

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE INFORMÁTICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

HENRIQUE FENSTERSEIFER

**Uma abordagem de analítica visual e  
clusterização para avaliação da qualidade  
da distribuição de energia elétrica**

Dissertação apresentada como requisito parcial  
para a obtenção do grau de Mestre em Ciência da  
Computação

Orientador: Prof. Dr. Marcelo Walter  
Co-orientador: Profa. Dra. Carla M.D.S. Freitas

Porto Alegre  
2020

## CIP — CATALOGAÇÃO NA PUBLICAÇÃO

Fensterseifer, Henrique

Uma abordagem de analítica visual e clusterização para avaliação da qualidade da distribuição de energia elétrica / Henrique Fensterseifer. – Porto Alegre: PPGC da UFRGS, 2020.

99 f.: il.

Dissertação (mestrado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR–RS, 2020. Orientador: Marcelo Walter; Co-orientador: Carla M.D.S. Freitas.

1. Analítica visual. 2. Visualização de informações. 3. Clusterização. 4. Qualidade da distribuição de energia. I. Walter, Marcelo. II. Freitas, Carla M.D.S.. III. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Rui Vicente Oppermann

Vice-Reitora: Prof<sup>a</sup>. Jane Fraga Tutikian

Pró-Reitor de Pós-Graduação: Prof. Celso Giannetti Loureiro Chaves

Diretora do Instituto de Informática: Prof<sup>a</sup>. Carla Maria Dal Sasso Freitas

Coordenadora do PPGC: Prof<sup>a</sup>. Luciana Salete Buriol

Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

*“Life is like riding a bicycle.  
To keep your balance you must keep moving.”*  
— ALBERT EINSTEIN

## **AGRADECIMENTOS**

Sou muito grato ao orientador Marcelo e à co-orientadora Carla, pela disponibilidade, pelo conhecimento transmitido, pelos conselhos e pelo tempo dedicado, que me geraram grande aprendizado. Agradeço também aos professores e aos colegas de aula, que tornaram o curso muito produtivo e agradável.

Agradeço aos colegas de trabalho pelo seu tempo investido, pelo conhecimento técnico transmitido e pelo apoio.

Por último, mas não menos importante, agradeço aos meus pais e à minha esposa, pelo suporte e motivação durante o desenvolvimento do curso. Da mesma forma, àqueles que, direta ou indiretamente, foram - e continuam sendo - essenciais e que me colocaram no caminho que estou seguindo, estando sempre comigo e ajudando a alcançar os meus objetivos.

## RESUMO

Sistemas de distribuição de energia são frequentemente visualizados por métodos simples que apresentam os dados essencialmente na sua forma quantitativa. Os requisitos cada vez mais exigentes dos consumidores e da própria agência reguladora, bem como a complexidade do sistema elétrico, exigem que o Gerenciamento da Qualidade da Distribuição de Energia acompanhe tais avanços. Para aprimorar este gerenciamento e a tomada de decisões, é importante uma análise adequada dos dados que estão à disposição, permitindo um entendimento intuitivo do sistema. Explorando a aplicação de Visualização de Informações ao suporte à tomada de decisões acerca da distribuição de energia, esta pesquisa apresenta uma forma visual e intuitiva para interpretar e analisar as dimensões do sistema elétrico que influenciam a qualidade da energia distribuída. Diferentes técnicas de analítica visual foram utilizadas, através de uma plataforma de visualização de dados especificamente voltada à qualidade da energia - tema pouco explorado na comunidade e que, em geral, não é suficientemente enfatizado pelos sistemas informatizados das companhias. Via entrevistas com funcionários especialistas de uma empresa de distribuição de energia elétrica, definimos as funcionalidades do sistema, bem como as informações a serem apresentadas. Apresentamos ainda um modelo de avaliação da qualidade de energia baseado em clusterização de consumidores, utilizando os algoritmos K-Means e DBSCAN. Para validação dos resultados, potenciais usuários de uma distribuidora de energia utilizaram o sistema. Os resultados obtidos mostraram que a abordagem é útil e permite agilizar e reduzir o esforço necessário para avaliar a qualidade da energia. Constatou-se também que a solução proposta, ainda de acordo com a opinião dos mesmos, poderia ser utilizada por outras distribuidoras e que foi de fácil aprendizado. O modelo de avaliação baseado em clusterização permitiu identificar ao menos 160 consumidores, de um total de 67.000, com características que podem causar distorção da qualidade de energia. Medições *in loco* comprovaram a identificação de consumidores com violação dos limites adequados de tensão, permitindo a criação de um modelo proativo para gerenciar a qualidade do sistema de distribuição.

**Palavras-chave:** Analítica visual. visualização de informações. clusterização. qualidade da distribuição de energia.

## **A visual analytics and clustering approach for quality assessment of electricity power distribution**

### **ABSTRACT**

Power distribution systems are often visualized by simple methods that present data primarily in quantitative form. The increasingly demanding requirements of customers and the regulatory agency itself and complexity of the electrical system, require that the Quality Management of Electric Power Distribution follows such advances. To improve quality management and decision-making, it is essential to analyze the available data properly, allowing an intuitive system understanding. Exploring Information Visualization to support decision-making about power distribution, this research presents a visual and intuitive way to interpret and analyze the electrical system's dimensions that influence power quality. Different visual analytical techniques were used, through a data visualization platform specifically focused on power quality - a topic that has been scarcely worked in the community so far and that, in general, is not sufficiently emphasized by the companies' computerized systems. Through interviews with specialist employees of a power distribution company, we defined the system's functionalities, and the kind of information to be presented. We also present a power quality assessment model based on customer clustering, using K-Means and DBSCAN algorithms. To validate results, potential users of a power distribution company used the system. The results obtained showed that the approach is useful and allows us to speed up and reduce the effort required to evaluate power quality. It was also found that the proposed solution, still in agreement with their opinion, could be used by other companies and was easy to learn. The clustering-based assessment model allowed the identification of at least 160 customers, out of a total of 67.000, with characteristics that can distort power quality. Measurements carried out proved the identification of customers with violation of the appropriate voltage limits, allowing the creation of a proactive model to the Distribution System Quality Management.

**Keywords:** visual analytics, information visualization, clustering, power distribution quality.

## LISTA DE ABREVIATURAS E SIGLAS

3D	Tridimensional
ANEEL	Agência Nacional de Energia Elétrica
BC	Bloco de carga
BT	Baixa Tensão
CSV	<i>Comma Separated Values</i>
DRC	Duração Relativa da transgressão para tensão Crítica
DRP	Duração Relativa da transgressão para tensão Precária
ERP	<i>Enterprise Resource Planning</i>
Fecoergs	Federação das Cooperativas de Energia, Telefonia e Desenvolvimento Rural do Rio Grande do Sul
GIS	<i>Geographic Information System</i>
HOG	<i>Histogram of Oriented Gradients</i>
kVA	Kilovoltampère
kW	Kilowatt
MT	Média Tensão
PC	<i>Principal Component</i>
PCA	<i>Principal Component Analysis</i>
PRODIST	Procedimentos de Distribuição
REN	Resolução Normativa
UC	Unidade Consumidora
SE	Subestação
SIG	Sistema de Informação Geográfica
SUS	<i>System Usability Scale</i>
SVM	Support Vector Machine

## LISTA DE FIGURAS

Figura 2.1	Visualização de contornos sobre um mapa geográfico .....	22
Figura 2.2	Visualização geográfica tridimensional do consumo médio de energia .....	23
Figura 2.3	Exemplo de técnicas de diagrama de cordas .....	24
Figura 2.4	Visualização da energia como campos de fluxo .....	24
Figura 2.5	Diagramas de contornos sob a rede elétrica .....	25
Figura 2.6	Interface de uma plataforma de visualização do sistema elétrico.....	25
Figura 2.7	Exemplos de técnicas de visualização na distribuição de energia .....	26
Figura 2.8	Representações visuais de dados variantes no tempo.....	27
Figura 2.9	Variação da tensão ao longo do tempo .....	27
Figura 2.10	Visualização conjunta de um esquema topológico e magnitudes de um dado específico na visualização do mapa tridimensional .....	28
Figura 3.1	Visualização de grafo e árvore de uma parte da rede de distribuição de energia.....	36
Figura 3.2	Representação visual da classificação das tensões adequada, crítica e precária.....	43
Figura 3.3	Representação esquemática da abordagem: visualização de informações ....	45
Figura 3.4	Tela de entrada da plataforma ELVIS: visão da área geográfica de per- missão da distribuidora (poligonais) e as reclamações (pontos vermelhos) .....	47
Figura 3.5	Visualização aproximada/detalhada do mapa: redes de média e baixa tensão, transformadores, consumidores, processos de qualidade e usinas de geração distribuída .....	47
Figura 3.6	Reclamações visualizadas por linhas de contorno.....	49
Figura 3.7	Visão geral do sistema elétrico: visualização em gráfico <i>sunburst</i> . <i>Impr</i> e <i>Proc</i> : siglas para Improcedente e Procedente. ....	50
Figura 3.8	Visão geral do sistema elétrico: <i>heatmap</i> .....	51
Figura 3.9	Dispersão de consumidores: gráfico de dispersão .....	52
Figura 3.10	UCs e consumo: radar e coordenadas polares .....	53
Figura 3.11	Carregamento dos transformadores: gráfico de barras empilhadas .....	54
Figura 3.12	Histórico de reclamações: gráfico de linhas .....	55
Figura 3.13	Reclamações por alimentador: <i>treemap</i> .....	56
Figura 3.14	Carregamento do transformador: gráfico de linhas .....	57
Figura 3.15	Balanceamento de fases do transformador: gráfico de barras .....	58
Figura 3.16	Vizinhança: gráfico de bolhas.....	59
Figura 3.17	Análise da vizinhança: mapa geográfico .....	61
Figura 3.18	Relatório de alertas da qualidade de energia do consumidor .....	62
Figura 4.1	Representação esquemática da abordagem para avaliação da qualidade de energia .....	64
Figura 4.2	Comparação da clusterização entre K-Means e DBSCAN.....	68
Figura 4.3	Matriz da Correlação de Pearson entre as dimensões de dados.....	69
Figura 4.4	Representação dos componentes principais do PCA.....	70
Figura 4.5	Método <i>Elbow</i> para definição da quantidade de clusters.....	71
Figura 4.6	Definição do parâmetro eps do DBSCAN.....	72
Figura 4.7	Visualização de um cluster específico no mapa geográfico.....	72
Figura 5.1	Avaliação das visualizações por tarefa: mínimos, máximos e média por questão .....	78



Figura 5.2	Avaliação subjetiva .....	79
Figura 5.3	Visualização bidimensional da clusterização com K-Means.....	82
Figura 5.4	Distribuição das dimensões de dados entre os clusters com K-Means.....	83
Figura 5.5	Visualização bidimensional da clusterização com DBSCAN .....	85
Figura 5.6	Distribuição das dimensões de dados entre os clusters com DBSCAN .....	86
Figura 5.7	Representatividade de cada dimensão para cada os onze componentes principais.....	86
Figura 5.8	Resultado do DBSCAN evidenciando o cluster 4, representado pela cor azul.....	87
Figura 5.9	Resultado da medição: gráfico de tensão máxima para UC trifásica .....	88
Figura 5.10	Resultado da medição: gráfico de tensão mínima para UC monofásica .....	88

## LISTA DE TABELAS

Tabela 1.1 Benefícios das técnicas de visualização de informações.....	16
Tabela 2.1 Técnicas de visualização propostas pelos trabalhos relacionados.....	29
Tabela 2.2 Temas apresentados pelas técnicas de visualização dos trabalhos relacionados .....	30

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>13</b>
1.1 Motivação.....	14
1.2 Escopo .....	17
1.3 Objetivos e resultados esperados .....	17
1.4 Contribuição .....	18
1.5 Organização do trabalho .....	19
<b>2 TRABALHOS RELACIONADOS</b> .....	<b>21</b>
<b>2.1 Visualização de Informações na área elétrica</b> .....	<b>21</b>
2.1.1 Visualizações geográficas .....	21
2.1.2 Visualizações não-geográficas .....	26
2.1.3 Visão geral .....	29
<b>2.2 Clusterização de dados na área elétrica</b> .....	<b>31</b>
2.2.1 Clusterização em qualidade de energia.....	32
2.2.2 Clusterização de consumidores.....	33
2.2.3 Visão geral .....	34
<b>3 VISUALIZAÇÃO DA QUALIDADE DA DISTRIBUIÇÃO DE ENERGIA</b> .....	<b>35</b>
<b>3.1 Conhecimento prévio sobre redes de distribuição</b> .....	<b>35</b>
<b>3.2 Descrição dos requisitos</b> .....	<b>37</b>
<b>3.3 Descrição do conjunto de dados</b> .....	<b>38</b>
3.3.1 Componentes do sistema elétrico.....	39
3.3.2 Consumidores .....	40
3.3.3 Processos de qualidade de energia .....	41
3.3.3.1 Funcionamento do processo da reclamação.....	42
3.3.3.2 Contextualização .....	43
3.3.4 Discussão do conjunto de dados .....	44
<b>3.4 Visualizações e funcionalidades implementadas</b> .....	<b>44</b>
3.4.1 Análise Geográfica.....	46
3.4.1.1 Simbologia .....	48
3.4.1.2 Densidade de reclamações e densidade de geração distribuída .....	49
3.4.2 Análise do sistema .....	49
3.4.2.1 Visão geral .....	50
3.4.2.2 Dispersão de consumidores.....	51
3.4.2.3 Unidades consumidoras e consumo de energia .....	52
3.4.2.4 Carregamento dos transformadores .....	53
3.4.2.5 Histórico de reclamações .....	54
3.4.2.6 Reclamações por alimentador .....	56
3.4.3 Análise de consumidores .....	56
3.4.3.1 Carregamento do transformador .....	57
3.4.3.2 Balanceamento de fases do transformador .....	57
3.4.3.3 Vizinhança.....	58
3.4.3.4 Análise geográfica da rede, equipamentos e vizinhos .....	59
3.4.4 Discussão .....	62
<b>4 MODELO DE AVALIAÇÃO DA QUALIDADE DE ENERGIA</b> .....	<b>64</b>
<b>4.1 Seleção dos dados de qualidade da energia</b> .....	<b>65</b>
<b>4.2 Algoritmos escolhidos</b> .....	<b>67</b>
<b>4.3 Procedimento</b> .....	<b>68</b>
4.3.1 Investigação e padronização .....	68
4.3.2 Redução de dimensionalidade com PCA.....	69

4.3.3 Clusterização.....	70
4.3.3.1 K-Means.....	70
4.3.3.2 DBSCAN .....	71
4.3.4 Visualização dos clusters .....	72
<b>5 AVALIAÇÃO E RESULTADOS .....</b>	<b>74</b>
<b>5.1 Resultados da visualização de informações .....</b>	<b>74</b>
5.1.1 Participantes .....	74
5.1.2 Procedimento .....	75
5.1.3 Discussão .....	76
<b>5.2 Resultados da clusterização .....</b>	<b>81</b>
5.2.1 Avaliação quantitativa .....	81
5.2.2 Avaliação qualitativa .....	83
5.2.3 Avaliação com medições.....	86
5.2.4 Discussão .....	89
<b>5.3 Limitações.....</b>	<b>90</b>
<b>6 CONCLUSÃO .....</b>	<b>91</b>
<b>6.1 Reflexões.....</b>	<b>92</b>
<b>6.2 Trabalhos futuros.....</b>	<b>93</b>
<b>REFERÊNCIAS.....</b>	<b>94</b>
<b>APÊNDICE A — QUESTIONÁRIO DE USABILIDADE .....</b>	<b>99</b>
<b>A.1 System Usability Scale - SUS.....</b>	<b>99</b>

## 1 INTRODUÇÃO

Por natureza, a complexidade do sistema elétrico, aliada à constante evolução e crescimento dos sistemas de distribuição de energia, criam conjuntos de dados cada vez maiores. Isto dificulta a realização de análises e a tomada de decisões por parte dos engenheiros e técnicos responsáveis pela operação e manutenção da rede elétrica, cujo objetivo é atender aos requisitos cada vez mais exigentes dos órgãos reguladores e dos próprios clientes (ZAIDI et al., 2017; YANG; GUO, 2007; OVERBYE; WEBER, 2015).

Atualmente, a energia elétrica é fundamental e isto pode ser percebido com a exigência cada vez maior dos consumidores na continuidade e na qualidade do fornecimento. Atendendo a esta premissa, a Agência Nacional de Energia Elétrica (ANEEL), estabelece normas e determina limites que as distribuidoras de energia precisam atender, tais como a Resolução Normativa nº 753 de 13 de dezembro de 2016, que estabelece os limites para os indicadores de continuidade para cada conjunto elétrico de cada distribuidora (AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA - ANEEL, 2016) e o Módulo 8 - Qualidade da Energia Elétrica, dos Procedimentos de Distribuição (PRODIST), referente aos procedimentos e limites adequados da energia (AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA - ANEEL, 2017). Do ponto de vista do cliente, problemas como interrupções de energia, tensão alta ou baixa, ou ainda oscilações de tensão, devem ser eventos mitigados ao seu limite mínimo, precavendo o registro de reclamações e promovendo a satisfação do cliente. O conjunto destes fatores acaba por criar a demanda por avanços no Gerenciamento da Qualidade da Distribuição de Energia Elétrica.

O volume e a complexidade de dados que o sistema de distribuição de energia pode abranger é proporcional à quantidade de dimensões: desde as subestações, linhas de transmissão, redes de distribuição, equipamentos, componentes e consumidores, relacionando entidades que influenciam a qualidade da energia, além das reclamações. Consequentemente fica mais complexa a identificação de tendências e detecção de situações anormais, representando um desafio para as distribuidoras de energia. Este tipo de análise faz com que seja necessária a utilização de técnicas para esta finalidade, as quais, porém, apenas são efetivamente úteis através de uma apresentação adequada das informações (DUTTA; OVERBYE, 2013; OLIVEIRA et al., 2017).

A análise de informações é uma atividade elementar para a tomada de decisões. Porém, ela pode se tornar muito maçante à medida em que o volume de dados a serem interpretados cresce (USTARIZ-FARFAN et al., 2012). Consequentemente, isto acarreta

maior demanda de tempo para a conclusão da atividade, que pode, inclusive, resultar numa provável tomada de decisão equivocada, tornando a atividade humana cada vez mais dependente de um recurso computacional apropriado (OLIVEIRA et al., 2017).

Abordagens visuais são formas adequadas para trabalhar com grandes quantidades de dados, dado que as pessoas naturalmente tendem a realizar análises visuais antes que outras formas de análise, como, por exemplo, ler um texto ou ouvir a explicação de um determinado assunto (MANGISI, 2015). Além da alta dimensionalidade dos dados, ou seja, do grande número de atributos de cada entidade, há a característica espaço-temporal do sistema elétrico, tornando a utilização de técnicas de visualização praticamente uma necessidade (CHEN; GUO; WANG, 2015).

Fundamentado nisto, explorar técnicas de Visualização de Informações para propor soluções para determinadas tarefas é uma tendência de pesquisa atualmente, utilizada em muitas áreas por consequência da sua capacidade de representação dos mais diversos tipos de dados. Trânsito e tráfego de veículos (CHEN; GUO; WANG, 2015), saúde e assistência médica (BOWMAN et al., 2016; STADLER et al., 2016), mudanças climáticas (HAROLD et al., 2016) e esportes (BASOLE; SAUPE, 2016) são exemplos da diversidade de estudos que estão sendo realizados como aplicação de visualização de informações, e ainda mais da relevância do tema na era da Ciência de Dados. Os conjuntos de dados, afinal, são cada vez maiores e torna-se necessário criar maneiras de extrair deles a maior quantidade de conhecimento e informação útil possível.

Dada a quantidade de técnicas propostas, é pertinente propor abordagens visuais e aplicação prática em áreas que permitam esta oportunidade, demonstrando como estas técnicas podem solucionar problemas existentes. Uma destas, é a área da distribuição de energia elétrica (ZAIDI et al., 2017), e no caso deste estudo, da sua qualidade.

## **1.1 Motivação**

Ao refletir sobre a visualização de informações na área de energia elétrica, percebe-se que existe uma tendência de apresentação dos dados em forma de tabela, inclusive nos sistemas informatizados das companhias. Poucas são as visualizações disponíveis que vão além da visualização geográfica ou esquemas elétricos, muito comuns em distribuidoras. Além disso, para a apresentação de alguns dados, os gráficos de barras ou linhas são as opções utilizadas. A comunidade acadêmica apontou que estas técnicas, sobretudo as formas textuais ou tabulares, costumam enfrentar problemas para representar dados

numerosos ou complexos, além de não evidenciar devidamente os dados que deveriam ser notados por quem está visualizando as informações. Isto abre espaço para interpretações desiguais por diferentes pessoas, interpretações equivocadas ou, ainda, deixar passar despercebida alguma informação relevante (ZAIDI et al., 2017; CUFFE; KEANE, 2015).

No âmbito dos sistemas elétricos, da mesma maneira, a distribuição de energia também é frequentemente visualizada por métodos simples que apresentam essencialmente os dados na sua forma quantitativa. Nigar, Agalgaonkar e Ciufu (2015) apontam que, em geral, são usadas representações de gráficos de barras, de pontos e mapas, as quais são focadas em mostrar o estado do sistema como um todo. Contudo, é preciso também focar na visualização dos efeitos de variáveis específicas do sistema, como a situação de um determinado equipamento. Outro trabalho constatou que a visualização de dados na área de transmissão de energia foi mais numerosa do que na área de distribuição de energia, motivando também a realização deste trabalho (NGA et al., 2012; CAPUDER; PANDZIC; KUZLE, 2017). Em termos gerais, enquanto a distribuição de energia representa a parcela do sistema elétrico que distribui a energia das subestações até os consumidores, a transmissão de energia é responsável por transportar a energia dos centros geradores até as subestações, com tensões mais elevadas e grandes distâncias.

Para aprimorar representações de dados nessa área, deve-se mostrar os benefícios que a Visualização de Informações pode disponibilizar através do emprego de uma abordagem que possa ir além daquilo que está atualmente disponível. A possibilidade de apresentar os dados das companhias de novas formas, principalmente aqueles que ainda não foram visualizados graficamente, aliada ao aprimoramento das técnicas existentes (como a geográfica), proporcionarão uma interpretação mais vantajosa dos dados e o favorecimento de análises de controle e qualidade do sistema - explorando o nicho existente entre as duas áreas (USTARIZ-FARFAN et al., 2012; NGA et al., 2012; KLUMP; DOOLEY; WU, 2003).

Sucharitha, Subash e Prakash (2014) levantaram os principais benefícios que a visualização de informações pode trazer para qualquer área de aplicação, resumidos na Tabela 1.1, ordenada pelo percentual de respostas.

Nesta tabela, pode-se perceber o impacto que as técnicas de visualização e análise de dados produzem em questões como melhor tomada de decisão e melhor análise de dados. Contudo, como visto nos parágrafos anteriores, métodos de visualização são insuficientemente explorados na área de distribuição de energia elétrica, tanto na literatura como também por empresas de distribuição de energia. Isto limita as inferências

Tabela 1.1: Benefícios das técnicas de visualização de informações

<b>Benefícios</b>	<b>Percentual (%)</b>
Melhor tomada de decisão	77
Melhor análise de dados	43
Compartilhamento de informações aprimorada	41
Capacidade aprimorada de autogerenciamento ao usuário final	36
Maior retorno do investimento	34
Economia de tempo	20
Redução da carga de operação da TI (Tecnologia da Informação)	15

Fonte: Sucharitha, Subash e Prakash (2014)

realizadas pelos usuários, o que fez emergir a possibilidade de aplicar técnicas de visualização voltadas ao sistema elétrico, com potencial para permitir avanços nos métodos de Gerenciamento da Qualidade da Distribuição de Energia.

A clusterização de dados é um método não supervisionado de aprendizado de máquina. Tem um caráter exploratório, identificando um ou mais agrupamentos de itens da população que possuem características similares entre si. No intuito de avaliar também proativamente a qualidade da energia, é possível utilizar este mecanismo para procurar identificar consumidores que possuem tendências de distorções na energia. O estudo de formas para identificação e tratamento proativo da energia destes consumidores são significantes no Gerenciamento da Qualidade.

Aplicada à área elétrica, grande parte dos trabalhos focou especificamente em perturbações na rede, qualidade da energia em subestações ou linhas de transmissão, ou na área comercial (ex. consumo, tarifa e preço). Contudo, a avaliação do agrupamento na presente pesquisa foi realizada com a clusterização dos consumidores, baseada nas informações representativas do seu suprimento de energia.

Foram levadas em consideração as informações de conexão à rede de cada consumidor, o consumo, características da vizinhança e os componentes elétricos próximos. Ou seja, os dados foram extraídos e processados de acordo com as características elétricas individuais de cada consumidor.

Neste contexto, uma visualização interativa dos dados, que permita uma interpretação adequada e uma inferência ágil das informações pelo usuário, tem um papel tão importante quanto a existência dos dados em si. A vantagem é que os dados podem ser representados visualmente, resultando em atividades eficientemente realizadas de análise e interpretação, além de explicitar possíveis tendências ocultas nos dados - diferentemente da forma baseada em texto (CHEN; GUO; WANG, 2015; ZAIDI et al., 2017). Conseqüentemente, o propósito desta dissertação é apresentar esta abordagem para a distribuição de energia e para a avaliação da qualidade da energia, contribuindo, ao final,



para ambas as áreas.

## **1.2 Escopo**

Este estudo focou na qualidade da energia voltada para os consumidores do grupo B, representados em grande parte pelos consumidores residenciais e comerciais. Ou seja, consumidores do grupo A não fazem parte do presente escopo, que são, em geral, grandes indústrias e por representarem uma pequena parcela do conjunto de consumidores (menos que 1% da distribuidora em estudo). Além disso, possuem outros mecanismos de controle de qualidade da energia, bem como métodos de telemetria já bastante utilizados por serem responsáveis por grande parte do consumo de energia distribuído pelas companhias.

A qualidade da energia analisada neste estudo envolve, principalmente, os limites adequados da tensão de fornecimento. Não são considerados os efeitos de distorções de energia ou corrente elétrica.

## **1.3 Objetivos e resultados esperados**

O objetivo principal deste trabalho é propor, implementar e validar uma abordagem visual para permitir avanços no Gerenciamento da Qualidade da Distribuição de Energia, pela utilização de técnicas de visualização de informações e aprendizado de máquina. O intuito é permitir a identificação de pontos de possível transgressão dos índices de qualidade, como também estudar as possíveis causas de reclamações. Sobre os resultados desta análise, os técnicos responsáveis pela distribuição da energia poderão definir ações e agir na resolução das ocorrências encontradas. Visualizações de relevância secundária, como identificação de possíveis segmentos de rede com superdimensionamento, também poderão ser evidenciados.

Com a apresentação das técnicas, pretende-se ainda permitir a descoberta de novas informações que, em função da quantidade de dimensões do sistema, podem passar despercebidas em técnicas de visualização simples. Além da quantidade de dados, a limitada exploração das técnicas de visualização também pode acarretar a não-identificação de padrões. Exemplo disto pode ser a comparação de eventualidades no sistema ao longo do tempo.

Embora o foco principal deste trabalho seja na área de Visualização de Informa-

ções, este estudo também envolveu um modelo de avaliação da qualidade de energia. O propósito é averiguar a possibilidade de encontrar conjuntos de consumidores com possível qualidade de energia abaixo da desejável, utilizando as principais dimensões que podem influenciar nisto através de técnicas de clusterização. O estudo adotou técnicas das áreas de *Visual Analytics* e *Knowledge Discovery*, pois as áreas ficam muito próximas no que tange a exibição de informações ao usuário final.

Outras finalidades também podem ser atingidas, como apoiar o planejamento da expansão do sistema elétrico, em função das informações de *status* do sistema serem úteis para o planejamento e priorização de obras em determinadas regiões.

Em resumo, os objetivos do trabalho são:

- Pesquisar o estado-da-arte da área de visualização de informações e clusterização relacionadas à área elétrica, elencando as técnicas utilizadas em cada trabalho e fundamentando as escolhas para a área deste estudo;
- Integrar, em um único local, todas as informações relevantes à qualidade da energia do sistema;
- Planejar e construir uma abordagem visual através de técnicas que atendam às necessidades de forma ágil e intuitiva, evitando a tradicional forma tabular ou de barras e linhas;
- Propor um modelo de avaliação da qualidade da distribuição de energia, com base em técnicas de aprendizado de máquina;
- Relatar as lições aprendidas na aplicação das técnicas e a percepção dos usuários quanto à nova abordagem.

## 1.4 Contribuição

Diferentemente de trabalhos na área de visualização que procuram criar novas técnicas, cuja contribuição é a técnica propriamente dita, a pesquisa deste estudo é focada em uma proposta de uma solução para uma situação específica e no seu desenvolvimento, ou seja, é um trabalho enquadrado como (*design study*) (SEDLMAIR; MEYER; MUNZNER, 2012).

