

UNIVERSIDADE FEDERAL DE ITAJUBÁ

PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA ELÉTRICA

Paulo Vitor Grillo de Souza

**Formação de Conjuntos de Sistemas de
Distribuição quanto aos Afundamentos de
Tensão**

Itajubá, Julho de 2016

UNIVERSIDADE FEDERAL DE ITAJUBÁ

PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA ELÉTRICA

Paulo Vitor Grillo de Souza

**Formação de Conjuntos de Sistemas de
Distribuição quanto aos Fundamentos de
Tensão**

Dissertação submetida ao Programa de Pós-Graduação em Engenharia Elétrica como parte dos requisitos para obtenção do título de **Mestre em Ciências em Engenharia Elétrica**.

Área de Concentração: Sistemas Elétricos de Potência

Orientador: Prof. Dr. José Maria de Carvalho Filho

Coorientador: Prof. Dr. Anderson Paulo de Paiva

Itajubá, Julho de 2016

DEDICATÓRIA

Aos meus pais, Paulo e Roseli.

“Cada sonho que você deixa para trás, é um pedaço do seu futuro que deixa de existir.”

Steve Jobs

AGRADECIMENTOS

A Deus, por todas as oportunidades que colocou em meu caminho, pela capacidade que me confiou, por me dar a saúde e força necessária para atingir os objetivos.

Aos meus pais, Paulo e Roseli, que me educaram com muito amor, me deram todo o apoio e depositaram confiança em mim para que eu atingisse este estágio da minha vida.

As minhas irmãs, Karina e Jessica, que apesar da distância, sempre me deram apoio.

A minha namorada, Fabiana, pelo companheirismo e compreensão.

Ao professor, José Maria, pela orientação e pela prontidão em ajudar.

Ao professor, Anderson Paulo de Paiva, pela coorientação, e auxílio em assuntos relacionados a estatística.

Ao professor Thiago Clé pelas sugestões criativas e revisões do trabalho.

A todos os mestres que me ensinaram e contribuíram com a minha evolução ao longo da minha vida.

Aos meus amigos que proporcionaram momentos de descontração essenciais.

À CAPES, ao CERIn e à UNIFEI pelo apoio financeiro e estrutural na pesquisa. Essa experiência foi muito especial para mim e não sairá da memória.

SUMÁRIO

DEDICATÓRIA.....	I
AGRADECIMENTOS.....	III
SUMÁRIO.....	IV
LISTA DE FIGURAS.....	VII
LISTA DE TABELAS.....	IX
LISTA DE ABREVIATURAS E SIGLAS.....	XI
RESUMO.....	XII
ABSTRACT.....	XIII
1. INTRODUÇÃO	1
1.1 Relevância do Tema.....	1
1.2 Estado da Arte.....	2
1.3 Objetivos e Contribuições da Dissertação	6
1.4 Estrutura da Dissertação	7
2. VARIACÕES DE TENSÃO DE CURTA DURAÇÃO	8
2.1 Considerações Iniciais	8
2.2 Definição de Afundamento de Tensão.....	8
2.3 Causas e Efeitos	10
2.4 Variáveis de Influência dos Parâmetros de AMT	11
2.4.1 Localização da Falta	11
2.4.2 Tipo de Falta.....	14
2.4.3 Taxas de Falta.....	15
2.4.4 Impedância de Falta.....	16
2.4.5 Tensão Pré-Falta.....	17
2.4.6 Conexão dos Transformadores	17
2.4.7 Sistemas de Proteção	18
2.4.8 Sistemas de Religamento.....	19
2.5 Indicadores.....	19
2.6 Considerações Finais	22
3. FUNDAMENTOS DA FORMAÇÃO DE CONJUNTOS.....	23
3.1 Considerações Iniciais	23
3.2 Medidas de Similaridade e Dissimilaridade	24
3.3 Métodos para Formação de Conjuntos (Clusters).....	25
3.3.1 Técnicas Hierárquicas Aglomerativas	26

3.3.2	Técnicas de Agrupamento Não Hierárquicas	31
3.4	Considerações Finais	33
4.	PROPOSTA DE METODOLOGIA PARA FORMAÇÃO DE CONJUNTOS	
	QUANTO AOS AFUNDAMENTOS DE TENSÃO	34
4.1	Considerações Iniciais	34
4.2	Resumo Passo a Passo da Metodologia	34
4.3	Escolha das Variáveis	34
4.4	Seleção das Variáveis	35
4.4.1	Coefficiente de Correlação Linear de Pearson.....	36
4.5	Implementação da Formação de Conjuntos.....	40
4.6	Comparação entre Métodos	40
4.6.1	Análise de Variância.....	41
4.7	Análise de Robustez dos Métodos	44
4.8	Considerações Finais	44
5.	ESTUDO DE CASO	45
5.1	Considerações Iniciais	45
5.2	Sistemas de Distribuição Hipotéticos	45
5.2.1	Escolha das Variáveis	46
5.2.2	Seleção das Variáveis	47
5.2.3	Implementação da Formação de Conjuntos.....	48
5.2.4	Análise de Robustez	51
5.3	Sistemas de Distribuição Reais (EDP Escelsa).....	53
5.3.1	Escolha das Variáveis	55
5.3.2	Seleção das Variáveis	55
5.3.3	Implementação da Formação de Conjuntos.....	56
5.3.4	Comparação entre Métodos	59
5.3.5	Análise de Robustez	61
5.4	Considerações Finais	64
6.	CONCLUSÕES.....	65
6.1	Conclusões Gerais.....	65
6.2	Proposta para Trabalhos Futuros	66
6.3	Artigos Publicados	66
	REFERÊNCIAS	68
	ANEXO A – Resolução de Exemplo de Formação de Conjuntos	74

A.1	Método de Ligação Simples (<i>Single Linkage</i>).....	74
A.2	Método de Ligação Completa (<i>Complete Linkage</i>).....	76
A.3	Método da Média das Distâncias (<i>Average Linkage</i>).....	78
A.4	Método de Ward	80
A.5	Método das k-médias (<i>k-means</i>).....	82
A.6	Comparação entre Métodos de Formação de Conjuntos	86
ANEXO B – 12 Sistemas de Distribuição		90
ANEXO C – Programa em Matlab		102

LISTA DE FIGURAS

Figura 2.1 - Exemplo de afundamento de tensão.....	8
Figura 2.2 - Área de vulnerabilidade para nó k de um sistema (KEMPNER, 2012).	12
Figura 2.3 - Método da distância crítica em sistema radial.	13
Figura 3.1 - Ilustração das técnicas hierárquicas aglomerativas e divisivas	26
Figura 3.2 - Modelo de um dendograma.	26
Figura 3.3 - Ilustração gráfica do método de ligação simples.	28
Figura 3.4 - Ilustração gráfica do método de ligação completa.....	29
Figura 3.5 - Ilustração gráfica do método da média das distâncias.	29
Figura 4.1 - Correlação linear positiva perfeita entre as variáveis X e Y (LIRA, 2004).....	37
Figura 4.2 - Correlação linear nula entre as variáveis X e Y (LIRA, 2004).....	37
Figura 4.3 - Correlação linear negativa perfeita entre as variáveis X e Y (LIRA, 2004).....	38
Figura 4.4 - Correlação não-linear entre as variáveis X e Y (LIRA, 2004).....	38
Figura 5.1 - Dendograma 36 sistemas (Ligação Simples).....	48
Figura 5.2 - Dendograma 36 sistemas (Ligação Completa).	49
Figura 5.3 - Dendograma 36 sistemas (Média das Distâncias).	49
Figura 5.4 - Dendograma 36 sistemas (Ward).....	50
Figura 5.5 - Resultado da análise de concordância para os 36 sistemas de distribuição.....	52
Figura 5.6 - Intervalo de confiança dos grupos quanto aos AMTs (Média das Distâncias)....	53
Figura 5.7 - Distribuição geográfica dos sistemas de distribuição.	54
Figura 5.8 - Dendograma Escelsa (Ligação Simples).	57
Figura 5.9 - Dendograma Escelsa (Ligação Completa).....	57
Figura 5.10 - Dendograma Escelsa (Média das Distâncias).....	58
Figura 5.11 - Dendograma Escelsa (Ward).	58
Figura 5.12 - Análise de variância para o método de ligação simples (Escelsa).....	60
Figura 5.13 - Análise de variância para o método de ligação completa (Escelsa).	60
Figura 5.14 - Análise de variância para o método da média das distâncias (Escelsa).....	60
Figura 5.15 - Análise de variância para o método de ward (Escelsa).	61
Figura 5.16 - Análise de variância para o método de <i>k-means</i> (Escelsa).	61
Figura 5.17 - Resultado da análise de concordância para os sistemas da Escelsa.....	62

Figura 5.18 - Intervalo de confiança dos grupos quanto aos AMTs (Ward).....	63
Figura 5.19 - Gráfico tridimensional (Ward).	64
Figura A.1 - Dendograma considerando o método de ligação simples.....	76
Figura A.2 - Dendograma considerando o método de ligação completa.	78
Figura A.3 - Dendograma considerando o método da médias das distâncias.	80
Figura A.4 - Dendograma considerando o método de Ward.....	82
Figura A.5 - Saída do método de <i>k-means</i>	85
Figura A.6 - Comparação entre métodos de formação de conjuntos.....	87
Figura A.7 - Comparação entre métodos (exemplo completo).	89

LISTA DE TABELAS

Tabela 2.1 - Classificação das Variações de Tensão de Curta Duração.....	9
Tabela 2.2 - Porcentagem de faltas para cada tipo de curto-circuito.....	14
Tabela 2.3 - Número de ocorrências percentual por tipo de falta.....	15
Tabela 2.4 - Taxas de faltas por 100 km/ano (GARCIA;ALVES, 2011).....	15
Tabela 2.5 - Taxas de faltas por 100 km/ano (CONRAD et al., 1991) e (BECKER et al., 1994).....	15
Tabela 2.6 - Número de Ocorrências por 100 km/ano (FONSECA, 1999).....	15
Tabela 2.7 - Taxa de falta para linhas de transmissão (*adotadas pela CELG) (RAMOS, 2009).....	16
Tabela 2.8 - Quantificação dos afundamentos de tensão NRS 048-2.	19
Tabela 2.9 - Estratificação das VTCDs com base nos níveis de sensibilidade das diversas cargas.	20
Tabela 2.10 - Fatores de ponderação e Fator de Impacto Base de acordo com a tensão nominal do barramento de distribuição.	21
Tabela 4.1 - Valores das variáveis X e Y.	39
Tabela 4.2 - Dados das amostras A, B e C	41
Tabela 4.3 - Variação total	42
Tabela 4.4 - Variação dentro das amostras.....	42
Tabela 4.5 - Variação entre amostras	42
Tabela 5.1 - Características dos sistemas de distribuição.....	45
Tabela 5.2 - Base de dados dos 36 sistemas de distribuição.	46
Tabela 5.3 - Correlação de Pearson entre variáveis e número médio de AMTs (36 sistemas).47	
Tabela 5.4 - Comparação <i>membership</i> (36 sistemas)	50
Tabela 5.5 - Base de dados da EDP Escelsa.....	55
Tabela 5.6 - Correlação de Pearson entre variáveis e número médio de AMTs (Escelsa).....	55
Tabela 5.7 - Comparação <i>membership</i> (Escelsa).....	59
Tabela A.1 - Medições das variáveis.....	74
Tabela A.2 - Centróides dos grupos (14) e (235).....	82
Tabela A.3 - Centróides dos grupos (124) e (35).	83

Tabela A.4 - Centróides dos grupos (12) e (345).	84
Tabela A.5 - Centróides dos grupos (123) e (45).	84
Tabela A.6 - Distância ao quadrado dos elementos para os grupos finais	85
Tabela A.7 - Classificação das companhias em dois grupos.	86
Tabela A.8 - Dados de 22 companhias e 8 variáveis.....	88

LISTA DE ABREVIATURAS E SIGLAS

ABRACE	Associação Brasileira de Grandes Consumidores Industriais de Energia e de Consumidores Livres
ABRADEE	Associação Brasileira de Distribuidores de Energia Elétrica
AMT	Afundamento Momentâneo de Tensão
ANEEL	Agência Nacional de Energia Elétrica
AP	Audiência Pública
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CBEMA	<i>Computer and Business Equipment Manufacturers Association</i>
CERIn	Centro de Excelência em Redes Elétricas Inteligentes
CP	Consulta Pública
EDP	Energias de Portugal
FAU	Fundação de Apoio Universitário
FUPAI	Fundação de Pesquisa e Assessoramento à Indústria
IEC	<i>International Electrotechnical Commission</i>
IEEE	<i>Institute of Electrical and Electronics Engineers</i>
ITIC	<i>Information Technology Industry Council</i>
ONS	Operador Nacional do Sistema Elétrico
PCA	<i>Principal Component Analysis</i>
P&D	Pesquisa e Desenvolvimento
PRODIST	Procedimentos de Distribuição
p.u.	por unidade
QEE	Qualidade da Energia Elétrica
RMS	<i>Root Mean Square</i> (valor eficaz)
SED	Subestação de Distribuição
SEMI	<i>Semiconductor Equipment and Materials Institute</i>
SRD	Superintendência de Regulação dos Serviços de Distribuição
UNIFEI	Universidade Federal de Itajubá
VTCD	Variação de Tensão de Curta Duração

RESUMO

Esta dissertação apresenta uma proposta de metodologia para formação de conjuntos de sistemas de distribuição quanto aos afundamentos momentâneos de tensão. O trabalho é focado em determinar qual o melhor método de formação de conjuntos a ser empregado em uma base de dados que considere as principais variáveis relacionadas com o número de afundamentos de tensão.

Para testar a metodologia proposta, foram utilizados dados reais de uma concessionária do sistema elétrico brasileiro, a Escelsa do grupo EDP. Foram monitoradas quinze barras de quatorze subestações da Escelsa. Para todos os casos, são simulados tanto os sistemas de transmissão quanto o de distribuição, assim como os sistemas de proteção envolvidos.

Para se ter segurança dos resultados, a metodologia foi criada tendo como base sistemas de distribuição hipotéticos, baseados no caso teste do IEEE 34 barras, onde há completa confiança nas variáveis utilizadas, e então foi aplicada a título de validação sobre dados reais.

Este estudo apresenta uma metodologia geral que pode ser estendida e aplicada em todos os sistemas de distribuição do Brasil, podendo servir de guia orientativo para o órgão regulamentador ANEEL, para estabelecer padrões de referência para o fenômeno. Além disso, destaca-se que o objetivo é mostrar através dos resultados que, para cada base de dados, há um método que melhor se ajusta e é este método que o procedimento criado se compromete a encontrar.

Palavras-Chave: Afundamento de tensão, Qualidade do produto, Formação de Conjuntos.

ABSTRACT

This master's thesis proposes a methodology for clustering distribution systems regarding voltage sags. The work is focused on determining the best clustering method to be used in a database that considers the main variables related to the number of voltage sags.

Aiming at testing the proposed methodology, real data from a Brazilian electric utility have been used, namely Escelsa owned by the EDP Group. Fifteen buses of fourteen Escelsa's substations were monitored. In all cases, the transmission and distribution systems are simulated, as well as the protection systems.

To have security in the results, the methodology was created based on hypothetical distribution systems, based on the IEEE test case 34 bars, where there is complete confidence in the used variables, and then it was applied to validation on real data.

This study presents a general methodology that can be extended and applied to the setting of all the national grid distribution systems and can serve as a guideline to the regulator ANEEL, to set benchmarks for the phenomenon. In addition, the goal is to show through the results for each database, there is a method that best fits and is this method that the established procedure undertakes to find.

Keywords: Voltage sag, Product quality, Clustering.

1. INTRODUÇÃO

1.1 Relevância do Tema

Com o desenvolvimento de novas tecnologias, que na maioria dos casos utiliza dispositivos eletrônicos, as cargas eletroeletrônicas estão prevalecendo nas indústrias em substituição às antigas cargas eletromecânicas. Esta transição permitiu às indústrias promover aumentos na produção e na eficiência. Porém, os equipamentos modernos em geral, possuem maior sensibilidade frente aos distúrbios de qualidade de energia elétrica, principalmente, aqueles relacionados às VTCDs (variações de tensão de curta duração).

Simplificadamente, quando ocorre VTCDs na rede elétrica algumas máquinas da planta industrial podem passar a funcionar fora dos padrões necessários para obtenção de um bom produto, ou até mesmo parar por completo o funcionamento, gerando, respectivamente, uma interrupção parcial ou total do processo. Independentemente do tipo de interrupção ocorrida no processo, há sempre atrelado a ela os prejuízos devido à perda de produtividade, perda de insumos, reparo e reposição de equipamentos danificados (CARVALHO FILHO, 2000).

Embora as VTCDs tenham grande impacto econômico negativo para as empresas, o (PRODIST, 2014) (Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional) elaborado pela ANEEL (Agência Nacional de Energia Elétrica), ainda não estabelece limites para este evento. Sendo assim, as concessionárias de energia elétrica ainda não são penalizadas se o consumidor industrial sofre paradas de processo.

Diante dos prejuízos que os consumidores industriais vêm sofrendo pela ocorrência desses eventos, estes estão pressionando as concessionárias de energia elétrica e principalmente a ANEEL para que se estabeleçam limites para o distúrbio, pois somente após a regulamentação pelo órgão competente, será possível cobrar das concessionárias de energia elétrica padrões mínimos da qualidade e induzi-las a promover mais investimentos na rede de distribuição.

Apesar de não estar em vigência, está em vias de vigorar uma nova revisão do PRODIST em que são estabelecidos indicadores e limites para as VTCDs. Tal alteração foi embasada em uma consultoria contratada pela SRD (Superintendência de Regulação dos Serviços de Distribuição) em 2013 e realizada pela FAU (Fundação de Apoio Universitário) e pelas contribuições recebidas no âmbito da Consulta Pública (CP N°18/2014). Para o fenômeno de VTCDs, a proposta consiste em um indicador único, denominado FI (Fator de Impacto). O limite proposto pela consultoria, foi o ponto de maior discussão na CP n°18/2014, sendo que a

maioria das contribuições foram contrárias ao estabelecimento de limites, pois alegaram que os dados de medição utilizados não são representativos, pois não houve tempo de medição suficiente para considerar a sazonalidade e não foram considerados dados de diferentes concessionárias. A ABRACE (Associação Brasileira de Grandes Consumidores Industriais de Energia e de Consumidores Livres) acenou que os limites propostos são muito conservadores não refletindo a qualidade do produto necessária a indústria. Já a ABRADDEE (Associação Brasileira de Distribuidores de Energia Elétrica) defende que deve haver um avanço no tema por meio da realização de medições permanentes nas SEDs (Subestações de Distribuição) franqueadas pelas distribuidoras, pois esta proposta garantirá um tempo de maturação e trará uma evolução gradual e equilibrada do tema no setor de distribuição. Apesar das controvérsias nas contribuições recebidas, a SRD, através da Audiência Pública (AP N°82/2015), deixa claro que é necessário definir limites para se aprimorar o regulamento e ressalta que aprimoramentos sempre serão necessários, mas a busca pela perfeição não deve inibir o avanço do assunto.

O grande problema é como estabelecer limites adequados para as VTCDs, sabendo da grande extensão territorial do Brasil, da consequente diversidade de relevo, vegetação e clima das regiões e do número de concessionárias presentes no sistema elétrico. Cabe algumas indagações, por exemplo: é viável estabelecer apenas um valor limite, que todos os sistemas de distribuição devem respeitar, ou é mais plausível estabelecer valores limites para agrupamentos de sistemas de distribuição que possuam características semelhantes do ponto de vista das variáveis que influenciam a ocorrência de VTCDs? A segunda proposição parece ser a mais coerente, mas para sua viabilidade, primeiramente é necessário formar os agrupamentos de sistemas de distribuição, a exemplo do que se faz na área de continuidade. É neste contexto que esta dissertação se insere, propondo uma metodologia para criação de conjuntos de sistemas de distribuição semelhantes quanto aos a fundamentos momentâneos de tensão.

1.2 Estado da Arte

Feito um levantamento em diversos acervos, foram encontrados artigos que utilizam algumas técnicas de análise multivariada para caracterizar os fenômenos relacionados a qualidade da energia elétrica. Depois de uma rígida seleção, dos artigos que mais estavam relacionados com o tema desta dissertação, foi feito um breve resumo de cada um destes trabalhos, de forma a apresentar como as técnicas estatísticas de análise multivariada estão sendo empregadas na área de qualidade da energia elétrica.

No trabalho de (LIRA et al., 2006), foi desenvolvido um algoritmo capaz de fazer uma

classificação automática da forma de onda do distúrbio, através de redes neurais artificiais. Oscilografias reais de tensão foram aplicadas no processo de construção do algoritmo. Os sinais são processados em dois estágios: primeiro, há a decomposição pela transformada wavelet até o quinto nível de decomposição, depois, os coeficientes resultantes da wavelet são processados pela PCA (*Principal Component Analysis*), reduzindo os elementos de entrada do classificador para uma dimensão muito inferior. A classificação é realizada usando uma combinação de 3 MLPs (*Multilayer Perceptron*) com arquiteturas diferentes. O algoritmo RPROP (*Resilient Backpropagation*) é aplicado para treinar a rede. São formadas as combinações nas redes e a decisão final do algoritmo corresponde à combinação de saída com maior valor. Os resultados mostraram-se bastante promissores para cinco tipos de perturbações testados: afundamentos, elevações, harmônicas, transitórios oscilatórios e interrupções, além do caso particular do sistema sem distúrbio.

Em (MELENDEZ; COLOMER, 2008), propõe-se um método para classificar afundamentos de tensão registrados em subestações de distribuição baseado em estatística. O objetivo é associar a forma de onda dos afundamentos de tensão com sua origem na rede (média tensão-MT ou alta tensão-AT). MPCA (*Multiway Principal Component Analysis*) foi usada como estratégia para redução de dimensão, enquanto que CBR (*Case Based Reasoning*) foi aplicada neste espaço de projeção para identificar eventos previamente registrados e diagnosticados. A capacidade do método proposto foi demonstrada com o uso de dados de um sistema real composto por cinco subestações do sistema elétrico da Catalonia (Espanha).

O artigo de (HUANG; LIN, 2010) apresenta uma nova abordagem para o reconhecimento dos sinais dos distúrbios de qualidade da energia através do uso de análise da componente principal e redes neurais. O algoritmo identifica a maioria dos distúrbios de qualidade da energia, como afundamentos, elevações, interrupções, transitórios, harmônicos e flutuação. O novo modelo incluiu principalmente três passos. Primeiramente, a transformada S é usada para analisar o sinal dos distúrbios, e 18 atributos distintos são extraídos desta transformação. No segundo estágio, a análise de componente principal é usada para reduzir a dimensão do conjunto de atributos. Finalmente, as componentes principais são usadas como vetores de entrada da rede neural que é responsável por classificar o sinal do distúrbio. Os resultados das simulações mostraram a validade e a eficiência do modelo sugerido.

O trabalho de (POZZEBON et al., 2010) propõe um método diferente de classificação de distúrbio de qualidade, combinando transformada discreta de *wavelet* (DWT), (PCA) e redes neurais. Esse método associa propriedades da técnica (MRA) da análise multiresolução com

desvio-padrão e o cálculo de média para extrair as características discriminadoras de sinais distorcidos em níveis diferentes de resolução. Posteriormente, um algoritmo PCA é usado para reduzir a dimensão do espaço das características pelo mapeamento do conjunto das características obtidas em um conjunto menor de elementos independentes. Então, uma rede de função de base radial (RBF) é empregada para executar a classificação das perturbações.

Outro trabalho que utiliza diversas técnicas estatísticas para análise dos afundamentos de tensão é (BARRERA; PAVAS; MELÉNDEZ, 2011), cujo artigo mostra uma análise da qualidade da energia elétrica da rede de distribuição de Bogotá usando quatro métodos estatísticos diferentes, sendo dois deles qualitativos e os outros dois quantitativos. Os dados foram coletados durante dois anos pela CODENSA. As estratégias quantitativas são baseadas no estudo da função densidade de probabilidade bem como na (PCA). A função densidade de probabilidade da magnitude dos afundamentos de tensão foi utilizada para caracterizar melhor sua ocorrência, e a técnica de PCA foi usada para melhorar a análise dos AMTs. Por outro lado, a abordagem qualitativa é focada em dois índices relacionados aos afundamentos: o SAI (*sag activity index*) e a quantidade da energia não suprida quando o distúrbio ocorre. Estes métodos foram aplicados para avaliar a qualidade de energia de 249 subestações. Como resultado, observa-se que as estratégias quantitativas permitem distinguir subestações com níveis de afundamentos de tensão significativamente diferentes em relação as outras subestações incluídas na análise. As estratégias qualitativas calculam um número real que representa a severidade dos níveis de afundamento nas subestações. Frequência e função de distribuição cumulativa são técnicas estatísticas úteis para a análise da qualidade da energia frente a AMTs, pois permite identificar subestações que possuem diferentes perfis de afundamentos de tensão.

É apresentado por (ROMERO; GALLEGO; PAVAS, 2012) uma nova metodologia para estimação de padrões de afundamento de tensão e agrupamento das zonas de falta em sistemas de alta e média tensão. A metodologia proposta compreende três estágios. Primeiro, modelagem da rede e simulação de faltas foram realizadas a fim de obter informações sobre os afundamentos de tensão causados por falhas no sistema de transmissão. Os padrões para AMTs foram identificados no segundo estágio, pelas médias do algoritmo de formação de conjuntos k-médias, permitindo a determinação das zonas de faltas. Usando a base de dados de medições de QEE de uma grande companhia de eletricidade de Bogotá, os AMTs foram classificados de acordo com os padrões previamente estabelecidos. No terceiro estágio da metodologia uma comparação entre simulação e medição foi realizada, permitindo a identificação dos afundamentos causados por faltas.

Foi desenvolvido por (BALOUJI; SALOR, 2014) um método para o agrupamento e classificação dos eventos de qualidade da energia, baseado na análise de autovalores. Pretende-se que tal método sirva para suprir as necessidades das aplicações futuras das redes elétricas inteligentes. O procedimento é fundamentado na técnica de agrupamento k-médias, que usa os valores de tensão RMS, como sugerido pela norma IEC 61000-4-30. Para reduzir o número de conjuntos formados e para aumentar a eficiência do agrupamento, foram aplicadas duas *Eingen-Analysis*, análise de componentes principais (PCA) e análise linear de discriminante (LDA) antes do processo de formação de conjuntos pelo método das k-médias.

Afundamento de tensão é um fenômeno típico de qualidade de energia, e identificar a fonte causadora deste distúrbio é importante para o monitoramento e gerenciamento dos afundamentos. Devido à correlatividade e à redundância das características, os métodos de identificação das fontes dos distúrbios são pouco precisos. Para resolver este problema, (WANG; WANG; JIA, 2014) propõe um método para redução das características dos eventos de afundamentos de tensão, baseado em análise de componentes principais (PCA). Foram analisadas as oscilografias de eventos de afundamentos de tensão causadas por faltas no sistema, partidas de motores de indução, energização de transformadores e três tipos de eventos resultantes do efeito combinado de fontes de AMTs. Os índices das características relacionadas aos afundamentos de tensão foram reduzidos através de PCA, que é uma técnica que permite transformar o espaço das características de muitas dimensões para poucas dimensões, mantendo a informação das características e eliminando as correlações entre as características. Os resultados de simulação mostraram que o método proposto reduz efetivamente o vetor de características iniciais, e a precisão da classificação e identificação das fontes causadoras de AMTs obtida com este método é melhor quando comparada aos métodos que não utilizam a técnica PCA.

Outra pesquisa relacionada com o tema em questão foi realizada por (HARIYANTO; ANGGORO; NOEGROHO, 2014), onde foi introduzido uma nova maneira de identificar o perfil de severidade da tensão de um barramento devido a curto-circuito ocorrido em determinado ponto do sistema de distribuição de energia. Um curto-circuito causa um afundamento na tensão por um período relacionado ao tempo de operação da proteção. Os dados contendo as duas variáveis (amplitude e duração) do afundamento de tensão devido ao curto-circuito foram gerados para cada barramento do sistema. Posteriormente, estes dados de todas as barras foram separados em conjuntos pelo método de *k-means*. E este processo produzirá barramentos de referência (que serão os centros dos conjuntos) e mostrará os barramentos

associados a cada referência. Para ser capaz de perceber a severidade do afundamento de tensão, os centros dos conjuntos serão convertidos para o índice de severidade do evento, que explica a severidade do AMT, com base na curva CBEMA-ITIC. Dados de um determinado barramento que está sujeito a eventos de afundamento de tensão serão classificados com base no conjunto que está associado. Assim, será obtido a frequência de eventos que são classificados em conjuntos particulares. Para melhor observar os dados recomenda-se o uso de histogramas.

O artigo mais correlato com o assunto desta dissertação foi publicado recentemente por (FILHO et al., 2016), onde os autores propõem a aplicação da análise de componentes principais para reduzir a dimensionalidade de um conjunto de 32 variáveis em 7 componentes principais que trazem praticamente toda a informação das variáveis originais, e a partir destas empregam a técnica de formação de conjuntos considerando o método de ligação de Ward e distância euclidiana para formar dois conjuntos. Os conjuntos obtidos foram satisfatórios quando analisados por um engenheiro que possui profundo conhecimento dos sistemas de distribuição considerados. Apesar do artigo apresentar uma comparação do método utilizado para formar conjuntos, não foi provocada uma perturbação nas variáveis para verificar a robustez dos métodos. Esta análise citada foi realizada nesta dissertação para complementar o trabalho.

1.3 Objetivos e Contribuições da Dissertação

O objetivo deste trabalho é propor uma metodologia capaz de identificar qual o melhor método de formação de conjuntos para se aplicar na base de dados considerada e assim permitir formar conjuntos robustos e confiáveis. O procedimento inicia-se com uma seleção de variáveis, seguida da implementação da divisão em conjuntos realizada pelos métodos de ligação simples, completa, média das distâncias, Ward e k-médias, passando por uma comparação estatística entre os métodos e encerrando com uma análise da robustez.

A grande contribuição da dissertação é propor e apresentar, uma possível metodologia de formação de conjuntos de sistemas de distribuição, capaz de ser estendida e aplicada em uma base de dados que contemple todo o sistema elétrico brasileiro. Com base nestes conjuntos será possível estabelecer limites apropriados que servirão de referência para as concessionárias e consumidores e poderão balizar eventuais penalizações por parte da ANEEL, caso tais limites sejam infringidos.

1.4 Estrutura da Dissertação

Este trabalho está dividido em seis capítulos, e a descrição de cada um deles é apresentada nos próximos parágrafos.

No presente capítulo 1 encontra-se a relevância do tema, o estado da arte mostrando como as técnicas de análise multivariada foram utilizadas nos trabalhos mais correlatos ao tema, o objetivo, as contribuições geradas pela proposta da dissertação e a estrutura do trabalho.

No capítulo 2, é feita uma revisão teórica sobre variações de tensão de curta duração, apresentando definições, causas, efeitos, variáveis de influência e alguns indicadores.

No capítulo 3, é feita uma revisão teórica dos principais métodos de formação de conjuntos, e para isto é mostrada a forma de calcular algumas medidas de similaridade usuais e os algoritmos dos métodos, seguido de exemplo de aplicação e comparações entre métodos.

No capítulo 4, apresenta-se a metodologia proposta para formação de conjuntos de sistemas de distribuição quanto aos fundamentos de tensão.

O capítulo 5, exhibe a aplicação da metodologia proposta em dados de sistemas de distribuição criados com base no caso teste do IEEE de 34 barras, bem como sua aplicação em 15 sistemas de distribuição reais da EDP Escelsa.

O capítulo 6 é destinado para as conclusões da dissertação, apresentação de propostas para trabalhos futuros e apresentação dos artigos publicados.

Finalmente, estão as referências utilizadas como base para realização deste trabalho e os anexos. O anexo A apresenta a resolução de um exemplo de formação de conjuntos utilizando os métodos considerados nesta dissertação, o anexo B mostra os parâmetros e os diagramas unifilares dos 12 sistemas de distribuição criados a partir do caso teste do IEEE de 34 barras e o anexo C apresenta uma rotina feita no *software* MATLAB para gerar 100 cenários de ruído em uma variável e formar conjuntos referentes a estes cenários considerando diversos métodos.

2. VARIAÇÕES DE TENSÃO DE CURTA DURAÇÃO

2.1 Considerações Iniciais

Este capítulo apresenta os conceitos básicos sobre variações de tensão de curta duração, mais especificamente sobre afundamentos de tensão considerando as definições, as causas, os efeitos sobre processos industriais, as variáveis de influência sobre seus parâmetros e algumas propostas de indicadores.

2.2 Definição de Afundamento de Tensão

Simplificadamente, o afundamento de tensão é um fenômeno onde o valor eficaz da tensão reduz em um curto intervalo de tempo e logo em seguida se recompõe, conforme pode ser observado na Figura 2.1.

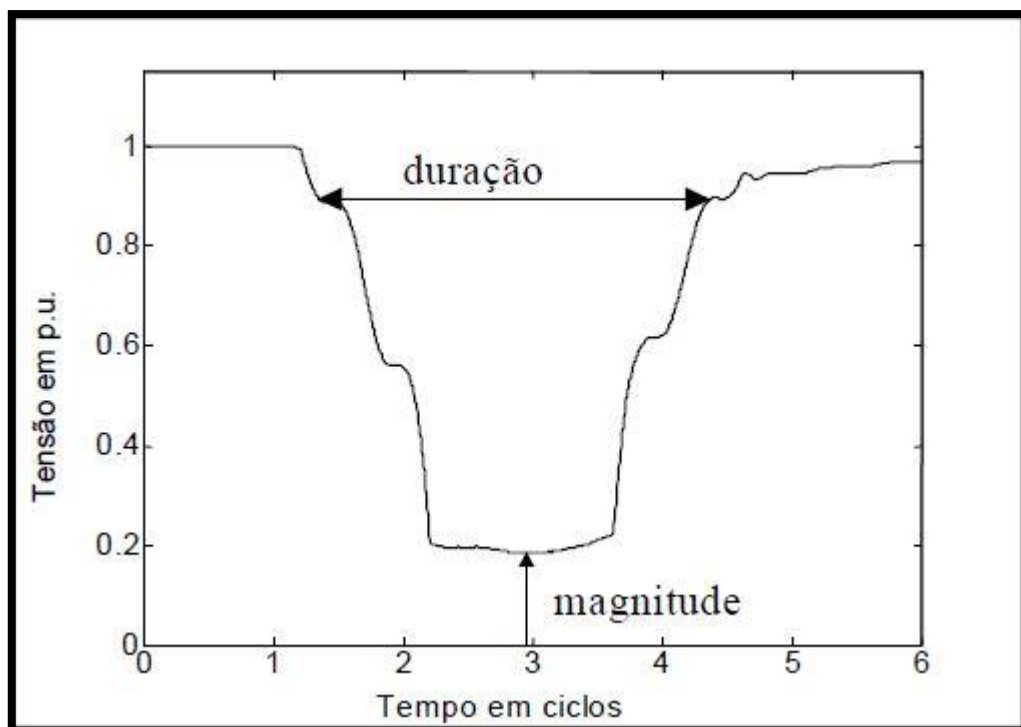


Figura 2.1 – Exemplo de afundamento de tensão.

Como se observa na Tabela 2.1, o afundamento de tensão se insere em uma das categorias de VTCDs e de acordo com sua duração pode ser classificado como sendo momentâneo ou temporário.

Tabela 2.1 – Classificação das Variações de Tensão de Curta Duração (PRODIST, 2014).

