scatterplots (MARTINEZ & MARTINEZ, 2002; HASTIE et al, 2009). A mesma abordagem foi utilizada por PESSANHA et. al. (2011)

Outros métodos de regressão não paramétrica como B-Spline e Kernel (MARTINEZ & MARTINEZ, 2002; HASTIE et al, 2009) também podem ser empregados na suavização da curva de carga, conforme apresentado por CHEN et al (2010). Esses autores utilizaram esses métodos também na detecção e correção de dados aberrantes, bem como no preenchimento de lacunas em curvas de carga.

XIAOXING & CAIXIN (2008) exploram o uso de técnicas de inteligência artificial na correção dos dados de curva de carga. Mais especificamente esses autores fazem o uso combinado de mapa de Kohonen, redes neurais de base radial e fuzzy clustering method (JANG et al, 1997). Nas atividades de correção, GRIGORAS et al. (2009) implementam um procedimento para preenchimento de lacunas de dados baseado em métodos de análise de agrupamentos (cluster analysis).

GUIRELLI (2006), em sua tese de doutorado sobre previsão de carga, obtém excelentes resultados com a aplicação da transformada wavelets (OGDEN, 1997) na filtragem das curvas de carga diárias. GUAN et al. (2009) e CHUNXIA (2010) também utilizam wavelets no tratamentos de dados de carga.

# Capítulo 4

## Metodologias de Filtragem

Neste capítulo são apresentadas as metodologias de filtragem de dados adotadas neste trabalho. Neste contexto existem dois objetivos: **identificação** e **correção**. Para identificação de falhas foi utilizada a transformada discreta *Wavelet* (TDW) proposta por DAUBECHIES (1988 e 1992). Para a correção de falhas foi aplicado o método de regressão não paramétrico LOESS - Locally Weighted Regression and Smoothing Scatterplots (MARTINEZ & MARTINEZ, 2002; HASTIE et al, 2009).

### 4.1. Transformada Wavelet

As *Wavelets* foram desenvolvidas inicialmente por HAAR (1910). As *Wavelets* de Haar permaneceram por muito tempo como a única base ortonormal de *Wavelets*. As *Wavelets* de Haar também são discretas.

Posteriormente, Mallat, Y. Meyer introduziu a primeira *Wavelet* suave. Estas são continuamente diferenciáveis, mas não possuem suporte compacto.

As bases para as aplicações atuais de *wavelet* foram fundamentadas por DAUBECHIES (1988 e 1992).

As transformadas de *Wavelets* podem ser vistas como ferramenta para decompor determinadas funções nas suas partes constituintes, permitindo analisar os dados em diferentes domínios de frequências. Além disso, pode-se utilizar funções que estão contidas em regiões finitas, facilitando a aproximação de dados com descontinuidades.

Os algoritmos de *Wavelets* processam dados em diferentes escalas ou resoluções. Independentemente do tipo de função analisada, as *Wavelets* fornecem uma técnica elegante na representação dos níveis de detalhes. Elas constituem uma ferramenta matemática para decompor funções hierarquicamente. Permitindo que uma função seja decomposta em duas componentes: uma é uma aproximação grosseira e a outra representa os detalhes.

Por causa de suas propriedades únicas, as *wavelets* são usadas em análise funcional, em singularidades ou oscilações locais em funções, em solução de equações

diferenciais, em reconhecimento de padrões, em compressão de dados, biologia, medicina, astronomia e outros.

Quando se utiliza a transformada de Fourier na análise de uma função, é extremamente difícil determinar quando um evento em particular ocorreu, pois as funções base utilizadas são localizadas em frequência, mas são aplicadas em todos os instantes. Desta forma, sugere-se que a análise de Fourier seja indicada para trabalhar com frequências que não evoluam com o tempo.

Uma alternativa para superar esta dificuldade é utilizar a transformada de *wavelet* utilizando funções que são locais, tanto no tempo como também na frequência. Assim, pode-se utilizar intervalos de tempo maiores quando deseja-se informações de baixa frequência e intervalos de tempo menores quando deseja-se informações de alta frequência.

#### 4.1.1. Transformada *Wavelet* Contínua e Discreta

A análise *wavelets* utiliza uma função protótipo denominada *wavelet* mãe. Na figura 4.1, são mostradas exemplos de funções *wavelet* mãe: Haar, Daubechies, Coiflet e Symmlet.

Essa função tem média zero e decai de forma oscilatória. Os dados são representados via superposição de versões dilatadas e transladadas da *wavelet* mãe.

A Transformada Contínua *Wavelet* (TCW) de uma função  $x(t)$ , com relação à *wavelet* mãe  $g(t)$  é dada por

$$TCW(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)g\left(\frac{t-b}{a}\right)dt \quad (4.1)$$

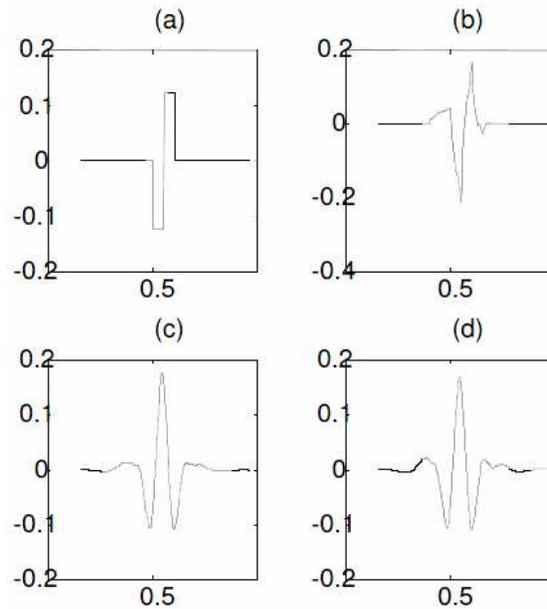
$a$  é o fator escala e  $b$  é o fator de translação.

Um coeficiente  $TCW(a,b)$ , em uma dada escala e fator de translação, representa como a função  $x(t)$  e a *wavelet* mãe dilatada/transladada se assemelham. Assim o conjunto de todos os coeficientes  $TCW(a,b)$  associados à função  $x(t)$ , é a representação *wavelet* do sinal com respeito a *wavelet* mãe  $g(t)$ .

Por exemplo, considere que o fator de escala de uma *wavelet* seja igual a  $2^{-j}$ . Quando  $j$  cresce, o fator de escala decresce (encolhimento) no tempo, as oscilações

crecem e a *wavelet* exibe alta frequência. De outra forma, se  $j$  decresce (expansão no tempo), as oscilações tornam-se mais lentas (baixa frequência). Esta é uma característica importante das *wavelets*. Componentes de alta frequência são analisadas em intervalos de tempo curtos e componentes de baixa frequência são analisadas em intervalos de tempo longos.

Com a análise *wavelets*, consegue-se caracterizar o comportamento de uma função localmente.



**Figura 4. 1 Wavelets mãe: (a) Haar, (b) Daubechies, (c) Coiflet, (d) Symmlet.**

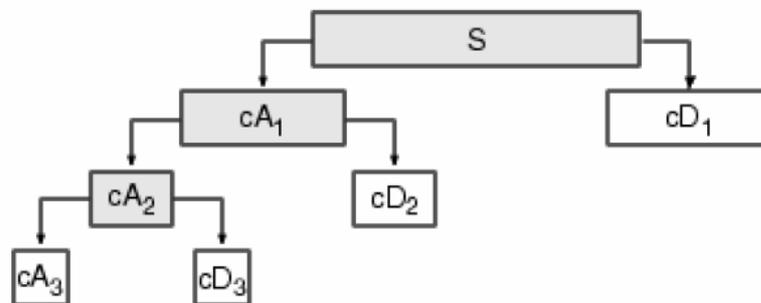
A *TCW* é conseguida dilatando-se e transladando-se a *wavelet* mãe continuamente. Pode-se também dilatar e transladar a *wavelet* mãe usando escalas e posições especiais baseadas em potência de 2. Esta forma é conhecida como Transformada Discreta *Wavelet* (*TDW*) e definida como:

$$TDW(m, k) = \frac{1}{\sqrt{a_o^m}} \sum_n x(n) g\left(\frac{k - nb_o a_o^m}{a_o^m}\right) \quad (4.2)$$

Os parâmetros  $a$  e  $b$  da equação 4.1 são funções do parâmetro inteiro  $m$ , isto é,  $a = a_o^m$  e  $b = nb_o a_o^m$ . Na equação 4.2,  $k$  é inteiro e também um parâmetro inicial.

### 4.1.2. Decomposição *Wavelet*

A *TWD* permite decompor e recompor uma função em duas componentes. Uma componente corresponde às altas frequências e outra componente corresponde às baixas frequências. O processo pode ser aplicado sucessivamente. Esse processo é exemplificado na figura 4.2



**S - sinal**

**cAn - componente de baixa frequência ou sinal filtrado**

**cDn - componente de alta frequência**

Figura 4. 2 Decomposição com Wavelet (Reis & Silva, 2004)

A decomposição pode ser repetida  $\log_2 N$  vezes, onde  $N$  é o número de amostras da função. Cada iteração do processo de decomposição gera um conjunto de coeficientes e o número de coeficientes diminui em cada iteração.

A figura 4.3 ilustra o processo de decomposição aplicado a uma curva de carga. São mostrados respectivamente: a curva original  $S$ , a curva decomposta após 5 repetições ( $a_5$ ) e as etapas de decomposição ( $d_1$  até  $d_5$ ).

A curva é integralmente reconstruída utilizando-se  $a_5$  e  $d_1$  até  $d_5$ .

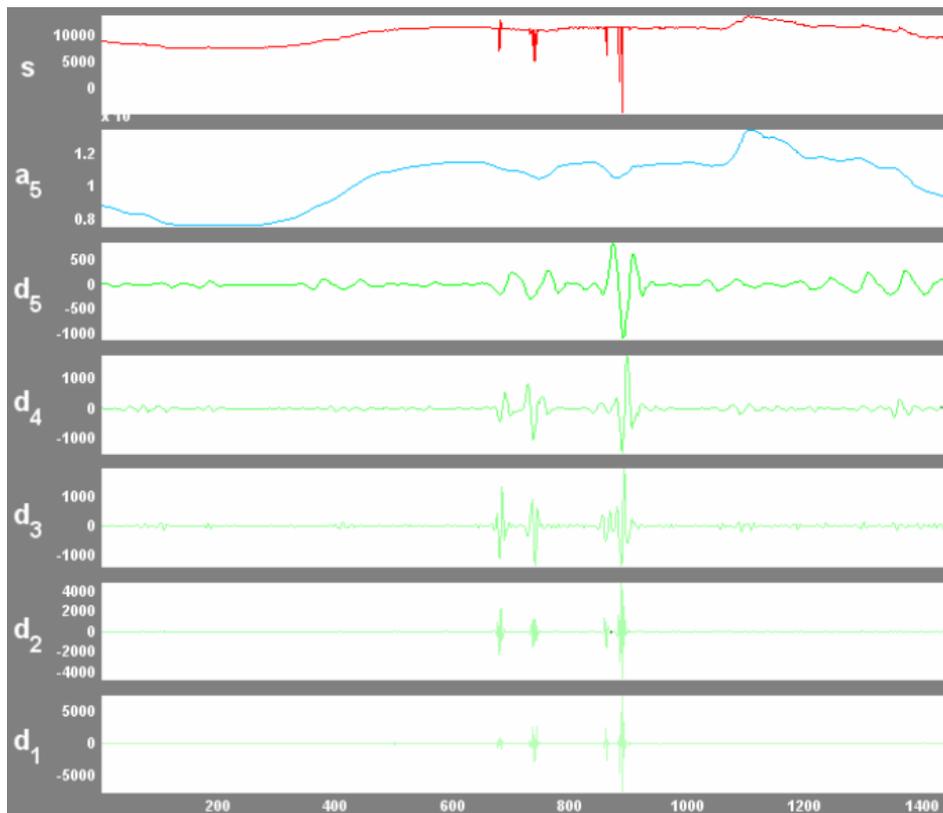


Figura 4. 3 Decomposição de curva de carga com TDW (GUIRELLI, 2006)

## 4.2.Boxplot

O *Boxplot* ou diagrama de caixa foi introduzido por TURKEY (1978). É uma ferramenta estatística útil para identificação de valores discrepantes sem conhecimento a priori da distribuição dos dados, segundo VICTORIA (2004).

Dado o conjunto de dados de entrada com  $n$  elementos. Ordenam-se esses dados em ordem crescente e calcula-se a mediana, o quartil inferior e superior. A figura 4.4 mostra um exemplo de *BoxPlot*

Esse cálculo é feito da seguinte forma: Se  $n$  for ímpar, a mediana será o elemento  $(n+1)/2$ , ou seja o elemento central. Se  $n$  for par a mediana será a média aritmética entre o elemento  $n/2$  e  $(n/2)+1$ . O quartil inferior é a mediana do conjunto dos elementos menores que a mediana do conjunto original. O quartil superior, de forma semelhante, é a mediana do conjunto dos elementos maiores do que a mediana do conjunto original. A distância interquartílica é definida como a diferença entre o quartil superior e o quartil inferior.

Para identificação dos valores discrepantes aplica-se o critério de *Whisker*, definido por TURKEY (1978). Esse critério estabelece uma cota superior, denominada máximo de *Whisker* igual ao quartil superior + 1,5 x (distância interquartílica).

De forma análoga é definido o mínimo de *Whisker* igual ao quartil inferior - 1,5 x (distância interquartílica).

Assim, um valor é considerado discrepante quando for maior ou igual ao máximo de *Whisker*; ou quando for menor ou igual ao mínimo de *Whisker*.

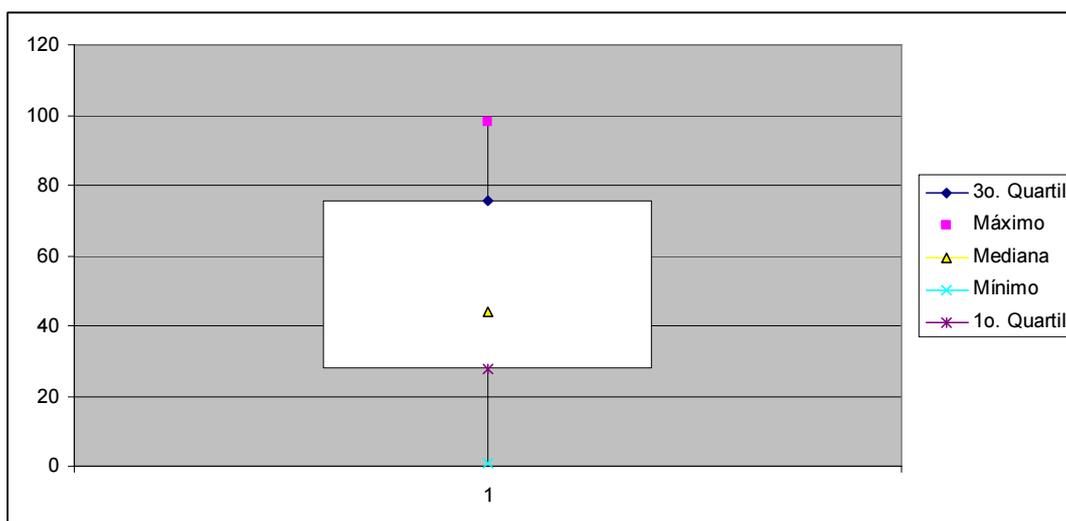


Figura 4. 4 Exemplo de BoxPlot

### 4.3.LOESS

LOESS - *Locally Weighted Regression and Smoothing Scatterplots* (MARTINEZ & MARTINEZ, 2002; HASTIE et al, 2009). Trata-se de um método de regressão não paramétrico. Útil na suavização de dados e na correção de descontinuidades.

O método é descrito passo a passo:

1) Dado um conjunto com  $N$  observações e uma janela temporal de tamanho  $k$ . Seja  $x$  o vetor de variáveis explicativas e  $y$  o vetor com as observações da variável dependente. No contexto de curvas de carga, o vetor  $y$  é uma curva típica com  $N$  observações de consumo, enquanto o vetor  $x$  é a seqüência  $1,2,3,\dots,N$  e representa os instantes das observações.

2) Para cada instante de tempo  $x_0$  identifique os  $k$  instantes  $x_i$  ( $i=1,k$ ) na vizinhança de  $x_0$  e denote esse conjunto por  $V(x_0)$ .

3) Calcule a maior distância entre  $x_0$  e o ponto  $x_i$  dentro da janela  $V(x_0)$ :

$$\Delta(x_0) = \text{máximo}_{x_i \in V(x_0)} \|x_0 - x_i\|$$

4) Pondere cada par  $(x_i, y_i)$ ,  $x_i$  em  $V(x_0)$  com base na seguinte função:

$$\text{peso}_i(x_0) = W\left(\frac{\|x_0 - x_i\|}{\Delta(x_0)}\right), \text{ onde } W(u) = \begin{cases} (1-u^3)^3 & 0 \leq u \leq 1 \\ 0 & \text{caso contrário} \end{cases}$$

5) Use o estimador de mínimos quadrados ponderados para obter uma estimativa  $\hat{y}$  para  $y$  no ponto  $x_0$  ajustado ao conjunto de observações que pertencem à vizinhança  $V(x_0)$ .

6) Repita os passos de 2 a 5 para cada instante de tempo no vetor  $x$ .

## 4.4. Aplicação Metodológica

No contexto das medições de curvas de carga, estas apresentam falhas: (i) erros de medição provocando pontos de descontinuidade na curva; ou (ii) perda de dados medidos provocando grandes lacunas na curva.

As fases cronológicas do processo de filtragem são mostradas da forma abaixo:

- (i) Identificação de falhas utilizando a transformada discreta *wavelet* (TDW) e *Boxplot*.
- (ii) Correção de falhas provocadas por erros de medição.

Na figura 4.5 é apresentado o fluxograma do processo de filtragem.

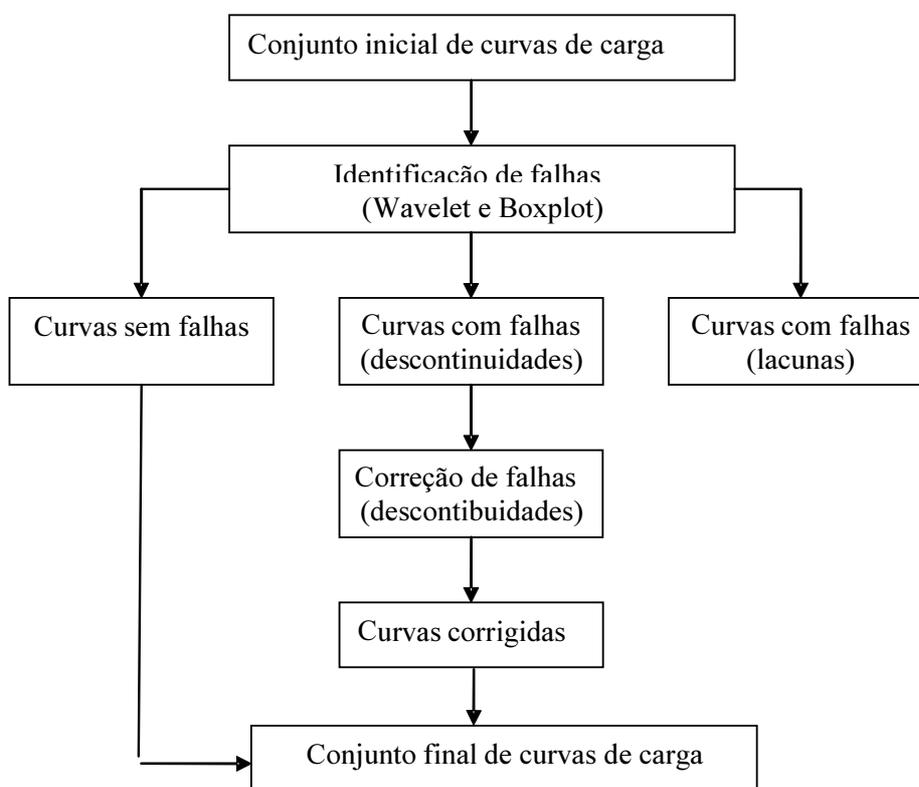


Figura 4. 5 Fluxograma da filtragem de dados

A fase inicial de identificação de falhas utiliza a transformada discreta *wavelet* (TDW) e o *Boxplot*. Conforme a seção 4.1, em síntese, com a TDW é possível dividir a curva de carga ( $C$ ) em duas componentes, uma de aproximação ( $A$ ) e outra de detalhe ( $D$ ). De tal forma que  $C = A + D$ .

