



**Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Curso de Engenharia de Energia**

**DESENVOLVIMENTO DE METODOLOGIA PARA  
DEFINIÇÃO DE LIMITES DE INDICADORES DE  
DESEMPENHO DE UNIDADES GERADORAS DE  
ENERGIA ELÉTRICA**

**Autoras: Juliany Martins da Silva  
Nathália Nóbrega de Oliveira Braga  
Orientador: Professor Doutor Augusto Brasil**

**Brasília, DF  
2014**



**JULIANY MARTINS DA SILVA  
NATHÁLIA NÓBREGA DE OLIVEIRA BRAGA**

**DESENVOLVIMENTO DE METODOLOGIA PARA DEFINIÇÃO DE LIMITES DE  
INDICADORES DE DESEMPENHO DE UNIDADES GERADORAS DE ENERGIA  
ELÉTRICA: SUBSÍDIO ÀS AÇÕES DE FISCALIZAÇÃO DA SUPERINTENDÊNCIA  
DE FISCALIZAÇÃO DOS SERVIÇOS DE GERAÇÃO DA AGÊNCIA NACIONAL  
DE ENERGIA ELÉTRICA**

Monografia submetida ao curso de graduação em Engenharia de Energia da Universidade de Brasília, como requisito parcial para obtenção do Título de Bacharel em Engenharia de Energia.

Orientador: Professor Doutor Augusto César Mendonça Brasil

Co-Orientador: Professor Doutor Marcus Vinícius Batistuta

**Brasília, DF  
2014**

## **CIP – Catalogação Internacional da Publicação**

Silva, Juliany M.; Braga, Nathália N. O.

Desenvolvimento de Metodologia para Definição de Limites de Indicadores de Desempenho de Unidades Geradoras de Energia Elétrica / Juliany Martins da Silva e Nathália Nóbrega de Oliveira Braga. Brasília: UnB, 2014. 103 p.: il ; 29,5 cm.

Monografia (Graduação) – Universidade de Brasília  
Faculdade do Gama, Brasília, 2014. Orientação: Augusto César  
Mendonça Brasil.

1. Indicadores de Desempenho. 2. Unidades Geradoras Hidrelétricas. 3. Limite. 4. Metodologia de Clusterização.  
I. Brasil, Augusto C. M. II. Desenvolvimento de Metodologia para Definição de Limites de Indicadores de Desempenho de Unidades Geradoras de Energia Elétrica.



**DESENVOLVIMENTO DE METODOLOGIA PARA DEFINIÇÃO DE LIMITES DE INDICADORES DE DESEMPENHO DE UNIDADES GERADORAS DE ENERGIA ELÉTRICA.**

**Juliany Martins da Silva  
Nathália Nóbrega de Oliveira Braga**

Monografia submetida como requisito parcial para obtenção do Título de Bacharel em Engenharia de Energia da Faculdade UnB Gama - FGA, da Universidade de Brasília, em (data da aprovação dd/mm/aa) apresentada e aprovada pela banca examinadora abaixo assinada:

---

**Prof. (Doutor): Augusto Brasil, UnB/ FGA**  
Orientador

---

**Prof. (Doutor): Marcus Vinícius Batistuta, UnB/ FGA**  
Co-Orientador

---

**Prof. (Doutor): Jorge Angarita Cormane, UnB/ FGA**  
Membro Convidado

---

**Prof. (Doutor): Luciano Gonçalves Noletto, UnB/ FGA**  
Membro Convidado

Brasília, DF  
2014

## AGRADECIMENTOS

Agradecemos ao nosso Orientador e Co-orientador, por toda a paciência, para todas as repetidas vezes que falamos: “Não, não entendi!”, por tomar este desafio para si e sempre nos estimular a fazer mais e melhor.

Agradecemos também aos Especialistas em Regulação Superintendência de Fiscalização dos Serviços de Geração da Agência Nacional de Energia Elétrica que nos orientaram e sempre estiveram dispostos a ajudar, em especial ao Issao Hirata, Sérgio Leite e ao Superintendente Alessandro Cantarino.

- Autoras.

Agradeço a Deus, início de tudo.

A minha mãe, início da minha vida.

Ao Anthony, início da minha nova vida.

À tia Marli, ao tio Paulo e aos irmãos Déborah e Gabriel, que viveram comigo essa nova vida.

À tia Pat, meu primeiro passo na UnB.

À professora Rita, minha mãe e minha mestra.

Ao Ti, irmão.

Ao meu pai Braga e à minha mãe Vera, que me deram uma família linda.

E a Julianny.

- Nathália Nóbrega.

Primeiramente agradeço a Deus, porque sinto a Tua presença a cada respiração e sei que é Ele a energia que permite que tudo aconteça.

A minha mãe, o Ser mais especial desse mundo, por ter sonhado os meus sonhos, confortado meu coração, enxugado minhas lágrimas e lutado muito para que nunca me faltasse oportunidade de novas conquistas.

A minha avó Francisca, minha mãezinha feita de amor, que sempre me traz a paz nos momentos que mais preciso.

Aos meus pais, Luciano e José Carlos, os melhores tios, meus amigos leais e exemplos de vida.

Ao professor Rudi por ter me mostrado o caminho e por ter estado sempre ao meu lado neste percurso.

Ao Netinho e ao Carlos Augusto que sempre me deram forças para fazer o meu melhor com reconhecimento e admiração.

E a minha amiga, companheira de todas as horas, sempre paciente e alegre, Nathália.

- Julianny Martins.

## RESUMO

A Agência Nacional de Energia Elétrica (ANEEL) é responsável pela fiscalização da produção de eletricidade no Brasil. Na intenção de suprir a necessidade da Superintendência de Fiscalização dos Serviços de Geração (SFG), foram definidos indicadores de desempenho utilizados para estimar a performance do parque de geração de energia elétrica interligado, o Sistema Elétrico Brasileiro (SEB). Este trabalho intenta desenvolver uma metodologia para analisar e avaliar esses indicadores de desempenho para grupos de unidades geradoras (UGs) que compõem as usinas hidrelétricas mais importantes do parque gerador, as classificadas como Tipo I pelo Operador Nacional do Sistema (ONS). Entre os vários indicadores, três foram selecionados e são usados neste trabalho: disponibilidade, taxa de falha e tempo médio de reparo. Atualmente, há limites definidos para esses indicadores com a intenção de classificar o desempenho das unidades geradoras, porém, esses limites foram definidos de forma empírica e temporária pelo ONS. Para definir tais limites formalmente, é necessária uma regulação formal da ANEEL e este trabalho tem a intenção de desenvolver uma metodologia científica para dar suporte a essa definição. Inicialmente, o parque elétrico foi agrupado em diferentes *clusters* considerando características físicas das unidades geradoras, como região, potência instalada, data de entrada em operação e tipo de turbina. Uma vez agrupadas, as unidades geradoras com características similares, foi desenvolvida uma comparação cujo objetivo era definir a performance mínima de desempenho para cada unidade.

Palavras-chave: Indicadores de Desempenho. Unidades Geradoras Hidrelétricas. Limite. Metodologia de Clusterização.

## **ABSTRACT**

The Brazilian Electricity Regulatory Agency (ANEEL) is responsible for the oversight of the electricity production in the country. In order to fulfill the need of the Superintendence of Generation Services Oversight (SFG), it was defined a set of performance indicators that are used to measure the performance of the power plants connected to the Brazilian national grid. This work aims to develop a methodology to analyze and evaluate these performance indicators for a group of generator units composed by the most important hydroelectric power plants in Brazil, those classified as Type I by the national grid operator (ONS). Among the various indicators, three were selected and used in this work: availability, failure rate and mean time to repair. Currently, there are limits defined for this indicators in order to classify the performance of the generating units, however, these limits were empirically and temporarily defined by the grid operator (ONS). To formally define such limits, it is necessary a formal regulation act from ANEEL and this work aimed to develop a scientific approach to support this definition. First, we grouped the power plants in different clusters, considering the physical characteristics of the generating units, such as region, installed power, age and type of turbine. Once the generating units were grouped with others with similar characteristics, we developed a comparative approach in order to define minimum performance levels for each unit.

**Keywords:** Performance Indicators. Hydroelectric Generating Units. Clustering Methodology.

## LISTA DE ILUSTRAÇÕES

Figura 1. Usina Hidrelétrica (UHE). Fonte: (GUIMARÃES, 2007).....	14
Figura 2. Planilha de Indicadores de Desempenho do RAD fornecida pelo ONS.....	31
Figura 3. Planilha confeccionada para utilização no Scilab: com apenas dados de interesse. .....	32
Figura 4. Distâncias Euclidiana e Manhattan. ....	36
Figura 5. Mapa de Manhattan. Fonte: < <a href="http://www.nuevayork.net/barrios-manhattan">http://www.nuevayork.net/barrios-manhattan</a> > acessado em 02 de junho de 2014. ....	37
Figura 6. Função <i>tabul</i> inserida no código do programa. ....	39
Figura 7. Saída da função <i>tabul</i> no console. ....	39
Figura 8. Programa Scilab para definição de faixas de potência e idade das UGs.....	41
Figura 9. Gráfico de idades com cálculo da distância entre pontos.....	41
Figura 10. Gráfico de potências com cálculo das distâncias entre pontos.....	42
Figura 11. Exemplo de Dendrograma. ....	47
Figura 12. Ilustração gráfica do método de ligação simples. ....	48
Figura 13. Ilustração gráfica do método de ligação completa.....	49
Figura 14. Benchmark. Fonte: (TANURE, 2000).....	52
Figura 15. Yardstick Competition. Fonte: (TANURE, 2000).....	53
Figura 16. Código do programa para escrita do arquivo *.csv.....	68
Figura 17. Arquivo escrito pelo comando <i>csvWrite</i> . ....	69
Figura 18. Planilha montada com os índices e <i>clusters</i> do arquivo gerado pelo Scilab e os parâmetros físicos. ....	70
Figura 19. Proporção de UGs em cada região no período de 1955 a 1984.....	72
Figura 20. Proporção de UGs em cada região no período de 1985 a 2012.....	72
Figura 21. Escolha do tipo de turbina de acordo com a queda e vazão do potencial e as potências obtidas. Fonte: < <a href="http://www.hacker.ind.br/produtos_turbinas_hidraulicas.php">http://www.hacker.ind.br/produtos_turbinas_hidraulicas.php</a> > acessado em 12 de maio de 2014. ....	73
Figura 22. Relação entre parâmetro físico Potência e indicador Disponibilidade. ....	75
Figura 23. Relação entre parâmetro físico tipo de turbina e indicador disponibilidade. ....	76
Figura 24. Relação entre parâmetro físico região e indicador disponibilidade.....	77
Figura 25. Relação entre parâmetro físico ano de entrada em operação e indicador disponibilidade.....	78
Figura 26. Relação entre parâmetro físico potência e indicador taxa de falhas.....	79
Figura 27. Relação entre parâmetro físico tipo de turbina e indicador taxa de falhas.....	80
Figura 28. Relação entre parâmetro físico região e indicador taxa de falhas. ....	81



Figura 29. “Curva da Banheira” .....	82
Figura 30. Relação entre parâmetro físico ano de entrada em operação e indicador taxa de falhas.....	83
Figura 31. Relação entre parâmetro físico potência e indicador tempo médio de reparo. ....	84
Figura 32. Relação entre parâmetro físico tipo de turbina e indicador tempo médio de reparo. ....	85
Figura 33. Relação entre parâmetro físico região e indicador tempo médio de reparo.....	86
Figura 34. Relação entre parâmetro físico ano de entrada em operação e indicador tempo médio de reparo.....	86
Figura 35. Gráfico de dispersão correlacionando a potência com o ano de entrada em operação com faixas de idade. ....	87
Figura 36. Gráfico de dispersão correlacionando a potência com o ano de entrada em operação com faixas de potência.....	88
Figura 37. Gráfico de dispersão correlacionando a potência com a região. ....	89
Figura 38. Gráfico de dispersão correlacionando a potência com o tipo de turbina.....	89
Figura 39. Gráfico de dispersão correlacionando o tipo de turbina com o ano de entrada em operação.....	91
Figura 40. Gráfico de dispersão correlacionando a região com o ano de entrada em operação.....	91
Figura 41. <i>Clusters</i> gerados no programa de Análise de <i>Clusters</i> .....	93
Figura 42. Projeção linear para definição de limites mínimos de disponibilidade para as UGs 147, 148, 149, 152, 171 e 79. ....	95
Figura 43. Projeção linear para definição de limites mínimos de taxa de falhas para as UGs 45, 82, 147 e 149.....	97
Figura 44. Projeção linear para definição de limites mínimos de disponibilidade para as UGs 147, 148, 149, 152, 171 e 79. ....	99

## LISTA DE SIGLAS

ANEEL .....	Agência Nacional de Energia Elétrica
CEDOC .....	Centro de Documentação da ANEEL
CGH .....	Microcentral de Geração Hidrelétrica
DISP .....	Indicador de Disponibilidade
LT .....	Linha de Transmissão
MME .....	Ministério de Minas e Energia
ONS .....	Operador Nacional do Sistema Elétrico
PCH .....	Pequena Central Hidrelétrica
RAD .....	Relatório de Avaliação de Desempenho de Manutenção
SAM .....	Sistema de Acompanhamento de Manutenção
SEB .....	Setor Elétrico Brasileiro
SFG .....	Superintendência de Fiscalização dos Serviços de Geração
SIN .....	Sistema Interligado Nacional
TCC .....	Trabalho de Conclusão de Curso
TF .....	Indicador Taxa de Falhas
TMR .....	Indicador Tempo Médio de Reparo
UHE .....	Usina Hidrelétrica
UG .....	Unidade Geradora

## LISTA DE TABELAS

Tabela 1. Proporção de valores indicados como normal e insatisfatório para DISP, TF e TMR de acordo com ONS e com o estudo de Fernandes et al. Fonte: (FERNANDES, COSTA e SILVA, 2013). .....	20
Tabela 2. Tabela com tempos de execução com e sem a função <i>gsort</i> . .....	37
Tabela 3. Valores de mediana para cada cluster para diferentes distribuições de elementos em função do número de clusters. ....	56
Tabela 4. Valores mid-range para cada cluster para diferentes distribuições de elementos em função do número de clusters. ....	56
Tabela 5. Projeção de valores para método de Regulação de Desempenho para Não Piorar (SRC, 2013). .....	57
Tabela 6. Distribuição dos elementos em <i>clusters</i> . .....	68
Tabela 7. Matriz de correlações com parâmetros físicos e indicadores. ....	71
Tabela 8. Média e desvio padrão da amostra para os indicadores DISP, TF e TMR. ....	74
Tabela 9. Média e desvio padrão da disponibilidade para faixas de potência. ....	75
Tabela 10. Média e desvio padrão da disponibilidade para tipos de turbina. ....	76
Tabela 11. Média e desvio padrão da disponibilidade para regiões. ....	77
Tabela 12. Média e desvio padrão da disponibilidade para faixas de idade. ....	78
Tabela 13. Média e desvio padrão da taxa de falhas para faixas de potência. ....	79
Tabela 14. Média e desvio padrão da taxa de falhas para tipos de turbina. ....	80
Tabela 15. Média e desvio padrão da taxa de falhas para regiões. ....	81
Tabela 16. Média e desvio padrão da taxa de falhas para faixas de idade. ....	83
Tabela 17. Média e desvio padrão do tempo médio de reparo para faixas de potência. ....	84
Tabela 18. Média e desvio padrão do tempo médio de reparo para tipos de turbina. ....	85
Tabela 19. Média e desvio padrão do tempo médio de reparo para regiões. ....	85
Tabela 20. Média e desvio padrão do tempo médio de reparo para faixas de idade. ....	87
Tabela 21. Valor de mediana, máximo e mínimo da disponibilidade do <i>cluster</i> 399. ....	94
Tabela 22. Dados de Disponibilidade para a UG 148 nos últimos 5 anos. ....	94
Tabela 23. Valores de Disponibilidade projetados para os próximos 5 anos e inclinação da reta de projeção. ....	94
Tabela 24. Dados de Disponibilidade para a UG 1 nos últimos 5 anos. ....	95
Tabela 25. Aplicação do método de regulação de desempenho para não piorar na UG 1, que apresenta bom desempenho. ....	96
Tabela 26. Valor de mediana, máximo e mínimo da taxa de falhas do <i>cluster</i> 399. ....	96
Tabela 27. Dados de Taxa de Falhas para a UG 139 nos últimos 5 anos. ....	96

Tabela 28. Valores de Taxa de Falhas projetados para os próximos 5 anos e inclinação da reta de projeção.....	96
Tabela 29. Dados de Taxa de Falhas para a UG 181 nos últimos 5 anos.....	97
Tabela 30. Aplicação do método de regulação de desempenho para não piorar na UG 181, que apresenta baixa taxa de falhas. ....	97
Tabela 31. Valor de mediana, máximo e mínimo do tempo médio de reparo do <i>cluster</i> 399. ....	98
Tabela 32. Dados de Tempo Médio de Reparo para a UG 43 nos últimos 5 anos. ....	98
Tabela 33. Valores de Tempo Médio de Reparo projetados para os próximos 5 anos e inclinação da reta de projeção. ....	98
Tabela 34. Dados de Tempo Médio de Reparo para a UG 9 nos últimos 5 anos. ....	99
Tabela 35. Aplicação do método de regulação de desempenho para não piorar na UG 9, que apresenta baixo tempo médio de reparo.....	99

## LISTA DE QUADROS

Quadro 1. Valores limites definidos pelo ONS para as faixas de qualidade de equipamentos e LT. ....	19
Quadro 2. Usinas e respectivas unidades geradoras excluídas da amostra em análise por filtragem de dados. ....	33
Quadro 3. Faixas de Potência e de Ano de Entrada em Operação. ....	42

## SUMÁRIO

<b>AGRADECIMENTOS</b> .....	<b>5</b>
<b>LISTA DE ILUSTRAÇÕES</b> .....	<b>8</b>
<b>LISTA DE SIGLAS</b> .....	<b>10</b>
<b>LISTA DE TABELAS</b> .....	<b>11</b>
<b>LISTA DE QUADROS</b> .....	<b>13</b>
<b>1. INTRODUÇÃO</b> .....	<b>10</b>
1.1. ASPECTOS GERAIS .....	11
1.1.1. A Agência Nacional de Energia Elétrica – ANEEL.....	11
1.1.2. O Operador Nacional do Sistema Elétrico – ONS .....	12
1.1.3. O Sistema Interligado Nacional – SIN.....	13
1.1.4. Usinas Hidrelétricas.....	13
1.1.5. A Superintendência de Fiscalização dos Serviços de Geração .....	15
1.1.6. Delimitação do objeto de estudo .....	15
1.1.7. A aquisição dos dados .....	18
1.1.8. A escolha dos Indicadores .....	18
1.1.9. Os valores limites dos Indicadores de Desempenho .....	19
<b>2. JUSTIFICATIVA</b> .....	<b>21</b>
<b>3. OBJETIVO</b> .....	<b>22</b>
3.1. OBJETIVO GERAL .....	22
3.2. OBJETIVOS ESPECÍFICOS .....	22
<b>4. CASOS INTERNACIONAIS</b> .....	<b>23</b>
4.1. PORTUGAL .....	23
4.2. ARGENTINA .....	24
4.3. COLÔMBIA .....	24
4.4. CONCLUSÃO.....	25
<b>5. METODOLOGIA</b> .....	<b>26</b>
5.1. CÁLCULO DOS INDICADORES DE DESEMPENHO .....	28
5.1.1. Disponibilidade .....	28
5.1.2. Taxa de Falhas.....	29
5.1.3. Tempo Médio de Reparo.....	29
5.2. O RELATÓRIO DE AVALIAÇÃO DO DESEMPENHO DE MANUTENÇÃO.....	30
5.2.1. Tratamento dos dados.....	31
5.2.2. Aplicação de filtros .....	32
5.2.2.1. Filtro para DISP = 0, TF = 8760 e TMR = 8760 .....	32
5.2.2.2. Filtro para DISP = 0, TF = 0 e TMR = 0 .....	33

5.2.3.	Normalização.....	34
5.2.4.	Medidas de Similaridade e Dissimilaridade.....	34
5.2.5.	A Função <i>gsort</i> .....	37
5.2.6.	A Função <i>tabul</i> .....	38
5.3.	DEFINIÇÃO DAS FAIXAS DE POTÊNCIA E IDADE.....	40
5.4.	CLUSTERIZAÇÃO.....	42
5.4.1.	<i>K-Means</i> .....	45
5.4.2.	Método Hierárquico.....	45
5.4.2.1.1.	<i>Método de Ligação Simples (Single Linkage)</i> .....	48
5.4.2.1.2.	<i>Método de Ligação Completa (Complete Linkage)</i> .....	49
5.4.3.	Vantagens e Desvantagens: <i>K-Means</i> ou Método Hierárquico?.....	51
5.5.	METODOLOGIA PARA REGULAÇÃO POR DESEMPENHO.....	51
5.5.1.	Benchmark.....	51
5.5.2.	Yardstick Competition.....	52
5.5.3.	Método de Regulação de Desempenho para Não Piorar.....	57
<b>6.</b>	<b>O PROGRAMA.....</b>	<b>59</b>
6.1.	PROGRAMA PARA GRÁFICOS DE ANÁLISE DE CORRELAÇÃO.....	59
6.2.	PROGRAMA PARA ANÁLISE DE <i>CLUSTERS</i> .....	65
6.2.1.	Escolha do Número de <i>Clusters</i> .....	67
6.2.2.	Identificação das Unidades Geradoras de Cada <i>Cluster</i> .....	68
<b>7.</b>	<b>CORRELAÇÃO E SELEÇÃO DE VARIÁVEIS.....</b>	<b>71</b>
7.1.	INFLUÊNCIA DOS PARÂMETROS FÍSICOS SOBRE OS INDICADORES.....	73
7.1.1.	Disponibilidade X Potência.....	74
7.1.2.	Disponibilidade X Tipo de Turbina.....	75
7.1.3.	Disponibilidade X Região.....	76
7.1.4.	Disponibilidade X Ano de Entrada em Operação.....	77
7.1.5.	Taxa de Falhas X Potência.....	78
7.1.6.	Taxa de Falhas X Tipo de Turbina.....	79
7.1.7.	Taxa de Falhas X Região.....	80
7.1.8.	Taxa de Falhas X Ano de Entrada em Operação.....	81
7.1.9.	Tempo Médio de Reparo.....	83
7.2.	CORRELAÇÃO ENTRE PARÂMETROS FÍSICOS.....	87
7.2.1.	Potência X Ano de Entrada em Operação.....	87
7.2.2.	Potência X Região.....	88
7.2.3.	Potência X Tipo de Turbina.....	89
7.2.4.	Tipo de Turbina X Ano de Entrada em Operação.....	90
7.2.5.	Região X Ano de Entrada em Operação.....	91
<b>9.</b>	<b>RESULTADOS E DISCUSSÃO.....</b>	<b>93</b>
9.1.	ANÁLISE DE <i>CLUSTERS</i> .....	93

9.2.	DEFINIÇÃO DOS LIMITES MÍNIMOS .....	93
9.2.1.	Disponibilidade .....	94
9.2.2.	Taxa de Falhas.....	96
9.2.3.	Tempo Médio de Reparo.....	98
9.3.	AVALIAÇÃO DAS UNIDADES GERADORAS .....	99
<b>11.</b>	<b>CONSIDERAÇÕES FINAIS.....</b>	<b>101</b>
<b>12.</b>	<b>SUGESTÃO DE CONTINUIDADE.....</b>	<b>102</b>
12.1.	ANÁLISE POR USINA.....	102
12.2.	OUTROS PARÂMETROS .....	102
12.3.	TÉCNICAS HIERÁRQUICAS AGLOMERATIVAS E MEDIDAS DE SIMILARIDADE 103	
12.4.	ANÁLISE PARA OUTROS EMPREENDIMENTOS DE GERAÇÃO DE ENERGIA 103	
12.5.	REALIZAÇÃO DE CONSULTA PÚBLICA PARA CONTRIBUIÇÃO DOS AGENTES DE GERAÇÃO .....	103



## 1. INTRODUÇÃO

Foram elaborados pelo Operador Nacional do Sistema Elétrico (ONS), com participação dos agentes de geração, e aprovados pela Agência Nacional de Energia Elétrica (ANEEL), documentos de caráter normativo, os Procedimentos de Rede, que definem os procedimentos e os requisitos necessários à realização das atividades de planejamento da operação eletroenergética, de administração da transmissão, de programação e de operação em tempo real no âmbito do Sistema Interligado Nacional (SIN). Nos Procedimentos de Redes são definidos diversos Indicadores de Desempenho, dentre eles, os Indicadores de Desempenho de Equipamentos e Linhas de Transmissão e das Funções Transmissão e Geração (ONS, 2011).

Provisoriamente, o ONS estabeleceu os limites dos Indicadores de Desempenho Disponibilidade (DISP), Taxa de Falhas (TF) e Tempo Médio de Reparo (TMR) para avaliação dos equipamentos e linhas de transmissão (LT) pertencentes às instalações dos agentes geradores e transmissores de energia elétrica do SIN. No conjunto, os valores mensais dos Indicadores DISP, TF e TMR sinalizam alguns aspectos importantes relativos à manutenção de um equipamento; a saber: disponibilidade, ausência de falhas e tempo de reparo para desligamentos forçados ou não programados. (FERNANDES, COSTA e SILVA, 2013)

Os limites ou valores de corte estabelecidos pelo ONS para cada um dos Indicadores DISP, TF e TMR resultam em três faixas de qualidade: normal, alerta e insatisfatória. Mensalmente, o ONS, por meio do Sistema de Acompanhamento de Manutenção – SAM, obtém os Indicadores de Desempenho das usinas hidrelétricas (UHEs), pequenas centrais hidrelétricas (PCHs) e usinas termelétricas (UTES), gerando o Relatório de Avaliação de Desempenho de Manutenção (RAD), com as análises de desempenho de manutenção dos equipamentos e LT da rede básica e das usinas programadas e despachadas centralizadamente pelo ONS.

Observando-se mais atentamente, a análise e interpretação desses resultados, considerando equipamentos de uma mesma classe, e as diversas instalações de um mesmo agente, permitem ao gestor uma visão ampla sobre o estado de manutenção dos equipamentos, instalações e agentes de geração e transmissão no país. Trata-se, portanto, de um instrumento de gestão de suma

importância para garantir a estabilidade do SIN, mitigando os seus riscos (KOÇOUSKI, 2011).

O presente estudo tem o intuito de suprir a demanda da SFG (Superintendência de Fiscalização dos Serviços de Geração) por uma fiscalização complementar às fiscalizações *in loco*, realizadas para inspeções da estrutura física dos empreendimentos e agentes correlacionados à geração de energia elétrica, e o monitoramento à distância, por meio de verificação da regularidade de documentos apresentados pelos agentes. A fiscalização por meio dos Indicadores de Desempenho possibilita o monitoramento de forma contínua, sem necessidade de deslocamento da equipe até a usina. A estrutura física de uma usina deve ser fiscalizada por diversos motivos, porém, essa fiscalização nem sempre é suficiente para garantir sua manutenção e seu bom desempenho.

## 1.1. ASPECTOS GERAIS

### 1.1.1. A Agência Nacional de Energia Elétrica – ANEEL

O amplo conjunto de reformas econômicas no Brasil na década de 90 acarretou conseqüentemente em mudanças no Setor Elétrico Brasileiro (SEB). No contexto geral, a diminuição expressiva da atuação empreendedora do Estado transferiu sua responsabilidade principal para o campo da regulação e fiscalização dos serviços enquanto foram delegadas à iniciativa privada as atividades econômicas que exigem regime especial. É nesse contexto que surgem as agências reguladoras, entre elas, a ANEEL.

As agências reguladoras são autarquias especiais caracterizadas por autonomia em relação ao poder público. A regulação contempla uma gama mais ampla de atribuições, relacionadas ao desempenho de atividades econômicas e à prestação de serviços públicos, incluindo diretrizes, fiscalização, intermediação em conflito e aplicação de eventuais sanções.

A ANEEL, criada em 1996, é uma autarquia sob regime especial, vinculada ao Ministério de Minas e Energia (MME), com a finalidade de regular e fiscalizar a produção, transmissão e comercialização de energia elétrica em conformidade com as políticas e diretrizes do Governo Federal. A ela compete garantir tarifas justas, zelar pela qualidade do serviço, exigir os investimentos necessários para a

adequada prestação do serviço, arbitrar conflito de interesses, estimular a competição em condições leais, assegurar a universalidade dos serviços, fiscalizar de forma ampla e promover a defesa do interesse do cidadão-consumidor.

Mais especificamente, entre as competências da ANEEL, está o estabelecimento de mecanismos de regulação e fiscalização para garantir o atendimento à totalidade do mercado de cada agente de distribuição e de comercialização de energia elétrica. É também função da agência intervir na prestação do serviço de energia elétrica e regular o serviço concedido, fiscalizando permanentemente sua prestação. Nessas competências se encaixa o objetivo desse trabalho de contribuir para a fiscalização da prestação do serviço de geração de energia elétrica.