Classificação	Denominação	Duração da Variação	Amplitude da tensão (valor eficaz) em relação à tensão de referência
Variação Momentânea de Tensão	Interrupção Momentânea de Tensão	Inferior ou igual a três segundos	Inferior a 0,1 p.u
	Afundamento Momentâneo de Tensão	Superior ou igual a um ciclo e inferior ou igual a três segundos	Superior a 0,1 p.u e inferior a 0,9 p.u
	Elevação Momentânea de Tensão	Superior ou igual a um ciclo e inferior ou igual a três segundos	Superior a 1,1 p.u
Variação Temporária de Tensão	Interrupção Temporária de Tensão	Superior a três segundos e inferior a três minutos	Inferior a 0,1 p.u
	Afundamento Temporária de Tensão	Superior a três segundos e inferior a três minutos	Superior a 0,1 p.u e inferior a 0,9 p.u
	Elevação Temporária de Tensão	Superior a três segundos e inferior a três minutos	Superior a 1,1 p.u

Porém, este evento tem definições ligeiramente diferentes nas normas existentes, por isto serão mostrados as definições propostas pelo IEEE, IEC e PRODIST.

A norma (IEEE 1159, 2009) define afundamento de tensão ou *voltage sag* como um decréscimo da tensão para um patamar entre 0,1 e 0,9 p.u. do valor eficaz da tensão nominal, com duração entre 0,5 ciclo e 1 minuto. Eventos cuja tensão remanescente seja inferior a 0,10 p.u. é considerado pelo IEEE como uma interrupção.

Em normas antigas da IEC, este evento era denominado como *voltage dip* e se referia à queda do valor RMS da tensão. Na norma (IEEE 493, 2007), é relatado que a (IEC 61000-4-30, 2006) passou a definir *voltage dip* como tensão remanescente, e que *dip* e *sag* passaram a ser sinônimos entre a IEC e o IEEE.

No Brasil, de acordo com os Procedimentos de Rede elaborados pelo (ONS, 2011), no item gerenciamento dos indicadores de desempenho da rede básica e de seus componentes, é definida VTCD como um evento aleatório de tensão caracterizado por desvio significativo, por curto intervalo de tempo, do valor eficaz da tensão.

Nos Procedimentos de Rede do ONS a classificação não se diferencia do PRODIST, exceto pelo limite superior do intervalo de duração da variação temporária de tensão que é de 1 minuto.

2.3 Causas e Efeitos

Os afundamentos de tensão no sistema elétrico são provenientes da energização de transformadores, ocorrência de curtos-circuitos (AYELLO et al., 1999), (CONRAD; LITTLE; GRIGG, 1991), (ORTMEYER; HIYAMA; SALEHFAR, 1996), partida de motores de grande porte (BOLLEN, 1994), e pela saída de grandes blocos de geração (STYVAKTAKIS; BOLLEN; GU, 2002) e (STYVAKTAKIS, 2002). No entanto, a causa mais recorrente são os curtos-circuitos, sobretudo na configuração das redes elétricas instaladas no Brasil, que possuem grandes extensões de linhas aéreas que estão sujeitas a intempéries. Nos sistemas de distribuição, há um agravante do problema, devido à ausência de cabos guarda para proteção da rede contra descargas atmosféricas. Por este motivo, grande parte dos trabalhos publicados na literatura referentes a afundamentos de tensão têm como foco o sistema de distribuição.

A ocorrência de curtos-circuitos em linhas aéreas deve-se em sua maioria à incidência de descargas atmosféricas, o que pode confirmar que a ocorrência de afundamentos de tensão está relacionada ao nível cerâmico da região onde as linhas estão instaladas. É importante salientar que nem todas as descargas atmosféricas resultam em curtos-circuitos e, conseqüentemente, em afundamentos de tensão. Além disso, a maioria dos sistemas elétricos são projetados para manter operação normal para perto de 95% das sobretensões de origem atmosférica (LEBORGNE, 2003).

O efeito dos afundamentos de tensão está diretamente relacionado com a sensibilidade dos equipamentos eletrônicos instalados nos consumidores, das especificidades de cada processo industrial e também dos sistemas de controle de processo envolvido.

Segundo (HONORATO, 2014), os principais efeitos dos afundamentos de tensão sob os equipamentos utilizados nas indústrias se manifestam por:

- Perda de programação de microprocessadores e de PLCs (circuito lógico programável);
- Desatracamento das bobinas de contadores e relés auxiliares, com conseqüentes desligamentos de cargas e equipamentos via lógica do sistema de controle;
- Desligamento de lâmpadas de descarga, como as de vapor de mercúrio, que levam cerca de alguns minutos para reacenderem;
- Variação de velocidade dos acionamentos CA (corrente alternada) e CC (corrente contínua), que dependendo do tipo de processo, poderá comprometer a qualidade do produto ou até provocar a parada de produção;

- Variação de torque do motor (CA e CC) com as mesmas implicações citadas anteriormente;
- Desligamento de acionamentos devido à atuação de dispositivos de proteção associados, que quando detectam condições de risco, promovem o bloqueio do disparo de tiristores ou até mesmo o desligamento imediato da fonte de alimentação;
- Falhas de comutação em pontes controladas, afetando os disparos dos tiristores;
- Queima de fusíveis e outros componentes, principalmente, nos acionamentos CC operando no modo regenerativo.

Para representar as sensibilidades das cargas frente a afundamentos de tensão, pode-se utilizar as curvas CBEMA, ITIC e SEMI.

2.4 Variáveis de Influência dos Parâmetros de AMT

Diversos fatores influenciam o afundamento de tensão. Associado ainda à aleatoriedade de algumas variáveis, a análise deste distúrbio se torna complexa (CARVALHO FILHO et al., 2002) e (CARVALHO FILHO, 2000). As principais variáveis de influência são:

- Localização da falta;
- Tipo de falta;
- Taxas de falta de linhas de transmissão e distribuição;
- Impedância de falta;
- Tensão pré-falta;
- Conexão dos transformadores entre o ponto de falta e a carga;
- Desempenho do sistema de proteção;
- Sistemas de religamento.

2.4.1 Localização da Falta

Os valores de magnitude do afundamento de tensão percebidos pelos consumidores, estão relacionados com a localização da falta. Em outras palavras, os barramentos mais próximos ao local onde ocorreu a falta, sofrem um afundamento na tensão mais intenso e a severidade do afundamento percebido é diminuída progressivamente à medida que se distancia do local de sua ocorrência.

Faltas ocorridas em sistemas de transmissão e subtransmissão, devido as suas características de abranger uma grande área geográfica e ser malhado, possuem maior potencial de serem sentidas por um número maior de consumidores, quando comparadas a faltas ocorridas

em sistemas de distribuição, que são mais concentrados geograficamente e possuem geralmente configuração radial.

Um conceito atrelado à localização da falta é a área de vulnerabilidade, definida como a região do sistema elétrico onde a ocorrência de curto-circuito poderá afundar a tensão de um barramento fixado para análise abaixo de limites pré-estabelecidos. Esta região consiste de nós e segmentos de linhas, e apesar da nomenclatura “área”, é usual informar tal dado com a unidade de medida (quilômetros de rede elétrica). Para determinar a área de vulnerabilidade é necessário simular em programas de cálculo de curto-circuito faltas aplicadas ao longo de todas as linhas e barramentos do sistema.

Para exemplificar, a Figura 2.2 ilustra duas áreas coloridas: a área de coloração escura representa a região onde a ocorrência de faltas, ocasionam afundamentos de tensão no nó k com magnitude igual ou inferior a 0,7 p.u. e a área de coloração clara, mostra a região onde a ocorrência de faltas acarretam afundamentos de tensão com magnitudes na faixa de 0,7 p.u. a 0,9 p.u.

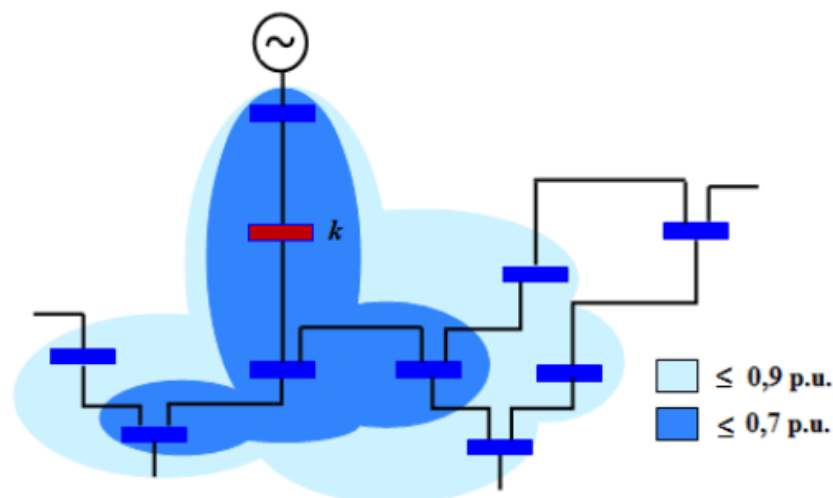


Figura 2.2 - Área de vulnerabilidade para nó k de um sistema (KEMPNER, 2012).

Outro conceito relacionado à localização da falta é a distância crítica, que corresponde à distância a partir da barra monitorada onde a ocorrência de falta gera afundamento de tensão. Este conceito é geralmente utilizado em sistemas elétricos radiais, característica típica de sistemas de distribuição.

A forma de determinação da distância crítica, possui uma vantagem em relação à área de vulnerabilidade, pois não é necessário simular faltas ao longo de todo o alimentador. Sua determinação se dá através de cálculos que necessitam como entrada os valores da impedância

do alimentador distribuída por unidade de comprimento, a impedância do circuito equivalente a montante do barramento monitorado e o limite de tensão remanescente desejado.

Por exemplo, aplicando-se tal conceito no sistema mostrado na Figura 2.3, o método é simples e se trata de um divisor de tensão. No cálculo, a carga é desprezada e a tensão pré-falta é adotada como 1 p.u. Assim, para a barra monitorada tem-se, para uma falta trifásica, a equação (2.1).

$$V_{barra_monitorada} = \frac{Z_J}{Z_J + Z_M} \quad (2.1)$$

Onde:

$V_{barra_monitorada}$: tensão da barra monitorada em p.u;

Z_J : impedância à jusante da barra monitorada, isto é, entre a barra monitorada e o ponto de falta, em Ohms ou em p.u;

Z_M : impedância à montante, isto é, entre a fonte e a barra monitorada, em Ohms ou em p.u.

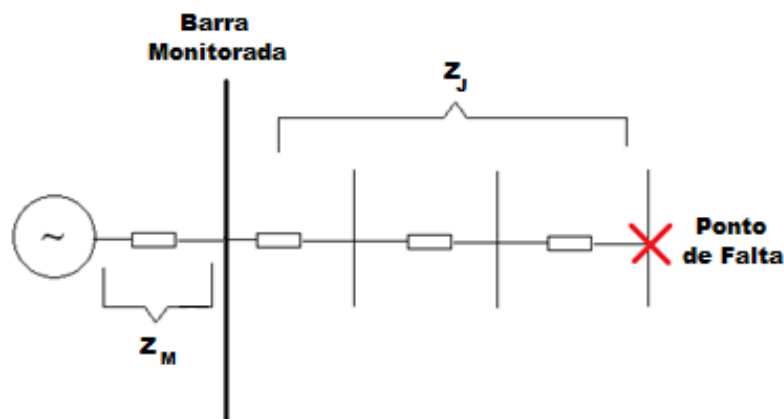


Figura 2.3 - Método da distância crítica em sistema radial.

Considerando $Z_J = z \times L$, sendo z a impedância do alimentador por unidade de comprimento e L a distância entre a barra monitorada e o ponto de falta, obtém-se (2.2).

$$V_{barra_monitorada} = \frac{zL}{zL + Z_M} \quad (2.2)$$

Com isto, a distância crítica pode ser calculada e todas as faltas que ocorrerem além da distância crítica irão causar afundamentos abaixo da tensão crítica. Por outro lado, as

ocorrências de faltas até a distância crítica irão provocar afundamentos mais severos e a severidade da falta se intensificará à medida que se aproximar da barra monitorada. Assim tem-se em (2.3).

$$L_{crítica} = \frac{Z_M}{z} \times \frac{V_{crítica}}{(1 - V_{crítica})} \quad (2.3)$$

Para estimar o número de afundamentos de tensão por ano inferiores a uma certa magnitude, basta multiplicar a distância crítica encontrada pela taxa de falta por comprimento e por ano (HONORATO, 2014).

2.4.2 Tipo de Falta

No sistema elétrico de potência trifásico e aterrado, as faltas podem ser: trifásicas (FFF), trifásicas à terra (FFFT), bifásicas (FF), bifásicas à terra (FFT) e fase-terra (FT).

As faltas trifásicas e trifásicas à terra são simétricas e provocam, então, afundamentos de tensão também simétricos. Elas causam, na maioria das vezes, afundamentos de tensão mais severos. Entretanto, são menos prováveis de ocorrer quando comparadas aos outros tipos de faltas.

As faltas bifásicas, bifásicas à terra e, sobretudo, as fase-terra, apresentam as maiores taxas de ocorrência, gerando afundamentos de tensão, em geral, menos severos, porém desequilibrados e assimétricos.

A cada tipo de falta está associada também uma probabilidade de ocorrência conforme o nível de tensão. Na Tabela 2.2, são mostradas porcentagens de faltas para cada tipo de curto-circuito. Estes dados foram utilizados no caso teste de (FONSECA, 1999) e não devem ser adotados como valores típicos.

Tabela 2.2 - Porcentagem de faltas para cada tipo de curto-circuito.

Nível de Tensão	Fase-Terra	Bifásica-Terra	Bifásica	Trifásica
138 kV	75%	13%	10%	2%
13,8 kV	55%	20%	20%	5%

Segundo (AMASIFEN, 2008), os tipos de curtos-circuitos em uma rede de distribuição se comportam com as probabilidades apresentadas na Tabela 2.3, valendo, entretanto, a mesma ressalva apresentada para a Tabela 2.2.

Tabela 2.3 - Número de ocorrências percentual por tipo de falta.

Fase-Terra	Bifásica-Terra	Bifásica	Trifásica
50%	20%	20%	10%

2.4.3 Taxas de Falta

A ocorrência de descargas atmosféricas em uma área é mensurada, em média, através do nível cerâmico, pois quanto maior o número de ocorrências de descargas atmosféricas maior é a exposição das linhas de transmissão e distribuição. Além das descargas atmosféricas, outras variáveis também afetam as taxas de falta, como: maresia, umidade e localização geográfica da rede elétrica (RAMOS, 2009).

A taxa de falta é um dado de difícil obtenção, pois para se ter representatividade estatística são necessárias medições de ocorrências por um longo período de tempo. Portanto, uma alternativa muito utilizada em estudos é considerar dados típicos obtidos em bibliografias sobre o tema.

Nas Tabelas 2.4 a 2.7, são mostrados alguns exemplos de dados de taxa média de faltas, considerando várias classes de tensão.

Tabela 2.4 - Taxas de faltas por 100 km/ano (GARCIA; ALVES, 2011).

Nível de Tensão	Taxa de Falta	Fase-Terra	Bifásica-Terra	Bifásica	Trifásica
500 kV	1	92%	0%	3%	5%
345 kV	2	92%	0%	3%	5%
230 kV	2	92%	0%	3%	5%
138 kV	6	75%	13%	10%	2%
69 kV	12	58%	25%	11%	6%
13,8 kV	50	78%	10%	9%	3%

Tabela 2.5 - Taxas de faltas por 100 km/ano (CONRAD; LITTLE; GRIGG, 1991) e (BECKER et al., 1994).

Nível de Tensão	Taxa de Falta	Fase-Terra	Bifásica-Terra	Bifásica	Trifásica
345 kV	2,31	91%	7%	1%	1%
230 kV	1,68	80%	17%	1,5%	1,5%
138 kV	2,98	73%	17%	6%	4%

Tabela 2.6 - Número de Ocorrências por 100 km/ano (FONSECA, 1999).

Nível de Tensão	Taxa de Falta
138 kV	5
13,8 kV	50

Tabela 2.7 - Taxa de falta para linhas de transmissão (*adotadas pela CELG) (RAMOS, 2009).

Nível de Tensão	Taxa de Falta
230 kV	2,32
138 kV	3,99
69 kV *	6

2.4.4 Impedância de Falta

Curtos-circuitos no SEP (Sistema Elétrico de Potência) dificilmente possuem impedância de falta nula (CABRAL, 2010). Geralmente, eles ocorrem devido a uma associação de resistências, como:

- Resistência do arco elétrico entre o condutor e a terra, para defeitos fase-terra;
- Resistência do arco elétrico entre dois ou mais condutores, para defeitos entre as fases;
- Resistência de contato devido à oxidação no local da falta;
- Resistência do pé de torre, para defeitos englobando a terra.

Para redes de transmissão, valores de resistência de falta entre 1 e 5 Ω são citados em (CARVALHO FILHO et al., 2002), (ORTMEYER; HIYAMA; SALEHFAR, 1996) e (BLACKBURN; DECKKER, 1987).

Segundo (DAGENHART, 2000), em linhas de distribuição, os valores típicos de resistência de falta são menores ou iguais a 40 Ω .

Já em (SOUSA; COSTA; JUNIOR, 2005), para sistemas aéreos de distribuição são estimados valores de resistência de falta para descargas atmosféricas, entre 0 e 10 Ω . Para defeitos provocados por árvores próximas das linhas são encontrados valores próximos de 70 Ω . Já faltas ocasionadas pela queda de estrutura resultam em resistências de falta entre 20 e 30 Ω .

Nas simulações de (MARTINEZ; MARTIN-ARNEDO, 2006), em redes de média e baixa tensão, a resistência de falta foi considerada dentro de uma distribuição normal cuja média era de 5 Ω e desvio-padrão de 1 Ω para cada fase em falta.

(AMASIFEN, 2008) vai além e especifica valores máximos de impedâncias de faltas para cada tipo de falta. Tais valores, fornecidos por uma concessionária, consistem em 10 Ω para as faltas trifásicas (impedância por fase), 30 Ω para faltas fase-terra, 20 Ω para faltas fase-fase e, nas faltas dupla fase-terra, 10 Ω para impedância entre fases e 30 Ω na impedância para

terra. Os eventos considerados nestas simulações foram restringidos para faltas originadas na rede primária de distribuição (13,8 kV).

No trabalho de (CABRAL, 2010) foram adotados para resistência de falta em sistemas de distribuição valores de 0, 1, 5, 15 e 25 Ω , sendo as probabilidades igualmente distribuídas para estes valores.

Por fim, o valor dessa variável não é de fácil determinação. No entanto, é uma variável que influencia nas intensidades dos afundamentos de tensão. No trabalho de (RAMOS, 2009) foi ressaltada a necessidade de representação desta variável aleatória para tornar os estudos mais próximos da realidade, pois valores nulos, normalmente adotados, tornam os eventos mais severos quando comparados com eventos que consideram esta impedância. As impedâncias aleatórias foram consideradas entre 0 e 15 Ω para faltas fase-terra no sistema de distribuição da CELG (MARTINEZ; MARTIN-ARNEDO, 2004).

2.4.5 Tensão Pré-Falta

O objetivo de qualquer distribuidora de energia elétrica é fornecer ao consumidor energia com tensão adequada segundo os limites normalizados, correspondente a 0,95 e 1,05 p.u. No entanto, o perfil de tensão em regime permanente está relacionado à curva de carga do sistema elétrico, bem como a existência de dispositivos responsáveis pelo controle de tensão, como banco de capacitores, reatores de linha, compensadores síncronos, etc.

Geralmente, a tensão se eleva em períodos de carga leve e apresenta redução nos períodos de carga pesada, ou seja, o perfil de tensão acompanha a variação da curva de carga diária.

Na maioria dos trabalhos envolvendo cálculos de curto-circuito em sistemas elétricos, atribui-se o valor de tensão pré-falta igual a 1,0 p.u. Porém, levando em conta a curva de carga do sistema, esta suposição feita é errônea, trazendo pequenos desvios nos valores calculados.

2.4.6 Conexão dos Transformadores

O modo de conexão dos transformadores existentes entre o local da falta e a barra na qual o consumidor está ligado, influencia diretamente as características do afundamento de tensão percebido pela carga. Simplificadamente, os transformadores podem ser agrupados em três categorias (BOLLEN, 1997).

- Primeira: aqueles, cujas tensões nas bobinas em um dos enrolamentos (primário ou secundário) são função da diferença fasorial (tensão composta) entre duas tensões aplicadas nas bobinas do outro enrolamento. Estes transformadores são os de conexão $Y-\Delta$, $\Delta-Y$, $Y_{\text{aterrado}}-\Delta$ e $\Delta-Y_{\text{aterrado}}$, que além de filtrarem a componente de sequência zero da tensão de frequência fundamental, introduzem defasamento angular entre as tensões primária e secundária;
- Segunda: são os transformadores que somente filtram as componentes de sequência zero da tensão de frequência fundamental, e que, geralmente do ponto de vista construtivo, são fabricados de modo a não introduzir defasamento angular, ou seja, com conexões $Y-Y$, $\Delta-\Delta$, $Y_{\text{aterrado}}-Y$, $Y-Y_{\text{aterrado}}$;
- Terceira: são aqueles que não filtram as componentes de sequência e geralmente, devido às mesmas razões citadas anteriormente, não introduzem defasamento angular. Pertencem a esta categoria os transformadores com as conexões $Y_{\text{aterrado}}-Y_{\text{aterrado}}$, $Y_{\text{aterrado}}-\Delta-Y_{\text{aterrado}}$. Neste último caso, o Δ é um enrolamento de compensação.

Em (CARVALHO FILHO, 2000), (OLIVEIRA et al., 2003) os autores enfatizam que os valores dos afundamentos de tensão vistos pela carga em decorrência de uma falta no sistema elétrico, dependem do efeito combinado da forma de conexão tanto do transformador como da carga.

2.4.7 Sistemas de Proteção

A duração dos afundamentos de tensão está diretamente relacionada ao desempenho dos sistemas de proteção. O tempo de sensibilização e de atuação dos relés, somado ao tempo de abertura e de extinção do arco pelos disjuntores, compõem o tempo total de desempenho do sistema de proteção.

O tempo de atuação dos relés é função de suas características de resposta (tempo dependente, tempo definido, entre outros), bem como da filosofia e dos ajustes adotados para se obter a seletividade desejada. Já o tempo de abertura e de extinção da corrente de curto-circuito dos disjuntores é função das características construtivas destes equipamentos (HONORATO, 2014).

Nos sistemas de distribuição, normalmente utilizam-se relés de sobrecorrente de fase e de neutro. Nos alimentadores primários são usados religadores, enquanto que nos ramais da rede de distribuição são empregadas chaves fusíveis.

2.4.8 Sistemas de Religamento

O número de afundamentos de tensão contabilizados está relacionado à existência de religadores no sistema de proteção, ao tipo de curto-circuito ocorrido (temporário ou permanente) e principalmente à filosofia de proteção adotada.

Existem duas maneiras de se contabilizar os afundamentos de tensão quando ocorrem religamentos. A primeira contabiliza todos os afundamentos medidos, resultando assim em um valor sobreestimado de ocorrência do fenômeno. A outra maneira lança mão de um procedimento chamado agregação temporal, que agrega eventos consecutivos ocorridos em um janela de tempo determinada e contabiliza apenas um evento.

2.5 Indicadores

Apesar de tantos estudos dedicados ao fenômeno de afundamentos de tensão, não há um consenso mundial quanto à forma de estabelecer indicadores para tal distúrbio. Algumas normas como a (IEEE P1564/D19, 2013), (IEC 61000-2-8, 2005), (IEC 61000-4-11, 2004), (NRS 048-2, 2003) apresentam propostas de tabelas discretizando faixas de duração e intensidade para contabilizar os eventos.

A norma Sul-Africana (NRS 048-2, 2003), além da discretização, apresenta valores limites para cada região da tabela de duração-intensidade em diferentes níveis de tensão. Em tal norma são definidas as categorias de afundamentos de tensão representadas na Tabela 2.8

Tabela 2.8 – Quantificação dos afundamentos de tensão segundo a NRS 048-2.

Magnitude da Tensão Residual	Duração do Afundamento de Tensão		
	20ms – 150ms	150ms – 600ms	0,6s – 3s
90% - 85%	Y		Z1
85% - 80%			
80% - 70%	X1	S	Z2
70% - 60%			
60% - 40%			
<40%	T		

Cada categoria apresenta diretrizes gerais sobre a ocorrência de afundamentos e a possibilidade de consumidores serem afetados, baseadas nas características da proteção dos sistemas elétricos e na compatibilidade das cargas (FILHO, 2016). Então, as regiões de magnitude e duração são assim definidas:

- Tipo Y – afundamentos esperados com frequência em sistemas típicos de média e alta tensão, para os quais as próprias unidades consumidoras devem assegurar suas devidas operações e ou proteções nesta faixa;
- Tipos X1 e X2 – afundamentos com durações típicas de atuação de proteção de sistemas de alta tensão em sistemas remotos com quantitativo significativo de eventos. Unidades consumidoras com equipamentos sensíveis devem buscar proteção para eventos notadamente na faixa X1;
- Tipo T – afundamentos ocasionados por faltas próximas à unidade consumidora afetada que não se espera ocorrer com grande frequência;
- Tipo S – afundamentos menos comuns que os dos tipos X e Y; curtos-circuitos em pontos remotos do sistema em situações onde são utilizados sistemas de proteção por impedância ou com lenta recuperação da tensão pós-evento;
- Tipos Z1 e Z2 – afundamentos muito incomuns em sistemas de alta tensão, podendo estar associados a deficiências no esquema de proteção, porém mais presentes na média tensão, com proteções de retaguarda ou térmicas.

A ANEEL, através da CP nº18/2014 e da AP Nº82/2015, propõe um indicador baseado na proposta feita pela norma sul-africana e na curva ITIC. Porém considerando as particularidades do sistema elétrico brasileiro, foi criada a Tabela 2.9 que apresenta 9 regiões de sensibilidade, visando correlacionar a importância de cada evento de VTCD com os níveis de sensibilidade de diferentes cargas.

Tabela 2.9 – Estratificação das VTCDs com base nos níveis de sensibilidade das diversas cargas.

Amplitude (p.u)	Duração						
	[16,67ms – 100ms]	(100ms – 300ms]	(300ms – 600ms]	(600ms – 1seg]	(1seg – 3seg]	(3seg – 1min]	(1min – 3min)
> 1,15	REGIÃO H			REGIÃO I			
(1,10 – 1,15]	REGIÃO H			REGIÃO I			
(0,85 – 0,90]	REGIÃO A						REGIÃO G
(0,80 – 0,85]	REGIÃO A						
(0,70 – 0,80]	REGIÃO B	REGIÃO D		REGIÃO F			
(0,60 – 0,70]	REGIÃO C						
(0,50 – 0,60]	REGIÃO C						
(0,40 – 0,50]	REGIÃO E						
(0,30 – 0,40]							
(0,20 – 0,30]							
(0,10 – 0,20]							
< 0,10	REGIÃO E						

O indicador chamado Fator de Impacto (FI), usado para caracterizar os barramentos de distribuição em função da incidência de eventos de VTCD, é apurado para um período de 30 dias consecutivos e é calculado pela equação (2.4).

$$FI = \frac{\sum_{i=A}^I (f_{ei} \times fp_i)}{FI_{BASE}} \quad (2.4)$$

Onde: f_{ei} é a frequência de ocorrência de eventos de VTCD, apuradas por meio de medição apropriada, em um período de 30 dias consecutivos, para cada região de sensibilidade i , com $i = A$ até I ;

fp_i é o fator de ponderação para cada região de sensibilidade;

FI_{BASE} é o fator de impacto base, calculado considerando os fatores de ponderação e a frequência limite de ocorrência de VTCD para cada região de sensibilidade.

O fator de ponderação (fp) para cada região de sensibilidade e também o fator de impacto base a ser utilizado é indicado na Tabela 2.10.

Tabela 2.10 – Fatores de ponderação e Fator de Impacto Base de acordo com a tensão nominal do barramento de distribuição.

Região de Sensibilidade	Fator de Ponderação (fp)	Fator de Impacto Base (FI_{BASE})	
		1,0 kV < Vn < 69 kV	69 kV < Vn < 230 kV
A	0,00	2,13	1,42
B	0,04		
C	0,07		
D	0,15		
E	0,25		
F	0,36		
G	0,07		
H	0,02		
I	0,04		

O valor limite para o indicador Fator de Impacto permitido para cada ponto analisado do sistema de distribuição, apurado por medição apropriada no período de 30 dias consecutivos, é de 1,0 p.u. Para os barramentos de distribuição com tensões nominais inferiores a 1 kV não são definidos limites de VTCDs.

Contribuições recebidas de diversas empresas na CP n°18/2014, solicitam esclarecimento sobre a forma de obtenção das regiões de sensibilidade apresentadas na Tabela

2.9 (regiões A à F), dos fatores de ponderação, das frequências máximas dos eventos por sensibilidade que resultam no FI_{BASE} . Portanto, trata-se de uma proposta apresentada pela ANEEL e que ainda encontra-se em discussão no setor elétrico brasileiro.

2.6 Considerações Finais

O presente capítulo apresentou a definição de afundamentos de tensão segundo a norma vigente no sistema de distribuição de energia brasileiro e para salientar as pequenas diferenças, foi mostrado também como as principais normas internacionais caracterizam o fenômeno.

Para compreender a origem dos AMTs, foram citados os principais fenômenos causadores deste (apontando os curto-circuitos originados por descargas atmosféricas como a principal causa), seguido de uma descrição dos efeitos que o distúrbio provoca nas cargas dos consumidores industriais.

Foram indicadas as variáveis que influenciam os parâmetros dos AMTs, a saber localização da falta, tipo de falta, taxas de falta de linhas de transmissão e distribuição, impedância de falta, tensão pré-falta, conexão dos transformadores entre o ponto de falta e a carga, desempenho do sistema de proteção, sistemas de religamento.

Para finalizar foi mostrada a discretização em regiões feita pela norma sul-africana (NRS 048-2, 2003), e a discretização proposta na AP N°82/2015 bem como o indicador FI – Fator de Impacto, com os respectivos limites de frequência de ocorrência por região da tabela.

3. FUNDAMENTOS DA FORMAÇÃO DE CONJUNTOS

Neste capítulo, são apresentados os conceitos gerais e definições básicas sobre os algoritmos matemáticos dos métodos de formação de conjuntos mais utilizados. No Anexo A é apresentada a resolução de um exemplo didático considerando os diversos métodos citados neste capítulo para ilustrar a maneira como o algoritmo é implementado nos *softwares* que possuem essa funcionalidade.

3.1 Considerações Iniciais

A análise de agrupamentos, também conhecida como análise de conglomerados, classificação ou *cluster*, tem como objetivo dividir os elementos da amostra, ou população, em grupos de forma que os elementos pertencentes a um mesmo grupo sejam similares entre si com respeito às variáveis (características) que neles foram medidas, e os elementos em grupos diferentes sejam heterogêneos em relação a estas mesmas características (MINGOTI, 2005).

A análise de agrupamento é uma técnica mais primitiva, onde não são feitas suposições sobre o número de grupos ou sobre a estrutura do grupo. O agrupamento é feito com base em semelhanças ou diferenças (distâncias). Os dados necessários são medidas de similaridade ou dados a partir do qual semelhanças possam ser computadas (JOHNSON; WICHERN, 2007).

Na maioria das aplicações práticas de formação de conjuntos, quem está realizando o trabalho sabe o suficiente do problema para distinguir o “bom” grupo do grupo “ruim”.

Mesmo sem o conhecimento matemático da teoria de formação de conjuntos, muitas vezes se é capaz de formar grupos quando se considera apenas duas ou três variáveis, pois há a possibilidade de representar os objetos em gráficos e fazer agrupamentos de forma visual. No entanto, a maioria dos trabalhos desenvolvidos no campo de análise multivariada, contém um número de variáveis superior a três. Então, a utilização de computadores é inevitável, e por este motivo diversos softwares estatísticos implementaram algoritmos que permitem a formação de conjuntos, como SAS, STATISTICA, MINITAB e SPSS.

O primeiro passo para análise de agrupamentos é desenvolver uma escala quantitativa para medir a associação (similaridade) entre os objetos em estudos. A seção 3.2 é dedicada à apresentação das medidas de similaridade mais utilizadas e a seção 3.3 descreve alguns dos métodos mais comuns utilizados para a formação de conjuntos.

3.2 Medidas de Similaridade e Dissimilaridade

Normalmente a escolha de uma medida de similaridade é muito subjetiva, então, considerações importantes para evitar a subjetividade são utilizadas, tais como: a natureza das variáveis (contínua, discreta, binária), a escala de medição (nominal, ordinal, intervalo) e os conhecimentos sobre o assunto em pesquisa.

Supondo um conjunto de dados com n elementos amostrais, tendo-se medido p variáveis aleatórias em cada um deles. Define-se para cada elemento amostral j o vetor de medidas X_j , conforme equação (3.1).

$$X_j = [X_{1j} \ X_{2j} \ \dots \ X_{pj}]', \quad j = 1, 2, \dots, n \quad (3.1)$$

Onde X_{ij} representa o valor medido da variável i para o elemento j . Para iniciar o processo de formação de conjuntos é necessário que se determine qual medida de similaridade ou dissimilaridade será utilizada. Há diversas medidas diferentes e a escolha delas afeta o resultado final dos agrupamentos formados. Apresenta-se a seguir as medidas mais usuais que são apropriadas para variáveis quantitativas, ressaltando que as medidas que serão mostradas são de dissimilaridade, ou seja, os elementos que estão sendo comparados serão mais similares, quanto menor forem os valores das medidas.

Distância Euclidiana: Considerando dois elementos X_l e X_k , $l \neq k$, a distância euclidiana entre eles é definida por (3.2).

$$d(X_l, X_k) = [(X_l - X_k)'(X_l - X_k)]^{\frac{1}{2}} = [\sum_{i=1}^p (X_{il} - X_{ik})^2]^{\frac{1}{2}} \quad (3.2)$$

Neste caso, cada variável pertencente ao vetor de observações é comparada entre os dois elementos amostrais.

Distância Generalizada ou Ponderada: Considerando os mesmos elementos da demonstração anterior, a distância generalizada entre eles é definida por (3.3).

$$d(X_l, X_k) = [(X_l - X_k)'A(X_l - X_k)]^{\frac{1}{2}} \quad (3.3)$$

$A_{p \times p}$ é uma matriz de ponderação, positiva definida. Geralmente, $A_{p \times p} = S_{p \times p}^{-1}$

(distância de Mahalanobis), sendo $S_{p \times p}$ a matriz de covariância amostral. No entanto, sem o conhecimento prévio da distinção dos conjuntos, a covariância amostral não pode ser calculada. Por esta razão, a distância Euclidiana é muitas vezes preferida para realizar a formação dos conjuntos.

Distância de Minkowsky: Considerando os mesmos elementos das demonstrações anteriores, a distância de Minkowsky entre eles é definida por (3.4).

$$d(X_l, X_k) = \left[\sum_{i=1}^p w_i |X_{il} - X_{ik}|^\lambda \right]^{\frac{1}{\lambda}} \quad (3.4)$$

Onde w_i são os pesos de ponderação para as variáveis. Para $\lambda = 1$, $d(X_l, X_k)$ se torna a distância chamada *city-block* ou Manhattan, e para $\lambda = 2$ tem-se a distância Euclidiana. Em geral, variando λ muda o peso dado às diferenças maiores e menores, por este motivo a métrica de Minkowsky é menos afetada pela presença de *outliers* na amostra do que a distância Euclidiana.

3.3 Métodos para Formação de Conjuntos (Clusters)

Mesmo com o auxílio de um computador com grande capacidade de processamento, é impossível verificar todas as possibilidades de agrupamentos em problemas onde o número de objetos e variáveis são consideráveis. Devido a este problema, foram criados uma variedade de algoritmos para formação de conjuntos que, sem olhar para todas as configurações possíveis de agrupamento, permitem encontrar agrupamentos razoáveis.