As componentes *wavelet* são sensíveis à descontinuidade e conseqüentemente qualquer perturbação nos dados são refletidas na componente de decomposição. Desta forma, para selecionar os pontos de descontinuidades, foi aplicado o *Boxplot* no logaritmo dos valores absolutos da componente de detalhe  $D = C - A$ . Observe que a componente de detalhe  $D$ , representa o erro da componente de aproximação. Assim os maiores valores absolutos de  $D$ , serão selecionados como pontos de descontinuidades. As curvas em que não foram identificadas anomalias são suavizadas e incluídas no conjunto final de curvas. A suavização é feita pelo método de regressão não paramétrico LOESS - *Locally Weighted Regression and Smoothing Scatterplots* (MARTINEZ & MARTINEZ, 2002; HASTIE et al, 2009).. Na seção 4.4.1, o procedimento de identificação de falhas é descrito em detalhes.

Na segunda fase, correção de falhas de descontinuidade, também é utilizado o método LOESS.

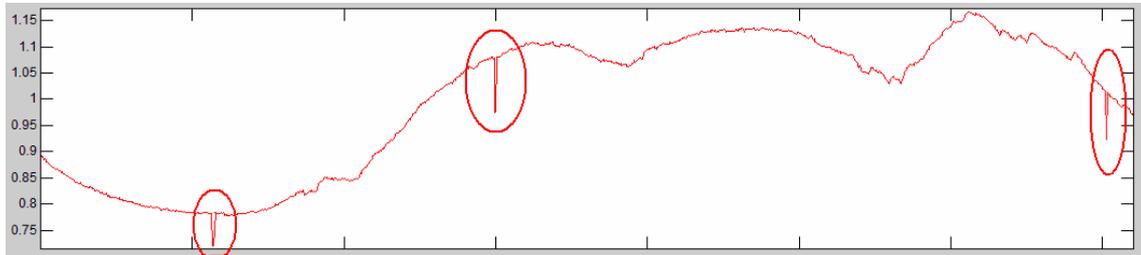
Cada ponto descontínuo é substituído momentaneamente pela regressão linear entre seus pontos contínuos vizinhos. Assim, nesta curva modificada é aplicado o método LOESS que suaviza a curva e fornece uma correção suave para os pontos descontínuos anteriores. Por até dez tentativas, aplica-se na curva corrigida o procedimento de identificação de falhas. Se ainda são identificadas falhas, reaplica-se o procedimento anterior. Se depois de dez tentativas de identificação/correção a curva ainda apresentar falha, a mesma é classificada como uma curva com falha. Na seção 4.4.2 é detalhado o procedimento de correção de falhas.

#### **4.4.1. Identificação de Falhas**

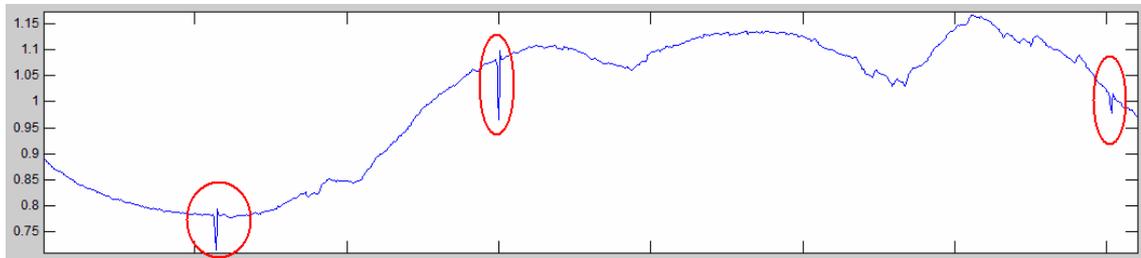
Nesta seção vamos discutir a identificação de falhas aplicando a transformada *wavelet* discreta (TWD) e o *Boxplot*. O objetivo desta abordagem é identificar se a curva de carga apresenta falhas e também identificar em que intervalo ocorrem estas falhas. Com a aplicação da TDW, divide-se a curva de carga ( $C$ ) em duas componentes tais que  $C = A + D$ . A componente  $A$  representa uma aproximação da curva de carga  $C$  e a

componente  $D$  representa o detalhe. Neste contexto, a componente  $D$  é melhor entendida como o erro da aproximação  $A$ , uma vez que  $D=C-A$ .

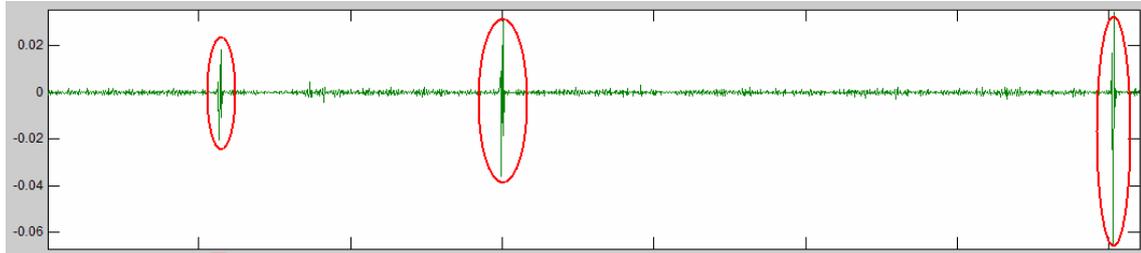
A figura 4.6 ilustra uma curva de carga com falhas e suas respectivas componentes de aproximação e detalhe.



(a) Curva de carga original



(b) Componente de aproximação  $A$



(c) Componente de detalhe  $D$

**Figura 4. 6 Decomposição de curva de carga**

Na curva de carga  $C$  da figura 4.6(a), podem-se identificar nos destaques: três pontos de descontinuidade. A componente de aproximação  $A$ , ilustrada na figura 4.6(b), não realiza qualquer correção, pois ainda apresenta os mesmos três pontos de descontinuidade.

A componente de detalhe  $D$  mostra a informação chave para identificação de falhas. Nos pontos de descontinuidades, conforme destaque da figura 4.6(c), a componente de detalhe em valor absoluto é muito maior do que qualquer outro ponto sem falha.

É necessário utilizar uma ferramenta automática para selecionar os pontos da componente de detalhe que representam efetivamente pontos de descontinuidade. Esta ferramenta é o *Boxplot* ou diagrama de caixa, introduzido por TURKEY (1978).

Segundo VICTORIA (2004), o *Boxplot* é uma das técnicas estatísticas mais simples para detecção de falhas sem conhecimento a priori da distribuição dos dados. Para esta metodologia, adotou-se o logaritmo dos valores absolutos da componente de detalhe  $D$  como o conjunto de dados de entrada para o *Boxplot*.

O mínimo de *Whisker*, definido na seção 4.2, não é aplicável neste contexto. Valores menores que este mínimo representam pontos de aproximação com erros reduzidos, ou seja, não são pontos de descontinuidade.

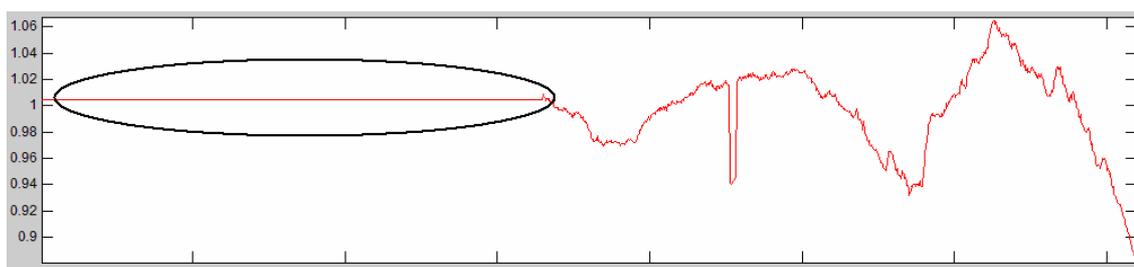
Os pontos de descontinuidade são caracterizados por grandes erros de aproximação. Desta forma foi aplicado o máximo de *Whisker*, definido na seção 4.2, para identificação de descontinuidade.

Um ponto maior ou igual ao máximo de *Whisker* é considerado um ponto de descontinuidade.

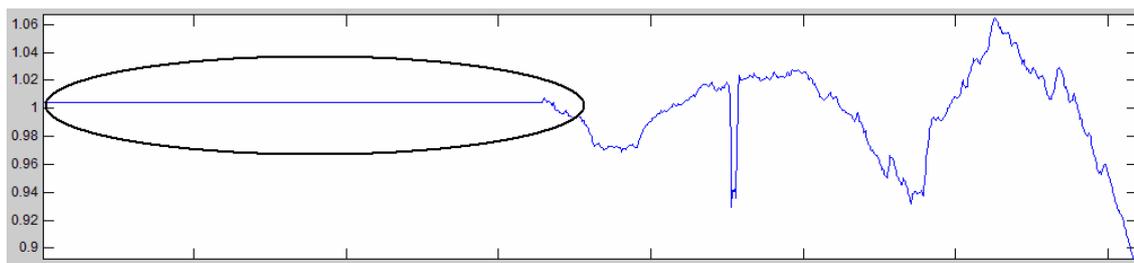
Para identificação de lacunas foi aplicado um critério mais simples. A figura 4.6 ilustra uma curva de carga com lacuna.

Na curva de carga  $C$  da figura 4.7(a) é destacada uma lacuna. A componente de aproximação  $A$  ilustrada na figura 4.7(b) conforme destaque não realiza qualquer correção apenas mantém a lacuna no mesmo valor da curva original que a princípio pode ser qualquer valor. Felizmente, na figura 4.7(c), a componente de detalhe  $D$  mostra a lacuna sempre com valor zero.

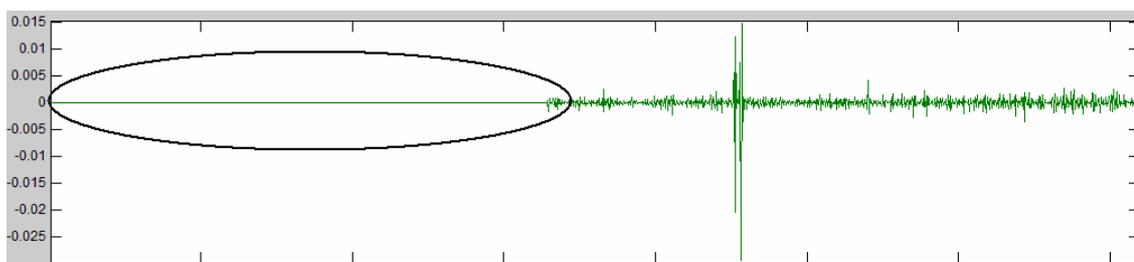
Foi definido o seguinte critério: sempre que forem identificados 10 elementos consecutivos com valor menor do que  $10^{-14}$  será considerado o início/término de uma lacuna.



(a) Curva de carga original



(b) Componente de aproximação *Wavelet* da curva de carga original



(c) Componente de detalhe  $D$

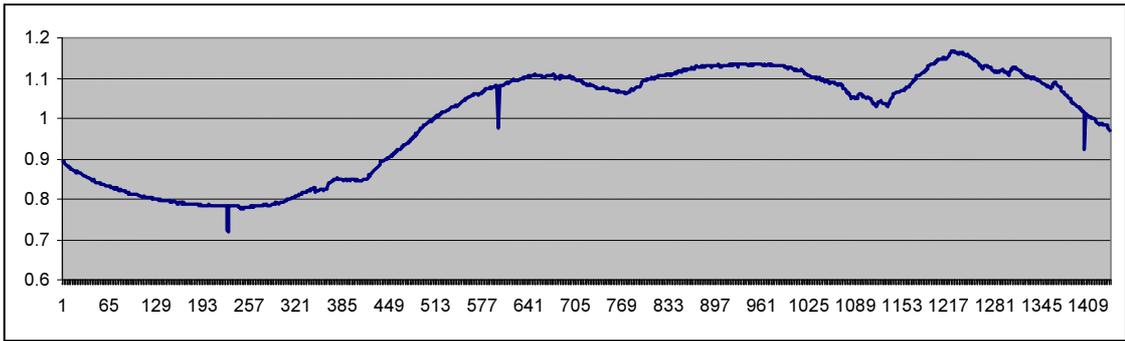
**Figura 4. 7** Decomposição de curva de carga com lacuna

#### 4.4.2. Correção de Falhas por descontinuidade

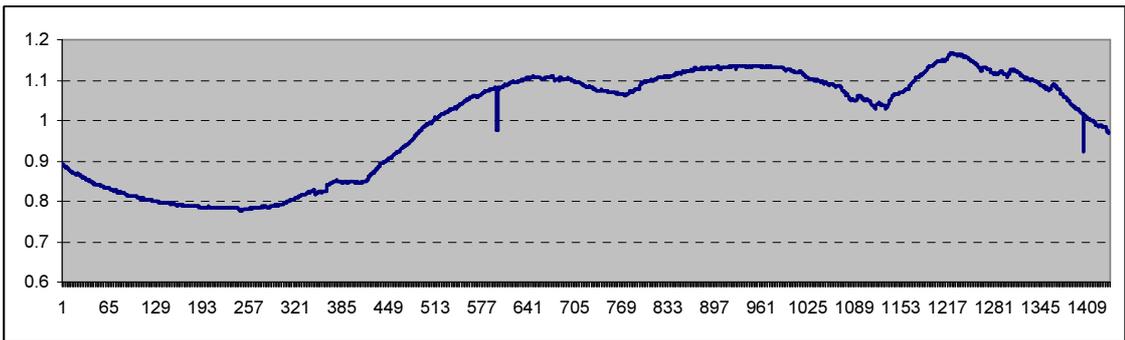
Nesta seção vamos apresentar uma abordagem para corrigir falhas por descontinuidade identificadas da forma como descrito na seção 4.4.1. O objetivo é aplicar o método LOESS na curva de carga. Cada ponto descontínuo é substituído pelo ponto suavizado determinado pelo LOESS. Deve-se aplicar novamente o procedimento de identificação de falhas por até 10 tentativas para garantir a efetiva correção.

Esta abordagem é simples e eficiente, porém se a correção não for alcançada, a curva será classificada como curva com falha.

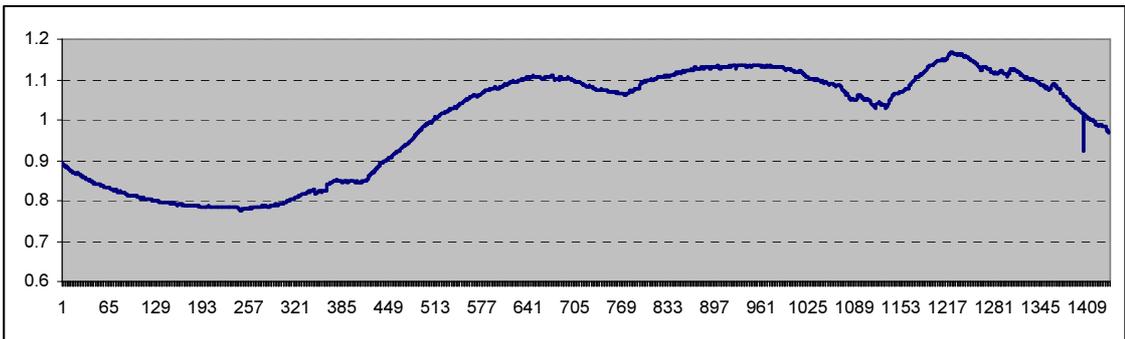
Na figura 4.8 é ilustrada a correção da curva de carga primeiramente apresentada na figura 4.6.



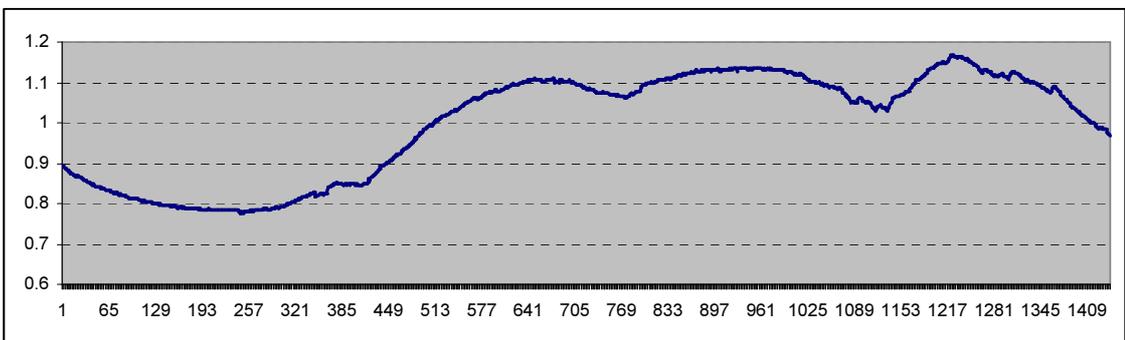
(a) Curva de carga original



(b) Correção 1



(c) Correção 2



(d) Correção 3

**Figura 4. 8 Correção de descontinuidade**

A figura 4.8(a) mostra a curva de carga original. Na primeira correção pelo LOESS, o primeiro ponto de descontinuidade é corrigido como mostra a figura 4.8(b). Repetindo-se o processo os demais pontos são corrigidos nas figuras 4.8 (c) e (d).

A curva corrigida é apresentada na figura 4.8(d).

# Capítulo 5

## Análise de Agrupamento

A análise de agrupamentos (*cluster analysis*) abrange uma ampla variedade de técnicas de classificação não supervisionada, usadas para identificar uma estrutura natural de agrupamentos (*cluster*) em objetos multidimensionais, ou seja, visa resolver o problema de como particionar um conjunto de  $N$  objetos em  $k$  classes mutuamente exclusivas, de tal forma que os objetos em uma mesma classe sejam semelhantes entre si, mas díspares dos objetos pertencentes às outras classes.

Atualmente grande quantidade de informação está disponível ao conhecimento humano. É inviável tratar toda informação de forma isolada e individualizada. A análise de agrupamentos fornece mecanismos para classificar esta informação em grupos (categorias) e assim ser possível tratá-las em conjunto e não individualmente.

No contexto dos problemas de análise de agrupamentos, não se conhece a estrutura dos grupos. Normalmente, são conhecidos o conjunto de observações, uma relação de semelhança entre elementos e o número de grupos. Quando o número de grupos não é conhecido, trata-se do problema de agrupamento automático. Ao utilizar uma das técnicas disponíveis em análise de agrupamentos, será possível identificar uma estrutura de grupos e também para cada elemento identificar sua pertinência em um grupo.