### **1.1.2. O Operador Nacional do Sistema Elétrico – ONS**

Após as mudanças ocorridas na década de 90 no Setor Elétrico Brasileiro, o governo federal lançou bases de um novo modelo para o SEB entre 2003 e 2004. Em termos institucionais, o novo modelo definiu a criação de uma entidade responsável pelo planejamento do setor elétrico, a Empresa de Pesquisa Energética (EPE); uma instituição com a função de avaliar permanentemente a segurança do suprimento de energia elétrica, o Comitê de Monitoramento do Setor Elétrico (CMSE); e uma instituição para dar continuidade às atividades do Mercado Atacadista de Energia (MAE), relativas à comercialização de energia elétrica no Sistema Interligado, a Câmara de Comercialização de Energia Elétrica (CCEE). Outras alterações importantes incluem a definição do exercício do Poder Concedente ao Ministério de Minas e Energia (MME) e a ampliação da autonomia do Operador Nacional do Sistema Elétrico (ONS, 2013).

O Operador Nacional do Sistema Elétrico é o órgão responsável pela coordenação e controle da operação das instalações de geração e transmissão de energia elétrica no Sistema Interligado Nacional (SIN), sob a fiscalização e regulação da Agência Nacional de Energia Elétrica (ANEEL). Para o exercício de suas atribuições legais e o cumprimento de sua missão institucional, o ONS desenvolve uma série de estudos e ações a serem exercidas sobre o sistema e seus agentes para manejar o estoque de energia de forma a garantir a segurança do suprimento contínuo em todo o país. O Operador Nacional é constituído por

membros associados e membros participantes, constituídos por empresas de geração, transmissão, distribuição e consumidores livres de grande porte. Também participam importadores e exportadores de energia, além do MME (ONS, 2013).

### **1.1.3. O Sistema Interligado Nacional – SIN**

Com tamanho e características que permitem considerar o SIN único em âmbito mundial, o sistema de produção e transmissão de energia elétrica do Brasil é um sistema hidrotérmico de grande porte, com forte predominância de usinas hidrelétricas e com múltiplos proprietários. O SIN é formado pelas empresas das regiões Sul, Sudeste, Centro-Oeste, Nordeste e parte da região Norte. Apenas 1,7% da energia requerida pelo país encontra-se fora do SIN, em pequenos sistemas isolados localizados principalmente na região amazônica (ONS, 2013).

### **1.1.4. Usinas Hidrelétricas**

As usinas hidrelétricas são as principais fontes geradoras de energia elétrica no Brasil, com participação superior a 85% de toda geração do SIN – Sistema Interligado Nacional. Geralmente, quando projetadas, as usinas hidrelétricas apresentam mais de uma unidade geradora, para melhor aproveitamento na geração de energia no caso de paradas devido às manutenções programadas e forçadas. Os revezamentos das manutenções programadas das unidades propiciam maior confiabilidade para o SIN (GUIMARÃES, 2007).

Estas unidades geradoras são formadas por centenas de equipamentos que compõem o conjunto e segundo um programa de manutenção suas manutenções preventivas são realizadas periodicamente.

As usinas hidrelétricas, Figura 1, são basicamente formadas por: barragem, sistema de adução, casa de força e vertedouro, que apresentam as seguintes funções segundo Tolmasquim (2005):

- a) Barragem – Armazenar água que é a matéria prima para gerar energia nos períodos de cheia e formação da queda d'água quando não houver desnível adequado para geração. O acúmulo desta água forma o reservatório que pode propiciar o controle das vazões naturais dos rios afluentes, condição esta estratégica na geração de energia.

- b) Sistema de adução – Captar a água do reservatório através de canais, túneis ou condutos metálicos, e conduzi-la até as turbinas que estão na casa de força.
- c) Casa de Força – Na casa de força estão as turbinas, que através da água vinda do sistema de adução transformam a energia cinética em mecânica, e o gerador transforma a energia mecânica em eletricidade. A água após passar pelas turbinas sai no canal de fuga, posteriormente voltando ao leito do rio seguindo seu curso normal.
- d) Vertedouro – Controlar o escoamento e medição das vazões, liberar o excesso do fluxo d'água que não pode ser armazenado no reservatório e absorvido pela geração da hidrelétrica, conseqüentemente garantindo a integridade da barragem devido a vazões máximas, assim mostrando ser fundamental para segurança de toda hidrelétrica.



Figura 1. Usina Hidrelétrica (UHE). Fonte: (GUIMARÃES, 2007).

A energia hidráulica ocorre da irradiação solar e da energia potencial gravitacional, sendo que o Sol e a força da gravidade colaboram com a evaporação, a condensação e a precipitação da água sobre a superfície terrestre. Assim, a água das chuvas em harmonia com a gravidade flui para os leitos dos rios e através do seu movimento produz energia cinética, que posteriormente pode ser convertido em

mecânica quando em contato com as turbinas das usinas hidrelétricas ligadas ao rotor de um gerador que converte a energia mecânica em energia elétrica.

#### **1.1.5. A Superintendência de Fiscalização dos Serviços de Geração**

Dentre as 25 superintendências que compõem a Agência Nacional de Energia Elétrica, a Superintendência de Fiscalização dos Serviços de Geração (SFG) está incumbida de fiscalizar todos os empreendimentos de geração de energia elétrica do país, contemplando tanto usinas em fase de construção, quanto em fase de operação (KOÇOUSKI, 2011). O principal objetivo da SFG é analisar (através do acompanhamento de relatórios fornecidos pelas empresas, fiscalização em escritório, monitoramento à distância e fiscalização em campo) se tais empreendimentos estão em conformidade com a legislação e com as normas vigentes dos procedimentos de operação, manutenção, conservação e segurança operacional.

#### **1.1.6. Delimitação do objeto de estudo**

Dentre os dispositivos de geração de energia elétrica (eólica, solar, hidráulica, termelétrica) fiscalizados pela SFG, foi escolhida a forma de geração hidráulica para ser abordada neste trabalho. Subdividindo-a, tem-se as microcentrais de geração hidrelétrica (CGHs) as pequenas centrais hidrelétricas e as usinas hidrelétricas. Essa classificação é feita de acordo com o potencial hidrelétrico de cada empreendimento e, de forma geral, as CGHs tem potência instalada de até 1MW, enquanto as PCHs trabalham com até 30MW e as UHEs compreendem os empreendimentos que trabalham acima de 30MW (ANEEL, 2008).

Os empreendimentos também são classificados, pelo ONS, a partir de sua modalidade de operação. De acordo com o Submódulo 26.1, dos Procedimentos de Rede do ONS: a modalidade de operação da usina é definida a partir da avaliação dos impactos verificados tanto na operação hidráulica e energética do SIN, como também na segurança da rede de operação (ONS, 2011).

As modalidades de operação de usinas são classificadas em 3 Tipos:

- a) Tipo I: serão consideradas nesta modalidade as usinas conectadas na rede básica – independente da potência líquida injetada no SIN e da natureza da

fonte primária; ou usinas cuja operação hidráulica possa afetar a operação de usinas Tipo I já existentes; ou ainda usinas conectadas fora da rede básica cuja máxima potência líquida injetada no SIN contribua para minimizar problemas operativos e proporcionar maior segurança para a rede de operação.

- b) Tipo II: serão consideradas nesta modalidade as usinas não classificadas como Tipo I, mas que afetam os processos de planejamento, programação da operação, operação em tempo real, normatização, pré-operação e pós-operação. As usinas classificadas nesta modalidade de operação se caracterizam por não ter de forma rotineira o seu programa de geração estabelecido de forma coordenada e centralizada pelo ONS, nem o seu despacho de geração coordenado estabelecido ou controlado pelo ONS em tempo real.
- c) Tipo III: as usinas, individualmente, não classificadas nas modalidades anteriores.

A Resolução Normativa N° 67, de 8 de junho de 2004 define as Instalações de Transmissão que integram a Rede Básica do Sistema Interligado Nacional; são estas: linhas de transmissão, barramentos, transformadores de potência e equipamentos de subestação em tensão igual ou superior a 230 kV, transformadores de potência com tensão primária igual ou superior a 230 kV e tensões secundária e terciária inferiores a 230 kV, bem como as respectivas conexões e demais equipamentos ligados ao terciário, a partir de 1° de julho de 2004.

A mesma Resolução define que são excluídos da rede básica, quando de uso de centrais geradoras, em caráter exclusivo ou compartilhado, ou de consumidores livres em caráter exclusivo: equipamentos em tensão menor que 230 kV; quando de uso exclusivo para importação e/ou exportação de energia elétrica; instalações de transmissão de energia elétrica destinadas a interligações internacionais e, ainda, instalações de transmissão de energia elétrica destinadas a interligações internacionais, conforme art. 21 do Decreto nº 7.246, de 28 de julho de 2010. (ANEEL, 2004)

Atualmente, o Brasil possui 197 UHEs em operação e 6 em construção (ANEEL, 2014). Dentre as usinas em operação, 133 usinas serão analisadas nesse trabalho, visto que são analisadas apenas as usinas hidrelétricas com modalidade

de operação do Tipo I, totalizando 549 unidades geradoras. O trabalho também se restringe às usinas que estão em operação e tem um histórico de dados de todas as suas unidades geradoras ou de parte delas e, particionando ainda mais o objeto do estudo, serão analisadas as unidades geradoras por usina, e não as usinas no geral, porém, a partir dos resultados obtidos para cada UG é possível expandir a análise para a usina, bem como para o agente de geração.

O agente de geração pode ser responsável por um ou mais empreendimentos de geração de energia, e a análise deste, considerando todos os seus empreendimentos permite uma visão da gestão realizada pelo agente e seu desempenho geral. Porém, visto que o desempenho de cada unidade varia, faz-se necessário conhecer o comportamento das unidades geradoras de uma usina antes de avaliá-la como um todo.

O Operador Nacional do Sistema Elétrico, através do Submódulo 25.8 dos Procedimentos de Rede apresenta Indicadores de Desempenho dos equipamentos: unidades geradoras (UGs), Compensadores Síncronos e Compensadores Estáticos, Transformadores e Linhas de Transmissão, além das Funções Geração e Transmissão. Dentre esses conjuntos de equipamentos, escolheu-se analisar as unidades geradoras, pois se acredita que a análise delas é suficiente para subsidiar as ações de fiscalização da SFG.

A unidade geradora é o equipamento eletromecânico mais importante de um empreendimento hidrelétrico. É composta, basicamente, pela turbina hidráulica e pelo gerador e, como equipamentos complementares principais: o regulador, a válvula, o quadro de comando e o volante. (ELETROBRÁS, 1985) No Submódulo 20.1 dos Procedimentos de Rede, uma unidade geradora é descrita como o equipamento instalado entre a tomada de água, o tubo de sucção e a bucha de baixa tensão do transformador elevador.

A Resolução Normativa nº 583, de 22 de outubro de 2013, que define a sistemática de medição da potência instalada e líquida em empreendimentos de geração de energia elétrica, define a unidade geradora como o conjunto de gerador conjugado a máquina(s) motriz(es) e respectivos equipamentos, que tenham como destino a conversão outras formas de energia em energia elétrica.



### **1.1.7. A aquisição dos dados**

Os dados utilizados na elaboração da análise de Indicadores feita nesse trabalho são fornecidos pelo ONS e pelos relatórios de fiscalização das usinas que constam arquivados na SFG. O ONS forneceu os números relativos aos três Indicadores DISP, TMR e TF, considerando um histórico de dados de 2008 a 2013, bem como a potência instalada, a quantidade de unidades geradoras e a região onde se localiza a usina, de acordo com a divisão do Sistema Interligado Nacional (SIN). Tais informações, salvo os valores dos três Indicadores, também se encontram nos processos de cada empreendimento, processos esses que também continham o tipo de turbina de cada UG e sua data de entrada em operação comercial.

### **1.1.8. A escolha dos Indicadores**

Entre os vários Indicadores de Desempenho para equipamentos e linhas de transmissão apresentados nos Procedimentos de Rede, Submódulo 25.8, do ONS, estão:

- a) Disponibilidade;
- b) Indisponibilidade para Manutenção Programada;
- c) Indisponibilidade para Manutenção Forçada;
- d) Taxa de Desligamento Forçado;
- e) Taxa de Falhas;
- f) Tempo Médio de Reparo.

O ONS definiu que teriam os seus limites estipulados e seriam apresentados no RAD os Indicadores: Disponibilidade, Taxa de Falhas e Tempo Médio de Reparo. Consequentemente, esses fazem parte do objeto de estudo deste trabalho.

A Disponibilidade indica a porção ou a porcentagem de tempo que o equipamento esteve em operação ou apto para operar no período de observação, com restrição ou não. Entende-se por Taxa de Falhas, na perspectiva de manutenção, a expressão da incidência de falhas nas horas de serviço de um equipamento ou um conjunto de equipamentos, no período considerado, referida para um ano padrão de 8.760 (oito mil e setecentos e sessenta) horas. O Tempo

Médio de Reparo inclui também a possível substituição pela equipe de manutenção do agente e é o tempo gasto, em média para que os equipamentos voltem a desempenhar as funções para as quais foi projetado. (ONS, 2011, p. Submódulo 20.1)

### 1.1.9. Os valores limites dos Indicadores de Desempenho

Os valores limites que definem as faixas de “normal”, “alerta” e “insatisfatório” para cada indicador de desempenho foram definidos provisoriamente pelo ONS e são mostrados no Quadro 1. Esses valores juntamente com os valores apresentados pelas usinas no RAD foram analisados por Fernandes et al, (2013). e foi constatado que para as unidades geradoras o valor mais bem ajustado às condições atuais das usinas hidrelétricas é o TMR; pois, para os Indicadores DISP e TF, os valores encontrados pelo estudo e os limites adotados pelo ONS foram bastante divergentes.

Quadro 1. Valores limites definidos pelo ONS para as faixas de qualidade de equipamentos e LT.

<b>Indicador</b>	<b>Normal</b>	<b>Alerta</b>	<b>Insatisfatório</b>
DISP	Maior que 95%	Entre 90% e 95%	Menor do que 90%
TF	Menor do que 2	Entre 2 e 4	Maior do que 4
TMR	Menor do que 180h	Entre 180 e 240h	Maior do que 240h

A metodologia utilizada por Fernandes et al. para definir o valor limite dos Indicadores de Desempenho propunha a utilização de um percentil de 5% de UGs hidrelétricas com valores indesejáveis de Disponibilidade, de Taxa de Falhas e de Tempo Médio de Reparo. Nesse contexto, a Tabela 1 apresenta, segundo a classificação do trabalho de Fernandes et al., que a maioria dos empreendimentos (95%) estaria na faixa normal para os valores de DISP e TF enquanto que, de acordo com o critério do ONS somente 60,3% delas estariam nessa faixa. Tamanha diferença mostra que o limite usado pelo ONS para DISP e TF é muito mais conservador do que o limite proposto por Fernandes et al.

Os valores conservadores definidos pelo ONS apontam que 21,9% das UGs tem os Indicadores DISP e TF num patamar insatisfatório, enquanto o percentil definido por Fernandes et al. limita a quantidade de unidades geradoras com o valores indesejáveis em 2,5%.

O valor de TMR apresenta um percentil divergente dos 5%, de faixa normal, relativos à DISP e à TF; pois o valor definido pelo ONS demonstrou ser menos conservador para esse indicador do que poderiam ser os valores de acordo com o estudo. O mesmo vale para a faixa insatisfatória do TMR, que se diferencia por ser menos conservadora do que o valor dado por Fernandes et al.

Tabela 1. Proporção de valores indicados como normal e insatisfatório para DISP, TF e TMR de acordo com ONS e com o estudo de Fernandes et al. Fonte: (FERNANDES, COSTA e SILVA, 2013).

Indicador	Normal		Insatisfatório	
	ONS	Dados	ONS	Dados
DISP	60,3%	95,0%	21,9%	2,5%
TF	38,9%	95,0%	37,3%	2,5%
TMR	95,8%	95,8%	3,4%	2,1%

## 2. JUSTIFICATIVA

Os Indicadores de Desempenho das famílias de equipamentos de uma usina de geração de energia elétrica são acompanhados pelo Operador Nacional do Sistema Elétrico e fiscalizados pela Agência Nacional de Energia Elétrica. Esse assunto é discutido há alguns anos; com pesquisas no Centro de Documentação da ANEEL (CEDOC), foram encontrados registros a partir de 1997 em que já se propõe modelos de cálculos de Indicadores de Desempenho.

Os limites para definir se o funcionamento das centrais de geração é satisfatório ou não, são deduzidos de forma empírica; portanto, há a necessidade de uma melhor representação da confiabilidade desses sistemas, que pode ser suprida através de uma análise estatística dos dados registrados nos anos anteriores.

Há três anos o IPEA (Instituto de Pesquisa Econômica Aplicada) criou a Chamada Pública PNPD Nº 84/2010 com o intuito de selecionar interessados para “Elaboração de Indicadores referentes aos Processos de Fiscalização do Parque Gerador de Energia Elétrica”. Essa iniciativa mostra que há o interesse, a nível econômico e de desenvolvimento do parque, de acompanhar melhor o seu funcionamento através do processo de fiscalização.

É de responsabilidade da ANEEL a definição desses valores limites (faixas normal, alerta e insatisfatória) através de um banco de dados consistente fornecido pelo ONS e com regulamentação específica definida pela ANEEL. (ONS, 2011, p. Submódulo 16.3) Portanto, o desenvolvimento do presente trabalho se justifica pelo interesse e pela necessidade de definir uma metodologia que justifique os valores limites dos Indicadores de Desempenho.

Os conjuntos, que servirão como ferramenta para auxiliar o regulador na definição do desempenho limite para as UGs, devem ser agrupados através de técnicas estatísticas exploratórias conhecidas como “*Cluster Analysis*” (TANURE, 2000). Essa metodologia se baseia na expectativa de que usinas semelhantes tenham desempenhos equivalentes, como sugerido na Nota Técnica nº0021/2011 (SRD, 2011) possibilitando-as, então, de serem divididas e analisadas em grupos. Os métodos que tratam da análise de agrupamentos permitem uma verificação global das variáveis definidas para cada elemento da população a ser estudada, evidenciando para este universo as características de similaridade dos seus elementos (TANURE, 2000).

### 3. OBJETIVO

#### 3.1. OBJETIVO GERAL

O objetivo geral, ao fim deste trabalho, é redefinir os limites dos Indicadores de Desempenho – Taxa de Falhas (TF), Taxa Média de Reparo (TMR) e Disponibilidade (DISP) – de unidades geradoras de energia elétrica de usinas, mais precisamente de usinas hidrelétricas do Tipo I.

#### 3.2. OBJETIVOS ESPECÍFICOS

Os objetivos específicos desta segunda parte do trabalho são:

- Obter gráficos para visualizar o comportamento das unidades geradoras de acordo com a variação de suas características a partir do novo método de cálculo proposto;
- Analisar a correlação entre os parâmetros entre si e entre parâmetros e indicadores através de gráficos e da Análise de Correlações.
- Redefinir as faixas de potência e idade a partir de uma metodologia matemática;
- Analisar os resultados dos gráficos e decidir se cabe ou não a redefinição dos limites dos indicadores de desempenho a partir da formação de *clusters*.

## 4. CASOS INTERNACIONAIS

### 4.1. PORTUGAL

Segundo Haddad *et al.*, (2000) o sistema elétrico português é supervisionado pela Entidade Regulamentadora do Sector Eléctrico – ERSE, que tem a função de verificar o cumprimento dos regulamentos da qualidade de serviço. Os parâmetros qualitativos da energia elétrica são estabelecidos contratualmente, cabendo à ERSE a fiscalização e a aplicação de penalidades às empresas que não cumprem o contrato de fornecimento de energia elétrica.

Nestes contratos são definidas as seguintes condições de venda:

- Potência disponível;
- Energia ativa;
- Serviços complementares, como produção de energia reativa, controle de tensão e de frequência;
- Serviços especiais, como controle por telemetria, partida por telemetria, compensação síncrona, reserva quente e fria e outros;
- Parâmetros dinâmicos como o tempo de partida, tempo de partida como compensador, tempo de transição gerador/compensador e vice versa, gradientes de subida e descida de carga, entre outros.

Além das cláusulas contratuais que estabelecem a qualidade dos serviços de geração, a ERSE possui os indicadores UNIPEDDE para avaliar o desempenho das centrais, mas que não são utilizados para penalização e fiscalização, diferentemente do caso do Brasil. São eles:

- Disponibilidade;
- Indisponibilidade não-programada;
- Número de disparos por 7000 horas (número de falhas para este período);
- Taxa de Sucesso nas Mudanças de Estado (específico para hidrelétricas);
- Taxa de sucesso de partidas (específico para termelétricas).

#### 4.2. ARGENTINA

De acordo com Haddad (2000), a agência reguladora do setor elétrico argentino, denominada Ente Nacional Regulador de la Electricidad – ENRE, não utiliza indicadores para avaliação da necessidade de fiscalização dos agentes, embora utilize três indicadores relativos a geração de energia elétrica: Disponibilidade, Taxa de falhas e TMRF, conforme as equações abaixo:

$$DISPONIBILIDADE = \frac{n^{\circ} \text{ de horas disponíveis no período}}{n^{\circ} \text{ de horas do período}} \quad (1)$$

$$FALHA = \frac{n^{\circ} \text{ de falhas registradas no período}}{n^{\circ} \text{ de equipamentos registrado no período}} \quad (2)$$

$$TMRF = \frac{\text{horas que um equipamento está disponível}}{n^{\circ} \text{ de falhas do período}} \quad (3)$$

Com exceção do indicador Taxa de falhas, que utiliza em seu denominador o número de equipamentos registrados no período, os demais indicadores são similares aos utilizados no Brasil.

#### 4.3. COLÔMBIA

Também na Colômbia, são utilizados indicadores para a fiscalização da qualidade da energia elétrica gerada, contudo esses indicadores não são utilizados como gatilhos para a fiscalização in loco das unidades geradoras.

O Centro Nacional de Despacho utiliza, para a supervisão do sistema elétrico colombiano, os indicadores Indisponibilidade de Curto Prazo das Máquinas Geradoras – ICP e Indisponibilidade Histórica de Unidades Geradoras – IH.

O ICP é a indisponibilidade de cada unidade geradora ocasionada por eventos diferentes das manutenções programadas nos últimos três anos. É calculado através da Equação 4.

$$ICP = 1 - \frac{(1 - IH)}{(1 - IMP)} \quad (4)$$

onde:

- IH é a indisponibilidade para cada máquina geradora ocasionada por limitações de sua capacidade efetiva e desconexões programadas ou não programadas, medida nos últimos três anos. O IH é calculado como sendo a diferença entre a capacidade efetiva da máquina geradora e a capacidade disponível horária da mesma, nos últimos três anos, e,
- IMP e a indisponibilidade da capacidade efetiva de cada máquina geradora atribuída às manutenções programadas nos últimos três anos.

Entretanto, para a análise de potência, os cálculos são efetuados sobre as horas onde a demanda é máxima, e, para análise da energia, sobre todas as horas do período, sendo que, na Colômbia, o período base adotado para as medições é três anos.

#### 4.4. CONCLUSÃO

É possível perceber, analisando o caso de Portugal e da Argentina, que há países que não tem o mesmo intuito do Brasil, mais especificamente da ANEEL, de utilizar os resultados dos Indicadores de Qualidade de Energia Elétrica como instrumento para embasar uma fiscalização, bem como embasar uma penalidade para os agentes produtores de energia.

Percebe-se também a falta de padronização do período de tempo considerado, comparando os países citados entre si e com o Brasil. Isso implica na necessidade dos países que comercializam energia elétrica de estabelecerem indicadores comuns, no intuito de homogeneizar a interpretação de indicadores e de qualidade de fornecimento de energia.

Segundo Haddad (2000), em alguns países não há fiscalização direta dos agentes, mas sim do não cumprimento das cláusulas contratuais de fornecimento de energia elétrica. Neste caso, o contrato é o instrumento que permite penalizar.



## 5. METODOLOGIA

Inicialmente, na primeira parte deste trabalho de conclusão de curso (TCC 1) foi realizada uma revisão bibliográfica sobre o uso de Indicadores de Desempenho no Sistema Elétrico Brasileiro e também em outros contextos, com o objetivo de conhecer a metodologia de tratamento dos dados. Foram encontradas várias aplicações de análises estatísticas para criação de valores limites, aos quais se deve observar uma provável situação de risco em contextos diferentes: na administração de empresas, na análise de qualidade de produtos, na manutenção industrial e no mercado financeiro de ações.

Um trabalho, em especial, guiou a parte inicial da elaboração deste trabalho de conclusão de curso. Trata-se do Relatório elaborado por um grupo de pesquisa do Departamento de Engenharia Elétrica da Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), coordenado pelo professor Cristiano Fernandes. Esse trabalho, intitulado “Elaboração de Indicadores referentes aos processos de fiscalização do parque de geração elétrico brasileiro”, é resultado da pesquisa associada à Chamada Pública PNPD Nº 84/2010 para a especificação e teste de uma nova metodologia desenvolvida com o objetivo de estabelecer os valores de corte de Indicadores de Desempenho utilizados pela ANEEL e pelo ONS, para avaliar equipamentos e linhas de transmissão pertencentes às instalações dos agentes geradores e transmissores de energia elétrica do Sistema Interligado Nacional (SIN).

Durante todas as etapas deste trabalho, o orientador e o co-orientador, ambos da Universidade de Brasília (UnB), e o supervisor da ANEEL acompanharam os procedimentos através de reuniões semanais e troca de informação via e-mail. Primeiramente houve a delimitação do objeto de estudo: ficou definido que seriam analisadas, prioritariamente, as unidades geradoras de usinas hidrelétricas do Tipo I. Em seguida, definiu-se dentre os Indicadores de Desempenho existentes, quais teriam seus limites redefinidos. Foram então selecionados os Indicadores de Disponibilidade, Taxa de falhas e Tempo Médio de Reparo. Também foi feito levantamento das características das usinas e das unidades geradoras, que serão:

- a) Potência: a potência das UGs varia entre 10MW e 527MW.
- b) Tempo em operação: as UGs analisadas possuem idades diversas, variando a entrada em operação desde 1924 até 2012.

- c) Região: as usinas e suas respectivas UGs foram agrupadas de acordo com a região do seu Centro de Operação, sendo estes: Sul (S), Sudeste (SE), Nordeste (NE) e Norte-Centro Oeste (NCO), de acordo com a classificação do ONS (ONS, 2013).
- d) Quantidade de UGs na usina: a quantidade de unidades geradoras varia de 2 UGs a 25 UGs por usina.
- e) Tecnologia hidráulica: entende-se por tecnologia hidráulica os tipos de turbina utilizados nas unidades geradoras, para as analisadas são utilizadas as do tipo Francis, Kaplan, Bulbo e Pelton.

O ONS conta com Centros de Operação, que coordenam, supervisionam e controlam toda a rede de operação do SIN. Os Centros de Operação utilizados no estudo foram: o Centro Regional de Operação Norte/Centro-Oeste (COSR-NCO), o Centro Regional de Operação Nordeste (COSR-NE), o Centro Regional de Operação Sudeste (COSR-SE), e o Centro Regional de Operação Sul (COSR-S). Há também o Centro Nacional de Operação do Sistema (CNOS), em Brasília (DF), que toma as decisões necessárias para que o suprimento energético em todo o país se dê de forma contínua, com segurança, qualidade e economicidade (ONS, 2013).

Os empreendimentos hidrelétricos estão submetidos a situações que intuitivamente poderiam ser um parâmetro de análise e de influência de um indicador, mas ao se observar num âmbito mais abrangente, não o são. Um desses parâmetros é a sazonalidade. A usina está submetida a ciclos da natureza como períodos chuvosos e de seca. Esses períodos influenciam nos níveis de reservatório e na operação das unidades geradoras, visto que dependendo da vazão de água, elas podem operar ou não.

Este trabalho usa dados que abrangem períodos anuais, nos quais a sazonalidade completa um ciclo, então é como se o efeito, de um ano para outro e ao final dos 5 anos que compreendem os dados do trabalho, se anulasse. Porém a sazonalidade não é irrelevante. Esse aspecto, entre um ano e outro, é utilizado para a comercialização de energia elétrica, no tocante à garantia física da usina hidrelétrica, quando, de 5 em 5 anos, os indicadores de manutenção definidos no projeto da usina, o TEIF e o TEIP, são ajustados.

Após definição dos parâmetros utilizados para agrupar as unidades geradoras de todas as usinas hidrelétricas do Tipo I que constam na superintendência, foi necessário colher e organizar os dados relativos a cada uma. A partir do material

disponível na SFG e da planilha do RAD fornecida pelo ONS com os dados de DISP, TMR e TF, essas informações foram filtradas e agrupadas para utilização no programa *software* Scilab.

O *software* Scilab é um *software* livre e de código aberto para computação numérica, proporcionando um ambiente de computação poderoso para aplicação na Engenharia. O Scilab tem acesso aos arquivos de MATLAB e Excel e permite a obtenção de gráficos em boa resolução de forma simples. (GADELHA) No Scilab também é possível fazer análises estatísticas através da programação de algoritmos e, por isso, é uma das principais ferramentas deste trabalho.

## 5.1. CÁLCULO DOS INDICADORES DE DESEMPENHO

Os Indicadores de Desempenho são calculados a partir das equações estabelecidas no Submódulo 25.8 dos Procedimentos de Rede do ONS. As equações para os Indicadores tratados neste trabalho são apresentadas nas equações 1 a 3. As equações apresentam um somatório de 1 a N, sendo N a quantidade de UGs da usina, sendo assim, para calcular a disponibilidade, taxa de falhas e tempo médio de reparo de cada unidade geradora faz-se  $N=1$ . Para o cálculo de Indicadores em bases anuais, adotam-se 8.760 (oito mil e setecentos e sessenta) horas para anos normais e 8.784 (oito mil e setecentos e oitenta e quatro) horas para anos bissextos. Os Indicadores do RAD são apresentados em base anual, desta forma, para exemplificar, o valor do mês de dezembro é do período de um ano, formado por todos os valores de janeiro a dezembro.