Os métodos de agrupamentos ou *clusters* são divididos em dois tipos: os não hierárquicos e os hierárquicos e estes são novamente classificados em aglomerativos e divisivos. O algoritmo dos métodos hierárquicos aglomerativos, iniciam considerando cada objeto como um conjunto, e em cada etapa agrupa os objetos ou conjunto de objetos que mais se assemelham. No final do processo, a um nível de similaridade pequeno, todos os objetos compõem um único conjunto. Em contrapartida, as técnicas hierárquicas divisivas funciona na direção oposta, ou seja, parte-se de um conjunto único que contém todos os objetos do estudo e a cada passo os separam de acordo com a dissimilaridade, encerrando com conjuntos formados por apenas um objeto. Um exemplo ilustrativo de ambos os processos está apresentado na Figura 3.1. A principal diferença entre os processos é que os não hierárquicos necessitam que o número de grupos que se deseja formar seja pré-estabelecido pelo pesquisador, enquanto o hierárquico não requer tal dado de entrada.

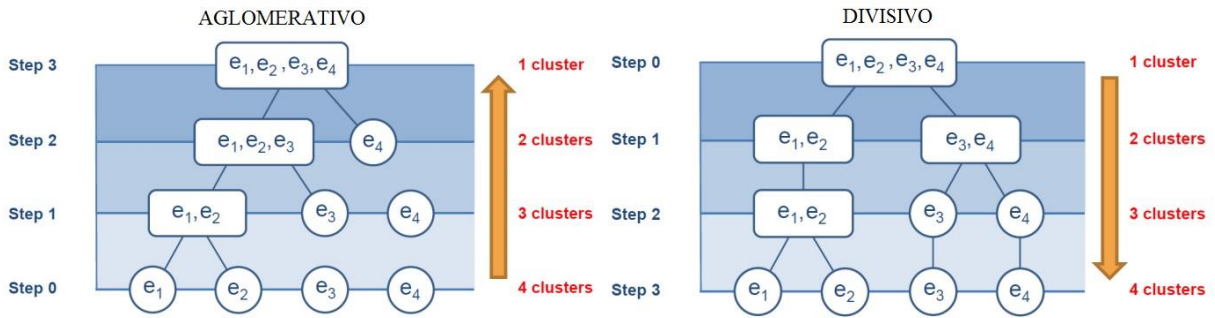


Figura 3.1 – Ilustração das técnicas hierárquicas aglomerativas e divisivas.

O resultado da aplicação de métodos hierárquicos pode ser exibido de forma gráfica por um diagrama bidimensional conhecido por dendograma, que mostra as uniões ou divisões feitas em níveis sucessivos. No eixo das abcissas, geralmente, estão os objetos que se deseja agrupar, e no eixo das ordenadas, são apresentadas as medidas de distância ou de similaridade. Um exemplo de um dendograma é apresentado na Figura 3.2.

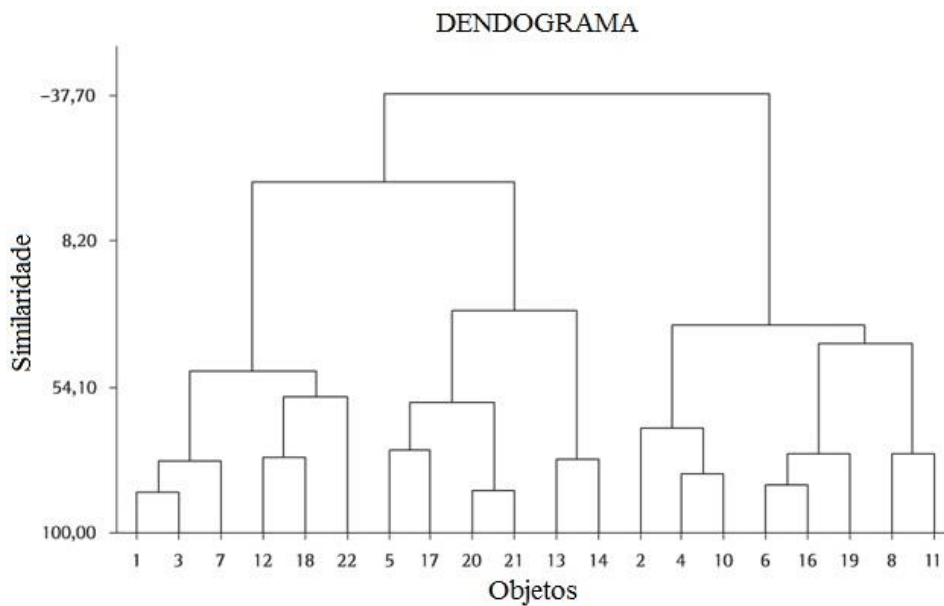


Figura 3.2 – Modelo de um dendograma.

3.3.1 Técnicas Hierárquicas Aglomerativas

Os principais passos para implementação dos métodos hierárquicos aglomerativos, estão listados a seguir:

1. Começa-se com n agrupamentos, cada um formado por um único objeto, então, cria-se uma matriz D simétrica $n \times n$ de distâncias (ou similaridades).
2. Procura-se na matriz de distâncias o par de agrupamentos mais próximos (maior

similaridade), supondo que a menor distância está entre os agrupamentos U e V.

3. Unir os agrupamentos U e V. Nomear o novo agrupamento formado como UV. Atualizar a matriz de distâncias, apagando as linhas e colunas correspondentes aos agrupamentos U e V, e adicionando a linha e a coluna referente às distâncias entre o agrupamento UV e os demais agrupamentos.
4. Repita os passos 2 e 3 por $n-1$ vezes. (No término do algoritmo, todos os objetos formaram um único agrupamento). Registrar os agrupamentos que estão sendo formados e em que nível (distância ou similaridade) as uniões ocorrem.

A escolha do número final de agrupamentos em que o total dos objetos devem ser divididos é subjetiva, sendo que há algumas técnicas para ajudar a determinar este número. O objetivo final é encontrar um número que representa a “partição natural” dos objetos que estão sendo agrupados.

Existem diversos métodos hierárquicos para formação de conjuntos. Na sequência serão apresentados os métodos mais utilizados e implementados na maioria dos *softwares* estatísticos. A melhor maneira de compreender os procedimentos de agrupamento é através de exemplos. Logo, as seções seguintes apresentam uma breve explicação sobre as entradas do algoritmo e seus componentes para cada método de formação de agrupamentos e o Anexo A mostra a resolução de um exemplo simples para ilustração do processo.

3.3.1.1 Método de Ligação Simples (*Single Linkage*)

A entrada para o algoritmo do método de ligação simples pode ser as distâncias ou similaridades entre os pares de objetos. Os agrupamentos são formados partindo de um objeto e unindo-o ao “vizinho” mais próximo (menor distância ou maior similaridade). Supondo que num determinado estágio do algoritmo de agrupamento existam dois grupos, um contendo os elementos amostrais 1, 3 e 7 e o outro contendo os elementos 2 e 6, assim:

$$C_1 = \{X_1, X_3, X_7\} \text{ e } C_2 = \{X_2, X_6\}$$

Então, a distância entre esses dois grupos será definida por:

$$d(C_1, C_2) = \min\{d(X_l, X_k), l \neq k, l = 1,3,7 \text{ e } k = 2,6\}$$

Na Figura 3.3 está representado graficamente o processo de comparação entre os agrupamentos C_1 e C_2 , onde a distância entre os elementos 1 e 6 é a menor em relação às demais.

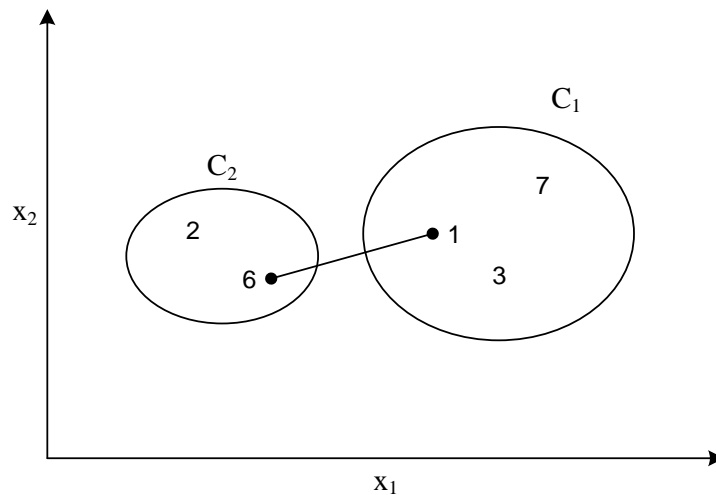


Figura 3.3 – Ilustração gráfica do método de ligação simples.

3.3.1.2 Método de Ligação Completa (*Complete Linkage*)

O método de ligação completa tem algoritmo muito parecido com o de ligação simples, com uma importante diferença, em cada estágio a matriz de distâncias (ou similaridade) entre os agrupamentos é calculada pelos dois elementos, um de cada agrupamento, que estão mais distantes. Desta forma, a técnica de ligação completa garante que os itens de um agrupamento estejam a uma distância máxima de outro agrupamento. Supondo novamente que num determinado estágio existam dois grupos, um contendo os elementos amostrais 1, 3 e 7 e o outro contendo os elementos 2 e 6, assim:

$$C_1 = \{X_1, X_3, X_7\} \text{ e } C_2 = \{X_2, X_6\}$$

Então, a distância entre esses dois grupos será definida por:

$$d(C_1, C_2) = \max\{d(X_l, X_k), l \neq k, l = 1, 3, 7 \text{ e } k = 2, 6\}$$

Em cada passo do algoritmo, as distâncias são calculadas para todos os pares de grupos, sendo, então, combinados num único grupo aqueles que apresentarem o menor valor da distância, isto é, o menor valor máximo. Na Figura 3.4, está representado graficamente o processo de comparação entre os agrupamentos C_1 e C_2 , onde a distância entre os elementos 2 e 7 é a maior em relação às demais.

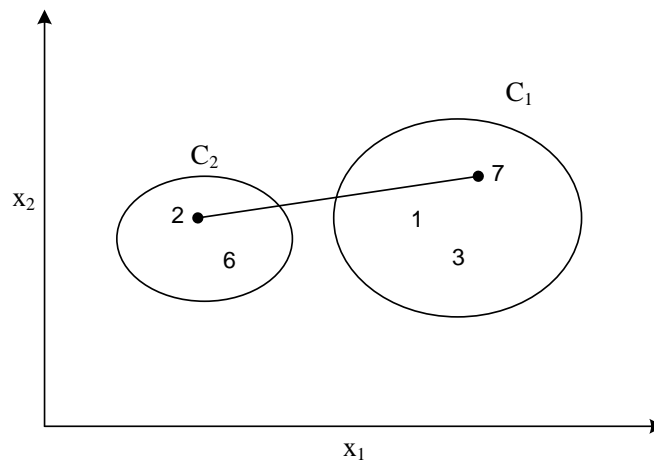


Figura 3.4 – Ilustração gráfica do método de ligação completa.

3.3.1.3 Método da Média das Distâncias (*Average Linkage*)

Este método trata a distância entre dois agrupamentos como a média das distâncias entre todos os pares de objetos, onde cada membro do par pertence a um agrupamento diferente. Se o agrupamento C_1 tem n_1 elementos e o agrupamento C_2 tem n_2 elementos, a distância entre eles será definida por (3.5).

$$d(C_1, C_2) = \frac{1}{n_1 n_2} \sum_{l \in C_1} \sum_{k \in C_2} d(X_l, X_k) \quad (3.5)$$

A distância entre os agrupamentos: $C_1 = \{X_1, X_3, X_7\}$ e $C_2 = \{X_2, X_6\}$ é igual a:

$$d(C_1, C_2) = \frac{1}{6} [d(X_1, X_2) + d(X_1, X_6) + d(X_3, X_2) + d(X_3, X_6) + d(X_7, X_2) + d(X_7, X_6)]$$

Na Figura 3.5, está representado graficamente o processo de comparação entre os agrupamentos C_1 e C_2 .

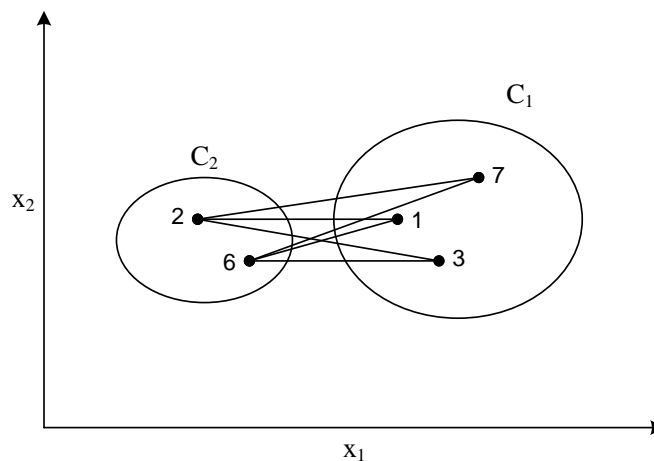


Figura 3.5 – Ilustração gráfica do método da média das distâncias.

3.3.1.4 Método de Ward

É desejado que os agrupamentos gerados sejam os mais heterogêneos possíveis e que os objetos dentro de cada agrupamento sejam homogêneos. Nos métodos anteriores, nota-se que, de um estágio para o seguinte, a qualidade da partição decresce, pois o nível de similaridade também decresce. Isso significa que a variação entre grupos diminui e a variação dentro dos grupos aumenta.

O método proposto por Ward é fundamentado justamente nesta mudança de variação entre os agrupamentos assim como dentro dos agrupamentos que estão sendo gerados a cada passo do algoritmo.

Assim, como nos demais métodos aglomerativos, inicialmente cada objeto é considerado como um agrupamento, e a cada passo do algoritmo calcula-se a distância Euclidiana de cada objeto da amostra presente no agrupamento em relação ao respectivo vetor de médias do agrupamento, conforme (3.6).

$$SS_i = \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)' (X_{ij} - \bar{X}_i) \quad (3.6)$$

Onde, n_i é o número de objetos no agrupamento C_i quando se está no passo k do processo de formação dos agrupamentos, X_{ij} é o vetor de observações do j -ésimo objeto amostral que pertence ao i -ésimo agrupamento, \bar{X}_i é o centróide do agrupamento, e SS_i representa a soma de quadrados correspondente ao agrupamento C_i . No passo k , a soma de quadrados total dentro dos grupos é definida por (3.7).

$$SSR = \sum_{i=1}^{g_k} SS_i \quad (3.7)$$

Onde g_k é o número de agrupamentos existentes quando se está no passo k .

A distância entre os agrupamentos C_l e C_i é definida por (3.8).

$$d(C_l, C_i) = \left[\frac{n_l n_i}{n_l + n_i} \right] (\bar{X}_l - \bar{X}_i)' (\bar{X}_l - \bar{X}_i) \quad (3.8)$$

Que corresponde à soma de quadrados entre os agrupamentos C_l e C_i . A cada passo do algoritmo, a união de todos os pares de agrupamentos é considerada, e os dois agrupamentos em que a combinação resultar em um menor crescimento de SSR são combinados.

3.3.1.5 Comentários Finais

Existem muitos métodos hierárquicos de formação de agrupamento além dos

apresentados nesta dissertação. Entretanto, todos os métodos aglomerativos seguem o mesmo algoritmo básico mostrado.

Como na maioria dos métodos de formação de agrupamentos, as fontes de erros e variações não são formalmente consideradas nos métodos hierárquicos, isto significa que o método será sensível a valores discrepantes (*outliers*).

Os procedimentos hierárquicos não prevêm em seus algoritmos a realocação de objetos que tenham sido agrupados incorretamente nos estágios iniciais do processo. Conseqüentemente, a configuração final dos agrupamentos deve sempre ser cuidadosamente analisada para ver se de fato faz sentido do ponto de vista do fenômeno que está sendo estudado.

Para um problema particular, é uma boa ideia tentar vários métodos de formação de agrupamentos e, com um dado método, testar as diferentes medidas de distâncias (similaridades). Se os agrupamentos formados por diversos métodos se repetirem, provavelmente foi encontrado o agrupamento natural para aquele fenômeno.

A estabilidade das soluções dos métodos hierárquicos podem ser testadas aplicando o algoritmo antes e depois da adição de pequenas perturbações nos dados. Caso os agrupamentos se mantenham semelhantes e com boa distinção, o agrupamento encontrado é robusto.

3.3.2 Técnicas de Agrupamento Não Hierárquicas

Os métodos não hierárquicos têm como meta achar diretamente uma partição de n objetos em k agrupamentos, de maneira que a partição atenda duas premissas: “coesão” interna (ou “semelhança” interna) e isolamento (ou separação) dos agrupamentos formados. Com o intuito de encontrar a melhor partição possível, algum critério de qualidade da partição deve ser utilizado. Mesmo com os melhores computadores não é possível analisar todas as partições de ordem k , e então decidir qual é a melhor. Portanto, o que os processos fazem é averiguar algumas partições possíveis objetivando descobrir a partição “quase ótima”.

As técnicas de agrupamento não hierárquicas foram desenvolvidas para agrupar somente objetos, portanto não deve ser aplicada para agrupamento de variáveis. Diferentemente dos métodos hierárquicos aglomerativos, o número de agrupamentos a ser formado deve ser informado previamente.

Outra particularidade dos métodos não hierárquicos é que em cada passo do algoritmo, os novos agrupamentos podem ser criados a partir da divisão ou da união de agrupamentos já combinados em estágios anteriores, conseqüentemente não é mais possível a construção de um

dendograma.

Métodos não hierárquicos podem ser aplicados em diversos conjuntos de dados, e têm uma maior capacidade de análise de conjunto de dados de maior porte, pois a matriz de distâncias (similaridades) não precisa ser determinada.

Os algoritmos não hierárquicos para formação de agrupamentos começam por partição inicial de objetos em agrupamentos ou por um conjunto inicial de sementes, que formarão os núcleos dos agrupamentos. Uma boa escolha para configuração de partida deve ser não tendenciosa, uma maneira de começar é selecionar aleatoriamente as sementes dentre os objetos ou particionar aleatoriamente os objetos em agrupamentos iniciais.

Na seção 3.3.2.1, há uma discussão do procedimento não hierárquico mais popular, conhecido como *k-means*.

3.3.2.1 Método das k-médias (*k-means*)

O funcionamento básico do método consiste na alocação de cada objeto amostral, no agrupamento cujo centróide (vetor de médias amostral) é o mais próximo do vetor de valores observados para o respectivo objeto. Segundo (MINGOTI, 2005) o método é composto por quatro passos:

1. Primeiramente escolhe-se k centróides, chamados de “sementes” ou “protótipos”, para se inicializar o processo de partição;
2. Cada elemento do conjunto de dados é, então, comparado com cada centróide inicial, através de uma medida de distância que, em geral, é a distância Euclidiana. O elemento é alocado ao grupo cuja distância é menor;
3. Depois de aplicar o passo 2 para cada um dos n elementos amostrais, recalcula-se os valores dos centróides para cada novo grupo formado, e repete-se o passo (2), considerando os centróides destes novos grupos;
4. Os passos (2) e (3) devem ser repetidos até que todos os elementos amostrais estejam “bem alocados” em seus grupos, isto é, até que nenhuma realocação de elementos seja necessária.

O resultado final do agrupamento dos objetos, está fortemente relacionado com a escolha inicial das partições ou a escolha das sementes.

3.4 Considerações Finais

Este capítulo apresentou os conceitos básicos sobre alguns métodos de formação de conjuntos, pois este entendimento será necessários para a compreensão da metodologia proposta que será exibida no capítulo 4.

4. PROPOSTA DE METODOLOGIA PARA FORMAÇÃO DE CONJUNTOS QUANTO AOS AFUNDAMENTOS DE TENSÃO

4.1 Considerações Iniciais

Neste capítulo, é apresentada a metodologia proposta para formação de conjuntos considerando afundamentos de tensão. Faz-se inicialmente um resumo dos passos da metodologia. Posteriormente, detalha-se cada etapa da metodologia proposta, seleção de variáveis, implementação da formação de conjuntos, verificação das respostas obtidas e análise de robustez dos métodos utilizados.

4.2 Resumo Passo a Passo da Metodologia

1. Seleciona-se através da experiência do pesquisador um universo de variáveis, que a priori, suspeita-se ter relação com o fenômeno estudado;
2. Através do coeficiente de correlação de Pearson, selecionam-se as variáveis para serem entrada dos algoritmos de formação de conjuntos;
3. Implementa-se em software a formação dos conjuntos através dos diversos métodos que se deseja investigar;
4. Através da análise de variância entre a resposta desejada e o membership de cada método, avaliar qual método gerou os melhores agrupamentos;
5. Verificar através da inserção de pequenos ruídos em alguma variável de entrada do algoritmo, se o método encontrado no item 4 é robusto. Caso seja encontrado dois ou mais métodos que geraram agrupamentos iguais no item 4, recomenda-se observar qual dentre os métodos apresenta maior concordância nos agrupamentos formados quando um ruído de mesma intensidade for considerado várias vezes.

4.3 Escolha das Variáveis

Dentre o universo de todas as variáveis possíveis, a primeira escolha de variáveis deve ser realizada pelo pesquisador, que com conhecimento suficiente no assunto estudado, pode listar as variáveis que suspeita-se ter influência no fenômeno.

Para o caso específico dos afundamentos de tensão, foi selecionado um conjunto de 9 variáveis. As variáveis e a relação que elas possuem com a frequência de afundamentos de

tensão, são:

- Potência de curto-circuito: esta variável é fruto do circuito equivalente que está a montante do sistema de distribuição estudado, e de forma simplificada mede a robustez do barramento, ou seja, potência de curto circuito alta mostra que o barramento tem menor propensão a sofrer afundamentos de tensão.
- Área de Vulnerabilidade: esta variável define a região do sistema elétrico onde ocorrência de faltas irá acarretar afundamento de tensão no barramento principal da subestação. Para cada tipo de falta (monofásica, bifásica, bifásica-terra e trifásica), há uma área de vulnerabilidade atrelada. Para o cálculo da área de vulnerabilidade equivalente, faz-se o somatório da multiplicação da área de vulnerabilidade relacionada a cada tipo de falta pela sua respectiva probabilidade de ocorrência.
- Taxa de Falta: esta variável quantifica o número de faltas que irá ocorrer em determinado comprimento de rede em um período de tempo, e normalmente é expressa em valores anuais.
- Área de Vulnerabilidade x Taxa de Falta: esta variável que de fato é uma combinação de duas variáveis, é a que se espera possuir maior relação com o número de afundamentos de tensão percebidos no barramento principal da subestação, pois toda falta que ocorrer dentro dos limites da área de vulnerabilidade causa afundamento de tensão na barra monitorada.
- Quilometragem da rede: esta variável possui relação direta com o número de faltas que ocorre em determinado sistema de distribuição, pois considerando uma taxa de falta constante, quanto maior o comprimento da rede maior a quantidade de faltas.

4.4 Seleção das Variáveis

Após a escolha das variáveis feita pelo pesquisador, é necessário averiguar se as variáveis escolhidas realmente possuem relação com o fenômeno, ou se não passa de uma suspeita infundada. Para realizar tal análise propõem-se que seja utilizado o coeficiente de correlação de Pearson entre as variáveis e a resposta, que nesta dissertação corresponde ao número de afundamentos de tensão. Esta etapa da metodologia é fundamental, pois, reduz a dimensionalidade do conjunto de variáveis que serão a entrada dos algoritmos de formação de conjuntos.

4.4.1 Coeficiente de Correlação Linear de Pearson

A análise de correlação é uma técnica importante para as diferentes áreas do conhecimento, não somente como resultado final, mas como uma das etapas para a utilização de outras técnicas de análise. Dentre as principais técnicas que utilizam o Coeficiente de Correlação estão a Análise de Confiabilidade, a Análise da Estrutura de Variância-Covariância e o Teste de Normalidade ou Gaussianidade.

O resultado da análise de correlação é um número, que evidencia a direção e a intensidade da relação linear ou não-linear entre duas variáveis. A vantagem desta análise é permitir observar a existência ou não de relação entre variáveis, sem que seja necessário ajustar uma função matemática. Uma propriedade desta técnica é que não há diferença entre a variável resposta e a explicativa, assim o coeficiente obtido da correlação X e Y é igual ao coeficiente obtido fazendo a correlação Y e X.

Interessante ressaltar que considerando um par de variáveis, existe a possibilidade do coeficiente de correlação apresentar forte correlação numérica, porém sabendo o pesquisador a priori que estas variáveis não possuem qualquer relação causal, ou seja, o número obtido é mera coincidência, neste caso a correlação entre as variáveis é espúria. Por exemplo, a correlação entre a variável mesada mensal em reais que apresenta valores 100, 200, 250 e 150 e a variável precipitação no sudeste em milímetros que apresenta valores 10, 20, 25 e 15 é perfeita, ou seja, coeficiente igual a 1, porém sabe-se que não há nenhuma relação de fato entre as variáveis envolvidas.

Segundo (LIRA, 2004), para o desenvolvimento teórico da análise de correlação, são feitas determinadas suposições sobre as variáveis envolvidas na análise e quando as suposições não forem atendidas para a análise de correlação, são possíveis os seguintes procedimentos, utilização de métodos não-paramétricos ou adequação dos dados às suposições através de uma transformação das variáveis envolvidas na análise.

O coeficiente de correlação é a medida que mostra o grau de relacionamento entre duas variáveis, tal medida é também conhecida por medida de intercorrelação, interdependência, relação ou associação.

Supondo duas variáveis X e Y, existem formas diferentes de correlação que podem existir entre elas, o caso mais comum e normalmente o primeiro que se investiga é a correlação simples. Quando uma variável puder ser obtida aproximadamente por uma equação de reta que envolva outra variável, diz-se que a relação entre elas é linear, então pode-se ajustar uma reta

da forma $Y=\alpha+\beta X$ aos dados. No entanto, caso não se verifique relação linear entre duas variáveis, não necessariamente significa que não exista correlação entre estas, visto que poderá haver correlação não-linear entre as mesmas.

A melhor forma de analisar o tipo de correlação entre variáveis é através do gráfico denominado diagrama de dispersão, que consiste em um plano cartesiano onde os eixos correspondem as duas variáveis em estudo e são plotados os pares (X, Y) de todas observações feitas. Nas Figuras 4.1, 4.2, 4.3 e 4.4 são mostrados diagramas de dispersão com os principais tipos de correlação entre duas variáveis.

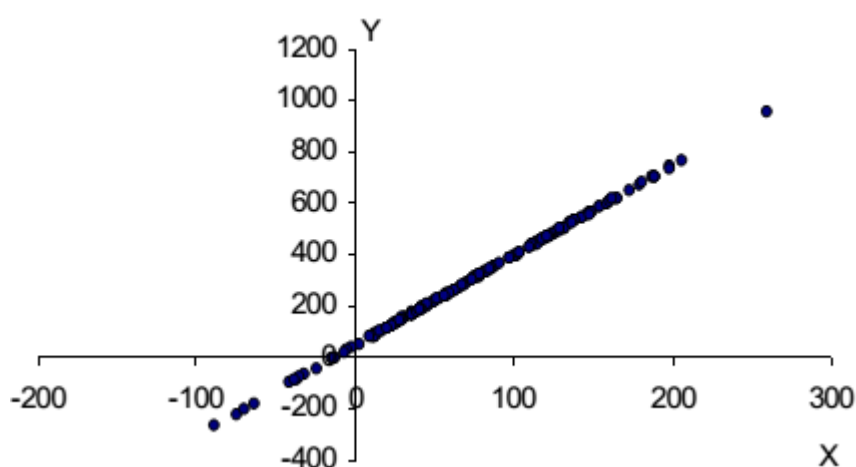


Figura 4.1 – Correlação linear positiva perfeita entre as variáveis X e Y (LIRA, 2004).

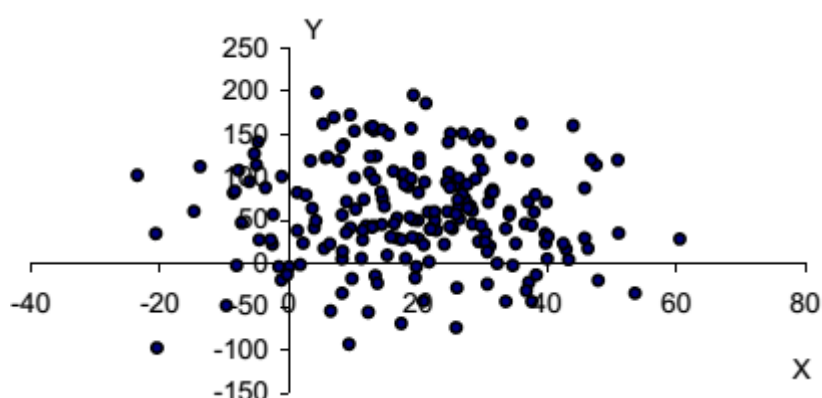


Figura 4.2 – Correlação linear nula entre as variáveis X e Y (LIRA, 2004).

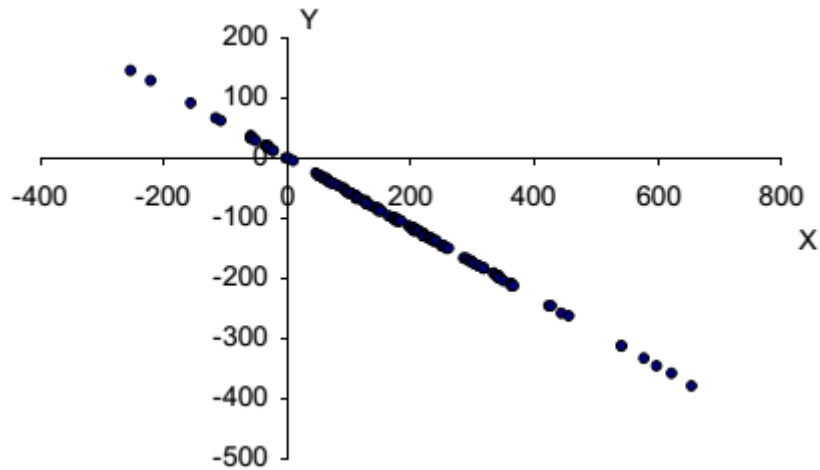


Figura 4.3 – Correlação linear negativa perfeita entre as variáveis X e Y (LIRA, 2004).

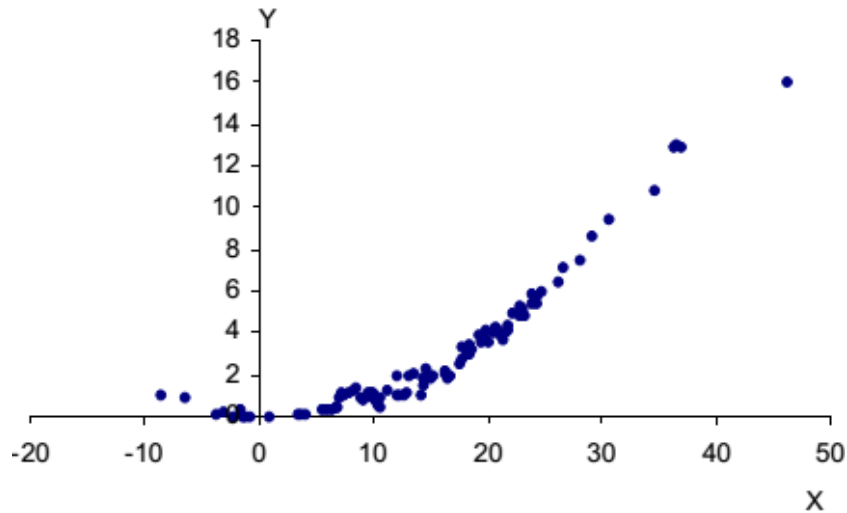


Figura 4.4 – Correlação não-linear entre as variáveis X e Y (LIRA, 2004).

O método mais conhecido e utilizado para quantificar a correlação entre duas variáveis é o coeficiente de correlação linear de Pearson, que também recebe o nome de coeficiente de correlação do momento produto ou simplesmente ρ . Há inúmeras aplicações para este coeficiente, dentre elas a Análise Fatorial, Análise de Componentes Principais, Análise de Confiabilidade, etc.

O coeficiente de correlação ρ é definido pela equação (4.1)

$$\rho = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (4.1)$$

Este coeficiente é adimensional e assume valores entre -1 e +1, $\rho = 1$ significa uma correlação perfeita entre as duas variáveis, $\rho = -1$ evidencia uma correlação negativa perfeita, ou seja, as variáveis são inversamente proporcionais, quando uma aumenta a outra diminui e $\rho=0$ indica que as duas variáveis não dependem linearmente uma da outra, mas não exclui a possibilidade de haver dependência não-linear entre elas, portanto quando ocorrer este resultado é necessário investigar outros tipos de relação entre as variáveis.

Para (COHEN, 1988), valores entre 0,10 e 0,29 podem ser considerados pequenos; escores entre 0,30 e 0,49 podem ser considerados como médios; e valores entre 0,50 e 1 podem ser interpretados como grandes. (DANCEY; REIDY, 2006) apontam para uma classificação ligeiramente diferente: $\rho = 0,10$ até 0,30 (fraco); $\rho = 0,40$ até 0,6 (moderado); $\rho = 0,70$ até 1 (forte). Seja como for, o certo é que quanto mais perto de 1 (independente do sinal) maior é o grau de dependência estatística linear entre as variáveis. No outro oposto, quanto mais próximo de zero, menor é a força dessa relação (FIGUEIREDO; SILVA, 2009).

Para correta utilização da correlação de Pearson é preciso atender algumas condições, primeiramente as duas variáveis devem possuir relacionamento linear, outra condição é que as variáveis sejam aleatórias e quantitativas (discretas ou contínuas), uma terceira condição é que os valores observados estejam normalmente distribuídos.

Para exemplificar o cálculo do coeficiente de correlação de Pearson, suponha as variáveis X e Y, apresentadas na Tabela 4.1.

Tabela 4.1 – Valores das variáveis X e Y.

X	Y
30	0,50
39	1,57
55	1,72
56	1,79
70	2,95

Primeiramente, verifica-se que ambas variáveis são quantitativas e contínuas e também possuem distribuição normal, então calcula-se a média de cada variável, o numerador e o denominador da expressão de ρ :

$$\bar{X} = \frac{30 + 39 + 55 + 56 + 70}{5} = \frac{250}{5} = 50$$

$$\bar{Y} = \frac{0,5 + 1,57 + 1,72 + 1,79 + 2,95}{5} = \frac{8,53}{5} = 1,71$$

$$\sum_{i=1}^5 (X_i - \bar{X})^2 = (30 - 50)^2 + (39 - 50)^2 + (55 - 50)^2 + (56 - 50)^2 + (70 - 50)^2 = 982$$

$$\sum_{i=1}^5 (Y_i - \bar{Y})^2 = (0,5 - 1,706)^2 + (1,57 - 1,706)^2 + (1,72 - 1,706)^2 + (1,79 - 1,706)^2 \\ + (2,95 - 1,706)^2 = 3,02772$$

$$\sum_{i=1}^5 (X_i - \bar{X})(Y_i - \bar{Y}) = (30 - 50)(0,5 - 1,706) + (39 - 50)(1,57 - 1,706) \\ + (55 - 50)(1,72 - 1,706) + (56 - 50)(1,79 - 1,706) + (70 - 50)(2,95 - 1,706) = 51,07$$

$$\rho = \frac{51,07}{\sqrt{982 \times 3,02772}} = 0,936$$

Como se observa o coeficiente de correlação de Pearson entre as variáveis X e Y é de 0,936, indicando forte relação linear entre estas variáveis.