Neste sentido, o presente estudo é classificado como *problem-driven research*, cujo objetivo é focar no domínio da aplicação solucionando problemas reais com usuários reais. Por isso, conforme sugerido por esses autores, esta pesquisa foi conduzida

seguindo implicitamente o framework de nove estágios proposto no referido trabalho: *learn, winnow, cast, discover, design, implement, deploy, reflect e write*.

A motivação supracitada permite realizar avanços na área de Visualização de Informações e na distribuição de energia. Neste contexto, este trabalho é focado em aplicar técnicas de visualização conhecidas e aceitas pela comunidade, além de técnicas de clusterização para construir um modelo de avaliação de qualidade da energia focada no consumidor.

Nesta linha de pesquisa, as contribuições deste estudo são:

- Um comparativo dos trabalhos relacionados da área, contrapondo as técnicas de visualização e os dados apresentados em cada um;
- A proposta de uma abordagem de analítica visual aplicada à análise da qualidade na distribuição de energia elétrica;
- A proposta de um modelo de avaliação da qualidade da distribuição de energia, a partir da clusterização de consumidores;
- A discussão das lições aprendidas e da percepção dos usuários, que servirão de fundamento para pesquisas posteriores na área.

Portanto, o presente estudo pode ser dividido em quatro principais etapas: 1) estudar trabalhos anteriores para fundamentar as decisões; 2) propor uma abordagem visual para a análise da qualidade da energia; 3) construir um modelo de avaliação de qualidade utilizando clusterização; e 4) avaliar as técnicas, a percepção do usuário e os resultados da clusterização.

## **1.5 Organização do trabalho**

Este trabalho será organizado em capítulos. No próximo capítulo temos os Trabalhos Relacionados, que apresenta os principais estudos realizados na interseção entre as áreas de Visualização de Informações e Clusterização de Dados com a área de Distribuição de Energia Elétrica. O Capítulo 3 descreve a concepção e a construção da abordagem visual e de análise do sistema elétrico, iniciando pela contextualização, elencando os requisitos e a descrição do conjunto de dados, e finalizando com o detalhamento das funcionalidades implementadas. Na sequência, o Capítulo 4 descreve o modelo de avaliação de qualidade da energia através de clusterização, com a especificação das dimensões de dados, o pré-processamento realizado e o detalhamento da sua implementação. Após,

o Capítulo 5 apresenta os resultados atingidos, através da discussão dos experimentos e métodos de avaliação. A conclusão e trabalhos futuros são apresentados no Capítulo 6.

## 2 TRABALHOS RELACIONADOS

Neste capítulo, são apresentados os trabalhos relacionados com este estudo. Foi feita uma revisão da literatura, apresentando os estudos realizados, as técnicas empregadas, e os resultados obtidos pelos autores. Estes trabalhos confirmaram a potencial contribuição no tópico escolhido e forneceram a perspectiva necessária para realizar escolhas durante o desenvolvimento das diferentes etapas do projeto. Este capítulo está dividido em seções, que tratam das duas principais partes do presente estudo. A primeira é referente a visualização e a segunda a clusterização.

### 2.1 Visualização de Informações na área elétrica

Esta seção trata da visualização de informações na área elétrica, trazendo os principais trabalhos encontrados com relação a qualidade de energia. Estes foram classificados em visualizações geográficas e não-geográficas.

#### 2.1.1 Visualizações geográficas

Esta subseção apresenta os principais resultados da comunidade acadêmica com relação a visualizações geográficas na área elétrica.

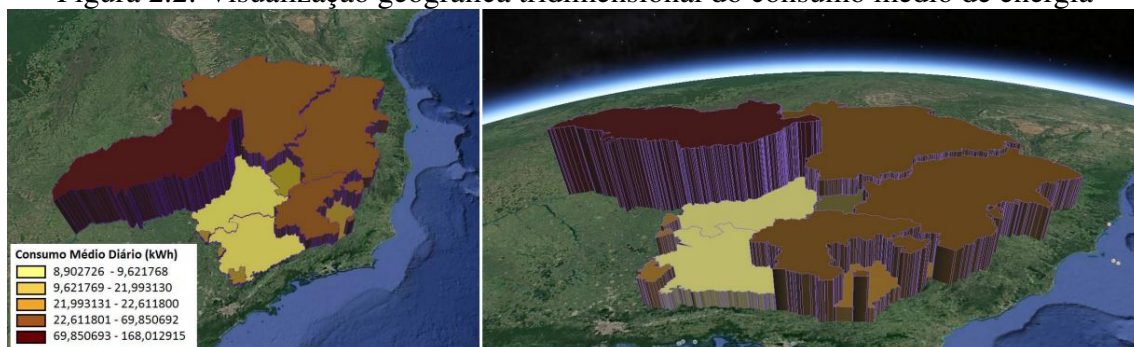
Capuder, Pandzic e Kuzle (2017) apontaram em seu trabalho que a literatura, em geral, afirma que o georreferenciamento das informações tornou-se uma necessidade para o planejamento de sistemas de distribuição. Mesmo possuindo os dados geográficos dos equipamentos, ainda são poucos os sistemas que combinam efetivamente os dados técnicos com os dados geográficos, incluindo em um local único as possibilidades de planejamento e estudo adequado. Estes sistemas são conhecidos como *Geographic Information System* - GIS. Os autores denominaram GIS 1.0 os sistemas geográficos simplificados que possuem apenas o mapa propriamente dito com pontos das localizações dos equipamentos, e GIS 2.0 aquele que agrega à versão anterior as informações proveitosas para análise efetiva dos dados, como as características técnicas e demais dados do sistema.

Complementando estes autores, Robbi (2000) afirma que o que diferencia os Sistemas de Informação Geográfica (SIG, ou GIS no inglês) das ferramentas de visualização, é a capacidade que os mesmos têm de utilizar a exploração para geração de conhecimento



zações geográficas bi e tridimensionais, para determinar magnitudes do sistema elétrico, como a densidade de consumidores e as médias de consumo em determinadas regiões - ou seja, o foco está direcionado à área comercial. A Figura 2.2 apresenta um exemplo de visualização geográfica tridimensional aplicada pelo autor. O trabalho de Klump, Doo-ley e Wu (2003) também aplicou visualização tridimensional sobre um mapa, juntamente com uma visualização de contornos.

Figura 2.2: Visualização geográfica tridimensional do consumo médio de energia



Fonte: Oliveira et al. (2017)

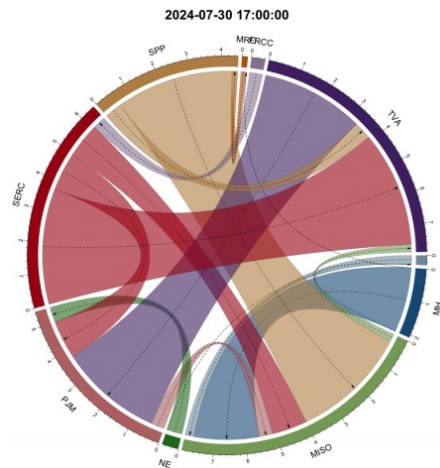
A maioria dos trabalhos deste tipo utiliza técnicas de visualização geográfica baseadas em mapa para melhor interpretação da localização geográfica da rede. O trabalho de Pourmirza e Brooke (2013), focado em Redes Inteligentes (*Smart Grids*), utilizou prioritariamente esta técnica e serve de exemplo.

A pesquisa de Molnar e Gruchalla (2018) desenvolveu uma técnica de agregação de fluxos. Os autores criaram equações matemáticas para calcular campos de vetores e gerar, como resultado, visualizações que agregam o fluxo de energia em um sistema de transmissão e distribuição de energia elétrica. Os resultados apontaram que é uma ótima técnica para ser agregada a outros tipos de visualizações, como o diagrama de cordas (ver Figura 2.3), pois permite que sejam realizadas análises mais aprofundadas considerando a origem e o destino de montantes de energia. A Figura 2.4 representa um dos resultados apresentados pelos autores.

Alguns estudos, em especial recentemente, têm trabalhado o assunto de Geração Distribuída. Este tema refere-se à geração de energia elétrica junto ou próxima aos consumidores, e em geral de pequeno porte, podendo ser consumida no mesmo local onde é gerada, além do excedente ser entregue ao sistema elétrico de distribuição. Neste caso, vale salientar que o consumo pode ser abatido do valor gerado, gerando economia para o titular. Porém, devido ao impacto da injeção de energia na rede, podem ser causados distúrbios, potencializados com a popularização do tema.

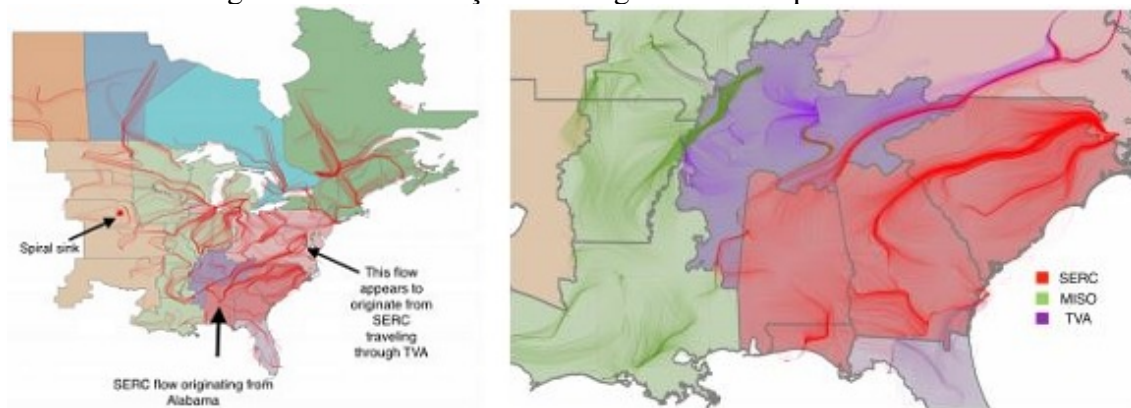
Nigar, Agalgaonkar e Ciufu (2015) realizaram um estudo para visualizar dados

Figura 2.3: Exemplo de técnicas de diagrama de cordas



Fonte: Molnar e Gruchalla (2018)

Figura 2.4: Visualização da energia como campos de fluxo



Fonte: Molnar e Gruchalla (2018)

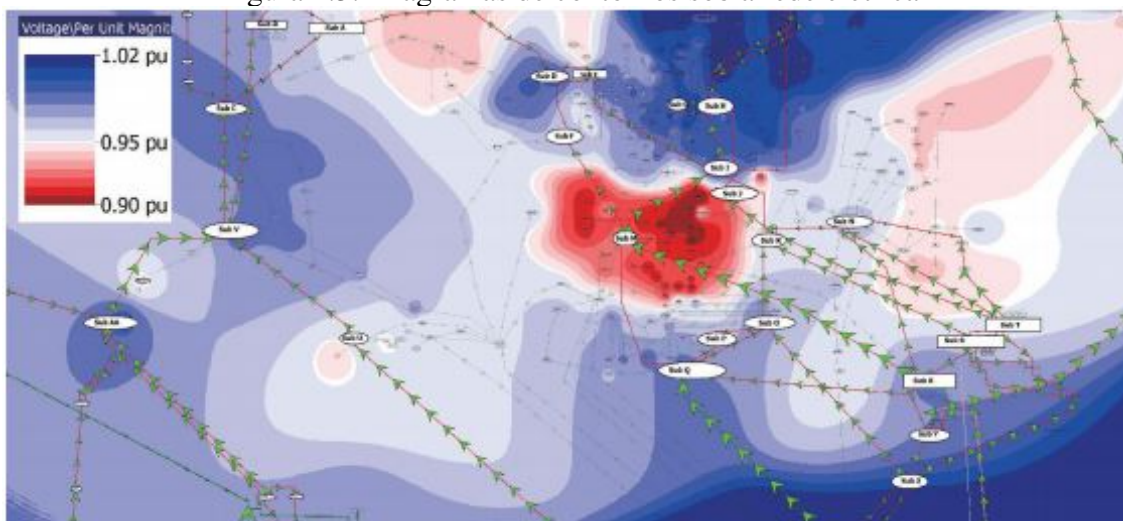
referentes à geração distribuída, ilustrando os efeitos da geração de energia na média tensão de redes de distribuição. A pesquisa de Overbye e Weber (2015), concentrada também na área de *Smart Grids*, mostrou que existem diversas técnicas de visualização que são utilizadas neste tema. As técnicas mais usuais são discutidas pelo autor, como gráfico de setores, contornos e geográfico. O autor apresenta ainda a técnica conhecida como *sparklines*, que reúne diversos gráficos pequenos e simples, geralmente no estilo de gráficos de linhas, porém sem eixos, que permitem uma aproximada - porém rápida - comparação das informações. Para a finalidade do estudo atual, a técnica mais relevante deste trabalho é a visualização de contornos logo abaixo à rede e a respectiva simbologia de equipamentos, representada pela Figura 2.5.

Vale ainda a consulta aos seguintes trabalhos, que compartilham da mesma característica de serem plataformas interativas com diversas técnicas de visualização, agregando várias informações de maneira fácil e intuitiva para acompanhar os dados do sistema elétrico. O primeiro, de Lu et al. (2019), apresenta dados de redes inteligentes, en-



quanto o segundo, de Lawson et al. (2018), é focado em monitorar o estado do sistema elétrico como um todo, integrando informações de diferentes bancos de dados (tensão, corrente, falta de energia, estados de equipamentos, entre outros). Este último é apresentado na Figura 2.6, ilustrando as quatro principais visualizações sugeridas pelos autores: A) mapa, B) linha do tempo, C) calendário de eventos e D) controle do mapa.

Figura 2.5: Diagramas de contornos sob a rede elétrica



Fonte: Overbye e Weber (2015)

Figura 2.6: Interface de uma plataforma de visualização do sistema elétrico



Fonte: Lawson et al. (2018)

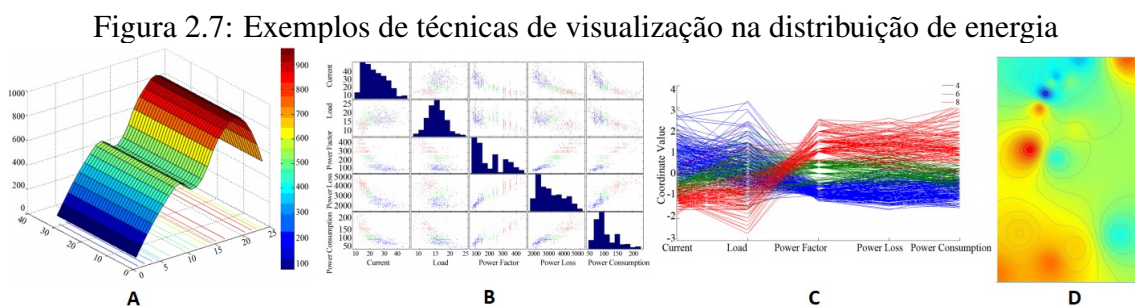
Além da visualização dos componentes elétricos, pesquisadores também desenvolveram soluções web para visualizar descargas elétricas. O trabalho de Silva (2014) propôs soluções de visualização deste tipo de fenômeno baseadas em técnicas geográficas através da colocação de marcadores sobre um mapa. Foi utilizada uma técnica de codificação de

cores, onde os marcadores de mesma cor representam as descargas ocorridas no mesmo momento.

### 2.1.2 Visualizações não-geográficas

Esta subsecção apresenta os trabalhos cujas visualizações não estão diretamente vinculadas a mapas ou geolocalização.

Nga et al. (2012) utilizaram técnicas de visualização voltadas para apresentar dados de redes inteligentes (*Smart Grids*). Os autores utilizaram superfícies de contorno 3D para apresentar dados de consumo de energia ao longo do tempo, onde os eixos representam o horário de um dia (1 a 24), o dia do mês (1 a 31) e o consumo de energia. O consumo está atrelado a um mapa de cores. A Figura 2.7-A apresenta esta visualização, onde é notável que o pico de consumo na região estudada é ao anoitecer. A Figura 2.7-B-C-D exemplifica outras técnicas empregadas pelos autores, respectivamente: matriz de dispersão, coordenadas paralelas e linhas de contorno (NGA et al., 2012). Vale consultar também o trabalho de Srinivasan e Reindl, que utilizou técnicas similares (SRINIVASAN; REINDL et al., 2015).



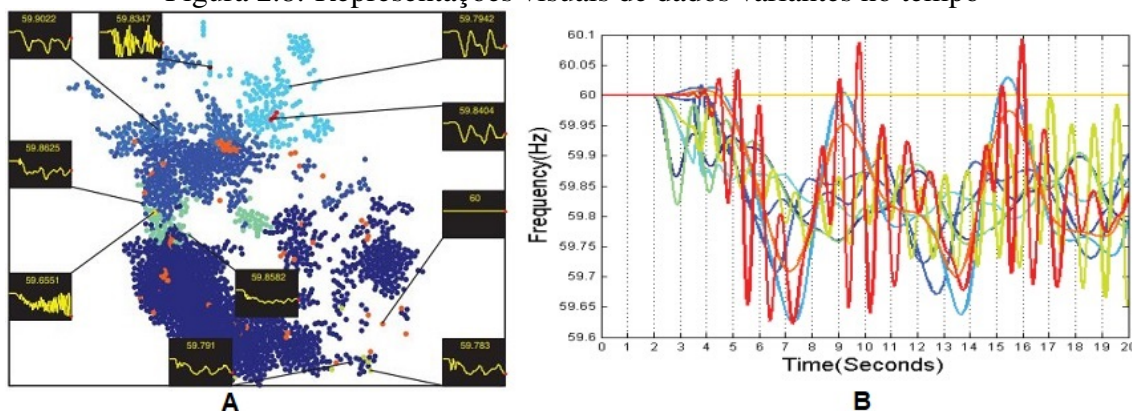
Fonte: Nga et al. (2012)

A literatura possui ainda trabalhos que apresentam técnicas de visualização de dados que variam no tempo. O estudo de Dutta e Overbye (2013) focou nesta característica, visualizando dados de grandes redes de distribuição em relação a dados variáveis no tempo. Ainda segundo os autores, técnicas de contorno são computacionalmente custosas, além de não apresentarem o dinamismo de uma variável do sistema, pois são de um instante de tempo específico. Em função disto, propuseram utilizar mapas de pontos, onde a cor indica o cluster ao qual os pontos pertencem, e os pontos de cada cluster tem sua variação ao longo do tempo representada em uma área deixada em evidência, no entorno do mapa. Assim é possível visualizar o agrupamento de pontos realizado pelo sistema,

vinculada com o comportamento de cada cluster ao longo do tempo. A Figura 2.8-A exibe esta visualização. Na sequência, também desenvolveram uma visualização simultânea de cada cluster em um gráfico de linhas, conforme apresenta a Figura 2.8-B.

Importante salientar que este trabalho foi classificado como não-geográfico em função da visualização prioritariamente considerar o agrupamento dos pontos em clusters, diferentemente de trabalhos que focam na localização geográfica dos equipamentos.

Figura 2.8: Representações visuais de dados variantes no tempo

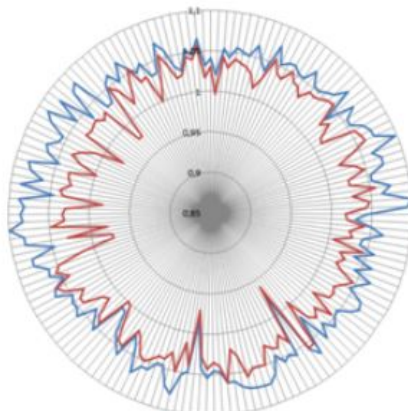


Fonte: Dutta e Overbye (2013)

A técnica denominada diagrama de círculos, geralmente aplicada de forma bidimensional, foi aplicada em 3D por Kumar e Singh para representar o fluxo de energia em linhas de transmissão, considerando variações de tensão de entrada e saída (KUMAR; SINGH, 2015).

Em Capuder, Pandzic e Kuzle (2017), os autores relacionam e analisam diversas técnicas de visualização já apresentadas até aqui por outros pesquisadores, além da visualização de gráfico radar que, neste caso, apresenta os picos de tensão ao longo de um período de tempo, apresentada pela Figura 2.9.

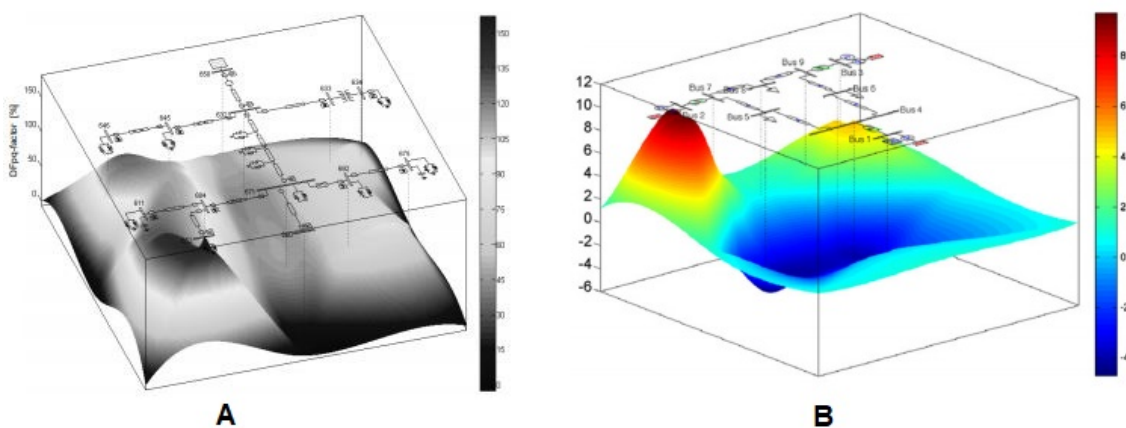
Figura 2.9: Variação da tensão ao longo do tempo



Fonte: Capuder, Pandzic e Kuzle (2017)

Voltado especificamente à perda de energia, o trabalho de Farfán, Palta e Tacca (2011) concentrou-se em desenvolver visualizações de sistemas elétricos apresentando dados topológicos e numéricos, sobrepostos a mapas bi e tridimensionais da perda da qualidade da energia. Utilizando a mesma técnica, Milano propôs visualizações relacionadas ao fluxo de energia (MILANO, 2009). Estas técnicas não foram baseadas em mapas geográficos, mas sim sobre o esquema elétrico propriamente dito. A Figura 2.10-A-B apresenta a visualização conjunta do esquema topológico e do respectivo mapa tridimensional de variáveis em análise, utilizada respectivamente em (FARFÁN; PALTA; TACCA, 2011) e (MILANO, 2009). O trabalho de Ustariz-Farfan et al. (2012) utilizou também esta técnica para representar perda de energia, além de visualizações de dados georreferenciados com base em cores para determinar regiões críticas.

Figura 2.10: Visualização conjunta de um esquema topológico e magnitudes de um dado específico na visualização do mapa tridimensional



Fonte: Farfán, Palta e Tacca (2011) e (MILANO, 2009)

Vale ao leitor conhecer também que finalidades complementares, como a área educacional, também estão sendo trabalhadas no nicho explorado nesta dissertação. Por isso, é importante relacionar o trabalho de Barcelos et al. (2013), que foi desenvolvido na área de realidade aumentada. Os autores aplicaram técnicas de visualização tridimensional de equipamentos de uma subestação de energia elétrica sobre um catálogo predefinido, com o intuito de auxiliar estudantes a conhecerem melhor os equipamentos, devido à complexidade e orçamento para estruturar laboratórios do tipo. A ideia é válida para possíveis trabalhos em realidade aumentada visualizando, ao invés de equipamentos, informações técnicas ou comerciais da rede em si, possuindo potencial utilização quando a realidade virtual e aumentada tornarem-se mais comuns no meio profissional.

### 2.1.3 Visão geral

Dada a literatura revisada até aqui, foi possível fundamentar a sequência deste estudo. Foram elaboradas duas tabelas comparativas, contrapondo os trabalhos relacionados com os sistemas atuais utilizados pela distribuidora de energia e também com a plataforma cuja abordagem deste trabalho será estruturada - nomeada como ELVIS. A Tabela 2.1 compara a relação das técnicas de visualização e a Tabela 2.2 as dimensões de dados apresentadas.

Tabela 2.1: Técnicas de visualização propostas pelos trabalhos relacionados

Técnicas de visualização propostas (com ou sem desenvolvimento de plataforma)															
Trabalho	Espacial 2D	Espacial 3D	Realidade aumentada	Linhas de contorno	Gráfico de setores	Diagrama Unifilar	Barras ou linhas	Gráfico espiral	Diagrama de dispersão	Colunas paralelas	Andrew Curve	Campo de vetores	Diagrama de cordas	Grafo ponderado	Outros
(KLUMP; WEBER, 2002)	✓	✓		✓	✓	✓									
(KLUMP; DOOLEY; WU, 2003)	✓	✓		✓	✓	✓	✓								
(MILANO, 2009)				✓		✓	✓								
(FARFÁN; PALTA; TACCA, 2011)				✓		✓									
(NGA et al., 2012)	✓			✓		✓	✓		✓	✓	✓				
(USTARIZ-FARFAN et al., 2012)	✓			✓		✓	✓								
(WOSNY et al., 2012)	✓														
(BARCELOS et al., 2013)			✓												
(DUTTA; OVERBYE, 2013)							✓		✓		✓				✓
(POURMIRZA; BROOKE, 2013)	✓					✓									
(CUFFE; KEANE, 2015)	✓					✓								✓	
(KUMAR; SINGH, 2015)				✓											
(MANGISI, 2015)	✓				✓		✓		✓						
(NIGAR; AGALGAONKAR; CIUFO, 2015)	✓			✓											
(OVERBYE; WEBER, 2015)	✓			✓	✓	✓	✓		✓		✓				✓
(SRINIVASAN; REINDL et al., 2015)	✓						✓								
(SUGINO et al., 2016)	✓													✓	
(CAPUDER; PANDZIC; KUZLE, 2017)	✓			✓	✓	✓	✓	✓	✓						
(OLIVEIRA et al., 2017)	✓	✓													
(ZAIDI et al., 2017)	✓				✓										
(LAWANSON et al., 2018)	✓			✓	✓		✓								
(MOLNAR; GRUCHALLA, 2018)												✓	✓		
(LU et al., 2019)	✓			✓	✓		✓	✓	✓	✓					
Sistema atual 1	✓					✓									
Sistema atual 2	✓					✓	✓								
ELVIS	✓			✓			✓	✓	✓	✓					✓

Fonte: O Autor

Para que fosse possível elaborar uma comparação prática das técnicas de visualização destes trabalhos, sem diversificar demais a quantidade de técnicas apresentadas, presumiu-se que era necessário agrupá-las de acordo com as suas similaridades. Isto é consequência de que as visualizações, em geral, podem possuir adaptações realizadas pelos autores para melhor atender aos objetivos específicos de cada trabalho. Logo, vale sugerir ao leitor que consulte as referências para maiores detalhes. Por exemplo, gráfi-

Tabela 2.2: Temas apresentados pelas técnicas de visualização dos trabalhos relacionados  
Principais temas/dados apresentados na forma visual

Trabalho	Carga do sistema	Densidade de UC's	Elementos da rede	Fluxo de energia	Linhas de transmissão	Dados técnicos	Smart Grid	Qualidade de energia	Geração Distribuída	Reclamações
(KLUMP; WEBER, 2002)				✓	✓					
(KLUMP; DOOLEY; WU, 2003)	✓		✓	✓		✓				
(MILANO, 2009)				✓		✓				
(FARFÁN; PALTA; TACCA, 2011)				✓		✓		✓		
(NGA et al., 2012)	✓					✓	✓			
(USTARIZ-FARFAN et al., 2012)						✓		✓		
(WOSNY et al., 2012)			✓		✓					
(BARCELOS et al., 2013)			✓							
(DUTTA; OVERBYE, 2013)						✓				
(POURMIRZA; BROOKE, 2013)							✓			
(CUFFE; KEANE, 2015)					✓	✓				
(KUMAR; SINGH, 2015)				✓						
(MANGISI, 2015)							✓			
(NIGAR; AGALGAONKAR; CIUFO, 2015)						✓		✓	✓	
(OVERBYE; WEBER, 2015)	✓		✓	✓		✓	✓			
(SRINIVASAN; REINDL et al., 2015)							✓			
(SUGINO et al., 2016)				✓		✓				
(CAPUDER; PANDZIC; KUZLE, 2017)	✓		✓			✓				
(OLIVEIRA et al., 2017)	✓	✓								
(ZAIDI et al., 2017)									✓	
(LAWANSON et al., 2018)	✓	✓	✓		✓	✓	✓	✓		
(MOLNAR; GRUCHALLA, 2018)				✓	✓					
(LU et al., 2019)	✓		✓	✓	✓	✓	✓	✓		
Sistema atual 1	✓	✓	✓	✓	✓	✓				
Sistema atual 2	✓				✓	✓				
ELVIS	✓	✓	✓		✓			✓	✓	✓

Fonte: O Autor

cos de barras e linhas foram unificados em função da similaridade, e linhas de contorno (2D) incluem também superfícies de contorno (3D). Foram considerados trabalhos que apresentam visualmente as técnicas empregadas - ou seja, técnicas constantes apenas na revisão literária ou descritas textualmente não foram consideradas.