### 5.1.1. Disponibilidade

A Disponibilidade é dada pelo somatório da multiplicação entre a potência de cada uma das unidades geradoras e o número de horas que ela está disponível dividido pelo somatório da mesma potência, sendo cada uma multiplicada pelo respectivo número total de horas de existência da unidade geradora no período de estudo, sendo dada em porcentagem. A expressão matemática que exprime o valor de Disponibilidade é dada pela Equação 5.

$$DISP = \frac{\sum_{i=1}^N (P_i \times HD_i)}{\sum_{i=1}^N (P_i \times HP_i)} \times 100 (\%) \quad (5)$$

Onde:

$HD_i$  = número de horas disponíveis da UGi;

$HP_i$  = número total de horas de existência da UGi no período considerado;

$P_i$  = potência efetiva da UGi homologada pela ANEEL, expressa em MW;

$N$  = número total de UGs; e

$i$  = contador do número de UGs.

### 5.1.2. Taxa de Falhas

A Taxa de Falhas considera o número de falhas de cada unidade geradora em relação ao número de horas de serviço da turbina utilizada. O valor encontrado é multiplicado pelo número de horas do ano. O cálculo desse indicador é feito através da expressão pela Equação 6.

$$TF = \frac{\sum_{i=1}^N NF_i}{\sum_{i=1}^N HX_i} \times 8.760 \quad (6)$$

Onde:

$NF_i$  = número de falhas da UGi;

$HX_i$  = número de horas da UGi conforme abaixo. Utiliza-se  $HX_i = HSi$  para número de horas de serviço do equipamento rotativo  $i$  e  $HX_i = HDi$  para número de horas disponíveis do equipamento não rotativo  $i$ ;

$i$  = contador do número de UGs;

$N$  = número total de UGs; e

8.760 (ou 8.784 para anos bisextos) = fator de anualização.

### 5.1.3. Tempo Médio de Reparo

O Tempo Médio de Reparo é o indicador que fornece a soma do número de horas em que cada unidade geradora esteve indisponível por motivo de manutenção forçada, em relação à soma de desligamentos forçados de cada unidade geradora, como na Equação 7.

$$TMR = \frac{\sum_{i=1}^N HIR_i}{\sum_{i=1}^N NDF_i} \quad (7)$$

Onde:

HIR<sub>i</sub> = número de horas em que a UG<sub>i</sub> ficou indisponível para operação e entregue à manutenção forçada;

NDF<sub>i</sub> = número de desligamentos forçados da UG<sub>i</sub>;

i = contador do número de UGs; e

N = número total de UGs.

A manutenção e o desligamento de uma unidade geradora são classificados como programado ou forçado. Uma manutenção programada é executada para cumprir um programa de manutenção preditiva, preventiva ou corretiva, ou ainda, para atender a manutenção de urgência. A manutenção pode ser considerada forçada quando é decorrente de desligamento forçado, executado para restabelecer sua condição satisfatória de operação.

O desligamento forçado de um componente de serviço acontece em condições não programadas, geralmente resultantes da ocorrência de uma condição de emergência que imponha o desligamento manual ou automático do equipamento para evitar risco de morte e/ou de dano a equipamento ou outras consequências indesejadas para o sistema elétrico. O ato de retirar de serviço o equipamento é o que diferencia um desligamento forçado de uma falha. A falha é o efeito ou consequência de ocorrência no equipamento que acarrete sua indisponibilidade operativa em condições não programadas e que, por isso, impede o equipamento ou de desempenhar suas funções em caráter permanente ou temporário. (ONS, 2011, p. Submódulo 20.1)

O caso de falha não inclui o desligamento forçado em certos casos, como, por exemplo, quando é percebido antecipadamente que o funcionamento do equipamento está fora de suas condições (temperaturas elevadas, ruídos incomuns) e este é levado para manutenção.

## 5.2. O RELATÓRIO DE AVALIAÇÃO DO DESEMPENHO DE MANUTENÇÃO

Os Indicadores do Relatório de Avaliação do Desempenho de Manutenção – RAD são disponibilizados em planilha de Excel, apresentada na Figura 2. Essa

planilha possui os valores de DISP, Indisponibilidade por Manutenção Forçada (INDISPMF), Indisponibilidade por Manutenção Programada (INDISPMP), TF, Taxa de Desligamento Forçado (TDF) e TMR.

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1															
2		USI 14 DE JULHO	RS												
3															
4				jan/2012	fev/2012	mar/2012	abr/2012	mai/2012	jun/2012	jul/2012	ago/2012	set/2012	out/2012	nov/2012	dez/2012
5	CERAN	UG 50 MW 14 DE JULHO UG1 RS		74,15925	74,39948	74,65923	74,79167	74,79167	74,79167	80,4178	88,86308	97,00592	99,36475	99,34711	99,36796
6	CERAN	UG 50 MW 14 DE JULHO UG2 RS		79,13813	81,61316	83,76859	91,96532	93,4781	93,49044	94,57669	95,53127	95,46714	95,7434	95,7267	95,85819
7		INSTALAÇÃO		76,64869	78,00632	79,21391	83,378495	84,134885	84,141055	87,497245	92,197175	96,23653	97,554075	97,536905	97,613088
8															
9		USI AGUA VERMELHA	SP												
10															
11				jan/2012	fev/2012	mar/2012	abr/2012	mai/2012	jun/2012	jul/2012	ago/2012	set/2012	out/2012	nov/2012	dez/2012
12	AES TIETÉ	UG 232 MW AGUA VERMELHA UG1 SP		97,24429	97,33796	97,33265	97,36699	97,23171	97,34176	97,31899	97,31899	94,87439	96,06576	95,97526	95,86274
13	AES TIETÉ	UG 232 MW AGUA VERMELHA UG2 SP		96,8769	96,98619	97,36775	97,36775	97,29432	97,40361	97,38084	97,38084	96,18454	96,95222	96,86248	96,86248
14	AES TIETÉ	UG 232 MW AGUA VERMELHA UG3 SP		87,96252	87,99939	87,94968	87,94968	87,88745	87,88745	87,85538	88,02709	96,1815	99,09874	99,64329	99,64329
15	AES TIETÉ	UG 232 MW AGUA VERMELHA UG4 SP		91,18779	90,8155	90,81777	90,76294	90,71797	93,35857	96,64143	96,70537	98,39614	98,40771	98,95226	98,95226
16	AES TIETÉ	UG 232 MW AGUA VERMELHA UG5 SP		98,29091	98,20962	97,99901	97,98345	97,08637	97,09168	98,00698	97,98232	97,97492	97,97492	98,54622	98,54622
17	AES TIETÉ	UG 232 MW AGUA VERMELHA UG6 SP		97,50419	97,511	97,42676	97,42676	96,53005	96,53537	97,01844	92,59753	92,5	92,46357	93,07149	93,21076
18		INSTALAÇÃO		94,8444333	94,8099433	94,8156033	94,809955	94,4579783	94,9364067	95,7036767	95,0020233	96,0185817	96,8271533	97,1751667	97,179625
19															
20		USI AIMORES	MG												
21															
22				jan/2012	fev/2012	mar/2012	abr/2012	mai/2012	jun/2012	jul/2012	ago/2012	set/2012	out/2012	nov/2012	dez/2012
23	CEMIG	UG 110 MW AIMORES UG1 MG		93,96518	94,50877	94,71387	93,75873	95,23509	95,12409	95,12542	96,60235	95,40984	95,40984	96,0576	96,80802

Figura 2. Planilha de Indicadores de Desempenho do RAD fornecida pelo ONS.

### 5.2.1. Tratamento dos dados

Uma vez que o ONS incumbiu a ANEEL de apresentar os valores limites para DISP, TF e TMR e os objetos de trabalho deste estudo são UGs de usinas hidrelétricas, esses valores foram filtrados e agrupados em outra planilha Excel. Foi necessária a confecção de nova planilha não apenas para selecionar e adicionar os dados de interesse, mas também para organizá-los de forma a adequar à leitura do programa. Para a aplicação no *software* Scilab, a planilha confeccionada não deveria conter quaisquer descrições ou dados que não os valores a ser plotados, assim a planilha foi simplificada para constar apenas os valores, sem células em branco, um exemplo é apresentado na Figura 3.

Para a execução dos programas, foi utilizada apenas a planilha apresentada na Figura 3, que contém o índice de 1 a 549, que representam cada unidade geradora na ordem do RAD; a potência de cada unidade geradora; a região sendo NCO (Norte-Centro Oeste) representada pelo número 1, NE (Nordeste) pelo número 2, S (Sul) pelo número 3 e o SE (Sudeste) pelo número 4; o tipo de turbina Francis é representado pelo número de 1, Kaplan pelo número 2, Pelton pelo número 3 e Kaplan pelo número 4; a idade, representada pelo ano de entrada em operação da

Unidade Geradora; o indicador Disponibilidade; o indicador Taxa de Falhas e o indicador Tempo Médio de Reparo.

	A	B	C	D	E	F	G	H
1	1	50	3	2	2008	85,04704	18,13625	7,33989
2	2	50	3	2	2008	80,60943	7,417253	3,382598
3	3	232	4	1	1978	96,7617	1,384523	13,14231
4	4	232	4	1	1978	91,00099	2,42374	13,99505
5	5	232	4	1	1979	95,79079	1,27184	5,931482
6	6	232	4	1	1979	94,78721	0,986517	1,546458
7	7	232	4	1	1979	94,51097	1,007208	4,60185
8	8	232	4	1	1979	94,39729	1,53759	75,94232
9	9	110	4	2	2005	92,38587	6,075707	4,614038
10	10	110	4	2	2005	94,60305	6,097975	10,37625
11	11	110	4	2	2005	93,49476	5,568065	26,22982
12	12	37	3	1	2009	97,31268	6,77991	4,700927
13	13	37	3	1	2009	97,05832	7,342638	2,66389

Figura 3. Planilha confeccionada para utilização no Scilab: com apenas dados de interesse.

### 5.2.2. Aplicação de filtros

Foi verificada a necessidade de aplicar filtros que retirassem da análise dados que não representam informações relevantes para a abordagem que está sendo feita.

#### 5.2.2.1. Filtro para DISP = 0, TF = 8760 e TMR = 8760

Quando DISP no período é igual a zero e for verificado que TMR e TF são diferentes de zero, significa dizer que a intervenção que retirou o equipamento de operação é para realização de manutenção forçada. Ficou estabelecida pelo ONS a seguinte convenção:

“Se na equação de TF o valor do numerador for igual a zero ou diferente de zero e o valor do denominador igual a zero (zero sobre zero ou número sobre zero) deveremos entender que (apesar de ser uma impossibilidade matemática) o resultado dessa razão será igual a 1, e assim o valor de TF nessa situação será 8760 ou 8784 (ano bissexto)” (FERNANDES, COSTA e SILVA, 2013).

Além das unidades geradoras descartadas da amostra através dos filtros (Quadro 2), também foi descartada da análise a UHE Itaipu e as suas unidades geradoras por ela possuir aspectos diferenciados, principalmente nos procedimentos de fiscalização, não sendo objeto de aplicação dos resultados deste estudo. Com a aplicação dos filtros, as unidades geradoras apresentadas no Quadro 2 foram retiradas.

Também foram excluídos os valores para a UG021 de Tucuruí, pois a Taxa de Falhas no ano de 2012 apresentou o valor de 876.000.000 ano<sup>-1</sup>, valor não possível para este indicador.

Quadro 2. Usinas e respectivas unidades geradoras excluídas da amostra em análise por filtragem de dados.

Ano	Usina	Unidade Geradora	Filtro
2008	SE251	UG12	DISP=0, TF=0 e TMR=0
2010	NE20	UG04	DISP=0, TF=0 e TMR=0
2009 2010	SE187	UG19	DISP=0, TF=0 e TMR=0
2010	SE301	UG03	DISP=0, TF=0 e TMR=0
2011	NE343	UG03	DISP=0, TF=0 e TMR=0
2011	NE346	UG05	DISP=0, TF=0 e TMR=0
2011	SE499	UG04	DISP=0, TF=0 e TMR=0
2013	NE343	UG02	DISP=0, TF=0 e TMR=0
2013	NE346	UG04, UG05 e UG08	DISP=0, TF=0 e TMR=0
2013	SE499	UG04	DISP=0, TF=0 e TMR=0
2008	SE410	UG01	DISP=0, TF= 8784 e TMR=8784
2009	NCO103	UG01, UG02 e UG03	DISP=0, TF=8760 e TMR=8760
2012	NE346	UG06, UG07 e UG08	DISP=0, TF=8760 e TMR=8760

#### 5.2.2.2. Filtro para DISP = 0, TF = 0 e TMR = 0

Este acontecimento pode ser entendido com a situação em que a unidade geradora se encontra fora de operação, durante todo o período analisado (um ano). A causa de esse equipamento estar fora de operação pode ser uma parada para manutenção programada ou para modernização do equipamento com tempo superior a um ano.



### 5.2.3. Normalização

O propósito da normalização é minimizar os problemas oriundos do uso de unidades e dispersões distintas entre as variáveis (PRATI). A normalização dos valores é relevante para que se possa comparar dados diferentes dando a eles o mesmo peso. Ela é calculada a partir da divisão do valor pelo desvio padrão da amostra para os valores de DISP, TF, TMR, Idade e Potência.

### 5.2.4. Medidas de Similaridade e Dissimilaridade

Para melhor entender as medidas de similaridade e dissimilaridade, inicia-se supondo que se tenha disponível um conjunto de dados constituído de  $n$  elementos amostrais, tendo-se medido  $p$ -variáveis aleatórias em cada um deles. O objetivo é agrupar esses elementos em  $g$  grupos. Para cada elemento amostral  $j$ , tem-se, portanto, o vetor de medidas  $X_j$  definido pela Equação 8.

$$X_j = [X_{1j} X_{2j} \dots X_{pj}]', j=1, 2, \dots, n. \quad (8)$$

onde  $X_{ij}$  representa o valor observado da variável  $i$  medida no elemento  $j$ .

Para que se possa proceder ao agrupamento de elementos, é necessário que se decida a priori a medidas de similaridade ou dissimilaridade que será utilizada. Existem várias medidas diferentes e cada uma delas produz um determinado tipo de agrupamento. Algumas medidas mais comuns, apropriadas para variáveis quantitativas, são apresentadas a seguir. Essas medidas são de dissimilaridade e, logo, quanto menor os seus valores, mais similares serão os elementos que estão sendo comparados (MINGOTI, 2013).

#### 5.2.4.1. Distância Euclidiana

A distância Euclidiana entre dois elementos  $X_l$  e  $X_k$ ,  $l \neq k$ , é definida pela Equação 9.

$$d(X_l, X_k) = [(X_l - X_k)'(X_l - X_k)]^{\frac{1}{2}} = [\sum_{i=1}^p (X_{il} - X_{ik})^2]^{\frac{1}{2}} \quad (9)$$

ou seja, os dois elementos amostrais são comparados em cada variável pertencente ao vetor de observações (MINGOTI, 2013).

#### 5.2.4.2. Distância generalizada ou ponderada

A distância generalizada entre dois elementos  $X_l$  e  $X_k$ ,  $l \neq k$ , é definida pela Equação 10.

$$d(X_l, X_k) = [(X_l - X_k)' A (X_l - X_k)]^{\frac{1}{2}} \quad (10)$$

onde  $A_{p \times p}$  é uma matriz de ponderação, positiva definida.

Quando a matriz  $A_{p \times p}$  é a matriz identidade, a distância generalizada é a distância Euclidiana; se  $A_{p \times p}$  é igual a  $S_{p \times p}^{-1}$ , tem-se a distância de Mahalanobis e quando  $A_{p \times p} = \text{diag}(\frac{1}{p})$ , tem-se a distância Euclidiana média. A escolha da matriz  $A_{p \times p}$  reflete o tipo de informação que o pesquisador deseja utilizar na ponderação das diferenças das coordenadas dos vetores que estão sendo comparados. Assim, uma matriz do tipo  $A_{p \times p} = \text{diag}((S_i^2)^{-1})$ , onde  $S_i^2$  é a variância amostral da  $i$ -ésima variável aleatória  $i = 1, 2, \dots, p$ , leva em consideração na ponderação apenas a diferença de variabilidade que pode haver entre as variáveis, uma vez que, neste caso, a distância entre  $X_l$  e  $X_k$ ,  $l \neq k$  pode ser reescrita como mostrado na Equação 11.

$$d(X_l, X_k) = \left[ \sum_{i=1}^p \frac{1}{S_i^2} (X_{il} - X_{ik})^2 \right]^{\frac{1}{2}} \quad (11)$$

Já uma matriz do tipo  $A_{p \times p} = S_{p \times p}^{-1}$  leva em consideração, na ponderação, as possíveis diferenças de variâncias e as relações lineares entre as variáveis, medidas em termos de covariâncias (MINGOTI, 2013).

#### 5.2.4.3. Distância Minkowski

Segundo Linden (2009) um dos tipos de medidas de similaridade existente é o Minkowski. A partir dela são calculadas as distâncias Euclidiana e Manhattan. Com ou sem a normalização e independentemente do método de cálculo das distâncias, a dissimilaridade (ou similaridade) entre os objetos descritos por variáveis escaladas

em intervalos são computadas baseado na distância entre cada par de objetos. Deve-se ressaltar que pares ordenados iguais foram plotados somente uma vez; visto que, se são iguais, a distância entre eles é zero.

A métrica mais geral é a Minkowski e dela que são definidas as métricas Euclidiana e Manhattan. Ela é dada pela Equação 12.

$$d(i, j) = \sqrt[q]{(|x_{i1} - x_{j1}|^q + |x_{i2} - x_{j2}|^q + \dots + |x_{ip} - x_{jp}|^q)} \quad (12)$$

#### 5.2.4.4. Distância Manhattan

A distância Manhattan é a soma dos módulos das diferenças entre todos os atributos de dois objetos em questão, quando  $q = 1$ , sendo dada pela Equação 13.

$$d(i, j) = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \dots + |x_{ip} - x_{jp}| \quad (13)$$

A Figura 4 apresenta graficamente uma comparação entre as três metodologias de cálculo de distância entre dois elementos. A linha fina representa a métrica euclidiana e é calculada com uma reta entre os dois pontos em análise. Já a distância Manhattan, linha cheia, é calculada por um cateto a cada vez, enquanto que a linha tracejada representa a distância de Chebyshev, que é dada pela maior das duas dimensões da distância.

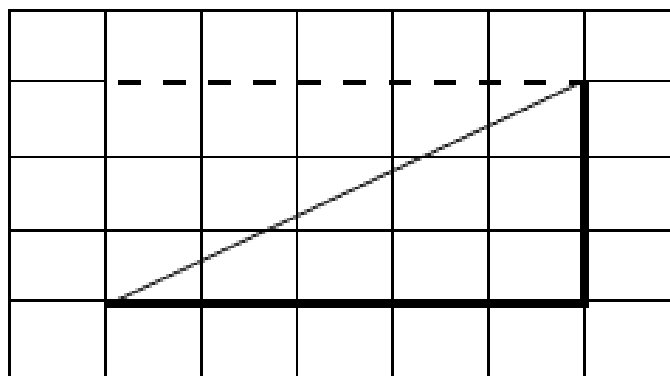


Figura 4. Distâncias Euclidiana e Manhattan.

Na Figura 5 encontra-se o mapa do bairro de Manhattan, na cidade de Nova Iorque. A composição das ruas, quase que somente paralelas e perpendiculares, justifica o nome da Métrica de Manhattan, que leva em consideração os catetos, representados no mapa pelas ruas. O programa elaborado nesta segunda parte do

trabalho de conclusão de curso (TCC 2) utiliza a distância Manhattan em vez da Euclidiana, pois sua expressão é mais simples e, por isso, facilita a programação.



Figura 5. Mapa de Manhattan. Fonte: <<http://www.nuevayork.net/barrios-manhattan>> acessado em 02 de junho de 2014.

### 5.2.5. A Função *gsort*

A função *gsort*, no Scilab, é utilizada para ordenar os valores da matriz distância em ordem crescente. Dessa forma, a leitura dos dados fica mais organizada e permite uma execução mais rápida. Como exemplo, segue o tempo de compilação cronometrado mostrado na Tabela 2 para gerar os gráficos de Disponibilidade x Idade, Taxa de falhas x Idade e Tempo médio de Reparo x Idade.

Tabela 2. Tabela com tempos de execução com e sem a função *gsort*.

	Tempo de execução com <i>gsort</i> (s)	Tempo de execução sem <i>gsort</i> (s)
DISP x Idade	18.564	19.611
TF x Idade	15.741	21.278
TMR x Idade	15.023	20.667

O cálculo do tempo foi feito a partir das funções *tic()* e *toc()* do Scilab. A primeira dispara um cronômetro enquanto a segunda lê o valor apontado por ele. A função *tic()* foi colocada no início do programa, antes de todos os comandos e a *toc()*, no fim, depois de todos os comandos. É possível perceber, a partir do exemplo dado, que a função *gsort*, ao ordenar os valores em ordem crescente, minimiza o tempo gasto para executar o programa.

#### 5.2.6. A Função *tabul*

A função *tabul* apresenta no console do Scilab uma matriz  $NCmin \times 2$ , a qual apresenta o número de algumas UGs (a primeira de cada *cluster*) e quantos pontos há no respectivo *cluster*. Essa função permite analisar a plotagem em mais de três dimensões, o que é muito útil, visto que o trabalho envolve até quatro parâmetros de plotagem que devem ser comparados ao mesmo tempo, além do indicador. A função *tabul* é apresentada no código, no fim do laço *for*, como mostrado Figura 6. O retorno que essa função oferece no console é mostrado na Figura 7.

O programa retorna a matriz  $10 \times 2$ , visto que foi escolhido  $NCmin = 10$ . Os valores da primeira coluna são os índices das UGs dispostos na coluna 1 da matriz T, enquanto ao lado estão a quantidade de UGs no respectivo *cluster*.

```

for co = 1:NT
    .....
    ..... conta = conta + 1;
    .....
    ..... i = DI(k(co), 1);
    ..... j = DI(k(co), 2);

    ..... if (C(i) ~= C(j))
    ..... plot2d([x(i) x(j)], [y(i) y(j)]);
    ..... xtitle('Disponibilidade x Idade', 'DISP', 'Idade');

    ..... M = M - 1;
    .....
    ..... Ci = C(i);
    ..... Cj = C(j);

    ..... for h = 1:N
    ..... if (C(h) == Cj)
    ..... C(h) = Ci;
    ..... end
    ..... end
    .....
    ..... if (M == NCmin) then
    ..... disp(tabul(C));
    ..... disp('Fim Clustering');
    ..... return;
    .....
    ..... end

```

Figura 6. Função *tabul* inserida no código do programa.

```

Início da Normalização

Fim da Normalização

Início Distâncias

Fim Distâncias

Início do Ordenamento

Fim do Ordenamento

Início Clustering

417. 504.
344. 2.
342. 2.
178. 1.
163. 18.
160. 1.
118. 1.
104. 7.
26. 1.
23. 1.

Fim Clustering

```

Figura 7. Saída da função *tabul* no console.

### 5.3. DEFINIÇÃO DAS FAIXAS DE POTÊNCIA E IDADE

Os valores de potência e idade (tempo desde o início da operação) das unidades geradoras são diversos e bastante específicos, por isso, faz-se necessária a definição de faixas de valores que classifiquem as unidades geradoras pela potência em: baixa, média ou alta, assim como por idade em: mais antigas, mais novas e intermediárias.

Buscou-se fazer essa classificação da forma mais natural possível, utilizando o método de clusterização para encontrar grupos com maior similaridade e localizar os *gaps*, espaços, maiores entre os valores. O método foi aplicado através do *software* Scilab, o programa, apresentado na Figura 8, normaliza os valores de potência e dos anos de entrada em operação das UGs, estes valores são apresentados em gráfico e as distâncias entre os pontos são calculadas apenas em uma dimensão, as distâncias são apresentadas por linhas que ligam os pontos. O gráfico gerado para a definição das faixas de idade é mostrado na Figura 9 e na Figura 10.

São observados nos gráficos *gaps*, espaços que distanciam valores de potência e idade, indicando que nesses espaços não possuem UGs que correspondem a estes valores. Assim, para a idade, as faixas podem ser definidas de modo a aproveitar os *gaps* visto que nenhuma UG pode ser construída com data de entrada em operação igual às encontradas nos espaços.

Porém, para a potência, faz-se necessário escolher um valor que divida as categorias de potência sem espaços, visto que futuramente podem ser construídas UGs com potências ainda não encontradas nesta base de dados.

```

Arquivo Editar Formatar Opções Janela Executar ?
faixas_idade.sce (C:\cluster\faixas_idade.sce) - SciNotes
faixas_idade.sce
1 //faixas_idade.sce
2 //Classificação Hierarquica em Clusters (1D) rápida
3 //Definindo Faixas de Idade.
4
5 clear;
6
7 stacksize('max'); //Para diminuir o tempo das contas
8
9 filename = ('C:\cluster\idade.csv');
10
11 T = csvRead(filename, ',', ',');
12
13 N = max(T(:,1)); //Número de Pontos;
14
15 M=N
16
17 C = 1:N; //Cada ponto é um cluster
18
19 x = T(:,2);
20 y = T(:,1);
21
22 disp('Início da Normalização');
23

```

Figura 8. Programa Scilab para definição de faixas de potência e idade das UGs.

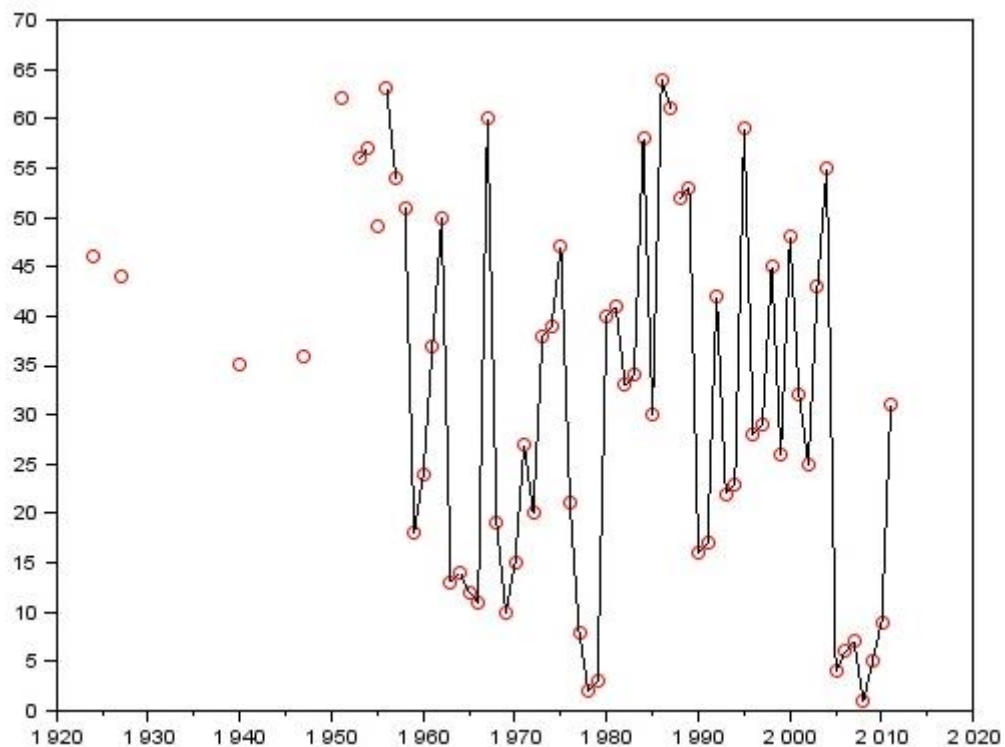


Figura 9. Gráfico de idades com cálculo da distância entre pontos.



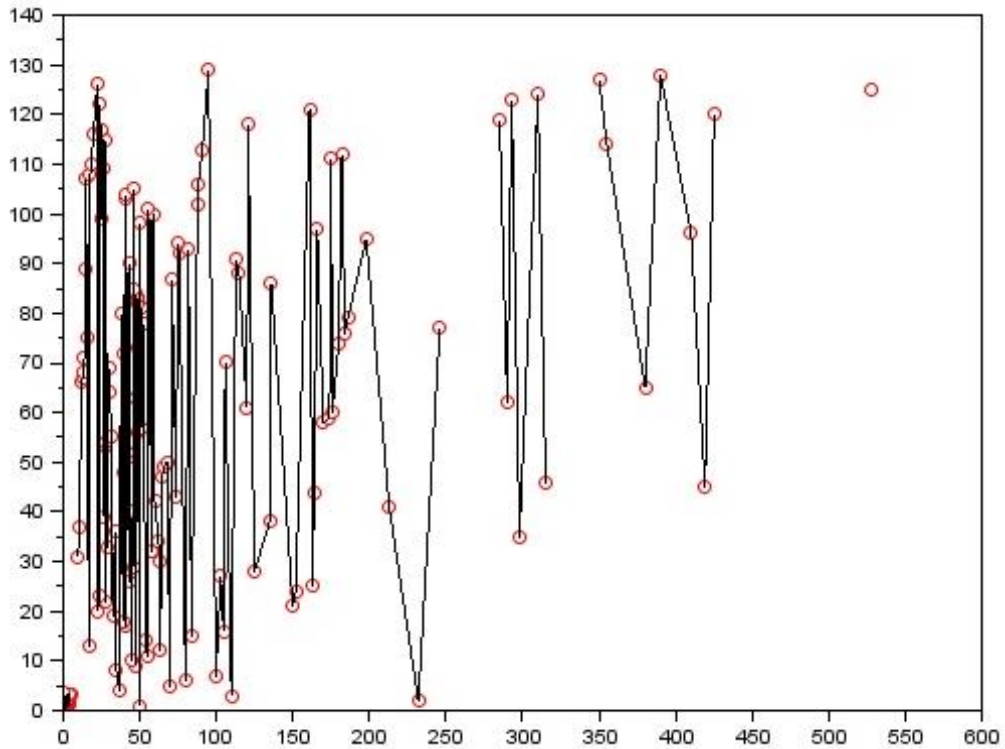


Figura 10. Gráfico de potências com cálculo das distâncias entre pontos.