4.5 Implementação da Formação de Conjuntos

Uma vez selecionada através da análise das correlações de Pearson, as variáveis de entrada para os algoritmos de formação de conjuntos. Deve ser implementada a formação de conjuntos através dos diversos métodos que se deseja testar, os métodos considerados neste trabalho foram o de ligação simples, ligação completa, média das distâncias, Ward e *k-means*, os algoritmos e particularidades de cada um deles foram apresentados no capítulo 3.

Como o número de variáveis e objetos considerados nos problemas de análise multivariada, costumam ser grandes, provavelmente, será necessário a utilização de algum *software*. Nesta pesquisa foi utilizado o MINITAB, mas como já citado, existem diversos outros programas no mercado que possuem a funcionalidade de formação de conjuntos.

4.6 Comparação entre Métodos

Este passo da metodologia visa determinar qual dentre os métodos aplicados é o melhor. O processo consiste em realizar uma análise de variância entre o *membership* gerado pelos métodos e os valores médios de AMTs de cada sistema de distribuição. Assim, o método que apresentar menor variação dentro dos grupos (*within*) e maior variação entre os tratamentos (*between*), ou seja, o método que apresentar maior valor para o teste de Fisher (F), será o melhor a ser aplicado àquela base de dados. Vale salientar que o método selecionado não é o melhor

em todas ocasiões. A conclusão única desta análise é que, para os dados considerados, aquela técnica foi a que apresentou melhor resultado. Sendo assim, se a base de dados for outra, mesmo que o fenômeno sob análise seja o mesmo, deve-se efetuar a análise seguindo os passos da metodologia proposta para verificar qual técnica deve ser empregada.

4.6.1 Análise de Variância

A análise de variância (ANOVA) é uma técnica estatística usada para avaliar afirmações sobre a média de amostras, fundamentalmente, visa verificar se existe diferença significativa entre as médias ou se é uma diferença casual.

As hipóteses nula e alternativa a serem testadas na análise de variância são:

H_0 (hipótese nula): as médias amostrais são iguais.

H_1 (hipótese alternativa): as médias amostrais são diferentes, ou seja, pelo menos uma das médias é diferente das demais.

Supondo que se deseja comparar as médias das amostras A, B e C apresentadas na Tabela 4.2.

Tabela 4.2 – Dados das amostras A, B e C.

	A	B	C
	5	9	10
	4	1	5
	6	8	8
	7	11	7
	8	6	10
Somatório	30	35	40
Médias	$\bar{X}_A = 6$	$\bar{X}_B = 7$	$\bar{X}_C = 8$

Inicialmente, calcula-se a variação total (SS: *Sum Squares*) que corresponde ao somatório do quadrado das diferenças de cada observação em relação à média geral. Efetuando este cálculo temos a Tabela 4.3.

$$\bar{X} \text{ (Média Geral)} = \frac{5 + 4 + 6 + 7 + 8 + 9 + 1 + 8 + 11 + 6 + 10 + 5 + 8 + 7 + 10}{15} = 7$$

Tabela 4.3 – Variação total.

X_i	$X_i - \bar{X} = x_i$	x_i^2
5	5-7=-2	4
4	4-7=-3	9
6	6-7=-1	1
7	7-7=0	0
8	8-7=1	1
9	9-7=2	4
1	1-7=-6	36
8	8-7=1	1
11	11-7=4	16
6	6-7=-1	1
10	10-7=3	9
5	5-7=-2	4
8	8-7=1	1
7	7-7=0	0
10	10-7=3	9
$\Sigma 105$	$\Sigma 0$	$\Sigma 96$

Em seguida calcula-se a variação dentro da amostra – *Within* (SSW), que corresponde ao somatório do quadrado das diferenças de cada observação em relação à média da amostra. Tal cálculo é explicitado na Tabela 4.4.

Tabela 4.4 – Variação dentro das amostras.

X_A	$X_A - \bar{X}_A$	$(X_A - \bar{X}_A)^2$	X_B	$X_B - \bar{X}_B$	$(X_B - \bar{X}_B)^2$	X_C	$X_C - \bar{X}_C$	$(X_C - \bar{X}_C)^2$
5	5-6=-1	1	9	9-7=2	4	10	10-8=2	4
4	4-6=-2	4	1	1-7=-6	36	5	5-8=-3	9
6	6-6=0	0	8	8-7=1	1	8	8-8=0	0
7	7-6=1	1	11	11-7=4	16	7	7-8=-1	1
8	8-6=2	4	6	6-7=-1	1	10	10-8=2	4
		10			58			18

Portanto $SSW = 10+58+18 = 86$. Posteriormente realiza-se o cálculo da variação entre as amostras – *Between* (SSB), que corresponde ao somatório das diferenças da média da amostra em relação à média geral, conforme mostra a Tabela 4.5.

Tabela 4.5 – Variação entre amostras.

\bar{X}_A	$\bar{X}_A - \bar{X}$	$(\bar{X}_A - \bar{X})^2$	\bar{X}_B	$\bar{X}_B - \bar{X}$	$(\bar{X}_B - \bar{X})^2$	\bar{X}_C	$\bar{X}_C - \bar{X}$	$(\bar{X}_C - \bar{X})^2$
6	6-7=-1	1	7	7-7=0	0	8	8-7=1	1
6	6-7=-1	1	7	7-7=0	0	8	8-7=1	1
6	6-7=-1	1	7	7-7=0	0	8	8-7=1	1
6	6-7=-1	1	7	7-7=0	0	8	8-7=1	1
6	6-7=-1	1	7	7-7=0	0	8	8-7=1	1
		5			0			5

Portanto $SSB = 5+0+5 = 10$ e nota-se que $SS = SSW + SSB = 86 + 10 = 96$. Os graus de liberdade – *Degree of Freedom* (DF) são a quantidade de informação que seus dados fornecem que você pode "gastar" para estimar os valores de parâmetros populacionais desconhecidos, e calcular a variabilidade dessas estimativas. Esse valor é determinado pelo número de observações em sua amostra e o número de parâmetros em seu modelo.

Aumentar seu tamanho amostral fornece mais informações sobre a população e, desta forma, aumenta os graus de liberdade em seus dados. Adicionar parâmetros ao seu modelo (aumentando o número de termos em uma equação de regressão, por exemplo) "gasta" informações dos seus dados, e reduz os graus de liberdade disponíveis para estimar a variabilidade das estimativas de parâmetro (SUPORTE AO MINITAB).

Como no exemplo realizado estima-se apenas um parâmetro, a média populacional, o tamanho amostral de n constitui n informações para a estimativa da média populacional e sua variabilidade. Um grau de liberdade é gasto estimando-se a média, e os $n-1$ graus de liberdade restantes estimam a variabilidade.

Assim sendo, o grau de liberdade para o total dos dados (DFSS) é $15 - 1 = 14$, o grau de liberdade para variabilidade dentro da amostra (DFSSW) é $(5 - 1) \times 3 = 12$ e o grau de liberdade para a variabilidade entre amostras (DFSSB) é $3 - 1 = 2$. Observa-se que $DFSS = DFSSW + DFSSB = 12 + 2 = 14$.

O cálculo da estatística de Fisher F é dado pela equação (4.2).

$$F = \frac{MS_B}{MS_W} = \frac{\frac{SSB}{DF_{SSB}}}{\frac{SSW}{DF_{SSW}}} = \frac{\frac{10}{2}}{\frac{86}{12}} = \frac{5}{7,17} = 0,70 \quad (4.2)$$

Para determinar qual a hipótese plausível, compara-se o F com o $F_{\text{crítico}}$ que é obtido a partir de uma função densidade de probabilidade, tendo como entradas o grau de liberdade da variabilidade dentro e entre amostras, bem como a nível de confiança. Se $F < F_{\text{crítico}}$ não se rejeita H_0 e o P-valor é extraído desta comparação.

No exemplo considerado para um $\alpha = 5\%$, $F_{\text{crítico}} = 3,89$, portanto F é menor que $F_{\text{crítico}}$ e não se rejeita H_0 o P-valor é igual a 0,517.

4.7 Análise de Robustez dos Métodos

Para verificar a robustez dos métodos de clusterização utilizados, propõem-se aplicar um pequeno ruído sobre uma das variáveis de entrada dos algoritmos e criar vários cenários de perturbação desta variável. Então, implementa-se a formação de conjuntos nos cenários criados, guardando o *membership* de cada método para cada cenário.

Assim, para descobrir o método mais robusto, realiza-se a análise de concordância por atributos, que parte do pressuposto que para detectar se uma dada classificação de um objeto é confiável, é necessário ter este objeto classificado várias vezes e por mais de um avaliador (BALTAR; OKANO). Nota-se que no caso em questão os avaliadores são os métodos de formação de conjuntos, os objetos são os sistemas de distribuição que foram classificados 100 vezes por cada método.

4.8 Considerações Finais

A metodologia proposta neste capítulo é inovadora e possui eficácia em encontrar o método de formação de conjuntos mais adequado a ser empregado. Para a validação da metodologia ora proposta, no capítulo subsequente esta será aplicada em duas bases de dados, uma contendo sistemas de distribuição hipotéticos e outra com sistemas de distribuição reais pertencentes a EDP Escelsa.

5. ESTUDO DE CASO

5.1 Considerações Iniciais

O presente capítulo tem o objetivo de aplicar e validar a metodologia estabelecida no capítulo 4, empregando-a em duas bases de dados distintas, uma referente a sistemas de distribuição hipotéticos e outra relacionada a sistemas de distribuição reais da EDP Escelsa. Todos os passos da metodologia são seguidos ordenadamente como está determinado no item 4.2. Os resultados obtidos mostram que a metodologia é eficaz para a determinação do método mais adequado para as bases de dados consideradas.

5.2 Sistemas de Distribuição Hipotéticos

Para ter completa confiança nos dados utilizados, optou-se por criar 12 sistemas de distribuição concebidos tomando como base o sistema de distribuição de 34 barras disponibilizado pelo IEEE. Para modelar diferentes redes de distribuição, foram utilizados sistemas contendo de 1 a 8 alimentadores semelhantes ao alimentador do caso teste do IEEE, considerou-se também diferentes potências para o transformador principal da subestação de distribuição. A Tabela 5.1 sintetiza as características dos 12 sistemas utilizados, cujos diagramas unifilares estão apresentados no Anexo B.

Tabela 5.1 – Características dos sistemas de distribuição.

Número de Alimentadores	Potência do Transformador (MVA)
1	2,5
	5,0
2	5,0
	7,5
3	7,5
	10,0
4	10,0
	12,5
5	12,5
	15,0
6	15,0
8	20,0

Para simular a diferença de nível cerâmico e de vegetação remanescente em diferentes regiões geográficas, cada um dos doze sistemas discriminados anteriormente foram simulados

considerando taxas de falta de 15, 30 e 60 faltas/100km/ano, gerando assim 36 sistemas de distribuição distintos.

5.2.1 Escolha das Variáveis

Conforme já mencionado no capítulo 4, a metodologia se inicia pela escolha das variáveis, que pela experiência do pesquisador, há uma suspeita que possuam relação com o fenômeno. Nesta etapa, foram selecionadas as variáveis: Potência de Curto-circuito, Área de Vulnerabilidade (monofásica, bifásica, bifásica-terra, trifásica e equivalente), Taxa de Falta, a interação Área de Vulnerabilidade x Taxa de Falta e quilometro de rede.

A Tabela 5.2 mostra os valores utilizados de cada uma das 9 variáveis para os 36 sistemas modelados, e o número médio de AMTs, que foi obtido a partir de simulação de Monte Carlo realizada no *software* ASPEN DistriView. Neste procedimento todas as variáveis envolvidas no cálculo (linha ou trecho sob falta, posição, tipo e resistência de falta), foram sorteadas de acordo com as respectivas distribuições de probabilidade (HONORATO, 2014). Nas simulações foram considerados 100 cenários (anos) de simulação.

Tabela 5.2 – Base de dados dos 36 sistemas de distribuição.

Subestações	Pcc (MVA)	AV Mono (km)	AV Bifásica (km)	AV Bi Terra (km)	AV Trifásica (km)	AV Equiv. (km)	Taxa de Falta	AV x Taxa	km Rede	Média de AMTs
1AL_15_2,5	32	90	90	90	90	90	15	1357	90	14
1AL_15_5	62	70	56	56	50	66	15	997	90	7
1AL_30_2,5	32	90	90	90	90	90	30	2714	90	27
1AL_30_5	62	70	56	56	50	66	30	1994	90	12
1AL_60_2,5	32	90	90	90	90	90	60	5427	90	54
1AL_60_5	62	70	56	56	50	66	60	3988	90	25
2AL_15_5	62	139	112	112	101	133	15	1994	181	14
2AL_15_7,5	78	119	94	94	87	113	15	1699	181	8
2AL_30_5	62	139	112	112	101	133	30	3988	181	28
2AL_30_7,5	78	119	94	94	87	113	30	3398	181	16
2AL_60_5	62	139	112	112	101	133	60	7976	181	59
2AL_60_7,5	78	119	94	94	87	113	60	6796	181	31
3AL_15_7,5	78	178	141	141	131	170	15	2549	271	13
3AL_15_10	100	95	114	114	114	99	15	1483	271	8
3AL_30_7,5	78	178	141	141	131	170	30	5097	271	27
3AL_30_10	100	95	114	114	114	99	30	2965	271	16
3AL_60_7,5	78	178	141	141	131	170	60	10194	271	54
3AL_60_10	100	95	114	114	114	99	60	5930	271	34
4AL_15_10	100	126	152	152	152	132	15	1977	362	12
4AL_15_12,5	121	104	132	132	132	110	15	1655	362	8
4AL_30_10	100	126	152	152	152	132	30	3953	362	24
4AL_30_12,5	121	104	132	132	132	110	30	3311	362	18
4AL_60_10	100	126	152	152	152	132	60	7907	362	47
4AL_60_12,5	121	104	132	132	132	110	60	6622	362	33
5AL_15_12,5	121	130	164	164	164	138	15	2069	452	11
5AL_15_15	140	116	144	144	149	122	15	1831	452	8

5AL_30_12,5	121	130	164	164	164	138	30	4138	452	21
5AL_30_15	140	116	144	144	149	122	30	3662	452	17
5AL_60_12,5	121	130	164	164	164	138	60	8276	452	44
5AL_60_15	140	116	144	144	149	122	60	7324	452	33
6AL_15_15	140	139	173	173	178	146	15	2197	543	10
6AL_30_15	140	139	173	173	178	146	30	4394	543	22
6AL_60_15	140	139	173	173	178	146	60	8789	543	41
8AL_15_20	173	158	180	180	195	163	15	2446	724	10
8AL_30_20	173	158	180	180	195	163	30	4892	724	22
8AL_60_20	173	158	180	180	195	163	60	9783	724	43

5.2.2 Seleção das Variáveis

Para selecionar as variáveis que melhor caracterizam o número de ocorrências do fenômeno, utilizou-se o processo estatístico de correlação de Pearson descrito no capítulo 4, entre as variáveis e o número médio de AMTs, obtendo-se como resultado os valores apresentados na Tabela 5.3.

Tabela 5.3 – Correlação de Pearson entre variáveis e número médio de AMTs (36 sistemas).

	Correlação	<i>P-Value</i>
Pcc	-0,105	0,541
AV Monofásica	0,224	0,190
AV Bifásica	0,110	0,522
AV Bi-Terra	0,110	0,522
AV Trifásica	0,086	0,620
AV Equivalente	0,205	0,229
Taxa de Falta	0,890	0,000
AVeq x Taxa	0,903	0,000
km rede	-0,008	0,965

Pode-se observar que as variáveis Taxa de Falta e AVeq x Taxa, possuem maior correlação com o número de AMTs, e somente estas apresentaram *p-value* satisfatório, ou seja, somente a correlação apresentada por elas são significativas. Portanto, estas são as variáveis selecionadas para realizar a clusterização dos sistemas de distribuição quanto aos afundamentos de tensão. Ressalta-se que, apesar da média de AMTs ser o objetivo final do agrupamento, ela não é usada como atributo no procedimento, pois em tese, o que se pretende é estabelecer agrupamentos de sistemas de distribuição com certas características semelhantes, e como

consequência espera-se possuir desempenho semelhante quanto ao número de AMTs.

5.2.3 Implementação da Formação de Conjuntos

Para facilitar o processo de análise de *cluster* utilizou-se o software (MINITAB, 2010), que possui os algoritmos dos métodos considerados. Então, implementando os algoritmos dos métodos de ligação simples, completa, média das distâncias e Ward, conforme apresentado no capítulo 3, e considerando a formação de dois grupos devido a quantidade reduzida de sistemas de distribuição. Os dendogramas obtidos estão representados respectivamente nas Figuras 5.1, 5.2, 5.3 e 5.4.

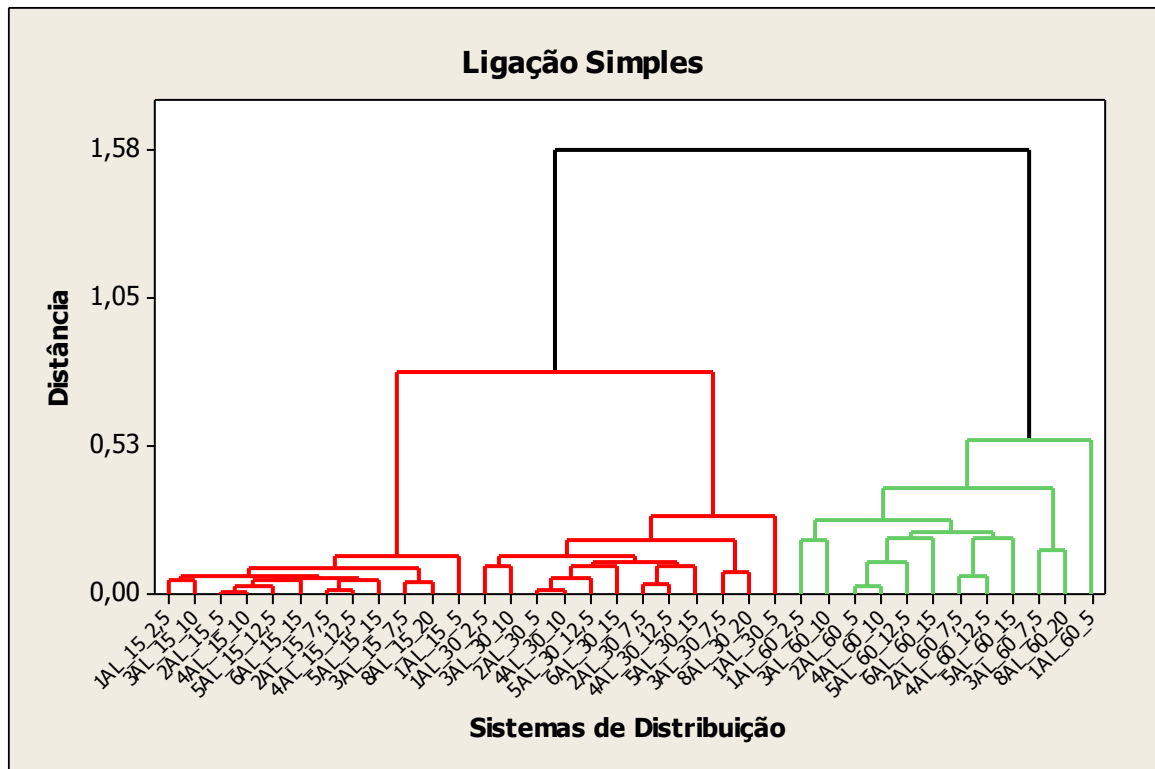


Figura 5.1 – Dendrograma 36 sistemas (Ligação Simples).

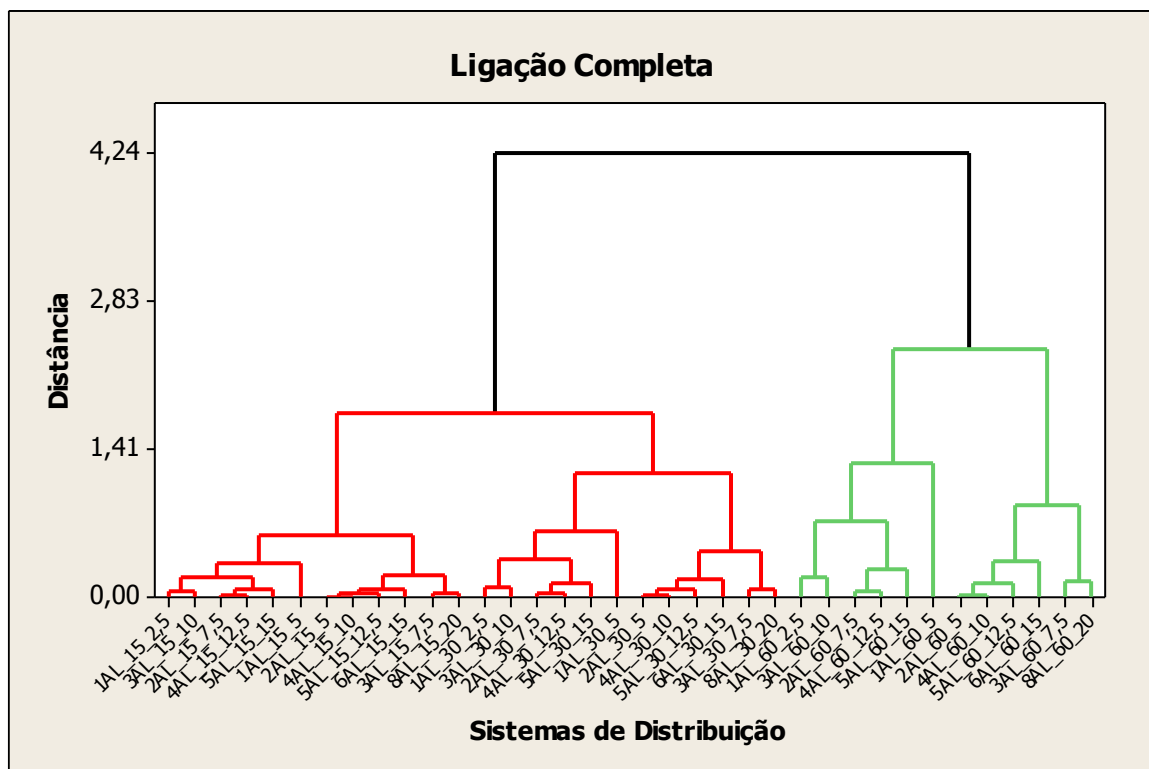


Figura 5.2 – Dendograma 36 sistemas (Ligação Completa).

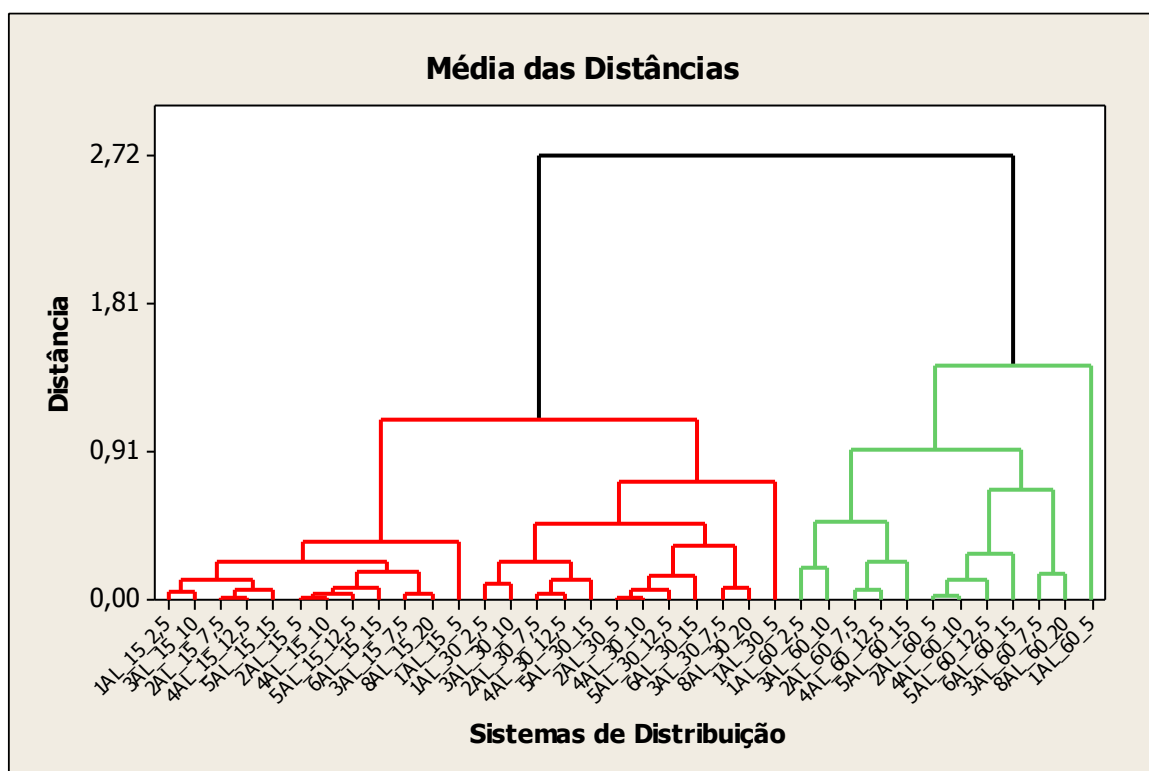


Figura 5.3 – Dendograma 36 sistemas (Média da Distâncias).

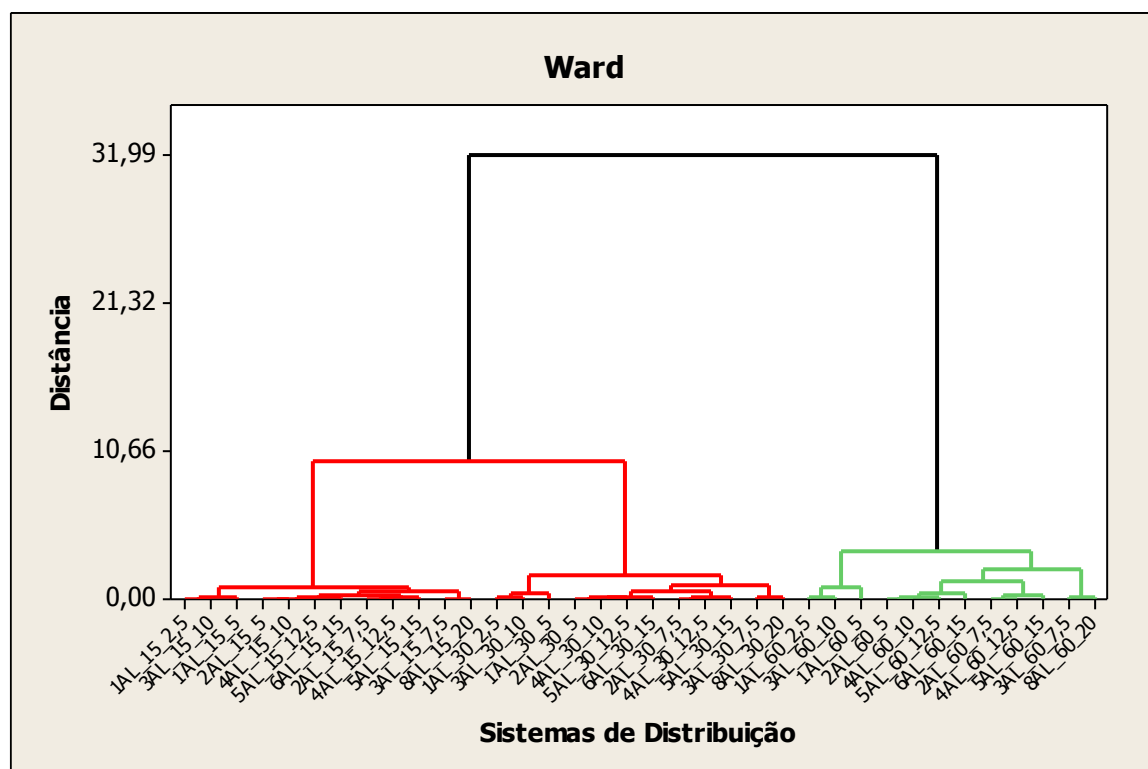


Figura 5.4 – Dendrograma 36 sistemas (Ward).

Nota-se, por uma inspeção visual dos dendrogramas, que os agrupamentos formados pelos métodos são idênticos, mas falta comparar com o método de *k-means*, portanto fez-se a Tabela 5.4 com o *membership* gerado com a implementação de cada algoritmo.

Tabela 5.4 – Comparação *membership* (36 sistemas).

Subestações	Simplex	Completa	Média	Ward	<i>k-means</i>
1AL_15_2,5	1	1	1	1	1
1AL_15_5	1	1	1	1	1
1AL_30_2,5	1	1	1	1	1
1AL_30_5	1	1	1	1	1
1AL_60_2,5	2	2	2	2	2
1AL_60_5	2	2	2	2	2
2AL_15_5	1	1	1	1	1
2AL_15_7,5	1	1	1	1	1
2AL_30_5	1	1	1	1	1
2AL_30_7,5	1	1	1	1	1
2AL_60_5	2	2	2	2	2
2AL_60_7,5	2	2	2	2	2
3AL_15_7,5	1	1	1	1	1
3AL_15_10	1	1	1	1	1
3AL_30_7,5	1	1	1	1	1
3AL_30_10	1	1	1	1	1
3AL_60_7,5	2	2	2	2	2
3AL_60_10	2	2	2	2	2

4AL_15_10	1	1	1	1	1
4AL_15_12,5	1	1	1	1	1
4AL_30_10	1	1	1	1	1
4AL_30_12,5	1	1	1	1	1
4AL_60_10	2	2	2	2	2
4AL_60_12,5	2	2	2	2	2
5AL_15_12,5	1	1	1	1	1
5AL_15_15	1	1	1	1	1
5AL_30_12,5	1	1	1	1	1
5AL_30_15	1	1	1	1	1
5AL_60_12,5	2	2	2	2	2
5AL_60_15	2	2	2	2	2
6AL_15_15	1	1	1	1	1
6AL_30_15	1	1	1	1	1
6AL_60_15	2	2	2	2	2
8AL_15_20	1	1	1	1	1
8AL_30_20	1	1	1	1	1
8AL_60_20	2	2	2	2	2

Pelos valores apresentados na tabela fica evidente que há completa concordância nos agrupamentos formados pelos diversos métodos. Tal concordância se deve em parte ao número reduzido de variáveis selecionadas para serem entrada dos algoritmos de formação de conjuntos.

A etapa seguinte da metodologia de comparação entre métodos se torna desnecessária, pois ela é responsável por determinar qual dentre os métodos gerou melhor agrupamento, e neste caso todos os métodos agruparam os sistemas de distribuição da mesma forma. Portanto, será feita em seguida a etapa de análise de robustez.

5.2.4 Análise de Robustez

Como o exemplo executado teve completa concordância entre os agrupamentos formados por todos os métodos, é apropriado verificar qual é o mais robusto através da aplicação de um pequeno ruído em alguma variável, criando vários cenários de perturbação e analisando em qual dos métodos a concordância nos agrupamento formados é maior.

Então, foram aplicados sucessivamente ruídos normalmente distribuídos com média 0 e desvio padrão igual a 0,03, 0,06, 0,1 e 0,2 na variável AV Eq x Taxa, pois está possui maior correlação com o número médio de AMTs. A intensidade do ruído é aumentada gradualmente para encontrar qual a magnitude necessária para causar “confundimento” nos agrupamentos gerados pelos métodos, neste caso o ruído com média 0 e desvio padrão de 0,2 foi suficiente para atrapalhar as classificações feitas pelos métodos.

Como o número de subestações consideradas é pequeno, o sorteio poderia resultar em um cenário tendencioso, portanto optou-se por criar 100 cenários de perturbação da variável escolhida. Como a análise desejada demanda a implementação de cada um dos cinco métodos estudados 100 vezes, foi criado um programa no *software* (MATLAB, 2011), capaz de realizar a classificação em conjuntos e salvar o *membership* para cada cenário e cada método considerado. O *script* deste programa encontra-se no Anexo C

Fazendo a análise de concordância entre os *membership* gerados pelos métodos nos 100 cenários de perturbação criados, obteve-se o gráfico da Figura 5.5.

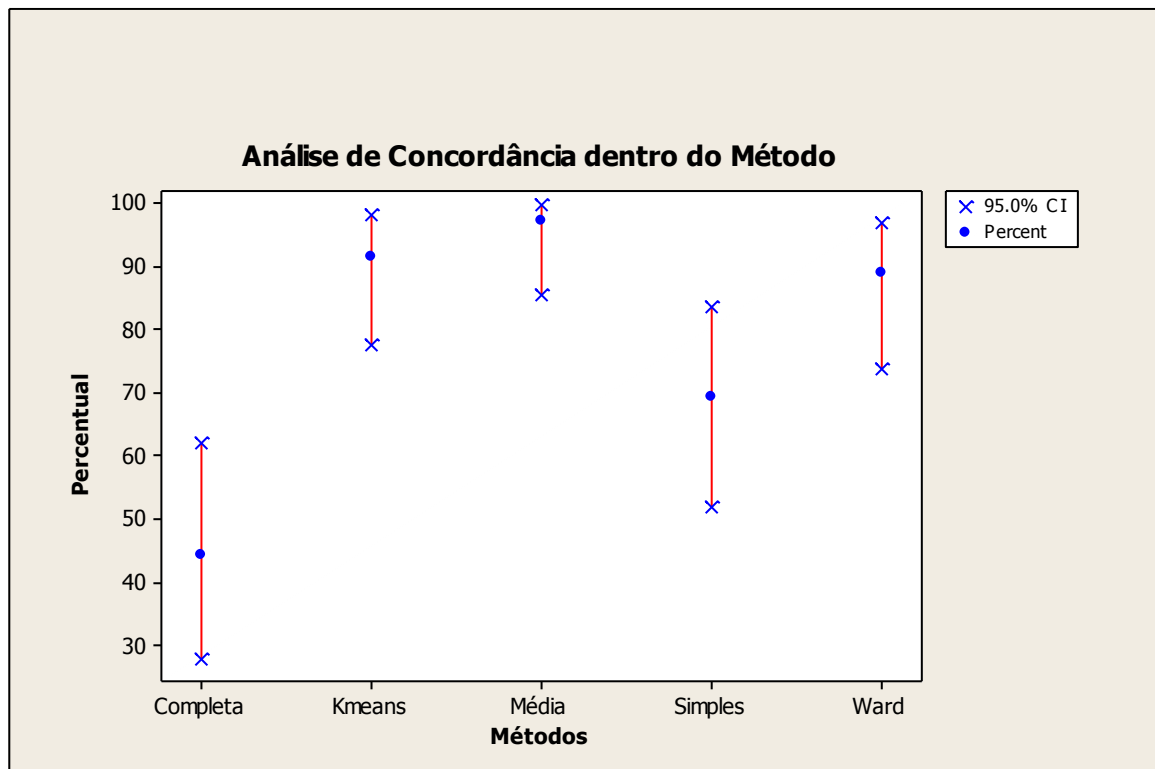


Figura 5.5 – Resultado da análise de concordância para os 36 sistemas de distribuição.

Dentre os métodos, o média das distâncias se mostrou mais robusto, pois apresenta maior concordância, em torno de 97%, com ele mesmo quando aplicado um ruído. Portanto, este é considerado o método mais adequado para formar agrupamentos nesta base de dados.

Conforme citado anteriormente, optou-se pela divisão em dois grupos, um contendo subestações com baixa e o outro com alta frequência de AMTs. Tal separação é confirmada pela Figura 5.6, que apresenta o intervalo de confiança de 95% para o número médio de AMTs de cada grupo formado pelo método da média das distâncias. Observa-se que o número médio de AMTs dos grupos são de aproximadamente 15 e 40, e não há sobreposição dos intervalos de confiança dos grupos.