Importante salientar também que, para a elaboração das comparações, foram consideradas todas técnicas de visualização que os autores utilizaram nos estudos. Ou seja, a tabela inclui estudos que apresentam técnicas de visualização na área de energia elétrica, porém podem não ter desenvolvido plataformas de visualização propriamente ditas.

De acordo com a Tabela 2.1, pôde-se perceber que a visualização baseada em mapa (espacial 2D e 3D), com a localização geográfica dos equipamentos, é considerada

em muitos trabalhos. Isto alinha-se à ideia de que é uma necessidade e uma expectativa dos usuários que as abordagens possuam esta funcionalidade. É perfeitamente entendível, pois, visto que o cotidiano dos técnicos envolve rotinas de inspeção e manutenção em campo, a orientação por mapa acaba por tornar a interação muito mais familiar, com o reconhecimento dos equipamentos que se está visualizando. Importante esclarecer que não foi um objetivo deste trabalho utilizar a maior quantidade possível de diferentes técnicas de visualização, foram escolhidas as mais adequadas para atender aos requisitos, e que serão detalhadas no próximo capítulo.

Em relação aos sistemas atualmente disponíveis na empresa, ficou visível com a Tabela 2.1, a reduzida quantidade de análises visuais que é possível realizar. Os sistemas possuem basicamente as visualizações espacial em duas dimensões, diagrama unifilar (esquemas elétricos) e gráficos de barras ou linhas. A Tabela 2.2 deixou evidente que as categorias de qualidade de energia, geração distribuída e reclamações não estão suficientemente cobertas por estes sistemas. O software que controla estes dados é o próprio Sistema de Gestão da empresa - ERP (do inglês, *Enterprise Resource Planning*), que não possui visualizações gráficas destes dados.

Ficou evidente também, de acordo com a Tabela 2.2, que não foi encontrado na literatura trabalhos que focassem na qualidade da energia distribuída a ponto de considerar, de fato, os elementos físicos da rede, geograficamente apresentados na rede de distribuição de energia. Na pesquisa realizada, não foram encontrados trabalhos que consideram componentes como o carregamento do transformador e perdas técnicas de energia, ou tampouco os tipos de consumidores e a existência de reclamações, para avaliar a qualidade da energia do sistema da distribuidora - de ponta a ponta, das subestações até os consumidores. Neste contexto, confirma-se a existência de um nicho de pesquisa a ser explorado.

## **2.2 Clusterização de dados na área elétrica**

Esta seção é voltada especificamente para o estudo dos trabalhos relacionados ao modelo de avaliação da qualidade da energia elétrica dos consumidores, baseado em técnicas de clusterização. Como Mulliez et al. (2018) afirmou, isto possibilita que operadores da distribuição da energia concentrem-se em um ou mais grupos (*clusters*), ao invés do montante inteiro de consumidores, para analisar a qualidade da distribuição. A seguir são apresentados os principais trabalhos realizados na área.

A consulta realizada levantou duas grandes áreas em que a clusterização foi abordada, no contexto desta dissertação. A primeira, voltada à qualidade de energia, porém baseada em dados obtidos principalmente da operação técnica da rede de distribuição (subestações, medições de grandezas elétricas, carga e consumo, entre outros). A segunda, voltada à clusterização de consumidores para fins comerciais e análise de mercado. A seguir são sucintamente apresentados estes trabalhos, subdivididos nestas categorias.

### 2.2.1 Clusterização em qualidade de energia

Esta subseção terá maior ênfase por estar mais alinhada à finalidade deste estudo e apresentará os principais trabalhos relacionados à clusterização ou mineração de dados voltada à qualidade de energia.

Asheibi et al. (2004) abordou mineração de dados não supervisionada, utilizando dados de medições de grandezas elétricas (como tensão, corrente e harmônicas) em transformadores em diferentes localizações: regiões sumariamente residenciais, comerciais ou industriais. O trabalho de Overbye e Weber (2015), já citado anteriormente e ilustrado na Figura 2.8-A, clusterizou dados no intuito de identificar pontos fora do padrão (*outliers*) em relação à geradores de energia.

Kapoor et al. (2018) desenvolveram seu trabalho com o objetivo de propor um método para identificar problemas de qualidade de energia identificando padrões através de aprendizado supervisionado, com *Support Vector Machine* (SVM) e Histograma de Gradientes Orientado (*Histogram of Oriented Gradients* (HOG)), calculando perturbações na rede principalmente em relação à harmônicas.

Outro trabalho interessante e relacionado ao presente estudo, avalia a qualidade da energia inclusive com o impacto da geração distribuída, diferenciando quando a mesma está ativa, desligada ou ainda quando são notadas alterações na carga da rede (JASIŃSKI; SIKORSKI; BORKOWSKI, 2019).

O trabalho de Mulliez et al. (2018) utiliza clusterização para avaliar a vida útil de transformadores de distribuição, utilizando dados como o seu carregamento com leituras horárias – dado que também será utilizado no nosso modelo. No caso do presente estudo, porém, serão utilizados dados de consumo dos últimos 12 meses para realizar os cálculos relativos a cada transformador.

Jasiński et al. (2016) estudaram a qualidade de energia com dados técnicos medidos em subestações, focando em distúrbios na rede, como: variação de frequência, alte-



rações ou flutuações de tensão e distorções harmônicas. Góes, Steiner e Peniche (2015) aplicaram técnicas de classificação em qualidade de energia, baseada em afundamentos de tensão, utilizando também outras técnicas de mineração de dados.

Neste contexto, muitos autores ainda utilizam técnicas de clusterização e redes neurais *fuzzy*, além de processamento de sinais. Podemos citar Duan et al. (2006), que combina qualidade de energia com o preço praticado, Seera et al. (2015) que propôs um modelo híbrido que utiliza rede neural fuzzy e árvore de clusterização, e Gargoom, Ertugrul e Soong (2005), que trabalharam com ferramentas de processamento de sinais para clusterização de eventos relacionados à qualidade da energia.

De fato, percebeu-se que a quantidade de trabalhos da área de clusterização na área elétrica em geral é relativamente grande, e este número é ainda maior caso o termo seja generalizado para mineração de dados. Para exemplificar, foram encontrados ainda estudos que focaram em geração distribuída (JUREEDI; ROSALINA; KUMAR, 2013), perdas de energia (GRIGORAS; CARTINA; ROTARU, 2010; GRIGORAS et al., 2011) e redes inteligentes (FLATH et al., 2012), entre outros que poderiam estender a variedade de aplicações. Contudo, foram citados aqui aqueles relacionados à qualidade.

### **2.2.2 Clusterização de consumidores**

Visto que esta dissertação é voltada à análise da qualidade para o usuário final, de fato devem ser relacionados a este os dados que serão utilizados para realizar a clusterização. Em função disto, é conveniente apresentar trabalhos relacionados à clusterização de consumidores, porém podendo não estar relacionado à qualidade da energia.

Clusterização de dados e outras técnicas de mineração têm sido utilizadas para identificação de padrões relacionados à fraudes e furtos de energia, conhecidos como perdas comerciais. Zubelli (2017) é um dos autores que utilizou clusterização voltada à esta finalidade.

Técnicas de classificação de consumidores estão sendo utilizadas em grande parte para análise de mercado e também para embasar estudos tarifários do serviço de distribuição de energia. Grigoras e Bobric (2013) desenvolveram seu trabalho nesta área, utilizando dados como o consumo dos clientes.

Vale citar ainda que a própria ANEEL utiliza métodos de clusterização para realizar a regulação do setor elétrico, determinando limites similares para as distribuidoras que possuem características parecidas em relação às unidades consumidoras. Mais detalhes

podem ser encontrados em (BERNARDO, 2013) e (PESSANHA; SOUZA; LAURENCEL, 2007), que comentam sobre a evolução da qualidade do serviço de distribuição de energia no Brasil.

### **2.2.3 Visão geral**

Como pôde ser visto, a análise de clusters é bastante utilizada pela comunidade para aprimorar as avaliações de qualidade de energia (JASIŃSKI; SIKORSKI; BORKOWSKI, 2019). Fica claro também que foram encontrados trabalhos que utilizaram dados de componentes de distribuição de energia (como subestações e transformadores, com base em medições de grandezas elétricas e perturbações na rede) e dados de consumidores (em grande parte, da área comercial). Porém, não foram identificados estudos que utilizaram simultaneamente ambos para a finalidade de análise da qualidade para o usuário final, a ponto de considerar, de fato, os dados das próprias unidades consumidoras.

É este ramo que será explorado, procurando prospectar a possibilidade de obter agrupamentos adequados, resultando, caso positivo, em um modelo de avaliação de qualidade da energia com uma abordagem centrada no consumidor final.

### 3 VISUALIZAÇÃO DA QUALIDADE DA DISTRIBUIÇÃO DE ENERGIA

Este capítulo apresenta a abordagem visual proposta pelo presente estudo. Inicia por uma breve introdução sobre redes de distribuição, na sequência a descrição dos requisitos do sistema e dos dados a serem visualizados, com posterior detalhamento das técnicas de visualização e funcionalidades implementadas. A clusterização será tratada no capítulo seguinte.

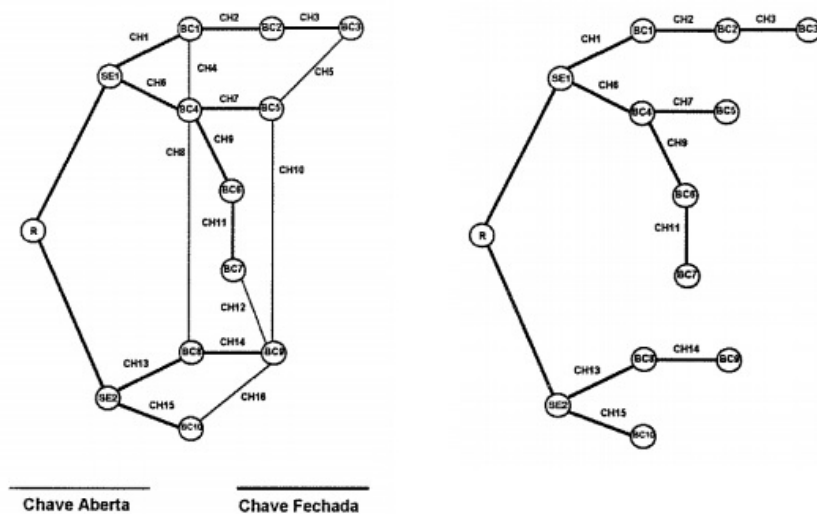
#### 3.1 Conhecimento prévio sobre redes de distribuição

Sob certo ponto de vista, os sistemas de distribuição de energia podem ser representados por conjuntos de vértices e arestas, que formam um grande grafo: os vértices são os postes onde estão alocados os equipamentos de distribuição (ou ainda consumidores, dependendo do tipo de visualização que se deseja empregar), e as arestas representam os trechos de rede (cabos) que conectam os equipamentos e consumidores. Estas redes, em seu estado normal de operação (energizadas), podem ser consideradas estruturas do tipo árvore, pois à medida em que a profundidade aumenta, os vértices não se conectam novamente.

Já considerando a rede desenergizada, a estrutura não é do tipo árvore em função de haver conexões que permitem alimentar energeticamente regiões por outros locais, através de chaveamento e redundância, com o objetivo de garantir a maior continuidade do fornecimento em caso de interrupção (na área elétrica, isto é conhecido como anel elétrico). A Figura 3.1 exemplifica este caso, utilizado em um estudo sobre perdas de energia nos sistemas de distribuição. Na imagem, o lado esquerdo mostra a estrutura do tipo grafo (com o chaveamento e a parte não energizada), e o lado direito a estrutura do tipo árvore (apenas com a parte energizada). Os nós (vértices) representam subestações (SE) ou blocos de carga (BC) do sistema (grandes grupos de consumidores, no caso daquele estudo).

Mesmo as redes sendo grandes e dinâmicas, passíveis de modificação estrutural e de seus componentes ao longo do tempo (KLUMP; DOOLEY; WU, 2003; MOLNAR; GRUCHALLA, 2018), as mudanças não são imediatas, e eventos que alteram a topologia da rede, quando em um estado normal, são eventuais. Por exemplo, uma região atendida por um determinado traçado de rede pode ser atendida por outro quando ocorrer uma interrupção, mediante técnicas citadas anteriormente. Para algumas finalidades, as apre-

Figura 3.1: Visualização de grafo e árvore de uma parte da rede de distribuição de energia



Fonte: Cavellucci et al. (1998)

sentações de dados em um determinado momento no tempo pode ser um problema, pois exibe apenas o estado do sistema naquele instante (DUTTA; OVERBYE, 2013).

Considerando isto, pode-se afirmar que este não é um problema nesta pesquisa, pois a qualidade da distribuição de energia deve considerar o estado do sistema em um instante normal, desde que seja recente para representar a topologia atual da rede. Desta forma, foi definida a premissa de que as técnicas de visualização poderão considerar a topologia de árvore (MOLNAR; GRUCHALLA, 2018).

Como explicado anteriormente, a visualização geográfica tornou-se uma expectativa dos usuários. Inclusive, autores afirmam que o georreferenciamento da visualização do sistema é uma necessidade em constante crescimento para as empresas que distribuem energia (CAPUDER; PANDZIC; KUZLE, 2017).

Desta forma, a utilização da plataforma é baseada em duas funcionalidades principais: 1) visualização geral e geográfica da rede elétrica, permitindo interações para aumentar o nível de detalhe sob demanda - como *zooming* para visualização de mais dados à medida em que o mapa é ampliado, e 2) visualizações derivadas dos componentes e equipamentos selecionados pelo usuário, através da interação sobre o mapa e seleção dos componentes que se deseja visualizar. Este será o princípio de funcionamento da abordagem proposta.

### 3.2 Descrição dos requisitos

Para entender as necessidades dos usuários, foi feito um levantamento de informações através de entrevistas com três engenheiros eletricitas e quatro técnicos em eletrotécnica, responsáveis pela manutenção do sistema elétrico e pelo tratamento das reclamações de qualidade de energia. Os engenheiros possuem, em média, 17 anos de experiência na empresa (desvio padrão (dp) 9.5), e os técnicos possuem 12 (dp 9.6). Ambos os períodos desconsideraram outros trabalhos na área elétrica. Não houve perguntas predefinidas, sendo que as entrevistas foram conduzidas no formato de diálogo, permitindo ampla interação e poucas limitações para o entrevistado, deixando clara a sua expectativa (se comparado a questionário, por exemplo) (YOUNG, 2002). O tempo de duração das entrevistas variou de acordo com a disponibilidade - as mais curtas cerca de 20 minutos enquanto as mais longas em torno de 1h30min.

Além do levantamento inicial, as entrevistas foram realizadas frequentemente, principalmente durante o planejamento da forma de abordagem visual: à medida em que funcionalidades relevantes ou grupo de funcionalidades foi implementado, foram realizadas validações e ajustes. Desta forma, a própria plataforma serviu como um excelente meio de comunicação entre os autores e os entrevistados, permitindo propor uma construção de abordagem visual intuitiva e personalizada de acordo com as atividades (YOUNG, 2002).

Durante as entrevistas, foram identificados os seguintes requisitos para construir uma abordagem visual que atinja os objetivos propostos:

- R1: Basear-se em visualizações geográficas regionais e detalhadas, exibindo as características relevantes do sistema naquela área;
- R2: Permitir interatividade, quanto à identificação, busca e extração de informações de elementos específicos;
- R3: Construir análises peculiares por consumidor, quanto às características da rede elétrica e equipamentos em seu entorno (dados do transformador, distância de equipamentos e outros atributos relevantes);
- R4: Apresentar dados de vizinhança com carga ou consumo elevado e a existência de geração distribuída;
- R5: Construir análises de carregamento e balanceamento de fases dos transformadores de distribuição;

- R6: Apresentar visualizações estruturadas das seguintes informações, vinculadas ao mapa, quando apropriado:
  - Reclamações críticas e precárias ao longo do tempo (estas classificações são definidas por norma e apresentadas na sequência do trabalho);
  - Distâncias dos consumidores da subestação e do transformador;
  - Carga dos transformadores ao longo dos últimos 12 meses;
  - Densidade das reclamações no mapa;
  - Densidade das usinas de geração distribuída;
  - Apresentar a queda de tensão para consumidores.
- R7: Apresentar o resultado do modelo de qualidade de energia, com o devido agrupamento dos consumidores de acordo com as suas características de qualidade de energia.

Os requisitos referentes a visualizações geográficas, dados dos transformadores e consumo foram os mais lembrados, principalmente devido a estas dimensões serem comuns para atender diferentes perfis de usuário. Conforme avaliado nos capítulos seguintes, além destes requisitos, intrinsecamente foi também definido que a ferramenta deveria ser intuitiva e de fácil aprendizado, bem como suficientemente ágil para a utilização adequada.

No planejamento da implementação do sistema, foi utilizada uma abordagem baseada em tarefas. Isto significa que as principais atividades realizadas pelos usuários para identificação, busca, visualização e estudo das informações procuraram ser amparadas pelas funcionalidades da plataforma. Alicerçado nestes requisitos, foram estudadas as técnicas de visualização mais apropriadas e planejada a forma de construção da abordagem visual. Obteve-se, ao final, a implementação de mais de 8 tipos de visualização, entre geográfica, de contornos, mapa de dispersão, radar, coordenadas polares, linhas/barras, bolhas e *treemap*, distribuídas em diversas funcionalidades.

### **3.3 Descrição do conjunto de dados**

A identificação dos dados realmente proveitosos é uma das tarefas mais importantes do projeto (DUTTA; OVERBYE, 2013). Esta seção descreve os dados que serão apresentados, como também define termos normativos da área de energia. Pode-se segregar os

dados em três partes principais: 1) Componentes do sistema elétrico; 2) Consumidores; e 3) Reclamações de qualidade da energia. Os dados foram selecionados com base nas entrevistas com os especialistas, sendo que foram agregadas novas dimensões à medida em que o projeto avançava.

Para embasar trabalhos futuros, é importante que sejam detalhadas as entidades utilizadas. Desta forma, nas subseções abaixo, serão explanadas separadamente cada classificação, entidades e atributos.

### 3.3.1 Componentes do sistema elétrico

Os componentes do sistema elétrico selecionados para este estudo foram estruturados nas entidades abaixo. Estes dados representam as características atuais do sistema elétrico da distribuidora, e pode-se dizer que são universais em empresas de distribuição de energia.

- Poligonais: Segmentos de reta baseados em coordenadas;
- Subestações: Nome da subestação;
- Alimentadores: Subestação, número e nome do alimentador;
- Postes: Identificação e Coordenadas;
- Transformadores: Coordenadas, Carregamento em kVA e pontos percentuais, Potência em kVA e Tipo (monofásico, bifásico ou trifásico);
- Reguladores de tensão: Identificação, Coordenadas e Potência;
- Capacitores: Identificação, Coordenadas e Capacidade nominal;
- Cabos: Identificação do poste de origem e do poste de destino, Comprimento do vão, Tensão (baixa, média ou alta tensão), Tipo (monofásico, bifásico ou trifásico), Com ou sem alma de alumínio, Seção, Material de composição (em geral, cobre ou alumínio) e Isolamento (nu, coberto, isolado);

Os dados desta categoria são primordiais para análise da qualidade da energia. Em função disto, merecem destaque nesta pesquisa. A entidade *Poligonais* define a área de atuação da distribuidora de energia. É formada por segmentos de reta - coordenadas de início e fim, que formam os polígonos onde a empresa é responsável pelo atendimento.

Em geral, problemas de energia nas subestações e alimentadores são imediatamente identificados e resolvidos com urgência através de monitoramento em tempo real e

automação. Porém, ainda são importantes neste trabalho para representar a hierarquia do sistema elétrico.

Contudo, a partir dos alimentadores, a relação custo/benefício para o monitoramento dos demais equipamentos torna-se elevado, e ainda são restritos os locais onde esse monitoramento é implantado. São exceções alguns reguladores e religadores de tensão, que são relativamente poucos na rede em função do custo e da capacidade de atendimento - cada um pode beneficiar inúmeros consumidores.

A visualização dos postes no sistema tem um mero objetivo de tornar mais realista a apresentação do mapa, como também facilitar a geolocalização dos operadores, facilitando o reconhecimento da área. Tecnicamente ele não representa influência nenhuma na qualidade da energia. Contudo, entidades como transformadores, reguladores de tensão, capacitores e cabos possuem diversos atributos que influenciam significativamente, e são cruciais para análise individual dos consumidores ou região que está sendo analisada.

### **3.3.2 Consumidores**

Referente aos consumidores, foram definidas como importantes as entidades e os atributos abaixo. Foram coletados todos os consumidores do sistema, possuindo conexões desde o ano de 1975.

- Consumidor: Identificação, Coordenadas, Transformador, Alimentador, Data de ligação, Corrente máxima do disjuntor, Nome do município, Tensão nominal, Endereço, Classe, Subclasse, Grupo, Fase, Possui ou não geração distribuída, Conjunto, Carga instalada;
- Unidades de geração distribuída: Identificação do consumidor, Nome do titular, Tipo (usina fotovoltaica, usina termelétrica, ou outro tipo de geração), Quantidade de módulos, Quantidade de inversores, Potência dos módulos, Potência dos inversores, Data da conexão, Fonte (exemplo no caso de termelétrica: biogás), Potência Aparente, Potência Instalada, Fases do inversor e Fase ligada;

Existem diversos atributos adicionais dos consumidores que servem a título de informação (como o endereço). Ao selecionar um consumidor específico, é importante ressaltar que a Fase, Carga Instalada, Geração Distribuída e a Tensão Nominal são as variáveis significantes nesta classificação, em função de influenciarem na análise da qualidade da energia.



A fase representa a quais fases da rede o consumidor está ligado (A, B, C, ou a combinação destas). A geração distribuída indica se esta unidade possui geração própria para complementar o consumo - em função disto, quando a geração é maior do que o consumo, o mesmo está entregando energia à rede e isto pode influenciar na qualidade da energia dos vizinhos, variando de acordo com a potência da usina.

A tensão nominal caracteriza a tensão que alimenta o consumidor (220V, 440V ou 380V) e também é uma variável importante que auxilia os operadores na análise da energia. As datas de ligação da unidade consumidora e data de conexão da unidade geradora, quando existentes, são importantes para comparações cronológicas da situação da rede para cruzamento com as reclamações registradas, além de apresentar visualmente o incremento e alterações do sistema elétrico ao longo do tempo.

Os atributos Classe, Subclasse e Grupo são definidos de acordo com a Resolução Normativa nº 414 da ANEEL. Cada um define a classificação da unidade consumidora com base na tensão de fornecimento e na finalidade da utilização da energia. Detalhes podem ser visualizados em Agência Nacional de Energia Elétrica - ANEEL (2010). Estas informações são relevantes pois o grupo e a classe podem indicar características de consumo, como industriais ou de serviços que demandam grande quantidade de energia e que conseqüentemente podem causar oscilações ou degradação da qualidade da energia na vizinhança.

### **3.3.3 Processos de qualidade de energia**

Esta classificação inclui todos os processos de qualidade de energia, desde o ano de 2007. Estes processos incluem reclamações e medições amostrais para a ANEEL.

- **Processos:** Identificação do processo, Origem (consumidor ou amostra para a ANEEL), Situação (processo concluído ou em andamento), Identificação da unidade consumidora, Identificação e nome do consumidor, Anormalidade descrita pela atendente ao registrar a reclamação, Data e hora do registro, Resultado da análise da reclamação, Data e hora da análise, Data e hora do encerramento, Trimestre e Ano do registro;
- **Medições de tensão:** Identificação da medição e do processo vinculado, Identificação da unidade consumidora e nome do titular, Data e hora do início da medição, Data e hora do fim da medição, Situação (encerrada ou em andamento), se foram

violados os limites adequados de tensão, Valor do índice de duração relativa da transgressão para tensão precária (DRP) e para tensão crítica (DRC), Se houve necessidade de intervenção para normalizar a tensão.

Além do próprio interesse do consumidor na qualidade da energia, o contexto é regulado através da determinação de requisitos normativos pela ANEEL, que regulamenta a distribuição de energia e penaliza as distribuidoras quando são extrapoladas estas exigências. A ANEEL determina, por exemplo, como deve funcionar o processo de reclamação de níveis de tensão, ou seja, reclamações da qualidade do fornecimento.

É válido ao leitor entender, embora sucintamente apresentada nesta dissertação, como é executado este processo em uma distribuidora de energia. Informações detalhadas podem ser encontradas no Módulo 8 - Qualidade da energia elétrica do PRODIST (AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA - ANEEL, 2017).

### *3.3.3.1 Funcionamento do processo da reclamação*

Ao identificar problemas no fornecimento da energia (como tensão baixa, tensão alta ou oscilações de tensão, em termos populares), o consumidor pode solicitar o registro de uma reclamação de qualidade de energia, junto à sua distribuidora. A distribuidora, em seguida, realiza uma medição instantânea da tensão, diretamente na conexão da unidade consumidora com o sistema elétrico, e poderá identificar se há ou não transgressão dos limites naquele instante de tempo. Caso não identificar, o consumidor pode solicitar uma medição com maior duração (168 horas) com equipamento registrador, que coleta as grandezas elétricas a cada 10 minutos, devendo haver no mínimo 1008 leituras válidas.

Caso não for identificada transgressão dos limites, o processo é encerrado. Caso houver, o processo continua até que o problema seja resolvido. Ao longo deste período, desde a reclamação até a comprovada solução do problema através de nova medição de 168 horas, o consumidor é compensado financeiramente na fatura de energia elétrica, com base nos cálculos definidos pela ANEEL (AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA - ANEEL, 2017). As possíveis soluções podem ser desde ajustes no ramal de ligação, ajustes no transformador ou mesmo obras de adequação da rede elétrica.

Preferencialmente, as medições instantâneas devem ser realizadas no dia e horário indicado pelo consumidor, pois a qualidade da energia pode ser horária, estando inadequada em determinados períodos do dia (como a ligação de um motor em um consumidor próximo). Em geral, estas eventualidades são registradas nas medições de 168 horas.

Este trabalho também coletou dados históricos de amostras de medição de tensão. A amostra da energia de consumidores é definida trimestralmente através de software da própria ANEEL, sendo que para cada unidade consumidora sorteada, deve ser realizada uma medição de 168 horas, também com 1008 leituras válidas (AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA - ANEEL, 2017). Estas informações são exibidas no mapa geográfico, tornando possível a identificação de consumidores próximos caso a consulta às medições realizadas seja necessário.

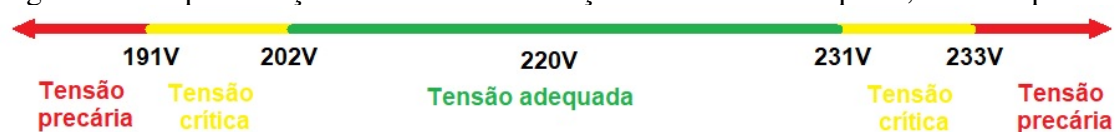
### 3.3.3.2 Contextualização

As reclamações são relevantes neste trabalho pois a sua representação visual no mapa pode indicar regiões onde há transgressão dos limites aceitáveis. O resultado da reclamação (se for procedente, no caso da tensão estar inadequada, ou improcedente, no caso da tensão estar dentro dos limites aceitáveis) é relevante de qualquer maneira na análise realizada pelos operadores.

A justificativa disto é que, mesmo a tensão estando dentro dos limites adequados, em determinados casos, distúrbios ainda podem existir, causando problemas na energia recebida pelo consumidor. Por exemplo, o consumidor pode não optar pela medição de 168 horas, e desta maneira uma reclamação improcedente continuar a causar problemas. Eventualmente, picos mínimos ou máximos de tensão, inferiores a 10 minutos, podem não ser registrados pelo equipamento registrador. Conseqüentemente a reclamação é encerrada como improcedente (pois está dentro dos limites) - contudo o consumidor pode continuar insatisfeito.

Conforme estabelece o Módulo 8 do PRODIST, os limites aceitáveis são de 202V a 231V para cada fase de 220V (utilizada na distribuidora em estudo). As medições realizadas podem resultar em tensão **adequada**: dentro dos limites mínimo e máximo, **precária**: fora dos limites adequados e entre 191V e 231V, ou **crítica**: menor que 191V ou maior que 233V (AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA - ANEEL, 2017). A Figura 3.2 representa esta classificação.

Figura 3.2: Representação visual da classificação das tensões adequada, crítica e precária



Fonte: O Autor

Os registros nas tensões crítica e precária são posteriormente calculados como ín-

lices, conhecidos por DRP e DRC, de acordo com as fórmulas e critérios estabelecidos pela ANEEL. São utilizados também para determinar os valores de compensação financeira ao consumidor. Os limites dos indicadores são de 3% para DRP e 0.5% para DRC (AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA - ANEEL, 2017).