Através da função *disp()* o programa apresenta o valor máximo e mínimo no *console* do *software* e, assim, foi possível concluir que os valores que definem as faixas da forma mais natural são os apresentados no Quadro 3.

Quadro 3. Faixas de Potência e de Ano de Entrada em Operação.

Faixas definidas	Potência	Faixas definidas	Ano de entrada em operação
Alta	Acima de 350 MW	Antigas	Até 1951
Média	Entre 250 MW e 350 MW	Menos antigas	Entre 1953 e 1987
Baixa	Abaixo de 250 MW	Novas	A partir de 1988

#### 5.4. CLUSTERIZAÇÃO

Para alcançar o objetivo do trabalho, definir os novos limites dos indicadores de qualidade, o método escolhido para o prosseguimento deste estudo é a clusterização. A intenção é conseguir organizar os dados em *clusters*, ou grupos,

que permitam encontrar um valor satisfatório para se redefinir o limite máximo da DISP e os mínimos do TMR e da TF.

Segundo Jain et al, clusterização é a classificação não-supervisionada de dados, que formam agrupamentos ou *clusters*. Ela representa uma das principais etapas de processos de análise de dados, denominada análise de *clusters*. A análise de *clusters* envolve, portanto, a organização de um conjunto de padrões (usualmente representados na forma de vetores de atributos ou pontos em um espaço multidimensional) em *clusters*, de acordo com alguma medida de similaridade (MOSCATO e VON ZUBEN). Os padrões pertencentes a um dado *cluster* devem guardar mais uma similaridade entre si do que em relação a padrões pertencentes a outros *clusters*.

De acordo com di Carlantonio, a clusterização permite determinar qual o número de grupos e os grupos existentes num conjunto de objetos. Com o conhecimento dos grupos existentes é possível analisar os elementos que compõem cada um deles. Também é possível identificar as características comuns aos seus elementos e, assim, criar uma classe para este grupo. Com as classes, pode-se classificar um novo objeto que pertença ao universo considerado.

“Então, o problema de *clustering* é descrito como: recebido um conjunto de dados, de objetos, tentar agrupá-los de forma que os elementos que compõem cada grupo sejam mais parecidos entre si do que parecidos com os elementos dos outros grupos. É colocar os iguais (ou quase) juntos num mesmo grupo e os desiguais em grupos distintos (DI CARLANTONIO, 2001).

A similaridade entre os dados de um mesmo agrupamento evidencia que a clusterização é uma ferramenta útil para o desenvolvimento deste trabalho. No início, os dados fornecidos não obedecem visivelmente a padrões relativos aos Indicadores, pois as usinas são independentes umas das outras, porém eles podem guardar semelhanças que os permitam classificar em faixas características, podendo ser analisados em conjunto, como por exemplo, uma faixa de valor de DISP para UGs que estão em uma região específica. Deve-se deixar claro que esse tipo de análise foi escolhido consciente da possibilidade, na pior das hipóteses, de que nenhuma família de dados forme *clusters*.

A análise feita com os dados fornecidos será do tipo exploratória (MOSCATO e VON ZUBEN), a qual permite a formulação de hipóteses e a tomada de decisão. A

formulação de hipóteses foi feita em cima da tentativa de unir duas ou mais características ao indicador, formando um *cluster*. A partir da formação de *clusters*, a decisão de um novo limite para estabelecer a qualidade de acordo com as características das UGs pôde ser tomada.

De acordo com Moscato & Von Zuben, as etapas de um processo de clusterização são:

- a) Representação dos padrões: envolve definição do número, do tipo e do modo de apresentação dos atributos que descrevem cada padrão. Nesse passo são definidos quais serão os dados utilizados para a formação de *clusters*, no caso, serão utilizados os Indicadores DISP, TF e TMR, que são dados por porcentagens, falhas/ano e ano, respectivamente. Para cada indicador de desempenho será feita uma análise por clusterização, identificando os possíveis conceitos inerentes a cada característica.
- b) Seleção de características: processo de identificação do subconjunto mais efetivo dos atributos disponíveis para descrever cada padrão. Para esse passo, foram escolhidas, com base nas possíveis consequências que poderiam ter sobre os Indicadores, as características: potência, região, número de unidades geradoras, idade e tipo de turbina.
- c) Extração de características: uso de uma ou mais transformações junto aos atributos de entrada de modo a salientar uma ou mais característica dentre aquelas que estão presentes nos dados. A formação de clusters pode acontecer com apenas uma característica ou com um conjunto dessas, isto é, pode-se utilizar a potência e a idade como fatores determinantes para a formação de *clusters*. No caso da utilização de mais características são utilizados mais dimensões no gráfico onde se formarão os *clusters*.
- d) Medida de similaridade: fornecida por uma função de distância definida entre pares de dados ou padrões. É possível incluir na medida de distância aspectos conceituais (qualitativos) ou então numéricos (quantitativos).
- e) Agrupamento: os grupos podem ser definidos como conjuntos *crisp* (um padrão pertence ou não pertence a um dado grupo) ou *fuzzy* (um padrão pode apresentar graus de pertinência aos grupos). O processo de agrupamento pode ser hierárquico, com um processo recursivo de junções ou separações de grupos, ou não hierárquico, com o emprego direto de técnicas de discriminação de *clusters*.

- f) Apresentação: deve permitir que um computador utilize o resultado de forma direta ou então deve ser orientada ao usuário, permitindo a visualização gráfica dos clusters e a compreensão de suas inter-relações, através da proposição de protótipos ou outras descrições compactas para os *clusters*.

#### 5.4.1. *K-Means*

Para encontrar a medida de similaridade, fazer o agrupamento e obter a apresentação dos *clusters*, o algoritmo *k-means*, que pode ser traduzido como k-médias, teve sua aplicação aos dados com o *software* Scilab como método de clusterização idealizada para dar continuidade ao TCC 1. Esse algoritmo faria o agrupamento de certo número de dados fornecidos em k grupos. Inicialmente são escolhidos k objetos e, a partir destes, são definidas as médias dos *clusters*. Para todos os outros objetos é feita a atribuição a um grupo, de acordo com a distância entre seu valor e o valor médio do grupo (FONSECA e BELTRAME), (DI CARLANTONIO, 2001).

Basicamente, com o algoritmo *k-means*, escolhe-se arbitrariamente k objetos da base de dados como os centros iniciais dos *clusters* e através de iterações são calculadas novas médias para cada *cluster* à medida que os outros objetos vão sendo atribuídos a cada *cluster* até que não haja mudança de objetos de um *cluster* para outro. (DI CARLANTONIO, 2001)

O método *k-means* foi pensado inicialmente como método ideal para alcançar os objetivos do TCC, porém, através da continuidade das pesquisas de metodologias passíveis de utilização no trabalho, decidiu-se por trocar o algoritmo pelo método Hierárquico. O item 5.4.3 explicará as vantagens e desvantagens de um método em relação ao outro.

#### 5.4.2. Método Hierárquico

As técnicas de clusterização são frequentemente classificadas em dois tipos: técnicas hierárquicas e não hierárquicas, sendo que as hierárquicas são classificadas em aglomerativas e divisivas. Ambas são utilizadas em análises exploratórias dos dados com o intuito de identificar possíveis agrupamentos e o valor

provado número de grupos. O presente trabalho utiliza-se da técnica hierárquica aglomerativa.

#### 5.4.2.1. Técnicas Hierárquicas Aglomerativas

As técnicas hierárquicas aglomerativas partem do princípio de que no início do processo de agrupamento tem-se  $n$  conglomerados, ou seja, cada elemento do conjunto de dados observado é considerado como sendo um conglomerado isolado. Em cada passo do algoritmo, os elementos amostrais vão sendo agrupados, formando novos conglomerados até o momento no qual todos os elementos considerados estão num único grupo. Portanto, no estágio inicial do processo de agrupamento, cada elemento amostral é considerado como um *cluster* de tamanho 1 e no último estágio de agrupamento tem-se apenas um único *cluster* constituído de todos os elementos amostrais.

Os passos principais para aplicação das técnicas hierárquicas aglomerativas podem ser resumidos da seguinte forma:

- a) Cada elemento constitui um *cluster* de tamanho 1. Portanto, tem-se  $n$  clusters.
- b) Em cada estágio do algoritmo de agrupamento, os pares de conglomerados mais “similares” são combinados e passam a constituir um único conglomerado. Apenas um novo conglomerado pode ser formado em cada passo. Dessa forma, em cada estágio do processo, o número de conglomerados vai sendo diminuído;
- c) Propriedade de hierarquia: Em cada estágio do algoritmo, cada novo conglomerado formado é um agrupamento de conglomerados formados nos estágios anteriores. Se dois elementos amostrais aparecem juntos num mesmo *cluster* em algum estágio do processo de agrupamento, eles aparecerão juntos em todos os estágios subsequentes, ou seja, uma vez unidos estes elementos não poderão ser separados;
- d) Devido à propriedade de hierarquia, é possível construir um gráfico chamado de Dendrograma ou Dendograma (Timm, 2002) que representa a árvore ou a história de agrupamento. O Dendrograma é um gráfico em forma de “árvore” no qual a escala vertical indica o nível de similaridade (ou dissimilaridade). No eixo horizontal, são marcados os elementos amostrais numa ordem

conveniente relacionada à história de agrupamento. As linhas verticais, partindo dos elementos amostrais agrupados, tem altura correspondente ao nível em que os elementos foram considerados semelhantes, isto é, a distância do agrupamento ou o nível de similaridade. Um exemplo de dendrograma é apresentado na Figura 11, este foi obtido a partir da análise de clusters para o indicador de disponibilidade que será apresentado ao longo deste trabalho.

A escolha do número final de grupos em que o conjunto de dados deve ser repartido é subjetiva, e isso se aplica ao trabalho realizado no TCC 2. Também se aplica à possibilidade de encontrar o número de grupos que esteja associado à “partição natural” dos elementos que estão sendo comparados e agrupados.

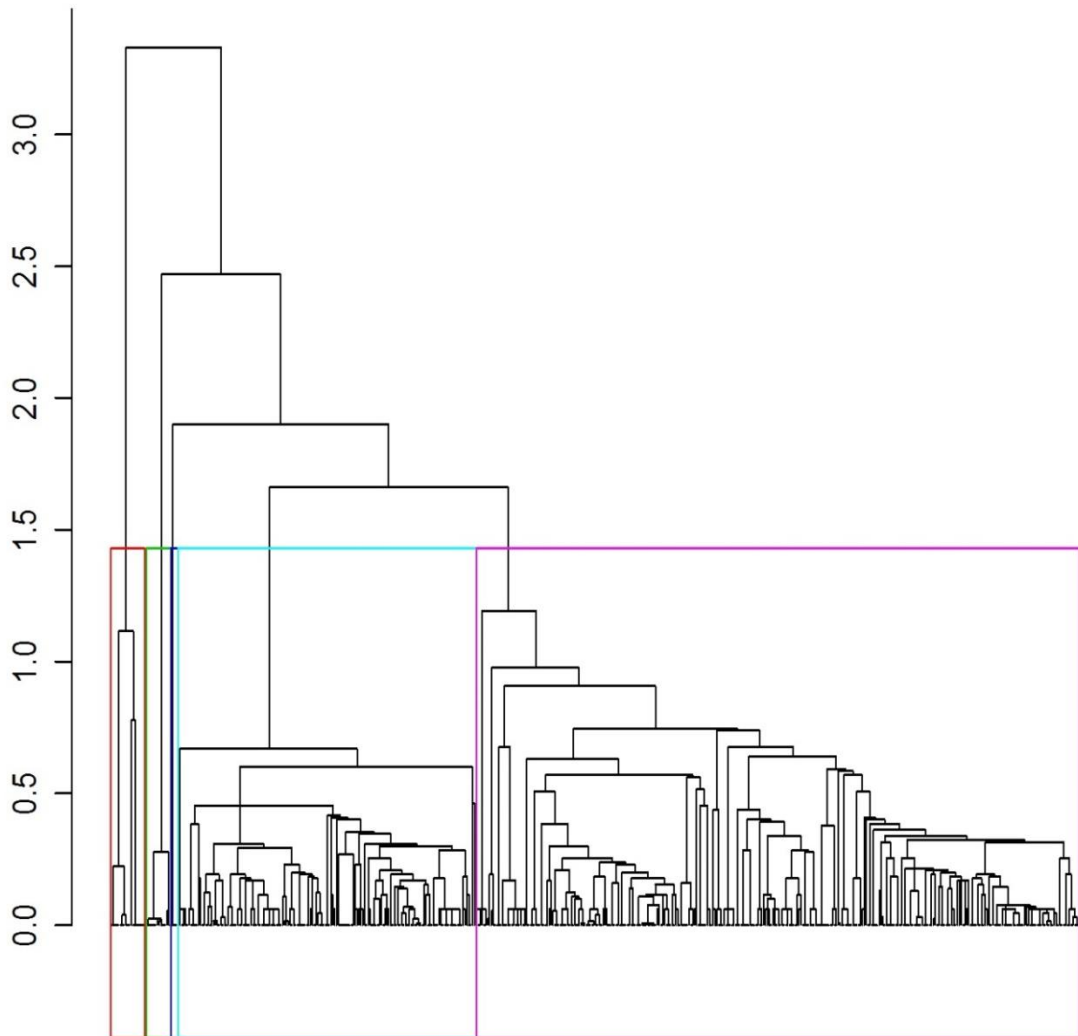


Figura 11. Exemplo de Dendrograma.

Dentre as técnicas hierárquicas aglomerativas, as mais utilizadas são o método de ligação simples, o método de ligação completa, o método das médias das distâncias e o método dos centróides. Para a realização deste trabalho foi escolhido o método de ligação simples, porém os quatro serão explicados abaixo.

#### 5.4.2.1.1. Método de Ligação Simples (Single Linkage)

Neste método, a similaridade entre dois clusters é definida pelos dois elementos mais parecidos entre si (SNEATH, 1957). Como ilustração, supõem-se que num determinado estágio do algoritmo de agrupamento se tenham dois grupos, um contendo elementos amostrais 1, 3, e 7 e outro contendo os elementos 2 e 6, isto é,  $C_1 = \{X_1, X_3, X_7\}$  e  $C_2 = \{X_2, X_6\}$ .

Então a distância entre esses dois grupos será definida pela Equação 14.

$$d(C_1, C_2) = \min\{d(X_l, X_k), l \neq k, l = 1, 3, 7 \text{ e } k = 2, 6\} \quad (14)$$

ou seja, é a distância entre os “vizinhos” mais próximos ou entre os elementos mais parecidos de cada conglomerado.

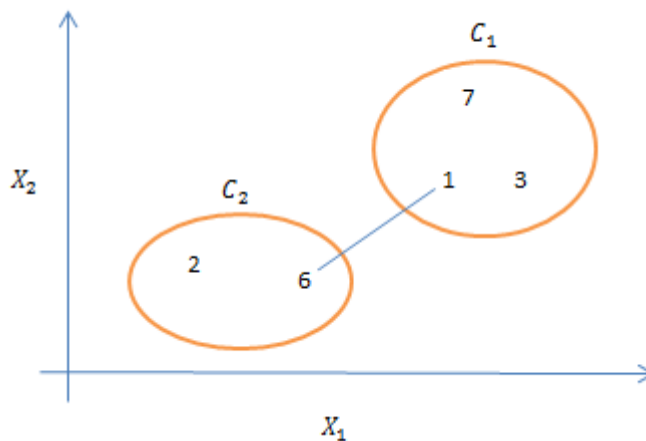


Figura 12. Ilustração gráfica do método de ligação simples.

Em cada estágio do processo de agrupamento, os dois conglomerados que são mais similares com relação à distância em Equação 14 são combinados em um único *cluster*. A Figura 12 mostra uma ilustração gráfica hipotética da comparação entre os conglomerados  $C_1$  e  $C_2$  no caso em que o número de *cluster* é 2 e a distância entre os elementos 6 e 1 é a menor entre os elementos de cada *cluster*,

caracterizando a distância entre os grupos no método de ligação simples (MINGOTI, 2013).

#### 5.4.2.1.2. Método de Ligação Completa (Complete Linkage)

Neste método, a similaridade entre dois conglomerados é definida pelos elementos que são menos semelhantes entre si (SNEATH, 1957). Por exemplo, considera-se os conglomerados:  $C_1 = \{X_1, X_3, X_7\}$  e  $C_2 = \{X_2, X_6\}$ .

Então a distância entre esses dois grupos será definida pela Equação 15.

$$d(C_1, C_2) = \max\{d(X_l, X_k), l \neq k, l = 1, 3, 7 \text{ e } k = 2, 6\} \quad (15)$$

Em cada estágio do processo de agrupamento, a medida na Equação 15 é calculada para todos os pares de grupos, sendo então, combinados num único aqueles que apresentarem o menor valor de máximo. A Figura 13 ilustra graficamente a comparação hipotética dos conglomerados  $C_1$  e  $C_2$  sendo a distância entre os elementos 2 e 7 aquela que caracteriza a distância entre os grupos no método de ligação completa.

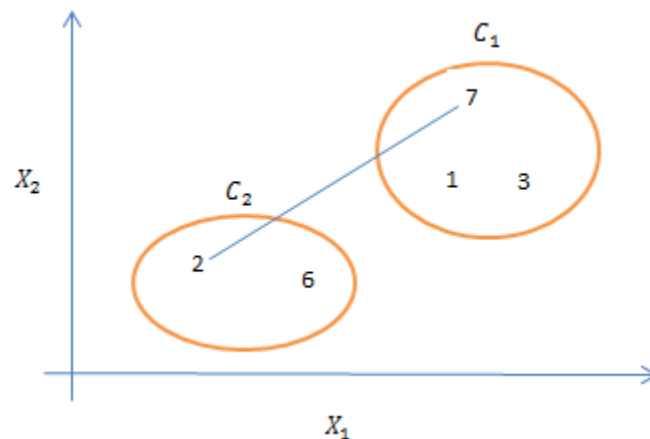


Figura 13. Ilustração gráfica do método de ligação completa.

#### 5.4.2.1.3. Método da média das distâncias (Average Linkage)

Segundo Mingoti (2013), esse método trata a distância entre dois conglomerados como a média das distâncias entre todos os pares de elementos que podem ser formados com os elementos dos dois conglomerados que estão sendo



comparados. Portanto, se o conglomerado  $C_1$  tem  $n_1$  elementos e o conglomerado  $C_2$  tem  $n_2$  elementos, a distância entre eles será definida pela Equação 16.

$$d(C_1, C_2) = \sum_{l \in C_1} \sum_{k \in C_2} \left( \frac{1}{n_1 n_2} \right) d(X_l, X_k) \quad (16)$$

Assim, a distância entre os conglomerados:  $C_1 = \{X_1, X_3, X_7\}$  e  $C_2 = \{X_2, X_6\}$  é igual à Equação 17.

$$d(C_1, C_2) = \frac{1}{6} [d(X_1, X_2) + d(X_1, X_6) + d(X_3, X_2) + d(X_3, X_6) + d(X_7, X_2) + d(X_7, X_6)] \quad (17)$$

#### 5.4.2.1.4. Método do centróide (Centroid Method)

Neste método, a distância entre dois grupos é definida como sendo a distância entre os vetores de médias, também chamados de centróides, dos grupos que estão sendo comparados. Assim, se  $C_1 = \{X_1, X_3, X_7\}$  e  $C_2 = \{X_2, X_6\}$ , os vetores de médias de  $C_1$  e  $C_2$  correspondentes são dados pelas Equações 18 e 19, respectivamente.

$$C_1 = \bar{X}_1 = \frac{1}{3} [X_1 + X_3 + X_7] \quad (18)$$

$$C_2 = \bar{X}_2 = \frac{1}{2} [X_2 + X_6] \quad (19)$$

e a distância entre  $C_1$  e  $C_2$  é definida pela Equação 20, que é a distância Euclidiana ao quadrado entre os vetores de médias amostral  $\bar{X}_1$  e  $\bar{X}_2$ .

$$d(C_1, C_2) = (\bar{X}_1 - \bar{X}_2)' (\bar{X}_1 - \bar{X}_2) \quad (20)$$

O método do centróide também pode ser usado com a distância Euclidiana usual entre os vetores de médias. Em cada passo do algoritmo do agrupamento, os conglomerados que apresentam o menor valor de distância são agrupados (MINGOTI, 2013).

### 5.4.3. Vantagens e Desvantagens: *K-Means* ou Método Hierárquico?

A principal vantagem do método hierárquico é a facilidade em lidar com qualquer medida de similaridade utilizada e a sua consequente aplicabilidade a qualquer tipo de atributo (numérico ou categórico), enquanto a principal vantagem, do ponto de vista deste trabalho, do método *k-means* é que nele, diferentemente dos métodos hierárquicos, os *clusters* podem ser melhorados gradativamente através de iterações.

Dentre as desvantagens do método hierárquico está o fato de apresentar imprecisão do critério de parada, visto que, a maioria dos algoritmos desta classe não faz iterações que revisitam os clusters formados ao longo de suas execuções. O método *k-means* apresenta como principal desvantagem o fato de a utilização do centróide como elemento representativo de um *cluster* ser conveniente para atributos numéricos. Para atributos geométrico e estatístico, pode receber muita influência de um elemento que se encontre próximo à fronteira do *cluster*.

Ponderando as vantagens e desvantagens entre métodos hierárquico e *k-means*, optou-se pela troca de método e pela utilização do hierárquico para gerar os clusters do programa, visto que, além de tudo, ele também deixa a compilação do código consideravelmente mais rápida.

## 5.5. METODOLOGIA PARA REGULAÇÃO POR DESEMPENHO

A técnica de análise para os clusters encontrados pelo método hierárquico é chamada de Regulação por Desempenho e nela se destacam os métodos *Yardstick Competition* e *Benchmark*, que serão explicados e exemplificados logo abaixo. Ambos os métodos podem ser utilizados no presente trabalho, porém, para sua finalização, foi escolhido o *Yardstick Competition*.

### 5.5.1. Benchmark

Modelo no qual as empresas são estimuladas a adotarem um padrão de desempenho baseado no comportamento do melhor elemento do universo analisado. Nesta abordagem, as empresas que apresentam desempenhos inferiores ao do melhor elemento são estimuladas a alcançarem este padrão de desempenho.

Para identificação do elemento de referência, a exemplo do Yardstick, pode-se adotar diferentes técnicas. A seguir descreve-se uma das possibilidades, visando tornar mais clara a abordagem do problema (TANURE, 2000).

A técnica consiste em determinar a reta que passa pelo ponto que representa o melhor desempenho das empresas analisadas. A Figura 14 representa esta forma de definir o padrão a ser adotado. A meta a ser atingida é representada pela norma do vetor que representa a menor distância entre o atual ponto de desempenho e a reta determinada.

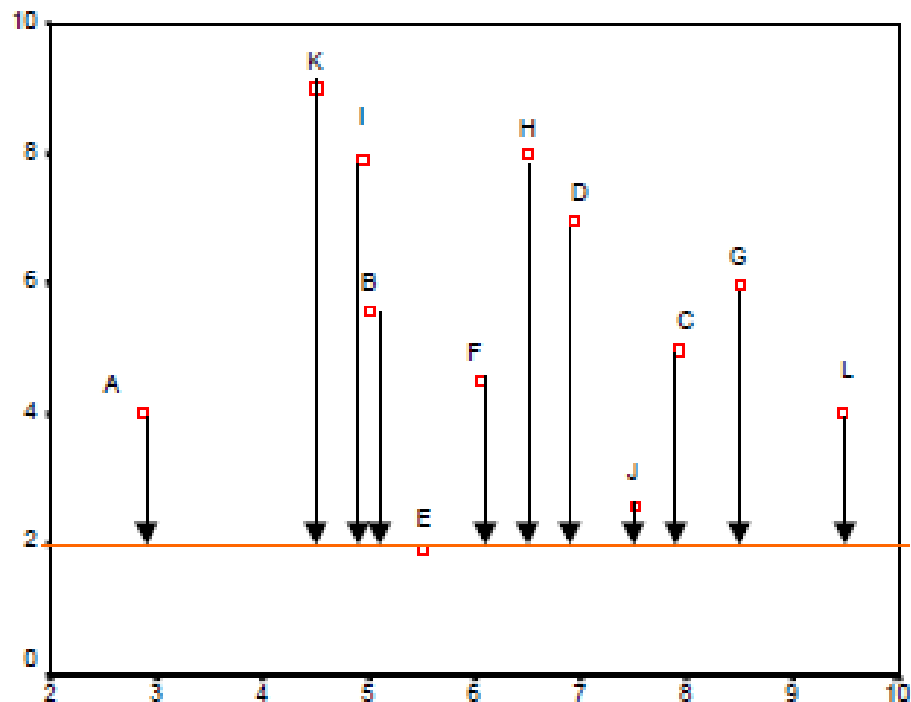


Figura 14. Benchmark. Fonte: (TANURE, 2000).

No exemplo apresentado na Figura 14 a empresa E é aquela de melhor desempenho dentro do universo analisado. Assim, as demais empresas devem evoluir seus padrões de desempenho visando atingir esta fronteira de eficiência (TANURE, 2000).

### 5.5.2. Yardstick Competition

O modelo de Yardstick Competition se baseia em empresas que são estimuladas a adotarem um desempenho baseado em um comportamento padrão. A fixação deste padrão decorre das análises de desempenho dos agentes que

compõem o universo considerado e do padrão que se julga razoável atingir. Nesta abordagem, as empresas que se situam abaixo da referência tem uma margem de folga para seu desempenho. As empresas situadas acima desta marca devem buscar melhorar seus indicadores para aproximar-se do valor de referência. Para identificação desta marca de referência pode-se adotar muitas técnicas diferentes, dependendo de cada contexto em que trabalha e do desempenho que se busca atingir. Analisa-se a seguir uma técnica passível de aplicação, que permitirá uma maior compreensão do problema (TANURE, 2000).

A técnica consiste em determinar a reta que passa pelo ponto médio das coordenadas que representam os desempenhos das empresas avaliadas. A Figura 15 representa esta forma de definir o padrão a ser adotado. Verifica-se que a meta a ser atingida é representada pela norma do vetor que apresenta a menor distância entre o atual ponto de desempenho e a reta determinada. Esta reta passa pelo valor de desempenho fixado como aquele a ser atingido (TANURE, 2000).

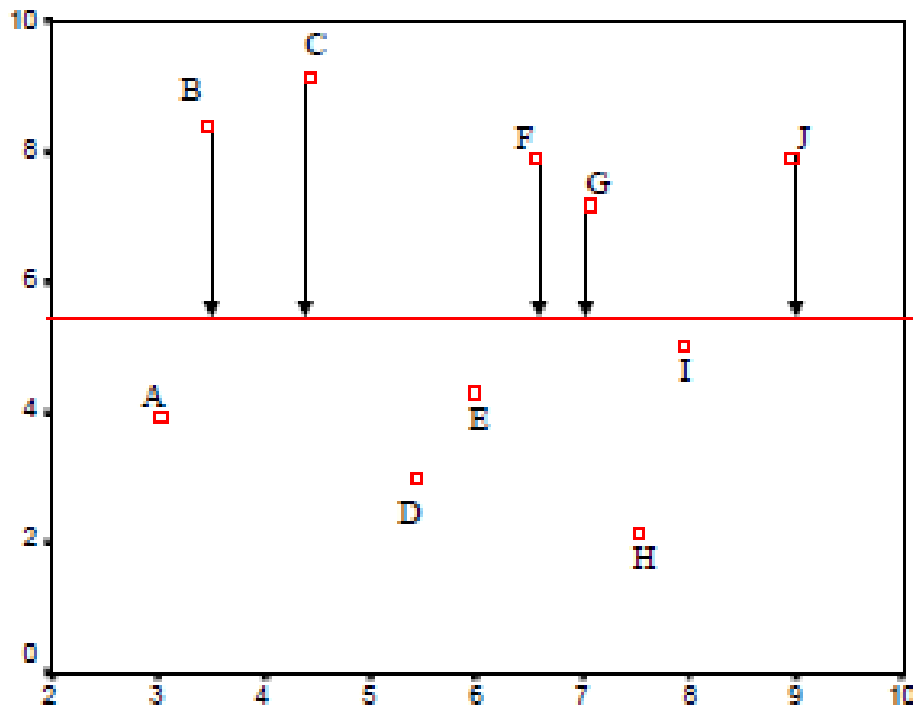


Figura 15. Yardstick Competition. Fonte: (TANURE, 2000).