A problemática de agrupamento já foi colocada na Grécia antiga por Aristóteles. Nos tempos modernos, uma das primeiras referências é de TRYON (1939). A aplicabilidade deste assunto é eclética. Há uma vasta quantidade de aplicações que carecem de técnicas de agrupamento em áreas como: biologia, entomologia, psicologia, educação, economia, pesquisa de mercado, geologia, planejamento urbano e regional, etc.

## 5.1. Definição dos dados de entrada

Seja  $X = \{x_1, x_2, \dots, x_n\}$  um conjunto de  $n$  observações. Para cada  $x \in X$ , supõem-se que existam  $p$  características mensuráveis. Define-se a matriz  $M$  ( $p \times n$ ) tal que

$$M = \begin{bmatrix} m_{11} & \cdots & \cdots & m_{1n} \\ m_{21} & \ddots & & m_{2n} \\ \vdots & & \ddots & \vdots \\ m_{p1} & \cdots & \cdots & m_{pn} \end{bmatrix} \quad (5.1)$$

## 5.2. Problema de análise de agrupamentos

Dado  $X = \{x_1, x_2, \dots, x_n\}$  um conjunto de  $n$  observações e um número inteiro  $m$ , tal que  $m \leq n$ . Determinar uma partição  $P_m = \{G_1, G_2, \dots, G_m\}$  de  $X$  contendo  $m$  grupos, satisfazendo as seguintes condições:

- (i) Para cada  $x_i \in X$ ; então  $x_i \in G_i \subset P_m$  se e somente se  $x_i \notin (P_m - G_i)$ .
- (ii) Para  $x_i, x_j \in X$ ; então  $x_i, x_j \in G_k \subset P_m$  se e somente se  $x_i$  e  $x_j$  são semelhantes
- (iii) Para  $x_i, x_j \in X$ ;  $G_i \neq G_j$  então  $x_i \in G_i, x_j \in G_j$  se e somente se  $x_i$  e  $x_j$  são díspares

Desta forma, a solução de um problema de análise de agrupamentos é encontrar uma partição de  $X$  que otimize uma função objetivo  $f(P)$ . Esta função quantifica a semelhança dentro do grupo e também a disparidade entre grupos para que as condições (ii) e (iii) sejam satisfeitas.

A semelhança e similaridade entre elementos de  $X$  são definidas por critérios ou métricas, que são divididas entre medidas de similaridade e medidas de dissimilaridade.

Nas medidas de similaridade quanto mais semelhante são os elementos, maior é o valor da medida e da mesma forma quanto menos semelhantes são os elementos, menor é o valor da medida. Um exemplo de medida de similaridade é o coeficiente de *Gower*.

Já nas medidas de dissimilaridade quanto mais semelhante são os elementos, menor é o valor da medida e quanto menos semelhantes são os elementos, maior é o valor da medida. A função de distância é um exemplo de medida de dissimilaridade.

### 5.2.1. Função de distância

É uma medida de dissimilaridade entre dois elementos.

Seja a função

$$d : \mathfrak{R}^n \rightarrow \mathfrak{R}^+$$

$$d(x, y) = d_{xy}$$

que satisfaz as seguintes condições:

(i)  $d_{xy} = 0$  se e somente se  $x = y$

(ii)  $d_{xy} = d_{yx}$

(iii)  $d_{xy} \leq d_{xz} + d_{zy}$  (*desigualdade triangular*)

são funções de distância as métricas de Minkowski ou normas  $l_p$

$$d_{xy} = d_p(x, y) = \|x - y\|_p = \sqrt[p]{\sum_{k=1}^n |x_k - y_k|^p}, \forall p \geq 1 \quad (5.2)$$

No caso de  $p=1$ , tem-se a norma 1 e para  $p=2$ , a norma euclidiana.

### 5.2.2. Coeficientes de similaridade

Seja a função

$$s_{ij} : \mathfrak{R}^n \rightarrow \mathfrak{R}^+$$

$$s(x_i, y_j) = s_{ij}$$

que satisfaz as seguintes condições:

(i)  $0 \leq s_{ij} \leq 1$  se  $x_i \neq y_j$

(ii)  $s_{ij} = 1$

(iii)  $s_{ij} = s_{ji}$

é denominada coeficiente de similaridade.

Pelas condições intrínsecas acima estabelecidas para coeficiente de similaridade, existem várias possibilidades de definição dos mesmos. Uma frequentemente usada é o coeficiente de *Gower*.

### 5.2.2.1. Coeficiente de Gower

Seja a função

$$s_{ij} : X \subset \mathfrak{R}^n \rightarrow \mathfrak{R}^+$$

$$s(x_i, x_j) = 1 - \frac{\|x_i - x_j\|}{R} \quad (5.3)$$

$$R = \text{MAX}_{x_r, x_s \in X} \{\|x_r - x_s\|\}$$

É possível construir uma métrica através de um coeficiente de similaridade. Utilizando o coeficiente de *Gower*, tem-se a seguinte métrica:

$$d_{ij} = \sqrt{1 - s_{ij}} \quad (5.4)$$

### 5.2.3. Medidas de dispersão interna

Um dos objetivos da análise de agrupamentos é a quantificação de semelhança e a reunião de elementos semelhantes em um mesmo grupo. Medir a dispersão dos elementos de um determinado grupo é aferir a semelhança entre eles. Uma medida de dispersão é a soma dos quadrados dentro do grupo, também denominada inércia intra grupos.

Soma dos quadrados dentro do grupo

Dado um grupo  $g_i$  com  $n_i$  elementos. Seja a função

$$W_i : X \subset \mathfrak{R}^n \rightarrow \mathfrak{R}^+$$

$$W_i(g_i) = \sum_{i=1}^{n_i} \sqrt{\|x_i - \bar{x}_i\|^2}$$

$$\bar{x}_i = \frac{1}{n_i} \sum_{i=1}^{n_i} x_i \quad (5.5)$$

$\bar{x}_i$  é definida como a média ou centróide do grupo.

$W_i$  é a soma dos quadrados das distâncias euclidianas entre cada elemento do grupo e a centróide do mesmo. Pode-se também calcular  $W_i$ , independentemente do cálculo da centróide.

$$W_i(g_i) = \sum_{i=1}^{n_i} \sum_{j=1}^{i-1} \sqrt{\|x_i - x_j\|} \quad (5.6)$$

## 5.2.4. Função objetivo

A solução de um problema de agrupamento é muitas vezes obtida pela resolução de um problema de otimização da seguinte forma:

Minimizar  $f(P_m)$

Onde  $P_m$  uma partição viável do conjunto de amostras.

Um critério natural e frequentemente usado em análise de agrupamento consiste em minimizar a soma dos quadrados dentro dos grupos.

## 5.2.5. Métodos de análise de agrupamentos

Estão disponíveis uma grande variedade de métodos para resolução de problemas de análise de agrupamentos. Pode-se categorizá-los em três categorias: métodos hierárquicos, métodos de realocação iterativa e métodos de programação matemática.

### 5.2.5.1. Métodos hierárquicos

Nestes métodos, inicialmente todos os  $n$  elementos da amostra são agrupados em  $n$  grupos contendo apenas um elemento cada. Depois, inicia-se processo iterativo em que selecionam-se os dois grupos mais semelhantes e estes são unificados em um novo

grupo. No final da  $k$ -ésima iteração, haverá uma partição contendo  $(n-k)$  grupos. O processo continua até que  $k=n-1$ , quando são obtidos dois grupos.

A cada passo do processo é definido uma hierarquia, pois um grupo definido em uma iteração é a união de dois grupos da iteração anterior.

A determinação do número de grupos  $m$ , é um problema em aberto dentro da análise de agrupamentos.

### **5.2.5.2. Métodos de realocação iterativa**

Dado o número de grupos  $m$  e uma partição qualquer  $P_m^0 = \{g^0_1, g^0_2, \dots, g^0_m\}$ . Calcula-se inicialmente os centros de cada grupo e cria-se uma nova partição  $P_m^1 = \{g^1_1, g^1_2, \dots, g^1_m\}$ , alocando cada elemento ao centro mais próximo. Repete-se o processo até que um teste de convergência seja satisfeito. Não há garantia que a partição encontrada seja ótima.

### **5.2.5.3. Métodos de programação matemática**

Nos métodos hierarquizados e de realocação iterativa, não há garantia de encontrar a partição ótima. A busca pela partição ótima utilizando a enumeração completa não é prática.

Os métodos de programação matemática assumem um determinado número de grupos  $e$ , essencialmente, buscam a otimização de uma função objetivo para medir a semelhança no interior de cada grupo  $e$  / ou a separação entre os grupos.

Existem vários métodos de programação matemática, destacando-se os de programação dinâmica, teoria dos grafos e programação inteira.

#### **5.2.5.3.1. Métodos da Soma dos Mínimos Quadrados**

Dentre os critérios usados na análise de agrupamento, o critério mais natural, intuitivo e freqüentemente adotado é o agrupamento por soma dos mínimos quadrados (MSSC), utilizando-se a medida de dispersão soma dos quadrados dentro do grupo. Este critério corresponde à minimização da soma dos quadrados das distâncias das

observações do grupo, ou equivalentemente, a minimização do grupo pela soma de quadrados. É um critério para ambos os objetivos: semelhança e não semelhança. De acordo com o Teorema de Huygens, minimizando a inércia de uma partição do grupo (semelhança dentro do grupo) é equivalente a maximizar a inércia entre grupos (separação entre grupos).

A análise de agrupamento pela formulação da soma dos mínimos quadrados (MSSC) produz um problema matemático de otimização global. É não diferenciável e também não convexo, com um grande número de mínimos locais.

Um novo método de programação matemática para agrupamento proposto por XAVIER (2010,2011) utiliza suavização hiperbólica (HSCM). Basicamente, o método executa a suavização da formulação de agrupamento não diferenciável min-min-soma. Esta técnica foi desenvolvida pensando numa adaptação do método de penalização hiperbólica originalmente introduzido por XAVIER (1982). Pela suavização, que significa fundamentalmente a substituição de um problema de nível intrinsecamente não diferenciável em dois níveis, por um problema irrestrito e continuamente diferenciável.

A idéia básica é a partição do conjunto de observações em duas partes não sobrepostas. Usando uma apresentação conceitual, o primeiro conjunto corresponde aos pontos de observação relativamente próximos a dois ou mais centróides. Este conjunto de observações denominadas pontos de borda podem ser gerenciados usando a abordagem de suavização apresentada anteriormente. O segundo conjunto corresponde a pontos de observação muito mais perto de um único centróide em comparação com os outros. Este conjunto de observações nomeadas pontos gravitacionais é gerido de uma forma direta e simples, proporcionando um desempenho muito mais rápido.

#### **5.2.5.4. Agrupamento pela soma dos mínimos quadrados**

Seja  $S = \{s_1, \dots, s_m\}$  um conjunto de  $m$  padrões ou observações em um espaço euclidiano de dimensão  $n$  a serem agrupados em um determinado número  $q$  de grupos disjuntos. Para formular o problema original de agrupamento como um problema de *min-sum-min*, procede-se da seguinte forma. Seja  $x_i, i = 1, \dots, q$  os centróides dos grupos, onde cada  $x_i \in R^n$ . O conjunto dessas coordenadas centróides será representado por  $X \in R^{nq}$ . Dado um ponto  $s_j \in S$ , inicialmente calcula-se a distância euclidiana de  $s_j$  ao *centroid* em  $X$  que está mais próximo.

Isso é dada por

$$z_j = \min_{x_i \in X} \|s_j - x_i\|_2 \quad (5.7)$$

O cálculo de  $z_j$  é exemplificado na figura 5.1. Nessa figura,  $s_j$  representa um observação conectada aos *centroids*  $x_1, x_2, x_3$  e  $x_4$ . A distância  $z_j$  está destacado em verde.

A metrica freqüentemente utilizada para aferir a qualidade de associação de um grupo a uma posição específica do centróide  $q$  é dada pela soma dos quadrados dessas distâncias, o que determina o problema MSSC (*Minimum Sum of Squares Clustering*):

$$\text{Minimizar} \quad \sum_{j=1}^m z_j^2 \quad (5.8)$$

sujeito a

$$z_j = \min_{i=1, \dots, q} \|s_j - x_i\|_2, j = 1, \dots, m$$

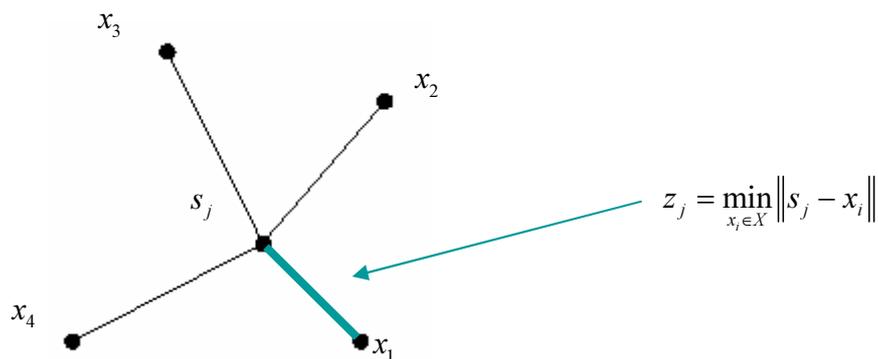


Figura 5. 1 Cálculo de  $Z_j$  ao *centroid* mais próximo

# Capítulo 6

## HSCM

Neste capítulo, é apresentado o método HSCM - *Hyperbolic Smoothing Clustering Method*, XAVIER(2010,2011).

Dado o problema MSSC (*Minimum Sum of Squares Clustering*):

$$\text{Minimizar} \quad \sum_{j=1}^m z_j^2 \quad (6.1)$$

sujeito a

$$z_j = \min_{i=1,\dots,q} \|s_j - x_i\|_2, j = 1, \dots, m$$

Considerando esta definição: cada  $z_j$  deve necessariamente satisfazer o seguinte conjunto de inequações:

$$z_j - \|s_j - x_i\|_2 \leq 0, i = 1, \dots, q \quad (6.2)$$

Substituindo as restrições de igualdade do problema 6.1 pelas inequações 6.2, produz-se o seguinte problema relaxado:

$$\text{Minimizar} \quad \sum_{j=1}^m z_j^2 \quad (6.3)$$

sujeito a

$$z_j - \|s_j - x_i\|_2 \leq 0, j = 1, \dots, m; i = 1, \dots, q$$

Desde que a variável  $z_j$  não é limitada inferiormente, a solução ótima do problema relaxado será  $z_j=0, j=1,\dots,m$ . A fim de obter o desejado, devemos modificar o problema 6.3. Primeiro devemos fazer  $\varphi(y)$  denotado por  $\max\{0,y\}$  e então observar que, do conjunto de inequações em (6.3) é dado por

$$\sum_{i=1}^q \varphi(z_j - \|s_j - x_i\|_2) = 0, j = 1, \dots, m \quad (6.4)$$

Usando (6.4) no lugar do conjunto das restrições de desigualdades em (6.3), teremos um problema equivalente mantendo a indesejável propriedade que  $z_j, j=1, \dots, m$  não possui limite inferior. Considerando que a função objetivo do problema (6.3) irá forçar cada  $z_j, j=1, \dots, m$ , para baixo, podemos pensar em um delimitador pela inclusão de uma perturbação  $\varepsilon$  em (6.4), conforme abaixo

$$\text{Minimizar} \quad \sum_{j=1}^m z_j^2 \quad (6.5)$$

sujeito a

$$\sum_{i=1}^q \varphi(z_j - \|s_j - x_i\|_2) \geq \varepsilon, j = 1, \dots, m$$

para  $\varepsilon > 0$ .

Desde que a viabilidade do problema (6.1) é o limite de (6.5) quando  $\varepsilon \rightarrow 0+$  podemos considerar resolver (6.1) por uma seqüência de problemas parecidos com (6.5) em que os valores de  $\varepsilon$  seguem uma trajetória decrescente e se aproximam de zero.

Analisando o problema (6.5), a definição da função  $\varphi$  impõe uma estrutura extremamente rígida e não diferenciável, o que torna a sua solução computacional muito difícil. Em vista disso, o método numérico adotado para resolver o problema (5.7), utiliza uma abordagem de suavização. A partir dessa perspectiva, vamos definir a função:

$$\phi(y, \tau) = (y + \sqrt{y^2 + \tau^2}) / 2 \quad (6.6)$$

para  $y$  real e  $\tau > 0$

A função  $\phi$  possui as seguintes propriedades:

(a)  $\phi(y, \tau) > \phi(y) \forall \tau > 0$

(b)  $\lim_{\tau \rightarrow 0} \phi(y, \tau) = \phi(y)$

(c)  $\phi(y, \tau)$  é um incremento convexo  $C^\infty$  na função de variável  $y$

Substituindo a função  $\varphi(y)$  no problema (6.5) pela função  $\phi$  definida em (6.6), obtém-se a seguinte formulação:

$$\text{Minimizar} \quad \sum_{j=1}^m z_j^2 \quad (6.7)$$

sujeito a

$$\sum_{i=1}^q \phi(z_j - \|s_j - x_i\|_2, \tau) \geq \varepsilon, j = 1, \dots, m$$

A distância euclideana  $\|s_j - x_i\|$  é o único componente não diferenciável do problema (6.7). Então para obter um problema completamente diferenciável é ainda necessário suavizar. Para esse propósito, vamos definir a função:

$$\theta(s_j, x_i, \gamma) = \sqrt{\sum_{l=1}^n (s_{j,l}^1 - x_{i,l}^1)^2 + \gamma^2} \quad (6.8)$$

para  $\gamma > 0$

A função  $\theta$  possui as seguintes propriedades

$$(a) \quad \lim_{\gamma \rightarrow 0} \theta(s_j, x_i, \gamma) = \|s_j - x_i\|_2$$

$$(b) \quad \theta \in C^\infty$$

pelo uso da função  $\theta$  no lugar da distancia  $\|s_j - x_i\|_2$ , o seguinte problema diferenciável é obtido.

$$\text{Minimizar} \quad \sum_{j=1}^m z_j^2 \quad (6.9)$$

sujeito a

$$\sum_{i=1}^q \phi(z_j - \theta(s_j, x_i, \gamma), \tau) \geq \varepsilon, j = 1, \dots, m$$

Então as propriedades da função  $\phi$  e da função  $\theta$  permitem procurar a solução do problema (6.5) pela solução de uma seqüência de subproblemas semelhantes a (6.9), produzidos pelo decréscimo dos parâmetros  $\gamma \rightarrow 0$ ,  $\tau \rightarrow 0$  e  $\varepsilon \rightarrow 0$

Desde que  $z_j \geq 0, j=1, \dots, m$ , na função objetivo, o processo de minimização irá reduzindo estes valores. Por outro lado, dado qualquer conjunto de centróides  $x_i, i=1, \dots, q$  usando a propriedade (c) da função de suavização hiperbólica  $\phi$ , as restrições do problema (6.9) são funções monótonas crescentes em  $z_j$ . Estas restrições serão ativas e o problema (6.9) irá ser equivalente ao problema:

$$\text{Minimizar} \quad \sum_{j=1}^m z_j^2 \quad (6.10)$$

sujeito a

$$h_j(z_j, x) = \sum_{i=1}^q \phi(z_j - \theta(s_j, x_i, \gamma), \tau) - \varepsilon = 0, j = 1, \dots, m$$

A dimensão do espaço da variável de domínio do problema (6.10) é  $(nq + m)$ . Como, em geral, o valor do parâmetro  $m$ , a cardinalidade do conjunto  $S$  das  $s_j$  observações, é grande, o problema (6.10) tem um grande número de variáveis. No entanto, tem uma estrutura separável, porque cada variável  $z_j$ , aparece apenas em uma restrição de igualdade. Portanto, como a derivada parcial de  $h(z_j, x)$  com respeito a  $z_j, j = 1, \dots, m$  não é igual a zero, pode-se utilizar o teorema da função implícita para calcular cada componente  $z_j, j = 1, \dots, m$  em função dos centróides  $x_i, i = 1, \dots, q$ . Desta forma segue, o problema irrestrito.

$$\text{Minimizar.} f(x) = \sum_{j=1}^m z_j(x)^2 \quad (6.11)$$

onde cada  $z_j(x)$  resulta do calculo do zero em cada equação

$$h_j(z_j, x) = \sum_{i=1}^q \phi(z_j - \theta(s_j, x_i, \gamma), \tau) - \varepsilon = 0, j = 1, \dots, m \quad (6.12)$$

Usando a propriedade (c) da função de suavização hiperbólica, cada termo  $\phi$  acima é estritamente crescente com a variável  $z_j$  e portanto a equação possui um único zero.