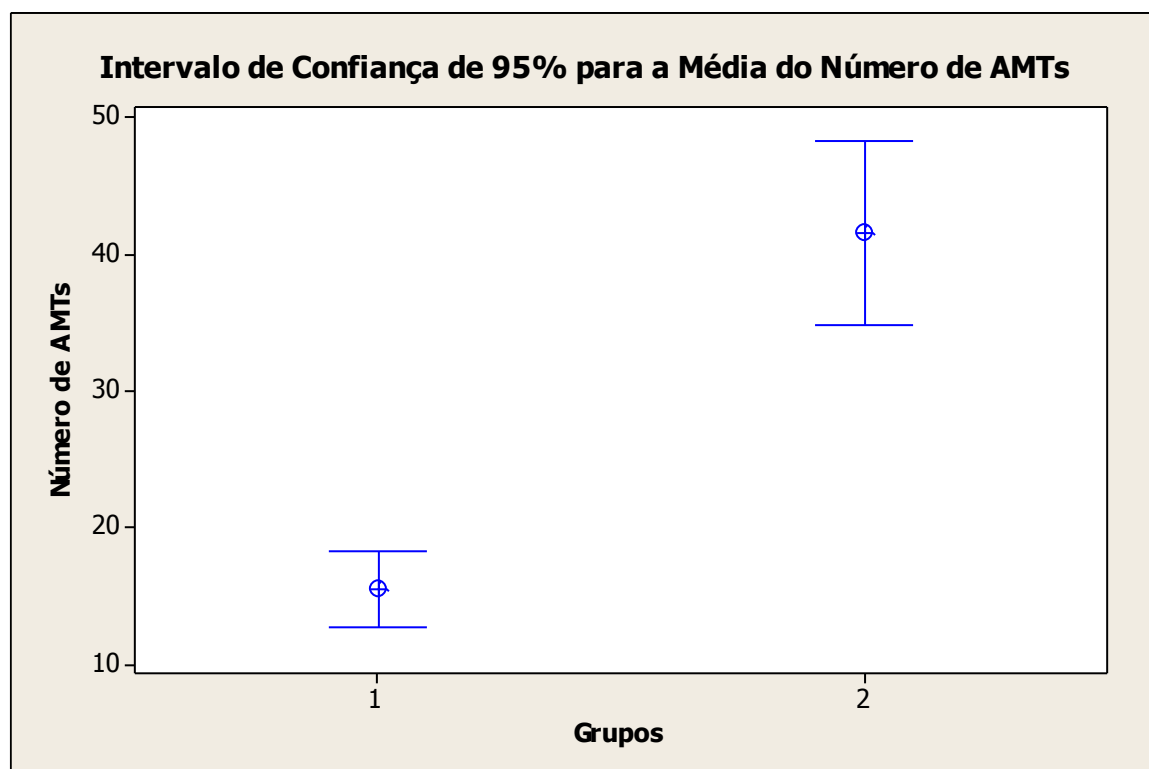


Figura 5.6 – Intervalo de confiança dos grupos quanto aos AMTs (Média das Distâncias).

5.3 Sistemas de Distribuição Reais (EDP Escelsa)

Considerou-se os dados de 15 sistemas de distribuição pertencentes a EDP Escelsa (Aracruz, Baixo Guandu, Barra do Sahy, Ecoporanga, Itarana, Jaguaré, João Neiva, Juncado, Linhares A, Linhares C, Nova Venécia, Paulista, Pinheiros, Santa Teresa e São Francisco), que estão situados ao norte do estado do Espírito Santo, como indica a Figura 5.7.



Figura 5.7 – Distribuição geográfica dos sistemas de distribuição.

Modelou-se os respectivos sistemas no software ASPEN DistriView, realizou-se simulações de curtos-circuitos baseadas em Monte Carlo, para representar cenários de faltas referentes a 100 anos de operação. A média do número de afundamentos de tensão obtida a partir da simulação foi utilizada nas análises.

5.3.1 Escolha das Variáveis

Como o fenômeno estudado é o mesmo (afundamentos de tensão), as variáveis escolhidas pelo pesquisador são idênticas as utilizadas na aplicação dos sistemas de distribuição hipotéticos. A Tabela 5.5 mostra os valores das 9 variáveis consideradas, bem como do número médio de AMTs de cada sistema de distribuição.

Tabela 5.5 – Base de dados da EDP Escelsa.

Subestações	Pcc (MVA)	AV Mono (km)	AV Bifásica (km)	AV Bi Terra (km)	AV Trifásica (km)	AV Equiv. (km)	Taxa de Falta	AV x Taxa	km Rede	Média de AMTs
Aracruz	75	78	63	63	63	74	235	17467	365	61
Baixo Guandu	155	63	116	115	100	74	126	9376	886	44
Barra Sahy	152	31	80	79	65	41	404	16713	165	77
Ecoporanga	39	201	82	82	78	174	90	15699	1087	117
Itarana	130	120	220	213	176	140	192	26804	1685	114
Jaguaré	184	84	134	130	115	94	206	19305	863	89
João Neiva	210	196	93	124	82	176	212	37266	903	114
Juncado	75	229	281	281	267	240	184	44222	625	172
Linhares A	178	145	256	257	245	170	279	47304	1800	169
Linhares C	180	60	103	103	95	69	474	32654	323	97
Nova Venecia	166	186	230	227	205	195	174	33889	1893	124
Paulista	66	130	57	55	53	114	97	11047	244	94
Pinheiros	99	165	214	210	188	175	148	25894	1173	135
Santa Tereza	87	166	154	141	109	160	203	32556	1137	109
São Francisco	81	304	243	248	235	291	159	46241	1442	165

5.3.2 Seleção das Variáveis

Seguindo a metodologia proposta no capítulo 4, o coeficiente de correlação de Pearson foi utilizado para selecionar as variáveis que descrevem melhor o distúrbio de afundamentos de tensão, obtendo-se como resultado os valores apresentados na Tabela 5.6.

Tabela 5.6 – Correlação de Pearson entre variáveis e número médio de AMTs (Escelsa).

	Correlação	<i>P-Value</i>
Pcc	-0,150	0,595
AV Monofásica	0,785	0,001

AV Bifásica	0,800	0,000
AV Bi-Terra	0,826	0,000
AV Trifásica	0,834	0,000
AV Equivalente	0,857	0,000
Taxa de Falta	-0,126	0,655
AVeq x Taxa	0,846	0,000
km rede	0,536	0,039

Pode-se observar que as variáveis AVeq x Taxa, km rede e todas as áreas de vulnerabilidade, possuem boa correlação com o número de AMTs e apresentaram *p-value* inferior ao limite de 0,05.

Para que o número de variáveis de entrada nos algoritmos não seja grande, dificultando a assertividade dos métodos, optou-se por selecionar as variáveis AVeq x Taxa, km rede e AV Equivalente, pois está é calculada levando em conta as demais áreas de vulnerabilidade e a probabilidade de ocorrência de cada tipo de falta.

5.3.3 Implementação da Formação de Conjuntos

Considerando as variáveis selecionadas no passo anterior da metodologia (AVeq x Taxa, km rede e AV Equivalente) e implementando os algoritmos para formação de conjuntos, obtém-se os dendogramas para os métodos de ligação simples, completa, média das distâncias e Ward mostrados, respectivamente, nas Figuras 5.8, 5.9, 5.10 e 5.11.

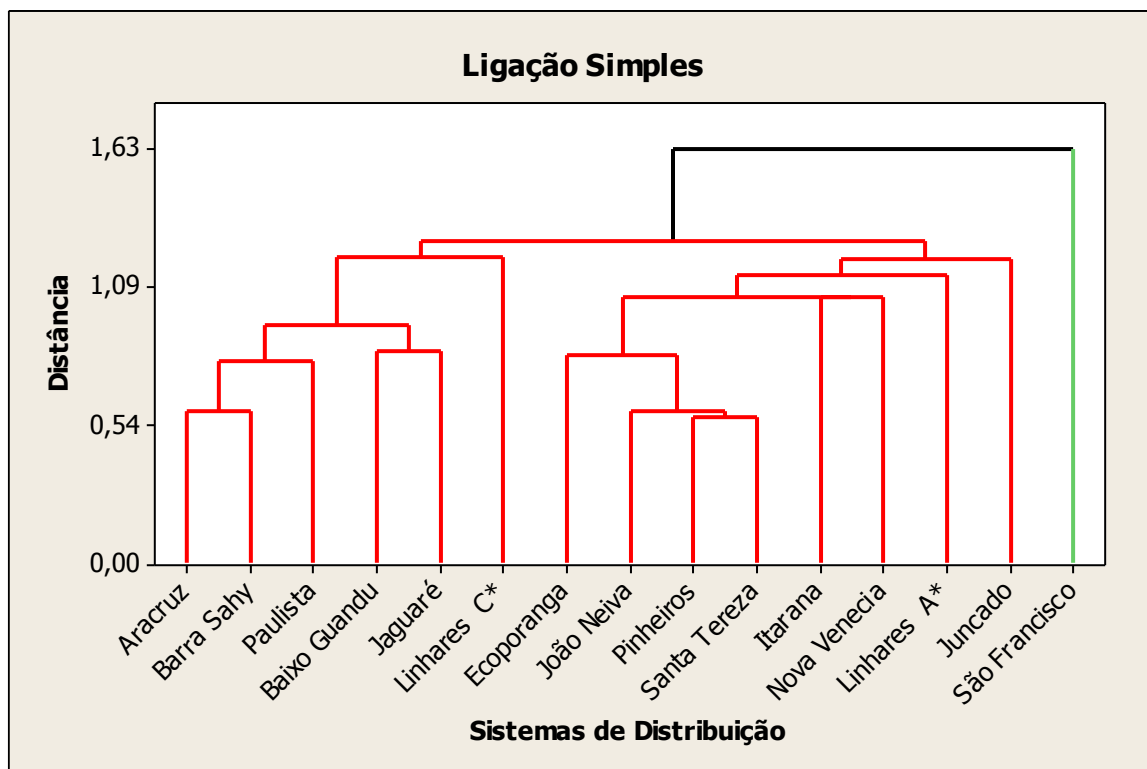


Figura 5.8 – Dendrograma Escelsa (Ligação Simples).

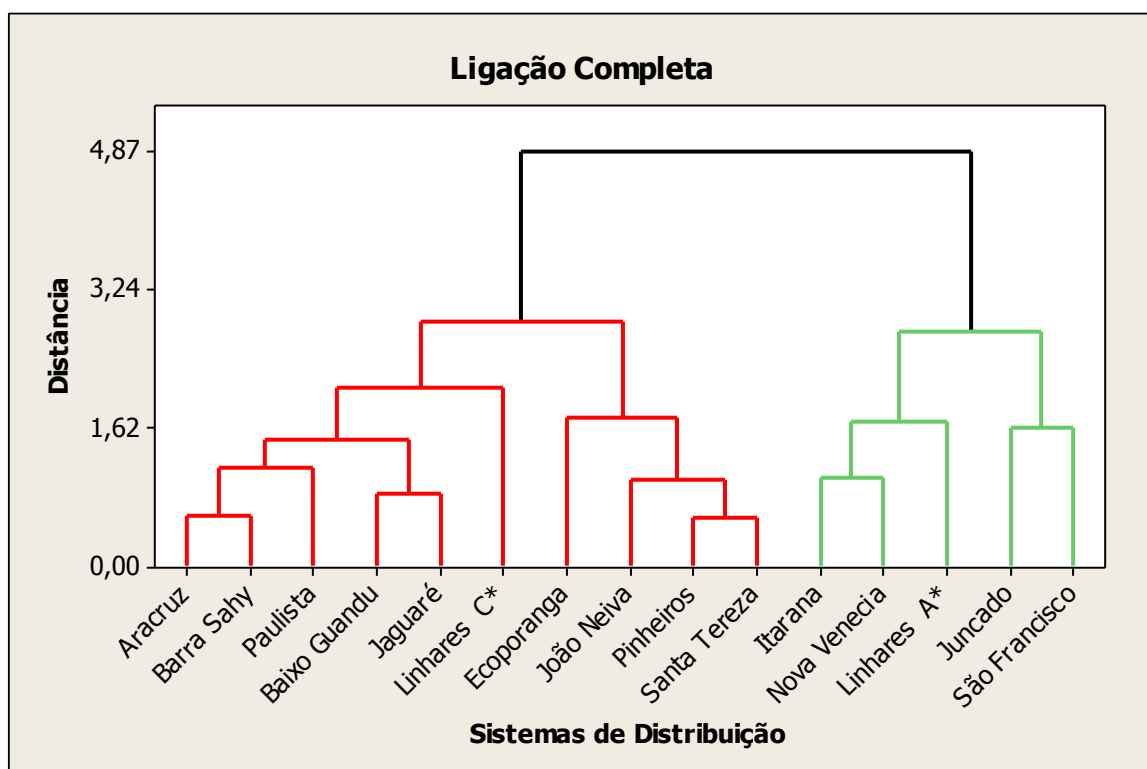


Figura 5.9 – Dendrograma Escelsa (Ligação Completa).

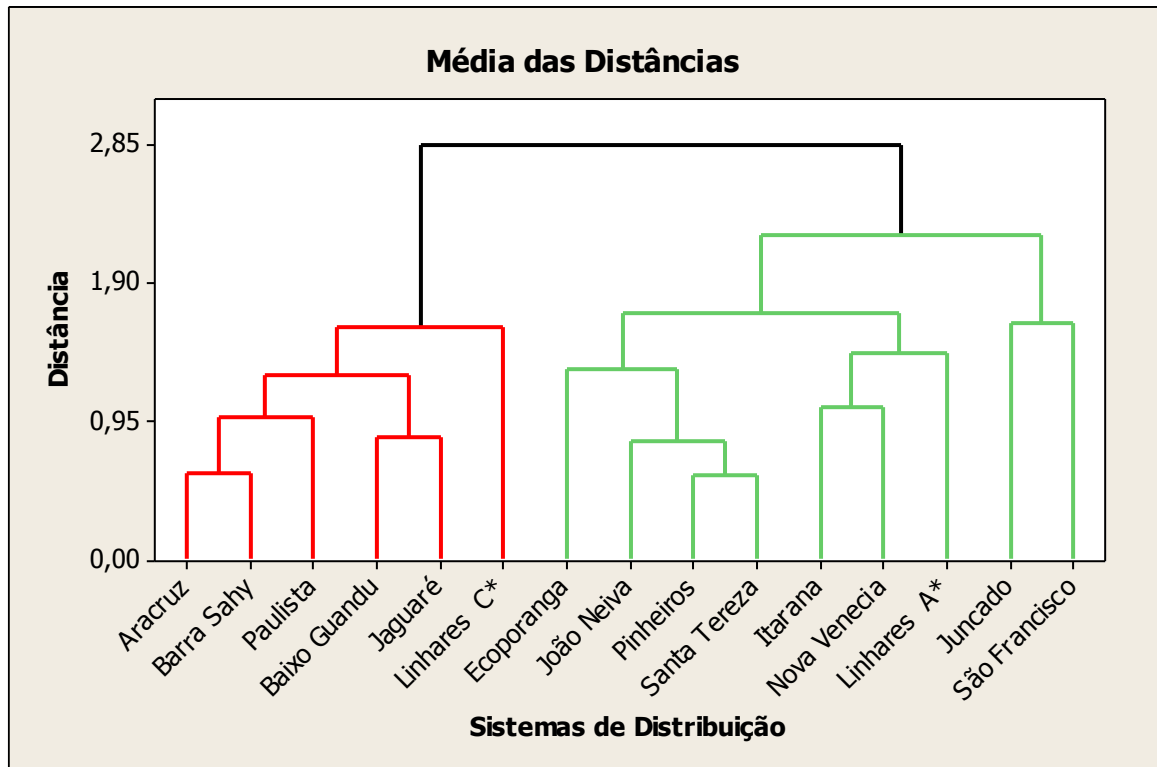


Figura 5.10 – Dendrograma Escelsa (Média das Distâncias).

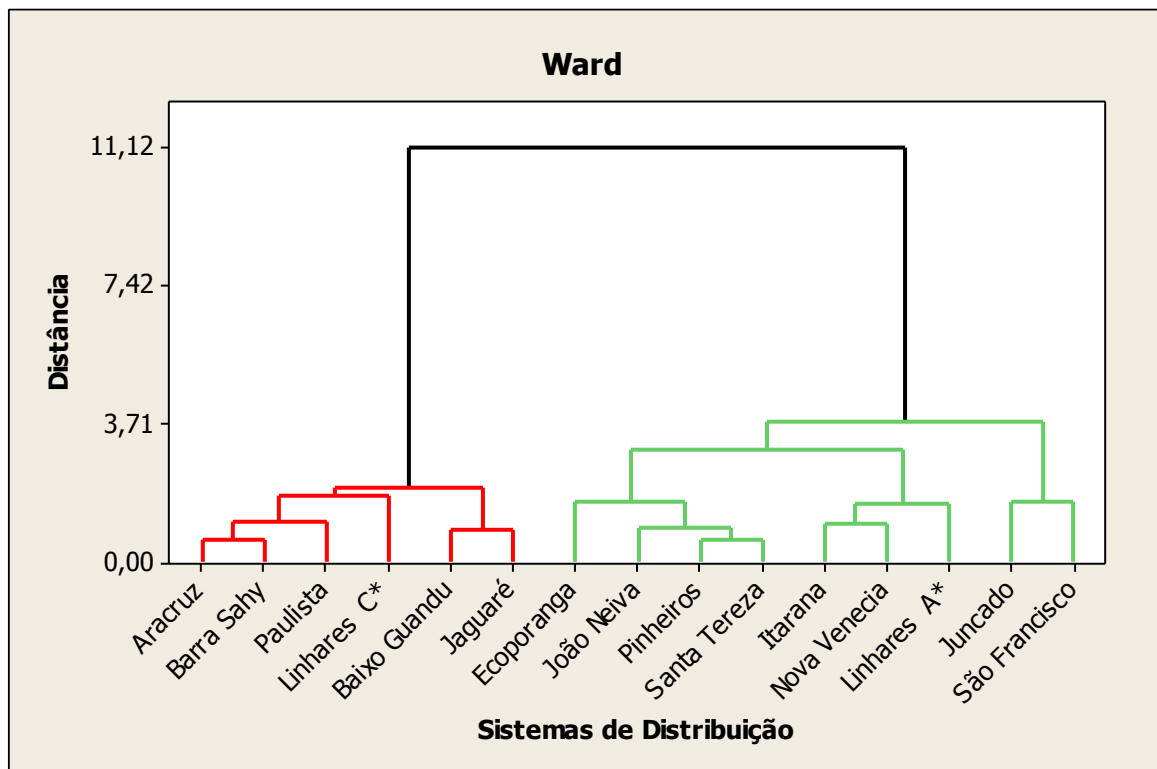


Figura 5.11 – Dendrograma Escelsa (Ward).

A inspeção visual dos dendogramas, aponta para uma coincidência nos agrupamentos formados pelos métodos da média das distâncias e Ward.

O método de *k-means* como já explicado no capítulo 3, não gera como saída um dendograma, sendo que a visualização do agrupamento formado é feita através do *membership* correspondente. A Tabela 5.7, mostra os valores de *membership* obtidos em cada método de formação de conjuntos utilizado.

Tabela 5.7 – Comparação *membership* (Escelsa).

Subestações	Simple	Completa	Média	Ward	<i>k-means</i>
Aracruz	1	1	1	1	1
Baixo Guandu	1	1	1	1	1
Barra Sahy	1	1	1	1	1
Ecoporanga	1	1	2	2	1
Itarana	1	2	2	2	2
Jaguare	1	1	1	1	1
João Neiva	1	1	2	2	2
Juncado	1	2	2	2	2
Linhares A*	1	2	2	2	2
Linhares C*	1	1	1	1	1
Nova Venecia	1	2	2	2	2
Paulista	1	1	1	1	1
Pinheiros	1	1	2	2	2
Santa Tereza	1	1	2	2	2
São Francisco	2	2	2	2	2

Nota-se que há boa concordância entre o *membership* gerado pelos métodos de ligação completa, média das distâncias, Ward e *k-means*, e há divergência destes em relação ao método de ligação simples.

5.3.4 Comparação entre Métodos

Para comparar os agrupamentos gerados pelos diversos métodos e para responder qual deles gerou melhor agrupamento, como propõe a metodologia deve ser feita a análise de variância entre o número médio de AMTs e o *membership* gerado por cada um dos métodos. Efetuando esta análise, os resultados obtidos são apresentados nas Figuras 5.12 a 5.16.

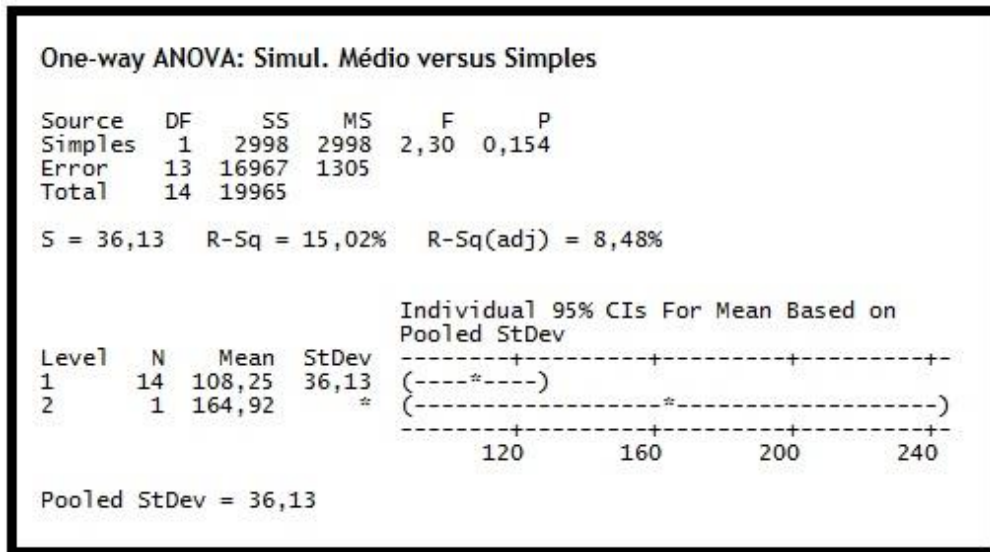


Figura 5.12 – Análise de variância para o método de ligação simples (Escelsa).

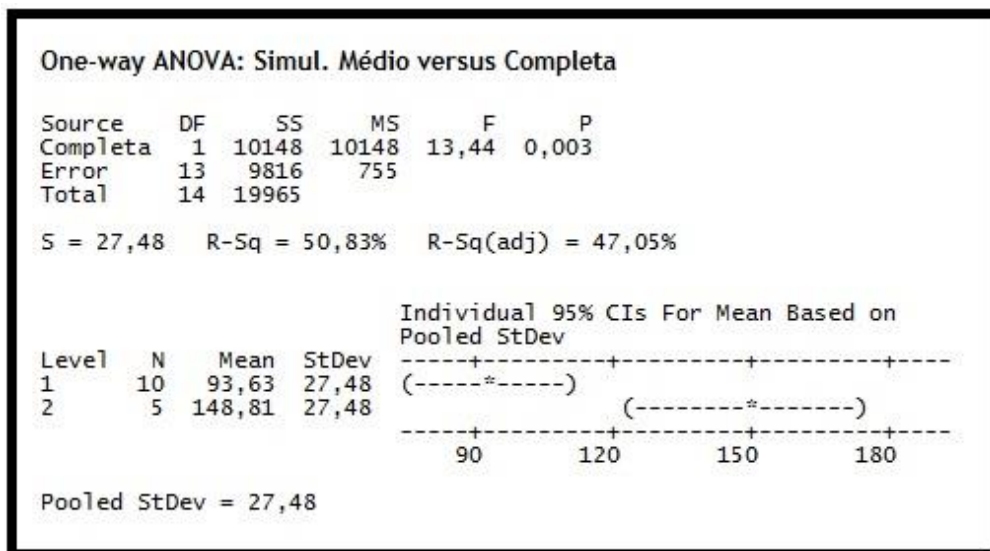


Figura 5.13 – Análise de variância para o método de ligação completa (Escelsa).

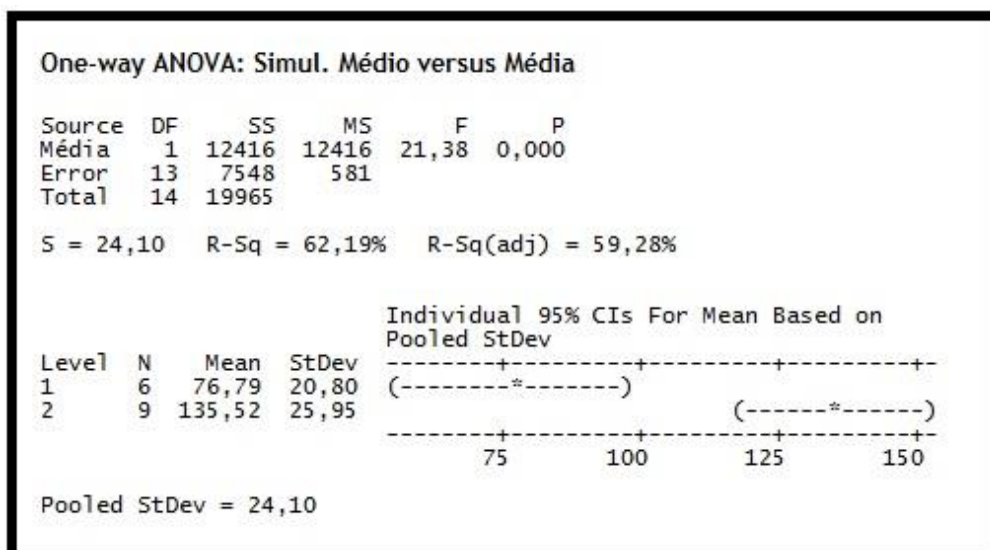


Figura 5.14 – Análise de variância para o método da média das distâncias (Escelsa).

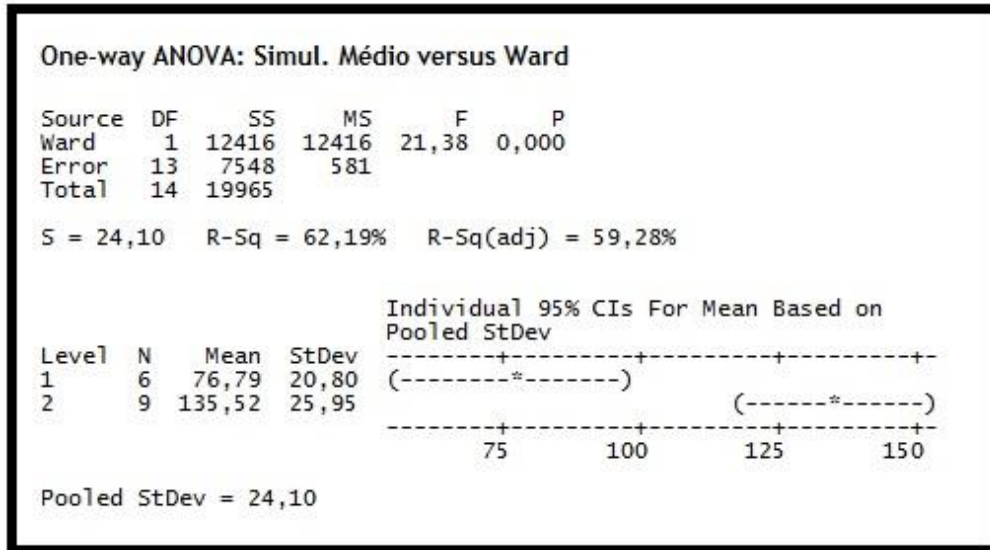
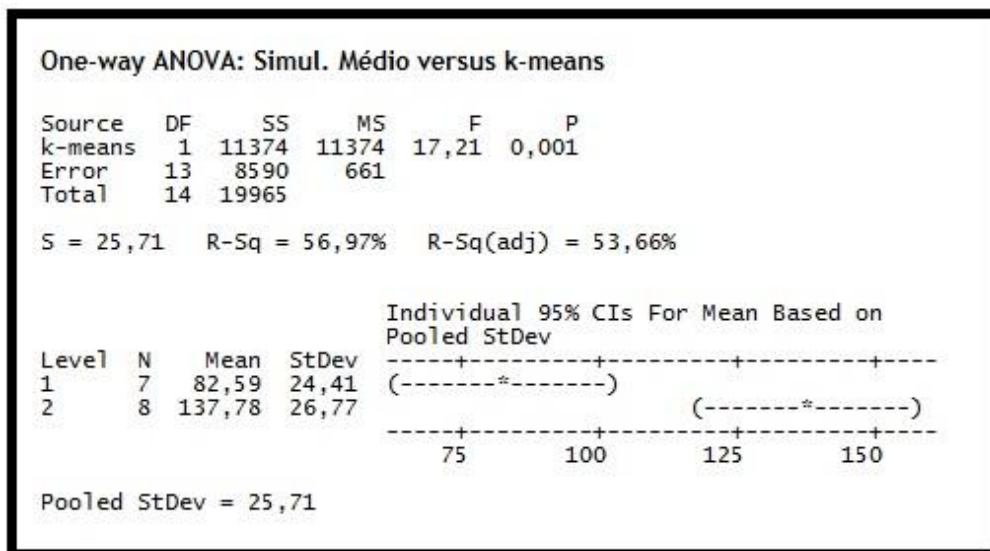


Figura 5.15 – Análise de variância para o método de Ward (Escelsa).

Figura 5.16 – Análise de variância para o método de *k-means* (Escelsa).

Pelos resultados da análise de variância, os métodos da média das distâncias e Ward, apresentaram maior valor para estatística $F=21,38$, portanto são os métodos que geraram melhores agrupamentos.

5.3.5 Análise de Robustez

Como salientado no passo 5 da metodologia proposta no capítulo 4, quando o resultado da análise de variância aponta como melhor mais de um método de formação de conjuntos, é conveniente aplicar um pequeno ruído criando vários cenários de perturbação em alguma variável e analisar em qual método a concordância nos agrupamento formados é maior.

Procedendo desta maneira, foi aplicado sucessivamente ruídos normalmente

distribuídos com média 0 e desvio padrão igual a 0,03, 0,06 e 0,1 na variável AV Equivalente, pois está possui maior correlação com o número médio de AMTs. A intensidade do ruído é aumentada gradualmente para encontrar qual a magnitude necessária para causar “confundimento” nos agrupamentos gerados pelos métodos, neste caso o ruído com média 0 e desvio padrão de 0,1 foi suficiente para atrapalhar as classificações feitas pelos métodos.

Fazendo a análise de concordância entre os *membership* gerados pelos métodos da média das distâncias e Ward que no estágio anterior na metodologia foram considerados os métodos que geraram melhor agrupamento, obteve-se o gráfico da Figura 5.17.

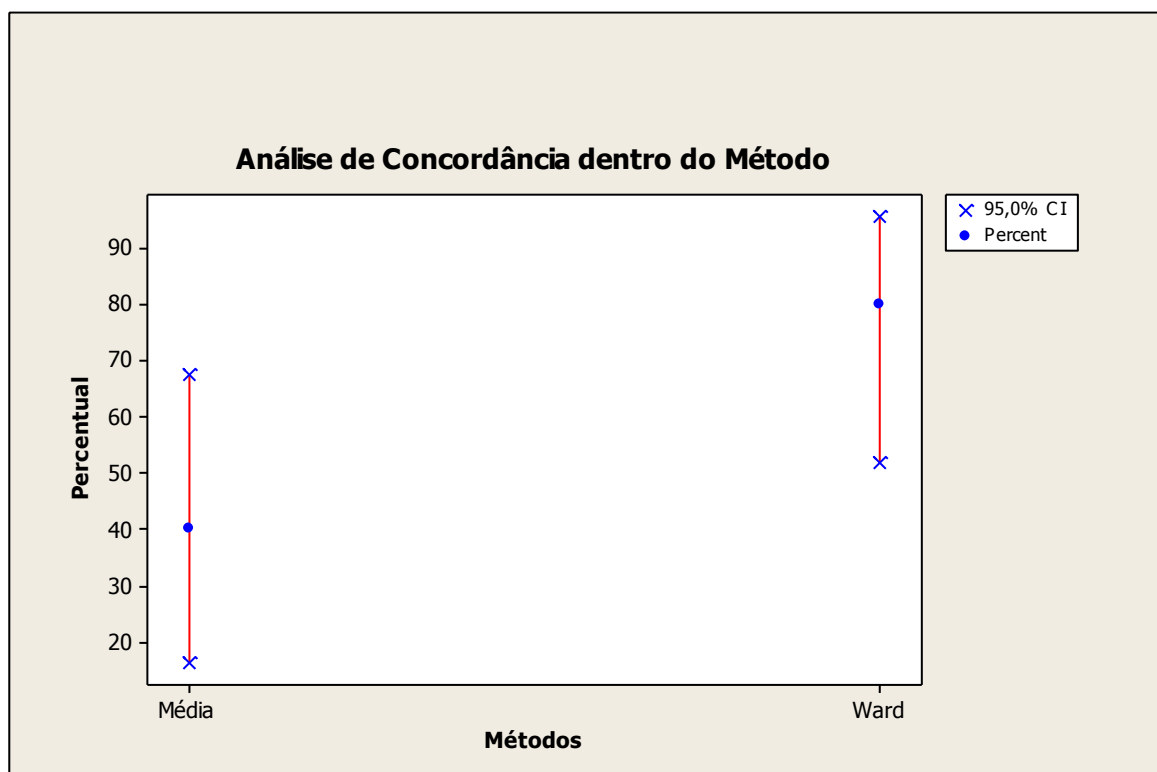


Figura 5.17 – Resultado da análise de concordância para os sistemas da Escelsa.

Nota-se que o método de Ward se mostrou mais robusto, possuindo uma concordância de aproximadamente 80% nos agrupamentos gerados. Portanto, este é o melhor método a ser utilizado na base de dados considerada.

O gráfico da Figura 5.18 mostra o intervalo de confiança de 95% para o número médio de AMTs considerando os agrupamentos formados pelo método de Ward. Nota-se que o grupo 1 possui a média de AMTs em um nível baixo e o grupo 2 possui alta frequência de AMTs, além disso, não há sobreposição do intervalo de confiança dos grupos, mostrando ser dois grupos distintos. Ressalta-se que a aplicação da metodologia proposta em uma base de dados

que contenha um número maior de sistemas de distribuição, permitirá a divisão dos objetos em um número maior de grupos sem prejuízo da separabilidade entre grupos.