Com base na figura acima, verifica-se que os conjuntos B, C, F, G e J devem evoluir seus desempenhos visando atingir a reta que define o desempenho desejável para o agrupamento estudado. Os conjuntos A, D, E, H e I encontram na

reta apresentada o limite de desempenho que podem atingir, devendo permanecer no semipleno inferior delimitado pela mesma (TANURE, 2000).

Neste exemplo, considerou-se que a meta foi estabelecida com base no desempenho médio das empresas, porém pode-se adotar como padrão de desempenho valores diferente como, por exemplo, medidas de posição (mediana, moda, percentis, etc) Cada aplicação poderá demandar um parâmetro de controle diferenciado (TANURE, 2000).

A partir de uma amostra de valores é possível determinar como padrão de desempenho, entre outros, a mediana dos valores ordenados e o *mid-range*, ou meio da faixa, que corresponde ao valor médio dos extremos da amostra ordenada (BASTISTA, 2013). O meio da faixa (*mid-range*) é determinado como a soma dividida por dois dos valores extremos de uma amostra de N elementos ordenados. Assim:

$$mid - range = \frac{a_1 + a_N}{2}$$

A mediana é definida ao se ordenar os elementos da amostra, sendo esta o valor que a divide ao meio, isto é, 50% dos elementos da amostra são menores ou iguais à mediana e os outros 50% são maiores ou iguais à mediana (SÓ MATEMÁTICA). A mediana (Me) é calculada por volumes ímpares de uma amostragem N como segue:

$$Me = a_{(N+1)/2}$$

e por volumes pares da amostragem:

$$Me = \frac{a_{N/2} + a_{(N/2)+1}}{2}$$

Como medida de localização, a mediana é mais robusta do que a média, pois não é tão sensível aos dados, isto é a mediana não é tão influenciada, como a média, por valores muito grandes ou muito pequenos que não caracterizam a amostra, mas que são desgarrados (*outliers*).

Quando a distribuição é simétrica, a média e a mediana coincidem. Se a distribuição dos dados for aproximadamente simétrica, a média aproxima-se da mediana, se for enviesada para a direita (alguns valores grandes como *outliers*), a média tende a ser maior que a mediana e se for enviesada para a esquerda (alguns valores pequenos como *outliers*), a média tende a ser inferior à mediana.

No caso deste trabalho, a meta estipulada para as UGs alcançarem foi a mediana de cada *cluster*. Tendo em vista o prazo de revisão da garantia física da usina e o conseqüente impacto comercial, o prazo para essa marca ser alcançada é de 5 anos. Foi feita uma projeção linear ligando os valores da média do histórico dos indicadores dos anos de 2009 a 2013, sendo essa média o valor no tempo zero da análise inicial até a mediana do histórico de cada *cluster*, correspondente ao valor que o indicador deve alcançar no tempo final igual a cinco anos. A partir dessa projeção, foi encontrado o valor do indicador de desempenho que a usina deve alcançar a cada ano, visto que a fiscalização desses valores pode ser realizada a qualquer momento.

Ou seja, foi escolhido o valor da mediana como o objetivo das unidades geradoras dentro de 5 anos, isto é, através de um projeção linear, as UGs devem elevar o indicador que possui atualmente até o valor de mediana da amostra. Desta forma, nenhuma unidade geradora está sendo reprovada pelo valor da mediana, mas valores intermediários advindos da projeção linear serão os seus limites, definidos individualmente.

A mediana foi escolhida em comparação com a média, visto que a média é muito sensível aos valores não representativos da amostra, os desgarrados. Assim como o *mid-range*, por utilizar os extremos considera com um peso elevado o valor mais baixo, que muitas vezes é zero ou próximo de zero, representando a indisponibilidade de uma UG que teve problemas naquele ano, e, junto ao valor máximo que é próximo de 100% representando as UGs com ótimo desempenho, resultam em valores baixos que não caracterizam um bom desempenho das UGs no SIN e, por isso, não foram utilizados como limites.

A média apenas foi utilizada para definir o valor inicial de indicador da unidade geradora na projeção linear, para isso foi utilizada a média dos últimos 5 anos, 2009 a 2013, e este valor médio projetado linearmente em 5 anos até a mediana do cluster ao qual pertence. A Tabela 6 da Seção 6.2.1. apresenta a distribuição das UGs em cada cluster, as Tabelas X e Y mostram os valores de medianas e *mid-range*, respectivamente, para comparação destes valores.



### 5.5.3. Método de Regulação de Desempenho para Não Piorar

Além do método de *Yardstick*, para projetar o bom desempenho das unidades geradoras será utilizado outro método de regulação por desempenho, chamado de Método de Regulação de Desempenho para Não Piorar.

Esse método foi utilizado no trabalho de redefinição de limites ligados ao setor de distribuição de energia elétrica e foi apresentado na Resolução Normativa 574 (SRC,2013). Os valores utilizados são relativos às reclamações que as distribuidoras recebiam. Foi estipulado que o ano 1 (2013) o número de reclamações poderia ser até 120% do maior valor histórico. Isso quer dizer que em 2013 poderia haver mais reclamações, ou seja, esse poderia ser o pior ano. Após o ano 1, os valores serão sempre relativos ao valor do ano anterior subtraídos de 25% da diferença entre o ano de maior número de reclamações e 120% do ano de menor número de reclamações. Ou seja, a tendência é que o número de reclamações se mantenha numa constante diminuição em relação ao ano anterior, porém com uma margem de piora em relação a ele próprio. A Tabela 5 permite que o método seja visualizado de acordo com o passar dos anos.

Tabela 5. Projeção de valores para método de Regulação de Desempenho para Não Piorar (SRC, 2013).

<b>2013 (A)</b>	<b>2014</b>	<b>2015</b>	<b>2016</b>	<b>≥ 2017 (B)</b>
120% maior valor histórico	valor anterior -25% x (A-B)	valor anterior -25% x (A-B)	valor anterior -25% x (A-B)	120% menor valor histórico

Usando o Indicador Disponibilidade como exemplo, o valor margem para a primeira coluna da tabela, representando o ano 1, inicia com 97,5% do valor relativo ao ano de pior desempenho da UG . O valor da quinta coluna, relativa ao ano 5, também é conhecido, pois ele equivale a 97,5% do valor do ano em que a disponibilidade esteve mais alta. A partir disso, é possível fazer uma projeção linear que mostre os valores acima dos quais a UG deve operar a cada ano no intervalo de 5 anos. Apesar do nome, esse método irá definir, para as unidades geradoras cujos valores dos indicadores estão num bom patamar, uma margem de queda de desempenho. Foi escolhido o percentil de 2,5% pois se acredita que esse valor corresponde a uma meta possível de ser alcançada.



Para as análises de Tempo Médio de Reparo e Taxa de falhas o raciocínio se inverte, visto que quanto menor for o valor desses indicadores, melhor é. Dessa forma, o valor da primeira coluna e da última é acrescido de 2,5%, que multiplicarão o valor de maior TMR e TF do histórico. Já o quinto valor é também acrescido de 2,5%, porém multiplica o valor de menor TMR e TF do histórico. A projeção linear é calculada da mesma forma que se calculou para DISP é calculada para TMR e TF, mudando apenas o percentil de partida e de final.

## 6. O PROGRAMA

O programa em Scilab, utilizando o Método Hierárquico de Clusterização, que permite a visualização da influência dos parâmetros sobre os indicadores é explicado no item 6.1, o exemplo utilizado para plotagem é o que permite a visualização da Potência por Idade. No item 6.2 é apresentado o programa utilizado para a análise de clusters, o exemplo utilizado para apresentação deste programa é da análise de clusters do indicador Taxa de Falhas, que considera os parâmetros físicos região, idade e potência.

### 6.1. PROGRAMA PARA GRÁFICOS DE ANÁLISE DE CORRELAÇÃO

Hair, Black, *et al* (2005) descreve medidas de similaridade baseadas em medidas de distância e em correlações e as compara. Os autores afirmam que “a similaridade entre objetos pode ser medida de diversas maneiras, mas três métodos dominam as aplicações de análise de agrupamentos: medidas correlacionais, medidas de distância e medidas de associação”. E ao comparar as medidas correlacionais e de distância, afirma que “As medidas de distância se concentram na magnitude dos valores e representam casos similares que estão próximos, mas podem ter padrões muito diferentes ao longo das variáveis.” e “Agrupamentos baseados em medidas correlacionais podem não ter valores similares, mas sim padrões similares”.

Neste trabalho a similaridade dos elementos foi medida pelos dois métodos: baseado em correlações e em distância, sendo que o primeiro permite uma análise inicial e o segundo foi efetivamente utilizado para formação de clusters, agrupando as unidades geradoras.

Como parte da análise baseada em correlações foram plotados, com o programa descrito nesta seção, gráficos de dispersão que não objetivam a formação de clusters, mas sim a análise do comportamento dos dados e a possível correlação linear, positiva ou negativa. A análise de clusters utiliza a medida de similaridade por distância, aplicada para os quatro parâmetros simultaneamente e formando grupos de elementos que podem ser comparados entre si por apresentarem características similares.

O código inicia na linha 4 com o comando *clear*, que cancela as variáveis que não estão protegidas. Ele remove as todas as variáveis nomeadas do ambiente exceto as protegidas pelo comando *predef*. Em seguida, na linha 6, o comando *tic()* é utilizado para ativar um cronômetro que contabilizará o tempo de execução do programa. Na linha 8, o comando *stacksize('max')* reserva a área de memória que as variáveis podem alocar no computador, maximizando-as e fazendo com que se possa criar *arrays* grandes. Esse comando tem a vantagem de diminuir o tempo dos cálculos. A linha 10 define que o arquivo cujo endereço está entre os parênteses se chamará *filename*. Na linha 12, a matriz T é criada a partir da leitura do arquivo *filename*, que está no formato \*.csv. O comando lê esse arquivo no formato de planilha Excel e o transforma em um arquivo de valores separados por ponto e vírgula. Além disso, ele especifica que a casa decimal é separada por vírgula e não ponto. No caso do trabalho, o arquivo *filename* é uma matriz, então T será uma matriz também.

A matriz T é composta por 549 linhas e 8 colunas. Cada linha corresponde a uma UG enquanto a primeira coluna tem os números no espaço amostral, na segunda coluna encontram-se as potências, na terceira coluna estão os índices relativos à região a qual pertence, a quarta coluna tem alocados os índices relativos ao tipo de turbina, bem como na quinta coluna são encontrados os anos de entrada em operação, a sexta, sétima e oitava colunas alocam os valores médios dos indicadores Disponibilidade, Taxa de Falhas e Tempo Médio de Reparo. Dessa forma, T é uma matriz toda numérica.

O arquivo *filename* é a planilha descrita no item 2, onde todos os dados necessários estão inseridos e são lidos de forma diferente para cada programa de acordo com os dados de entrada necessários. O parâmetro N é a leitura da primeira coluna da matriz T, então será do tamanho do número de linhas da matriz T indicado pelo sinal de dois pontos dentro dos parênteses, ou seja, terá o valor igual ao número de unidades geradoras e será também a quantidade de pontos no gráfico.

```

0001 //Nome do programa
0002 //Descrição do programa
0003
0004 clear;
0005
0006 tic();
0007
0008 stacksize('max');
0009
0010 filename = ('Endereço do arquivo')

```

```

0011
0012 T = csvRead(filename, ',', ',');
0013
0014 N = max(T(:,1));

```

Em seguida, na linha 16, declara-se que o parâmetro M vale inicialmente o valor de N. M será o número inicial de *clusters*, visto que pelo método hierárquico aglomerativo, cada dado da amostra corresponde inicialmente a um *cluster*. Na linha 18, define-se uma variável C que vai de 1 até o número N, ou o número de UGs. As linhas 20 e 21 declaram as variáveis x e y, estas serão os dados utilizados na análise, específica de cada programa gerado neste trabalho.

Neste exemplo as variáveis x e y serão obtidas a partir da leitura da quinta e da segunda coluna de dados desta matriz, respectivamente. Assim, no gráfico gerado por este programa o ano de entrada em operação ficará ao longo do eixo x e a potência ao longo do eixo y.

```

0016 M=N
0017
0018 C = 1:N;
0019
0020 x = T(:,5);
0021 y = T(:,2);

```

Inicia-se agora o procedimento de normalização. O comando da linha 23, bem como o da linha 35, será escrito no console do Scilab. As linhas 26 e 27 calculam a média aritmética dos valores de x e y e serão utilizadas no cálculo abaixo, dos desvios padrões de x e y. Esses procedimentos são necessários para a normalização, que é realizada nos comandos das linhas 32 e 33.

```

0023 disp('Início da Normalização');
0024
0025 //Acha Desvios-Padrão
0026 mx = (1/N).*sum(x);
0027 my = (1/N).*sum(y);
0028 dvx = sqrt((1/N).*sum((x - mx).^2));
0029 dvy = sqrt((1/N).*sum((y - my).^2));
0030
0031 //Normalização
0032 x = (1/dvx)*x;
0033 y = (1/dvy)*y;
0034
0035 disp('Fim da Normalização');

```

Os comandos abaixo iniciarão os cálculos para formar a matriz distância. Na linha 37, declara-se que NT possui o valor de uma matriz NxN, excluindo-se a sua diagonal e os elementos abaixo dela. Isso forma uma matriz triangular superior. Na linha 39, a matriz DI é montada com duas colunas, e o número de linhas equivalente

ao valor de NT. Ela é preenchida com *%inf*, que representa o maior valor na precisão double no Scilab. A mesma coisa é feita para o vetor DD logo abaixo e para a matriz D, NxN, na linha 42. Na linha 41, declara-se que o valor *conta* é zero.

```
0037 //Calcula Matriz de Distâncias
0038 NT = (N^2 - N)/2;
0039 DI = %inf*ones(NT,2);
0040 DD = %inf*ones(NT,1);
0041 conta = 0;
0042 D = %inf*ones(N,N);
```

Abaixo inicia a formação da matriz distância, sendo ela uma triangular superior cuja diagonal foi desconsiderada. O laço *for* determina que ocorram iterações de 1 até o valor anterior a N. Na linha 47, j terá exatamente NT pontos, que é tudo que vem acima da diagonal principal. Abaixo, comentado, está o comando para que a distância seja calculada segundo a métrica euclidiana. Porém no trabalho foi utilizada a métrica abaixo, Manhattan. O comando da linha 53 forma a matriz distância. *Conta*, abaixo, é o contador que se iguala ao valor de NT. Ele tem o tamanho da sequência do triângulo superior da matriz. Os comandos das linhas 44 e 57 imprimem no *console* o início e o fim do cálculo da distância.

```
0044 disp('Início Distâncias');
0045
0046 for i = 1:(N-1)
0047     for j = (i+1):N
0048         //D(i,j) = sqrt((x(i)- x(j))^2 + (y(i)-y(j))^2); //Euclidiana
0049         D(i,j) = abs(x(i)- x(j)) + abs(y(i)-y(j)); //Manhattan
0050         conta = conta + 1;
0051         DI(conta,1) = i; //linha
0052         DI(conta,2) = j; //coluna
0053         DD(conta) = D(i,j);
0054     end
0055 end
0056
0057 disp('Fim Distâncias');
```

Para visualizar o gráfico com os valores reais e não com os valores normalizados, deve-se multiplicar as variáveis x e y pelo desvio padrão, *dvx* e *dvy*, respectivamente. O comando *scf(1)* na linha 62 está relacionado ao *plot*, que vem logo em seguida. O *scf* determina que somente uma janela seja aberta para plotar os gráficos. O comando *plot* determina a plotagem dos valores pré-determinados de x e y na forma de bolinhas (*o*) vermelhas (*r*).

Visto que para idade e potência foram definidas faixas, é interessante que essas sejam plotadas no gráfico para facilitar a visualização desta divisão natural,

para este exemplo serão apresentados os comandos para visualizar as faixas da potência. Para isso define-se  $r$  para criar as linhas que cortam as faixas na linha 65 e com a função `plot` nas linhas 67 e 69 uma reta com o valor constante que divide as faixas é desenhada no gráfico.

```
0059 x=x*dvx;
0060 y=y*dvy;
0061
0062 scf(1);
0063 plot(x,y,'ro');
0064
0065 r=1920:1:2020;
0066 scf(1);
0067 plot(r,250,'k>');
0068 scf(1);
0069 plot(r,350,'k>');
```

O início do ordenamento dos menores aos maiores valores de distância se dá na linha 72, utilizando a função `gsort`, que foi explicada no item 0. A função `gsort` retorna dois *arrays*, o *DO* é o *array* ordenado das distâncias do triângulo superior e o *k* é a nova ordem das UGs. Dentro dos parênteses da função, o *DD* é a matriz que será ordenada, o *g* indica que o que será ordenado é um vetor de pilhas (matriz) e o *i* indica que a organização será em ordem crescente. Os comandos das linhas 71 e 73 imprimem no *console* o início e o fim do cálculo da distância.

```
0071 disp('Início do Ordenamento');
0072 [DO k] = gsort(DD,'g','i'); //Ordena
0073 disp('Fim do Ordenamento');
```

A próxima etapa é a clusterização. Ela inicia zerando o contador *conta*. O passo seguinte é a determinação do valor de *NCmin*, que é a quantidade de *clusters* que serão formados pelo programa ao analisar as distâncias. Embora os gráficos em duas dimensões não estejam sendo utilizados para a análise de *clusters*, para os gráficos que relacionam os parâmetros físicos entre si, os cálculos da distância e agrupamento pela similaridade foram mantidos para visualização de como funciona o método. Além de que os grupos formados nesses gráficos tiveram uma boa distribuição e permitiram uma análise da divisão natural das características relacionadas.

Para definir o número de *clusters* que deveriam ser formados foram feitas análises visuais de como se distribuíam os pontos e escolheu-se o número de *clusters* que diminuíssem as distâncias *interclusters*, isto é, que entre um *cluster* e

outro as distâncias não fossem muito grandes e ainda mantivesse o menor número de *clusters* possível, para evitar *clusters* com poucos elementos.

O laço *for* iniciado na linha 80, gera as iterações *co* que vão de 1 até o valor de *NT*, que é o número de elementos de *DO* e de *K*. Em seguida, define-se que *conta* receberá seu valor anterior acrescido de uma unidade, começando no valor de 1. Na linha 84, registra-se o elemento de *DI* encadeado de *i* e, em seguida, na linha 85, o elemento de *DI* encadeado do valor *j*.

```
0076 disp('Início Clustering');
0077 conta = 0;
0078 NCmin = 4;
0079
0080 for co = 1:NT
0081
0082     conta = conta + 1;
0083
0084     i = DI(k(co),1);
0085     j = DI(k(co),2);
```

Dando continuidade à clusterização, na linha 87 se inicia um laço *if* que compara se são diferentes os elementos *i* e *j* da matriz *C*. Nesse comando ele apenas concatena dois pontos se estes não já pertencerem ao mesmo *cluster*, ou seja, esse comando verifica se um ponto já não foi unido a um outro *cluster*. O comando seguinte plota um gráfico 2D, de linhas entre os dois pontos que se juntaram ao mesmo *cluster* devido à função anterior. *xtitle* imprime no gráfico primeiramente o título, depois o nome do eixo *x* e em seguida o nome do eixo *y*. Na linha 91, a cada vez que se junta dois *clusters*, o número de *clusters* diminui de um.

Nas linhas 93 e 94, são criadas duas variáveis auxiliares, *Ci* e *Cj*. Em seguida inicia-se outro laço *for* cujo número de iterações é o número de pontos *N*. O laço *if* que começa logo abaixo verifica os pontos que pertencem a *C(h)*, que foi associado ao *C(j)* na linha 94. Em seguida o programa avalia se esse ponto já pertence ao *cluster* ao qual está sendo associado. Se sim, esse ponto também já pertence ao *cluster Ci*. Os dois últimos laços *for* e *if* são encerrados nas linhas 99 e 100. Em seguida um laço *if* é iniciado e compara se são iguais o número de *cluster M* e o número de *cluster NCmin*. Esse é o critério de parada visto que *M* é o número inicial de *clusters*, que vai sendo decrescido de 1 até se igualar a *NCmin*. A função da linha 103, *tabul*, imprime a matriz *C*, que terá o número de linhas igual ao valor definido de *NCmin*. Tanto a função *tabul* quanto a matriz *C* foram mais bem explicadas no item 5.2.6. A matriz *C* só será impressa se for constatado na linha 102 que *M* é igual a

*NCmin*. Em seguida é impresso no console que o processo de clusterização foi finalizado e todos os laços são encerrados. A função *toc()* lê o cronômetro iniciado pela função *tic()* no início do programa.

```

0087     if (C(i) ~= C(j))
0088         plot2d([x(i) x(j)], [y(i) y(j)]);
0089         xtitle('Potência x Ano de Entrada em Operação', 'Ano de Entrada em
Operação', 'Potência');
0090
0091         M = M - 1;
0092
0093         Ci = C(i);
0094         Cj = C(j);
0095
0096         for h = 1:N
0097             if(C(h) == Cj)
0098                 C(h) = Ci;
0099             end
0100         end
0101
0102         if (M == NCmin) then
0103             disp(tabul(C));
0104             disp('Fim Clustering');
0105             return;
0106
0107         end
0108     end
0109 end
0110
0111 toc();

```

Quando o início e o fim de uma etapa do programa são dispostos no *console* do Scilab, a visualização permite que se identifique em qual fase de execução ele está e caso haja um erro, esse erro será apresentado pelo Scilab depois do início de um procedimento e antes do fim dele. Assim, fica mais fácil identificar e corrigir o problema.

## 6.2. PROGRAMA PARA ANÁLISE DE *CLUSTERS*

O programa para análise de *clusters* assemelha-se com o descrito na seção anterior, para análise de correlação. A maior diferença está no número de dimensões da análise, na correlação apenas dois parâmetros, ou um indicador e um parâmetro, eram analisados, gerando gráficos em duas dimensões. Na análise de *clusters* são analisados todos os quatro parâmetros, ou seja, quatro dimensões, impossibilitando a visualização por gráfico.



As diferenças entre o programa da análise de correlação e da análise de *clusters* serão pontualmente apresentadas nesta seção. A primeira mudança deve acontecer na declaração das variáveis, em vez de declarar apenas as variáveis *x* e *y*, também serão declaradas as variáveis *w* e *z*. Nas linhas 20 a 23 do programa são atribuídos a cada variável os valores de uma coluna da matriz *T*, a saber, a variável *x* será dada pela leitura da segunda coluna, onde estão as potências; a variável *y*, dada pela terceira coluna, conterà os índices relativos às regiões; a variável *w*, dada pela quarta coluna, conterà os índices relativos aos tipos de turbina; a variável *z*, dada pela quinta coluna, conterà os anos de início da operação.

```
0019
0020 x = T(:,2);
0021 y = T(:,3);
0022 w = T(:,4);
0023 z = T(:,5);
0024
```

A normalização também será modificada pela adição de novas variáveis, a média e o desvio padrão deverão ser calculados também para as variáveis *w* e *z*, além de *x* e *y*, como é mostrado nas linhas 28 a 41.

```
0025 disp('Início da Normalização');
0026
0027 //Acha Desvios-Padrão
0028 mx = (1/N)*sum(x);
0029 my = (1/N)*sum(y);
0030 mw = (1/N)*sum(w);
0031 mz = (1/N)*sum(z);
0032 dvx = sqrt((1/N)*sum((x - mx).^2));
0033 dvy = sqrt((1/N)*sum((y - my).^2));
0034 dvw = sqrt((1/N)*sum((w - mw).^2));
0035 dvz = sqrt((1/N)*sum((z - mz).^2));
0036
0037 //Normalização
0038 x = (1/dvx)*x;
0039 y = (1/dvy)*y;
0040 w = (1/dvw)*w;
0041 z = (1/dvz)*z;
0042
0043 disp('Fim da Normalização');
```

O cálculo das distâncias deverá englobar as novas variáveis inseridas, o novo cálculo da distância Manhattan está na linha 56 do programa:

```
0052 disp('Início Distâncias');
0053
0054 for i = 1:(N-1)
0055     for j = (i+1):N
0056         D(i,j) = abs(x(i)- x(j)) + abs(y(i)-y(j)) + abs(w(i)- w(j)) + abs(z(i)- z(j));
```

```

0057     conta = conta + 1;
0058     DI(conta,1) = i; //linha
0059     DI(conta,2) = j; //coluna
0060     DD(conta) = D(i,j);
0061     end
0062 end
0063
0064 disp('Fim Distâncias');

```

### 6.2.1. Escolha do Número de *Clusters*

No programa, o número de *clusters* deve ser escolhido antes da execução do mesmo, para a escolha deste número deve-se considerar a distribuição dos dados entre os *clusters*. Ao escolher apenas a formação de 1 *cluster*, este conterà todos os elementos da amostra. À medida que o número de *clusters* é aumentado, os elementos se distribuem nos *clusters* de forma que os mais similares continuem unidos e os pontos mais distantes formem outros *clusters*.

A Tabela 6 apresenta o número de elementos em cada *cluster* quando se varia o número de *clusters* de 2 a 20, sendo mostrados apenas os que apresentaram maiores alterações. Foi escolhida a distribuição com 15 *clusters*, pois, esta quantidade de *clusters* proporcionou uma distribuição onde não apenas um grupo continha a maioria de todos os elementos, mas dividiram-se em pelo menos dois grupos maiores e alguns grupos com quantidades significativas de elementos.

A Tabela 6 continua até o número 20 para mostrar que a tendência é aumentar o número de *clusters* com poucos elementos e manter os dois *clusters* maiores. Para simplificar o procedimento de regulação, é ideal uma boa distribuição dos pontos nos *clusters*, isto é, evitar grupos com quantidades muito pequenas de elementos.

É possível perceber, que alguns *clusters* não variam a quantidade de elementos entre os passos, tais como 4, 6, 9 e 11, estes podem ser visualizados como aqueles que estão mais distantes do grupo com as UGs de características mais comuns. Suas características são bastante incomuns e próximas e por isso manter-se-ão unidas.

Tabela 6. Distribuição dos elementos em *clusters*.

Número de <i>Clusters</i>	2	4	6	7	8	10	14	15	16	17	20
Quantidade de UGs por <i>cluster</i>	9	9	6	6	6	6	6	6	6	6	6
	540	14	9	9	9	9	9	9	9	9	9
		515	505	336	117	117	4	190	190	184	16
		11	14	14	336	298	11	4	4	4	6
			4	169	14	6	117	11	11	11	4
			11	11	4	32	40	117	117	117	11
					52	14	6	40	40	40	117
					11	4	32	6	6	6	26
						52	14	32	6	6	6
						11	4	14	14	32	6
							8	4	4	14	32
							11	8	8	4	14
							243	11	11	8	4
							44	53	53	11	8
								44	34	53	8
									10	34	11
									10	53	
										34	
										10	
										168	

### 6.2.2. Identificação das Unidades Geradoras de Cada *Cluster*

A saída do programa, obtida através da função *tabul*, é a quantidade de unidades geradoras existentes em cada *cluster*. Porém, é necessário identificar quais são as unidades geradoras que estão em cada *cluster* e quais são os valores de seus indicadores, para isso é necessário escrever um arquivo com o comando *csvWrite()*, o código utilizado para a escrita deste programa é encontrado na Figura 16.

```

105 .....ind=[1:1.0:549];
106 .....Q = [ind' C'];
107 .....csvWrite(Q, 'C:\cluster\todososparametros.csv', 'ascii(9)', ',');

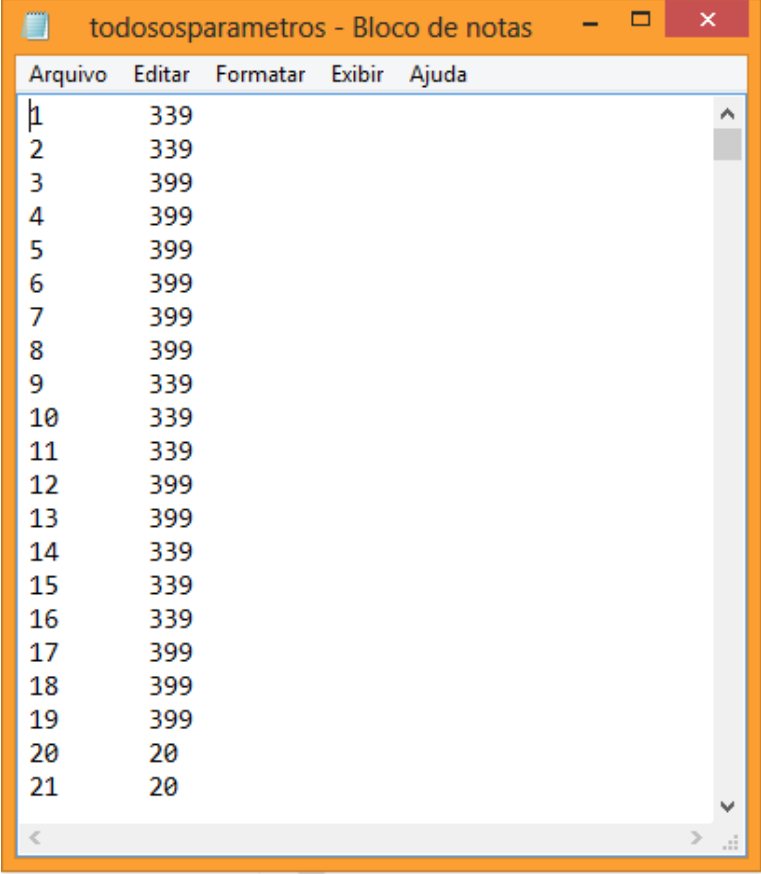
```

Figura 16. Código do programa para escrita do arquivo \*.csv.