Novamente, devido ao teorema da função implícita, as funções  $z_j(x)$  possuem todas as derivadas com relação as variáveis  $x_i, i = 1, \dots, q$  e dessa forma, é possível calcular o gradiente da função objetivo do problema (6.11).

$$\nabla f(x) = \sum_{j=1}^m 2z_j(x) \cdot \nabla z_j(x), \quad (6.13)$$

onde

$$\nabla z_j(x) = \frac{-\nabla h_j(z_j, x)}{\frac{\partial h_j(z_j, x)}{\partial z_j}}, \quad (6.14)$$

$\nabla h_j(z_j, x)$  e  $\frac{\partial h_j(z_j, x)}{\partial z_j}$  são obtidos das equações (6.6), (6.8) e (6.12).

Desta forma é fácil resolver o problema (6.11) utilizando qualquer método baseado em derivadas de primeira ordem. Finalmente deve-se enfatizar que o problema (6.11) é definido em um espaço  $(nq)$ -dimensional. É um problema pequeno desde que o número de clusters  $q$  é em geral pequeno para aplicações reais.

A solução do problema original de agrupamento pode ser obtido pelo uso do Algoritmo de agrupamento com suavização hiperbólica (HSC), descrito abaixo:

Inicialização: Escolha valores iniciais para  $x^0, \gamma^0, \tau^1$  e  $\varepsilon^1$

Escolha valores  $0 < \rho_1 < 1$ ;  $0 < \rho_2 < 1$ ;  $0 < \rho_3 < 1$  e seja  $k=1$

Principal: Repita até que critério de parada seja satisfeito  
 Resolva problema (6.11) com  $\gamma=\gamma^k, \tau=\tau^k$  e  $\varepsilon = \varepsilon^k$   
 começando com um ponto inicial  $x^{k-1}$  e seja  $x^k$  a solução  
 obtida  
 seja  $\gamma^{k+1}=\rho_1 \gamma^k$ ;  $\tau^{k+1}=\rho_2 \tau^k$ ;  $\varepsilon^{k+1}=\rho_3 \varepsilon^k$ ;  $k:=k+1$

Este é um método de suavização, a solução do problema de clusterização é obtida em teoria pela solução de uma seqüência infinita de problemas de otimização. No algoritmo HSC, cada problema minimizado é irrestrito e de baixa dimensão.

Os parâmetros  $\rho_1, \rho_2$  e  $\rho_3$  respectivamente fazem com que  $\gamma, \tau$  e  $\varepsilon$  se aproximem de zero, portanto, as restrições dos subproblemas como dadas em (6.9) tendem para (6.5). Dessa forma, o algoritmo faz com que  $\varepsilon$  vá para zero, assim, de

forma simultânea, o problema resolvido (6.5) aproxima-se gradualmente do problema original (6.1).

Segundo MENDES (2012), o método HSCM é **paralelizável**. A idéia principal está no cálculo da função  $f(x)$  do problema irrestrito (6.11). Conforme a equação (6.15), cada cálculo de  $z_j(x)^2$  depende apenas de  $x$ , ou seja, da solução da iteração atual do método.

$$\text{Minimizar. } f(x) = \sum_{j=1}^m z_j(x)^2 = z_1(x)^2 + z_2(x)^2 + \dots + z_m(x)^2 \quad (6.15)$$

A equação (6.12) representa o cálculo de cada  $z_j(x)^2$ . O somatório de (6.15) possui  $m$  parcelas independentes, correspondentes as  $m$  observações. Em problemas práticos, o número de observações é grande. Os cálculos das  $m$  raízes de (6.12) são independentes entre si. Dessa forma, o cálculo dessas  $m$  raízes pode ser feito em paralelo.

De forma semelhante pode-se aplicar a mesma metodologia para o cálculo do gradiente de  $f$  em (6.13)

## Capítulo 7

# Experimentos Computacionais

Neste capítulo serão apresentados os experimentos computacionais aferidos. Foram utilizados dados reais do Sistema Integrado Nacional, fornecidos pelo ONS - Operador Nacional do Sistema Elétrico. Especificamente, são dados diários de carga a nível nacional, entre os anos de 2005 e 2006, totalizando 730 curvas de carga.

A discretização de cada curva de carga diária é feita minuto a minuto, ou seja, cada curva é um vetor de 1.440 posições. Em suma, são dados de grande dimensão.

Os procedimentos da metodologia de filtragem descritos no capítulo 4 foram implementados em linguagem C++, a implementação para a *TDW* foi utilizada uma biblioteca *opensource* denominada *wavelet1d* C++ 1D/2D DWT IMPLEMENTATION FOR WIN32 AND LINUX [2013?]

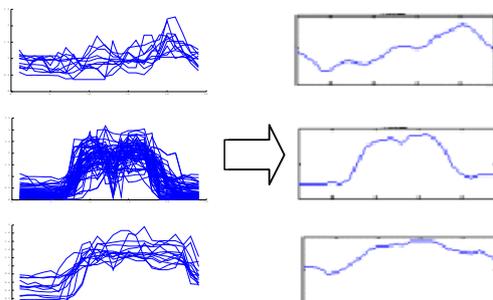
Para a análise de agrupamento, foi utilizado o método de suavização hiperbólica XAVIER (2010, 2011) denominado HSCM. O método HSCM é detalhado no capítulo 6.

O procedimento de identificação de falhas, descrito na seção 4.4.1, analisou 730 curvas. Foram identificadas 239 curvas sem falhas e em 491 curvas foram identificadas falhas. Dessas curvas com falhas, 115 curvas apresentaram falhas por perda de dados e 376 curvas com falhas por descontinuidade.

As curvas com falhas foram analisadas pelo procedimento de correção de falha, descrito na seção 4.4.2. Foram então corrigidas 490 curvas e apenas 1 curva não foi possível efetivar a correção, que foi eliminada do conjunto de dados.

No contexto da análise de agrupamentos, o objetivo é identificar perfis típicos das curvas de cargas que não apresentaram falhas ou, então, que sofreram processos de correção. Essas tipologias são de suprema importância para previsão de carga. Como será apresentado no próximo capítulo, também apresentam aplicação para correção de falhas de curva de carga ocorridas por perda de dados.

Na análise de agrupamento aplicada em curvas de carga, deseja-se agrupar curvas semelhantes e aferir curvas típicas que representam o comportamento de cada grupo. Essas curvas típicas são também denominadas na literatura como *centroids*.



**Figura 7. 1 Grupos e seus *centroids***

Na figura 7.1, no lado esquerdo, são apresentados três conjuntos com número expressivo de observações. No lado direito, estão apresentados os *centroids*, que basicamente representam um comportamento médio de cada um dos três conjuntos.

O método HSCM foi executado com número de grupos variando de 2 a 10. As seguintes médias foram calculadas: BSS (*Between Sum of Squares*), WSS (*Within Sum of Squares*), CS (*Compactness and Separation*) e o tempo de processamento. Os resultados são mostrados na tabela 7.1. Foi utilizado um computador PC com 512MB de memória RAM e com processador Intel Celeron de 2.7GHz.

Numero de Grupos	BSS	WSS	CS	Tempo (minutos)
2	51%	49%	1.696,00	1
3	62%	38%	1.317,20	2
4	68%	32%	1.122,00	3
5	75%	25%	858,80	5
6	77%	23%	802,90	16
7	78%	22%	767,00	13
8	79%	21%	723,10	23
9	83%	17%	589,80	36
10	84%	16%	556,80	51

**Tabela 7. 1 Medidas de validação e tempo de processamento**

O termo WSS é a variância dentro dos grupos e corresponde ao valor da função objetivo do problema 6.1. O termo BSS é a variância entre os grupos. Pelo teorema de Huygens a soma desses 2 termos é igual à variância total do conjunto de observações, que é constante. O termo CS é a razão entre WSS e a menor distancia entre dois *centroids*. Quanto menor for CS mais homogêos são os grupos e maior separação entre grupos.

Na tabela 7.1, o valor mínimo da medida CS indica que a solução com dez aglomerados é a melhor.

## 7.1. Apresentação dos grupos

As figuras de 7.2(a) até 7.2(j) mostram os 10 agrupamento aferidos. Nessas figuras os *centroids* são destacados em verde.

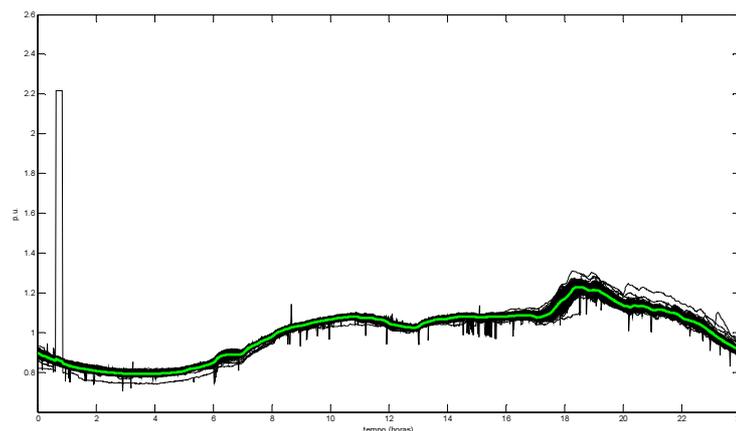
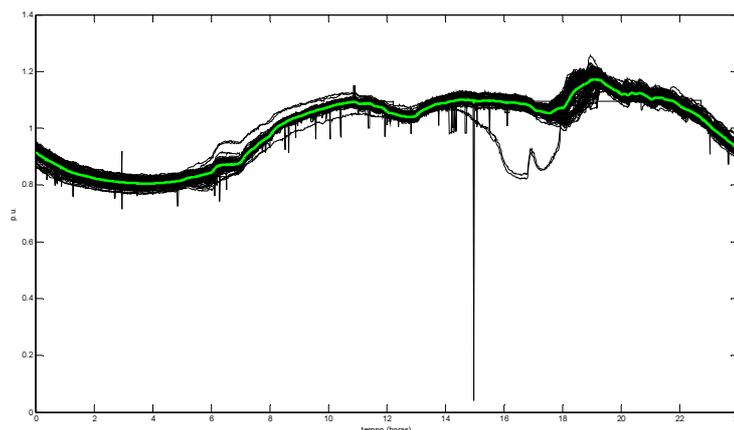


Figura 7. 2 (a) Grupo 1 - Dias Úteis – Frequência = 161

O grupo 1 basicamente é constituído por observações associadas aos dias úteis da semana, no intervalo, de terça à sexta-feira, conforme a composição: 25,47% de terças-feiras, 26,09% de quartas-feiras, 22,98% de quintas-feiras, 24,22% de sextas-feiras. De uma forma diferenciada dos outros dias da semana, a participação de segundas-feiras é rarefeita, somente 1,24%.

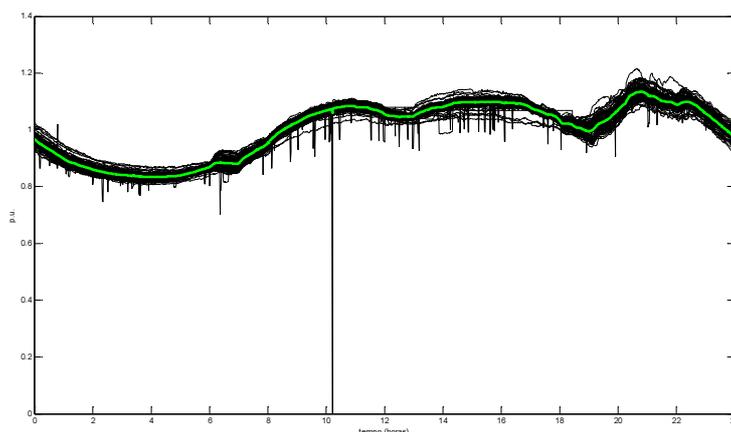
No contexto das estações do ano, o grupo 1 é constituído por observações de primavera, outono e inverno, conforme a composição: 4,97% de primavera, 40,99% de outono e 54,04% de inverno, sem qualquer observação ocorrida no verão. Trata-se de um grupo “outono-inverno”.



**Figura 7. 2 (b) Grupo 2 – Dias Úteis – Frequência = 118**

O grupo 2 basicamente é constituído por observações associadas aos dias úteis da semana, no intervalo de terça à sexta-feira, conforme a composição: 22,88% de terças-feiras, 22,88% de quartas-feiras, 22,88% de quintas-feiras e 23,73% de sextas-feiras. De uma forma diferenciada dos outros dias da semana, a participação de segundas-feiras é pequena 7,63%.

No contexto das estações do ano, o grupo 2 é constituído por observações de todas as estações do ano, conforme a composição: 22,03% de primavera, 31,36% de verão, 30,51% de outono e 16,10% inverno, ou seja, é um grupo bem distribuído entre as quatro estações do ano.

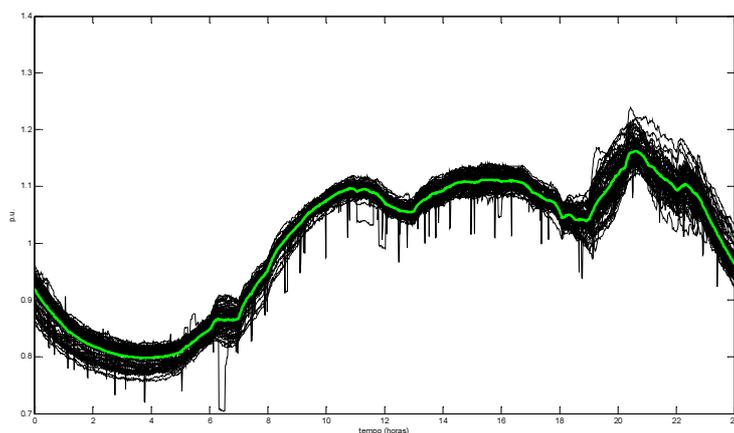


**Figura 7. 2 (c) Grupo 3 – Dias Úteis – Frequência = 86**

O grupo 3 basicamente é constituído por observações associadas aos dias úteis da semana, no intervalo de terça à sexta-feira, conforme a composição: 17,44% de terças-

feiras, 20,93% de quartas-feiras, 27,91% de quintas-feiras e 29,07% de sextas-feiras. De uma forma diferenciada dos outros dias da semana, a participação de segundas-feiras é pequena 4,65%.

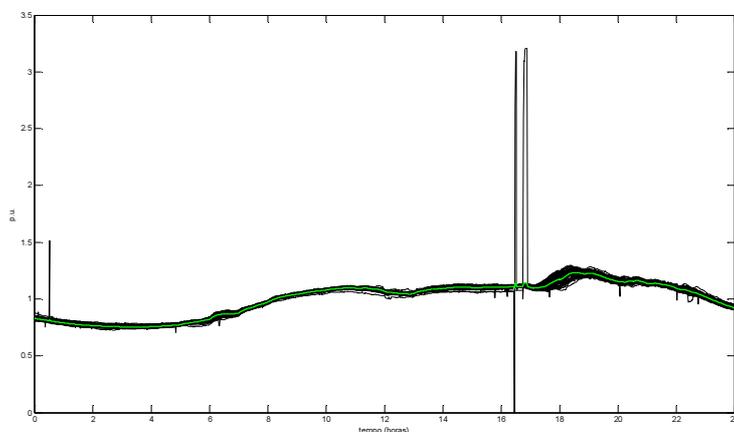
No contexto das estações do ano, o grupo 3 é constituído somente por observações pertencentes ao verão e à primavera, conforme a composição: 22,09% de primavera e 77,91% de verão. Trata-se de um grupo basicamente “verão”.



**Figura 7. 2 (d) Grupo 4 – Dias Úteis – Frequência = 72**

O grupo 4 é constituído por observações associadas a todos os dias úteis da semana, conforme a composição: 38,89% de segundas-feiras, 20,83% de terças-feiras, 16,67% de quartas-feiras, 12,50% de quintas-feiras e 11,11% de sextas-feiras.

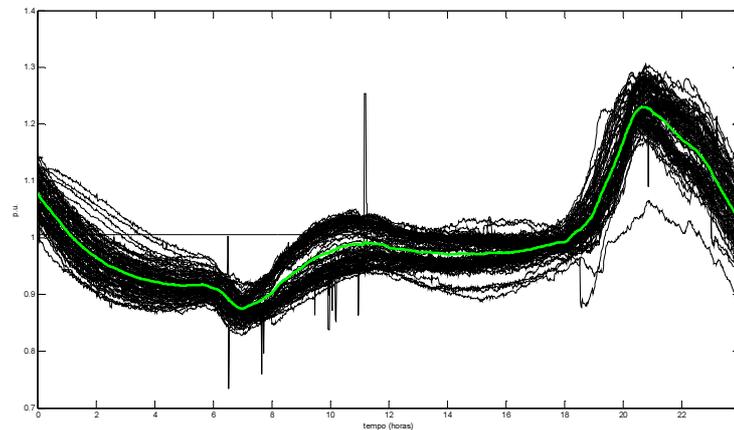
No contexto das estações do ano, o grupo 4 é constituído por somente observações pertencentes ao verão e à primavera, conforme a composição: 83,33% de primavera e 16,67% de verão. Trata-se de um grupo basicamente “verão”.



**Figura 7. 2 (e) Grupo 5 – Segunda-feira Típica – Frequência = 60**

O grupo 5 é constituído principalmente por observações associadas as segundas-feiras, totalizando 95,00% das observações. Em contrapartida os demais dias úteis da semana apresentam pouca participação, conforme a composição: 1,67% de terças-feiras, 1,67% de quintas-feiras e 1,67% de sextas-feiras.