### **3.3.4 Discussão do conjunto de dados**

Numericamente, são 14 entidades representadas pelos dados, com mais de 100 dimensões caso somadas. O volume de unidades consumidoras da distribuidora em estudo é de aproximadamente 70.000, com 380 usinas de geração distribuída (com grande crescimento atualmente). São mais de 5.300 processos de qualidade de energia (3.054 reclamações e 2.269 amostras) e mais de 6.000 medições de 168h. Em relação aos equipamentos de distribuição com relação a qualidade, são 21 reguladores de tensão, 68 bancos de capacitores e mais de 8.100 transformadores distribuídos pela sua área. Importante salientar que este estudo pode ser aplicado a outras distribuidoras, além de ser totalmente escalável, pois altera-se, na prática, o volume de dados.

Dada a apresentação e descrição do conjunto de dados, é possível perceber a alta dimensionalidade dos mesmos e os desafios para encontrar meios adequados para sua visualização de maneira intuitiva e suficientemente detalhada.

Também é notável, através da identificação dos dados selecionados, a abordagem de qualidade de energia centrada no fornecimento até o consumidor final, diferentemente de outros trabalhos que a consideram até as subestações ou, em alguns casos, até o transformador.

### **3.4 Visualizações e funcionalidades implementadas**

Esta seção apresentará as funcionalidades implementadas, juntamente com as técnicas de visualização que foram empregadas. Para ser flexível e proporcionar interoperabilidade, a plataforma foi baseada em Web. Isto facilita o acesso, sem requerer instalações de software, além de ser independente da plataforma ou sistema operacional que o usuário utiliza (YANG; GUO, 2007; SILVA, 2014). Os dados não representam informações críticas quanto à segurança da informação, principalmente por não serem utilizados dados pessoais para a finalidade ou, ainda, informações estratégicas da companhia. Além

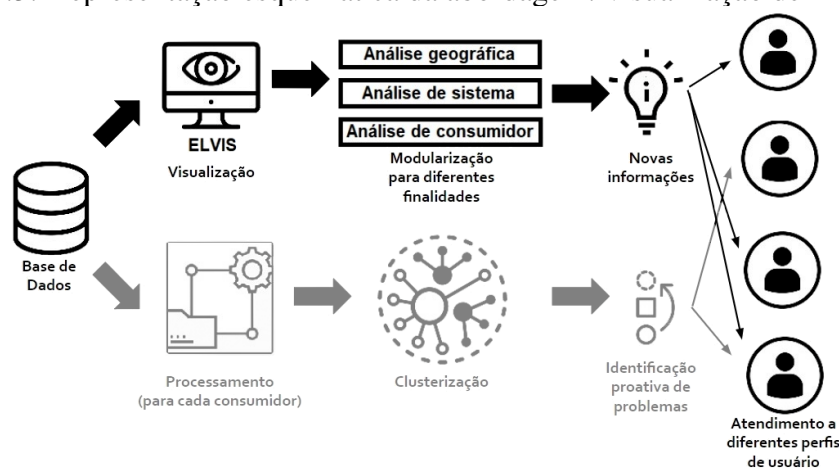
disso, o acesso à plataforma permanece através da rede interna da distribuidora. Os dados foram exportados dos sistemas informatizados da empresa no formato CSV, que por sua vez foram preprocessados e adaptados para serem consumidos pela plataforma.

Um ponto negativo a ser considerado pode ser a limitada capacidade de processamento no navegador (YANG; GUO, 2007). Porém, acredita-se que isto não é um problema no projeto atual, devido ao custo computacional não ser tão elevado a ponto de prejudicar a usabilidade dado que processos mais caros computacionalmente são realizados durante a inicialização do sistema. Os experimentos realizados, detalhados nos capítulos seguintes, confirmaram que foi obtida uma ótima usabilidade e o desempenho não foi criticado. Em versões futuras, considerando o aumento do volume de dados, poderão ser consideradas outras formas para aprimorar o processamento.

O sistema desenvolvido foi, após os experimentos, intitulado pelos próprios usuários de ELVIS - *ELectric energy VISualization*). A representação esquemática do ELVIS é mostrada na Figura 3.3 (parte superior), sendo sua abordagem visual modularizada para diferentes finalidades, dividida em 3 grandes seções:

- Análise geográfica, que incorpora o mapa e a localização geográfica das redes e equipamentos, além de outras visualizações e funcionalidades;
- Análise de sistema, cuja função é permitir análises voltadas ao sistema de distribuição de energia como um todo;
- Análise de consumidor, onde o usuário escolhe uma unidade consumidora para iniciar a análise. Nesta seção, todos os dados e visualizações são processadas de acordo com as características da rede elétrica peculiares a este consumidor.

Figura 3.3: Representação esquemática da abordagem: visualização de informações



As seções supracitadas são utilizadas para organizar a sequência de descrição, finalizando-se este capítulo com uma seção de discussão.

### 3.4.1 Análise Geográfica

Visto que os dados tem a característica de possuírem uma relação muito próxima à localização geográfica, a aplicação necessariamente precisa atender a esta necessidade - como também visto na revisão literária. Portanto, a visualização principal da abordagem será baseada em mapas. As técnicas tridimensionais possuem grande poder de permitir ao usuário uma ótima interpretação dos dados, devido à sua escalabilidade de atributos (ver (OLIVEIRA et al., 2017)). Porém, ao passo em que os dados tornam-se mais volumosos, a visualização torna-se confusa e mais difícil de interpretar. Atrelada à diminuição da área das formas geográficas tridimensionais, técnicas bidimensionais são as mais indicadas nestes casos.

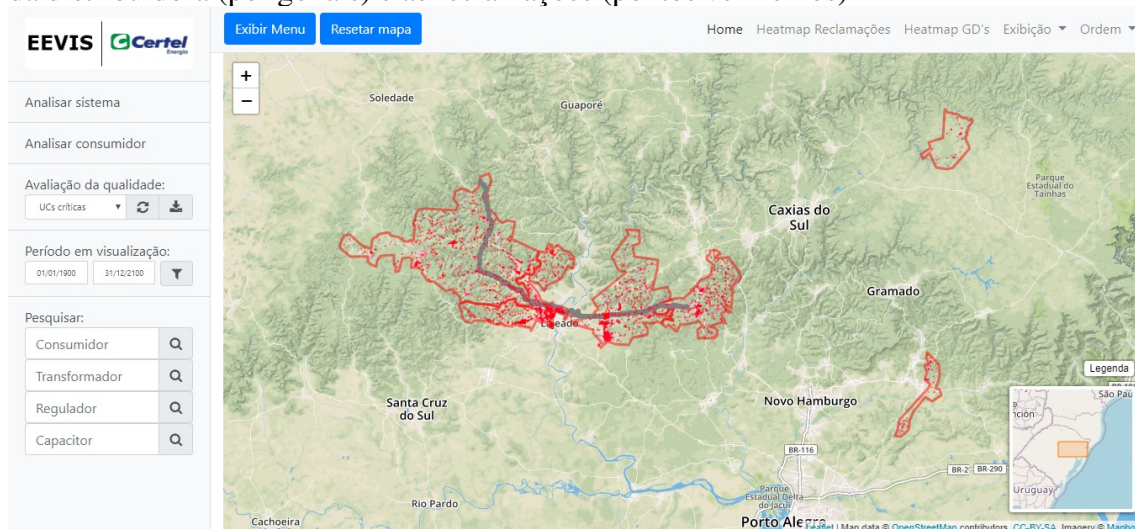
Visto isto, conclui-se que a representação geográfica será bidimensional, referenciando a localização dos diversos componentes da rede elétrica. A própria visualização da estrutura e hierarquia da rede é facilitada através do georreferenciamento dos componentes, facilitando a compreensão (CAPUDER; PANDZIC; KUZLE, 2017).

Como tem-se uma elevada quantidade de dados, neste projeto foram utilizados direcionamentos sugeridos por autores da área de *Big Data*. A primeira visualização, ao acessar a plataforma, é uma visão geral espacial do sistema elétrico com a área de atuação da distribuidora e todos os registros de reclamações identificados sobre a mesma. Isto atende à premissa de criar uma visão geral primeiro, enquanto a visualização é de alto nível, para depois aumentar a quantidade de informações à medida em que o usuário deseja aproximar-se de determinada região, evitando que sejam apresentados detalhes de informação que não são pertinentes neste momento (SHULL, 2013; OVERBYE; WEBER, 2015). Esta técnica, com detalhes sob demanda, evita que todas as informações sejam apresentadas ao mesmo tempo, obtendo uma observação adequada do conjunto - menos é mais neste caso (MILANO, 2009). Ao passo em que o detalhamento aumenta, o mesmo autor ainda cita que outras técnicas de visualização são necessárias para especializar cada tipo de análise (SHULL, 2013). Isto foi atendido através dos menus *Analisar sistema* e *Analisar consumidor*, cada qual com visualizações complementares à espacial-geográfica, que serão tratadas em seções específicas a seguir.

As Figuras 3.4 e 3.5 exibem a primeira tela ao acessar a aplicação e o nível de

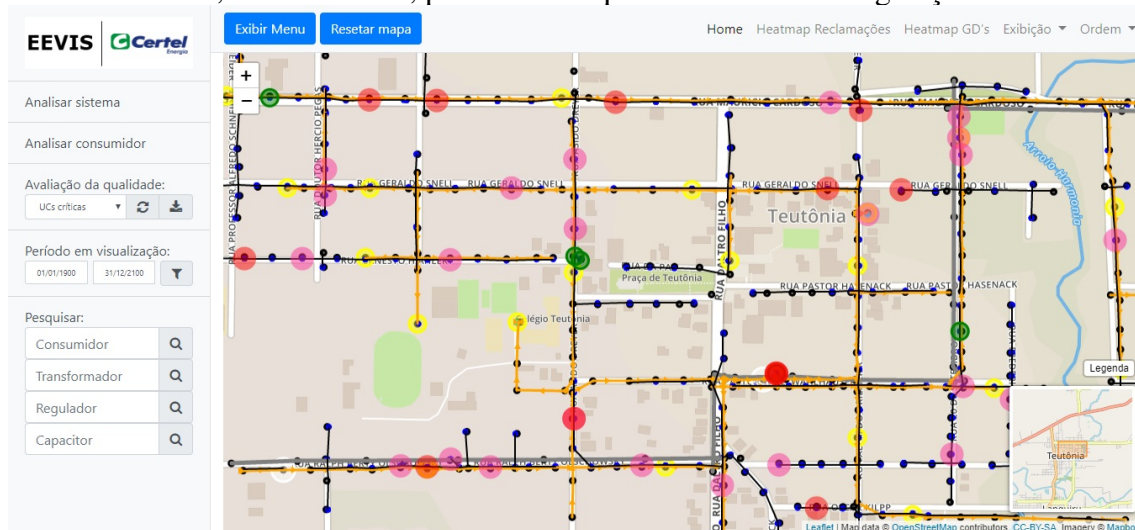
detalhe aumentado pelo *zoom*, com mais informações apresentadas.

Figura 3.4: Tela de entrada da plataforma ELVIS: visão da área geográfica de permissão da distribuidora (poligonais) e as reclamações (pontos vermelhos)



Fonte: O autor

Figura 3.5: Visualização aproximada/detalhada do mapa: redes de média e baixa tensão, transformadores, consumidores, processos de qualidade e usinas de geração distribuída



Fonte: O autor

O sistema de mapas utilizado foi o OpenStreetMap<sup>1</sup>, de dados abertos e sem custo de utilização. Como apoio ao mapa, foi incluído:

- Campos de pesquisa para consumidores, transformadores, reguladores de tensão e capacitores, que levam a visão do mapa até a rede onde o item buscado situa-se;
- Um mini-mapa, no canto inferior direito, com uma visualização mais distante para facilitar a localização;

<sup>1</sup>OpenStreetMap: <https://www.openstreetmap.org/>

- Logo acima, uma legenda que é exibida ao passar o mouse, explicando o que cada cor representa no mapa;
- No canto superior direito, uma opção para escolher a visibilidade (visível ou não) das dimensões disponíveis;
- Ao lado, uma opção para escolher a ordem de visualização das dimensões, para os casos de sobreposição; e
- Um *spinner*, visível enquanto o mapa está em processamento (a interação sobre o mapa continua possível).

Desta maneira, até aqui foram atendidos os requisitos R1 e R2 (descritos na página 37).

#### 3.4.1.1 Simbologia

A eficiência com que as atividades serão realizadas são resultantes da qualidade dos mapas, das visualizações e da interação necessária (ROBBI, 2000). Em função disto, a abordagem procurou utilizar uma simbologia intuitiva de glifos, cores e tamanhos.

A simbologia padrão dos equipamentos elétricos não foi tomada rigorosamente como premissa, pois, assim como pode tornar confortável a utilização do sistema por alguns usuários em um primeiro momento, pode ser mais adequada a utilização de símbolos que são mais intuitivos para outro grupo de usuários, como estudantes ou principalmente usuários não-especialistas (KLUMP; DOOLEY; WU, 2003). Desta forma, atende-se ambos os públicos, sendo utilizável por um número maior de usuários.

A simbologia utilizada no mapa baseou-se em uma biblioteca JavaScript conhecida como Leaflet<sup>2</sup>.

Os cabos foram representados por segmentos de reta. Diferentes cores simbolizam baixa, média ou alta tensão. O tipo de tracejado indica a quantidade de fases do vão de rede (monofásico, bifásico ou trifásico). O fluxo de energia é representado por setas inseridas sobre os segmentos de reta. As demais informações são simbolizadas por círculos, sendo que as diferentes entidades são simbolizadas por diferentes cores. Quando considerado necessário, como será visto adiante, o tamanho dos círculos representa a magnitude da informação. Ainda, ao clicar sobre qualquer componente, informações adicionais são apresentadas no formato de *tooltip*.

---

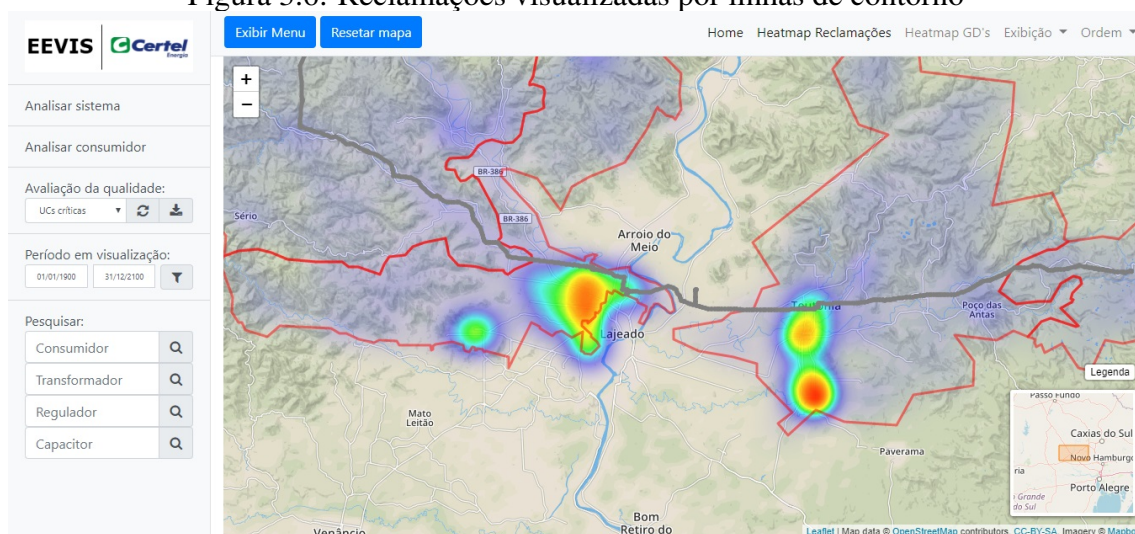
<sup>2</sup>Leaflet: <https://leafletjs.com/>



### 3.4.1.2 Densidade de reclamações e densidade de geração distribuída

Para atender ao quarto e quinto itens do R6 (descritos na página 37), foi inserido sobre o mapa uma visualização de contornos, para apresentar informações a respeito da densidade geográfica de reclamações dos consumidores e também da densidade das unidades de geração distribuída. A Figura 3.6 exemplifica esta visualização, e ambas são importantes para deixar notáveis as regiões com maior incidência destes atributos.

Figura 3.6: Reclamações visualizadas por linhas de contorno



Fonte: O autor

Nessa figura fica evidente que a maior quantidade de reclamações está localizada nos centros urbanos. Esta visualização é dinâmica, e isto significa que, ao realizar operações de *zooming*, os contornos são atualizados. Portanto, ao aproximar o mapa, é possível identificar as regiões mais críticas, seja na área rural ou urbana.

### 3.4.2 Análise do sistema

Dados os requisitos identificados com os potenciais usuários, percebeu-se que estes poderiam ser classificados como informações referentes ao sistema elétrico (globais) ou à qualidade da energia para consumidores específicos. Isto levou a crer que o ideal seria também separar estas duas seções na plataforma. Portanto, foram criados módulos separados para cada um. Os gráficos desenvolvidos utilizaram as bibliotecas JavaScript Chart.js<sup>3</sup>, Plotly<sup>4</sup> e D3.js<sup>5</sup>.

<sup>3</sup>Chart.js: <https://www.chartjs.org/>

<sup>4</sup><https://plotly.com/javascript/>

<sup>5</sup><https://d3js.org/>

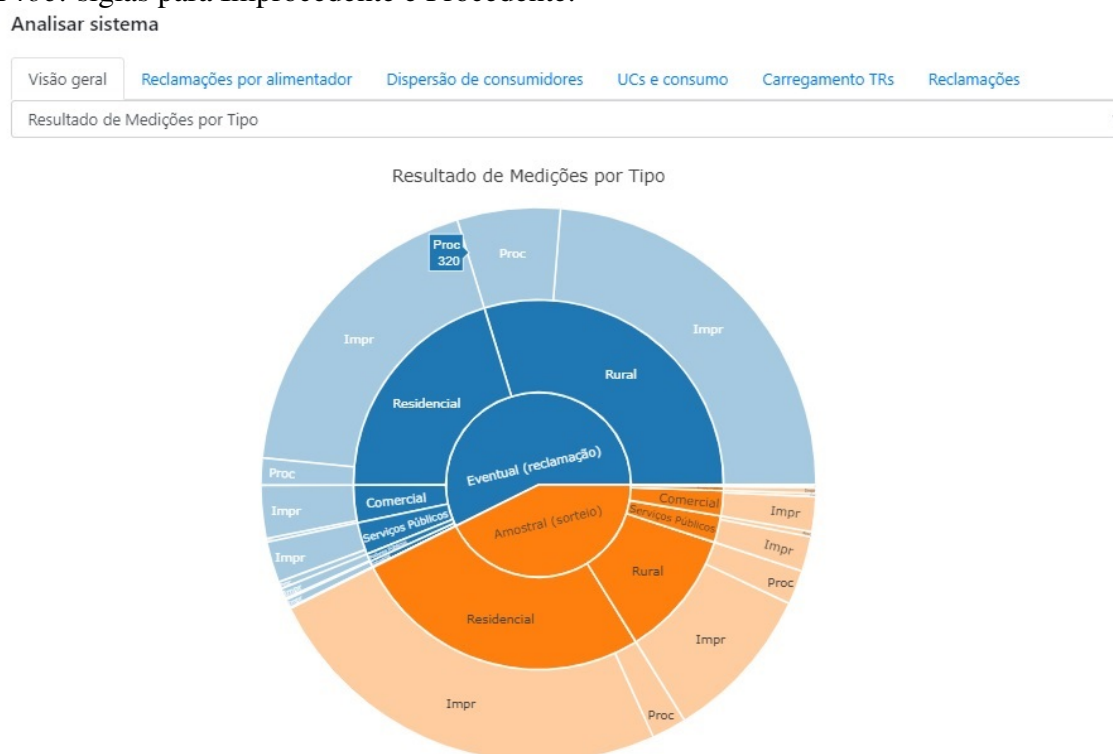
A seção *Análise do sistema* foi desenvolvida como uma caixa de diálogo (janela modal), subdividida em abas, e incorpora as seguintes visualizações:

### 3.4.2.1 Visão geral

Esta primeira parte da visualização do sistema elétrico traz uma visão geral a respeito das reclamações e seus resultados e foi dividida em duas partes: 1) avaliar o resultado das medições de tensão por classe do consumidor e, na sequência, por resultado; e 2) avaliar as reclamações por consumo dos clientes e por anormalidade da reclamação.

As Figuras 3.7 e 3.8 apresentam estas visualizações. Na primeira, foi utilizada a técnica *Sunburst* e, na segunda, a técnica *Heatmap*. Foi criada uma opção de seleção através de menu, pelo qual o usuário escolhe qual deseja avaliar - evitando o excesso de informações em uma única tela.

Figura 3.7: Visão geral do sistema elétrico: visualização em gráfico *sunburst*. *Impr* e *Proc*: siglas para Improcedente e Procedente.

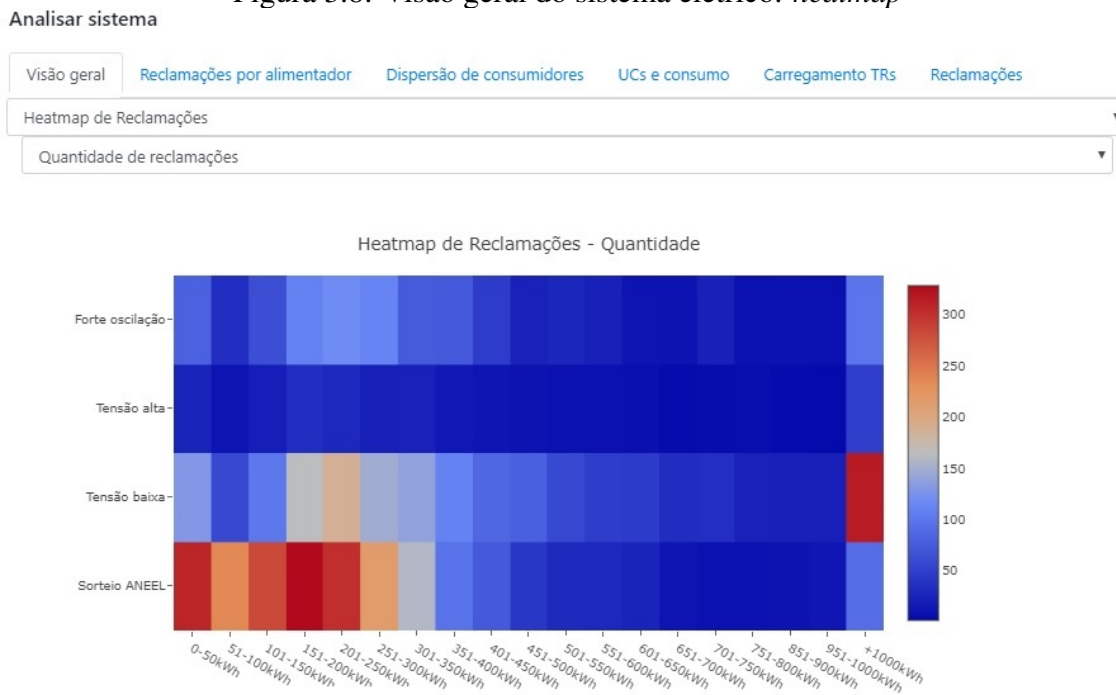


Fonte: O autor

É possível interagir sobre a visualização de *Sunburst*, selecionando a hierarquia que seja visualizar com mais detalhes. As dimensões selecionadas aqui levaram em consideração aquelas que os usuários determinaram como importantes, além daquelas que ainda não haviam sido solicitadas, associando-as para permitir um novo tipo de análise aos operadores.

Para a visualização de *Heatmap* foram criadas três categorias para representar a cor da célula: a quantidade de reclamações, a média do DRC e a média do DRP, também selecionáveis por menu de opções.

Figura 3.8: Visão geral do sistema elétrico: *heatmap*



Fonte: O autor

Estas representações não foram um requisito dos usuários. Foi criada com a intenção de utilizar técnicas de visualizações menos comuns, além da potencialidade que estas podem proporcionar às tarefas. Fica notável, por exemplo, que a quantidade de reclamações da classe rural é a mais elevada, e que a maior parte das reclamações é avaliada como improcedente.

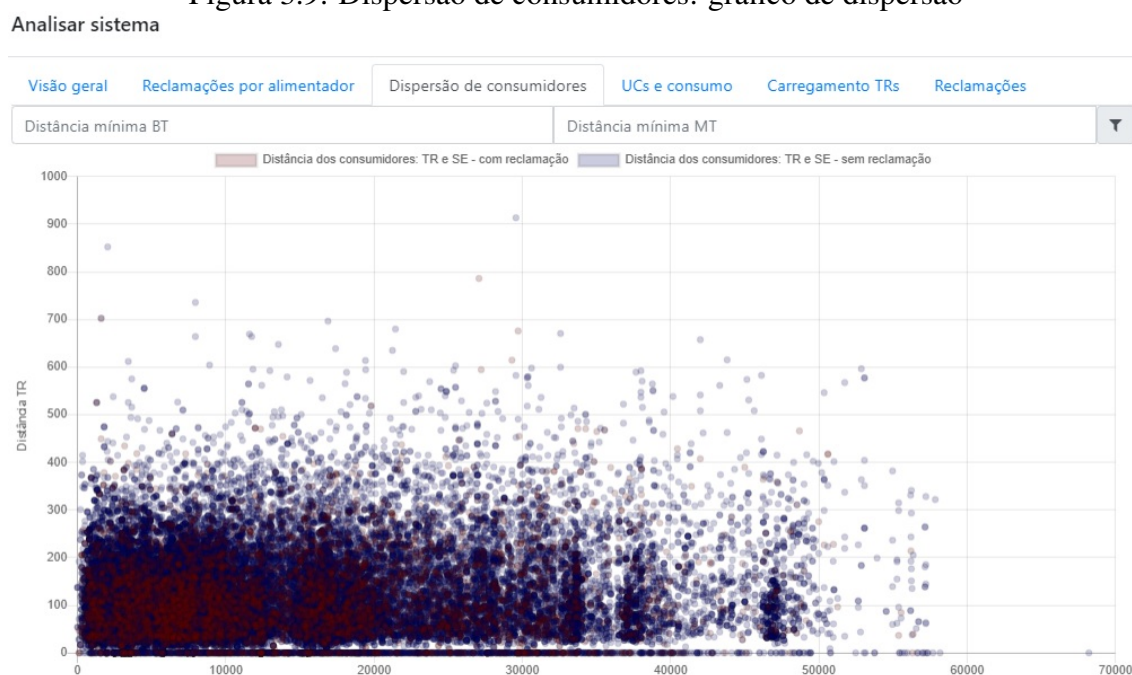
#### 3.4.2.2 Dispersão de consumidores

O objetivo desta visualização é apresentar as distâncias que todos os consumidores estão de duas origens: da subestação (rede de média tensão) e do transformador (rede de baixa tensão), procurando identificar aqueles mais distantes. Como são duas dimensões, isto levou a escolher o tipo de visualização de *Gráfico de dispersão*. Cada consumidor é identificado por um ponto, e as cores diferenciam aqueles que já reclamaram da qualidade de energia daqueles que não reclamaram.

Quanto mais distante dos eixos  $x$  e  $y$ , mais distante da subestação ou do transfor-

mador o consumidor está. Esta forma de apresentar os dados facilmente permite identificar pontos isolados, ou seja, consumidores que podem ter problemas com sua energia devido a estarem distantes da origem, principalmente da rede de baixa tensão (BT). Quanto mais distante, maior a queda de tensão. A Figura 3.9 apresenta esta visualização, que atende o segundo item do requisito R6. As distâncias são informadas em metros e é possível filtrar pela distância mínima para ambos eixos, sendo o gráfico redimensionado neste caso. Ao clicar sobre um ponto específico, é exibida a distância de cada origem, além da identificação da unidade consumidora.

Figura 3.9: Dispersão de consumidores: gráfico de dispersão



Fonte: O autor

### 3.4.2.3 Unidades consumidoras e consumo de energia

Nesta opção, existem três visualizações. A principal, apresenta o consumo de energia dos últimos 12 meses, separado nas cinco classes de consumidores: residencial, industrial, comercial, rural e poderes públicos - cada qual com sua cor. O tipo de gráfico escolhido foi o *gráfico radar*, aproveitando a sua continuidade entre os 12 meses do ano e a comparabilidade entre todas as classes.

Nas visualizações inferiores, são apresentados: o número total de Unidades Consumidoras (UCs), ao lado esquerdo, e a proporção de consumo de cada uma das classes em relação ao total, ao lado direito. A visualização utilizada aqui foi o gráfico de *coordenadas polares*, em função de que este contrasta a magnitude dos valores entre as

classes. Neste caso, poderia ser utilizado diversos outros tipos que também teriam esta característica, mas foi escolhido especificamente para dar diversidade à plataforma (não foi utilizado em outro local), para instruir os operadores também desta forma de visualização. Eventualmente, estes mesmos podem sugerir visualizações de outros dados baseadas naquelas que já estão utilizando. A Figura 3.10 exibe a caixa de diálogo com estes gráficos.