No arquivo escrito em \*.csv, cada unidade geradora está identificada pelo índice, representado pela *array ind*, e apresenta o *cluster* em que está inserida, o *array C*. A matriz *Q* transpõe os *arrays* de forma a criar duas colunas para serem

escritas no arquivo que ao ser aberto em bloco de notas tem a aparência da Figura 17.

A função `csvWrite` tem como entrada a matriz com os dados desejados no arquivo, o nome do local onde deverá ser guardado o arquivo, a forma de separação dos dados, neste caso a separação é feita por tabulações com o comando “`ascii(9)`” e o símbolo de decimal dos dados, neste caso é a vírgula.



Arquivo	Editar	Formatar	Exibir	Ajuda	
1					339
2					339
3					399
4					399
5					399
6					399
7					399
8					399
9					339
10					339
11					339
12					399
13					399
14					339
15					339
16					339
17					399
18					399
19					399
20					20
21					20

Figura 17. Arquivo escrito pelo comando `csvWrite`.

Visto que a ordem do índice respeita a ordem da planilha, utiliza-se a planilha da Figura 3, utilizada para o programa, para identificar qual é a unidade geradora e quais são os valores dos indicadores, montando uma planilha conforme apresentado na Figura 18, onde ao aplicar os filtros na coluna B podemos separar cada *cluster* e seus elementos.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Índice	Cluster	Potência	Região	Turbina	Ano	2008	2009	2010	2011	2012	2013
2	1	339	50	3	2	2008	99,4934	87,59475	54,71461	74,03139	99,36798	95,0801
3	2	339	50	3	2	2008			56,97393	70,64498	95,85819	98,96062
4	3	399	232	4	1	1978	96,63972	98,95472	93,21461	97,9024	95,86274	97,996
5	4	399	232	4	1	1978	98,95055	99,30289	88,37595	97,53501	96,86248	64,97907
6	5	399	232	4	1	1979	93,24416	98,49391	97,7344	87,96252	99,64329	97,66648
7	6	399	232	4	1	1979	93,10337	98,2188	97,8417	91,18779	98,95226	89,41933
8	7	399	232	4	1	1979	92,98668	84,2618	98,76027	98,29091	98,54622	94,21994
9	8	399	232	4	1	1979	92,48216	86,9804	98,61454	97,60635	93,21076	97,48954
10	9	339	110	4	2	2005	86,06387	90,8204	94,13299	93,24011	96,80802	93,24981
11	10	339	110	4	2	2005	91,20313	93,10483	96,02568	95,44806	97,89674	93,93988
12	11	339	110	4	2	2005	92,24746	88,08086	95,12709	97,51465	93,87447	94,12405
13	12	399	37	3	1	2009			94,97127	97,56183	98,76063	97,957
14	13	399	37	3	1	2009			94,16096	98,30137	99,71672	96,05422
15	14	339	70	4	2	2006	82,2744	94,18455	99,47451	97,28158	98,91982	88,83048
16	15	339	70	4	2	2007	81,37504	92,49905	99,75209	95,27245	99,05908	89,93493
17	16	339	70	4	2	2007	88,23638	89,31773	97,72241	93,30708	95,50129	99,19292
18	17	399	80	4	1	2006	96,94198	97,59932	99,69292	94,77207	94,46968	98,80898
19	18	399	80	4	1	2006	96,35018	99,4102	99,69901	95,33638	97,7755	97,3868
20	19	399	80	4	1	2006	97,36206	99,65601	99,87766	95,30898	97,18446	98,91572
21	20	20	100	2	2	1977	98,86726	77,98059	99,29129	84,47374	96,80726	99,71081
22	21	20	100	2	2	1977	99,06819	96,26046	99,87405	79,15354	97,93185	98,34551
23	22	20	100	2	2	1977	98,38665	77,24201	67,58238	93,38071	96,75414	98,62881
24	23	20	100	2	2	1977	50,03245	71,81355		0	19,63627	98,21442
25	24	339	35	4	2	2009			92,41058	92,59075	98,34605	97,56906
26	25	339	35	4	2	2009			89,34056	90,45719	92,71877	98,66895
27	26	339	35	4	2	2010			93,02416	19,06145	68,85037	96,14612
28	27	339	35	4	2	2010			94,46518	89,01294	93,22025	99,38604
29	28	339	48	4	2	1969	95,08462	97,23992	98,63185	93,64631	95,3588	90,67618
30	29	339	48	4	2	1966	97,87701	97,77169	96,51865	96,78976	87,91382	90,36206

Figura 18. Planilha montada com os índices e *clusters* do arquivo gerado pelo Scilab e os parâmetros físicos.

## 7. CORRELAÇÃO E SELEÇÃO DE VARIÁVEIS

Ferramentas de análise de correlação são utilizadas para examinar pares de variáveis de medida a fim de determinar se as duas variáveis de medida tendem a se mover juntas. A correlação entre variáveis pode ser positiva, significando que os valores altos de uma variável tendem a ser associados aos valores altos da outra, ou negativa, se os valores baixos de uma variável tendem a ser associados aos valores altos da outra. Quando os valores das duas variáveis tendem a não estar relacionados, a correlação é próxima de zero.

O coeficiente de correlação de Pearson é uma medida de relacionamento natural que expressa a similaridade de duas variáveis relativas à “relação linear”. Quanto maior for o valor absoluto, mais relacionadas de forma linear são as variáveis em questão, ou seja, maior a proximidade (MINGOTI, 2013). Na Tabela 7 é apresentada a matriz de correlações entre parâmetros e indicadores e de ambos entre si. A Tabela 7 foi gerada no programa Excel do Pacote Microsoft Office 2010, através da Análise de Dados, e utiliza a mesma base de dados utilizada no programa de formação de clusters.

Tabela 7. Matriz de correlações com parâmetros físicos e indicadores.

	<i>DISP</i>	<i>TF</i>	<i>TMR</i>	<i>Potência</i>	<i>Região</i>	<i>Turbina</i>	<i>Idade</i>
<i>DISP</i>	1						
<i>TF</i>	-0,197104	1					
<i>TMR</i>	-0,343134	-0,007625	1				
<i>Potência</i>	0,146496	-0,218747	0,0689008	1			
<i>Região</i>	-0,048173	-0,306214	0,0578481	-0,26834	1		
<i>Turbina</i>	-0,121085	-0,058475	0,1290151	-0,347741	0,1856106	1	
<i>Idade</i>	0,2326166	0,2565399	-0,09436	0,1507219	-0,394013	-0,10143	1

Embora os valores de correlação encontrados não possuam valores muito elevados, os maiores, acima de 10% de correlação, serão considerados para a análise da influência dos parâmetros físicos sobre os indicadores. Numa primeira análise pode-se retirar as seguintes informações:

- A correlação de valor absoluto mais alto é entre a região e a data de entrada em operação da UG, isso se deve pela exploração dos potenciais hidráulicos ter iniciado predominantemente no Sudeste e depois expandido para outras regiões.

Nas primeiras três décadas da amostra (de 1924 a 1954) havia UGs apenas no Sudeste, nas três décadas seguintes (de 1955 a 1984) houve expansão para as regiões Sul e Nordeste em mesma escala, 13% das unidades geradoras com início de operação nesta data, e para a região Norte e Centro Oeste, 8%, como mostrado na Figura 19.

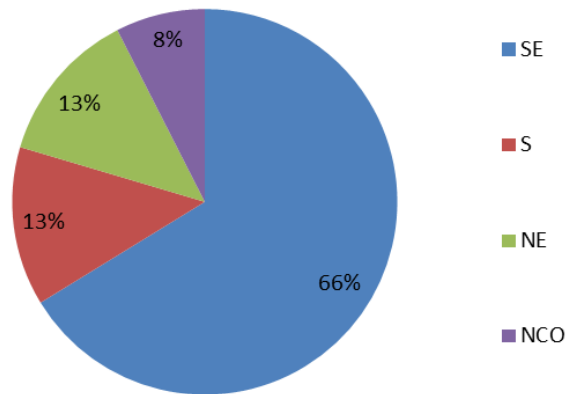


Figura 19. Proporção de UGs em cada região no período de 1955 a 1984.

No período que compreende 1985 a 2012 houve uma maior exploração dos potenciais hidroelétricos nas regiões Norte e Centro Oeste em mesma escala, 13% das unidades geradoras com início de operação nesta data, e para a região Norte e Centro Oeste, que até então representava a menor parcela de UGs.

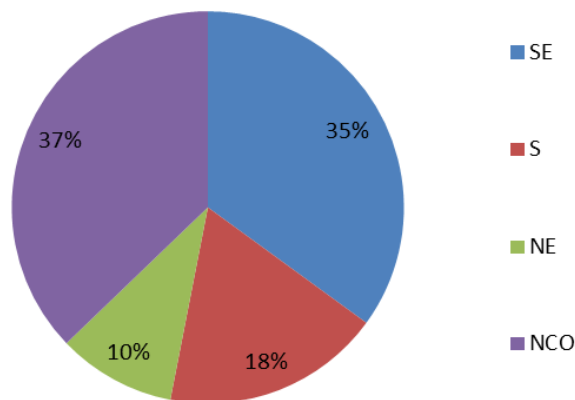


Figura 20. Proporção de UGs em cada região no período de 1985 a 2012.

Outra correlação de valor absoluto considerável nessa amostra é da potência com o tipo de turbina e esta é explicada pela própria tecnologia de turbina ter uso especificado pelo potencial local e, assim como apresentado na Figura 21, cada turbina tem uma faixa de potências que podem ser geradas.

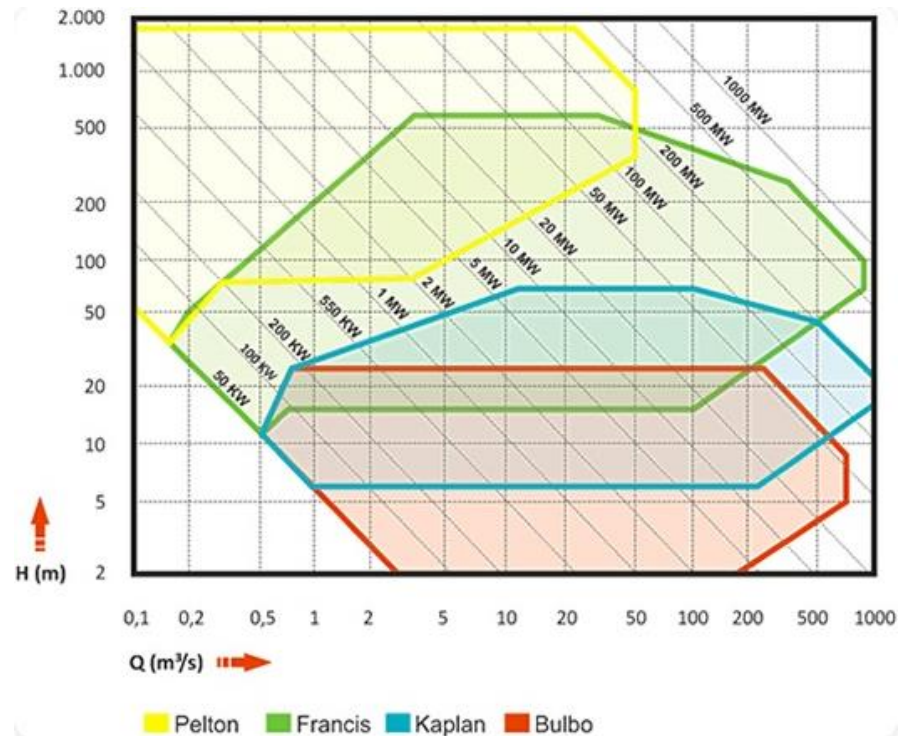


Figura 21. Escolha do tipo de turbina de acordo com a queda e vazão do potencial e as potências obtidas. Fonte: <[http://www.hacker.ind.br/produtos\\_turbinas\\_hidraulicas.php](http://www.hacker.ind.br/produtos_turbinas_hidraulicas.php)> acessado em 12 de maio de 2014.

- A disponibilidade e o tempo médio de reparo mostraram uma relação maior do que entre a disponibilidade e a taxa de falhas. Isso mostra que a gestão da manutenção bem realizada, diminuindo o tempo médio de reparo consequentemente aumenta a disponibilidade do equipamento significativamente.

Ainda que a taxa de falhas do equipamento esteja relacionada com a disponibilidade, esta correlação é menor do que quando relacionada com o tempo médio de reparo.

Os resultados obtidos com a correlação podem ser visualizados por meio de gráficos em que são relacionados os parâmetros físicos com os indicadores, assim como os parâmetros físicos entre si.

### 7.1. INFLUÊNCIA DOS PARÂMETROS FÍSICOS SOBRE OS INDICADORES

Para a verificação da influência dos parâmetros físicos sobre os indicadores serão utilizados valores médios dos indicadores DISP, TF e TMR. Visto que os valores do RAD são apresentados em base anual, cada ano é representado pelo



valor do mês de dezembro, e desta forma tem-se, para cada indicador, seis valores, um para cada ano. Para essa análise decidiu-se por utilizar a média dos valores de todos os anos, representando cada UG com um único valor representativo para cada indicador. Visto que a distribuição dos valores durante os anos não segue nenhuma tendência, o ideal seria manter este valor sempre o melhor e conseqüentemente o mais constante possível, neste caso, a média é considerada uma boa métrica.

Os resultados obtidos com o TCC 1 permitiram observar que os parâmetros físicos escolhidos para a análise de *clusters* poderiam influenciar nos indicadores de diversas formas. Para esta segunda parte do trabalho, também foi utilizado o programa de análise de *clusters* para observar a relação entre os parâmetros e os indicadores Disponibilidade, Taxa de Falhas e Tempo Médio de Reparo.

Para uma análise inicial, tem-se que a amostra inteira tem uma Disponibilidade média de 90,53% e um desvio padrão de 8,81, a Taxa de Falhas média da amostra é de 7,10 falhas por ano e o desvio padrão de 12,08, o Tempo Médio de Reparo médio da amostra é de 54,56 horas e o desvio padrão foi de 147,06, como apresentado na Tabela 8. Essa mesma análise foi realizada para relacionar os indicadores com os parâmetros físicos, a média e o desvio padrão foi calculada para cada faixa de potência e idade, para as regiões e para os tipos de turbina.

Tabela 8. Média e desvio padrão da amostra para os indicadores DISP, TF e TMR.

	Disponibilidade (%)	Taxa de Falhas (ano <sup>-1</sup> )	Tempo Médio de Reparo (horas)
Média	90,53	7,10	54,56
Desvio Padrão	8,81	12,08	147,06

### 7.1.1. Disponibilidade X Potência

A disponibilidade relaciona-se com a potência de acordo com o apresentado visualmente na Figura 22 e com os valores da Tabela 9, pode-se observar que as unidades geradoras com potências mais baixas tendem a apresentar alguns casos com baixa disponibilidade. E ainda que as unidades geradoras com alta e média potência não possuem casos de disponibilidades muito baixas.

O fato de UGs de mais alta potência apresentarem disponibilidades mais altas pode ser explicado pela robustez das máquinas: quanto maior o porte da unidade

geradora, mais energia ela estará fornecendo ao SIN e, conseqüentemente, maior a importância desta no atendimento à demanda. Assim, as máquinas de empreendimentos de grande porte devem ser mais robustas e os agentes responsáveis por estas devem ter uma gestão eficiente para evitar a indisponibilidade.

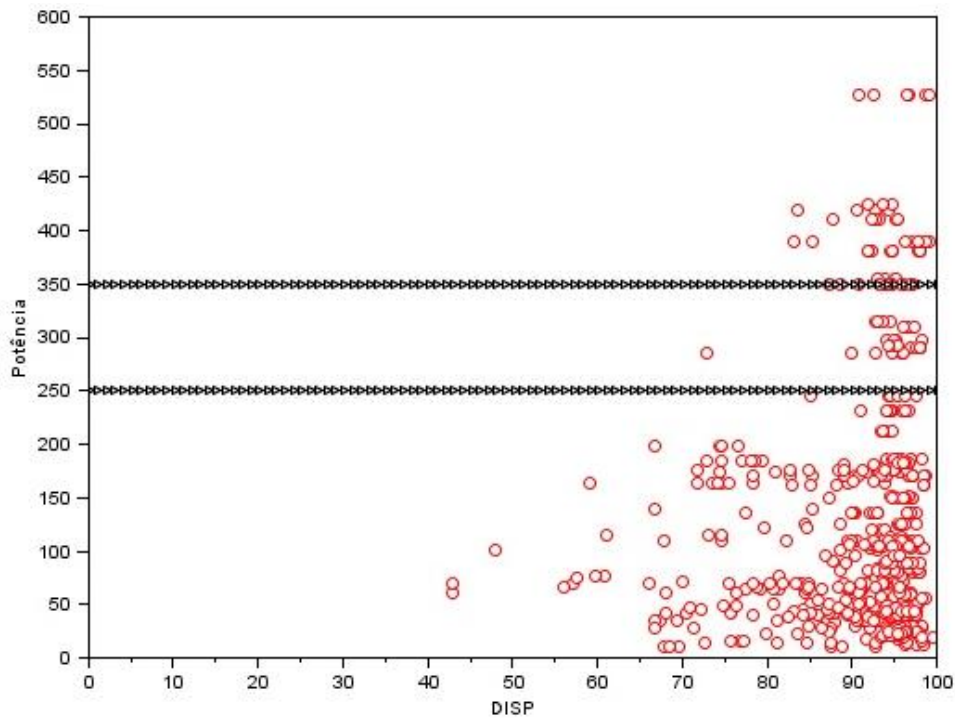


Figura 22. Relação entre parâmetro físico Potência e indicador Disponibilidade.

Tabela 9. Média e desvio padrão da disponibilidade para faixas de potência.

<b>Faixas de Potência</b>	<b>Média (%)</b>	<b>Desvio Padrão (%)</b>
Baixa potência	89,89	9,23
Média potência	94,24	4,31
Alta potência	94,18	3,85

### 7.1.2. Disponibilidade X Tipo de Turbina

A relação entre a disponibilidade e o tipo de turbina apresentada na Figura 23, embora sua situação não esteja clara visualmente, pode-se observar, analisando a Tabela 10, que as turbinas Francis apresentam a disponibilidade média maior, seguido das turbinas Kaplan, o tipo de turbina Pelton apresenta valor médio de disponibilidade ainda mais baixo e a do tipo Bulbo apresenta o menor valor médio.

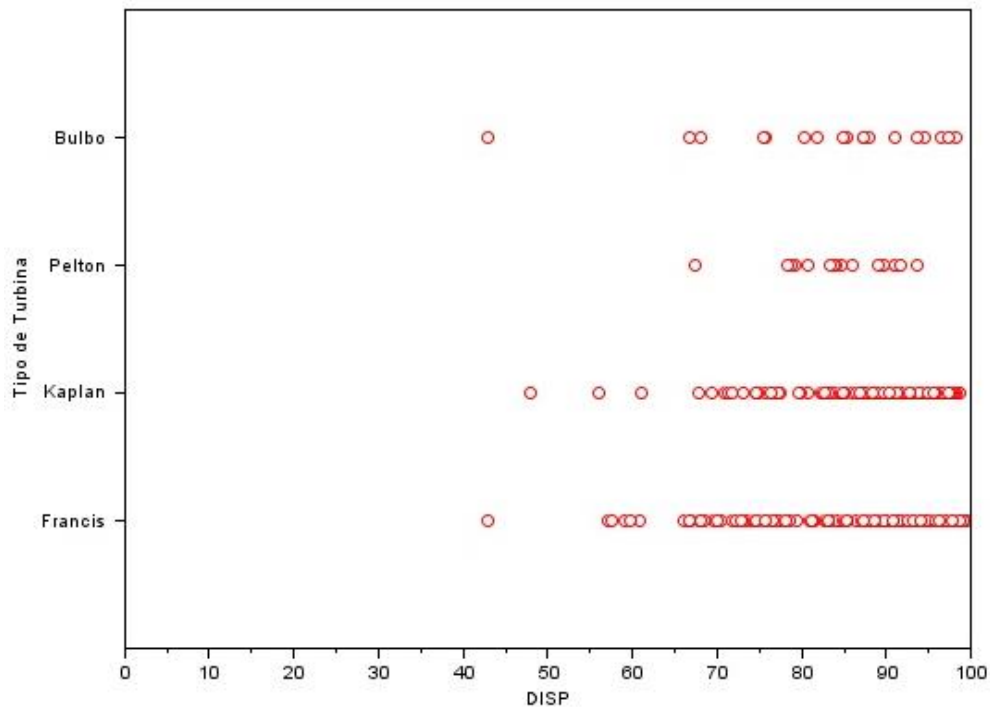


Figura 23. Relação entre parâmetro físico tipo de turbina e indicador disponibilidade.

Tabela 10. Média e desvio padrão da disponibilidade para tipos de turbina.

<b>Tipo de Turbina</b>	<b>Média (%)</b>	<b>Desvio Padrão (%)</b>
Francis	91,35	8,64
Kaplan	90,29	8,21
Pelton	85,12	6,59
Bulbo	83,36	13,06

### 7.1.3. Disponibilidade X Região

A disponibilidade e a região estão relacionadas no gráfico na Figura 24 e com os valores da Tabela 11, é possível constatar que o Nordeste apresentou valores mais baixos de disponibilidade, mas também apresentou um desvio padrão mais elevado, representando uma maior dispersão das UGs. O Sul apresentou o maior valor de disponibilidade média e o menor desvio padrão, indicando que as UGs dessa região apresentaram melhor desempenho.

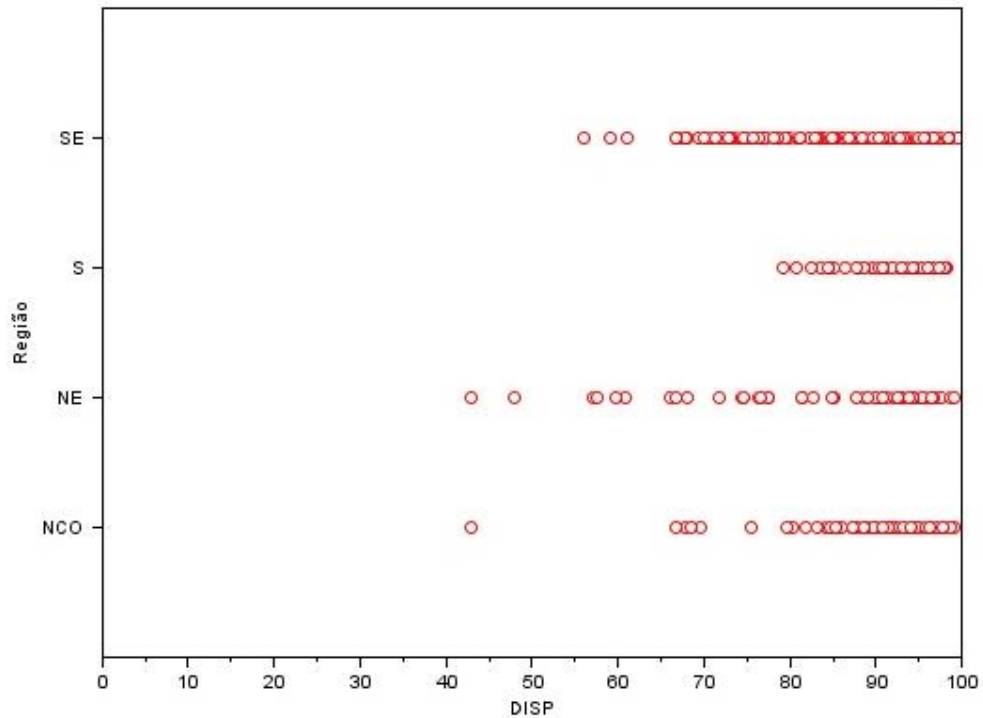


Figura 24. Relação entre parâmetro físico região e indicador disponibilidade.

Tabela 11. Média e desvio padrão da disponibilidade para regiões.

Região	Média (%)	Desvio Padrão (%)
Norte-Centro Oeste	91,69	7,67
Nordeste	85,19	13,70
Sul	93,55	4,20
Sudeste	90,25	8,50

#### 7.1.4. Disponibilidade X Ano de Entrada em Operação

As unidades geradoras estão mais concentradas nas faixas de idade intermediárias e mais novas, poucas estão na faixa das mais antigas, assim, o gráfico da Figura 25 não tem uma fácil visualização e pode levar a concluir que as mais antigas têm as melhores disponibilidades.

Na Tabela 12 pode-se perceber que as mais antigas possuem a menor média de disponibilidade e o desvio padrão baixo, mas por ter uma menor quantidade de elementos, o fato de ter UGs com baixa disponibilidade tem um peso alto na média. As mais novas têm a maior média de disponibilidade e o menor desvio padrão, apresentando o melhor desempenho. As de idade intermediária, embora apresentem um valor de disponibilidade bastante próximo das mais antigas, possui uma dispersão maior dos elementos, como pode ser visto no desvio padrão desta faixa.

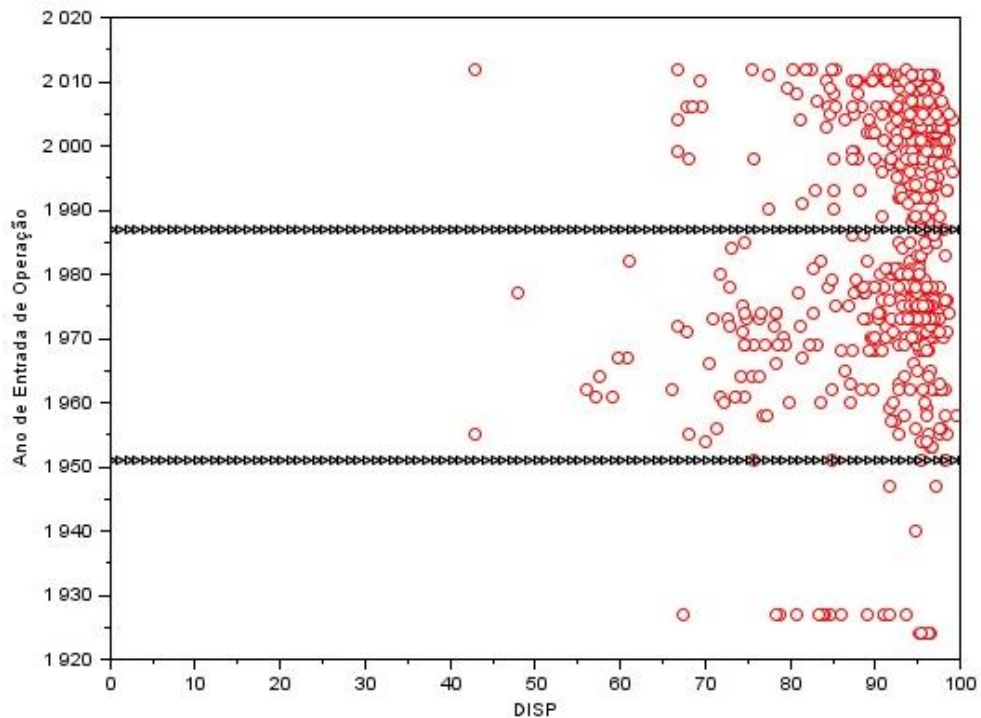


Figura 25. Relação entre parâmetro físico ano de entrada em operação e indicador disponibilidade.

Tabela 12. Média e desvio padrão da disponibilidade para faixas de idade.

Faixas de Idade	Média (%)	Desvio Padrão (%)
Mais antigas	88,44	7,88
Intermediárias	88,97	10,22
Mais novas	92,24	6,92

### 7.1.5. Taxa de Falhas X Potência

A partir da análise do gráfico na Figura 26 é possível perceber que a maior concentração de UGs está na faixa de baixa potência e que são elas que apresentam maior dispersão quanto ao valor de Taxa de Falhas. As UGs que se encontram nas faixas de potência mais altas não possuem tamanha variação de valor da Taxa de Falhas, como pode ser observado ao se comparar os valores de média e desvio padrão da Tabela 13. Um dos motivos para que isso ocorra é a robustez da máquina, refletindo o que acontece com a Disponibilidade.