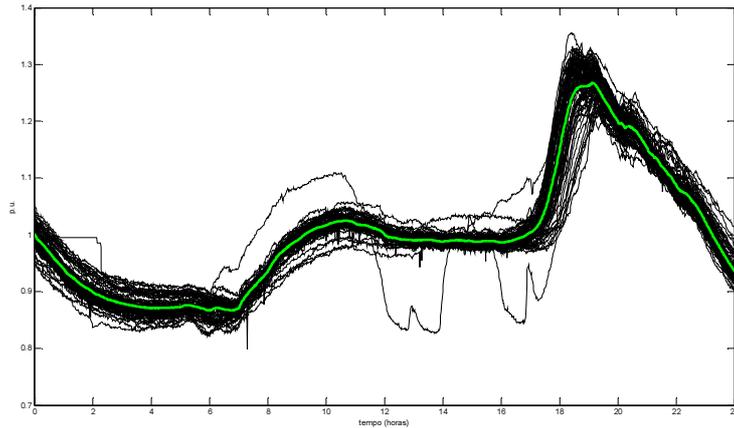
No contexto das estações do ano, o grupo 5 é constituído por observações pertencentes à primavera, ao outono e ao inverno, conforme a composição: 15,00% de primavera, 43,33% de outono e 41,67% de inverno. Trata-se de um grupo “outono-inverno”.



**Figura 7. 2 (f) Grupo 6 – Sábados e Domingos – Frequência = 76**

O grupo 6 é constituído por observações associadas aos sábados e aos domingos, conforme a composição: 46,74% de sábados e 44,74% de domingo. De uma forma diferenciada dos dias de final de semana, os dias úteis apresentam pouca contribuição: 2,63% de segunda-feira, 2,63% de terça-feira e 3,95% de quarta-feira.

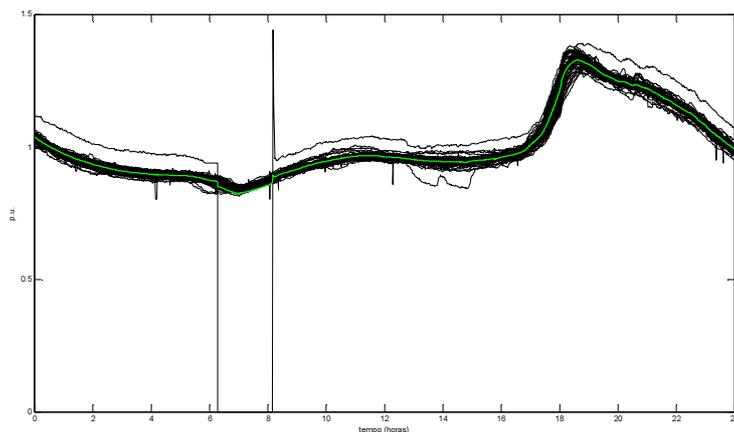
No contexto das estações do ano, o grupo 6 é constituído por observações ocorridas na primavera e no verão, conforme a composição: 46,05% de dias primavera, 53,95% de verão. Trata-se de um grupo “primavera-verão”.



**Figura 7.2 (g) Grupo 7 – Sábado Típico – Frequência =75**

O grupo 7 é principalmente constituído por observações associadas aos sábados, totalizando 93,33% das observações. De uma forma diferenciada, as demais observações associadas aos dias úteis da semana apresentam pouca contribuição, conforme a composição: 1,33% de segundas-feiras, 1,33% de terças-feiras, 1,33% de quartas-feiras, 1,33% de quintas-feiras e 1,33% de sextas-feiras.

No contexto das estações do ano, o grupo 7 é constituído por observações de todas as estações do ano: 14,67% de dias primavera, 10,67% de dias de verão, 37,33% de dias de outono e 37,33 dias de inverno, ou seja, é um grupo bem distribuído entre as quatro estações.

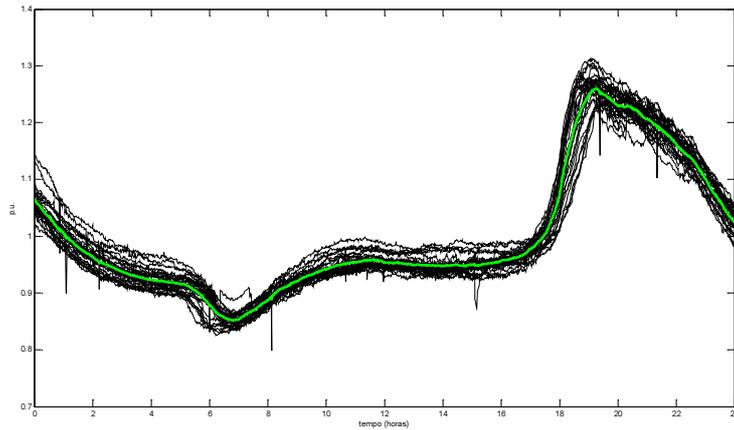


**Figura 7.2 (h) Grupo 8 – Domingo Típico – Frequência = 53**

O grupo 8 é principalmente constituído por observações associadas aos domingos, totalizando 92,40% das observações. De uma forma diferenciada, as demais observações associadas aos dias da semana segunda-feira e quinta-feira apresentam

pouca contribuição, conforme a composição: 1,89% de segundas-feiras e 5,66% de quintas-feiras.

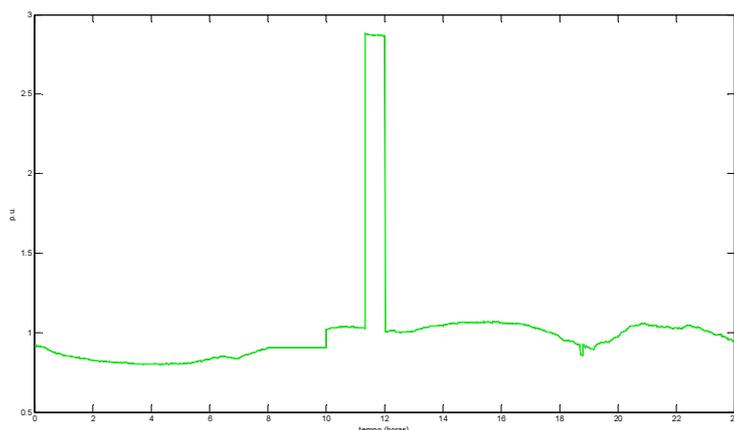
No contexto das estações do ano, o grupo 8 é constituído por observações ocorridas na primavera, no outono e no inverno: 9,43% de dias primavera, 41,51% de dias de outono e 49,06 de dias de inverno. Trata-se basicamente de um grupo “outono-inverno”.



**Figura 7. 2 (i) Grupo 9 – Domingo Típico – Frequência = 28**

O grupo 9 é principalmente constituído por observações associadas aos domingos, totalizando 78,57% das observações. De uma forma diferenciada, as demais observações associadas aos dias da semana de terça à sexta-feira apresentam pouca contribuição, conforme a composição: 3,57% de terças-feiras, 3,57% de quartas-feiras, 7,14% de quintas-feiras e 7,14% de sextas-feiras.

No contexto das estações do ano, o grupo 9 é constituído por observações de todas as estações do ano: 25,00% de dias primavera, 35,71% de dias de verão, 35,71% de dias de outono e 3,57 dias de inverno. Trata-se basicamente de um grupo “primavera-verão-outono”.



**Figura 7. 2 (j) Grupo 10 – observação anômala – Freqüência = 1**

O grupo 10 apresenta apenas uma curva com lacunas muito expressivas e completamente destoantes das demais observações. Em resumo, trata-se de uma observação dominada por essas lacunas. Ademais, um grupo com apenas uma observação definitivamente não possui qualquer significância para efeitos práticos. Essa situação ressalta a importância do tratamento de dados de curva de carga especificados no capítulo 4 para eliminação desse efeito deletério.

Na tabela 7.2 Resumo, são mostradas as participações percentuais dos dias da semana nos grupos.

Grupo	Freqüência	Domingo	Segunda	Terça	Quarta	Quinta	Sexta	Sábado
1	161	0.00	1.24	25.47	26.09	22.98	24.22	0.00
2	118	0.00	7.63	22.88	22.88	22.88	23.73	0.00
3	86	0.00	4.65	17.44	20.93	27.91	29.07	0.00
4	72	0.00	38.89	20.83	16.67	12.50	11.11	0.00
5	60	0.00	95.00	1.67	0.00	1.67	1.67	0.00
6	76	44.74	2.63	2.63	3.95	0.00	0.00	46.05
7	75	0.00	1.33	1.33	1.33	1.33	1.33	93.33
8	53	92.40	1.89	0.00	0.00	5.66	0.00	0.00
9	28	78.57	0.00	3.57	3.57	7.14	7.14	0.00
10	1	0.00	0.00	100.00	0.00	0.00	0.00	0.00

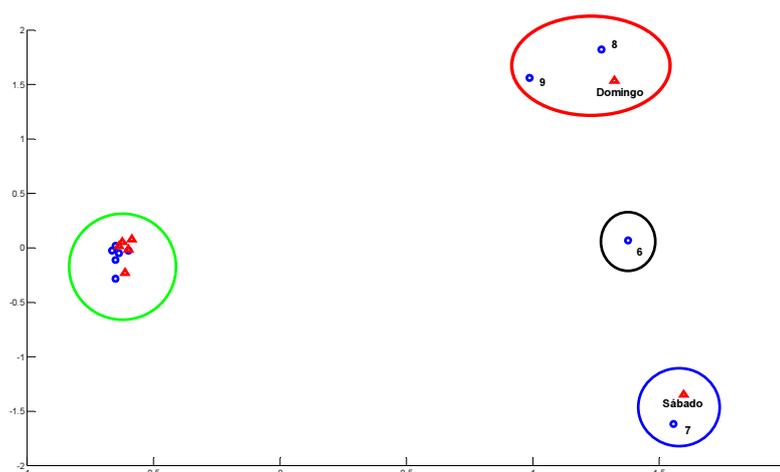
**Tabela 7. 2 Resumo - Participação percentual dos dias de semana nos grupos**

Os grupos 1, 2, 3, 4 e 5 são formados por diferentes combinações de dias úteis. Os três primeiros grupos são do tipo “terça-quarta-quinta-sexta”. O grupo 4 é mais “forte segunda”, enquanto o grupo 5 é quase puramente “segunda”. O grupo 6 é formado por curvas de carga de final de semana. No grupo 7 há alta frequência de sábados e os grupos 8 e 9 apresentam grande concentração de domingos. O grupo 10 apresenta apenas uma observação com lacunas. Essa situação ressalta a importância do tratamento de dados de curva de carga especificados no capítulo 4 para eliminação desse efeito deletério.

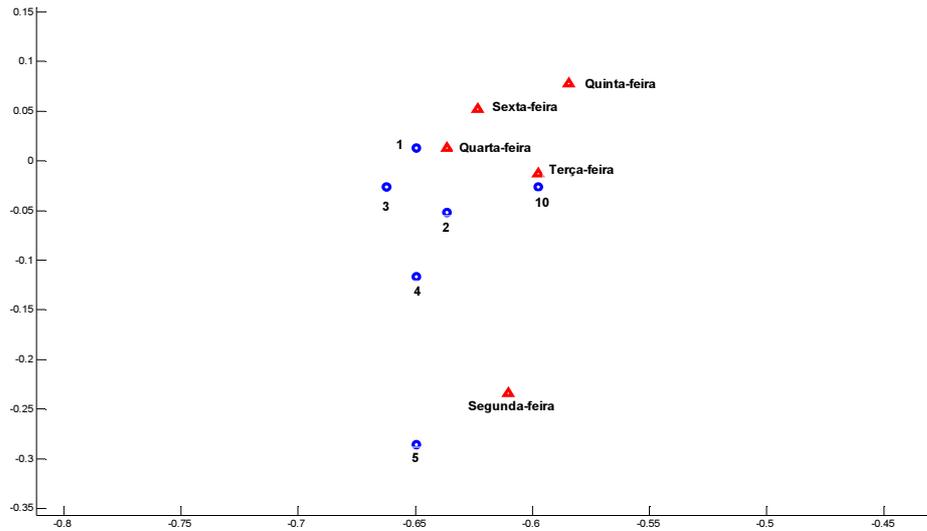
As associações entre grupos e dias da semana mostradas na tabela 7.2 são melhores visualizadas com o gráfico da figura 7.3, obtido pela análise de correspondência simples (Lebart et al, 2004).

Análise de correspondência é uma técnica de análise exploratória de dados adequada para analisar tabelas de duas ou mais entradas, levando em conta algumas medidas de correspondência entre linhas e colunas. Basicamente, converte uma matriz de dados não negativos em um tipo particular de representação gráfica em que as linhas e colunas da matriz são simultaneamente representadas em dimensão reduzida, isto é, por pontos no gráfico. Esse gráfico é denominado mapa percentual. Esse mapa concentra a maior parte do grau de associação entre grupos e dias da semana.

Na figura 7.3, os pontos marcados com bolas azuis representam cada um dos 10 grupos e os marcados em triângulos em vermelho representam os sete tipos de dia da semana.



**Figura 7. 3 - Mapa Percentual – Demonstra associação entre grupos e dias da semana. Círculo verde ampliado na figura 7.4**



**Figura 7. 4 Mapa Percentual – Ampliação do círculo verde da figura 7.3**

Na figura 7.3, os sábados e os domingos estão em regiões distintas entre si. Respectivamente demarcados com círculos azul e vermelho. Isso caracteriza que as assinaturas dos dias úteis, dos sábados e dos domingos são distintas entre si.

Enquanto isso nessa mesma figura 7.3, os dias úteis estão todos concentrados numa mesma região, demarcada por círculo verde, do mapa percentual. A figura 7.4 basicamente amplia esse círculo verde.

Dessa forma, os dias úteis apresentam comportamento semelhante entre si e distintos do comportamento típico do final de semana.

Vemos na figura 7.4, que os grupos 1, 2, 3, 4 e 5 marcados com círculos azuis estão situados na região destacada do círculo verde da figura 7.3. Consoante aos valores mostrados na tabela 7.2 Resumo, vemos na figura 7.4 que os dias úteis, representados por triângulos vermelhos, também se situam nessa mesma região destacada, sendo que terças, quartas, quintas e sextas-feiras bem próximas entre si e segundas-feiras um pouco mais distante.

Na figura 7.3, no círculo azul temos somente um grupo, o grupo 7, que fica em posição próxima ao sábado, demarcado por círculo azul. Em harmonia com a tabela 7.2 Resumo, o grupo 7 apresenta o percentual de 93,33 % de sábados.

Na figura 7.3, as posições dos grupos 8 e 9 estão próximas à posição do domingo, conforme demarcação com círculo vermelho. Dessa forma, a nossa análise mostra a

grande presença de curvas de domingo nesses grupos. Respectivamente apresentam o percentual de 92,40% e 78,67%, conforme a tabela 7.2.

Na figura 7.3, o grupo 6 é demarcado com círculo preto. Conforme a tabela 7.2, apresenta os percentuais: 46,05% de sábado e 44,74% de domingo. Assim, conforme o mapa percentual da figura 7.3, o grupo 6 está aproximadamente a mesma distância do sábado e do domingo.

Na tabela 7.3, são mostradas as participações percentuais das estações do ano nos grupos.

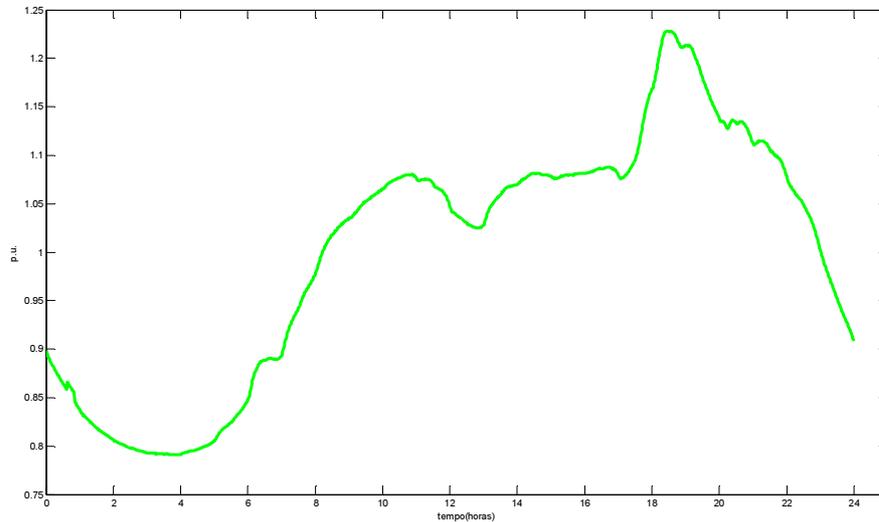
Grupo	Frequência	Primavera	Verão	Outono	Inverno
1	161	4.97	0.00	40.99	54.04
2	118	22.03	31.36	30.51	16.10
3	86	22.09	77.91	0.00	0.00
4	72	83.33	16.67	0.00	0.00
5	60	15.00	0.00	43.33	41.67
6	76	46.05	53.95	0.00	0.00
7	75	14.67	10.67	37.33	37.33
8	53	9.43	0.00	41.51	49.06
9	28	25.00	35.71	35.71	3.57
10	1	0.00	100.00	0.00	0.00

**Tabela 7. 3 Participação Percentual das estações do ano nos grupos**

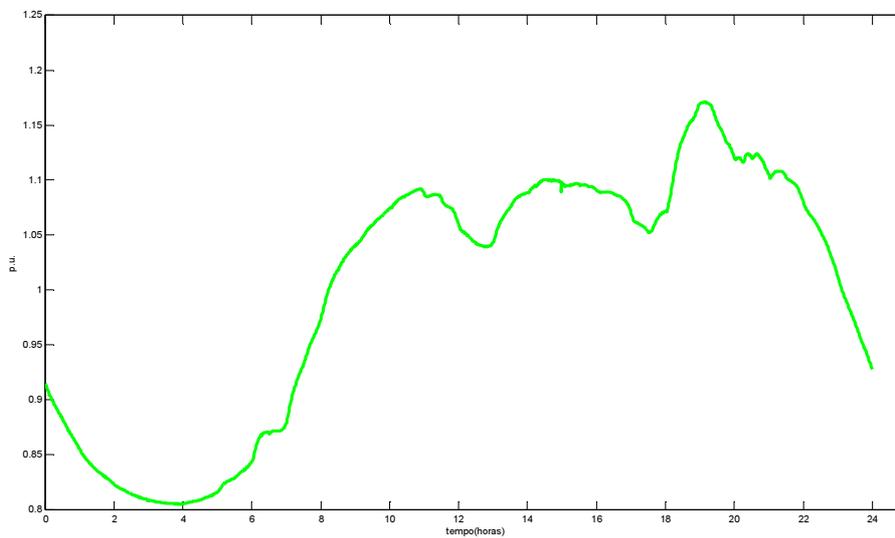
Os percentuais de participação das estações do ano nos grupos são informações relevantes para o entendimento do efeito climático no consumo relativo de energia elétrica. Posteriormente será demonstrado o efeito positivo, em termos de economia de energia, da aplicação do horário de verão.

## 7.2. Apresentação dos *Centroids*

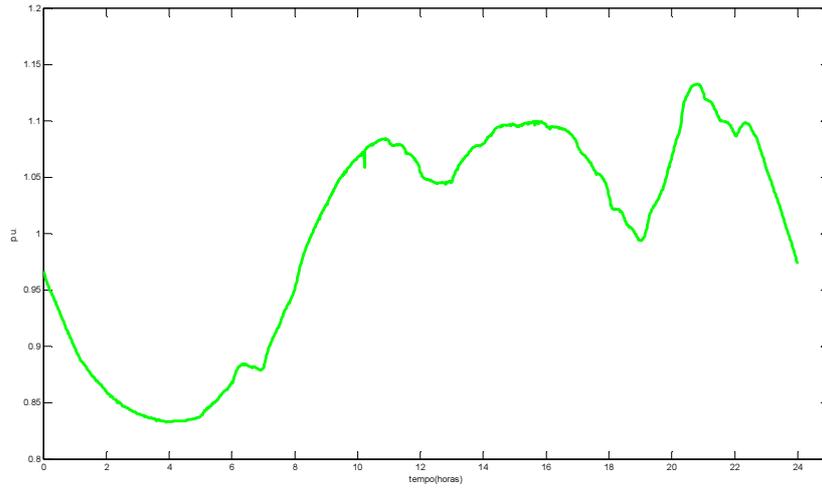
As figuras de 7.5(a) até 7.5(j) mostram os 10 *centroids* aferidos. Os *centroids* representam cada um dos comportamentos típicos da carga.