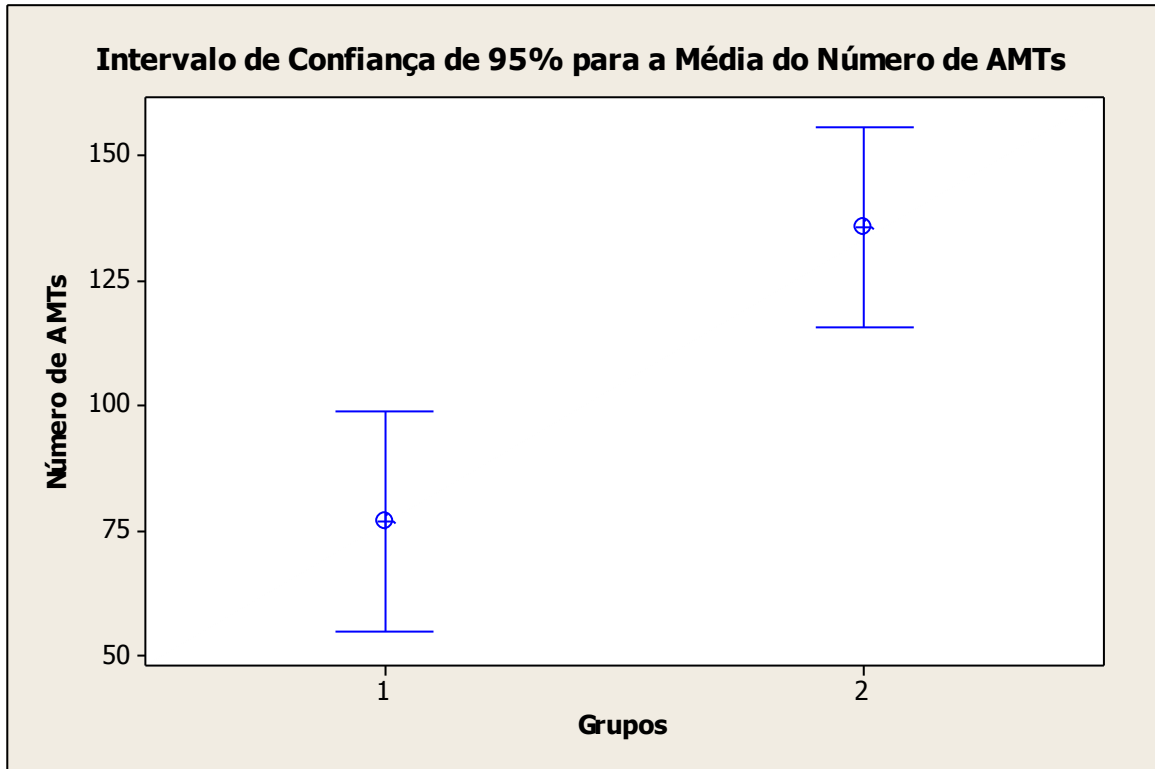


Figura 5.18 – Intervalo de confiança dos grupos quanto aos AMTs (Ward).

Outra visualização possível dos agrupamentos, é o gráfico tridimensional, onde os sistemas de distribuição pertencentes ao mesmo grupo é representado por um símbolo igual e os eixos correspondem as três variáveis utilizadas como entrada nos algoritmos de formação de conjuntos. A Figura 5.19 mostra o gráfico tridimensional relativo aos agrupamentos formados pelo método de Ward. Observa-se que visualmente a divisão em grupos feita pelo método de Ward é coerente, quando observada no gráfico a distância entre elementos pertencentes ao mesmo grupo.

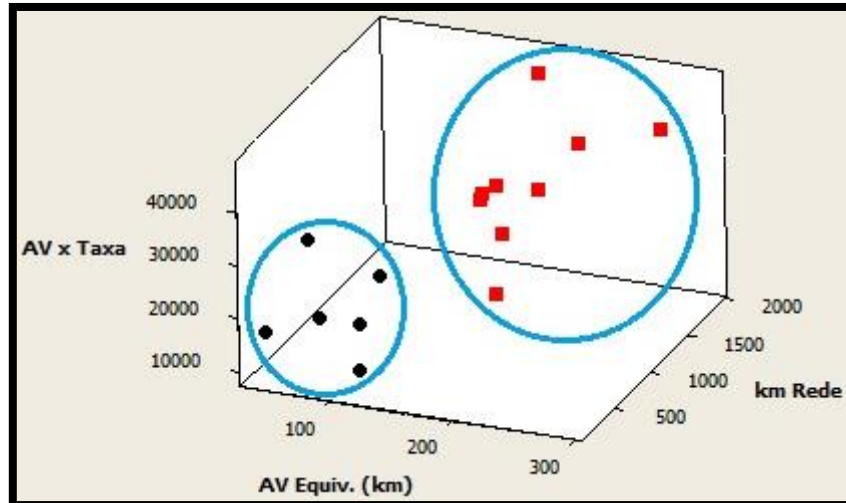


Figura 5.19 – Gráfico tridimensional (Ward).

5.4 Considerações Finais

Nota-se que o método que melhor agrupou os sistemas de distribuição reais da EDP Escelsa foi diferente daquele encontrado considerando os sistemas de distribuição hipotéticos. Conforme já mencionado, a metodologia proposta é capaz de encontrar o método mais adequado a cada base de dados, como os dados são diferentes de um caso para o outro é possível que o método de formação de conjuntos que melhor se ajusta também seja diferente.

Observa-se também que na base de dados dos sistemas de distribuição hipotéticos, um dos passos da metodologia teve de ser suprimido diante da completa coincidência nos agrupamentos formados por todos os métodos. Porém, ressalta-se que este resultado ocorreu devido ao número de variáveis de entrada, de sistemas de distribuição e de grupos a serem formados serem pequenos. Então, para o propósito que a metodologia está sendo estabelecida, que é a aplicação em todos os sistemas de distribuição brasileiros a fins de gerar limites adequados para a frequência de afundamentos de tensão, provavelmente tal resultado não irá ocorrer, pois o número dos fatores citados serão muito maiores.

6. CONCLUSÕES

6.1 Conclusões Gerais

O afundamento de tensão é o distúrbio relacionado à qualidade da energia elétrica que causa maior impacto nos consumidores industriais que possuem cargas sensíveis, gerando prejuízos expressivos. Assim sendo, a ABRACE tem enviado solicitações fundamentadas a ANEEL para regulamentação do fenômeno. Como consequência, a revisão 8 do Módulo 8 do PRODIST, mesmo com posicionamento contrário da ABRADDEE, está em vias de regulamentar os AMTs, empregando a princípio, conforme descrito nesta dissertação, o indicador Fator de Impacto.

Contudo, concordando com o passo inicial necessário em qualquer regulamentação, acredita-se que de forma semelhante ao ocorrido com os indicadores de qualidade do serviço (DEC e FEC), a norma proposta poderá evoluir para formação de conjuntos de sistemas de distribuição com característica semelhantes para afundamento de tensão.

Neste contexto, este trabalho de dissertação vem ao encontro da necessidade do setor elétrico, apresentando de maneira didática a proposição de uma metodologia para formação de conjuntos robustos de sistemas de distribuição quanto aos afundamentos de tensão.

A metodologia para formação de conjuntos proposta, como visto no estudo de caso é robusta, pois diferentemente da maior parte dos trabalhos publicados, o método de clusterização considerado melhor é estabelecido através de métodos estatísticos, livres de qualquer tipo de intervenção ou escolha do pesquisador.

A capacidade da metodologia de ser estendida a uma base de dados maior que contemple outros sistemas de distribuição é um importante fator, pois viabiliza a sua utilização pelo órgão regulamentador para gerar limites adequados de frequência de afundamentos de tensão para cada conjunto de sistemas de distribuição.

A contribuição deste trabalho em estabelecer agrupamentos confiáveis para o fenômeno de afundamento de tensão, visa auxiliar os diversos agentes do setor elétrico, sobretudo os consumidores e concessionárias.

Os consumidores industriais, poderão saber quais sistemas de distribuição pertencem a conjuntos que apresentam baixa frequência de AMTs e, assim, escolher estrategicamente a melhor localidade para instalação da indústria.

Indiretamente, a metodologia ajudará os consumidores industriais já instalados, uma vez que as concessionárias de energia elétrica que possuem sistemas de distribuição com desempenho inferior ao esperado para o conjunto, poderá sofrer penalização e será pressionada a investir na rede de modo a diminuir o número de AMTs, e conseqüentemente reduzir as paradas de processo na indústria.

As empresas distribuidoras de energia elétrica que realizam com frequência manutenções preventivas e investe em atualizações da rede, também se beneficiarão, já que a quantidade de afundamentos de tensão em seus sistemas de distribuição estarão abaixo dos demais sistemas pertencentes a um mesmo conjunto, sendo isto um fator de *marketing* para empresa firmar contrato com grandes consumidores de energia elétrica.

6.2 Proposta para Trabalhos Futuros

Como trabalhos futuros, pode-se implementar novos exemplos de aplicação contemplando maior número de sistemas de distribuição pertencentes a diferentes empresas, permitindo-se assim ter maior representatividade estatística e melhor separação entre conjuntos.

Por último, e não menos importante, pode-se utilizar a metodologia para estabelecer padrões de referência para a frequência de afundamentos de tensão em cada conjunto de sistemas de distribuição, por exemplo considerando como referência o percentil 20% do número de AMTs apresentados pelos sistemas de distribuição pertencentes a um conjunto.

6.3 Artigos Publicados

Diretamente relacionado com o tema desta dissertação, foi publicado o artigo:

FILHO, J. M.; CARVALHO, J. M.; SOUZA, P. V.; TOMASIN, S. A PCA-based Approach for Substation Clustering for Voltage Sag Studies in the Brazilian New Energy Context. **Electric Power Systems Research**, v. 136, p. 31–42, 2016.

De forma mais abrangente no assunto VTCDs, durante o curso de mestrado foram publicados os artigos:

OLIVEIRA, T. C.; CARVALHO, J. M.; HONORATO, K. S.; RIBEIRO, H. K.; SOUZA, P. V.; FILHO, J. M.; PEREIRA, N.; CARNEIRO, J. R. Proposta de Método para a Validação de Resultados de Simulação e Medição de VTCDs. **Conferência Brasileira sobre Qualidade da Energia Elétrica – CBQEE**, 2015.

OLIVEIRA, T. C.; CARVALHO, J. M.; HONORATO, K. S.; RIBEIRO, H. K.; SOUZA, P. V.; FILHO, J. M.; PEREIRA, N.; CARNEIRO, J. R. Proposta de Padrões de Referência para Indicadores de Afundamentos de Tensão. **Conferência Brasileira sobre Qualidade da Energia Elétrica – CBQEE**, 2015.

OLIVEIRA, T. C.; CARVALHO, J. M.; RIBEIRO, H. K.; SOUZA, P. V. Influência da Agregação Temporal nos Indicadores de Afundamentos de Tensão. **Seminário Nacional de Produção e Transmissão de Energia – XXIII SNPTEE**, 2015.

REFERÊNCIAS

AMASIFEN, J. C. C. **Metodologias para Avaliação de Riscos e dos Custos de Interrupções em Processos Causados por Faltas em Sistemas de Distribuição de Energia Elétrica**. Escola Politécnica da Universidade de São Paulo, 2008.

ASPEN. **DistriView**, Disponível em:

<http://www.aspeninc.com/web/index27cb.html?option=com_content&view=article&id=87&Itemid=68>

AYELLO, F. P. et al. Influência do Sistema de Proteção na Qualidade da Energia. **III Concladis - Congresso Latino Americano de Distribuição de Energia Elétrica**. São Paulo: 1999

BALOUJI, E.; SALOR, O. Eigen-Analysis Based Power Quality Event Data Clustering and Classification. **5th IEEE PES International Conference and Exhibition on Innovative Smart Grid Technologies**. Ankara, Turkey: 2014

BALTAR, V.; OKANO, V. **Análise de Concordância - Kappa**. Disponível em: <<http://www.lee.dante.br/pesquisa/kappa/>>. Acesso em: 14 maio. 2016.

BARRERA, V. N.; PAVAS, A. M.; MELÉNDEZ, J. F. Power Quality Assessment of the Bogotá Distribution Network Focused on Voltage Sag Analysis. **2nd IEEE PES International Conference and Exhibition on Innovative Smart Grid Technologies**. 2011

BECKER, C. et al. Proposed Chapter 9 for Predicting Voltage Sags (Dips) in Revision to IEEE Std 493, The Gold Book. **IEEE Transactions on Industry Applications**, v. 30, n. 3, p. 805–821, 1994.

BLACKBURN, J. L.; DECKKER, M. Protective Relaying. **Electrical Engineering and Electronics, New York**, n. 37, 1987.

BOLLEN, M. H. J. The Influence of Motor Reacceleration on Voltage Sags. **IEEE Transactions on Industry Applications**, v. 30, n. N°3, p. 805–821, 1994.

BOLLEN, M. H. J. Characterization of Voltage Sags Experienced by Three - Phase Adjustable - Speed Drive. **IEEE Transactions on Industry Applications**, 1997.

CABRAL, R. J. **Análise Numérica de Curto Circuito Utilizando Componentes Simétricas e Componentes de Fases para Obter Índices de Afundamentos de Tensão**. Universidade Federal do Rio Grande do Sul, 2010.

CARVALHO FILHO, J. M. **Uma Contribuição à Avaliação do Atendimento a Consumidores com Cargas Sensíveis – Proposta de Novos Indicadores**. Escola Federal de Engenharia de Itajubá, 2000.

CARVALHO FILHO, J. M. et al. Comparative Analysis between Measurements and Simulations of Voltage Sags. **IEEE – PES - 10th International Conference on Harmonics and Quality of Power, Rio de Janeiro, Brazil**, 2002.

COHEN, J. **Statistical Power Analysis for the Behavioral Sciences**. Hillsdale, NJ: Erlbaum, 1988.

CONRAD, L.; LITTLE, K.; GRIGG, C. Predicting and Preventing Problems Associated with Remote Fault - Clearing Voltage Dips. **IEEE Transactions on Industry Applications**, v. 27, n. 1, p. 167–172, 1991.

CP N°18/2014. Disponível em: <<http://www.aneel.gov.br/consultas-publicas>>.

DAGENHART, J. B. The 40 Ω ground-fault phenomenon. **IEEE Transactions on Industry Applications**. NJ: **IEEE**, v. 36, n. 1, p. 30–32, 2000.

DANCEY, C.; REIDY, J. **Estatística Sem Matemática para Psicologia: Usando SPSS para Windows**. Porto Alegre: Artmed, 2006.

FIGUEIREDO, D. B.; SILVA, J. A. Desvendando os Mistérios do Coeficiente de Correlação de Pearson (r). **Revista Política Hoje**, Vol. 18, p. 115–146, 2009.

FILHO, J. M. et al. A PCA-based Approach for Substation Clustering for Voltage Sag Studies in the Brazilian New Energy Context. **Electric Power Systems Research**, v. 136, p. 31–42, 2016.

- FILHO, J. M. **Agrupamento de Subestações para Estudos de Afundamentos de Tensão por Análise de Componentes Principais**. Universidade Federal de Itajubá - UNIFEI, 2016.
- FONSECA, V. R. DE C. **Cálculo Estocástico do Afundamento de Tensão**. Pontifícia Universidade Católica de Minas Gerais, 1999.
- GARCIA, C. P.; ALVES, M. F. Dimensionamento de Compensadores Série Considerando a Frequência de Ocorrência dos Afundamentos. **Anais do IX CBQEE**, p. 255–259, 2011.
- HARIYANTO, N.; ANGGORO, B.; NOEGROHO, R. New Probabilistic Approach for Identification Event Severity Index Due To Short Circuit Fault. **IEEE International Conference on Electrical Engineering and Computer Science**. Bali, Indonesia: 2014
- HONORATO, K. S. **Simulações Estocásticas de Variações de Tensão de Curta Duração em Sistemas de Distribuição**. Universidade Federal de Itajubá - UNIFEI, 2014.
- HUANG, N. H. N.; LIN, L. L. L. Power Quality Disturbances Recognition Based on PCA and BP Neural Network. **Power and Energy Engineering Conference (APPEEC)**, Asia-Pacific. 2010
- IEC 61000-2-8. **Electromagnetic Compatibility (EMC) - Part 2-8: Environment - Voltage dips and short interruptions on public electric power supply systems with statistical measurement results**, 2005.
- IEC 61000-4-11. **Electromagnetic Compatibility (EMC) - Part 4-11: Testing and measurement techniques - Voltage dips, short interruptions and voltage variations immunity tests**, 2004.
- IEC 61000-4-30. **Electromagnetic Compatibility (EMC) - Testing and measurement techniques - quality measurements methods**. International Electrotechnical Commission, 2006.
- IEEE 1159. **IEEE Recommended Practice for Monitoring Electric Power Quality**, 2009.
- IEEE 493. **IEEE Recommended Practice for the Design of Reliable Commercial Power (The Gold Book)**, 2007.

IEEE P1564/D19. **IEEE Guide for Voltage Sag Indices**, 2013.

JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis**. Sixth ed. Pearson Prentice Hall, 2007.

KEMPNER, T. R. **A Robustez de um Sistema de Distribuição e a Alocação de Medidores de Qualidade da Energia Elétrica Frente aos Afundamentos de Tensão**. Escola de Engenharia de São Carlos da Universidade de São Paulo, 2012.

LANDIS, J.; KOCH, G. **The Measurement of Observer Agreement for Categorical Data**. Biometrics, 1977.

LEBORGNE, R. C. **Uma Contribuição à Caracterização da Sensibilidade de Processos Industriais Frente a Afundamentos de Tensão**. Universidade Federal de Itajubá, 2003.

LIRA, M. M. S. et al. **Improving Disturbance Classification by Combining Multiple Artificial Neural Networks**. International Joint Conference on Neural Networks. Vancouver, Canada: 2006

LIRA, S. A. **Análise de Correlação: Abordagem Teórica e de Construção dos Coeficientes com Aplicações**. Universidade Federal do Paraná, 2004.

MARTINEZ, J. A.; MARTIN-ARNEDO, J. Voltage Sag Stochastic Prediction Using an Electromagnetic Transients Program. **IEEE Transactions on Power Delivery**, v. 19, n. 4, p. 1975–1982, 2004.

MARTINEZ, J. A.; MARTIN-ARNEDO, J. Voltage Sag Studies in Distribution Networks — Part II : Voltage Sag Assessment. **IEEE Transactions on Power Delivery**, v. 21, n. 3, p. 1679–1688, 2006.

MATLAB. **Version 7.12.0.635 (R2011a)**. The MathWorks, Inc., , 2011.

MELLENDEZ, A. K. J.; COLOMER, M. Z. J. Classification of Voltage Sags based on Multiway Principal Component Analysis and Case Based Reasoning. **Proceedings of the 17th World Congress The International Federation of Automatic Control**. Seoul, Korea: 2008

MINGOTI, S. A. **Análise de Dados Através de Métodos de Estatística Multivariada: Uma Abordagem Aplicada**. 2ª. ed. Belo Horizonte: Editora UFMG, 2005.

MINITAB, 16. **Minitab Statistical Software Release 16.1.0.0**. State College USA, , 2010.

NRS 048-2. **Electricity Supply - Quality of Supply Part 2: Voltage characteristics, compatibility levels, limits and assessment methods**, 2003.

OLIVEIRA, T. C. et al. Análise da Influência da Conexão de Transformadores Δ -Y aterrado na Propagação de Afundamentos de Tensão. **Seminário Brasileiro sobre Qualidade de Energia Elétrica - SBQEE - Aracajú**. 2003

ONS. **Sub-módulo 2.8: Gerenciamento dos Indicadores de Desempenho da Rede Básica e de seus Componentes**.

ORTMEYER, T. H.; HIYAMA, T.; SALEHFAR, H. Power Quality Effects of Distribution Systems Faults. **Electrical Power & Energy Systems**, v. 18, n. 5, p. 323–329, 1996.

POZZEBON, G. G. et al. Neural Classification of Power Quality Disturbances: An Application of the Wavelet Transform and Principal Component Analysis. **9th IEEE/IAS International Conference on Industry Applications**. São Carlos, Brasil: 2010

PRODIST. **Agência Nacional de Energia Elétrica – ANEEL Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional – PRODIST Módulo 8 – Qualidade da Energia Elétrica**.

RAMOS, A. C. L. **Avaliação dos Impactos da Impedância de Falta e da Geração Distribuída em Estudos de Afundamentos de Tensão**. UFG - Universidade Federal de Goiás, 2009.

ROMERO, M.; GALLEGRO, L.; PAVAS, A. Fault Zones Location on Distribution Systems Based on Clustering of Voltage Sags Patterns. **15th IEEE International Conference on Harmonics and Quality of Power**. 2012

SOUSA, M.; COSTA, C. A.; JUNIOR, I. H. P. Sistema de Localização Automática de Falhas em Linhas de Transmissão. **Anais do XVIII SNPTEE - SEMINÁRIO NACIONAL DE PRODUÇÃO E TRANSMISSÃO DE ENERGIA ELÉTRICA**, 2005.

STYVAKTAKIS, E. **Automating Power Quality Analysis**. Chalmers University of Technology - Goteborg, Sweden, 2002.

STYVAKTAKIS, E.; BOLLEN, M. H. J.; GU, I. Y. H. Automatic Classification of Power System Events Using RMS Voltage Measurements. **IEEE, Power Engineering Society Summer Meeting**, v. 2, p. 824–829, 2002.

SUPORTE AO MINITAB, 17. **O que são graus de liberdade?** Disponível em: <<http://support.minitab.com/pt-br/minitab/17/topic-library/basic-statistics-and-graphs/introductory-concepts/basic-concepts/df/>>. Acesso em: 13 jun. 2016.

WANG, N.; WANG, S.; JIA, Q. The Method to Reduce Identification Feature of Different Voltage Sag Disturbance Source Based on Principal Component Analysis. **IEEE Conference and Expo Transportation Electrification Asia-Pacific (ITEC Asia-Pacific)**. IEEE, 2014.

ANEXO A – Resolução de Exemplo de Formação de Conjuntos

O exemplo que será apresentado a seguir, foi retirado de um exemplo maior, com mais objetos e variáveis do (JOHNSON; WICHERN, 2007). Tal redução foi feita para permitir a execução passo a passo do algoritmo, sem torná-la muito extensa. Este mesmo exemplo será utilizado para demonstrar todos os demais métodos estudados neste documento, para permitir uma comparação entre os resultados de cada método. Supondo que foram medidas duas variáveis X_1 e X_2 para cada uma das cinco empresas de energia elétrica. Os valores obtidos são apresentados da Tabela A.1.

Tabela A.1 – Medições das variáveis.

Elementos	X_1	X_2
1 – Hawaiian	1,13	1,29
2 – New England	0,61	0,74
3 – Pacific	-0,35	-0,65
4 – San Diego	-1,49	-1,15
5 - United	0,10	-0,23

A.1 Método de Ligação Simples (*Single Linkage*)

A partir das variáveis medidas, calcula-se a matriz de distâncias entre os pares das cinco empresas de energia elétrica, utilizando a equação (3.2)

$$\begin{array}{c}
 \begin{array}{ccccc}
 & 1 & 2 & 3 & 4 & 5 \\
 \begin{array}{c}
 1 \\
 2 \\
 3 \\
 4 \\
 5
 \end{array}
 & \left[\begin{array}{ccccc}
 0 & & & & \\
 0,75 & 0 & & & \\
 2,43 & 1,69 & 0 & & \\
 3,58 & 2,83 & 1,24 & 0 & \\
 1,83 & 1,09 & 0,62 & 1,84 & 0
 \end{array} \right]
 \end{array}
 \end{array}$$

Considerando inicialmente cada objeto como um grupo, o processo de agrupamento começa combinando os dois elementos mais próximos (menor distância):

$$\min(d_{ik}) = d_{53} = 0,62$$

Então os objetos 5 e 3 são unidos para formação do grupo (35). Para implementar o próximo passo do algoritmo de agrupamento, calcula-se as distâncias entre o grupo (35) e os objetos remanescentes 1, 2 e 4. A menor distância entre eles são:

$$d_{(35)1} = \min\{d_{31}, d_{51}\} = \min\{2,43 ; 1,83\} = 1,83$$

$$d_{(35)2} = \min\{d_{32}, d_{52}\} = \min\{1,69 ; 1,09\} = 1,09$$

$$d_{(35)4} = \min\{d_{34}, d_{54}\} = \min\{1,24 ; 1,84\} = 1,24$$

Excluindo as linhas e colunas da matriz de distâncias correspondente aos objetos 3 e 5, e adicionando a linha e coluna relacionada ao conjunto (35), obtém-se a nova matriz de distâncias:

$$\begin{array}{c} (35) \\ (35) \left[\begin{array}{cccc} (35) & 1 & 2 & 4 \\ 0 & & & \\ 1 & 1,83 & 0 & \\ 2 & 1,09 & 0,75 & 0 \\ 4 & 1,24 & 3,58 & 2,83 & 0 \end{array} \right] \end{array}$$

Agora, a menor distância entre os pares de grupos é $d_{(12)} = 0,75$, então o grupo (1) é combinado com o grupo (2) para formar o novo grupo (12). Calculando, vem:

$$d_{(12)(35)} = \min\{d_{1(35)}, d_{2(35)}\} = \min\{1,83 ; 1,09\} = 1,09$$

$$d_{(12)4} = \min\{d_{14}, d_{24}\} = \min\{3,58 ; 2,83\} = 2,83$$

A nova matriz de distâncias, fica:

$$\begin{array}{c} (35) \\ (35) \left[\begin{array}{ccc} (35) & (12) & 4 \\ 0 & & \\ (12) & 1,09 & 0 \\ 4 & 1,24 & 2,83 & 0 \end{array} \right] \end{array}$$

A menor distância entre os pares de grupos é $d_{(12)(35)} = 1,09$, assim o grupo (12) é unido ao grupo (35) formando o novo grupo (1235). Neste ponto tem-se dois grupos distintos, (1235) e (4). A distância entre o “vizinho” mais próximo é:

$$d_{(1235)4} = \min\{d_{(35)4}, d_{(12)4}\} = \min\{1,24 ; 2,83\} = 1,24$$

A matriz final das distâncias, é:

$$\begin{array}{c} (1235) \\ (1235) \left[\begin{array}{cc} (1235) & (4) \\ 0 & \\ (4) & 1,24 & 0 \end{array} \right] \end{array}$$

No último estágio, os grupos (1235) e (4) são combinados para formar um único grupo contendo todos os cinco objetos, (12345). O dendograma da Figura A.1, ilustra o processo de agrupamento hierárquico, onde a união dos grupos e o nível de distância em que elas ocorrem estão claramente visíveis.

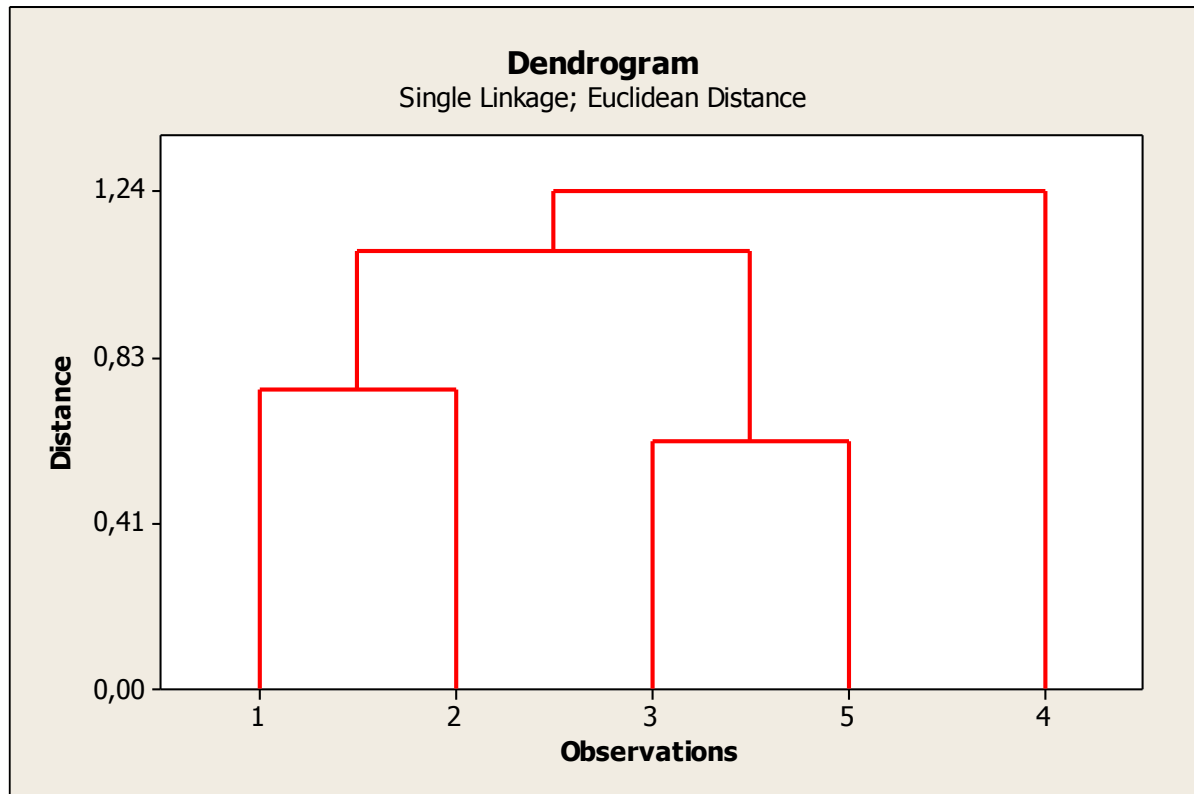


Figura A.1 – Dendrograma considerando o método de ligação simples.

A.2 Método de Ligação Completa (*Complete Linkage*)

Retornando aos dados do exemplo anterior, tem-se a matriz de distâncias:

	1	2	3	4	5
1	0				
2	0,75	0			
3	2,43	1,69	0		
4	3,58	2,83	1,24	0	
5	1,83	1,09	0,62	1,84	0

No primeiro estágio, os objetos 3 e 5 são combinados, pois eles são os mais similares.

Gera-se o grupo (35). No segundo estágio, calcula-se:

$$d_{(35)1} = \max\{d_{31}, d_{51}\} = \max\{2,43 ; 1,83\} = 2,43$$

$$d_{(35)2} = \max\{d_{32}, d_{52}\} = \max\{1,69 ; 1,09\} = 1,69$$

$$d_{(35)4} = \max\{d_{34}, d_{54}\} = \max\{1,24 ; 1,84\} = 1,84$$

E a matriz de distâncias modificada, torna-se:

$$\begin{array}{c}
 (35) \\
 (35) \left[\begin{array}{cccc}
 & (35) & 1 & 2 & 4 \\
 & 0 & & & \\
 1 & 2,43 & 0 & & \\
 2 & 1,69 & 0,75 & 0 & \\
 4 & 1,84 & 3,58 & 2,83 & 0
 \end{array} \right]
 \end{array}$$

A próxima combinação ocorre entre os grupos mais similares, 1 e 2, resultando no grupo (12), então, tem-se:

$$d_{(12)(35)} = \max\{d_{1(35)}, d_{2(35)}\} = \max\{2,43 ; 1,69\} = 2,43$$

$$d_{(12)4} = \max\{d_{14}, d_{24}\} = \max\{3,58 ; 2,83\} = 3,58$$

E a matriz de distâncias resultante é:

$$\begin{array}{c}
 (35) \quad (12) \quad 4 \\
 (35) \left[\begin{array}{ccc}
 & (35) & (12) & 4 \\
 & 0 & & \\
 (12) & 2,43 & 0 & \\
 4 & 1,84 & 3,58 & 0
 \end{array} \right]
 \end{array}$$

A próxima união produzirá o grupo (345), com a distância:

$$d_{(345)(12)} = \max\{d_{(35)(12)}, d_{4(12)}\} = \max\{2,43 ; 3,58\} = 3,58$$

A matriz final das distâncias, é:

$$\begin{array}{c}
 (345) \quad (12) \\
 (345) \left[\begin{array}{cc}
 & (345) & (12) \\
 & 0 & \\
 (12) & 3,58 & 0
 \end{array} \right]
 \end{array}$$

No estágio final, os grupos (12) e (345) formarão o grupo (12345). O dendograma do processo é mostrado na Figura A.2.

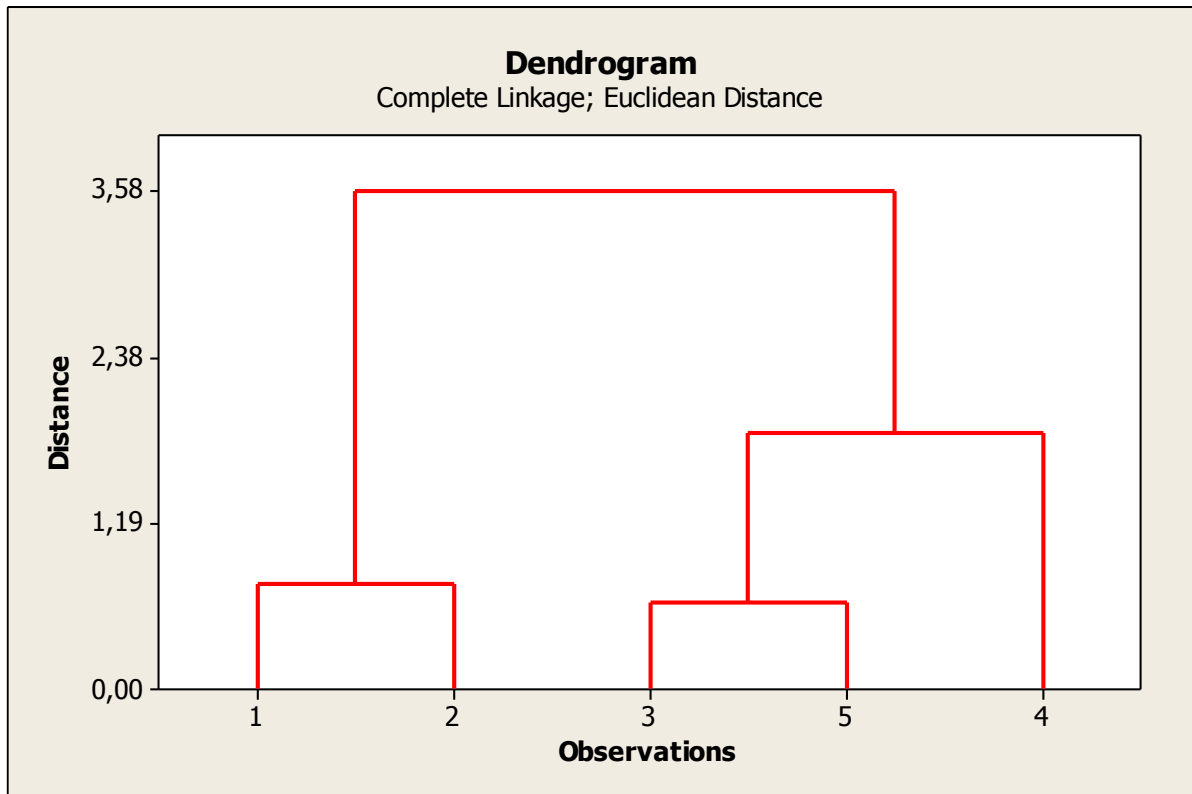


Figura A.2 – Dendrograma considerando o método de ligação completa.

A.3 Método da Média das Distâncias (*Average Linkage*)

Retornando à matriz de distâncias dos exemplos anteriores:

	1	2	3	4	5
1	0				
2	0,75	0			
3	2,43	1,69	0		
4	3,58	2,83	1,24	0	
5	1,83	1,09	0,62	1,84	0

No primeiro estágio, os objetos 3 e 5 são combinados, pois eles são os mais similares.

Gera-se o grupo (35). No segundo estágio, calcula-se:

$$d_{(35)1} = \frac{1}{2}(d_{31} + d_{51}) = \frac{2,43 + 1,83}{2} = 2,13$$

$$d_{(35)2} = \frac{1}{2}(d_{32} + d_{52}) = \frac{1,69 + 1,09}{2} = 1,39$$

$$d_{(35)4} = \frac{1}{2}(d_{34} + d_{54}) = \frac{1,24 + 1,84}{2} = 1,54$$

E a matriz de distâncias modificada, torna-se:

$$(35) \begin{matrix} & (35) & 1 & 2 & 4 \\ (35) & \left[\begin{array}{cccc} 0 & & & \\ 2,13 & 0 & & \\ 1,39 & 0,75 & 0 & \\ 1,54 & 3,58 & 2,83 & 0 \end{array} \right] \end{matrix}$$

A próxima combinação ocorre entre os grupos mais similares, 1 e 2, resultando no grupo (12), então, tem-se:

$$d_{(12)4} = \frac{1}{2}(d_{14} + d_{24}) = \frac{3,58 + 2,83}{2} = 3,2$$

$$d_{(12)(35)} = \frac{1}{4}(d_{13} + d_{15} + d_{23} + d_{25}) = \frac{2,43 + 1,83 + 1,69 + 1,09}{4} = 1,76$$

A matriz de distâncias resultante é:

$$\begin{matrix} & (35) & (12) & 4 \\ (35) & \left[\begin{array}{ccc} 0 & & \\ 1,76 & 0 & \\ 1,54 & 3,20 & 0 \end{array} \right] \\ (12) & & & \\ 4 & & & \end{matrix}$$

A próxima união produzirá o grupo (345), com a distância:

$$d_{(12)(345)} = \frac{1}{6}(d_{13} + d_{14} + d_{15} + d_{23} + d_{24} + d_{25}) = \frac{2,43 + 3,58 + 1,83 + 1,69 + 2,83 + 1,09}{6} = 2,24$$

A matriz final das distâncias é:

$$\begin{matrix} & (345) & (12) \\ (345) & \left[\begin{array}{cc} 0 & \\ 2,24 & 0 \end{array} \right] \\ (12) & & \end{matrix}$$

No estágio final os grupos (12) e (345) formarão o grupo (12345). O dendograma do processo é mostrado na Figura A.3.