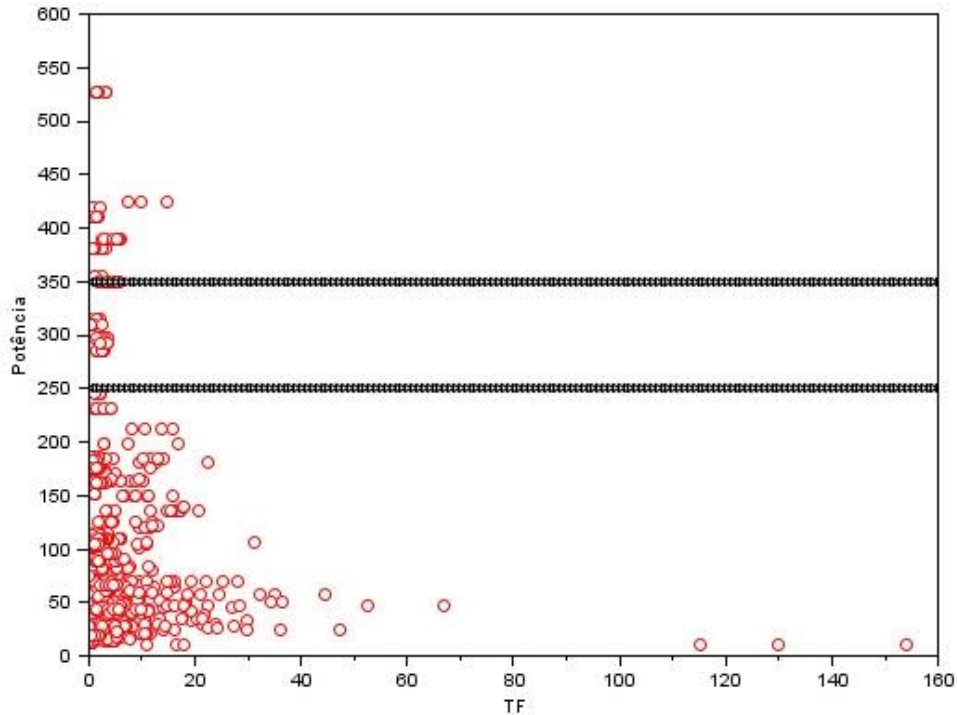


Figura 26. Relação entre parâmetro físico potência e indicador taxa de falhas.

Tabela 13. Média e desvio padrão da taxa de falhas para faixas de potência.

<b>Faixas de Potência</b>	<b>Média (ano<sup>-1</sup>)</b>	<b>Desvio Padrão (ano<sup>-1</sup>)</b>
Baixa potência	7,85	12,90
Média potência	2,55	1,43
Alta potência	2,97	2,75

### 7.1.6. Taxa de Falhas X Tipo de Turbina

Quando a taxa de falhas é relacionada ao tipo de turbina, percebe-se nitidamente a partir do gráfico, que as turbinas do tipo Francis são as únicas que apresentam taxa de falhas muito altas. Um dos motivos para que isso ocorra é a quantidade de dados na amostra, as turbinas em maior quantidade são as do tipo Francis.

As turbinas do tipo Pelton são as que apresentam a menor média de taxa de falhas, bem como o menor desvio padrão, permitindo concluir que são as que apresentam melhor resultado.

As turbinas do tipo Bulbo e Kaplan apresentam uma distribuição semelhante ao longo do eixo da taxa de falhas, porém a média e o desvio padrão das turbinas tipo bulbo são consideravelmente maiores do que as do tipo Kaplan. Isso se deve ao

tamanho da amostra: Há muito mais turbinas Kaplan do que turbinas Bulbo sendo utilizadas pelas UHEs estudadas.

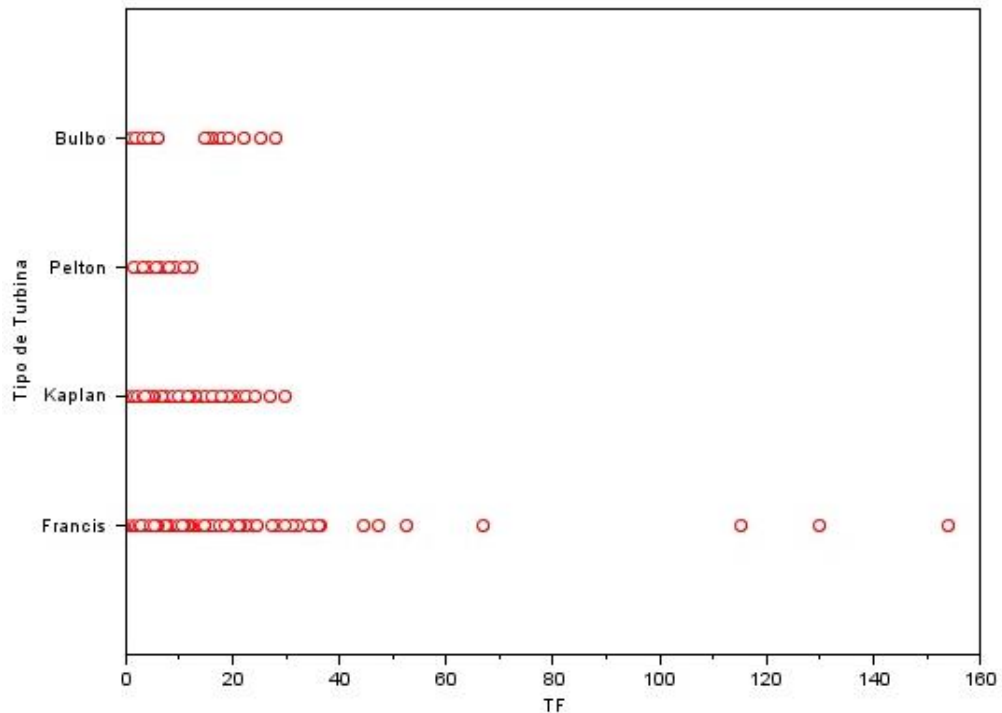


Figura 27. Relação entre parâmetro físico tipo de turbina e indicador taxa de falhas.

Tabela 14. Média e desvio padrão da taxa de falhas para tipos de turbina.

Tipo de Turbina	Média (ano <sup>-1</sup> )	Desvio Padrão (ano <sup>-1</sup> )
Francis	7,27	14,48
Kaplan	6,47	6,14
Pelton	6,26	2,94
Bulbo	10,22	9,31

### 7.1.7. Taxa de Falhas X Região

A taxa de falhas apresentou valores médios muito próximos e valores baixos e parecidos de desvio padrão, entre as regiões Nordeste, Sul e Sudeste. Apenas a região Norte-Centro Oeste mostrou uma alta dispersão dos elementos e valores muito altos que, conseqüentemente, elevaram a média da taxa de falhas nessa região (Figura 28).

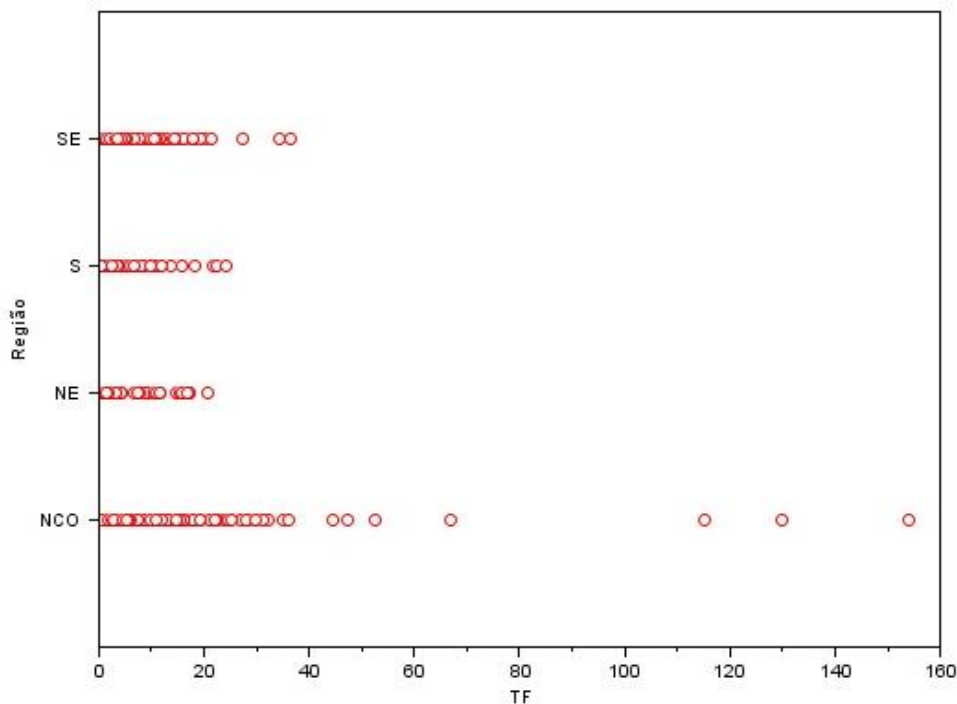


Figura 28. Relação entre parâmetro físico região e indicador taxa de falhas.

Tabela 15. Média e desvio padrão da taxa de falhas para regiões.

Região	Média (ano <sup>-1</sup> )	Desvio Padrão (ano <sup>-1</sup> )
Norte-Centro Oeste	15,42	22,00
Nordeste	5,42	5,72
Sul	4,73	5,08
Sudeste	4,61	4,99

### 7.1.8. Taxa de Falhas X Ano de Entrada em Operação

A taxa de falhas relaciona-se com a idade como apresentado na Figura 30 e na Tabela 16, onde se pode perceber que as UGs mais novas apresentam taxas de falhas mais elevadas. A explicação para esta situação é conhecida para empreendimentos hidrelétricos e é denominada a “Curva da Banheira”, apresentada na Figura 29, esta curva mostra de que forma as idades de um empreendimento hidrelétrico e, conseqüentemente, da unidade geradora influenciam a Taxa de Falhas.

As instalações e o desempenho das unidades geradoras podem ser analisados através do modelo da “Curva da Banheira”, que apresentam três fases distintas:



1ª Fase – Juventude (Mortalidade Infantil): causada geralmente por causas relacionadas a projeto e fabricação.

2ª Fase – Maturidade (Vida Útil): as causas que afetaram a 1ª fase foram eliminadas, assim as taxas de falhas tornam-se estáveis.

3ª Fase – Obsolescência (Envelhecimento): nesta fase os equipamentos começarão a falhar devido a problemas relacionados ao limite de sua vida útil, apresentando como causa fundamental a fadiga e o envelhecimento (GUIMARÃES, 2007).

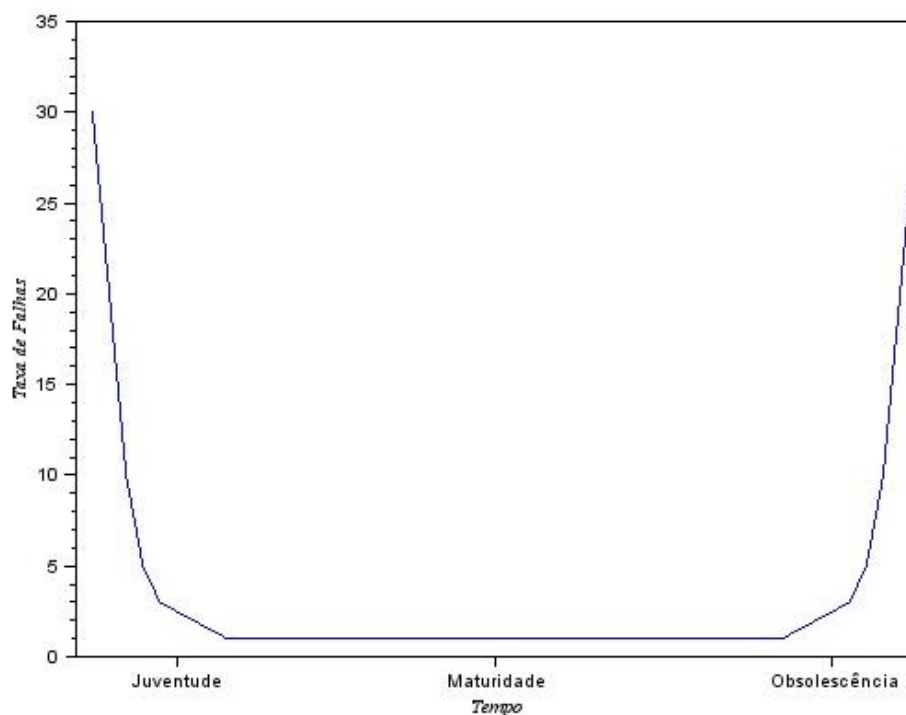


Figura 29. “Curva da Banheira”.

Guimarães, et al. (2006) ressaltou que manutenções preventivas e melhorias nos processos e projetos reduzem as ocorrências de falhas, sendo possível até mesmo eliminar a “Mortalidade Infantil”, 1ª fase, e como consequência, tem-se o aumento da vida útil das instalações.

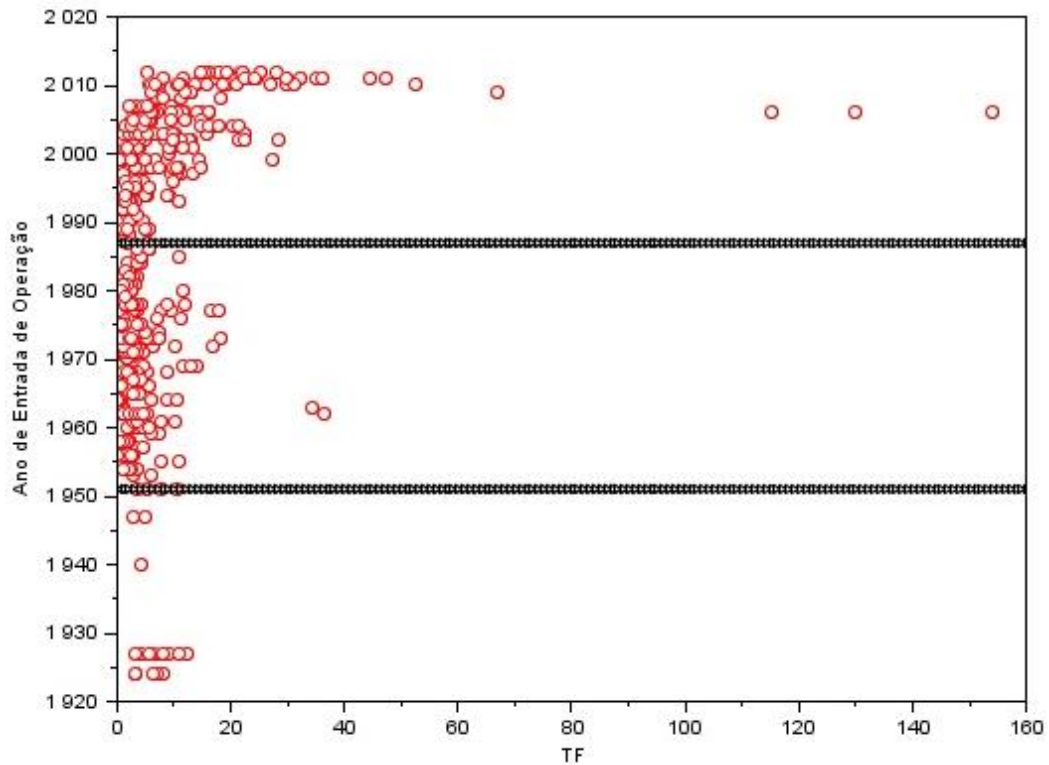


Figura 30. Relação entre parâmetro físico ano de entrada em operação e indicador taxa de falhas.

Tabela 16. Média e desvio padrão da taxa de falhas para faixas de idade.

Faixas de Idade	Média (ano <sup>-1</sup> )	Desvio Padrão (ano <sup>-1</sup> )
Mais antigas	6,53	2,52
Intermediárias	3,57	4,21
Mais novas	10,56	16,12

### 7.1.9. Tempo Médio de Reparo

O indicador de tempo médio de reparo apresenta valores muito altos e valores muito baixos, na ordem de décimos e milhares, isto aumenta a dispersão dos pontos e os valores de desvios padrão tornam-se muito elevados.

À medida que média aumenta, o desvio padrão também tende a aumentar, indicando que um ponto de valor muito elevado, como os da ordem de milhares, influencia muito na média e, por isso, torna-se inválida a análise por este tipo de métrica.

A análise visual também é prejudicada, visto que os valores elevados fazem com que os menores se aglomerem no início do gráfico e não permite visualizar

estes com detalhe. Assim como os pontos de altos valores se espalham por toda a área do gráfico pela alta dispersão e não são representativos para análise visual.

Além disto, os valores de correlação apresentados na Tabela 7 mostram que os parâmetros escolhidos neste trabalho não tiveram grande influência sobre o indicador TMR.

Da Figura 31 a Figura 34 é mostrada a relação dos parâmetros potência, turbina, região e idade com o tempo médio de reparo, respectivamente, bem como as tabelas com os valores de médias e desvios padrão de cada análise de correlação.

#### 7.1.9.1. Tempo Médio de Reparo x Potência

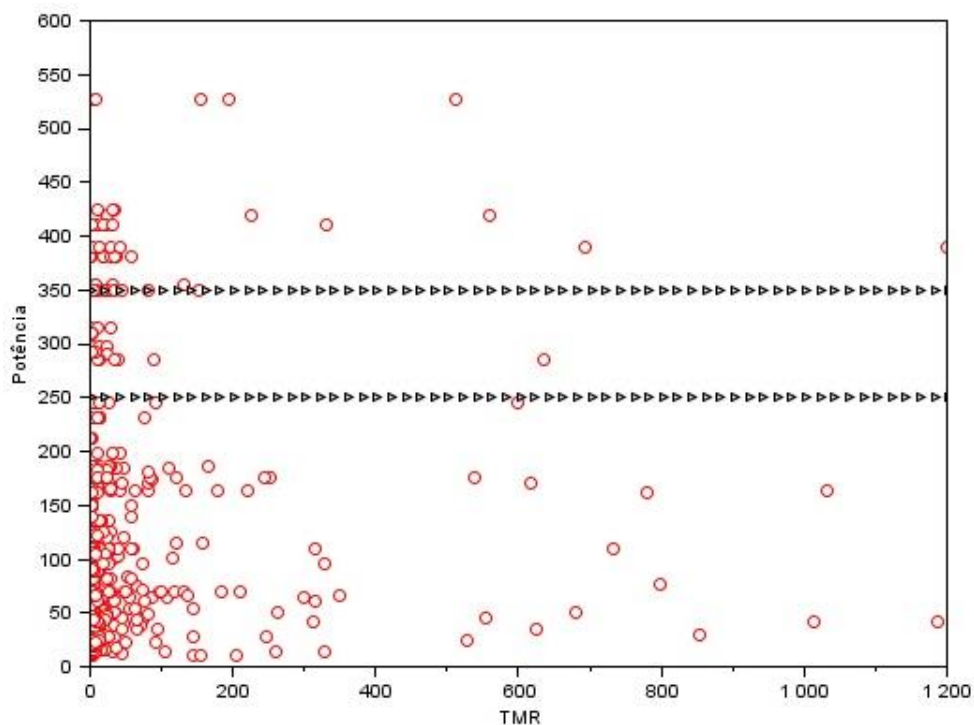


Figura 31. Relação entre parâmetro físico potência e indicador tempo médio de reparo.

Tabela 17. Média e desvio padrão do tempo médio de reparo para faixas de potência.

<b>Faixas de Potência</b>	<b>Média (horas)</b>	<b>Desvio Padrão (horas)</b>
Baixa potência	51,19	139,11
Média potência	36,89	102,80
Alta potência	106,80	229,18

## 7.1.9.2. Tempo Médio de Reparo x Tipo de Turbina

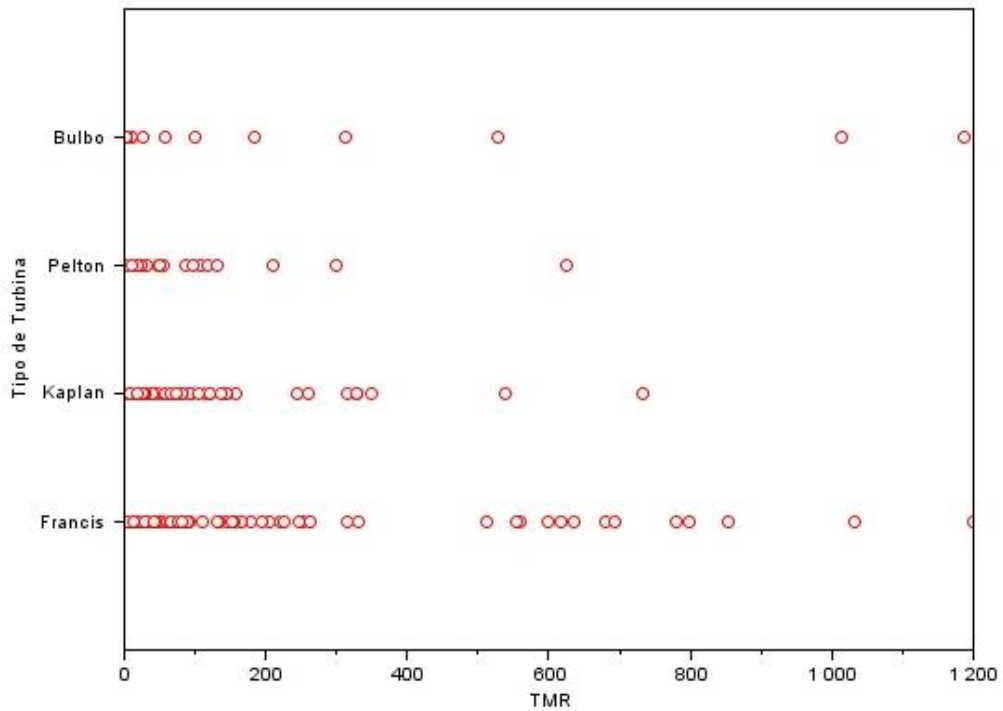


Figura 32. Relação entre parâmetro físico tipo de turbina e indicador tempo médio de reparo.

Tabela 18. Média e desvio padrão do tempo médio de reparo para tipos de turbina.

<b>Tipo de Turbina</b>	<b>Média (horas)</b>	<b>Desvio Padrão (horas)</b>
Francis	52,27	147,92
Kaplan	39,28	90,43
Pelton	109,46	149,80
Bulbo	173,35	344,64

## 7.1.9.3. Tempo Médio de Reparo x Região

Tabela 19. Média e desvio padrão do tempo médio de reparo para regiões.

<b>Região</b>	<b>Média (horas)</b>	<b>Desvio Padrão (horas)</b>
Norte-Centro Oeste	38,51	128,37
Nordeste	73,32	162,28
Sul	25,68	75,61
Sudeste	65,34	164,34

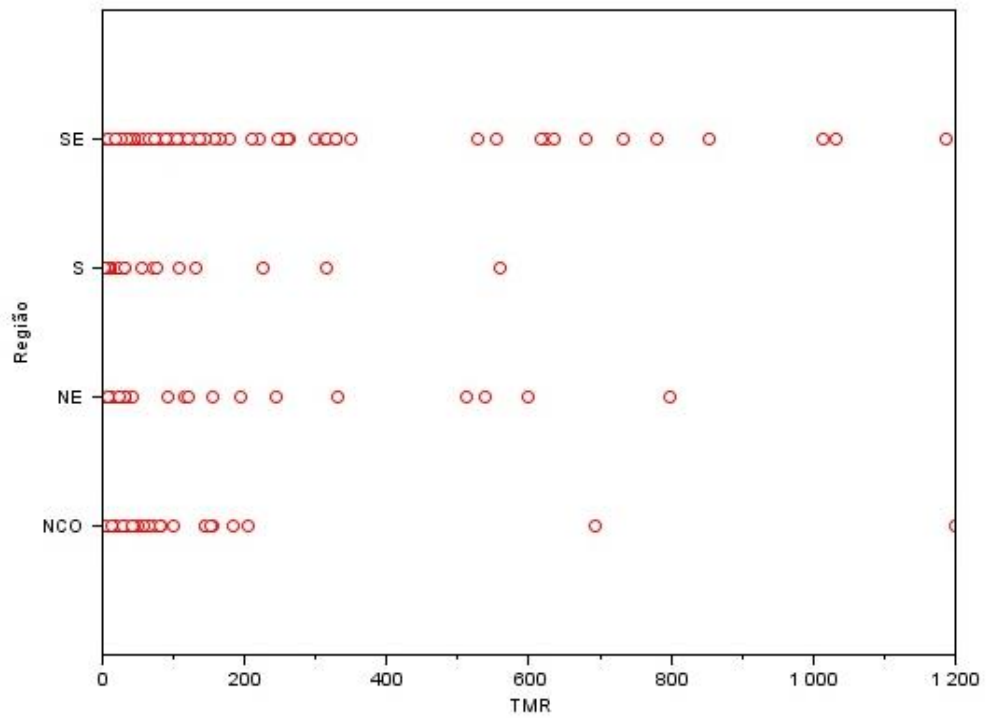


Figura 33. Relação entre parâmetro físico região e indicador tempo médio de reparo.

#### 7.1.9.4. Tempo Médio de Reparo x Ano de Entrada em Operação

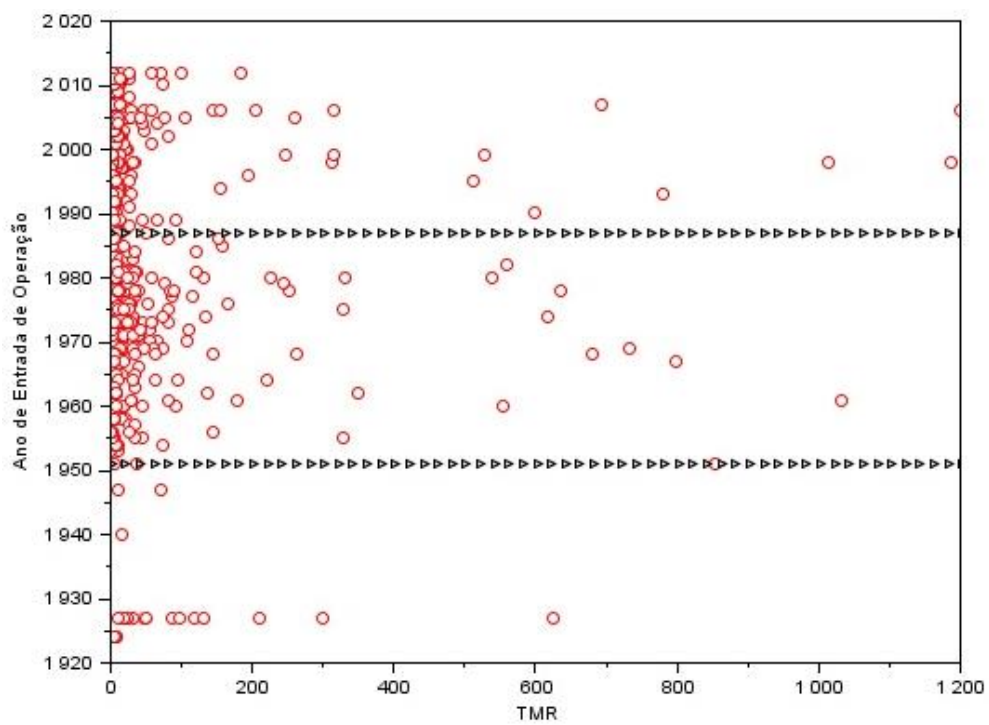


Figura 34. Relação entre parâmetro físico ano de entrada em operação e indicador tempo médio de reparo.

Tabela 20. Média e desvio padrão do tempo médio de reparo para faixas de idade.

Faixas de Idade	Média (horas)	Desvio Padrão (horas)
Mais antigas	107,91	201,57
Intermediárias	59,26	135,42
Mais novas	44,80	150,92

## 7.2. CORRELAÇÃO ENTRE PARÂMETROS FÍSICOS

Os resultados obtidos com a correlação podem também ser visualizados por meio de gráficos em que são relacionados os parâmetros físicos.

### 7.2.1. Potência X Ano de Entrada em Operação

Na Figura 35 é apresentado o gráfico que mostra a relação entre a potência e a data de entrada em operação das unidades geradoras, na figura estão interligados os grupos que mais se assemelham, formando quatro *clusters*, bem como as faixas de idade definidas naturalmente no Item 5.3.

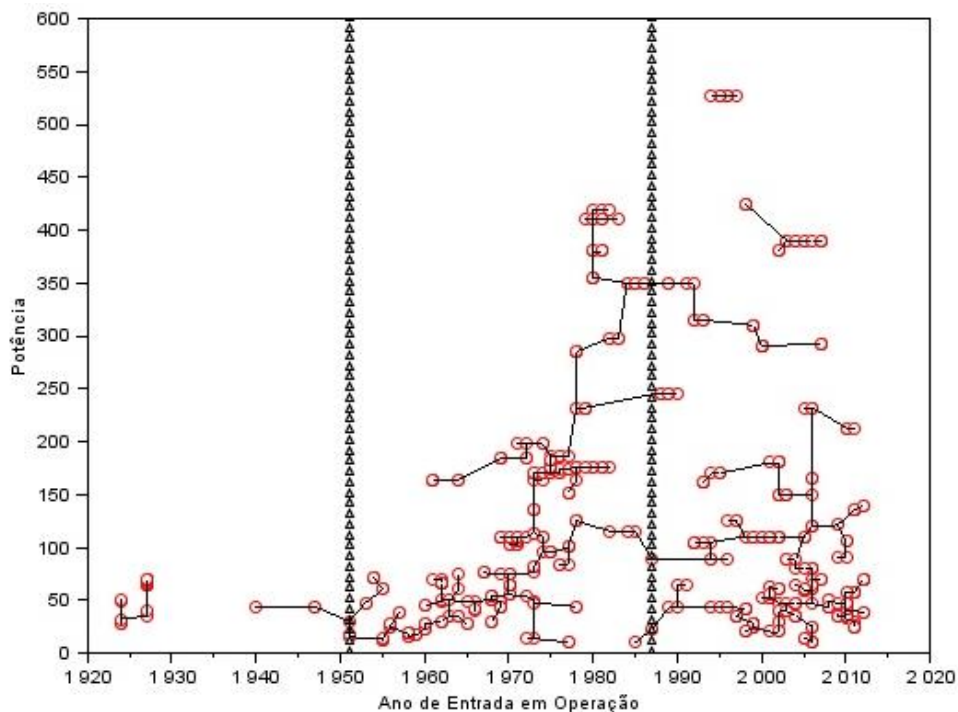


Figura 35. Gráfico de dispersão correlacionando a potência com o ano de entrada em operação com faixas de idade.