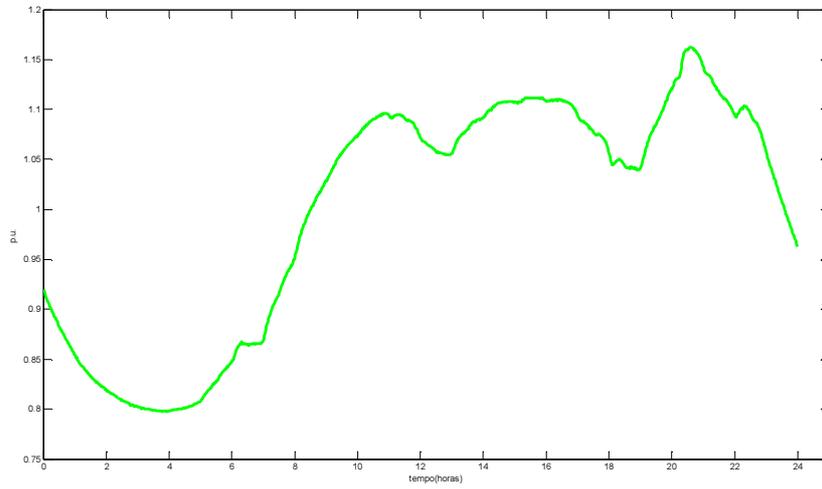
**Figura 7.5 (a) Centroid referente ao grupo 1 – Dias Úteis – Frequência = 161**



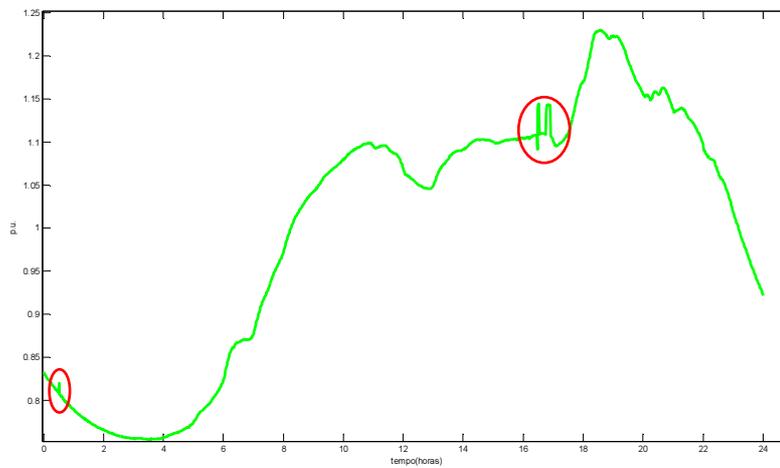
**Figura 7.5 (b) Centroid referente ao grupo 2 – Dias Úteis – Frequência = 118**



**Figura 7.5(c) Centroid referente ao grupo 3 – Dias Úteis – Freqüência = 86**



**Figura 7.5 (d) Centroid referente ao grupo 4 – Dias Úteis – Freqüência = 72**

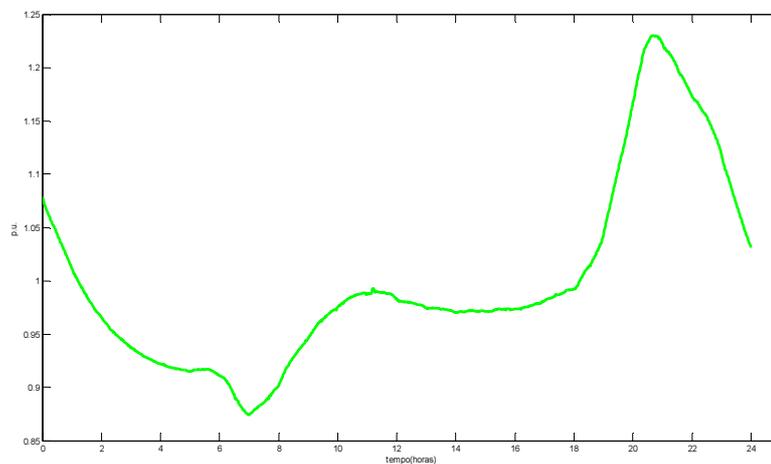


**Figura 7.5(e) Centroid referente ao grupo 5 – Segunda-feira Típica – Freqüência = 60**

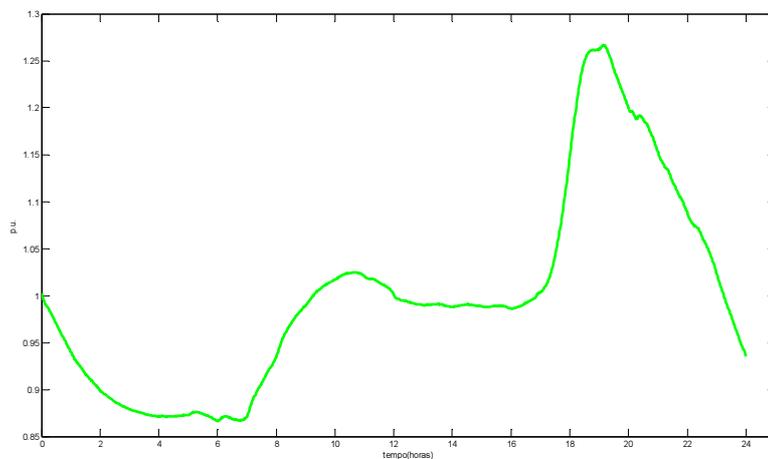
Os *centroids* dos grupos 1, 2, 3, 4 e 5 visualizados respectivamente nas figuras de 7.5(a) até 7.5(e) apresentam aproximadamente mínimo global próximo às 04:00hs. Em seguida, o consumo relativo aumenta até próximo ao meio-dia. Posteriormente ocorrem dois mínimos locais, em horários próximo às 14:00hs e às 16:00hs. O consumo relativo torna a subir atingindo o pico (máximo global) em torno das 20:00hs. Esse é um comportamento esperado para dias úteis e confirmado pelos percentuais da tabela 7.2

Resumo.

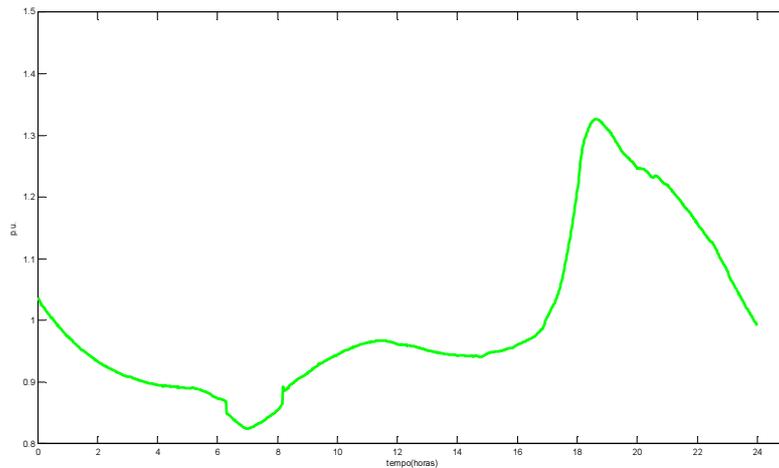
O *centroid* do grupo 5 apresenta falhas por descontinuidade. Na figura 7.5(e), essas falhas são assinaladas com círculos vermelho.



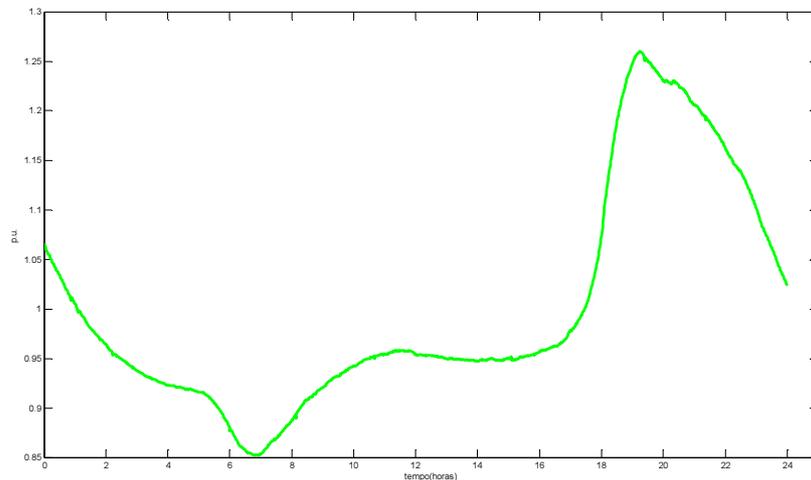
**Figura 7.5 (f) Centroid referente ao grupo 6 – Sábado e Domingo – Freqüência = 76**



**Figura 7.5 (g) Centroid referente ao grupo 7 – Sábado Típico – Freqüência = 75**

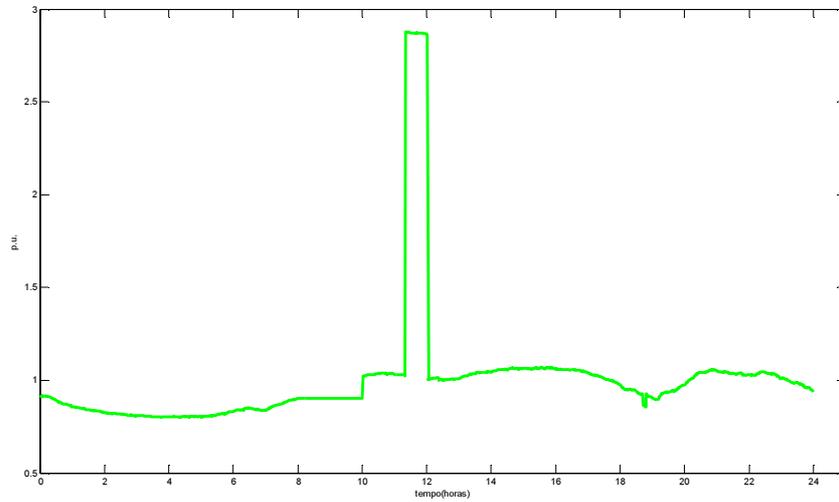


**Figura 7.5 (h) Centroid referente ao grupo 8 – Domingo Típico – Frequência = 53**



**Figura 7.5 (i) Centroid referente ao grupo 9 – Domingo Típico – Frequência = 28**

Os *centroids* dos grupos 6, 7, 8 e 9 visualizados respectivamente nas figuras de 7.5(f) até 7.5(i) apresentam aproximadamente mínimo global próximo às 07:00hs. Em seguida, o consumo relativo aumenta até próximo ao meio-dia. Posteriormente o consumo relativo é aproximadamente constante e dessa forma **não** ocorrem mínimos locais no intervalo de 11:00hs até às 17:00hs. O consumo relativo torna a subir atingindo o pico (máximo global) em torno das 20:00hs. Esse é um comportamento esperado para finais de semana e confirmado pelos percentuais da tabela 7.2 Resumo.

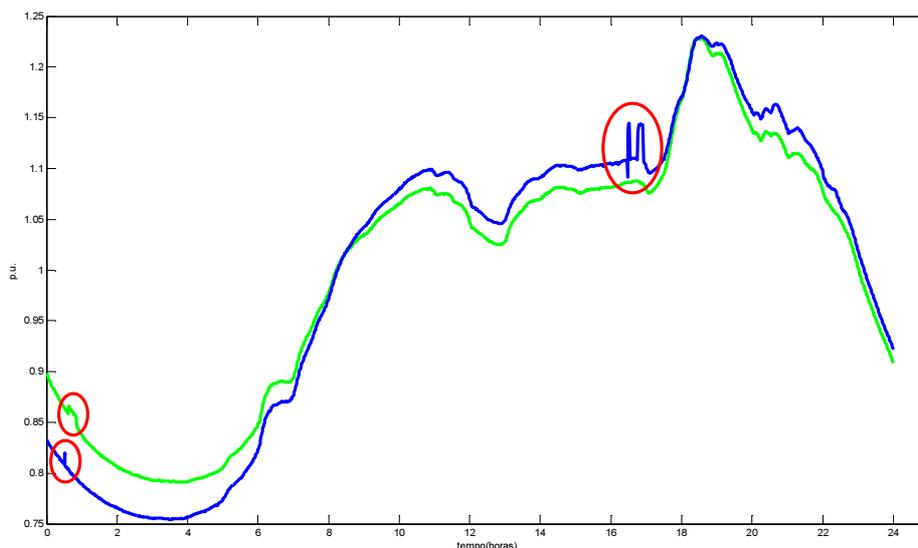


**Figura 7.5 (j) Centroid referente ao grupo 10 – observação anômala – Frequência = 1**

O grupo 10 apresenta apenas uma curva com lacunas e assim não tem qualquer expressividade. Dessa forma, *centroid* e observação se confundem. Um grupo com apenas uma observação definitivamente não possui qualquer significância para efeitos práticos. Essa situação ressalta a importância do tratamento de dados de curva de carga especificados no capítulo 4 para eliminação desse efeito deletério.

### 7.3. Comparação entre *centroids*

As figuras de 7.6 até 7.8 mostram gráficos comparativos entre *centroids* de diferentes grupos. A figura 7.6 mostra o efeito nocivo das falhas na identificação de grupos e seus respectivos *centroids*. As figuras 7.7 e 7.8 mostram os efeitos do *horário-de-verão* no consumo relativo de energia.



**Figura 7. 6 – Comparativo entre o *centroid* do grupo 1 desenhado em verde e o *centroid* do grupo 5 desenhado na cor azul. As falhas por descontinuidades são destacadas com círculos vermelhos.**

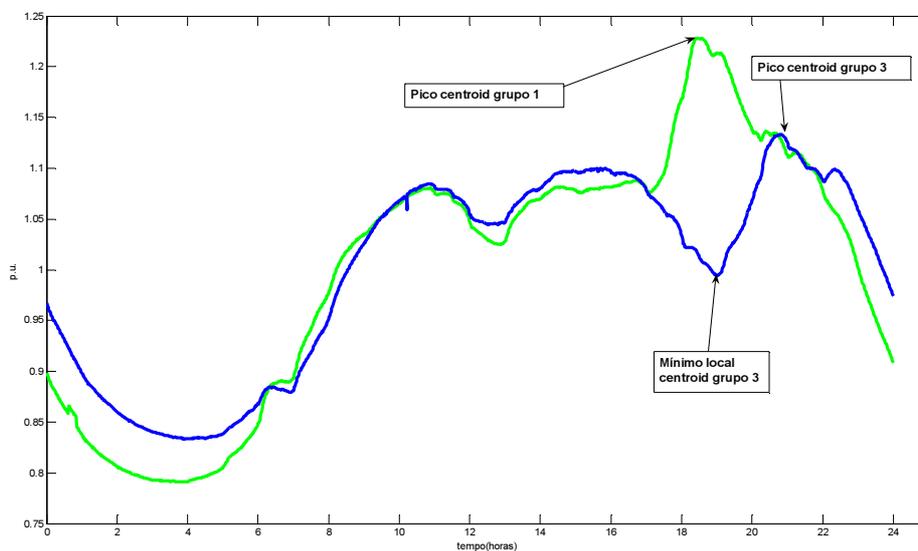
Na figura 7.6, o *centroid* do grupo 1 é apresentado em verde. Apresenta uma distribuição aproximadamente equitativa entre os dias úteis, com exceção das segundas-feiras. O percentual de segundas-feiras é de apenas 1,24%. Em contrapartida, o *centroid* do grupo 5 em azul, apresenta percentual de 95,00% de segundas-feiras.

Na análise da figura 7.6, fica evidente que o consumo relativo de energia tem certo retardo nas segundas-feiras em relação aos demais dias da semana.

Os *centroids* dos grupos 1 e 5, embora diferentes tem alguma similaridade. Deve-se ressaltar a presença deletéria de falhas no *centroid* do grupo 5 às 16:00hs que justifica as maiores diferenças.

Assim, essas falhas distorcem os cálculos de distancia e, conseqüentemente, elevam incorretamente o consumo. Dessa forma, uma vez mais, fica patente a imprescindibilidade da etapa de filtragem.

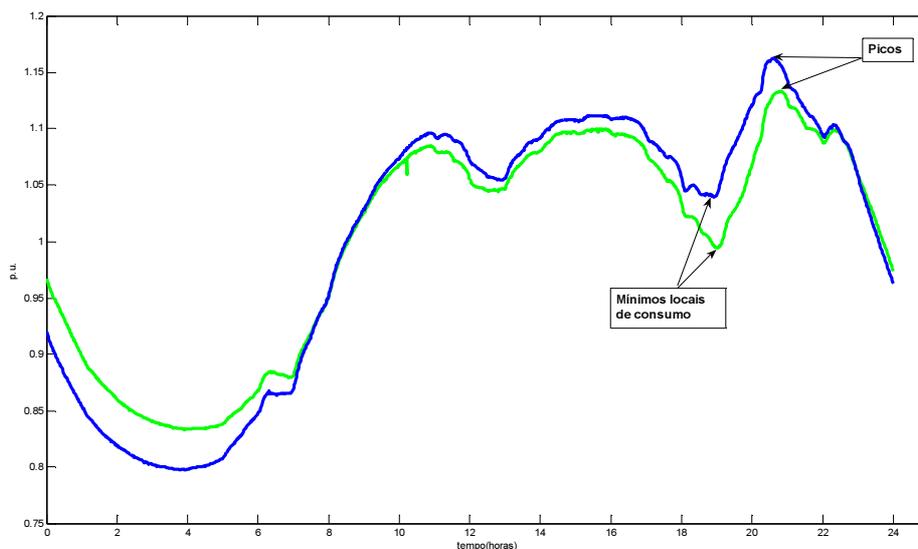
A seguir, as figuras 7.7 e 7.8 mostram os efeitos do *horário-de-verão* no consumo relativo de energia.



**Figura 7. 7 – Comparativo entre o centroid do grupo 1 desenhado em verde e o centroid do grupo 3 desenhado em azul com 77,91% de dias de verão. O efeito de redução do pico do grupo 3 é ocasionado pelo *horário-de-verão*.**

Na figura 7.7, o *centroid* do grupo 1, em verde, não apresenta dias de verão. O *centroid* do grupo 3, em azul, apresenta 77,91% de dias de verão. Depois das 16:00hs o consumo relativo do *centroid* do grupo 3 diminui, enquanto o consumo relativo do *centroid* do grupo 1 aumenta. No horário próximo ao pico do *centroid* do grupo 1, o *centroid* do grupo 3 está próximo ao mínimo local de consumo relativo. Isso explicita a ocorrência concreta de economia de energia elétrica, proporcionada pelo *horário-de-verão*, entre às 16:00hs e 20:00hs.