Figura 3.10: UCs e consumo: radar e coordenadas polares



Fonte: O autor

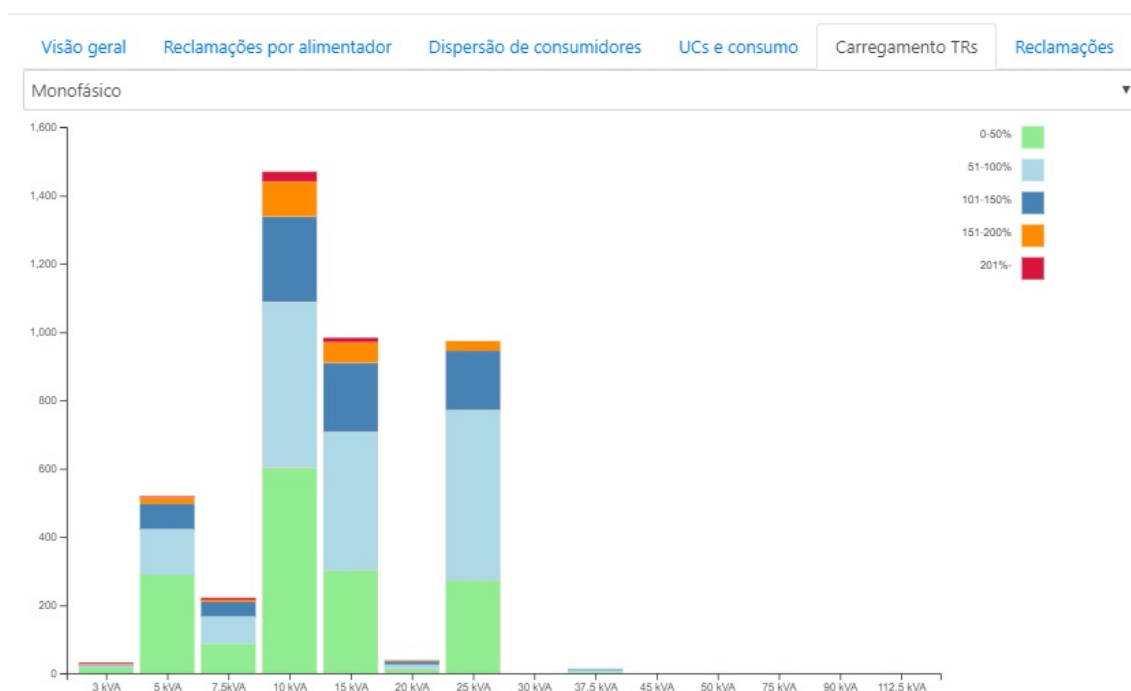
É possível interagir com cada um deles: clicar sobre os itens da legenda alterna a sua visibilidade, e o gráfico é atualizado mantendo apenas aqueles selecionados. Isto permite a adequada comparação quando uma das classes possui valores significativamente maiores, pois ao deixá-lo invisível, a visualização é aproximada para os demais.

#### 3.4.2.4 Carregamento dos transformadores

É importante avaliar como está o carregamento geral dos transformadores, principalmente se existem muitos sobrecarregados. Neste contexto, foi optado em utilizar o

*Gráfico de barras empilhadas.* A Figura 3.11 ilustra esta visualização, onde o eixo vertical representa a quantidade de transformadores e o eixo horizontal a potência. A legenda, à direita, exibe o significado da cor de cada pilha. Ao clicar em um item da legenda, o mesmo é retirado e o gráfico é atualizado. A visualização também é separada pelo tipo do transformador: mono, bi ou trifásico, cuja seleção é feita com o menu que fica sobre o gráfico. Ao passar o mouse sobre alguma barra, é explicitada a quantidade de transformadores nesta categoria.

Figura 3.11: Carregamento dos transformadores: gráfico de barras empilhadas  
Analisar sistema



Fonte: O autor

Esta análise permite rapidamente identificar o carregamento geral entre todos os transformadores da distribuidora. Isto inclui, além daqueles sobrecarregados (com provável necessidade de troca por outro de maior potência, para evitar a queima e garantir a qualidade da energia), os transformadores sobredimensionados, cujo carregamento está abaixo de 50% e podem eventualmente ser substituídos por outro de menor potência - gerando economia de recursos financeiros.

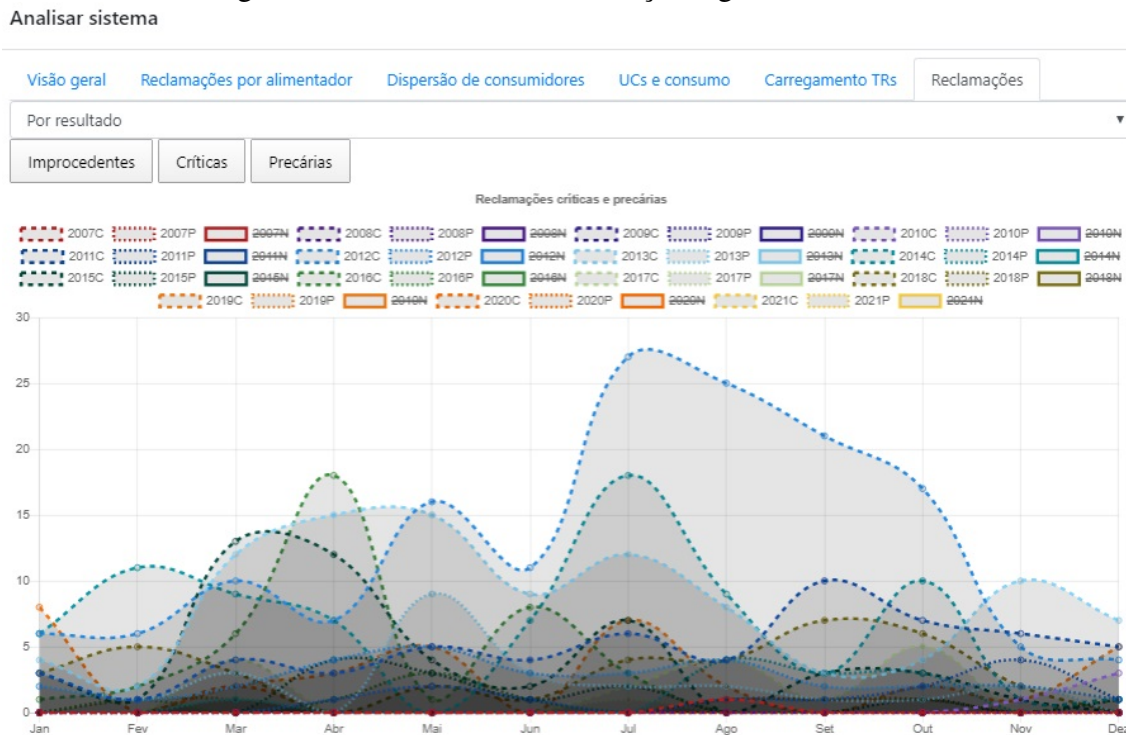
#### 3.4.2.5 Histórico de reclamações

O menu para análise do histórico de reclamações inclui dois gráficos. O primeiro exibe a evolução da quantidade total de reclamações, enquanto o segundo segrega a quantidade total em quantidade de reclamações improcedentes (energia adequada), reclama-

ções críticas e reclamações precárias.

O histórico contempla três dimensões: quantidade mensal das reclamações, mês e ano da reclamação, e tipo (improcedente, crítica ou precária). A visualização escolhida foi *Gráfico de linhas*, utilizando diferentes cores e serrilhados.

Figura 3.12: Histórico de reclamações: gráfico de linhas



Fonte: O autor

Ao passar o mouse sobre os pontos, é exibida a quantidade exata das reclamações daquele mês e ano, para aquela classificação. É possível também interagir de forma a escolher quais os anos ficam visíveis, flexibilizando as comparações possíveis. Foi criado um chaveamento para escolher qual visualização o usuário deseja: reclamações totais ou por resultado (improcedentes, críticas e precárias). Nesta última, o usuário pode optar em desmarcar o tipo da reclamação sem precisar desmarcar cada ano. Isto permitirá agilidade na interação e também comparar anos diferentes não-contínuos.

Estas visualizações do ELVIS atendem o primeiro item do requisito R6 (descritos na página 37). A Figura 3.12 exibe a visualização 'Por resultado', exibindo o histórico das reclamações Críticas e Precárias.

Havia sido estudado utilizar o *Gráfico radar com mapa de calor*: os eixos radiais representariam o ano, os eixos verticais os meses e as cores a quantidade de reclamações. Porém, ainda faltaria uma dimensão (improcedente, crítica ou precária). Para melhor comparação entre os meses, ambos gráficos utilizaram linhas.

### 3.4.2.6 Reclamações por alimentador

Aqui é apresentada a quantidade total de reclamações para cada alimentador. Inicialmente, havia sido utilizado o gráfico de setores (popularmente, pizza), porém não é o mais adequado para a maioria das visualizações - principalmente quando houver fatias muito pequenas ou de tamanho muito similar. Sendo assim, constatou-se que uma técnica de visualização adequada para esta necessidade é a *treemap*. A Figura 3.13 ilustra o resultado. Ao passar o mouse sobre cada bloco, é identificado o alimentador com a respectiva quantidade de reclamações.

Figura 3.13: Reclamações por alimentador: *treemap*



Fonte: O autor

### 3.4.3 Análise de consumidores

Este módulo é voltado exclusivamente para a análise visual das características do sistema elétrico peculiares de cada consumidor. Foi desenvolvido de maneira idêntica ao módulo anterior, sendo exibido uma caixa de diálogo do tipo janela modal. Este módulo atende o requisito R3 (descritos na página 37).

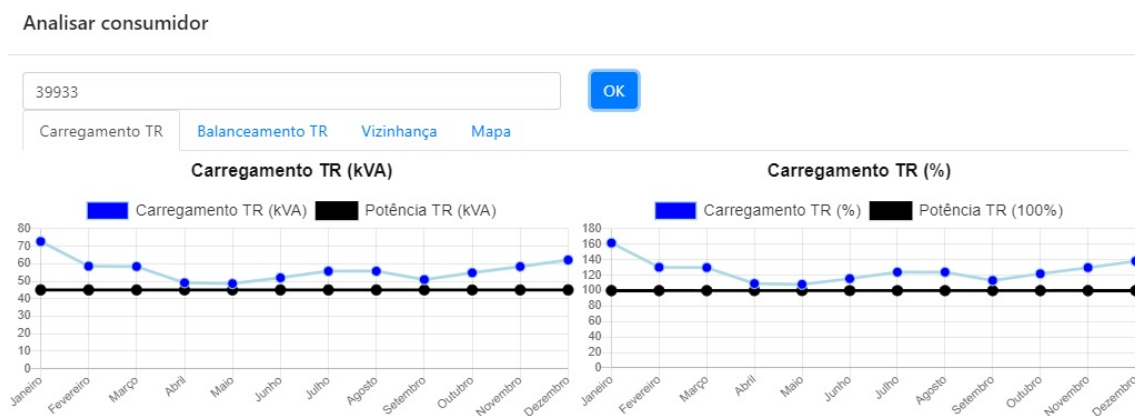
Ao escolher esta opção, o usuário precisa digitar a unidade consumidora que deseja analisar. Feito isto, o sistema prepara quatro tipos de visualização, detalhadas a seguir.



### 3.4.3.1 Carregamento do transformador

Esta visualização apresenta o carregamento do transformador que alimenta o consumidor em análise. A Figura 3.14 apresenta esta visualização, que atende o requisito R5 e o terceiro item do R6 (descritos na página 37). O tipo de gráfico utilizado foi o *gráfico de linhas*, que facilmente permite identificar a variação do carregamento ao longo dos últimos 12 meses.

Figura 3.14: Carregamento do transformador: gráfico de linhas



Fonte: O autor

O carregamento é calculado de acordo com o consumo de todos consumidores ligados ao transformador, em relação a sua potência. A cada ligação nova de consumidor conectada ao transformador, é realizado um cálculo do carregamento. Mesmo assim, o aumento de carga dos consumidores é muito comum e natural. Em função disto, eventualmente os transformadores ficam sobrecarregados e podem causar problemas de qualidade de energia, além de poderem queimar.

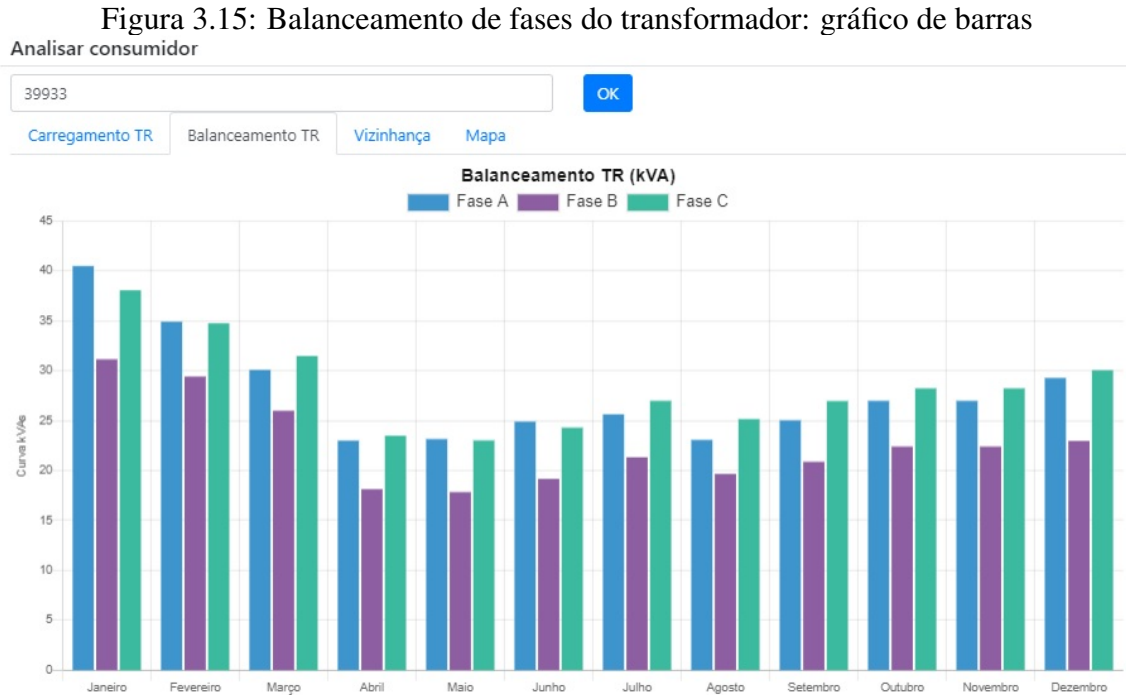
O gráfico da esquerda mostra o carregamento em kVA, enquanto o da direita apresenta em pontos percentuais, ambos de linha azul. A linha preta exibe a potência do transformador (à esquerda em kVA e à direita, evidentemente, em 100%).

### 3.4.3.2 Balanceamento de fases do transformador

Quando os transformadores são bifásicos ou trifásicos, pode haver desbalanceamento de suas fases. Isto ocorre porque os consumidores podem estar ligados à diferentes fases, e eventualmente alguns com carga elevada podem estar ligados em uma mesma fase. Isto também pode causar problemas na energia, sendo que uma das soluções é distribuir a ligação dos consumidores nas fases mais adequadas.

A Figura 3.15 apresenta esta visualização, que atende o requisito R5 (descritos na

página 37) e utilizou o *gráfico de barras*. O eixo vertical representa a curva kVAs para cada fase, cujo conceito é explicado na seção seguinte. Foi escolhido este gráfico porque as três fases são muito bem representadas pelas barras, inclusive nos 12 meses, permitindo a adequada comparação temporal e entre as fases.



Fonte: O autor

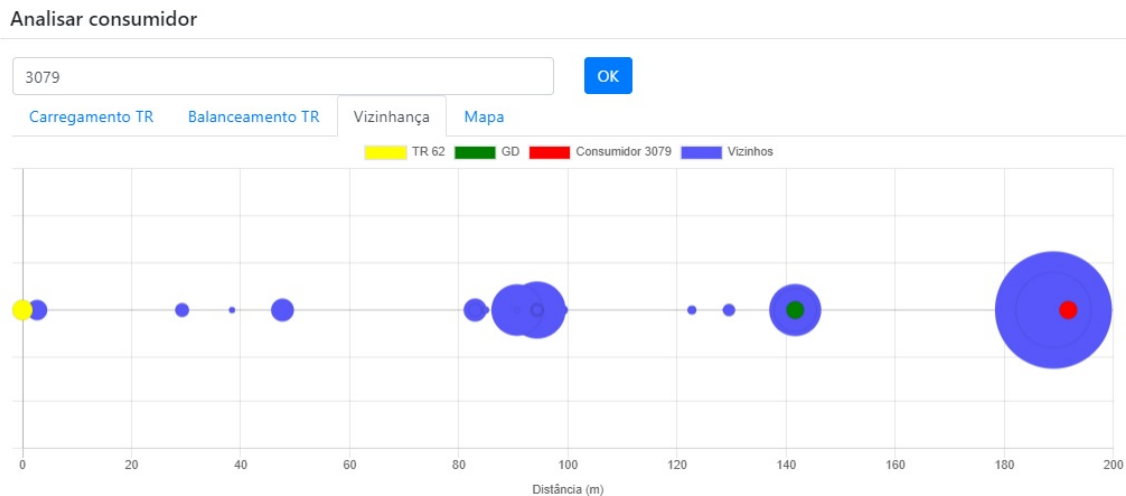
Caso necessário, é possível alternar a visibilidade de cada uma das fases, clicando sobre a legenda correspondente.

### 3.4.3.3 Vizinhança

Aqui são representadas todas as unidades consumidoras e geradoras que ficam na mesma rede de baixa tensão do consumidor em análise, deixando evidente aqueles que ficam mais próximos ao transformador. Foi escolhido o *gráfico de bolhas* para apresentar estas informações: todos os elementos são círculos, e a sua magnitude é representada pelo seu diâmetro. A Figura 3.16 ilustra esta visualização, que atende ao requisito R4 (descritos na página 37) em conjunto com a visualização da próxima subseção.

Junto ao eixo vertical, é exibido o transformador, em amarelo. A partir dele, todos os demais elementos são distribuídos sobre o eixo horizontal, de acordo com a sua distância do transformador. Cada cor ilustra um tipo de dado: em amarelo: transformador, em vermelho: consumidor em análise, em azul: consumidores vizinhos, e em verde: unidades geradoras. O tamanho do círculo é representado pela curva kVAs dos consumidores

Figura 3.16: Vizinhança: gráfico de bolhas



Fonte: O autor

ou potência da usina geradora.

Esta visualização permite agilmente identificar se existem consumidores na rede de BT com carga instalada elevada ou ainda unidades geradoras de energia, que podem causar distúrbios na rede.

#### 3.4.3.4 Análise geográfica da rede, equipamentos e vizinhos

A última visualização apresentada leva a análise dos dados dos consumidores novamente para o mapa. Após visualizar os gráficos anteriores, o usuário pode clicar sobre uma opção que fecha a caixa de diálogo e reapresenta a visualização geográfica. Esta visualização atende o último item do requisito R6 (descritos na página 37), atendendo-o de forma completa.

Nesta visualização, todas as informações que estavam visíveis inicialmente são eliminadas (com exceção do mapa), enquanto a visualização é aproximada para a região onde o consumidor está localizado. Ao destacar a região, são exibidos os seguintes dados, especificamente relacionados ao consumidor em análise:

- O consumidor em análise, representado por um círculo vermelho opaco;
- Os vizinhos de baixa tensão (grupo B), em azul;
- Os vizinhos de média tensão (grupo A), em azul claro;
- As unidades geradoras de energia, em verde;
- O transformador que alimenta o consumidor em análise e seus vizinhos, em amarelo;

- A rede de BT específica do consumidor, em preto;
- Todos os caminhos da rede de MT, em laranja, cuja distância a partir do consumidor atinja até 2km;
- A queda de tensão para cada consumidor da rede de BT, simbolizada pelo tamanho da borda de seu círculo;
- A média queda de tensão das fases na rede de MT, em todos os fins de rede (o valor exato de cada fase pode ser visualizando clicando sobre o vão de rede);
- Os reguladores de tensão e os capacitores que ficam dentro dos 2km supracitados;
- As reclamações que foram registradas até então nos trechos de rede apresentados.

Importante salientar que o diâmetro dos elementos que simbolizam os consumidores é representado pela sua demanda estimada de energia. Esta, foi calculada de acordo com o Critérios de Elaboração de Projetos, da Federação das Cooperativas de Energia, Telefonia e Desenvolvimento Rural do Rio Grande do Sul (FEDERAÇÃO DAS COOPERATIVAS DE ENERGIA, TELEFONIA E DESENVOLVIMENTO RURAL DO RIO GRANDE DO SUL, 2014).

A curva kVAs é uma das formas utilizadas para transformar o consumo (kWh), em um valor estimado de demanda (kVA), para realizar os cálculos de estimativa do carregamento do sistema elétrico. Foi optado em utilizar este valor pois: a) a carga instalada não é necessariamente utilizada a todo instante; b) o consumo do mês não representa apropriadamente a demanda que realmente é utilizada. Portanto, se cada um destes valores fosse utilizados isoladamente, teria grande probabilidade de não estar tão próximo à realidade quanto a curva kVAs pode representar. Está clara a importância de utilizar a demanda neste caso.

A distribuidora deste estudo utiliza este método, contudo é importante salientar ao leitor que existem outras técnicas para estimar seu valor, além da curva kVAs, que podem ser utilizados em seu lugar.

A distância de 2km foi definida com base em análise do engenheiro eletricista responsável. Esta distância foi considerada adequada para representar as possíveis ocorrências que podem causar distúrbios ao consumidor, concomitantemente à menor perda de performance no processamento - visto ao passo que a distância aumenta, mais tempo de processamento será necessário. A complexidade computacional de identificar todos os possíveis caminhos, comparando-os a todos os trechos existentes na rede, é extremamente elevada, sendo proporcionalmente maior conforme a distância aumenta (foi necessário

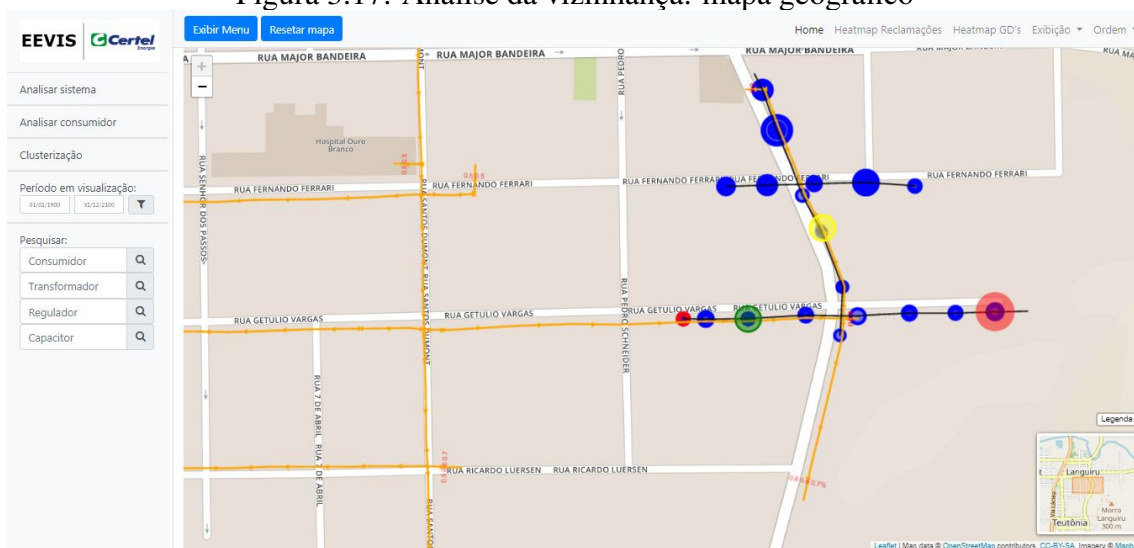
utilizar algoritmo recursivo para tal).

O algoritmo recursivo desenvolvido para percorrer as redes baseou-se na entidade *Cabos*, onde cada linha do dataset representa um trecho de rede (ou seja, cada vão entre dois postes). Este vão, além de outros atributos, possui identificações únicas de início e fim, que são os mesmos valores que um consumidor ou os componentes da rede elétrica possuem caso estejam conectado neste ponto específico (através do qual permite-se vinculá-los). Cada vão também possui a identificação do seu comprimento, pelo qual a soma acumulada da recursão foi utilizada com o limite de 2km. Desta forma é possível atravessar a rede elétrica ao entorno de um consumidor: quando chega-se ao limite da distância definida, volta-se uma etapa da recursão e procura-se um novo possível caminho, até não serem mais encontrados e volta-se ao ponto de partida.

Nesta nova visualização de mapa, o tipo de tracejado da linha das redes continua representando a quantidade de fases. Todos os elementos, ao serem clicados, apresentam informações adicionais (como identificação e informações técnicas) na forma de *tooltip*.

A Figura 3.17 apresenta uma região da distribuidora que possui a maioria destas informações. Esta abordagem atende o requisito R3, R4 e também o sexto item do R6 (descritos na página 37). De forma complementar, enquanto os dados são processados para gerar as visualizações, é gerado um relatório que contém avisos e alertas para os operadores, aliando as informações visualizadas com informações numéricas e técnicas (como sua distância do consumidor e a potência dos equipamentos).

Figura 3.17: Análise da vizinhança: mapa geográfico




Fonte: O autor

Neste relatório, são listados todos os equipamentos, consumidores, unidades geradoras e reclamações. Os consumidores que aparecem no relatório são apenas aqueles que

podem significar alguma influência considerável na qualidade de energia da vizinhança. O critério definido foi que a demanda, calculada de acordo com a curva kVAs, seja maior do que 50kW (demandas altas podem significar distúrbios na rede). A Figura 3.18 ilustra o exemplo de um relatório gerado, que pode direcionar o usuário a uma das visualizações apresentadas até aqui.

Figura 3.18: Relatório de alertas da qualidade de energia do consumidor

	<p><b>CARACTERÍSTICAS ELÉTRICAS DA PROXIMIDADE:</b></p> <p>ALERTA: Queda de tensão na BT até o consumidor: 5.16%</p> <p>ALERTA: O vizinho 3034 é ABCN, possui uma curva de kVAs de 65.72 e é da classe/subclasse COMERCIAL/COMERCIAL</p> <p>ALERTA: Houve reclamação 2771714 da UC 58362 no dia 27/6/2014</p> <p>ALERTA: Existe uma GD na UC 31813 com potencia de 3</p> <p>AVISO: O consumidor está a uma distância de 191.74m do transformador nº 62</p> <p>AVISO: O consumidor está a uma distância de 3445.56m do alimentador 12</p> <p>AVISO: O transformador tem as fases ABC e potência de 112.5kVA</p> <p>ALERTA: O transformador está com carregamento de 177.93 kVA de uma potência de 112.5 kVA (158.16%)</p> <p>ALERTA: Queda de tensão da MT do até no transformador: A:0.57%   B:0.55%   C:0.67%</p>
---	--

Fonte: O autor

### 3.4.4 Discussão

Retomando os requisitos levantados, baseado no que foi apresentado até aqui, pode-se afirmar que todos aqueles relacionados à visualização foram atendidos. A usabilidade e o modelo de avaliação da qualidade serão tratados nos capítulos seguintes. Foram elaborados diversos tipos de visualizações, criando uma abordagem de analítica visual totalmente personalizada para a análise da qualidade da energia. Com o passar do tempo, acredita-se que gradativamente poderão ser levadas estas técnicas de apresentação de dados para esta área.

A variedade de técnicas utilizadas não foi considerada um requisito, porém naturalmente os tipos de dados levaram a plataforma a adquirir esta característica. Para cada requisito foi analisada a técnica de visualização que melhor atendia a finalidade. Algumas técnicas poderiam ter sido reutilizadas em alguns casos. Entretanto, acredita-se que o que foi desenvolvido é bastante adequado em cada caso, tendo sido validado na avaliação de usabilidade realizada.

Para tornar ágil a interação do usuário, foi utilizado um mecanismo conhecido como *Web Worker*. Este permitiu que funções que exigem muito processamento sejam executadas fora do fluxo de execução principal (*main thread*), conseqüentemente evi-

tando que a interface seja bloqueada. À medida em que os dados são processados, os mesmos são apresentados na tela. No que se refere à qualidade da energia, a abordagem permitiu que principais análises realizadas na rotina dos operadores sejam amparadas pela plataforma.