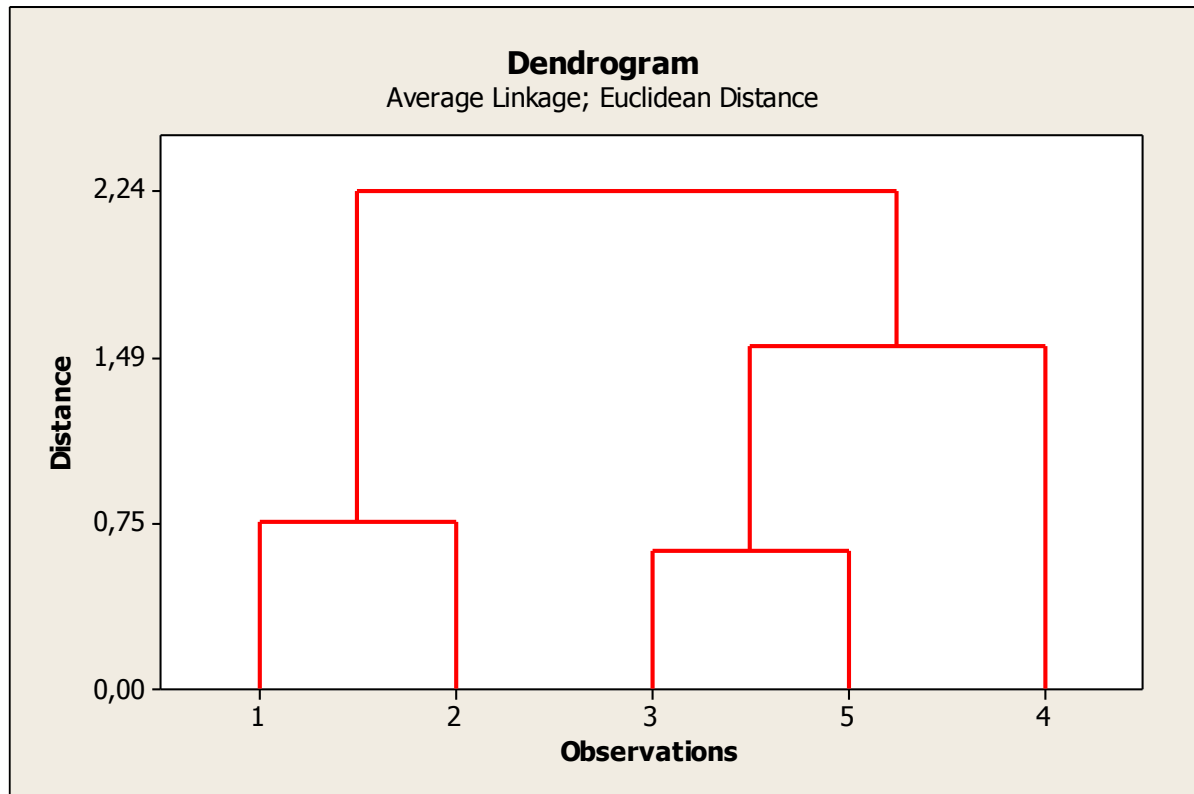


Figura A.3 – Dendrograma considerando o método da médias das distâncias.

A.4 Método de Ward

Retornando aos dados do exemplo considerado, para a implementação do método de Ward, apesar do dendrograma ser mostrado em termos das distâncias Euclidianas, o critério para realizar o agrupamento de elementos, corresponde ao menor SSR (soma de quadrados total dentro dos grupos). Então, partindo-se do estado inicial em que cada elemento é considerado um grupo, obtêm-se uma matriz que não mais corresponde à matriz de distâncias entre os elementos, cada célula desta matriz corresponde ao valor do SSR se os elementos da linha e da coluna fossem agrupados.

	1	2	3	4	5
1	0				
2	0,28	0			
3	2,96	1,43	0		
4	6,39	4,00	0,77	0	
5	1,67	0,60	0,19	1,70	0

No primeiro estágio, os objetos 3 e 5 são combinados, pois a união deles gerou o menor valor de $SSR = 0,19$. Gera-se o grupo (35). No segundo estágio, calcula-se novamente os valores de SSR considerando o novo grupo formado. Para exemplificar, a seguir é mostrado o cálculo feito para preencher uma célula da matriz de SSRs:

$$\begin{aligned}
 SSR_{(35)1} &= SSi_{(35)1} + SSi_2 + SSi_4 \\
 &= [(1,13 - 0,5)^2 + (1,29 - 0,42)^2 + (-0,35 - 0,5)^2 + (-0,65 - 0,42)^2 \\
 &\quad + (0,1 - 0,5)^2 + (-0,23 - 0,42)^2] + 0 + 0 = 3,59
 \end{aligned}$$

E a matriz de SSRs, torna-se:

$$\begin{array}{c}
 \begin{array}{c} (35) \\ 1 \\ 2 \\ 4 \end{array} \begin{array}{c} (35) \\ 1 \\ 2 \\ 4 \end{array} \begin{array}{c} 1 \\ 2 \\ 4 \end{array} \begin{array}{c} 1 \\ 2 \\ 4 \end{array} \begin{array}{c} 1 \\ 2 \\ 4 \end{array} \\
 \left[\begin{array}{cccc}
 0 & & & \\
 3,59 & 0 & & \\
 1,64 & 0,47 & 0 & \\
 1,97 & 6,58 & 4,19 & 0
 \end{array} \right]
 \end{array}$$

A próxima combinação ocorre entre os grupos em que a união acarreta menor aumento do SSR, 1 e 2, resultando no grupo (12). Recalculando a matriz de SSRs, tem-se:

$$\begin{array}{c}
 \begin{array}{c} (35) \\ (12) \\ 4 \end{array} \begin{array}{c} (35) \\ (12) \\ 4 \end{array} \begin{array}{c} (12) \\ 4 \end{array} \begin{array}{c} (12) \\ 4 \end{array} \\
 \left[\begin{array}{ccc}
 0 & & \\
 3,57 & 0 & \\
 2,05 & 7,31 & 0
 \end{array} \right]
 \end{array}$$

A próxima união produzirá o grupo (345). A matriz final de SSRs é:

$$\begin{array}{c}
 \begin{array}{c} (345) \\ (12) \end{array} \begin{array}{c} (345) \\ (12) \end{array} \begin{array}{c} (12) \end{array} \\
 \left[\begin{array}{cc}
 0 & \\
 8 & 0
 \end{array} \right]
 \end{array}$$

No estágio final os grupos (12) e (345) formarão o grupo (12345) com SRR=8. O dendograma do processo é mostrado na Figura A.4.

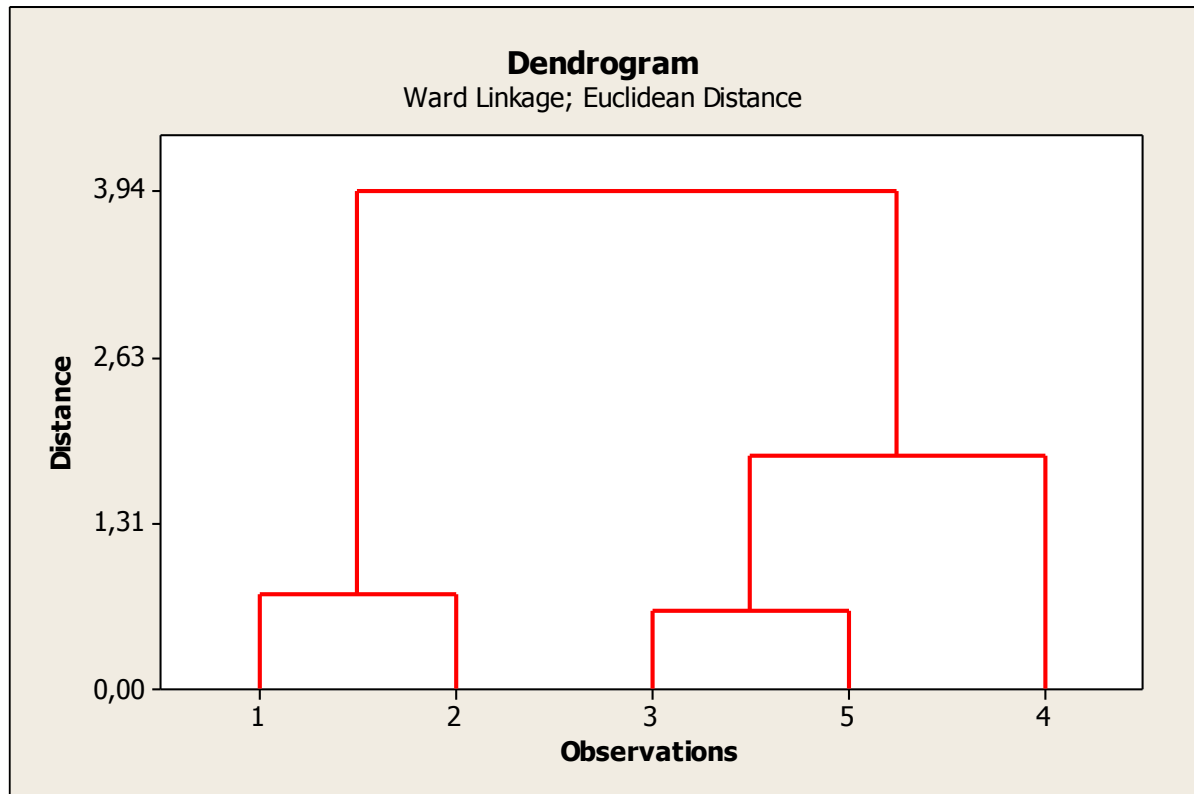


Figura A.4 – Dendrograma considerando o método de Ward.

A.5 Método das k-médias (*k-means*)

Retomando os mesmos dados do exemplo utilizado nos demais métodos, de acordo com a Tabela A.1, pretende-se na sequência ilustrar a aplicação do método de *k-means*.

O objetivo é dividir esses elementos em $k=2$ grupos de modo que elementos dentro de um grupo estejam mais próximos entre si do que estão dos elementos do outro grupo. Para implementar o método das k-médias, arbitra-se uma partição dos elementos em dois grupos, como por exemplo (14) e (235), e calcula-se as coordenadas (\bar{x}_1, \bar{x}_2) dos centróides (média) destes grupos. Assim, no passo 1, tem-se a Tabela A.2.

Tabela A.2 – Centróides dos grupos (14) e (235).

Grupos	Coordenadas do centróide	
	\bar{x}_1	\bar{x}_2
(14)	$\frac{1,13 + (-1,49)}{2} = -0,18$	$\frac{1,29 + (-1,15)}{2} = 0,07$
(235)	$\frac{0,61 + (-0,35) + 0,1}{3} = 0,12$	$\frac{0,74 + (-0,65) + (-0,23)}{3} = -0,046$

No passo 2, calcula-se a distância Euclidiana de cada elemento para o centróide de seu respectivo grupo. Se um elemento pertencente a um grupo passar a fazer parte de um outro

grupo, os centróides dos grupos devem ser recalculados antes de prosseguir o algoritmo, e a i -ésima coordenada do centróide pode ser recalculada usando as equações (A.1) e (A.2).

$$\bar{x}_{i,novo} = \frac{n\bar{x}_i + x_{ji}}{n+1} \text{ se o } j\text{-ésimo elemento foi adicionado ao grupo} \quad (\text{A.1})$$

$$\bar{x}_{i,novo} = \frac{n\bar{x}_i - x_{ji}}{n-1} \text{ se o } j\text{-ésimo elemento foi removido do grupo} \quad (\text{A.2})$$

Onde n é o número de elementos do grupo “antigo” com centróide $\bar{x}' = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p)$. Considerando os grupos iniciais (14) e (235), as coordenadas dos centróides são (-0,18 ; 0,07) e (0,12 ; -0,046) respectivamente. Supondo que o elemento 2 com coordenadas (0,61 ; 0,74) é movido para o grupo (14). Os novos grupos são (124) e (35), com os novos centróides apresentados na Tabela A.3.

Tabela A.3 – Centróides dos grupos (124) e (35).

Grupos	Coordenadas do centróide	
	\bar{x}_1	\bar{x}_2
(124)	$\frac{2(-0,18) + 0,61}{2 + 1} = 0,083$	$\frac{2(0,07) + 0,74}{2 + 1} = 0,293$
(35)	$\frac{3(0,12) - 0,61}{3 - 1} = -0,13$	$\frac{3(-0,046) - 0,74}{3 - 1} = -0,44$

Retornando ao agrupamento inicial do passo 1, calcula-se as distâncias ao quadrado:

Se 2 não é movido para o novo grupo:

$$d^2(2, (14)) = (0,61 + 0,18)^2 + (0,74 - 0,07)^2 = 1,07$$

$$d^2(2, (235)) = (0,61 - 0,12)^2 + (0,74 + 0,046)^2 = 0,86$$

Se 2 for movido para o novo grupo:

$$d^2(2, (124)) = (0,61 - 0,083)^2 + (0,74 - 0,293)^2 = 0,48$$

$$d^2(2, (35)) = (0,61 + 0,13)^2 + (0,74 + 0,44)^2 = 1,94$$

Como 2 está mais próximo do centróide do grupo (124) do que do grupo (235), a realocação suposta é efetuada. Continuando o processo, considera-se a realocação de 4, e têm-se os novos centróides mostrados na Tabela A.4.

Tabela A.4 – Centróides dos grupos (12) e (345).

Grupos	Coordenadas do centróide	
	\bar{x}_1	\bar{x}_2
(12)	$\frac{3(0,083) + 1,49}{3 - 1} = 0,87$	$\frac{3(0,293) + 1,15}{3 - 1} = 1,01$
(345)	$\frac{2(-0,13) - 1,49}{2 + 1} = -0,58$	$\frac{2(-0,44) - 1,15}{2 + 1} = -0,68$

Então, calcula-se as distâncias ao quadrado:

Se 4 não é movido para o novo grupo:

$$d^2(4, (124)) = (-1,49 - 0,083)^2 + (-1,15 - 0,293)^2 = 4,55$$

$$d^2(4, (35)) = (-1,49 + 0,37)^2 + (-1,15 + 0,832)^2 = 1,35$$

Se 4 for movido para o novo grupo:

$$d^2(4, (12)) = (-1,49 - 0,87)^2 + (-1,15 - 1,01)^2 = 10,23$$

$$d^2(4, (345)) = (-1,49 + 0,58)^2 + (-1,15 + 0,68)^2 = 1,05$$

Nota-se que 4 está mais próximo do centróide do grupo (345), então 4 é realocado para o grupo (345). Agora, tem-se os grupos (12) e (345) com os respectivos centróides (0,87 ; 1,01) e (-0,58 ; -0,68). Analisando a realocação de 3, tem-se os valores dos centróides calculados na Tabela A.5.

Tabela A.5 – Centróides dos grupos (123) e (45).

Grupos	Coordenadas do centróide	
	\bar{x}_1	\bar{x}_2
(123)	$\frac{2(0,87) - 0,35}{2 + 1} = 0,46$	$\frac{2(1,01) - 0,65}{2 + 1} = 0,46$
(45)	$\frac{3(-0,58) + 0,35}{3 - 1} = -0,69$	$\frac{3(-0,68) + 0,65}{3 - 1} = -0,69$

Então, calcula-se as distâncias ao quadrado:

Se 3 não é movido para o novo grupo:

$$d^2(3, (12)) = (-0,35 - 0,87)^2 + (-0,65 - 1,01)^2 = 4,24$$

$$d^2(3, (345)) = (-0,35 + 0,58)^2 + (-0,65 + 0,68)^2 = 0,05$$

Se 3 for movido para o novo grupo:

$$d^2(3, (123)) = (-0,35 - 0,46)^2 + (-0,65 - 0,46)^2 = 1,89$$

$$d^2(3, (45)) = (-0,35 + 0,69)^2 + (-0,65 + 0,69)^2 = 0,12$$

Visto que 3 está mais próximo do centróide do grupo (345) comparado ao grupo (123), 3 não será movido para este novo grupo suposto. Continuando este algoritmo, observa-se que não há mais realocações plausíveis, então os k=2 grupos finais são (12) e (345), cujas distâncias ao quadrado dos elementos aos grupos formados estão apresentadas na Tabela A.6.

Tabela A.6 – Distância ao quadrado dos elementos para os grupos finais.

Grupos	Distância ao quadrado				
	Elemento				
	1	2	3	4	5
(12)	0,14	0,14	4,24	10,23	2,13
(345)	6,76	3,43	0,05	1,05	0,67

A soma de quadrados da distância dos elementos para o centróide do grupo são, Grupo (12): $0,14+0,14=0,28$ e Grupo (345): $0,05+1,05+0,67=1,77$. Implementando este exemplo no software (MINITAB, 2010), obtém-se os resultados apresentados na Figura A.5.

```

K-means Cluster Analysis: C15; C16
Standardized Variables
Final Partition
Number of clusters: 2

          Within  Average  Maximum
          cluster  distance  distance
          Number of  sum of  from  from
          observations  squares  centroid  centroid
Cluster1          2    0,280    0,374    0,374
Cluster2          3    1,774    0,691    1,027

Cluster Centroids
Variable  Cluster1  Cluster2  Grand
          centroid
C15       0,8701   -0,5801   0,0000
C16       1,0132   -0,6755  -0,0000

Distances Between Cluster Centroids

          Cluster1  Cluster2
Cluster1  0,0000    2,2259
Cluster2  2,2259    0,0000

```

Figura A.5 – Saída do método de k-means.

A coluna nomeada *Number of observations* contabiliza o número de elementos que cada grupo possui. No caso, o grupo 1 possui dois elementos (12) e o grupo 2 possui três elementos (345). A coluna *Within cluster sum of squares* apresenta a soma das distâncias ao quadrado dos elementos pertencentes ao grupo em relação ao centróide do grupo, por exemplo o valor 1,774 corresponde à soma dos valores das distâncias do elemento 3, 4 e 5 em relação ao centróide do grupo (345), e tais valores são apresentadas na Tabela A.6. Na coluna *Average distance from centroid*, é mostrada a distância média dos elementos em relação ao centróide. Para o grupo 1, que possui apenas dois elementos, a distância média é a distância de um dos elementos para o centróide. Como o valor apresentado não está elevado ao quadrado basta considerar a raiz quadrada do valor 0,14 apresentado na Tabela A.6, para obter o valor 0,374. A coluna *Maximum distance from centroid* exibe a maior distância entre os elementos e o centróide do grupo, no grupo 2 do exemplo considerado, observa-se na Tabela A.6 que o elemento mais distante do centróide é o elemento 4, e novamente necessita-se extrair a raiz quadrada de 1,05 para obter o valor de 1,027.

Na segunda parte da saída de resultados, são apresentadas as coordenadas dos centróides dos grupos formados e na parte final é mostrada a matriz das distâncias euclidianas entre estes centróides.

Como o método de *k-means* não possui como saída o dendograma para identificação dos conjuntos formados, criou-se a Tabela A.7 com o *membership* (número que rotula a qual conjunto o elemento pertence) de cada companhia, que corresponde a uma identificação do grupo que determinada companhia pertence.

Tabela A.7 – Classificação das companhias em dois grupos.

Companhias	X ₁	X ₂	Membership
1 - Hawaiian	1,13	1,29	1
2 - New_England	0,61	0,74	1
3 - Pacific	-0,35	-0,65	2
4 - San_Diego	-1,49	-1,15	2
5 - United	0,1	-0,23	2

A.6 Comparação entre Métodos de Formação de Conjuntos

Tendo como uma possível meta a partição das cinco empresas de energia elétrica consideradas no exemplo, em dois agrupamentos, apresentam-se na Figura A.6 a os dendogramas obtidos aplicando cada um dos métodos considerados.

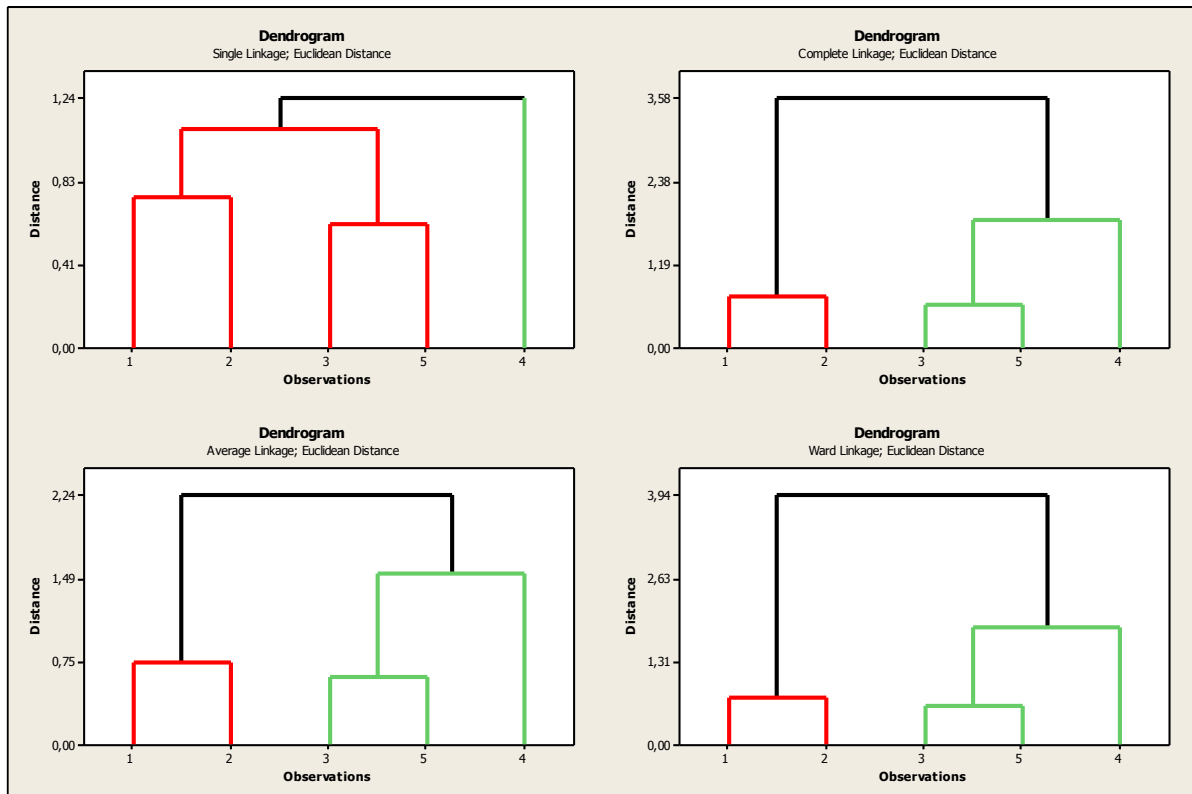


Figura A.6 – Comparação entre métodos de formação de conjuntos.

Observa-se pelos resultados obtidos, exceto o método de ligação simples, todos os demais, obtiveram como resultado a mesma divisão de agrupamentos (12) e (345). Isto possibilita ter segurança no resultado encontrado, e deixa evidente que foram encontrados dois agrupamentos naturais.

Para explorar mais alguns detalhes do processo de formação de conjuntos, optou-se por resolver com todos os métodos analisados, o exemplo completo do livro (JOHNSON; WICHERN, 2007), contendo 22 empresas de energia elétrica e 8 variáveis, mostradas na Tabela A.8. Para resolução foi utilizado o software (MINITAB, 2010), cujos dendogramas resultantes estão apresentados na Figura A.7.

Tabela A.8 – Dados de 22 companhias e 8 variáveis.

Companhias	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
1 - Arizona	-0,29	-0,68	-0,42	-0,58	-0,53	0,05	-0,71	-0,85
2 - Boston	-1,21	-0,19	0,82	0,21	-0,33	-1,08	0,79	0,81
3 - Louisiana	1,71	2,08	-1,34	-0,89	0,05	0,08	-0,71	-0,08
4 - Commonwealth	-0,51	0,21	0,00	-0,22	-0,94	-0,70	1,33	-0,72
5 - Consolidated	2,04	-0,86	0,58	-1,30	-0,72	-1,58	0,21	1,69
6 - Florida	1,12	1,23	-1,39	0,68	-1,74	0,62	0,63	0,25
7 - Hawaiian	0,57	0,65	0,17	2,38	-0,33	-0,36	-0,71	0,99
8 - Idaho	-0,08	-0,68	1,86	0,01	0,02	1,17	-0,71	-1,43
9 - Kentucky	1,22	1,01	0,00	0,77	1,27	-0,14	-0,71	-0,43
10 - Madison	0,03	0,74	0,70	-0,89	-0,17	-0,69	1,62	-0,86
11 - Nevada	-1,97	-1,44	0,12	-1,23	1,05	2,40	-0,71	-0,60
12 - New_England	0,09	0,07	0,24	1,13	0,15	-0,78	-0,71	1,43
13 - Northern	0,19	0,88	0,75	-0,73	1,01	-0,49	2,27	-1,04
14 - Oklahoma	-0,13	0,56	-1,75	-1,61	-0,59	0,21	-0,71	-0,93
15 - Pacific	-0,84	-1,40	-0,10	1,17	-1,07	-0,69	-0,66	0,53
16 - Puget	0,25	-0,37	2,03	-0,22	1,91	1,99	-0,71	-0,87
17 - San_Diego	-1,92	-1,93	-0,78	1,10	1,85	-0,90	-0,22	1,47
18 - The_Southern	-0,35	0,83	-0,44	-0,06	-0,17	0,35	-0,71	0,01
19 - Texas	0,25	0,43	-1,56	-0,67	-1,71	1,29	-0,71	-0,84
20 - Wisconsin	0,47	0,47	-0,49	0,66	0,08	-0,46	1,73	-0,72
21 - United	-0,40	-0,95	0,87	0,90	0,08	-0,64	-0,71	1,82
22 - Virginia	-0,24	-0,64	0,14	-0,60	0,85	0,33	0,87	0,37

Onde: X₁: Razão de cobertura à taxa fixa.

X₂: Taxa de retorno sobre o capital.

X₃: Custo por capacidade instalada (kW) no local.

X₄: Fator de carga anual.

X₅: Crescimento da demanda de pico em kWh de 1974 a 1975.

X₆: Vendas (kWh por ano).

X₇: Percentual de energia nuclear.

X₈: Custo total de combustível.

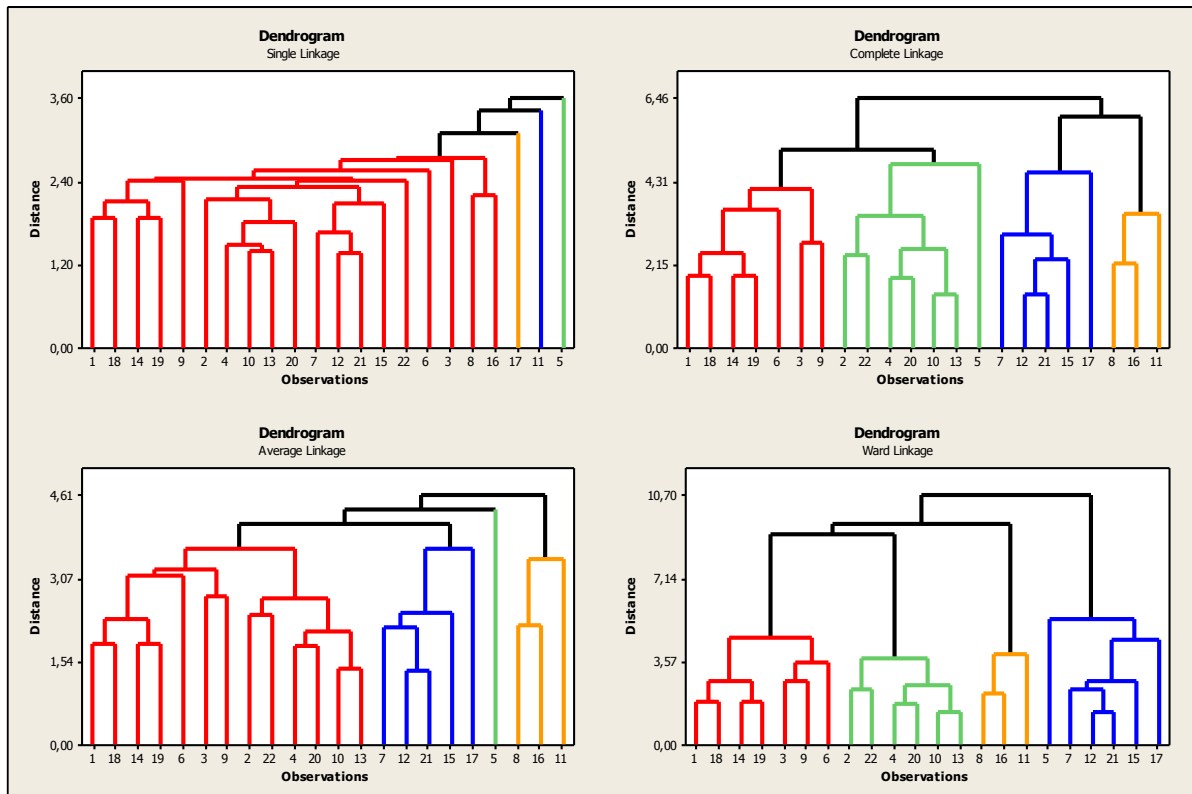


Figura A.7 – Comparação entre métodos (exemplo completo).

Nota-se, a partir da Figura A.7, que novamente o método de ligação simples possui um resultado muito diferente dos demais, sendo esta diferença uma possível tendência de discrepância deste método em relação aos demais. É importante também mencionar que a empresa 5, está presente cada vez em um agrupamento distinto de acordo com o método implementado, tal comportamento deve-se ao fato desta empresa possuir valores para as variáveis que estão justamente entre os valores característicos de dois ou mais agrupamentos.

O interessante das técnicas de formação de conjuntos é a sua capacidade de depreender informações que não foram inseridas como entradas no algoritmo. Por exemplo, analisando os grupos formados neste exemplo e conhecendo a localização das empresas no país (EUA) percebe-se que em geral os agrupamentos obtidos estão de acordo com a localização geográfica, e não é de se espantar que isto ocorra, pois empresas instaladas em áreas próximas, usam o mesmo tipo de combustível e dividem o mesmo mercado, conseqüentemente, as variáveis consideradas contém intrinsicamente esta informação.