Na figura 7.8, os *centroids* dos grupos 3 e 4 apresentam uma distribuição aproximadamente equitativa entre os dias úteis. Ambos apresentam curvas semelhantes e pertencem apenas as estações do ano: primavera e verão. O maior percentual de primavera corresponde ao grupo 4, totalizando 83,33%. O maior percentual de verão corresponde ao grupo 3, totalizando 77,91%.



**Figura 7. 8 – Comparativo entre o centroid do grupo 3 desenhado em verde com 77,91% de dias de verão e o centroid do grupo 4 desenhado na cor azul com 83,33% de dias de primavera. O consumo reduzido do grupo 3 em relação ao grupo 4 é ocasionado pelo *horário-de-verão*.**

Na figura 7.8, que faz comparação entre os grupos 3 e 4, o *centroid* do grupo 3 é desenhado em verde enquanto o *centroid* do grupo 4 é desenhado em azul. Os picos ocorrem em horários próximos e de forma semelhante, também com os mínimos locais de consumo. O consumo relativo do grupo 3 é menor do que o consumo relativo do grupo 4. Essa diferença no consumo relativo é devido ao percentual mais elevado de verão no grupo 3 em relação ao grupo 4. Esse fenômeno demonstra empiricamente o efeito de redução do consumo de energia na período de 18:00hs às 22:00hs, o que mostra o acerto de implantação da política de *horário-de-verão*.

## 7.4. Síntese

Ao final desse capítulo, devemos em suma registrar que, de forma geral, os resultados computacionais da determinação das tipologias de curva de carga corroboram o conhecimento apriorístico sobre os perfis das curvas de carga, nos diferentes dias da semana.

# Capítulo 8

## Aplicações de tipologias

Neste capítulo, são apresentados exemplos de aplicações das tipologias aferidas pela análise de agrupamentos em curvas de carga.

Primeiramente é apresentada uma metodologia para correção de lacunas. Nesse contexto, são utilizadas as tipologias aferidas pelo conjunto de técnicas apresentadas no presente trabalho. O processo ocorre em duas fases: *(i)* Identificação, para cada curva com falha, o perfil típico que apresenta o melhor ajuste, ou seja, apresenta a menor distancia; *(ii)* Correção das lacunas utilizando projeção por regressão linear do perfil identificado na lacuna.

As tipologias de curvas de carga são de relevante importância para previsão de carga. Independentemente do método de previsão adotado, estas tipologias servem para a modelagem de tendências e de sazonalidades da carga. Isto é fundamental para que a previsão seja feita com qualidade e, conseqüentemente, o atendimento da demanda de energia elétrica seja atendido com segurança.

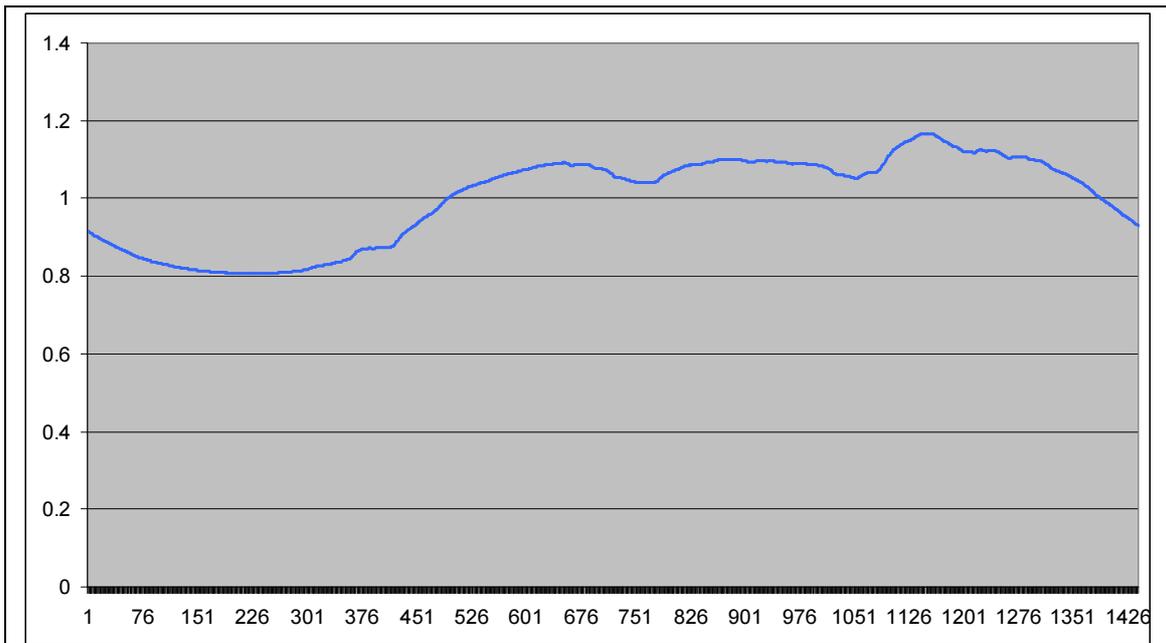
### 8.1. Correção de Lacunas

Nesta seção será mostrada uma abordagem para corrigir lacunas, que são ilustradas na figura 4.7(a). Como os dados a serem corrigidos foram perdidos, não é possível corrigi-los utilizando informação da própria curva, como aplicamos na seção 4.4 que trata de identificação e correção de falhas.

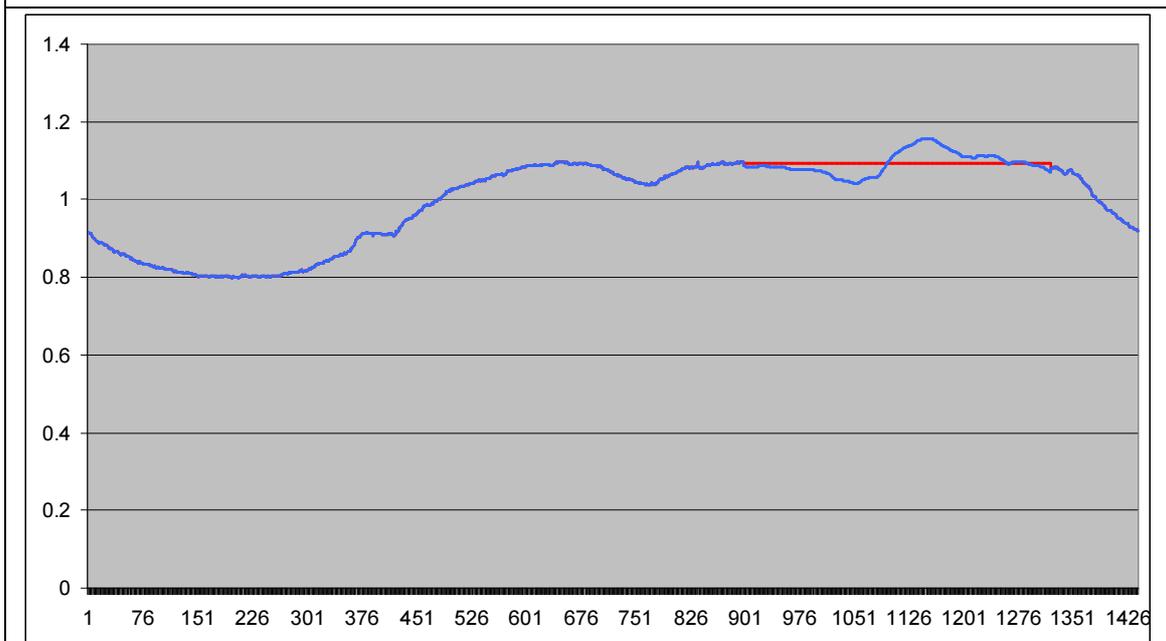
Partimos da hipótese que a informação para correção de lacunas está no conjunto de curvas usadas para determinação das tipologias, desde que o comportamento regular da carga é sazonal, conforme evidências empíricas de longos anos. Para podermos identificar qual curva é o melhor ajuste para corrigir uma lacuna inteira, utilizamos a análise de agrupamento. Nesse contexto, a análise de agrupamento determina tipologias que serão utilizadas na correção de lacunas. Assim, as tipologias são fundamentais para a efetiva correção de lacunas. Conforme descrito no capítulo 6, foi utilizado o método de suavização hiperbólica XAVIER (2010).

O processo ocorre em duas fases: (i) Identificação, para cada curva com falha, do perfil típico que apresenta o melhor ajuste, ou seja, apresenta a menor distância euclidiana; (ii) Correção da lacuna utilizando projeção por regressão linear do perfil identificado na lacuna.

Na figura 8.1 é demonstrado a tipologia utilizada para o preenchimento da lacuna. Na figura 8.2, são apresentadas: a curva original e a curva mesma curva com lacunas preenchidas.



**Figura 8.1** Tipologia utilizada para preenchimento das lacunas



**Figura 8.2** Curva de carga com lacunas preenchidas

## 8.2.Previsão de carga

O planejamento do sistema elétrico visa atender a uma série de objetivos. Dentre esses, destaca-se, atender à demanda por energia elétrica. Isto deve ser feito de forma confiável e econômica.

O planejamento da operação é um problema complexo e para resolvê-lo necessita-se de uma cadeia de modelos organizados em uma estrutura hierárquica, onde cada nível abrange um período do horizonte de planejamento (Fortunato et al, 1990).

No planejamento de longo prazo (de 2 até 5 anos à frente) o objetivo é definir uma estratégia de geração para atender à demanda de forma confiável ao mínimo custo esperado de operação durante o período de planejamento. A formulação desta estratégia passa pela otimização dos intercâmbios entre áreas geográficas e pela coordenação hidrotérmica com a definição dos montantes de geração hidráulica e térmica. Já em um horizonte de médio prazo (período de 1 mês até 1 ano à frente), as metas de geração estabelecidas na estratégia de longo prazo são individualizadas por usina, com a fixação de metas de geração semanais. Por fim, no horizonte de curto prazo (período de uma semana até 1 mês à frente) as metas semanais são desagregadas em metas diárias para cada usina. No curto prazo também são realizadas as atividades relacionadas com a programação da operação, cuja finalidade consiste em estabelecer um cronograma que atenda às metas do planejamento da operação energética e que seja viável do ponto de vista elétrico. A programação da operação inclui a programação semanal em bases diárias e a programação diária em base horária, também conhecida como pré-despacho.

Uma informação fundamental em todas as etapas do planejamento da operação é a previsão da demanda por energia elétrica (ou carga). No planejamento da operação energética de longo prazo são suficientes previsões com discretização mensal, enquanto que no planejamento de curto prazo, são necessárias previsões para uma hora até uma semana à frente, em bases horárias (Gross & Galiana, 1987). As previsões de curto prazo incluem projeções da ponta diária (máxima carga), os valores da carga em certos instantes de tempo e o consumo de energia elétrica em certos períodos horários. Também fazem parte deste escopo as previsões de curtíssimo prazo para horizontes de 10 até 30 minutos à frente.

As previsões de carga de curto-prazo também são úteis em análises de contingência conduzidas offline com a finalidade de detectar condições operativas futuras que representem risco para o sistema elétrico. A partir destas análises o operador pode

preparar antecipadamente as ações corretivas necessárias para a operação segura do sistema (Gross & Galiana, 1987; Lee et al, 1992).

A previsão de carga em um horizonte de curto prazo ou STLF (*short-term load forecasting*) constitui uma área de intensa pesquisa e conta uma vasta literatura técnica contendo uma variedade de métodos, os quais são descritos de forma panorâmica por Gross & Galiana, (1987), Liu et al (1996), Lotufo & Minussi (1999), Alfares & Nazeeruddin (2002) e Hahn et al (2009). A variedade de técnicas é o resultado da busca dos operadores por métodos capazes de gerar previsões mais precisas que permitam fazer um melhor uso dos recursos disponíveis, ou seja, operar o sistema elétrico de forma ótima.

Os principais métodos para previsão de carga podem ser classificados em métodos estatísticos e métodos baseados em técnicas de inteligência artificial, os quais incluem a lógica *fuzzy*, as redes neurais e as máquinas de vetor de suporte.

Estas técnicas são gerais e podem ser utilizadas desde a previsão de carga em uma barra da rede elétrica, em uma regional, em um subsistema até a previsão da carga agregada do sistema.

Os modelos de previsão de carga podem utilizar ou não os dados meteorológicos, em geral, dados de temperatura máxima, temperatura mínima e amplitude térmica como variáveis explicativas da carga.

A modelagem da tendência e da sazonalidade da carga são necessárias para a previsão de carga em um horizonte de curto prazo, independentemente do método adotado.

As sazonalidades da carga são caracterizadas por: ciclos com periodicidades diárias e semanais definidos em função do dia da semana e período do ano, conforme ilustrado na Figura 8.3 que mostra variações da carga ao longo de uma semana do verão e de uma semana do inverno e pelas demandas no sábado e domingo menores que nos outros dias da semana. Além da diferenciação entre dias úteis e dias de fim de semana, a modelagem da carga também deve reconhecer e prever o comportamento diferenciado da demanda nos feriados e nos dias especiais que ocorrem antes e após os feriados, cujos perfis de demanda comportam-se de forma diferente dos demais dias da semana.

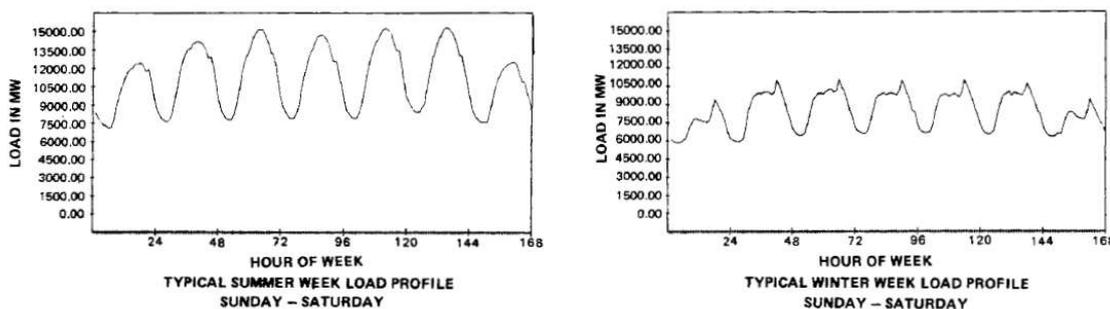


Figura 8. 3 Padrões semanais típicos da carga (Fonte: Gross & Galiana (1987))

As sazonalidades da carga são aferidas aplicando-se a análise de agrupamentos aos dados de consumo. Espera-se que no mesmo grupo sejam selecionados curvas de carga com padrão de consumo semelhante.

O desempenho de qualquer método de previsão depende da qualidade dos dados de entrada, por esta razão a aplicação destes métodos deve ser precedida por um tratamento dos dados com a finalidade de identificar e corrigir os eventuais valores aberrantes e lacunas de dados.

### 8.3. Cálculo de Tarifas

Nesta seção é apresentado o uso de curvas de cargas aplicado ao cálculo de tarifas de energia elétrica. Nesse contexto, tarifa corresponde ao valor pago pelo consumidor proporcional ao seu próprio consumo individual. Este valor remunera a concessionária de distribuição de energia elétrica.

Existem várias modalidades tarifárias. Algumas destas modalidades possuem preços diferentes dependendo do período do dia. Há valores diferenciados para período de ponta (horário entre 18:00hs e 20:00hs) e fora de ponta (demais horários). O período de ponta corresponde ao período de maior consumo.

A fixação das tarifas é baseada em metodologia PESSANHA (2001) de custos marginais na qual a receita permitida da concessionária de distribuição é repartida em subgrupos tarifários e postos tarifários de acordo com a contribuição dos clientes em cada subgrupo para as demandas de ponta no sistema da concessionária de distribuição.

Normalmente uma concessionária de distribuição possui muitos clientes. Idealmente, a análise da contribuição de todos os clientes para a demanda de ponta

forneceria a melhor tarifa. Infelizmente, não é possível registrar o consumo individual no tempo de todos os clientes.

Devido a esta indisponibilidade, utiliza-se uma amostra aleatória estratificada por classe e faixa de consumo. As classes de consumo são residencial, comercial e industrial. As faixas de consumo são baixa tensão, média tensão e alta tensão.

Para cada consumidor selecionado na amostra, é feito o registro temporal de seu consumo. Este registro individual é a curva de carga correspondente a cada consumidor da amostra. Cada curva de carga individual pode ser vista como a manifestação de um ou mais perfis típicos de uma classe de consumo.

Assim aplica-se a análise de agrupamentos neste conjunto de curvas de carga para obter grupos homogêneos. Estes grupos possuem curvas médias que correspondem aos perfis típicos. O comportamento da demanda por classe de consumo é caracterizada por estes perfis típicos. A figura 8.2 exemplifica estes perfis típicos.

Estes perfis típicos representam o conjunto de consumidores de uma classe de consumo e contém a informação fundamental para desagregação da receita dos subgrupos e postos tarifários. Desta forma, é necessário extrapolar os perfis típicos para o mercado real. Esta extrapolação é o ajuste dos perfis típicos ao mercado conforme demonstrado em PESSANHA (2001).

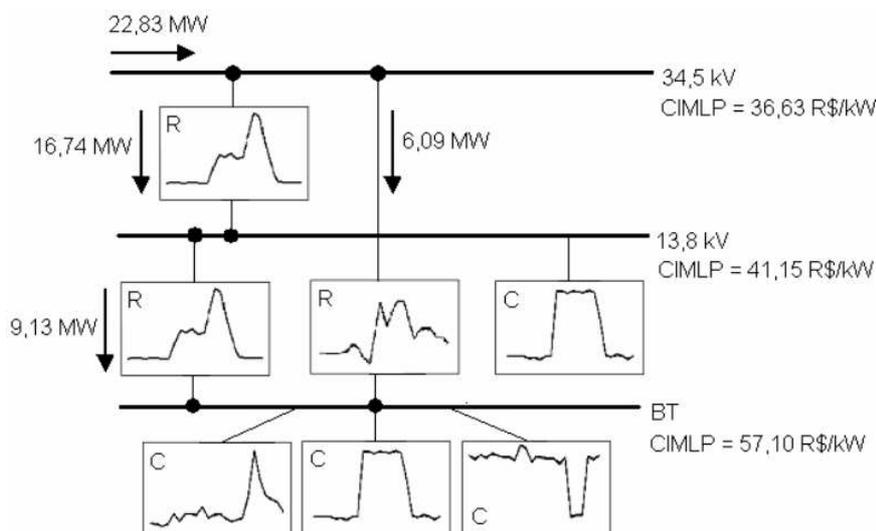


Figura 8. 4 Níveis de tensão e Perfis típicos de clientes e redes (Fonte: PESSANHA 2004)

# Capítulo 9

## Conclusões

O recente aumento de consumo de energia elétrica no Brasil tem como razão o aumento do poder aquisitivo da população. Desta forma, há um círculo virtuoso. A população compra mais equipamentos e a indústria produz mais bens. Assim, o consumo de energia elétrica aumenta.

Cada consumidor, com o seu consumo, determina a quantidade de energia elementar que deve ser gerada para o seu próprio suprimento individual. A análise do comportamento dos consumidores foi feita através de sua representação em função do tempo; denominada curva de carga. As curvas de carga demonstram o consumo agregado ao longo do tempo.

Normalmente as medições destas curvas de carga apresentam falhas: **descontinuidades** e **lacunas**. Desta forma, foi apresentado o tema filtragem de dados. Os objetivos da filtragem são **identificação** e **correção** das falhas supracitadas. Foi proposta uma nova metodologia de filtragem de dados de curvas de carga.