É interessante que o leitor acesse os vídeos demonstrativos da abordagem visual, disponíveis nos endereços abaixo:

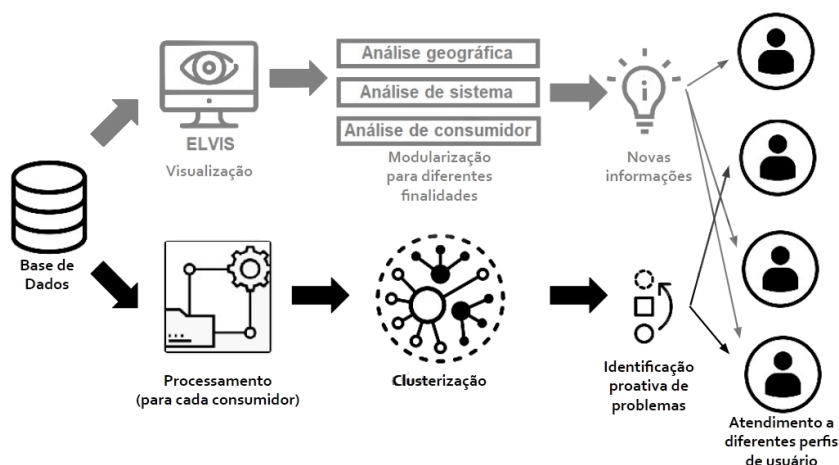
- Análise geográfica: <<https://vimeo.com/445221067>>;
- Análise do sistema: <<https://vimeo.com/445221068>>;
- Análise do consumidor: <<https://vimeo.com/445221069>>.

#### 4 MODELO DE AVALIAÇÃO DA QUALIDADE DE ENERGIA

Este capítulo descreve como foi desenvolvido o modelo de avaliação da qualidade da energia para os consumidores, através de aprendizado de máquina não supervisionado. Como descrito em capítulos anteriores, o objetivo é agrupar os consumidores em conjuntos, por similaridade em relação à qualidade da energia. Em seguida, serão analisados os clusters criados, averiguando a existência de clusters que representem aqueles consumidores com possível má-qualidade de energia. Com isto, será possível, por exemplo, avaliar de fato a existência de problemas de energia e incluir a região destes consumidores em planejamentos de obras, para corrigir tais distúrbios.

A Figura 4.1 (parte inferior) ilustra o esquema adotado para o modelo de avaliação da qualidade da energia. Para a mesma base de dados da visualização de informações, foram processados os dados do sistema elétrico para cada consumidor (na prática, foi automatizada a opção de *Análise do consumidor* da parte de visualização). Ao final, foi obtida uma base de dados personalizada, com informações inerentes à conexão com a rede elétrica de cada consumidor. Isto significa que foram utilizados dados individuais na clusterização, no intuito de identificar problemas de energia de forma proativa.

Figura 4.1: Representação esquemática da abordagem para avaliação da qualidade de energia



Fonte: O autor

Neste capítulo, a primeira seção trata das dimensões selecionadas e a elaboração da base de dados. A seção seguinte descreve a escolha dos algoritmos de redução de dimensionalidade e clusterização. A última seção analisa os clusters gerados.



#### 4.1 Seleção dos dados de qualidade da energia

A clusterização é uma forma de aprendizado de máquina não-supervisionada, que permite a identificação automática de padrões e similaridades entre uma série de informações - no caso desta pesquisa, de consumidores de energia (PANAPAKIDIS; MOSCHAKIS, 2019).

Para agrupar os consumidores em clusters, primeiramente é necessário selecionar e calcular os dados que caracterizam a sua qualidade de energia. O primeiro passo, é estabelecer quais são as dimensões relevantes que serão utilizadas para o processamento (PANAPAKIDIS; MOSCHAKIS, 2019). Desta forma, os próprios dados descritos no capítulo anterior foram processados individualmente e formaram uma base de dados personalizada, elencando as características elétricas específicas de cada consumidor.

Na prática, os algoritmos elaborados para analisar os consumidores (ver seção 3.4.3) foi processado para cada um destes, de forma adaptada para resultar em um arquivo CSV. O tempo de processamento de mais de 65.000 consumidores de BT foi de aproximadamente 30 horas<sup>1</sup>. A maior parte deste tempo é consequência da complexidade de percorrer as redes de BT e MT - onde foi utilizada recursividade (conforme descrito no capítulo anterior), não sendo um acesso trivial. É um processamento significativo, contudo é o que permite identificar quais detalhes caracterizam a rede elétrica de cada consumidor.

As dimensões que compõe a base de dados são, em geral, universais para as distribuidoras de energia e são anaixo listadas:

- UC: Identificação da Unidade Consumidora;
- Balanceamento de fases do transformador (*balanc*);
- Carregamento do transformador (*carreg*);
- Queda de tensão na rede de BT até o consumidor (*queda\_bt*);
- Queda de tensão na rede de MT que conecta ao transformador (*queda\_mt*);
- Distância do transformador (*dist\_bt*);
- Distância do alimentador (*dist\_mt*);
- Distância do capacitor (*dist\_cap*);
- Distância do regulador (*dist\_reg*);
- Soma total das curvas kVAs, das unidades consumidoras da rede de baixa tensão do

---

<sup>1</sup>Especificações do computador: processador Intel Core i5, 8GB de memória RAM

consumidor (*kvas\_bt*);

- Soma total das curvas kVAs, das unidades consumidoras de alta tensão de até 2 km a partir do consumidor (*kvas\_mt*);
- Soma da potência das unidades de geração distribuída, existentes na rede de baixa tensão que atende o consumidor (*gd\_bt*);

O nome entre parênteses representa a abreviatura que será utilizada nas imagens da Seção 5.2.2 Avaliação qualitativa. Algumas destas dimensões são explícitas (extraídas diretamente do sistema técnico da distribuidora de energia, como a queda de tensão e o carregamento do transformador) e outras implícitas (derivadas do processamento dos dados brutos extraídos, como o balanceamento de fases e as distâncias e somas de itens específicos).

O balanceamento de fases foi calculado pela divisão da fase com maior curva kVAs pela fase com a menor. Ou seja, quanto mais próximo de 1, mais balanceada está a rede. Deste modo, em redes monofásicas o índice será sempre 1. Esse cálculo considera o equilíbrio de fases nas unidades consumidoras, porque o consumo de energia não é realizado por fase.

O carregamento do transformador e as quedas de tensão foram exportados em pontos percentuais, permitindo a comparação entre transformadores de diferentes potências. As distâncias foram expressas em metros e as quedas de tensão em pontos percentuais. Foi mantida a distância de 2 km, de acordo com o explicado na seção 3.4.3.4.

As reclamações propriamente ditas não foram consideradas para clusterização. Como não há treinamento do algoritmo, não são utilizadas informações prévias (PANAPAKIDIS; MOSCHAKIS, 2019). Em função disto, precisaram ser desconsideradas as reclamações na base de dados, pois, quando uma reclamação é procedente, a rede elétrica próxima ao reclamante eventualmente é modificada. Consequentemente, as características originais da rede estão diferentes, e a base de dados ficaria incoerente: as reclamações indicariam má qualidade da energia, porém as características da rede elétrica já não representam mais isto.

Neste caso, uma solução seria construir uma base de dados considerando a rede elétrica no momento de cada reclamação (que não está disponível facilmente e exige tempo considerável para extração). Contudo, esta forma possibilitaria o desenvolvimento de técnicas de aprendizado de máquina supervisionado - utilizando as reclamações.

## 4.2 Algoritmos escolhidos

Esta seção detalha as escolhas realizadas para os algoritmos de clusterização e também de redução de dimensionalidade. Optou-se em utilizar técnicas bastante conhecidas e difundidas na comunidade acadêmica, visto que esta pesquisa não focou em analisar as técnicas propriamente ditas, e utilizando algoritmos com abordagens de funcionamento diferentes, para aumentar a probabilidade de melhores resultados.

O algoritmo K-Means (DING; HE, 2004; LLOYD, 1982) é um dos mais conhecidos e aplicados para a clusterização de dados, sendo de uso geral e apresenta um bom desempenho. Baseia-se na distância euclidiana entre os pontos: quanto mais próximo um item ficar do outro, mais similares estes são – e vice-versa.

É muito comum utilizar algoritmos de redução de dimensionalidade juntamente com clusterização (ver (DING; HE, 2004), principalmente quando há alta dimensionalidade nos dados, para minimizar a distorção entre todas as dimensões e minimizar ruídos e *outliers* que prejudicam os resultados (VANDERPLAS, 2016).

Assim, optou-se em utilizar o algoritmo *Principal Components Analysis* (PCA), utilizado com K-Means muito frequentemente. O algoritmo identifica as componentes principais do conjunto de dados (cujo número é igual ou menor à quantidade de variáveis), considerando aquelas com maior variância e correlação. Após, as dimensões são reduzidas através de transformações ortogonais, procurando perder a menor quantidade de informação possível (e mantendo a maior similaridade com as características originais) (DING; HE, 2004; VANDERPLAS, 2016).

Além da ampla utilização, possui agilidade e versatilidade (no sentido de já ter comprovado que pode ser utilizado com dados de diversos tipos), permitindo conhecer os principais padrões antes mesmo de iniciar a clusterização (VANDERPLAS, 2016).

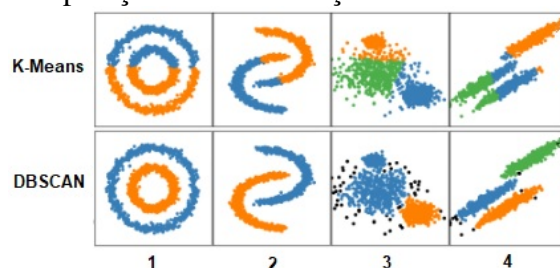
Algumas pesquisas avaliam especificamente estas técnicas aplicadas juntas. É o caso do trabalho de Ding e He (2004), que aplicaram PCA e K-means. Estes concluíram que a redução de dimensionalidade está muito relacionada ao resultado do aprendizado de máquina não-supervisionado. Detalhes de funcionamento dos algoritmos K-Means e PCA podem ser encontrados no trabalho de Chicco, Napoli e Piglione (2006).

Contudo, devido a existirem algoritmos com diferentes formas de operação, K-means pode não ser a melhor escolha. Por isso foi aplicado também o algoritmo DBSCAN (do inglês, *Density-Based Spatial Clustering of Applications with Noise*) (ESTER et al., 1996).

Diferente do K-Means, o algoritmo DBSCAN não tem a quantidade de clusters predefinida como parâmetro. A clusterização é realizada de acordo com a densidade dos dados, através da definição de raio máximo da vizinhança e quantidade mínima de itens nesta vizinhança. Como este consiste na densidade em si, tem bons resultados também em geometrias não convexas, diferente do K-Means que baseia-se na distância euclidiana (SCHUBERT et al., 2017).

Para exemplificar, veja a Figura 4.2. Fica evidente a influência da geometria dos dados para a formação dos clusters, sendo o DBSCAN indicado também à geometrias não-convexas.

Figura 4.2: Comparação da clusterização entre K-Means e DBSCAN



Fonte: Adaptado de (SCIKIT-LEARN DEVELOPERS, 2019)

### 4.3 Procedimento

A clusterização foi realizada na aplicação web Jupyter Notebook<sup>2</sup>, que permite o desenvolvimento e a análise de resultados de forma ágil e prática. Foi utilizado Python 3, com as bibliotecas *pandas*, *numpy*, *sklearn*, *matplotlib* e *plotly*, que são extensamente utilizadas e atendem de uma forma muito funcional os objetivos deste tipo de pesquisa.

As subseções seguintes explicam como foi realizada a clusterização, considerando normalização, redução de dimensionalidade e clusterização.

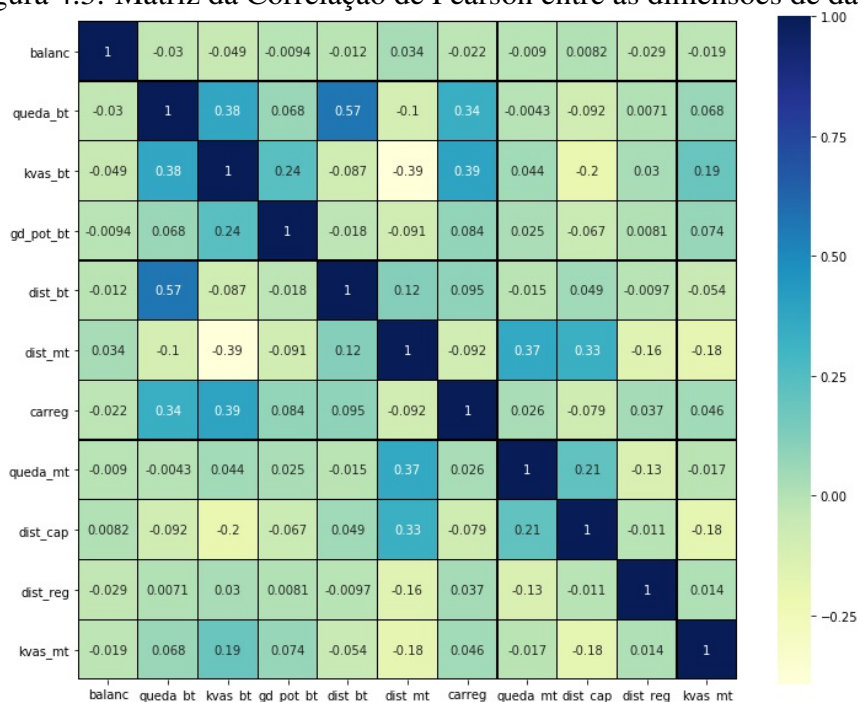
#### 4.3.1 Investigação e padronização

Antes de iniciar as transformações dos dados, foi realizada uma apuração de possíveis padrões pré-existent. Para isto, foi construída a matriz do Coeficiente de Correlação de Pearson, ilustrada pela Figura 4.3.

Naturalmente foram encontrados alguns padrões que representam características

<sup>2</sup>Jupyter.org: <https://jupyter.org/>

Figura 4.3: Matriz da Correlação de Pearson entre as dimensões de dados



Fonte: O autor

elétricas. As principais correlações encontradas foram relacionadas à queda de tensão com a distância, tanto na BT como na MT. Também foi identificado que há correlação da soma das curvas kVAs com a queda de tensão na BT e também com o carregamento do transformador.

A inexistência de muitas correlações nos dados, e a inexistência correlações significativas, já permite inferir que será necessário mais do que apenas dois ou três componentes principais do PCA para reduzir a dimensionalidade com menor perda de variância, devido à falta de dimensões que possam significativamente diferenciar os dados.

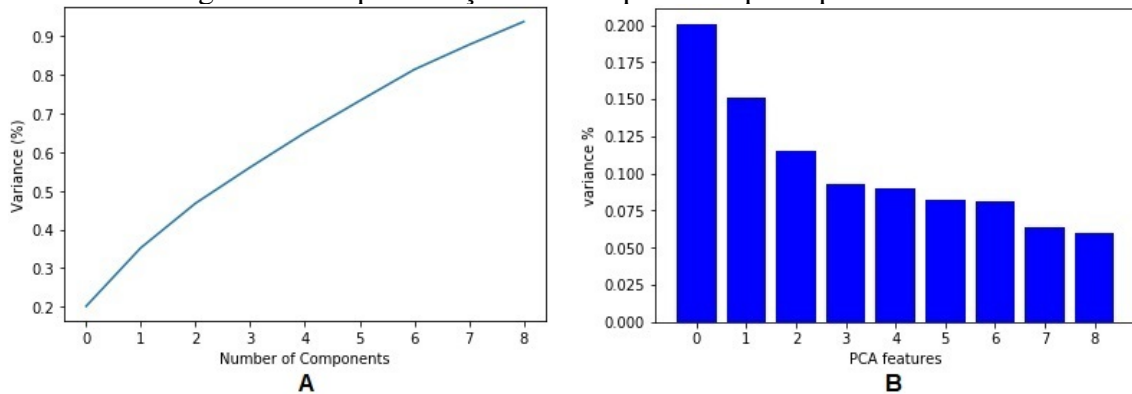
Após esta análise preliminar os dados foram padronizados para unidades de desvio padrão, para posterior aplicação dos algoritmos.

### 4.3.2 Redução de dimensionalidade com PCA

É fundamental procurar manter a maior parte das características do conjunto original ao reduzir a dimensionalidade. Com menos dimensões, a interpretação é facilitada e a clusterização pode ser otimizada. Para aplicar o algoritmo PCA, é necessário definir a quantidade de componentes principais que se deseja aplicar. Este número será a nova quantidade de dimensões da base de dados, que é a menor quantidade de dimensões possível mantendo a maior similaridade possível ao conjunto.

Caso fosse possível manter apenas duas ou três dimensões, seria possível visualizar os dados bi ou tridimensionalmente. Porém, como exibido pela Figura 4.4-A, dois e três componentes principais representariam respectivamente apenas 35% e 47% da variância dos dados - que é um valor baixo.

Figura 4.4: Representação dos componentes principais do PCA



Fonte: O autor

Em outra análise, ilustrada pela Figura 4.4-B, é possível ver que a variação dos três primeiros componentes principais são de 20%, 15% e 11%, respectivamente. Do quarto ao sétimo, a variância decresce 9% a 8%, a partir do qual cai a 6%. Este contraste entre a variância do sétimo e oitavo componentes principais (onde a queda é mais acentuada) levou a escolher sete componentes principais, que em conjunto representam 81% da variância original das onze dimensões e são um bom resultado para a redução de dimensionalidade.

### 4.3.3 Clusterização

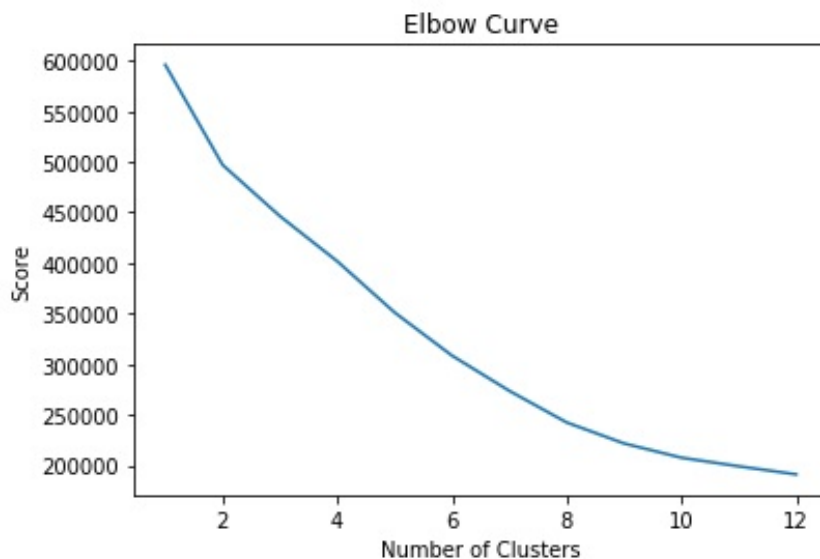
Esta subseção descreve a aplicação dos algoritmos K-Means e DBSCAN.

#### 4.3.3.1 K-Means

Como entrada para o algoritmo K-means é necessário apenas informar o número desejado de clusters - " $k$ ". Para identificar a melhor escolha, foi utilizado o método *Elbow* (do inglês, cotovelo), procurando minimizar o erro de clusterização (PANAPAKIDIS; MOSCHAKIS, 2019). Seu funcionamento baseia-se na soma das distâncias do cluster mais próximo. Quanto mais clusters, mais o valor tende a zero. A Figura 4.5 representa o resultado do método.

Desta forma, foi optado em utilizar oito clusters, identificado pelo ponto em que

Figura 4.5: Método *Elbow* para definição da quantidade de clusters



Fonte: O autor

o aumento da quantidade começa a tornar-se um pouco menos significativa para as distâncias mencionadas acima. Na sequência foi aplicado K-Means, exportando os consumidores e seus clusters em formato CSV. Isto permitiu analisar as características de cada cluster. Os resultados são apresentados no capítulo seguinte.

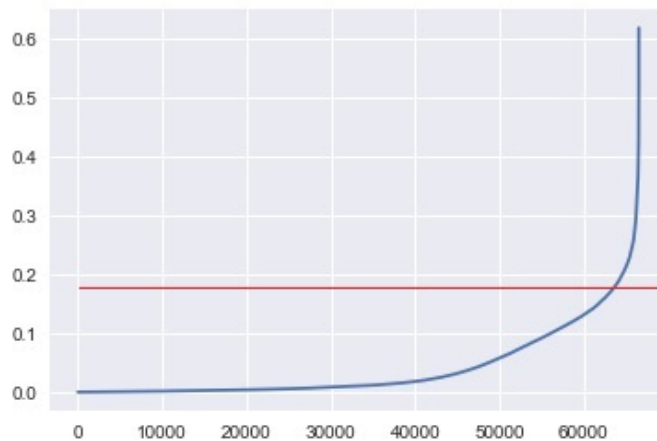
#### 4.3.3.2 DBSCAN

O valor para o parâmetro *eps* (distância máxima entre dois pontos para ser considerado dentro de uma vizinhança) foi determinado baseado no trabalho de Rahmah e Sitanggang (2016). Estes sugerem calcular a distância dos  $n$  vizinhos mais próximos para cada ponto, utilizando o algoritmo *NearestNeighbors*. No eixo que a curva do gráfico é mais acentuada, há mais contraste entre densidade dos pontos. A Figura 4.6 apresenta esta ilustração. É possível identificar que os valores indicados ficam entre 0.1 e 0.2. Com base nos experimentos realizados, o parâmetro *eps* foi definido como 0.18.

O valor de *min\_samples* (que representa a quantidade mínima de pontos para formar uma vizinhança) foi escolhido através da exploração sistemática de seus valores, avaliando o número adequado de clusters e o resultado do índice *Davies-Bouldin* (DAVIES; BOULDIN, 1979). O parâmetro foi definido como 105.

Valores muito baixos ou muito elevados para estes parâmetros eventualmente resultavam em maiores índices *Davies-Bouldin*, porém com pouca ou muita quantidade de clusters e consumidores em cada cluster. Consequentemente, o resultado final pretendido seria inviabilizado. Ao final da aplicação deste algoritmo, foram obtidos 11 clusters. Os

Figura 4.6: Definição do parâmetro eps do DBSCAN



Fonte: O autor

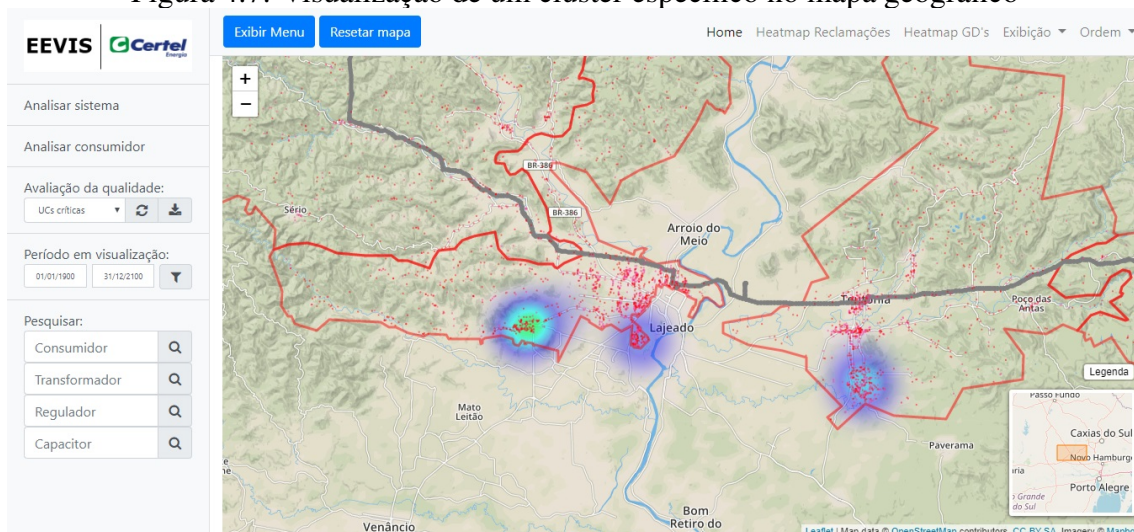
resultados são apresentados no capítulo seguinte.

#### 4.3.4 Visualização dos clusters

Com o resultado da aplicação da clusterização dos consumidores, estes foram alimentados na plataforma de visualização, em seção específica denominada *Clusterização*.

Neste módulo, foi criada uma opção para selecionar o cluster que se deseja visualizar, a partir do qual é criada uma visualização de linhas de contorno sobre o mapa geográfico para visualizar o cluster escolhido - ver Figura 4.7.

Figura 4.7: Visualização de um cluster específico no mapa geográfico



Fonte: O autor

Também foi criada uma opção para exportar um arquivo CSV que contém a lista de UC's e o seu respectivo cluster. Isto permitirá ao usuário facilmente identificar os



consumidores que deseja tratar ou priorizar com eventuais operações na rede.

## 5 AVALIAÇÃO E RESULTADOS

Este capítulo avalia os resultados obtidos nas duas partes da pesquisa. A primeira seção apresenta o procedimento e resultados da parte da visualização de informações, enquanto a seção seguinte trata dos resultados da avaliação da qualidade com clusterização.

### 5.1 Resultados da visualização de informações

A boa usabilidade requer que o usuário possua um mínimo conhecimento prévio do que está sendo visualizado, para conseqüentemente bem entender o funcionamento da abordagem visual que foi construída e atingir os resultados que se espera (MILANO, 2009).

Portanto, nesta seção, foram analisados os resultados da validação com os usuários - todos com conhecimento técnico comprovado e que trabalham na área, para avaliar a usabilidade e sua avaliação subjetiva. Para isto, foi utilizado um questionário próprio e também o questionário SUS (do inglês, *System Usability Scale*, ver Apêndice A) (BROOKE et al., 1996).

Esta seção está subdividida na descrição dos participantes, do procedimento, nos resultados obtidos e por final, na discussão.

#### 5.1.1 Participantes

Os participantes são funcionários de uma empresa de distribuição de energia elétrica. Ao total, nove pessoas participaram do experimento, todas do sexo masculino. A idade variou de 28 a 49 anos (média 36 e desvio padrão 7.14). Em relação à formação, foram três Engenheiros (33%), sendo dois Engenheiros Eletricistas e um Engenheiro de Controle e Automação (um deles possui Mestrado, 11%). Os outros seis são Técnicos em Eletrotécnica (67%), sendo que dois estão cursando Engenharia Elétrica (22%) e um é formado em Administração de Empresas (11%).

A ocupação dos mesmos variou entre Projetista de rede (quatro, 44%), Supervisor (dois, 22%), Analista (um, 11%), Eletrotécnico (um, 11%) e Coordenador (um, 11%). O tempo de experiência nesta função foi de: três participantes de até 5 anos, quatro de até 10 anos, e dois com mais de 10 anos de experiência.

Todos utilizam sistemas de informática diariamente. Em relação à utilização prévia da plataforma avaliada, seis possuíam experiência prévia (67%) e três não (33%). Quando a plataforma estava praticamente concluída, em torno de uma semana antes dos experimentos, ela foi liberada para o uso pela distribuidora. Portanto, aqueles que possuíam alguma experiência prévia, foram aqueles que já haviam acessado o sistema ao menos uma vez.

### 5.1.2 Procedimento

O experimento foi realizado em horário marcado com cada participante, em um mesmo computador. Cada experimento levou aproximadamente 26 minutos. Não havia limite de tempo para realizar o experimento. Todos assinaram um termo de consentimento e participação, explicando o procedimento e as intenções do estudo. O experimento iniciou com o preenchimento de um questionário pessoal e situacional (cujos dados foram detalhados na seção anterior). Em seguida, cada participante recebeu uma breve treinamento de 5 minutos, explicando as visualizações e funcionalidades disponíveis.

Para efetivamente avaliar a percepção dos usuários utilizando a abordagem visual, foram criadas situações hipotéticas de análise do sistema ou de um consumidor específico - por exemplo, simulando a reclamação de um consumidor. Desta forma, as tarefas tiveram objetivos claros e bem definidos. Foram aplicadas 12 tarefas, listadas a seguir, para que o participante utilizasse cada visualização da plataforma e extraísse informações da mesma, através unicamente de análise e interpretação. A ordem das tarefas foi a mesma para todos os participantes, como também o nível de dificuldade foi considerado similar para todas. As tarefas iniciaram pelas visualizações geográficas, passando em seguida para o Módulo Analisar Sistema e, por final, para o Módulo Analisar Consumidor. Não houve respostas corretas ou incorretas, devendo o usuário sempre informar a sua interpretação e se existiam ou não características relevantes a respeito da qualidade da energia na sua análise. Após cada tarefa, os participantes avaliaram a sua percepção.

1. Utilize a visualização geográfica e navegue pela rede. Aproxime a visualização da região que deseja. Veja a legenda ao lado direito e identifique os componentes na rede.
2. Extraia informações relevantes da visão geral das reclamações.
3. Identifique os alimentadores que mais possuem reclamações.

4. Verifique se existem consumidores muito distantes do transformador ou da subestação.
5. Analise o consumo e a quantidade de UCs da classe industrial em relação à classe residencial.
6. Analise o carregamento geral dos transformadores trifásicos da distribuidora, e a quantidade sobrecarregada.
7. Avalie o histórico de reclamações críticas.
8. Avalie na rede da distribuidora as regiões onde há maior densidade de usinas de Geração Distribuída.
9. Avalie como está o carregamento do transformador que alimenta o consumidor 63928.
10. Avalie como está o balanceamento de fases do transformador que alimenta o consumidor 77524.
11. Avalie a vizinhança do consumidor 38122 e identifique se existe algo que possa impactar negativamente na qualidade da sua energia.
12. Avalie, no mapa, a rede da região do consumidor 7202 e identifique se existe algo que possa impactar positiva ou negativamente na qualidade da sua energia.