ANEXO B – 12 Sistemas de Distribuição

Parâmetros da Subestação de Entrada			
Pcc Trifásico (MVA)	X/R Trifásico	Pcc Monofásico (MVA)	X/R Monofásico
676	4,33	467	4,02

Parâmetros dos Transformadores Δ - Y_{aterrado} (69 kV - 24,9 kV)					
Potência (MVA)	Z% (p.u)	X/R	Ângulo (rad)	R (p.u)	X (p.u)
2,5	0,07	8	1,446	0,009	0,069
5	0,07	12	1,488	0,006	0,070
7,5	0,08	15	1,504	0,005	0,080
10	0,08	17	1,512	0,005	0,080
12,5	0,08	18	1,515	0,004	0,080
15	0,08	19	1,518	0,004	0,080
20	0,08	21	1,523	0,004	0,080

Configurações para Linhas Aéreas				
Código	Sequência de Fases	Condutor - ACSR		Tipo de Estrutura
		Fase	Neutro	
300	B A C N	1/0	1/0	500
301	B A C N	#2 6/1	#2 6/1	500
302	A N	#4 6/1	#4 6/1	510
303	B N	#4 6/1	#4 6/1	510
304	B N	#2 6/1	#2 6/1	510

Descrição dos Códigos de Configuração			
Nó DE	Nó PARA	Comprimento(ft.)	Código
800	802	2580	300
802	806	1730	300
806	808	32230	300
808	810	5804	303
808	812	37500	300
812	814	10	301
814	850	29730	300
816	818	1710	302
816	824	10210	301
818	820	48150	302
820	822	13740	302
824	826	3030	303

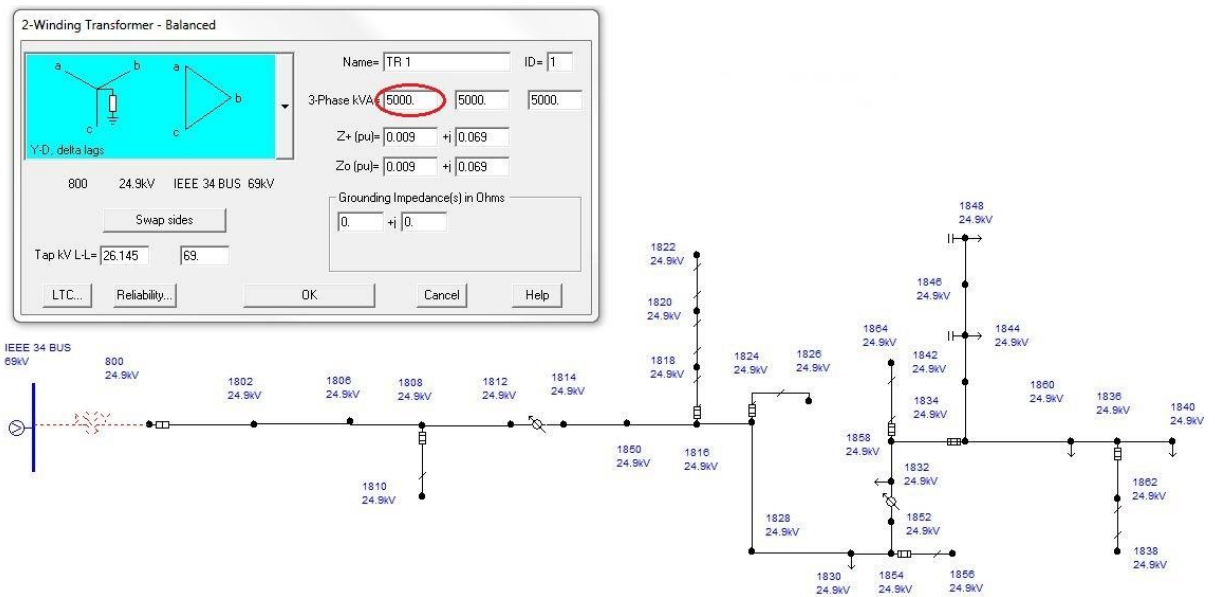
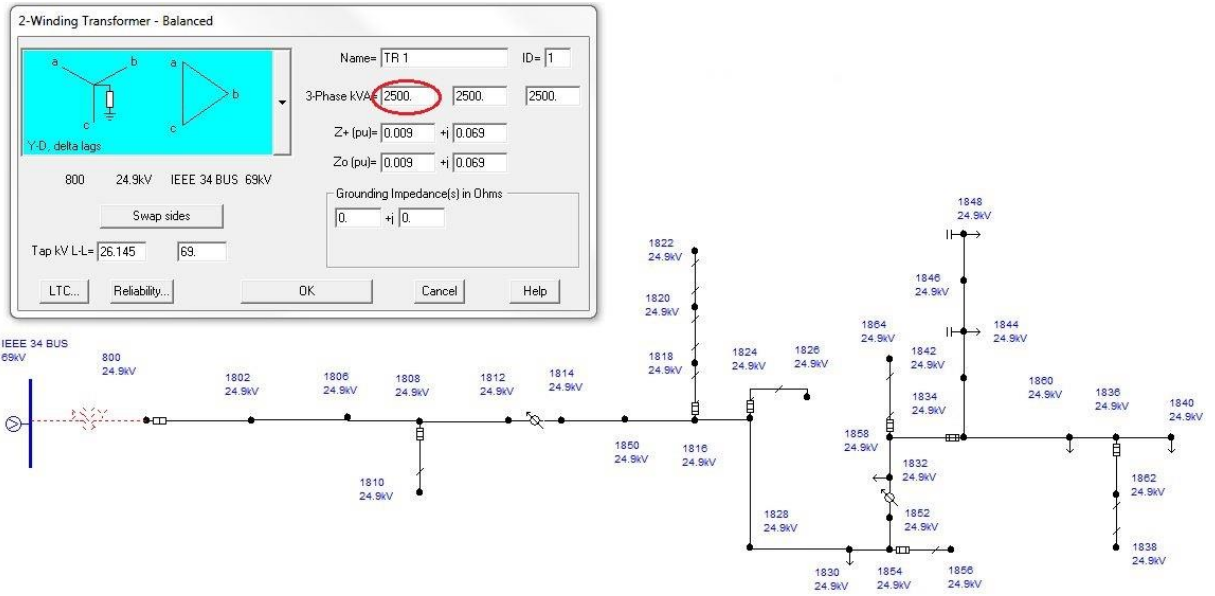
824	828	840	301
828	830	20440	301
830	854	520	301
832	858	4900	301
834	860	2020	301
834	842	280	301
836	840	860	301
836	862	280	301
842	844	1350	301
844	846	3640	301
846	848	530	301
850	816	310	301
852	832	10	301
854	856	23330	303
854	852	36830	301
858	864	1620	302
858	834	5830	301
860	836	2680	301
862	838	4860	304

Cargas Distribuidas								
Nó DE	Nó PARA	Modelo Carga	Fase A (kW)	Fase A (kVAr)	Fase B (kW)	Fase B (kVAr)	Fase C (kW)	Fase C (kVAr)
802	806	Y-PQ	0	0	30	15	25	14
808	810	Y-I	0	0	16	8	0	0
818	820	Y-Z	34	17	0	0	0	0
820	822	Y-PQ	135	70	0	0	0	0
816	824	D-I	0	0	5	2	0	0
824	826	Y-I	0	0	40	20	0	0
824	828	Y-PQ	0	0	0	0	4	2
828	830	Y-PQ	7	3	0	0	0	0
854	856	Y-PQ	0	0	4	2	0	0
832	858	D-Z	7	3	2	1	6	3
858	864	Y-PQ	2	1	0	0	0	0
858	834	D-PQ	4	2	15	8	13	7
834	860	D-Z	16	8	20	10	110	55
860	836	D-PQ	30	15	10	6	42	22
836	840	D-I	18	9	22	11	0	0
862	838	Y-PQ	0	0	28	14	0	0
842	844	Y-PQ	9	5	0	0	0	0
844	846	Y-PQ	0	0	25	12	20	11
846	848	Y-PQ	0	0	23	11	0	0

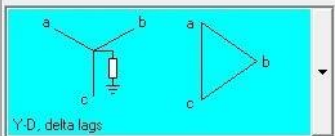
Cargas Pontuais							
Nó	Modelo Carga	Fase A (kW)	Fase A (kVAr)	Fase B (kW)	Fase B (kVAr)	Fase C (kW)	Fase C (kVAr)
860	Y-PQ	20	16	20	16	20	16
840	Y-I	9	7	9	7	9	7
844	Y-Z	135	105	135	105	135	105
848	D-PQ	20	16	20	16	20	16
832	D-I	164	92	164	92	164	92
830	D-Z	10	5	10	5	25	10

Capacitores Shunts			
Nó	Fases		
	A (kVAr)	B (kVAr)	C (kVAr)
844	100	100	100
848	150	150	150
Total	250	250	250

Parâmetros - Reguladores de Tensão						
	Regulador					
	1			2		
Segmento DE - PARA	814 - 850			852 - 832		
Localização	814			852		
Fases	A - B - C			A - B - C		
Conexão	Trifásico, Estrela Aterrado			Trifásico, Estrela Aterrado		
Fases Monitoradas	A-B-C			A-B-C		
Largura de Faixa	2.0 volts			2.0 volts		
Razão de Transformação do TP (N_{pt})	120			120		
Corrente Nominal no Primário do TC (CT_p)	100			100		
Ajustes do Compensador	Fase A	Fase B	Fase C	Fase A	Fase B	Fase C
R - %	2,7	2,7	2,7	2,5	2,5	2,5
X - %	1,6	1,6	1,6	1,5	1,5	1,5
Nível de Tensão	122	122	122	124	124	124



2-Winding Transformer - Balanced



Name= TR 1 ID= 1

3-Phase kVA= 5000. 5000. 5000.

Z+ (pu)= 0.006 +j 0.07

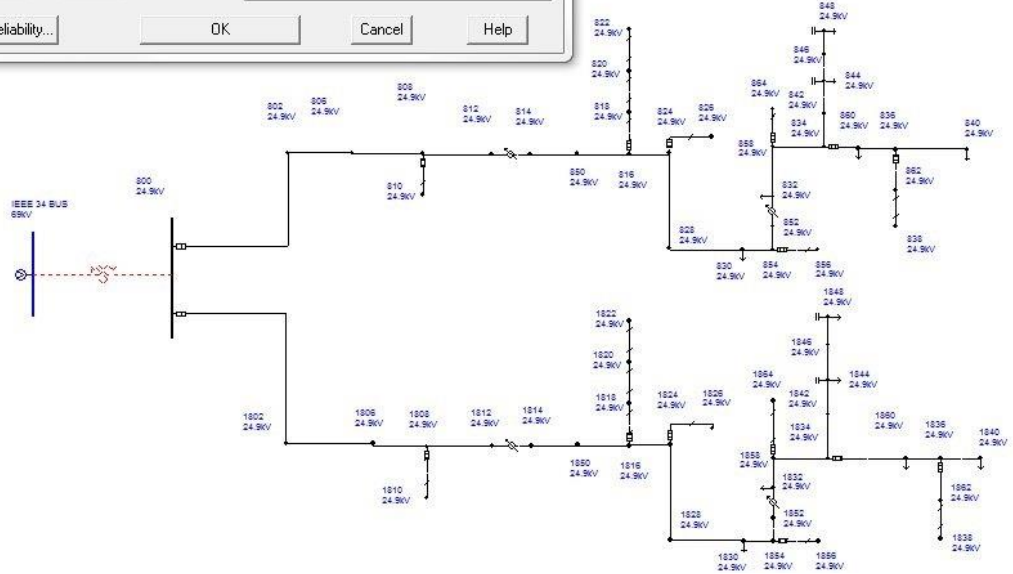
Zo (pu)= 0.006 +j 0.07

Grounding Impedance(s) in Ohms
0. +j 0.

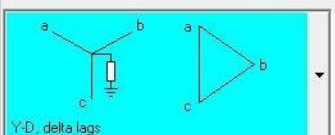
800 24.9kV IEEE 34 BUS 69kV

Tap kV L-L= 26.145 69.

LTC... Reliability... OK Cancel Help



2-Winding Transformer - Balanced



Name= TR 1 ID= 1

3-Phase kVA= 7500. 7500. 7500.

Z+ (pu)= 0.006 +j 0.07

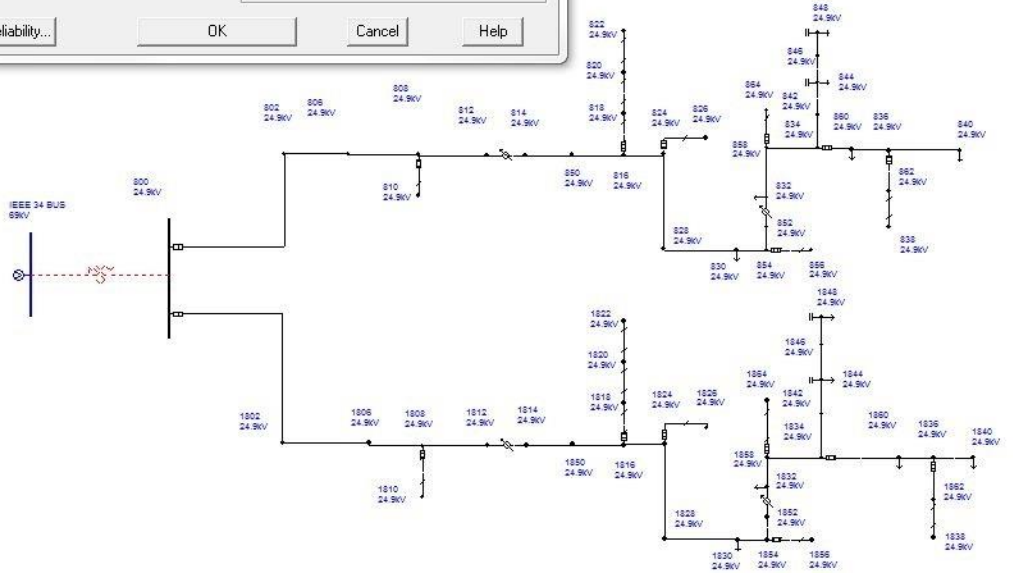
Zo (pu)= 0.006 +j 0.07

Grounding Impedance(s) in Ohms
0. +j 0.

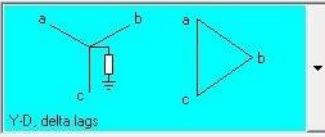
800 24.9kV IEEE 34 BUS 69kV

Tap kV L-L= 26.145 69.

LTC... Reliability... OK Cancel Help



2-Winding Transformer - Balanced



Name= TR 1 ID= 1

3-Phase kVA= 7500 7500 7500

Z+ (pu)= 0.005 +j 0.08

Zo (pu)= 0.005 +j 0.08

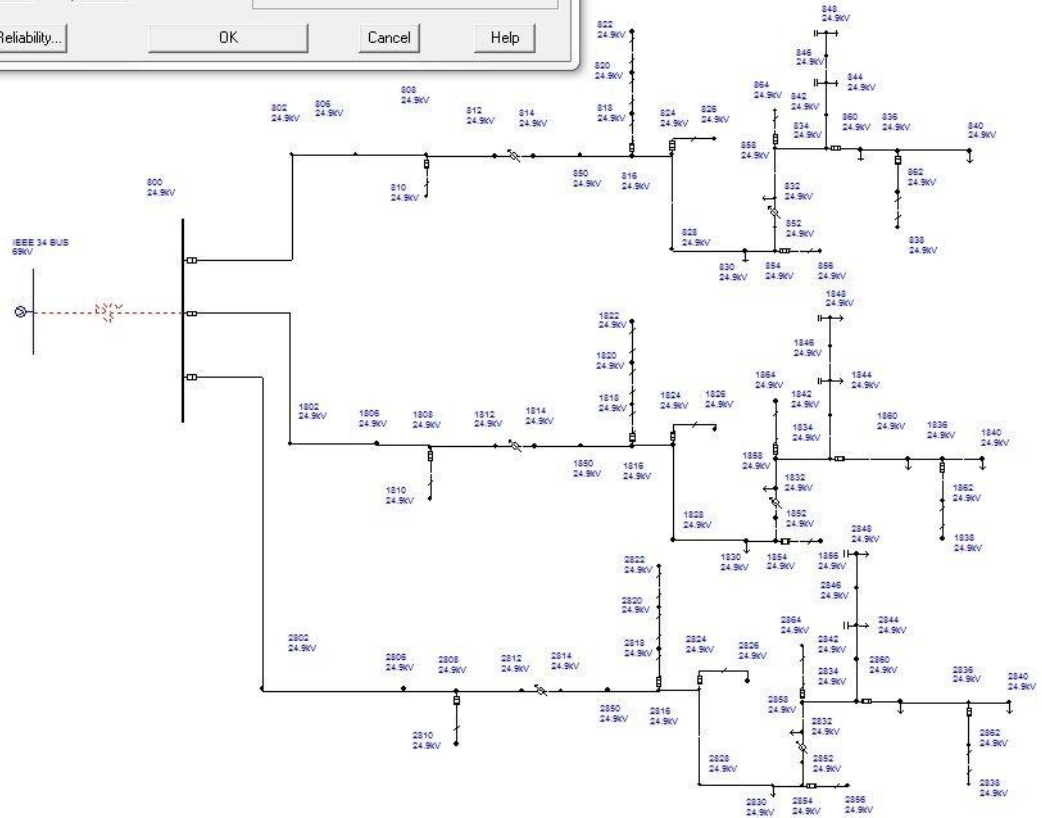
Grounding Impedance(s) in Ohms
0. +j 0.

800 24.9kV IEEE 34 BUS 69kV

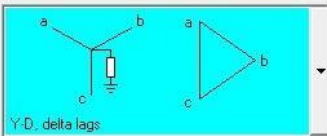
Swap sides

Tap kV L-L= 26.145 69.

LTC... Reliability... OK Cancel Help



2-Winding Transformer - Balanced



Name= TR 1 ID= 1

3-Phase kVA= 10000 10000 10000

Z+ (pu)= 0.005 +j 0.08

Zo (pu)= 0.005 +j 0.08

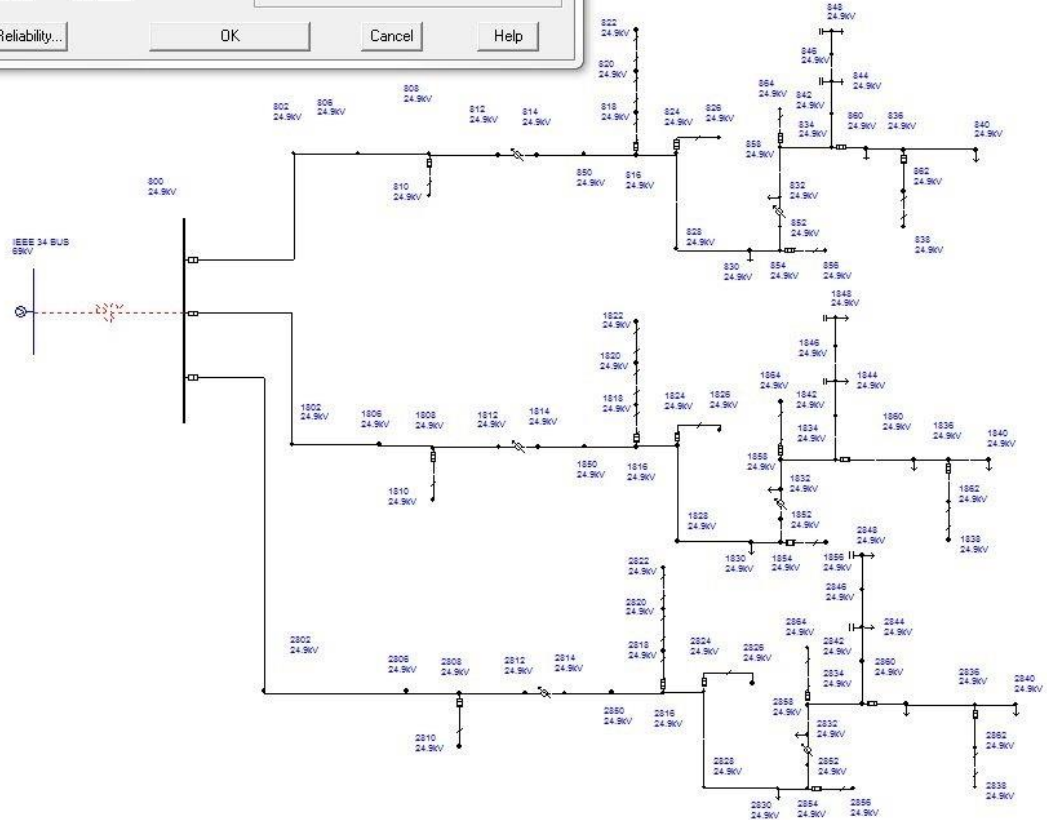
Grounding Impedance(s) in Ohms
0. +j 0.

800 24.9kV IEEE 34 BUS 69kV

Swap sides

Tap kV L-L= 26.145 69.

LTC... Reliability... OK Cancel Help



2-Winding Transformer - Balanced

Name= TR 1 ID= 1

3-Phase kVA= 10000. 10000. 10000.

Z+ (pu)= 0.005 +j 0.08

Zo (pu)= 0.005 +j 0.08

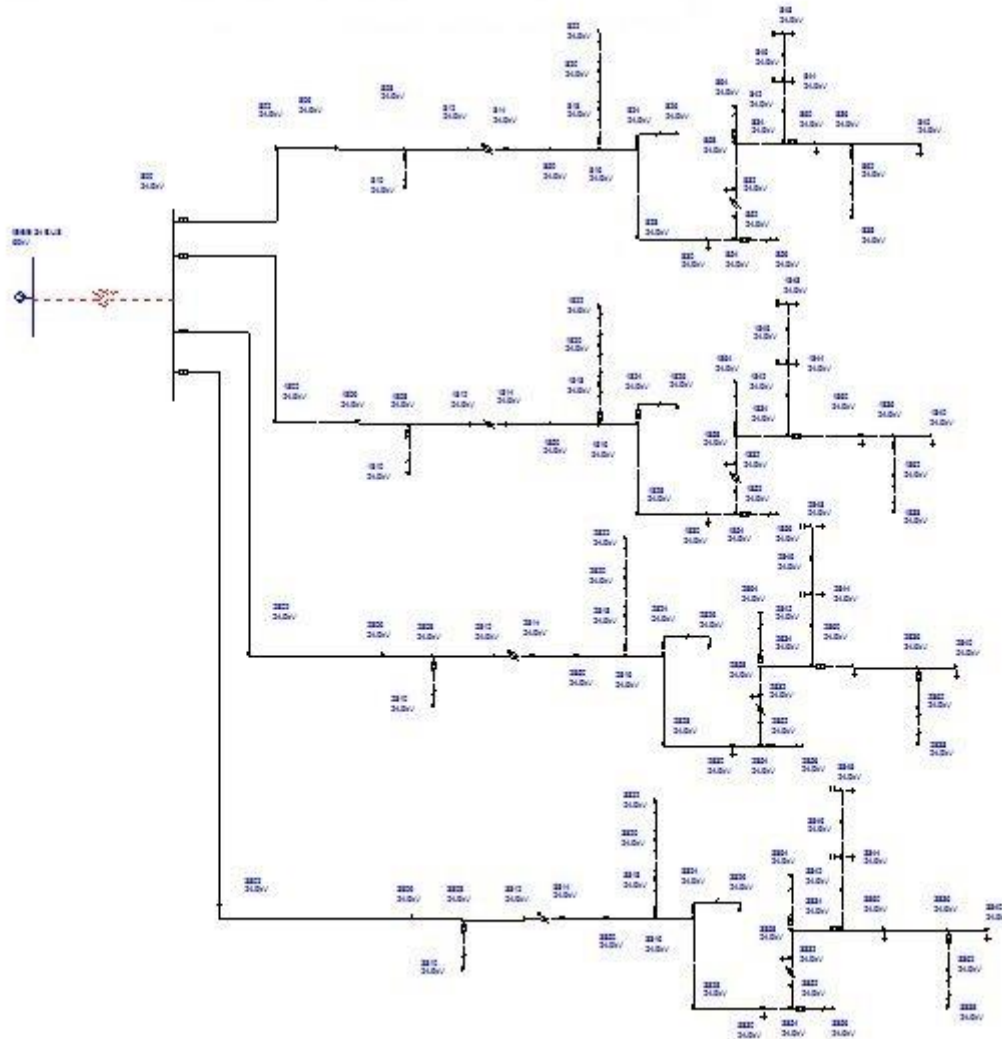
Grounding Impedance(s) in Ohms
0. +j 0.

800 24.9kV IEEE 34 BUS 69kV

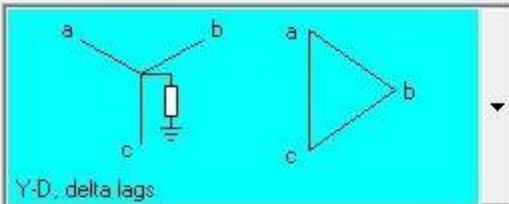
Swap sides

Tap kV L-L= 26.145 69.

LTC... Reliability... OK Cancel Help



2-Winding Transformer - Balanced



Y-D, delta lags

Name= TR 1 ID= 1

3-Phase kVA 12500. 12500. 12500.

Z+ (pu)= 0.005 +j 0.08

Zo (pu)= 0.005 +j 0.08

Grounding Impedance(s) in Ohms

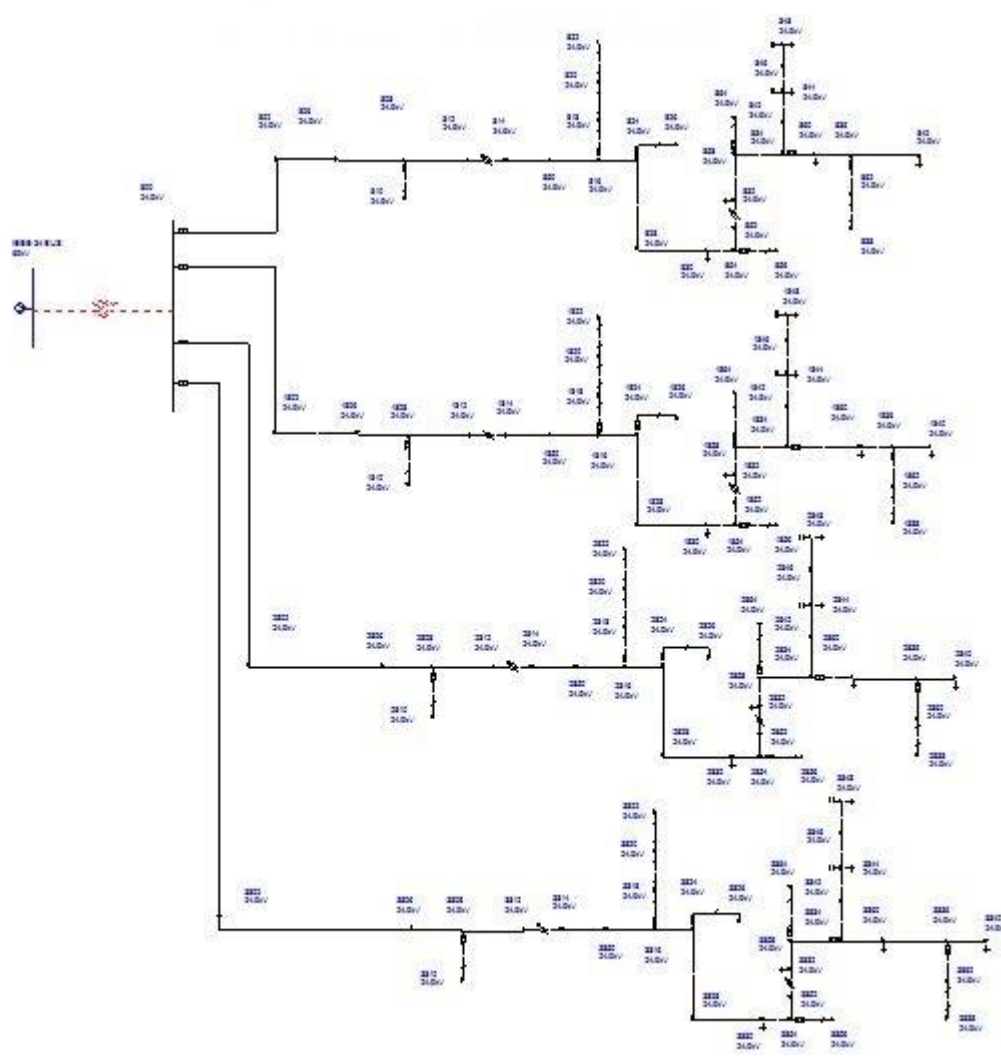
0. +j 0.

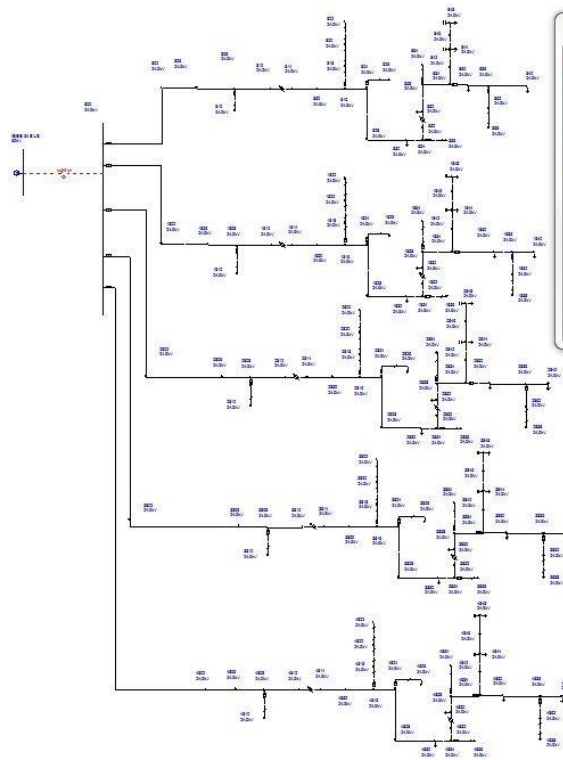
800 24.9kV IEEE 34 BUS 69kV

Swap sides

Tap kV L-L= 26.145 69.

LTC... Reliability... OK Cancel Help





2-Winding Transformer - Balanced

Y-D, delta lags

800 24.9kV IEEE 34 BUS 69kV

Swap sides

Tap kV L-L= 26.145 69.

LTC... Reliability... OK Cancel Help

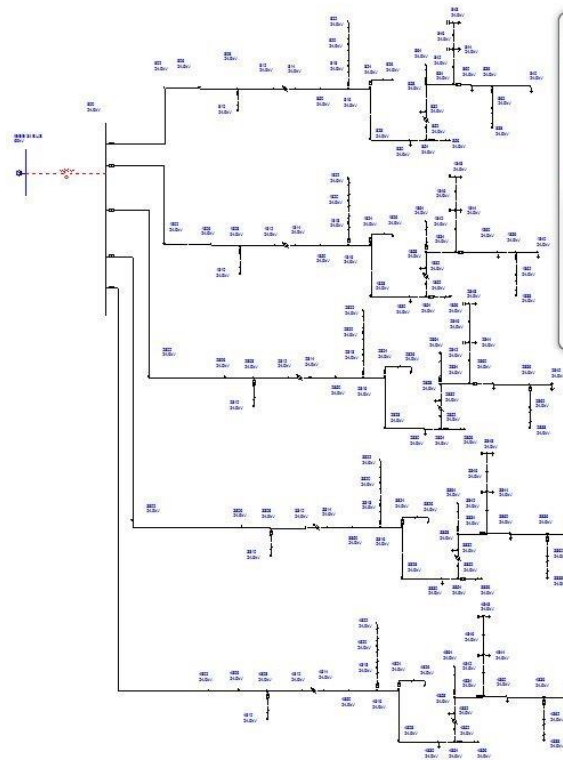
Name= TR 1 ID= 1

3-Phase kVA= 12500. 12500. 12500.

Z+ (pu)= 0.004 +j 0.08

Zo (pu)= 0.004 +j 0.08

Grounding Impedance(s) in Ohms
0. +j 0.



2-Winding Transformer - Balanced

Y-D, delta lags

800 24.9kV IEEE 34 BUS 69kV

Swap sides

Tap kV L-L= 26.145 69.

LTC... Reliability... OK Cancel Help

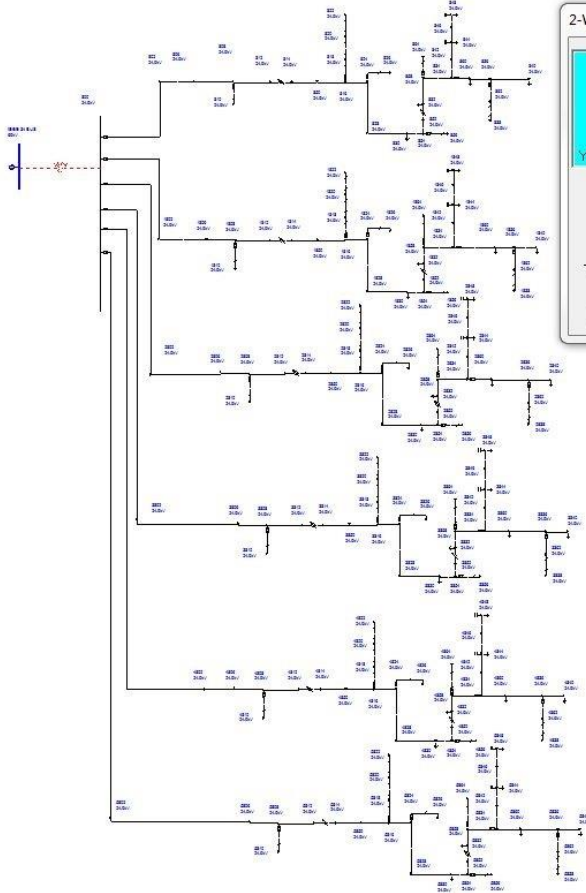
Name= TR 1 ID= 1

3-Phase kVA= 15000. 15000. 15000.

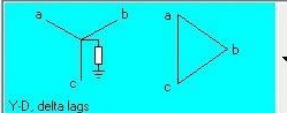
Z+ (pu)= 0.004 +j 0.08

Zo (pu)= 0.004 +j 0.08

Grounding Impedance(s) in Ohms
0. +j 0.



2-Winding Transformer - Balanced



Name= TR 1 ID= 1

3-Phase kVA= 15000. 15000. 15000.

Z+ (pu)= 0.004 +j 0.08

Zo (pu)= 0.004 +j 0.08

Grounding Impedance(s) in Ohms

0. +j 0.

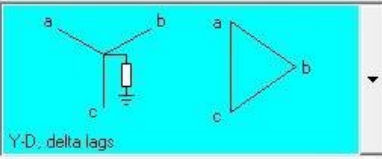
800 24.9kV IEEE 34 BUS 69kV

Swap sides

Tap kV L-L= 26.145 69.

LTC... Reliability... OK Cancel Help

2-Winding Transformer - Balanced



Y-D, delta lags

800 24.9kV IEEE 34 BUS 69kV

Swap sides

Tap kV L-L= 26.145 69

LTC... Reliability... OK Cancel Help

Name= TR 1 ID= 1

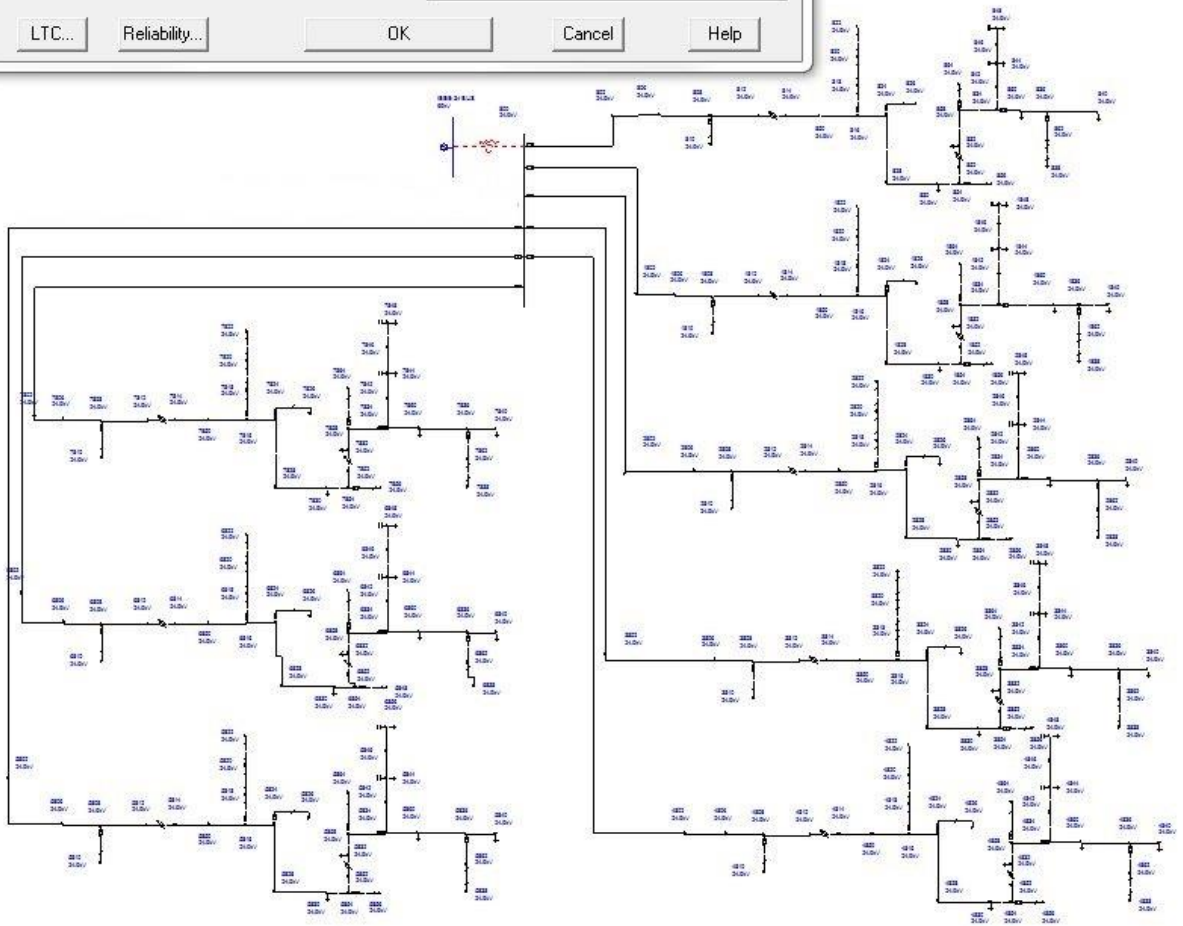
3-Phase kVA= 20000 20000 20000

Z+ (pu)= 0.004 +j 0.08

Zo (pu)= 0.004 +j 0.08

Grounding Impedance(s) in Ohms

0 +j 0



ANEXO C – Programa em Matlab

```

clear all
clc

%      AV  Taxa  AVxTaxa

Vars=[74  235  17467;
      74  126  9376;
      41  404  16713;
      174  90  15699;
      140  192  26804;
      94  206  19305;
      176  212  37266;
      240  184  44222;
      170  279  47304;
      69  474  32654;
      195  174  33889;
      114  97  11047;
      175  148  25894;
      160  203  32556;
      291  159  46241];

% pré-alocação de variáveis

membership=zeros(1500,5);
X=zeros(15,300);
k=1;
j=1;

while k<300
    X(:,k)=(Vars(:,1));
    f=0.03*randn(15,1);
    X(:,k+1)=Vars(:,2).*f+Vars(:,2);
    X(:,k+2)=(Vars(:,1).*X(:,k+1));
    S=linkage(X(:,k:k+2),'single','seuclidean');
    membership(j:j+14,1)=cluster(S,'maxclust',3);
    C=linkage(X(:,k:k+2),'complete','seuclidean');
    membership(j:j+14,2)=cluster(C,'maxclust',3);
    A=linkage(X(:,k:k+2),'average','seuclidean');
    membership(j:j+14,3)=cluster(A,'maxclust',3);
    W=linkage(X(:,k:k+2),'ward','seuclidean');
    membership(j:j+14,4)=cluster(W,'maxclust',3);
    membership(j:j+14,5)=kmeans(zscore(X(:,k:k+2)),3);
    k=k+3;
    j=j+15;
end

```