A primeira fase desta metodologia de filtragem de dados trata da identificação de falhas. Para identificação de falhas por **descontinuidade** foram aplicadas a transformada discreta *Wavelet* (TDW) e o *Boxplot*.

Com o uso da transformada discreta *Wavelet* (TDW), foi possível dividir a curva de carga ( $C$ ) em duas componentes, uma de aproximação ( $A$ ) e outra de detalhe ( $D$ ). De tal forma que  $C = A + D$ .

As componentes *wavelet* são sensíveis a descontinuidades e, conseqüentemente, qualquer perturbação nos dados são refletidas na componente de detalhe. Desta forma, para selecionar os pontos de descontinuidade, foi aplicado o *Boxplot* no logaritmo dos valores absolutos da componente de detalhe  $D = C - A$ . Assim os maiores valores absolutos de  $D$ , serão selecionados como pontos de descontinuidades.

Para identificação de **lacunas** foi aplicado um critério mais simples. Na presença de lacunas, a componente de aproximação  $A$  não realiza qualquer correção, apenas mantém as lacunas no mesmo valor da curva original. Nos procedimentos intrínsecos ao transformada discreta *Wavelet*, nas lacunas, a componente de detalhe  $D$  sempre apresenta valor zero.

Foi definido o seguinte critério para identificação de lacunas: foram identificados por 10 elementos consecutivos com valor menor do que  $10^{-14}$  os início/término de lacuna.

A segunda fase da metodologia de filtragem de dados trata da correção de falhas de **descontinuidade**. Nesta fase, foi aplicado o método LOESS.

Para seu uso, em um procedimento anterior, cada ponto descontínuo é substituído momentaneamente pela interpolação linear entre seus pontos contínuos vizinhos. A seguir, nessa curva modificada é aplicado o método LOESS que suaviza a curva e fornece uma correção suave para os pontos descontínuos anteriores.

Foi utilizada a técnica de análise de agrupamentos, para identificar os perfis de carga. Nesse contexto, a análise de agrupamentos determina tipologias que apresentam aplicabilidade na correção de lacunas, na previsão de carga e no cálculo de tarifas.

Outra questão proposta neste trabalho foi o procedimento de correção de **lacunas**. Em síntese, não existem informações na própria curva de carga para preencher estas lacunas. Foi adotada a seguinte hipótese: as informações necessárias para correção de lacunas estão no conjunto de curvas usadas para a determinação de tipologias. Desde que o comportamento da carga é sazonal, conforme evidências empíricas.

Do ponto de vista do planejamento e da operação do sistema elétrico, é fundamental ter informações futuras do consumo para antecipar medidas a serem tomadas para garantir o fornecimento de energia de forma adequada e com qualidade. Isto é feito através de métodos de previsão, que tem como um de seus fundamentos mais importantes os perfis típicos de carga.

Previsões de carga precisas são fundamentais para a operação econômica e segura de um sistema elétrico de potência. Se por um lado, previsões de carga superestimadas tendem a elevar os custos operacionais do sistema com a necessidade de uma maior reserva operativa, por outro lado, previsões subestimadas contribuem para reduzir a reserva operativa comprometendo a segurança do sistema.

A necessidade de previsões precisas tem conduzido ao desenvolvimento de um amplo conjunto de metodologias para previsão de carga, sobretudo de metodologias para previsão de curto prazo.

No contexto da previsão de carga, tem-se sempre necessidade de informações do consumo passado para estimar o consumo futuro. Essa estimativa, independentemente da metodologia de previsão adotada, depende fundamentalmente da qualidade das informações presentes no conjunto de curvas de carga.

Neste trabalho, foi considerado o estudo do comportamento dos consumidores brasileiros de energia elétrica, segundo a análise de dados de curvas de carga com discretização, minuto a minuto, e visando definir metodologias para tratamento destes dados de carga.

Foi desenvolvida uma nova metodologia para construção de arquétipos de curvas de carga de forma consistente, ou seja, impermeável a presença de erros. Foram apresentados resultados computacionais com discretização, minuto a minuto, enquanto usualmente a discretização é feita de hora em hora. Essa metodologia é uma alternativa consistente, com sofisticadas técnicas de análise de agrupamento e aplicabilidade em dados de registro de carga a nível nacional. A metodologia proposta é geral e pode ser utilizado em outros tipos de dados funcionais (RAMSAY, 2009). Assim, podemos usá-la pela análise dos registros de velocidade do vento a fim de construir tipologias características do comportamento do regime de ventos em dado espaço geográfico.

Os registros de velocidade do vento são informações importantes para o planejamento e geração de energia eólica. A energia eólica é a transformação da energia cinética dos ventos em energia elétrica. No caso do Brasil, embora o país tenha uma matriz elétrica predominantemente hidráulica, portanto, limpa e renovável, o aproveitamento do potencial eólico do país, avaliado em 143,5 GW a 50 metros de altura do solo (AMARANTE et al., 2001), oferece um importante complemento ao regime de geração hidrelétrica nos períodos de estiagem, contribuindo para assegurar a segurança energética do país e diversificar a matriz energética brasileira.

A grande vantagem da energia eólica é o baixo impacto ambiental, entretanto, essa depende do regime de ventos para produção eficaz de energia elétrica. Para produção em larga escala, são instalados centenas de geradores eólicos numa mesma área geográfica. Essa área é denominada parque eólico.

Por exemplo, para determinar a melhor localização de um parque eólico é necessário analisar os registros de velocidade do vento. Do ponto de vista da operação desses parques, é fundamental estimar períodos do dia onde a velocidade do vento é baixa. Essa estimativa permite a harmonização entre as diferentes formas de geração de energia para então manter a produção de energia elétrica estável e segura. Desta forma, as metodologias propostas são ferramental necessário para essa importante tarefa.

Em uma forma análoga, podemos usar a metodologia de tratamento de dados temporais apresentados neste trabalho para análise de registros de temperatura. Segundo SOUZA (2013), a temperatura apresenta influencia não linear sobre o consumo de

energia elétrica, no curto prazo. Assim, os registros de temperatura também são informações importantes para a operação do sistema. Esses registros, como também os perfis típicos de carga são ferramenta necessário para os métodos de previsão de carga. As metodologias propostas são aplicáveis aos registros de temperatura. Por isso, essas curvas de temperatura podem ser usadas com o objetivo de melhorar a qualidade da previsão de carga de energia elétrica. Produzindo perfis ou curvas típicas nss diferentes estações do ano.

Dessa forma, foi preenchida uma lacuna no conjunto de ferramentas de planejamento do setor elétrico brasileiro através da provisão de uma metodologia alternativa com adequação necessária para tratar dessas importantíssimas questões intrínsecas ao estudo de carga. Acredita-se ser esta a maior contribuição deste trabalho.

# Referências Bibliográficas

ALFARES, H.K.; Nazeeruddin, M. Electric load forecasting: literature survey and classification of methods, *International Journal of Systems Science*, v. 33, number 1, PP. 23-34, 2002.

ALSAFIH, H.A.; DUNN, R.; , "Determination of coherent clusters in a multi-machine power system based on wide-area signal measurements," **Power and Energy Society General Meeting, 2010 IEEE** , vol., no., pp.1-8, 25-29 July 2010.

AMARANTE, O.A.C., Brower, M., Zack, J. e Sá, A.L., *Atlas do Potencial Eólico Brasileiro*, Centro de Pesquisas de Energia Elétrica, Brasília, 2001.

CHEN, J.; LI, Wenyuan; Lau, A.; Cao, J.; Wang, K. Automated load curve data cleansing in power systems, *IEEE Transactions on Smart Grid*, vol. 1, no. 2, September, 2010.

CHUNXIA, Y. A data pretreatment technique about power system load modeling, *IEEE International Conference and Computer Application and System Modeling*, 22-24 oct, Shanxi, Taiyuan, 2010.

CLEVELAND, W.S. 1979 "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*, Vol. 74, pp. 829-836.

C++ 1D/2D DWT IMPLEMENTATION FOR WIN32 AND LINUX. Wavelet2d / Wavelet2s Libraries. [S.l.]: Google Code, [2013?]. Disponível em: <<https://code.google.com/p/wavelet1d/>>. Acesso em: 31 out 2013.

DAUBECHIES I., 1988 *Orthonormal Basis of Compactly Supported Wavelets*, *Comm. Appl. Math.*, 41, pp. 909-996, 1998. G.

DAUBECHIES I., 1992 *Ten Lectures on Wavelets*, CBMS – NSF Regional Conferences Series in Applied Mathematics, 1992.

DUAN Q.; HUIWEN D.; DI Ye; , "Weak voltage area recognition method in power system based on fuzzy subtract clustering," **Fuzzy Systems and Knowledge Discovery (FSKD)**, 2010 Seventh International Conference on , vol.4, no., pp.1662-1667, 10-12 Aug. 2010

EMPRESA DE PESQUISA ENERGÉTICA (BRASIL). Balanço energético nacional 2013 - ano base 2012 : relatório síntese. Rio de Janeiro: EPE, 2013. 55p. Disponível em: <[https://ben.epe.gov.br/downloads/S%C3%ADntese%20do%20Relat%C3%B3rio%20Final\\_2013\\_Web.pdf](https://ben.epe.gov.br/downloads/S%C3%ADntese%20do%20Relat%C3%B3rio%20Final_2013_Web.pdf)>.

Acesso em: 25 out 2013

FORTUNATO, L.A.M.; Neto, T.A.A.; Albuquerque, J.C.R.; Pereira, M.V.F. Introdução ao Planejamento da Expansão e Operação de Sistemas de Produção de Energia Elétrica, Niterói, Universidade Federal Fluminense, EDUFF, 1990.

GENG G.; JIAQI LIANG; HARLEY, R.G.; RUIQIAN QU; , "Load profile partitioning and dynamic reactive power optimization," **Power System Technology (POWERCON)**, 2010 *International Conference on* , vol., no., pp.1-8, 24-28 Oct. 2010

GRIGORAS, G.; Cartina, G.; Bobric, E.C.; Barbulescu, C. Missing data treatment of the load profiles in distribution networks, IEEE Bucharest Power Tech Conference, June 28th – July 2nd, Bucharest, Romania, 2009.

GROSS, G; Galiana, F.D. Short-Term Load Forecasting, Proceedings of IEEE, vol. 75, no. 12, pp. 1558– 1573, Dec. 1987.

GUAN, C.; Luh, P.B.; Coolbeth, M.A.; Zhao, Y.; Michel, L.D.; Chen, Y.; Manville, C.J.; Friedland, P.B.; Rourke, S.J. Very short-term load forecasting: multilevel wavelet neural networks with data pre-filtering, IEEE Power & Energy Society General Meeting, Storrs, 26-30 July, 2009.

GUIERELLI, C. R. 2006 *Previsão de carga de curto prazo de áreas elétricas através de técnicas de inteligência artificial* D.Sc. Thesis, Escola Politécnica-USP, São Paulo.

HAAR, A. *Zur Theorie der Orthogonalen Funktionen-Systeme*, Math. Ann, 69, pp. 331-371, 1910.

HAHN, H.; Meyer-Nieberg, S.; Pickl, S. Electric load forecasting methods: Tools for decision making, European Journal of Operational Research, 199, pp. 902–907, 2009.

HARTINGAN, J.A., *Clustering Algorithms*, John Wiley and Sons, Inc. New York, NY, 1975

HASTIE, T.; Tibshirani, R.; Friedman, J. The elements of statistical learning: data mining, inference and prediction, Second edition, Springer, 2009.

JANG, J.S.R.; Sun, C.T.; Mizutani, E Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence, Prentice Hall Inc, 1997.

JOHNSON, R.A. AND WICHERN, D.W.; 1998 *Applied Multivariate Analysis, Forth Edition, Prentice Hall*, New Jersey.

LI, Yang Short-Term Load Forecasting Based on LS-SVM Optimized by BCC Algorithm, 15th International Conference on Intelligence System Applications to Power Systems, Curitiba, Dezembro, 2009

LIMA, J. M.; SCHILING, M. TH.; COURTOUKE, C.; RODRIGUES, LAUBE, R..D. “Uma nova filosofia para o controle de tensão da rede básica”, **XI Simpósio de Especialistas em Planejamento da Operação e Expansão Elétrica**, Belém, 2009

LOTUFO, A.D.P.; Minussi, C.R. Electric power systems load forecasting: a survey, Powertech, Budapest, 1999.

MACQUEEN, J. B.; “Some Methods for classification and Analysis of Multivariate Observations”, **Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability**, Berkeley, University of California Press 1, 1967.

MARTINEZ, W.L.; Martinez, A.R. Computational statistics handbook with matlab, Chapman & Hall/CRC, 2002.

MENDES, M. S., 2012, *Agrupamento Via Suavização Hiperbólica com Arquitetura CUDA*. Dissertação de M.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.

NIZAM, M.; MOHAMED, A.; HUSSAIN, A.; , "Determining voltage unstable area in power systems using kohonen neural network," **Power Engineering Conference, 2007. IPEC 2007. International** , vol., no., pp.59-63, 3-6 Dec. 2007

NUHANOVIC, A.; GLAVIC, M.; PRLJACA, N.; "Validation of a clustering algorithm for voltage stability analysis on the Bosnian electric power system," **Generation, Transmission and Distribution**, IEE Proceedings- , vol.145, no.1, pp.21-26, Jan 1998

OGDEN, R. T. 1997, *Essential Wavelets for Statistical Applications and Data Analysis* 1.ed. Columbia, Birkhauser.

OLIVEIRA, H.M. 2007 *Análise de Sinais para Engenheiros: uma abordagem via Wavelets*, 1ed. Rio de Janeiro, BRASPORT

OPERADOR NACIONAL DO SISTEMA ELÉTRICO (BRASIL). [**Totais do consumo nacional de energia elétrica de 2005 à 2009**]. Rio de Janeiro, RJ: ONS, 2013. Arquivo digital 53,3 MBytes. Relatório técnico.

PESSANHA, J.F.M.; Velasquez, R.M.G.; Melo, A.C.G.; Caldas, R.P.; Tanure, J.E.P.S., Santos, P.E.S.; **Metodologia e Aplicação do Cálculo dos Custos Marginais de Fornecimento e das Tarifas de Uso dos Sistemas de Distribuição** Latin America Power & Gas, Rio de Janeiro 2001

PESSANHA, J.F.M.; Velasquez, R.M.G.; Melo, A.C.G.; Caldas, R.P. **Técnicas de Cluster Analysis na construção de tipologias de curva de carga**, XV Seminário Nacional de Distribuição de Energia Elétrica, Salvador, 2002.

PESSANHA, J.F.M., Huang, J.L.C., Pereira, L.A.C. , Passos Júnior, R., Castellani, V.L.O. **Metodologia e sistema computacional para cálculo das tarifas de uso dos sistemas de distribuição**, XXXVI SBPO, São João del Rey - MG, 2004.

PESSANHA, José Francisco Moreira; JUSTINO, Thatiana Conceição; MACIEIRA, Maria Elvira Piñeiro. **Métodos de previsão de carga elétrica de curto prazo**. Rio de Janeiro, RJ: CEPEL, 2011. 51p. Relatório técnico 42908/2011.

PESSANHA, José Francisco Moreira; JUSTINO, Thatiana Conceição; MACIEIRA, Maria Elvira Piñeiro. **Metodologia de tratamento dos dados de carga elétrica**. Rio de Janeiro, RJ: CEPEL, 2011. 45p. Relatório técnico 42907/2011.

PRICE, Phillip N. *Methods for analyzing electric load shape and its variability*. Berkeley, CA: Lawrence Berkeley National Laboratory, 2010. 54p. Relatório técnico LBNL-3713E.

RAMSAY, J.; Hooker, G. ; Graves, S. **Functional Data Analysis with R and MATLAB (Use R!)** 1st ed. New York: Springer, 2009. 202p.

REIS, A.J.R.; SILVA, A.P.A. *APLICAÇÃO DA TRANSFORMADA WAVELET DISCRETA NA PREVISÃO DE CARGA A CURTO PRAZO VIA REDES NEURAIS*, Revista de controle e automação, vol. 15, no.1, pp.101-108, 2004

SALEH, A.O.M.; LAUGHTON, M.A.; "Cluster analysis of power-system networks for array processing solutions," **Generation, Transmission and Distribution**, IEE Proceedings C , vol.132, no.4, pp.172-178, July 1985

SINGH, H.K.; SRIVASTAVA, S.C.; , "A reduced network representation suitable for fast nodal price calculations in electricity markets," **Power Engineering Society General Meeting**, 2005. IEEE , vol., no., pp. 2070-2077 Vol. 2, 12-16 June 2005

SOUZA, L. F. S. ; Wazlawick, R. S. ; Oliveira, C. M. ; Luca, L. A. D. A Influência da Temperatura no Comportamento da Carga de Curto Prazo e na Precisão da Sua Previsão. XI Simpósio Brasileiro de Automação Inteligente, Fortaleza, 2013

SPATH, H. 1980 *Cluster Analysis Algorithms for Data Reduction and Classification*, Ellis Horwood, Upper Saddle River, NJ.

SUAREZ-FARINAS, M.; Souza, R.L.; Souza, R.C. A methodology to filter time series: applications to minute-by-minute electric load series, *Pesquisa Operacional*, v.24, n.3, pp. 355- 371, Setembro a Dezembro, 2004.

TAN, P.N.; Steinbach, M.; Kumar, V. *Introdução ao Data Mining Mineração de Dados*, Editora Ciência Moderna, Rio de Janeiro, 2009.

THUILLARD, M. 2001, "Wavelets in Soft Computing" **World Scientific Series in Robotics and Intelligent Systems**, Vol.25

TORRES, G. LABERT; SILVA FILHO, D.; MORAES, C.H.V.;"An Intelligent System for Wrong Data Detection and Correction for Demand Forecasting Purpose"; 2006 IEEE PES **Transmission and Distribution Conference and Exposition Latin America**, Venezuela, 2006

TRYON, R. C. 1939 *Cluster Analysis*. Ann Arbor, Michigan, Edwards Bros,

TUKEY, J. W., MCGILL, R. and LARSEN, W. A. (1978), "Variations of Box Plots". *The American Statistician*, 32(1) 22-16.

VICTORIA H. , JIM A., A Survey of Outlier Detection Methodologies, *Artificial Intelligence Review*, v.22 n.2, p.85-126, October 2004

VISWANATH, P.A.; GOEL, L.; WANG, P.; , "Application of Fuzzy Clustering Technique to Reduce the Load Data in Reliability Evaluation of Restructured Power Systems," **Emerging Trends in Engineering and Technology (ICETET)**, 2009 2nd International Conference on , vol., no., pp.543-548, 16-18 Dec. 2009

XAVIER, A. E. 1982 Penalização hiperbólica: Um novo método para Resolução de Problemas de Otimização, M.Sc. Thesis, COPPE, Rio de Janeiro.

XAVIER, A. E., "The hyperbolic smoothing clustering method", **Pattern Recognition**, Volume 43, Issue 3, March 2010, Pages 731-737, ISSN 0031-3203

XAVIER, A. E., XAVIER, V. L. 2011 "Solving the minimum sum-of-squares clustering problem by hyperbolic smoothing and partition into boundary and gravitational regions", *Pattern Recognition*, v.44, p. 70-77.

XIAOXING, Z.; CAIXIN, S.; "Dynamic intelligent cleaning model of dirty electric load data" **Energy Conversion and Management** 49 (2008) 564-569.

YANG, J.; Stenzel, J. Historical load curve correction for short-term load forecasting, 7th International Power Engineering Conference, 29 Nov - 02 Dec, Singapore, 2005.