Foi utilizada a escala Likert de cinco pontos (1 - Discordo fortemente, 5 - Concorde fortemente), quanto à questão 'Achei útil a visualização [tipo e finalidade]'. Ao final, foi solicitado o preenchimento do questionário de percepção e o questionário SUS (BROOKE et al., 1996).

### 5.1.3 Discussão

Russom et al. (2011) realizou uma pesquisa com companhias de diversas áreas, onde foi apontado que alguns dos motivos que organizações trocam de plataforma analítica - quando já a possuem - é porque são pouco adequadas para os usuários finais ou para análise visual. Portanto, percebe-se que para este tipo de estudo é de suma importância a validação com usuários.

No questionário situacional, 100% dos entrevistados concordam ou concordam fortemente com a sentença *os sistemas atuais que utilizo no meu trabalho possuem muitos dados em forma de tabela ou texto, e que poderiam possuir mais visualizações gráficas dos dados*. Destes, 88% concordam ou concordam fortemente que analisar os dados do

meu trabalho de forma visual poderia atender melhor as necessidades da minha rotina.

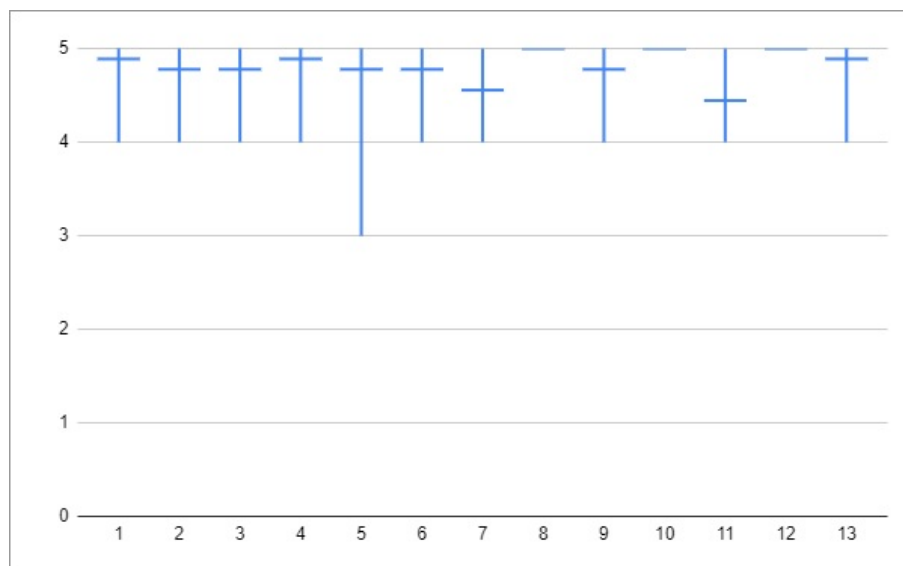
A Figura 5.1 mostra os resultados das avaliações para cada uma das tarefas propostas. A numeração das perguntas segue a lista abaixo:

1. Achei útil a visualização geográfica do sistema elétrico.
2. Achei útil a visão geral dos processos de qualidade com Sunburst e Heatmap.
3. Achei útil a visualização de reclamações por alimentador com Treemap.
4. Achei útil a visualização da distância do TR e da SE dos consumidores com Gráfico de Dispersão.
5. Achei útil a visualização de UC's e consumo com Gráfico Radar e Coordenadas Polares.
6. Achei útil a visualização do carregamento dos transformadores com Gráfico de Barras Empilhadas.
7. Achei útil a visualização do histórico de reclamações com Gráfico de Linhas.
8. Achei útil a visualização de Linhas de Contorno (mapa de calor) para geração distribuída / reclamações.
9. Achei útil a visualização do carregamento de um transformador específico com Gráfico de Linhas.
10. Achei útil a visualização do balanceamento das fases com Gráfico de Barras.
11. Achei útil a visualização da vizinhança de um consumidor com Gráfico de Bolhas.
12. Gostei e achei útil a elaboração do relatório ao analisar um consumidor.
13. Achei útil a visualização do mapa geográfico de um consumidor específico.

Embora não sendo baixas, as menores pontuações foram em relação às visualizações do histórico de reclamações com gráfico de linhas (Questão 7, média 4.5, desvio padrão 0.52) e vizinhança do consumidor com gráfico de bolhas (Questão 11, média 4.4 e desvio padrão 0.53). Em análise de causa, pode-se compreender que para o primeiro houve uma certa dificuldade no entendimento do gráfico em função da quantidade de informações, enquanto para o segundo foi sugerida a aplicação de mais filtros, bem como uma menor familiaridade com esta técnica de visualização. Para o gráfico de bolhas foi sugerida ainda a inclusão de mais legendas explicativas (por exemplo, que o tamanho do círculo representa a curva kVAs).

As pontuações mais altas (média 5) foram relativas às visualizações do mapa de calor na visualização geográfica, balanceamento de fases com gráfico de barras e do re-

Figura 5.1: Avaliação das visualizações por tarefa: mínimos, máximos e média por questão



Fonte: O autor

latório resumindo a análise de todo o entorno de um consumidor específico. Prováveis causas deste resultado podem incluir o quanto as visualizações e informações são novas para os participantes, que antes não conheciam ou não possuíam outras formas de obter tais informações ou de modo prático, como é o caso do mapa de calor. Outro fator relevante que aumentou a pontuação destas visualizações é o quanto os usuários sentiram-se confortáveis na análise. Que é o caso, por exemplo, do gráfico de barras do balanceamento de fases - visto que é um tipo mais comum e que os participantes estavam mais familiarizados, e por isso naturalmente não exigiu o esforço intelectual necessário para compreender determinados tipos de gráfico pela primeira vez.

O tempo médio do experimento pode ser considerado relativamente baixo. Contudo, as causas disto podem ser a combinação da clara definição de cada tarefa, a objetividade de cada uma, bem como a experiência dos participantes. Em função de cada visualização possuir uma finalidade específica, não é possível comparar o tempo de realização de cada tarefa entre elas. Contudo, sempre é possível extrair informações relevantes quanto ao tempo. O tempo médio de realização de todas as tarefas foi de 13min59s (desvio padrão 5min12s). Por faixa de tempo de experiência, aqueles com até 5 anos levaram em média 11min41s, com até 10 anos 15min08s, e mais de 10 anos 10min32s.

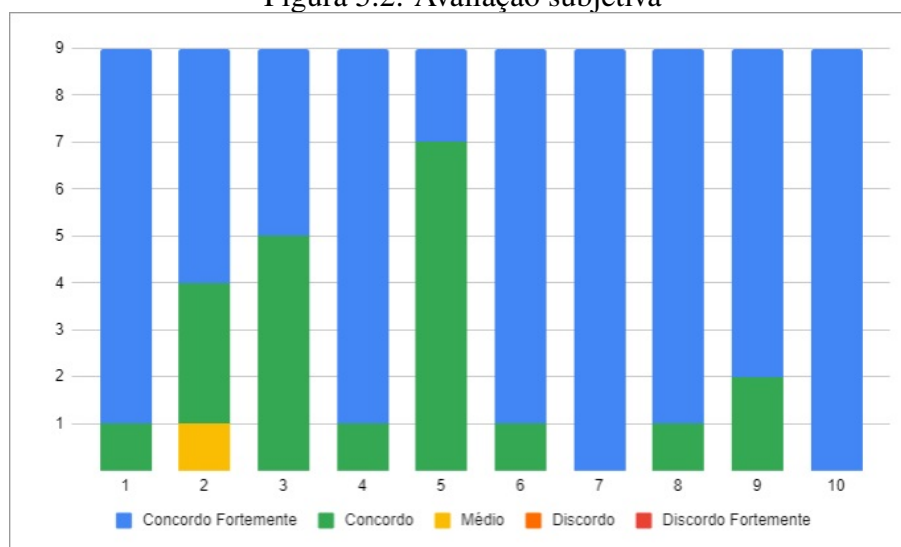
Avaliando isoladamente, o tempo médio para técnicos foi de 15min51s (desvio padrão 4min37s) e para graduados ou graduandos foi de 10min40s (desvio padrão 4min53s). Partindo do pressuposto que participantes graduados são mais habituados a trabalhar com visualizações gráficas (trabalhadas durante a graduação independente do contexto), este

dado pode indicar também que a familiaridade da área elétrica com a área Visualização de Informações não é tão estreita.

A Figura 5.2 apresenta a avaliação subjetiva da plataforma. Os itens 1 a 10 são representados pela questões abaixo:

1. Acho a ferramenta interativa.
2. Achei que as visualizações foram eficazes.
3. Achei que a quantidade de interação disponibilizada foi suficiente.
4. Achei o sistema relevante.
5. Achei que as visualizações permitem uma melhor avaliação se comparadas com os sistemas atuais.
6. Eu utilizaria este sistema na minha rotina de trabalho relacionada à qualidade de energia.
7. Eu acho que este sistema agilizaria o meu trabalho e reduziria o esforço necessário para avaliar a qualidade de energia.
8. Eu acho que este sistema poderia ser utilizado por outras empresas da área.
9. Eu acho que visualizações, como estas que foram implementadas, são importantes para tornar eficiente a rotina de trabalho.
10. Eu acho que visualizações, como estas que foram implementadas, são importantes para entender melhor os dados com que trabalhamos.

Figura 5.2: Avaliação subjetiva



Fonte: O autor

Somando este resultado com o anterior, vemos que a plataforma obteve uma aceitação significativa. Além da utilidade das visualizações, pode-se afirmar que a limitação

de comparabilidade dos participantes com outras abordagens visuais possa ter também alavancado os resultados.

Todas as questões obtiveram ótima concordância. As mais significativas foram em relação à redução do esforço e ao melhor entendimento dos dados quando utilizados com a plataforma avaliada.

A avaliação da usabilidade com o questionário SUS confirmou os dados coletados nos questionários anteriores. O resultado final atingido foi de 88.6. A questão com menor pontuação foi *Considerarei o sistema mais complexo do que o necessário*. Foi percebido durante as entrevistas que o efeito que provavelmente causou este resultado foi que alguns participantes não lembravam o local com a visualização mais adequada para atender à tarefa solicitada ou tinham alguma dificuldade para entender a visualização que atende à tarefa proposta.

Lembrando que o questionário SUS possui perguntas com viés negativo, houve uma grande discordância com as afirmações: *Acho que necessitaria de ajuda de um técnico para conseguir utilizar este sistema* e *Achei que este sistema tinha muitas inconsistências*. Como são usuários especialistas com conhecimento técnico na área, os participantes consideraram não precisar a ajuda de um técnico para auxiliar no uso do sistema. Além disso, foram evidenciadas poucas ou nenhuma inconsistência durante as entrevistas.

Os comentários gerais dos participantes sobre as visualizações que mais gostaram ou mais acharam útil mostram o nível de utilidade cobrindo diferentes perfis de usuário. Em geral, foi percebido que cada participante escolheu a visualização que mais é aplicável para a sua rotina, não podendo ser evidenciada uma visualização comum que mais agradasse ou fosse útil igualmente a todos participantes. Por exemplo, funcionários envolvidos com Geração Distribuída preferiram visualizar a densidade das usinas no mapa geográfico com mapa de calor, enquanto participantes envolvidos com a análise de reclamações preferiram as visualizações do módulo 'Análise do consumidor'.

A coleta de críticas e sugestões evidenciou a possibilidade de ser dada continuidade à evolução da plataforma. Afirmações como *Incluir o resumo de quantidade total de micro [geração] por região e fase que estão ligadas* e *Sugiro informar ou mapear áreas onde foram efetuadas limpezas de rede*. foram assinaladas. Sugestões bem específicas como cores e filtros adicionais para alguns gráficos também foram registradas. Por exemplo, para evidenciar a sobrecarga de transformadores com outra cor e filtros por subestação na quantidade de reclamações por alimentador.

Dados os resultados apresentados, pode-se confirmar a utilidade e aplicabilidade



da abordagem proposta.

## 5.2 Resultados da clusterização

Esta seção avaliará os clusters gerados pelos dois algoritmos, com base em avaliação quantitativa e qualitativa. As duas últimas seções tratam da avaliação de consumidores com a realização de medições *in-loco* e a discussão dos resultados.

A implementação baseou-se também na exploração da normalização ou não dos dados (convertendo-os para a escala de 0 a 1). Para o algoritmo DBSCAN, o resultado foi aprimorado, e é consequência disto que a sua visualização deste é esférica. Lembrando que ambos foram padronizados para a escala de unidades de desvio padrão.

### 5.2.1 Avaliação quantitativa

A avaliação quantitativa é realizada com algoritmos que calculam índices sobre os clusters gerados. Este procedimento é conhecido como validação. Foram utilizados índices internos, analisando a estrutura dos grupos gerados<sup>1</sup>. Foram aplicados dois algoritmos: o *índice Silhouette* (ROUSSEEUW, 1987) e o *índice Davies-Bouldin* (DAVIES; BOULDIN, 1979). O primeiro é calculado para cada ponto, enquanto o segundo para cada cluster.

O *índice Silhouette* é calculado para cada valor do conjunto, utilizando a distância euclidiana média intra-cluster e a distância euclidiana média entre o cluster mais próximo e varia no intervalo [-1,1]. Quando está próximo a 0, significa que há sobreposição de clusters. Quando negativo, é um indicativo de que há dados atribuídos a clusters errados ou mais similares a outro cluster. Quanto mais próximo de 1, melhor (ROUSSEEUW, 1987). O índice para K-Means resultou em 0.25 e para DBSCAN em -0.3105. O K-Means com normalização ficou em 0.21 (menos adequado) e o DBSCAN sem normalização em 0.01 (melhor, mas como explicado abaixo, o índice Davies-Bouldin é mais aplicável neste caso).

É notável que o índice para o algoritmo K-Means foi bastante superior. Sendo 1.0 a pontuação ideal para um conjunto de dados, a pontuação obtida indica que não houve sobreposição ou atribuição incorreta dos clusters, podendo ser considerada apta para ser

---

<sup>1</sup>Visto que não há comparações externas a serem realizadas, não é possível utilizar índices externos

estudada. É possível observar, de acordo com a Figura 5.3, que há uma adequada divisão de clusters entre a população.

Figura 5.3: Visualização bidimensional da clusterização com K-Means



Fonte: O autor

Já em relação ao algoritmo DBSCAN, a pontuação foi diferente. Contudo, é fundamental avaliar a forma como o algoritmo atua. O DBSCAN realiza a clusterização com base na densidade de seus pontos, clusterizando, portanto, também geometrias não convexas. Neste caso, há pontos que podem ficar distantes do centroide e que inclusive podem aproximar-se mais de outros clusters do que àquele ao qual pertence, causando a diminuição de sua pontuação. Portanto, o resultado do DBSCAN continua válido.

O *índice Davies-Bouldin* calcula a similaridade média entre cada cluster e seu cluster mais parecido, utilizando a relação da distância intra-cluster e entre-clusters. Quanto mais próximo de 0 (menor valor possível), melhor o resultado - pois significa que os clusters idealmente estão compactos e distantes entre si (DAVIES; BOULDIN, 1979). O índice para K-Means resultou em 1.25 e para DBSCAN em 1.16. O K-Means com normalização ficou em 1.44 e o DBSCAN sem normalização em 1.30.

Os dois algoritmos ficaram com uma pontuação similar, levemente melhor para o DBSCAN. Isto significa que os clusters de ambos algoritmos, em relação à distância entre si e dispersão, estão parecidos segundo o cálculo utilizado.

Importante salientar para trabalhos futuros que a aplicação do K-Means sem PCA resultou em índices menos adequados de *Silhouette* e *Davies Bouldin*, respectivamente: 0.17 (quanto maior, melhor) e 1.41 (quanto menor, melhor).

Não é possível determinar o melhor algoritmo para esta pesquisa baseado em apenas estas análises. Na sequência, será avaliado qualitativamente o resultado da clusterização, que tem um papel também muito relevante em função de poder ser avaliada a utilização prática.

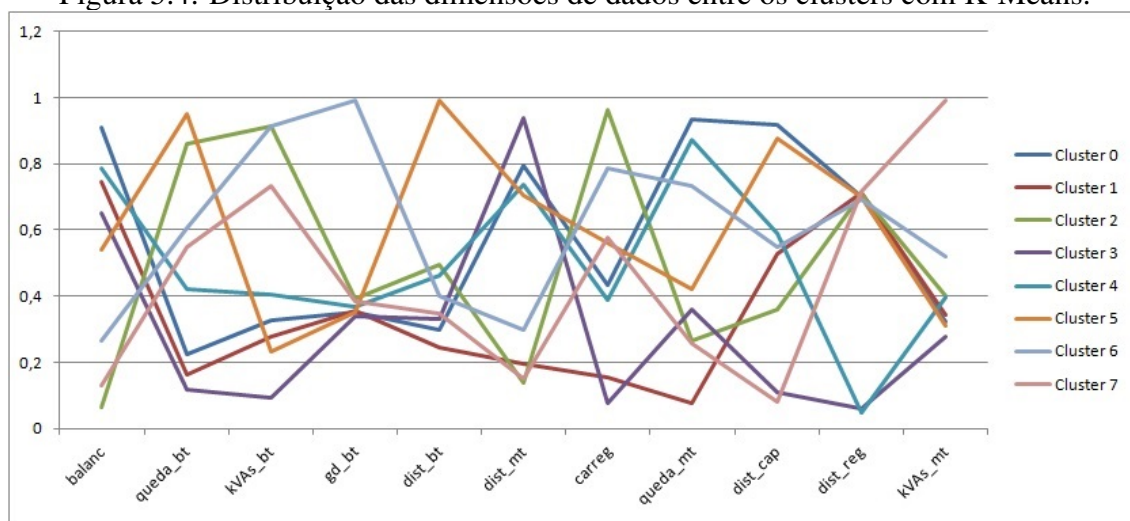
## 5.2.2 Avaliação qualitativa

Após extrair a identificação do cluster para cada unidade consumidora, foi possível avaliar as características de cada conjunto. O resultado do K-Means, devido a sua forma de funcionamento, foi mais igualmente distribuído: a média ficou em 8317 consumidores por cluster, sendo que o menor possui 1555 consumidores.

Para visualizar os resultados da forma mais adequada possível, foi aplicado o PCA sobre o resultado da clusterização. Os dois primeiros componentes principais foram plotados em uma visualização bidimensional, ilustrado pela Figura 5.3.

Para este algoritmo, os valores que caracterizam má qualidade de energia ficaram distribuídos entre diversos clusters. Isto é exibido na Figura 5.4, evidenciando que os maiores valores de cada dimensão, em geral, ficaram em diferentes clusters. Por exemplo, o desbalanceamento das fases (*balanc*) foi característico do cluster 0, enquanto a elevada queda de tensão (*queda\_bt*) e distância do transformador (*dist\_bt*) do cluster 5, entre outros.

Figura 5.4: Distribuição das dimensões de dados entre os clusters com K-Means.



Fonte: O autor

Neste contexto, avalia-se que a quantidade de consumidores é muito elevada para efetivamente poder selecionar e priorizar os ajustes para os consumidores críticos. Além

disso, a análise foi dificultada em função da clusterização não discriminar adequadamente as características de má-qualidade de energia.

Uma diferença considerável entre os algoritmos, além da forma de seu funcionamento, é a identificação de ruídos e *outliers*. Enquanto o K-Means obrigatoriamente precisa incluir todos consumidores em algum cluster, o DBSCAN os identifica e isola.

Foi isto que aconteceu na sua aplicação. A maioria dos consumidores foi classificada como ruído - em torno de 90%. Isto deixou evidente a característica ruidosa da base de dados.

Procurando diminuir a quantidade de ruído, foi explorado e identificado que valores diferentes para *eps* e *min\_samples* permitiam diminuir consideravelmente o ruído para menos de 10%. Contudo, todos estes consumidores antes considerados como ruído foram alocados em um único cluster demasiadamente grande, com mais de 60000 consumidores.

É factível observar que a junção desta gama de consumidores como ruído ou em um único cluster é consequência de todas as variáveis serem contínuas, o que causa uma distribuição muito regular dos dados no espaço, produzindo uma clusterização muito mais concentrada.

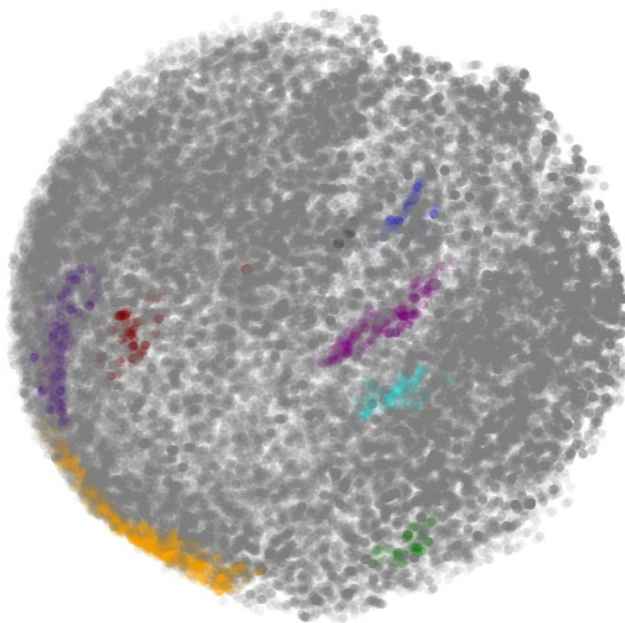
Portanto, o estudo foi continuado com os parâmetros originais. A grande quantidade de ruído não é um ponto negativo, pois a quantidade enxuta de consumidores nos clusters foi também mais adequada para a finalidade desta pesquisa. Os clusters ficaram, em média, com 303 consumidores - o que permite muito objetivamente identificar aqueles com características realmente críticas.

A Figura 5.5 apresenta a visualização do resultado deste algoritmo. Como são diversas dimensões, também foi utilizado o PCA, com os dois primeiros componentes principais. A cor cinza representa os consumidores identificados como ruído, e as demais cores representam efetivamente os clusters. Embora exista uma perda considerável para visualizar os dados de forma bidimensional (43% da variância original), já é possível notar que os clusters são bem concentrados. A visualização esférica é consequência da normalização para este algoritmo, além da padronização.

Dentre todos os 11 clusters, foi claramente identificado um que apresenta características que podem indicar má qualidade de energia, representado pela linha azul que atravessa todas dimensões quase sempre pelo topo conforme a Figura 5.6. É possível notar que o cluster 4 apresenta diversas dimensões cujo valor está considerável.

O cluster é constituído por 160 consumidores, identificados com as maiores médias para a maioria das variáveis. Algumas destas dimensões foram características tam-

Figura 5.5: Visualização bidimensional da clusterização com DBSCAN



Fonte: O autor

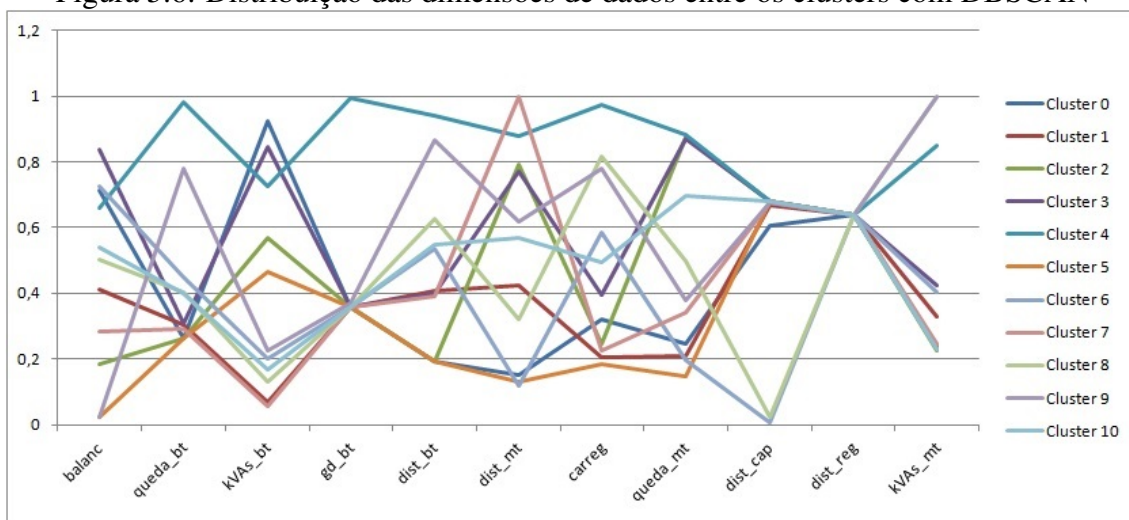
bém de outros clusters, contudo, isoladamente. Relevante mencionar que, como o objetivo foi encontrar clusters que agregassem características de má-qualidade de energia, este estudo focou no cluster que mais atendeu a esta premissa, não sendo analisados, em detalhe, os resultados de cada cluster individualmente. A visão geral de cada um é possível ser analisada pela Figura 5.6.

É possível avaliar as dimensões que mais representam a visualização bidimensional através da utilização do PCA. A Figura 5.7 apresenta este resultado: células cuja cor fica mais à esquerda ou direita da legenda representam valores mais distantes da média. Isto permite avaliar a clusterização sob outra perspectiva, elaborando uma visualização do cluster obtido com a dimensão que mais o representa.

Na Figura 5.6 fica evidente que a dimensão *Geração Distribuída* contrasta entre o cluster 4 e os demais. Com a Figura 5.7, identifica-se que o componente principal que representa a *Geração Distribuída* é o Oitavo PC. Por conseguinte, pode-se visualizá-lo bidimensionalmente através do primeiro e oitavo para ilustrar com mais evidência o cluster 4 - ver Figura 5.8.

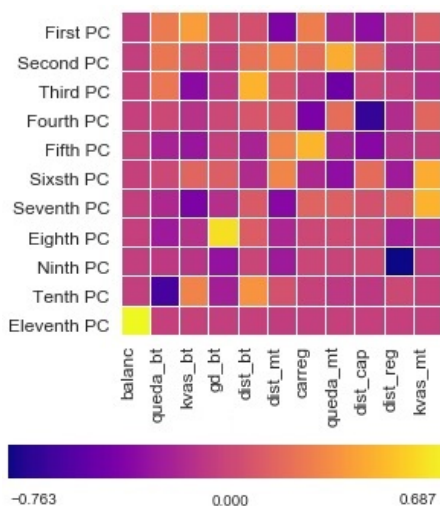
Importante salientar que esta imagem bidimensional, com estes componentes principais, possui 23% da variância original - não sendo fidedigna à geometria de fato. O cluster 4 é ilustrado em azul.

Figura 5.6: Distribuição das dimensões de dados entre os clusters com DBSCAN



Fonte: O autor

Figura 5.7: Representatividade de cada dimensão para cada os onze componentes principais



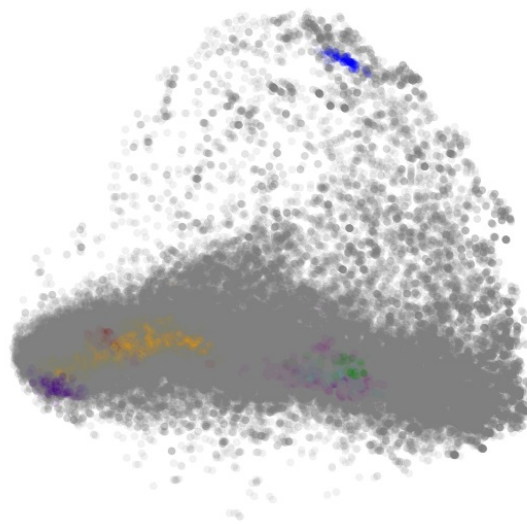
Fonte: O autor

### 5.2.3 Avaliação com medições

Para averiguar os resultados obtidos com a clusterização, foram aleatoriamente selecionados consumidores do cluster 4 de transformadores diferentes para a realização de medições *in-loco*. As medições seguiram o mesmo padrão de medição realizado para os processos de qualidade de energia, portanto resultando também em índices de DRP (limite 3%) e DRC (limite 0.5%) (AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA - ANEEL, 2017).

De um total de 160 consumidores nesse cluster, foram realizadas 9 medições - de acordo com a disponibilidade de equipamentos e tempo da distribuidora. As medições foram realizadas no período de 15–24 de junho de 2020, sendo sete dias para cada UC,

Figura 5.8: Resultado do DBSCAN evidenciando o cluster 4, representado pela cor azul.



Fonte: O autor

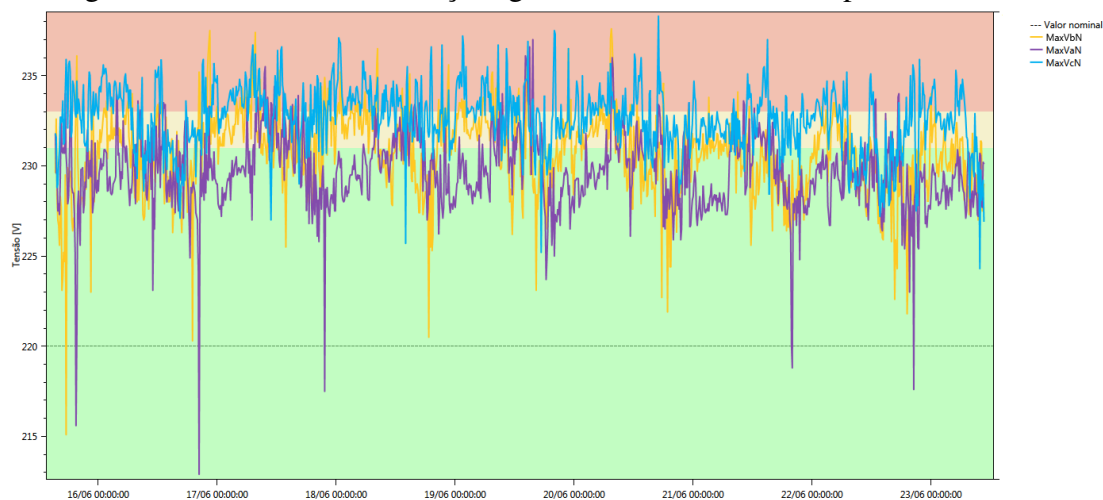
e utilizando analisadores e registradores portáteis de grandezas elétricas. Cada medição resultou em um gráfico de tensões mínimas, médias e máximas, visto que a leitura é realizada para cada 10 minutos.

Das nove medições realizadas, três não registraram problemas de energia, enquanto as outras seis continham ao menos uma violação dos limites de tensão (de 202V a 231V), ou ficavam muito próximo ao limite de DRP/DRC, ou estavam violando também estes índices. Vale lembrar que isto pode não significar que haverá uma reclamação ou que existe de fato algum problema que pode ser percebido pelo consumidor, que obviamente já teria registrado uma reclamação para os casos tão sensíveis. Todavia, isto permite que seja realizado o trabalho preventivo pela distribuidora. Mesmo com a seleção de consumidores com a clusterização, as medições são importantes para, além de confirmar os resultados, permitir o planejamento adequado da solução baseado em dados medidos.

A Figura 5.9 exhibe o resultado de uma medição. Este é um exemplo de tensão alta, que viola os limites adequados de tensão, exigindo ajuste. Esta é uma UC trifásica (cada cor de linha representa uma fase medida): a área verde representa os limites adequados de tensão, a área amarela a área crítica e a área laranja a área precária. Nota-se que existe ao menos uma fase que viola claramente os limites adequados. Os índices para esta UC resultaram em:

- Fase 1: DRP 1.42% e DRC 0% (adequado);
- Fase 2: DRP 9.96% e DRC 0.53% (violado);
- Fase 3: DRP 36.44% e DRC 6.49% (violado).

Figura 5.9: Resultado da medição: gráfico de tensão máxima para UC trifásica



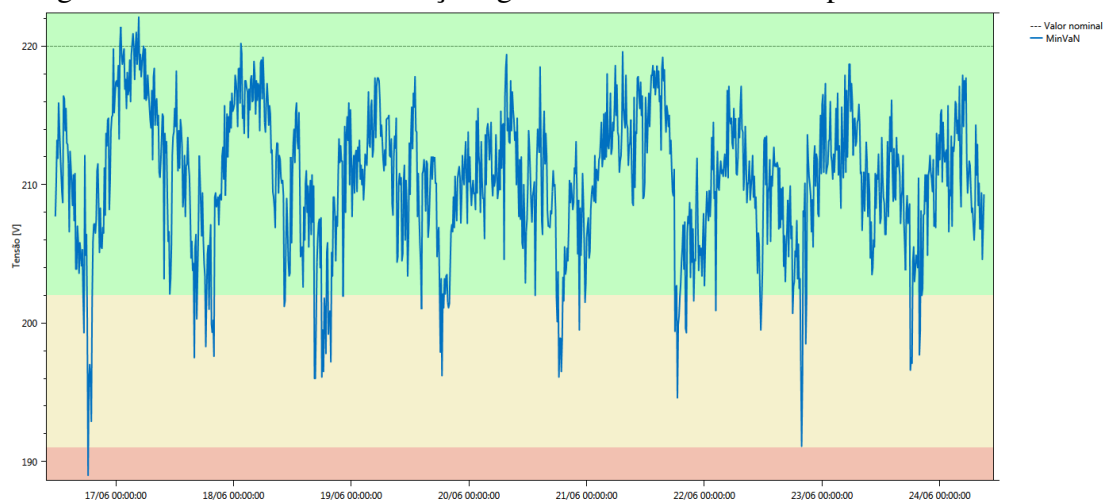
Fonte: O autor

O exemplo da Figura 5.10 é o inverso, sendo que foi evidenciado tensão baixa. Esta é uma UC monofásica, por isso existe apenas uma tensão no gráfico. Tensões baixas, em geral, são mais perceptíveis para o consumidor durante a utilização de equipamentos. As cores seguem o mesmo padrão anterior. Os índices para esta UC resultaram em:

- Fase 1: DRP 0.20% e DRC 0.50% (não violado).

Os índices não foram violados, tanto para DRP quanto para DRC, considerando a tensão média nos intervalos de 10 minutos. Contudo, conforme apontado no gráfico da Figura 5.10 que apresenta as tensões mínimas, este resultado indica que a qualidade da energia está inadequada.

Figura 5.10: Resultado da medição: gráfico de tensão mínima para UC monofásica



Fonte: O autor

Portanto, mediante as medições realizadas, foi possível identificar consumidores



que possuem índices fora dos limites adequados e podem possuir problemas perceptíveis na qualidade da sua energia. Através disto, distribuidoras podem realizar este tipo de seleção e priorização de consumidores, realizar análises de tensão e agir proativamente quando necessário para adequar os resultados encontrados.

#### **5.2.4 Discussão**

Sendo a mineração de dados um trabalho exploratório, foram sistematicamente aplicados valores diferentes para os parâmetros de cada algoritmo. Aqueles com melhores resultados quantitativa e qualitativamente foram apresentados aqui.

Ficou evidente a característica ruidosa da base de dados. A quantidade de dimensões, além da própria natureza dos dados de energia serem variáveis contínuas, acabou por produzir efeitos deste tipo. Além disso, a geometria do espaço e a forma de atuação dos algoritmos sugerem que o DBSCAN seja mais aplicável para este estudo.

Foi identificado um cluster que classifica 160 consumidores com características bastante evidentes que indicam má-qualidade de energia. Este resultado permite que estes consumidores sejam analisados de forma aprofundada pelos responsáveis técnicos e, caso comprovado, sejam realizadas adequações na rede.

O resultado obtido já permite a identificação proativa de consumidores que merecem atenção. Sobre isso, é observável que a clusterização pode ser um conveniente modo de encontrar consumidores com tendências de problemas na energia, de forma automática, e priorizá-los para medição e adequação da qualidade da energia.

Por mais concreta que seja a base de dados utilizada na clusterização, o passo seguinte à identificação destes consumidores é aplicar mecanismos técnicos que comprovem a existência de distorções na energia - por exemplo, a medição dos consumidores selecionados. Dadas as medições realizadas, foi possível encontrar de fato UC's que continham violações dos índices adequados, fortalecendo as conclusões desta pesquisa.

Dados os resultados mencionados acima, é possível afirmar que o objetivo de propor um modelo de avaliação da qualidade da distribuição de energia foi satisfatoriamente atingido.

### **5.3 Limitações**

A parte de visualização do presente estudo focou na análise da qualidade da energia global e inerente a cada consumidor, abrangendo toda a rede de distribuição. Esta pesquisa não focou em perturbações do tipo de harmônicas ou outras relacionadas. Também não fez parte a análise de interrupções de energia.

O modelo de avaliação com clusterização enfatizou consumidores do grupo B. Não incluiu consumidores do grupo A, que são, em geral, grandes indústrias, além de serem em muito menor quantidade - na distribuidora em estudo representam menos de 1% da quantidade total de consumidores.

## 6 CONCLUSÃO

Na presente pesquisa foi proposta uma nova abordagem de analítica visual e clusterização para aprimorar o Gerenciamento da Qualidade da Distribuição de Energia Elétrica. Esta construção foi baseada em conhecimentos técnicos transmitidos por profissionais da área de distribuição de energia, que participaram durante todo o processo - desde o planejamento até a avaliação final.

Na revisão literária, foram explorados sistematicamente os trabalhos anteriores relacionados à qualidade de energia, onde foi percebido que, em geral, os trabalhos continham um viés do sistema elétrico para analisar a qualidade da energia - tanto para visualização quanto para clusterização. Neste sentido, foi identificado o nicho existente para a avaliação da qualidade sob a perspectiva do consumidor final.

As ideias foram centradas em propor análises visuais do sistema elétrico neste cenário, modelando uma plataforma para esta finalidade. Esta inclui diversos mecanismos de Visualização de Informações, resumindo visualmente os atributos que influenciam a qualidade da energia aos consumidores, segregadas sob os conceitos de sistema geral e também peculiar a cada consumidor. A parte de clusterização foi realizada com dois algoritmos, categoricamente escolhidos por seu *modus-operandi*: K-Means e DBSCAN. Estes algoritmos consumiram uma base de dados construída a partir do processamento individual de cada consumidor, de acordo com as características elétricas exclusivas da sua conexão, formando uma base de dados personalizada.

Resultados mostraram que a abordagem visual obteve uma grande taxa de aceitação, onde todas as visualizações foram consideradas úteis pelos entrevistados, além de todos terem concordado ou concordado fortemente em a plataforma ser relevante. Este alcance foi confirmado através do questionário SUS, que atingiu 88.6. Apesar do resultado ser bastante positivo, não é sinônimo de finalização: naturalmente muitos entrevistados sugeriram novas implementações a serem feitas, com a inclusão de novas funcionalidades ou mais dados, de modo a dar continuidade à ferramenta. Além disso, a plataforma foi considerada apta a ser utilizada por outras distribuidoras, o que permitirá o atingimento de tais benefícios independente do tamanho da empresa.

Os resultados da clusterização deixaram evidentes a característica ruidosa da base de dados. Além disso, a geometria do espaço de dados e a forma de operação dos algoritmos resultaram na escolha do DBSCAN para este estudo. Em contrapartida, foi possível evidenciar que a clusterização pode ser um meio muito conveniente de identificar con-

sumidores com tendências de problemas na qualidade da energia. Foi possível identificar um cluster que classifica 160 consumidores com claras características que podem indicar problemas, bem como que as medições realizadas comprovaram que foi possível encontrar de fato UC's com violações dos índices adequados, possibilitando que estes sejam analisados detalhadamente e realizados ajustes na rede elétrica.

Concluindo, a soma destes resultados possibilita realizar análises práticas e proativas do sistema, mostrando que o objetivo de propor um modelo para avaliar a qualidade do sistema de distribuição de energia foi satisfatoriamente atingido. Além das contribuições supracitadas, é importante mencionar a própria perspectiva do consumidor para analisar e avaliar o sistema elétrico, onde a individualização destes representa um papel crucial na análise de ponta-a-ponta do sistema elétrico e que também permitirá avanços no Gerenciamento da Qualidade na Distribuição de Energia.

## **6.1 Reflexões**

Mesmo que o principal impacto deste estudo seja voltado à qualidade da energia, e a consequente satisfação dos consumidores conforme já bem detalhado anteriormente, existem também outros benefícios. Visto o trabalho preventivo que pode ser realizado pelas distribuidoras, existe um impacto econômico-financeiro que permite que o planejamento dos ajustes na rede sejam realizados de acordo com os recursos e o tempo julgados necessários, já atentos à situação elétrica de regiões específicas.

Interessante apreciar como a pesquisa e desenvolvimento em uma empresa pode trazer melhorias que vão muito além do escopo propriamente dito. Mudança de cultura, por exemplo. Durante o desenvolvimento deste estudo, principalmente nas etapas finais, foi notável o envolvimento dos funcionários da distribuidora enquanto participantes voluntários. Críticas construtivas, sugestões, melhorias, várias ideias para a evolução da plataforma foram sugeridas, mesmo além da qualidade da energia, motivados pela possibilidade de melhorar processos e rotinas de trabalho.

A visão centrada no consumidor trabalhada aqui, por exemplo, após os experimentos realizados, levou à mobilização do setor responsável pelos projetos de rede a solicitar informações dos transformadores mais sobrecarregados para planejamento de obras de melhoria, para ajustar tal situação priorizando os mais críticos.

Sem dúvida isto poderá ser palco para qualquer organização onde sejam realizados tais estudos, seja da área elétrica ou não, mostrando o potencial que existe entre as

entidades academia e empresa e as possibilidades de avanço mútuo.

## **6.2 Trabalhos futuros**

Embora a presente pesquisa tenha sido voltada à análise considerando diversas dimensões de dados relevantes à qualidade de energia, ainda há muito a ser explorado. Indica-se aqui a importância de ser dada continuidade incluindo visualizações de outras grandezas elétricas, como fator de potência ou perdas de energia. A realização de simulações de modificação na estrutura das redes, prevendo a qualidade de energia em diferentes layouts, também seria de grande relevância.

Também é significativo explorar técnicas de visualização aplicadas a outra grande expectativa dos consumidores (além da qualidade da energia): a continuidade do fornecimento. A visualização de dados referentes às interrupções, além de suas causas e consequências, conduzirão a importantes contribuições para ambas as áreas de visualização e elétrica.

Em relação ao modelo de avaliação da qualidade, como dito anteriormente, a avaliação poderá ter resultados aprimorados caso as características da rede elétrica dos consumidores que reclamarem sejam registradas antes do ajuste para corrigir os problemas da energia. Isto resultará em um banco de dados específico das características de reclamantes, abrindo novos caminhos a serem percorridos utilizando técnicas de aprendizado de máquina supervisionado.

Além disso, a utilização de perfis de consumo dos clientes poderá levar em consideração outras características que permitam aprimorar ainda mais os resultados, tanto da visualização quanto da clusterização.

## REFERÊNCIAS

- AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA - ANEEL. **Resolução Normativa Nº 414, de 9 de setembro de 2010**. 2010. Acesso em: 04 set. 2019. Disponível em: <<http://www2.aneel.gov.br/cedoc/ren2010414.pdf>>.
- AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA - ANEEL. **Resolução Normativa nº 753, de 13 de dezembro de 2016**. 2016. Acesso em: 04 set. 2019. Disponível em: <<http://www2.aneel.gov.br/cedoc/ren2016753.pdf>>.
- AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA - ANEEL. **Módulo 8 – Qualidade da Energia Elétrica**. Procedimentos de Distribuição - PRODIST, 2017. Acesso em: 04 set. 2019. Disponível em: <<http://www.aneel.gov.br/modulo-8>>.
- ASHEIBI, A. et al. Power quality data analysis using unsupervised data mining. **Australasian Universities Power Engineering Conference (AUPEC 2004) Brisbane: AUPEC.**, 2004.
- BARCELOS, M. A. et al. Uso de realidade aumentada na visualização de componentes de subestações de energia elétrica. In: SN. **XI Conferência de Estudos em Engenharia Elétrica**. [S.l.], 2013.
- BASOLE, R. C.; SAUPE, D. Sports data visualization [guest editors' introduction]. **IEEE Computer Graphics and Applications**, IEEE, v. 36, n. 5, p. 24–26, 2016.
- BERNARDO, N. Evolução da gestão da qualidade de serviço de energia elétrica no Brasil. **Rio de Janeiro: Projeto de Graduação-Universidade Federal do Rio de Janeiro**, 2013.
- BOWMAN, R. L. et al. Gliovis data portal for visualization and analysis of brain tumor expression datasets. **Neuro-oncology**, Oxford University Press US, v. 19, n. 1, p. 139–141, 2016.
- BROOKE, J. et al. Sus-a quick and dirty usability scale. **Usability evaluation in industry**, London–, v. 189, n. 194, p. 4–7, 1996.
- CAPUDER, T.; PANDZIC, H.; KUZLE, I. Visualization of distribution system planning for engineers and decision-makers. **Turkish Journal of Electrical Engineering & Computer Sciences**, The Scientific and Technological Research Council of Turkey, v. 25, n. 4, p. 2871–2884, 2017.
- CAVELLUCCI, C. et al. **Buscas informadas baseadas em grafos para a minimização das perdas em sistemas de distribuição de energia elétrica**. [S.l.]: [sn], 1998.
- CHEN, W.; GUO, F.; WANG, F.-Y. A survey of traffic data visualization. **IEEE Transactions on Intelligent Transportation Systems**, IEEE, v. 16, n. 6, p. 2970–2984, 2015.
- CHICCO, G.; NAPOLI, R.; PIGLIONE, F. Comparisons among clustering techniques for electricity customer classification. **IEEE Transactions on Power Systems**, IEEE, v. 21, n. 2, p. 933–940, 2006.

- CUFFE, P.; KEANE, A. Visualizing the electrical structure of power systems. **IEEE Systems Journal**, IEEE, v. 11, n. 3, p. 1810–1821, 2015.
- DAVIES, D. L.; BOULDIN, D. W. A cluster separation measure. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, n. 2, p. 224–227, 1979.
- DING, C.; HE, X. K-means clustering via principal component analysis. In: **Proceedings of the twenty-first international conference on Machine learning**. [S.l.: s.n.], 2004. p. 29.
- DUAN, X. et al. Synthetic evaluation of power quality based on fuzzy cluster analysis. In: IEEE. **2006 International Conference on Power System Technology**. [S.l.], 2006. p. 1–6.
- DUTTA, S.; OVERBYE, T. J. Information processing and visualization of power system wide area time varying data. In: IEEE. **2013 IEEE Computational Intelligence Applications in Smart Grid (CIASG)**. [S.l.], 2013. p. 6–12.
- ESTER, M. et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In: **Kdd**. [S.l.: s.n.], 1996. v. 96, n. 34, p. 226–231.
- FARFÁN, A. J. U.; PALTA, E. A. C.; TACCA, H. E. Three-dimensional maps of power quality loss based in the power tensor theory. In: **Simposio Internacional sobre la Calidad de la Energía Eléctrica-SICEL**. [S.l.: s.n.], 2011. v. 6.
- FEDERAÇÃO DAS COOPERATIVAS DE ENERGIA, TELEFONIA E DESENVOLVIMENTO RURAL DO RIO GRANDE DO SUL. **Critérios de Elaboração de Projetos**. 2014. Acesso em: 28 jun. 2020. Disponível em: <<https://www.fecoergs.com.br/pagina.php?cont=padroesProjetos>>.
- FLATH, C. et al. Cluster analysis of smart metering data. **Business & Information Systems Engineering**, Springer, v. 4, n. 1, p. 31–39, 2012.
- GARGOOM, A. M.; ERTUGRUL, N.; SOONG, W. A comparative study on effective signal processing tools for optimum feature selection in automatic power quality events clustering. In: IEEE. **Fourtieth IAS Annual Meeting. Conference Record of the 2005 Industry Applications Conference, 2005**. [S.l.], 2005. v. 1, p. 52–58.
- GÓES, A. R. T.; STEINER, M. T. A.; PENICHE, R. A. Classification of power quality considering voltage sags in distribution systems using kdd process. **Pesquisa Operacional**, SciELO Brasil, v. 35, n. 2, p. 329–352, 2015.
- GRIGORAS, G.; BOBRIC, E.-C. Clustering based approach for customers' classification from electrical distribution systems. In: **Proc. of International Conference on Energy and Environment, Bucharest, Romania**. [S.l.: s.n.], 2013.
- GRIGORAS, G. et al. The efficiency of the clustering techniques in the energy losses evaluation from distribution networks. **International Journal of Mathematical Models and Methods in Applied Sciences**, v. 5, n. 1, p. 133–141, 2011.
- GRIGORAS, G.; CARTINA, G.; ROTARU, F. Using k-means clustering method in determination of the energy losses levels from electric distribution systems. **World Scientific and Engineering Academy and Society (WSEAS)**, p. 52–56, 2010.

HAROLD, J. et al. Cognitive and psychological science insights to improve climate change data visualization. **Nature Climate Change**, Nature Publishing Group, v. 6, n. 12, p. 1080, 2016.

JASIŃSKI, M.; SIKORSKI, T.; BORKOWSKI, K. Clustering as a tool to support the assessment of power quality in electrical power networks with distributed generation in the mining industry. **Electric Power Systems Research**, Elsevier, v. 166, p. 52–60, 2019.

JASIŃSKI, M. et al. Cluster analysis of long-term power quality data. In: IEEE. **2016 Electric Power Networks (EPNet)**. [S.l.], 2016. p. 1–6.

JUREEDI, N.; ROSALINA, K.; KUMAR, N. P. Clustering analysis and its application in electrical distribution system. **International Journal of Electrical, Electronics and Computer Systems (IJEECS)**, 2013.

KAPOOR, R. et al. Detection of power quality event using histogram of oriented gradients and support vector machine. **Measurement**, Elsevier, v. 120, p. 52–75, 2018.

KLUMP, R.; DOOLEY, G.; WU, W. Displaying aggregate data, interrelated quantities, and data trends in electric power systems. In: IEEE. **36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the**. [S.l.], 2003. p. 10–pp.

KLUMP, R. P.; WEBER, J. D. Real-time data retrieval and new visualization techniques for the energy industry. In: IEEE. **Proceedings of the 35th Annual Hawaii International Conference on System Sciences**. [S.l.], 2002. p. 712–717.

KUMAR, P.; SINGH, A. K. Transmission line power flow visualization using 3-dimensional circle diagrams. In: IEEE. **2015 Annual IEEE India Conference (INDICON)**. [S.l.], 2015. p. 1–6.

LAWANSON, T. et al. Improving power distribution system situational awareness using visual analytics. In: IEEE. **SoutheastCon 2018**. [S.l.], 2018. p. 1–6.

LLOYD, S. Least squares quantization in pcm. **IEEE transactions on information theory**, IEEE, v. 28, n. 2, p. 129–137, 1982.

LU, Q. et al. Electricvis: visual analysis system for power supply data of smart city. **The Journal of Supercomputing**, Springer, p. 1–21, 2019.

MANGISI, R. **An investigation into visualisation and forecasting of real-time electrical consumption based on smart grid data**. Tese (Doutorado), 2015.

MILANO, F. Three-dimensional visualization and animation for power systems analysis. **Electric Power Systems Research**, Elsevier, v. 79, n. 12, p. 1638–1647, 2009.

MOLNAR, S.; GRUCHALLA, K. Visualizing electrical power systems as flow fields. In: EUROGRAPHICS ASSOCIATION. **Proceedings of the Workshop on Visualisation in Environmental Sciences**. [S.l.], 2018. p. 63–71.

MULLIEZ, E. et al. Life characterization of power distribution transformers using clustering techniques. In: **PHM Society European Conference**. [S.l.: s.n.], 2018. v. 4, n. 1.



- NGA, D. V. et al. Visualization techniques in smart grid. **Smart Grid and Renewable Energy**, Scientific Research Publishing, v. 3, n. 03, p. 175, 2012.
- NIGAR, Y.; AGALGAONKAR, A.; CIUFO, P. Visualising the effect of dg on voltage profile in medium voltage distribution networks. In: IEEE. **2015 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)**. [S.l.], 2015. p. 1–5.
- OLIVEIRA, B. A. et al. Plataforma para visualização espacial de dados em redes de distribuição de energia elétrica. In: **IX Congresso de Inovação Tecnológica em Energia Elétrica (IX CITENEL)**. [S.l.: s.n.], 2017.
- OVERBYE, T. J.; WEBER, J. Smart grid wide-area transmission system visualization. **Engineering**, Elsevier, v. 1, n. 4, p. 466–474, 2015.
- PANAPAKIDIS, I. P.; MOSCHAKIS, M. N. Consumer load profile determination with entropy-based k-means algorithm. **International Journal of Energy and Power Engineering**, v. 13, n. 3, p. 144–149, 2019.
- PESSANHA, J. F. M.; SOUZA, R. C.; LAURENCEL, L. d. C. Um modelo de análise envoltória de dados para o estabelecimento de metas de continuidade do fornecimento de energia elétrica. **Pesquisa Operacional**, SciELO Brasil, v. 27, n. 1, p. 51–83, 2007.
- POURMIRZA, Z.; BROOKE, J. M. Monitoring and visualising a neighbourhood area sub-grid. In: **SMARTGREENS**. [S.l.: s.n.], 2013. p. 127–131.
- RAHMAH, N.; SITANGGANG, I. S. Determination of optimal epsilon (eps) value on dbscan algorithm to clustering data on peatland hotspots in sumatra. In: IOP PUBLISHING. **IOP Conference Series: Earth and Environmental Science**. [S.l.], 2016. v. 31, n. 1, p. 012012.
- ROBBI, C. Sistema para visualização de informações cartográficas para planejamento urbano. **São José dos Campos**, 2000.
- ROUSSEEUW, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. **Journal of computational and applied mathematics**, North-Holland, v. 20, p. 53–65, 1987.
- RUSSOM, P. et al. Big data analytics. **TDWI best practices report, fourth quarter**, v. 19, n. 4, p. 1–34, 2011.
- SCHUBERT, E. et al. Dbscan revisited, revisited: why and how you should (still) use dbscan. **ACM Transactions on Database Systems (TODS)**, ACM New York, NY, USA, v. 42, n. 3, p. 1–21, 2017.
- SCIKIT-LEARN DEVELOPERS. **Overview of clustering methods**. 2019. Acesso em: 31 mai. 2020. Disponível em: <<https://scikit-learn.org/stable/modules/clustering.html>>.
- SEDLMAIR, M.; MEYER, M.; MUNZNER, T. Design study methodology: Reflections from the trenches and the stacks. **IEEE transactions on visualization and computer graphics**, IEEE, v. 18, n. 12, p. 2431–2440, 2012.
- SEERA, M. et al. Power quality analysis using a hybrid model of the fuzzy min–max neural network and clustering tree. **IEEE transactions on neural networks and learning systems**, IEEE, v. 27, n. 12, p. 2760–2767, 2015.

SHULL, F. Getting an intuition for big data. **IEEE software**, IEEE, v. 30, n. 4, p. 3–6, 2013.

SILVA, G. A. Visualização científica de descargas elétricas atmosféricas. **Universidade Estadual Paulista (UNESP)**, 2014.

SRINIVASAN, D.; REINDL, T. et al. Real-time display of data from a smart-grid on geographical map using a gis tool and its role in optimization of game theory. In: IEEE. **2015 IEEE Innovative Smart Grid Technologies-Asia (ISGT ASIA)**. [S.l.], 2015. p. 1–6.

STADLER, J. G. et al. Improving the efficiency and ease of healthcare analysis through use of data visualization dashboards. **Big data**, Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA, v. 4, n. 2, p. 129–135, 2016.

SUCHARITHA, V.; SUBASH, S.; PRAKASH, P. Visualization of big data: its tools and challenges. **International Journal of Applied Engineering Research**, v. 9, n. 18, p. 5277–5290, 2014.

SUGINO, S. et al. Power-flow simulation with visualization function based on ieeec common data format. In: IEEE. **2016 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)**. [S.l.], 2016. p. 301–304.

USTARIZ-FARFAN, A. et al. Visualizing two-and three-dimensional maps for power quality loss assessment. In: IEEE. **2012 IEEE 15th International Conference on Harmonics and Quality of Power**. [S.l.], 2012. p. 909–914.

VANDERPLAS, J. **Python data science handbook: Essential tools for working with data**. [S.l.]: "O'Reilly Media, Inc.", 2016.

WOSNY, G. C. et al. **Proposta de base cartográfica para linhas de transmissão de energia elétrica**. 2012.

YANG, L.; GUO, Z. An integrated mis using interactive graphics interface for distribution system. In: IEEE. **2007 IEEE Power Engineering Society General Meeting**. [S.l.], 2007. p. 1–7.

YOUNG, R. R. Recommended requirements gathering practices. **CrossTalk**, v. 15, n. 4, p. 9–12, 2002.

ZAIDI, F. et al. Information management of a power distribution network in real time through gis mapping. In: IEEE. **2017 Australasian Universities Power Engineering Conference (AUPEC)**. [S.l.], 2017. p. 1–6.

ZUBELLI, F. S. **Métodos de Inteligência Computacional para Clusterização de Consumidores no Setor de Energia Elétrica**. Tese (Doutorado) — Universidade Federal do Rio de Janeiro, 2017.

## APÊNDICE A — QUESTIONÁRIO DE USABILIDADE

### A.1 *System Usability Scale - SUS*

1. Acho que gostaria de utilizar este sistema com frequência.
2. Considerei o sistema mais complexo do que o necessário.
3. Achei o sistema fácil de utilizar.
4. Acho que necessitaria de ajuda de um técnico para conseguir utilizar este sistema.
5. Considerei que as várias funcionalidades deste sistema estavam bem integradas.
6. Achei que este sistema tinha muitas inconsistências.
7. Suponho que a maioria das pessoas aprenderia a utilizar rapidamente este sistema.
8. Considerei o sistema muito complicado de utilizar.
9. Senti-me muito confiante ao utilizar este sistema.
10. Tive que aprender muito antes de conseguir lidar com este sistema.
11. Você tem alguma crítica e/ou sugestão referente ao sistema?