É possível observar que as potências das UGs aumentaram com o passar dos anos, isto se deve à crescente demanda de energia que levou à necessidade de explorar potenciais hidráulicos maiores e, conseqüentemente, aumentou o porte dos empreendimentos e equipamentos.

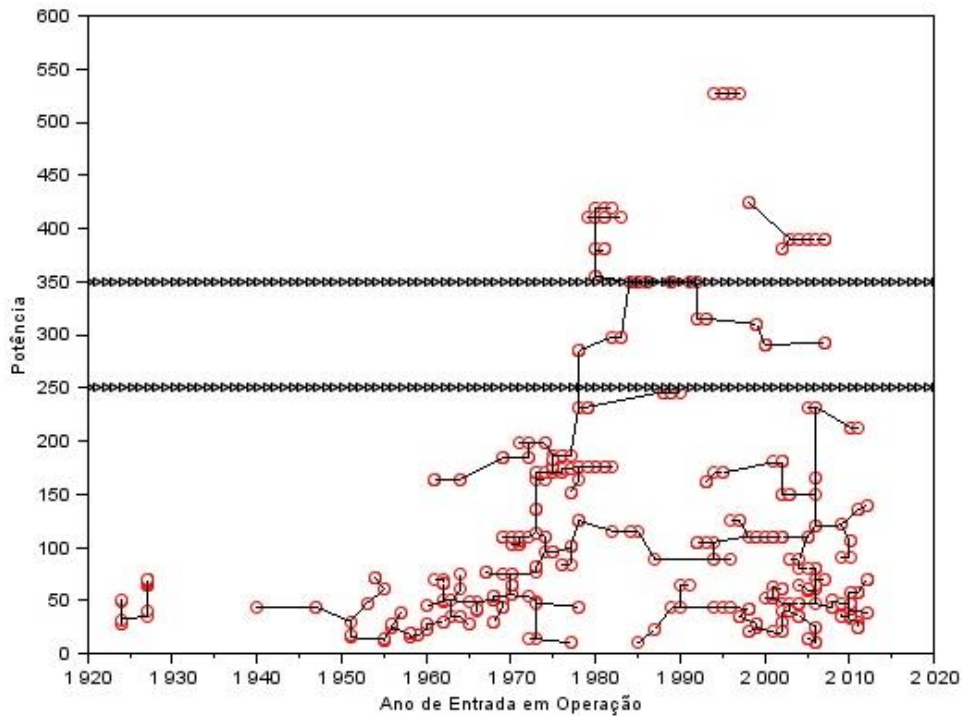


Figura 36. Gráfico de dispersão correlacionando a potência com o ano de entrada em operação com faixas de potência.

### 7.2.2. Potência X Região

Na Figura 37, a potência e a região das unidades geradoras são relacionadas e é possível perceber que, na região Sudeste não há unidades geradoras de alta potência, limitando-se às de baixa e média potência. Também pode-se constatar que as regiões Nordeste e Norte-Centro Oeste não possuem UGs com média potência, enquanto a região Sul tem UGs em todas as faixas de potências. Estas constatações podem estar relacionadas às próprias características do local que permitam ou não a utilização de equipamentos de maior porte.

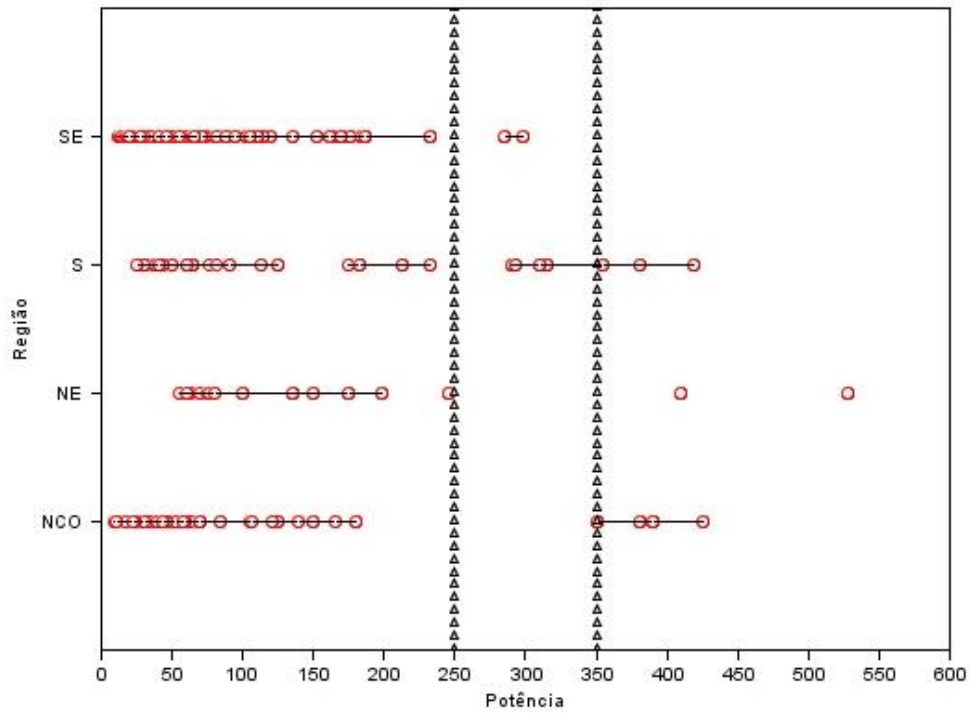


Figura 37. Gráfico de dispersão correlacionando a potência com a região.

### 7.2.3. Potência X Tipo de Turbina

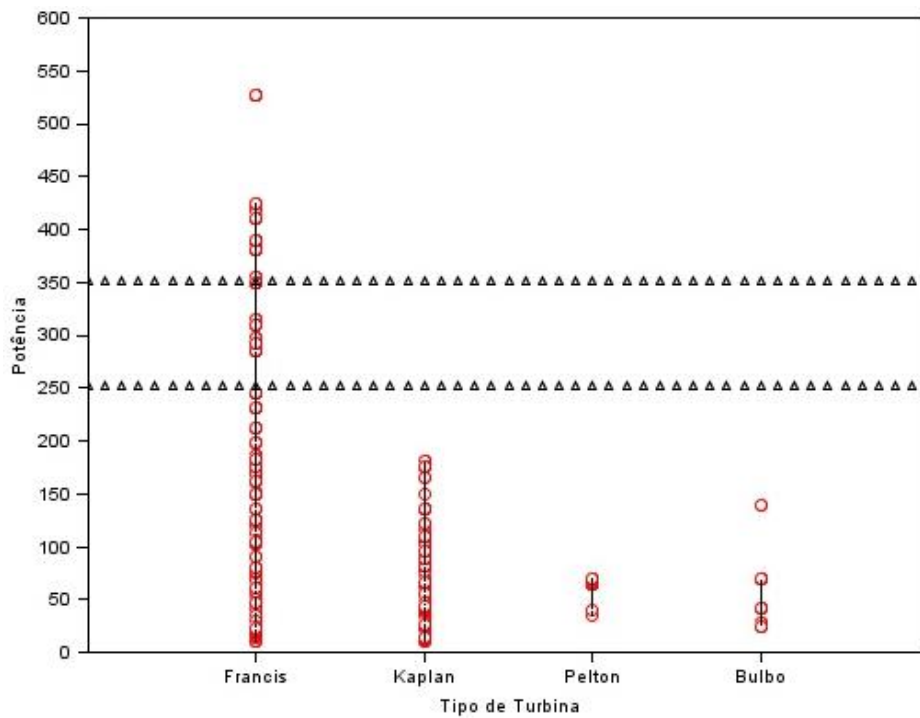


Figura 38. Gráfico de dispersão correlacionando a potência com o tipo de turbina.



Ao relacionar a potência com o tipo de turbina da unidade geradora é possível perceber que as turbinas de maior porte, são todas do tipo Francis, assim como as mais comuns, cerca de 60% de todas as unidades geradoras da amostra utiliza esta tecnologia hidráulica.

As turbinas Pelton e Bulbo aparecem em menor quantidade e também são as tecnologias das UGs de menores potências. Enquanto a turbina do tipo Kaplan, ainda que estejam todas na faixa de potências mais baixas, tem unidades geradoras que se aproximam de 200 MW.

Estas informações são mostradas visualmente no gráfico de dispersão da Figura 38, onde também é mostrada a divisão das unidades geradoras dos diferentes tipos de turbina nas faixas de potência.

#### **7.2.4. Tipo de Turbina X Ano de Entrada em Operação**

A turbina Francis e a turbina Pelton foram as tecnologias utilizadas nas primeiras unidades geradoras de Usinas Hidrelétricas do Tipo I no Brasil, a partir da década de 50 a turbina Kaplan passou a ser utilizada, sendo a sua variação, as turbinas do tipo Bulbo sendo utilizadas apenas ao fim da década de 90, como mostrado na Figura 39.

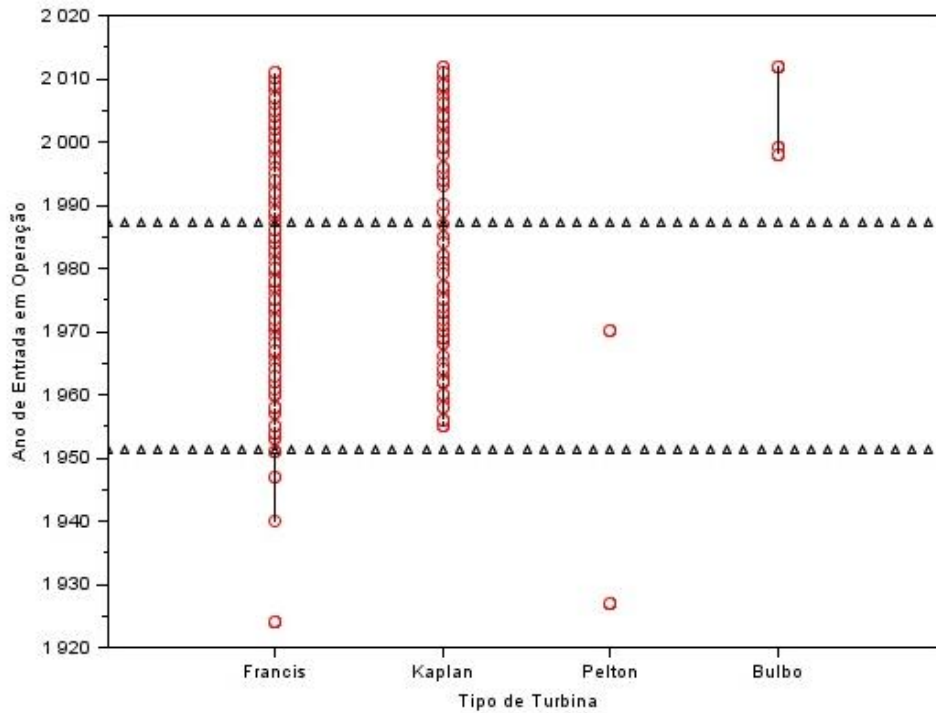


Figura 39. Gráfico de dispersão correlacionando o tipo de turbina com o ano de entrada em operação.

### 7.2.5. Região X Ano de Entrada em Operação

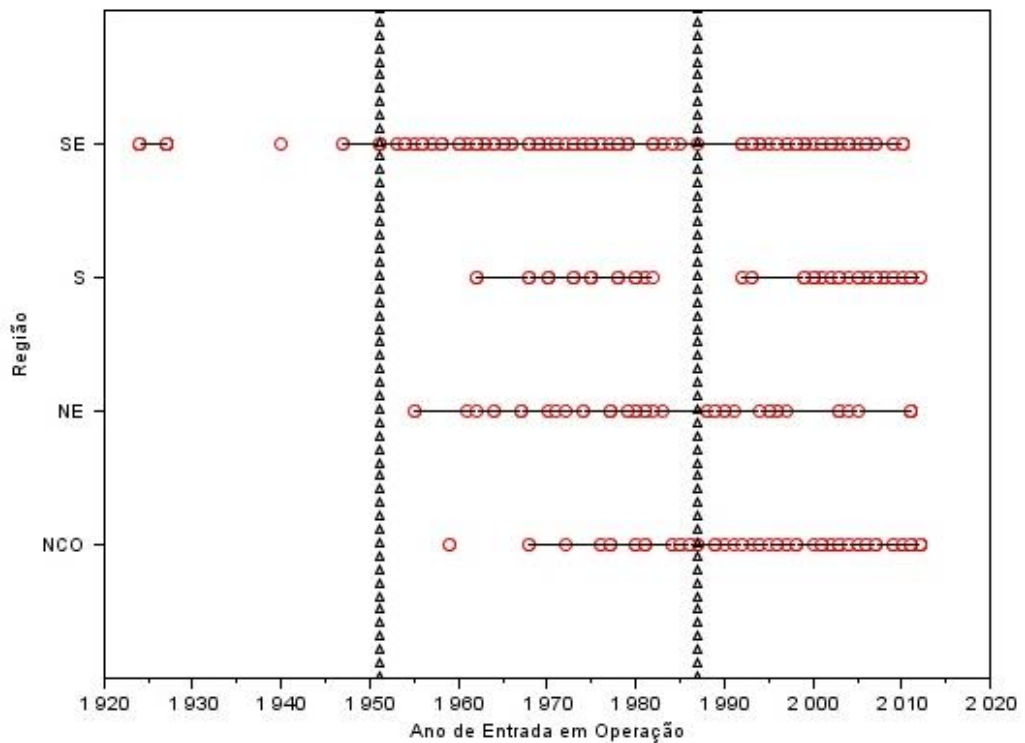


Figura 40. Gráfico de dispersão correlacionando a região com o ano de entrada em operação.

O gráfico apresentado na Figura 40 confirma o que foi apresentado na correlação pelo coeficiente de Pearson, em que a região Sudeste foi a primeira localização das unidades geradoras analisadas neste trabalho e, com o passar dos anos as outras regiões foram sendo empreendidas.

## 9. RESULTADOS E DISCUSSÃO

### 9.1. ANÁLISE DE *CLUSTERS*

A partir do programa de clusterização e do método de escolha do número de *clusters* descritos no item 6.2.1, foram obtidos os *clusters* relacionando todos os parâmetros físicos em análise neste trabalho (Figura 41), ou seja, a data de entrada em operação, a região em que se encontra a potência da UG e seu tipo de turbina.

#### Início Clustering

515.	6.
456.	9.
399.	190.
352.	4.
344.	11.
339.	117.
286.	40.
280.	6.
226.	32.
160.	14.
149.	4.
106.	8.
63.	11.
60.	53.
20.	44.

#### Fim Clustering

Figura 41. *Clusters* gerados no programa de Análise de *Clusters*.

### 9.2. DEFINIÇÃO DOS LIMITES MÍNIMOS

Encontrados os *clusters*, definidas as medianas de cada *cluster* e identificadas as UGs que estão abaixo e acima da mediana do seu agrupamento, torna-se necessário ir adiante com esses dados, de forma a estipular a melhora das usinas que não tem um desempenho satisfatório e manter as usinas que estão acima da mediana nesse nível ou melhores.

Como foi dito anteriormente, o método de regulação do desempenho das unidades geradoras será dado pelo método de *Yardstick Competition*, definindo a mediana como o limite para garantir o bom desempenho. Sendo assim, para cada

um dos 15 *clusters* gerados tem-se um valor de mediana, de cada indicador, que definirá o objetivo a ser alcançado em 5 anos para as unidades geradoras que estão abaixo deste valor.

### 9.2.1. Disponibilidade

Como exemplo para a disponibilidade, toma-se o *cluster* 399, com 190 unidades geradoras, a Tabela 21 apresenta os valores da mediana, máximo e mínimo deste grupo.

Tabela 21. Valor de mediana, máximo e mínimo da disponibilidade do *cluster* 399.

<b>Cluster 399</b>	
Mediana	96,50571
Máximo	99,97203
Mínimo	2,27359

O caso mais crítico, com menor disponibilidade, com o qual será feito o estudo de caso deste *cluster* é a 148ª UG que tem os seguintes dados históricos de disponibilidade:

Tabela 22. Dados de Disponibilidade para a UG 148 nos últimos 5 anos.

<b>Índice</b>	<b>2008</b>	<b>2009</b>	<b>2010</b>	<b>2011</b>	<b>2012</b>	<b>2013</b>	<b>Média (2009-2013)</b>
148	89,77	97,85	47,59	2,27	69,54	93,50	62,15

A projeção linear da disponibilidade média dos 5 anos antecedentes no período atual até chegar à mediana nos próximos 5 anos dar-se-á como na Tabela 23 e na Figura 42.

Tabela 23. Valores de Disponibilidade projetados para os próximos 5 anos e inclinação da reta de projeção.

	<b>Ano 1</b>	<b>Ano 2</b>	<b>Ano 3</b>	<b>Ano 4</b>	<b>Ano 5</b>	<b>Inclinação</b>
<b>Valor de DISP</b>	69,02	75,89	82,76	89,63	96,50	6,87

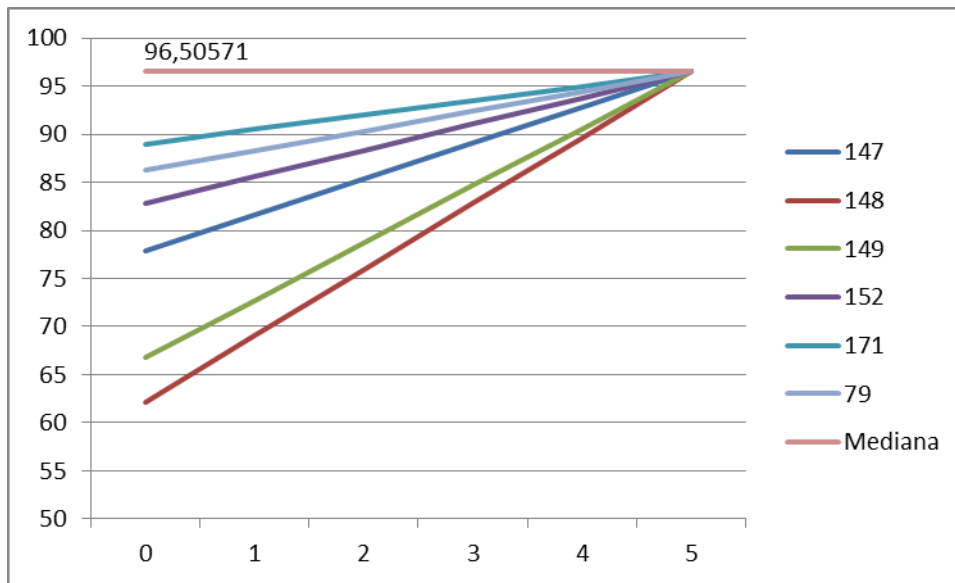


Figura 42. Projeção linear para definição de limites mínimos de disponibilidade para as UGs 147, 148, 149, 152, 171 e 79.

De forma a complementar este método de regulação, foi utilizada uma análise do histórico individual da UG, visando definir limites para que as UGs que já estão com bom desempenho não o ultrapassem caso piorem, ou ainda, para que unidades geradoras que tiveram anos incomuns, não os façam de referência e possam manter o seu melhor desempenho.

Para exemplificar este método, toma-se uma unidade geradora do *cluster* 399 que está com um bom desempenho, como a 1ª UG que está com uma média dos 5 anos antecedentes um pouco acima da mediana, como mostrado na Tabela 24.

Tabela 24. Dados de Disponibilidade para a UG 1 nos últimos 5 anos.

Índice	2008	2009	2010	2011	2012	2013	Média (2009-2013)
1	96,64	98,95	93,21	97,90	95,86	98,00	96,79

O segundo método consiste em utilizar o melhor e o pior ano da amostra para definir o limite para o próximo ano e para o 5º ano e, ao fazer a projeção linear, definir os objetivos dos anos intermediários. Na Tabela 25 estão os valores obtidos com a projeção linear do segundo método e os valores objetivados nos anos intermediários.

Tabela 25. Aplicação do método de regulação de desempenho para não piorar na UG 1, que apresenta bom desempenho.

<b>Valor Mínimo</b>	<b>2014</b>	<b>2015</b>	<b>2016</b>	<b>2017</b>	<b>2018</b>	<b>Valor Máximo</b>
93,21	90,88	93,12	94,24	95,36	96,48	98,95

### 9.2.2. Taxa de Falhas

Para o exemplo da Taxa de Falhas, também será analisado o *cluster* 399, com 190 unidades geradoras, a Tabela 26 apresenta os valores da mediana, máximo e mínimo deste grupo.

Tabela 26. Valor de mediana, máximo e mínimo da taxa de falhas do *cluster* 399.

<b>Cluster 399</b>	
Mediana	2,71
Máximo	99,04
Mínimo	0

O caso mais crítico, com maior taxa de falhas, com o qual será feito o estudo de caso deste *cluster* é a 139ª UG que tem os seguintes dados históricos de taxa de falhas:

Tabela 27. Dados de Taxa de Falhas para a UG 139 nos últimos 5 anos.

<b>Índice</b>	<b>2008</b>	<b>2009</b>	<b>2010</b>	<b>2011</b>	<b>2012</b>	<b>2013</b>	<b>Média (2009-2013)</b>
139	14,60	5,04	99,02	95,64	1,83	3,33	40,97

A projeção linear da taxa de falhas média dos 5 anos antecedentes no período atual até chegar à mediana nos próximos 5 anos dar-se-á como na Tabela 23 e na Figura 43.

Tabela 28. Valores de Taxa de Falhas projetados para os próximos 5 anos e inclinação da reta de projeção.

	<b>Ano 1</b>	<b>Ano 2</b>	<b>Ano 3</b>	<b>Ano 4</b>	<b>Ano 5</b>	<b>Inclinação</b>
<b>Valor de TF</b>	33,32	25,67	18,01	10,36	2,71	-7,65

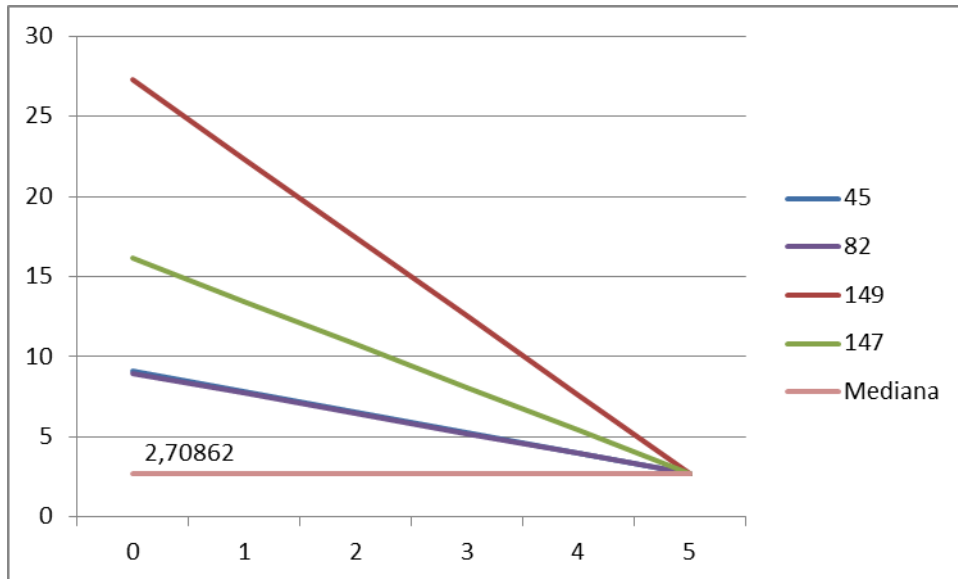


Figura 43. Projeção linear para definição de limites mínimos de taxa de falhas para as UGs 45, 82, 147 e 149.

Para exemplificar o segundo método de regulação, toma-se uma unidade geradora do *cluster* 399 que apresenta uma baixa taxa de falhas, como a 181ª UG que está com uma média dos 5 anos antecedentes menor que a mediana, como mostrado na Tabela 29.

Tabela 29. Dados de Taxa de Falhas para a UG 181 nos últimos 5 anos..

Índice	2008	2009	2010	2011	2012	2013	Média (2009-2013)
181	0	2,21	0	1,09	1,07	1,38	1,15

Na Tabela 30 estão os valores obtidos com a projeção linear do segundo método e os valores objetivados nos anos intermediários. Sabe-se que zero é um valor extremamente restritivo, por isso, como será visto na próxima Seção, nesta situação, o método de limite da mediana do *cluster* deve ser usado para definir os limites mínimos.

Tabela 30. Aplicação do método de regulação de desempenho para não piorar na UG 181, que apresenta baixa taxa de falhas.

Valor Máximo	2014	2015	2016	2017	2018	Valor Mínimo
2,21	2,26	1,36	0,90	0,45	0	0



### 9.2.3. Tempo Médio de Reparo

Como exemplo para o Tempo Médio de Reparo também será analisado o *cluster* 399, com 190 unidades geradoras, a Tabela 31 apresenta os valores da mediana, máximo e mínimo deste grupo.

Tabela 31. Valor de mediana, máximo e mínimo do tempo médio de reparo do *cluster* 399.

<b>Cluster 399</b>	
Mediana	3,97
Máximo	4955,85
Mínimo	0

O caso mais crítico, com maior tempo médio de reparo, com o qual será feito o estudo de caso deste *cluster* é a 43ª UG que tem os seguintes dados históricos de tempo médio de reparo:

Tabela 32. Dados de Tempo Médio de Reparo para a UG 43 nos últimos 5 anos.

Índice	2008	2009	2010	2011	2012	2013	Média (2009-2013)
43	70,95	207,23	307,75	117,88	4267,58	1209,94	1222,08

A projeção linear do tempo médio de reparo dos 5 anos antecedentes no período atual até chegar à mediana nos próximos 5 anos dar-se-á como na Tabela 33 e na Figura 44.

Tabela 33. Valores de Tempo Médio de Reparo projetados para os próximos 5 anos e inclinação da reta de projeção.

	Ano 1	Ano 2	Ano 3	Ano 4	Ano 5	Inclinação
Valor de TMR	978,46	734,84	491,21	247,59	3,97	-243,62

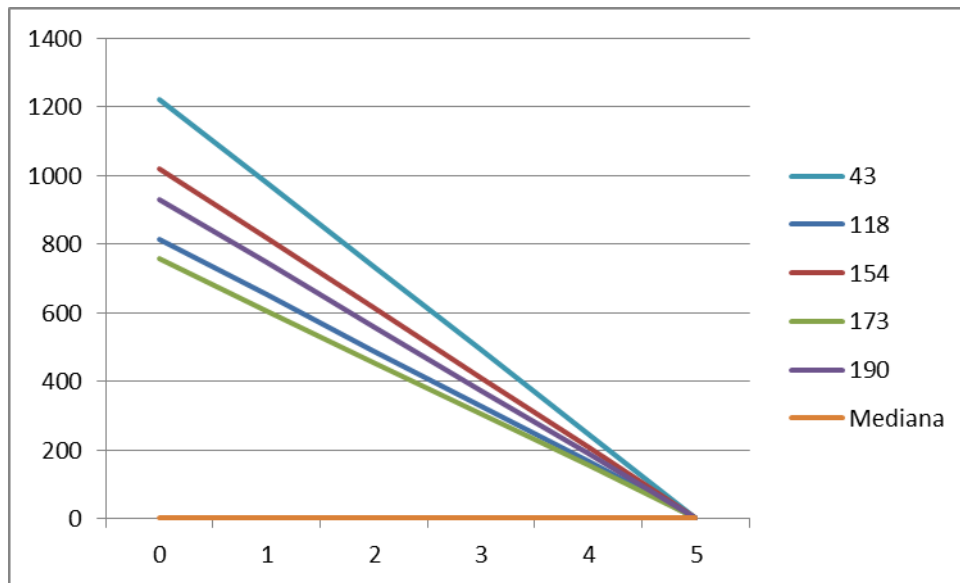


Figura 44. Projeção linear para definição de limites mínimos de disponibilidade para as UGs 147, 148, 149, 152, 171 e 79.

Para exemplificar o método de não piorar, toma-se uma unidade geradora do *cluster* 399 que está com um bom desempenho, como a 9ª UG que está com uma média dos 5 anos antecedentes um pouco abaixo da mediana, como mostrado na Tabela 34.

Tabela 34. Dados de Tempo Médio de Reparo para a UG 9 nos últimos 5 anos.

Índice	2008	2009	2010	2011	2012	2013	Média (2009-2013)
9	1,64	1,22	0,89	3,50	4,73	7,14	3,50

Na Tabela 35 estão os valores obtidos com a projeção linear do método de não piorar e os valores objetivados nos anos intermediários.

Tabela 35. Aplicação do método de regulação de desempenho para não piorar na UG 9, que apresenta baixo tempo médio de reparo.

Valor Mínimo	2014	2015	2016	2017	2018	Valor Máximo
7,14	7,32	4,76	3,48	2,20	0,91	0,89

### 9.3. AVALIAÇÃO DAS UNIDADES GERADORAS

A avaliação do desempenho das unidades geradoras é feita com base em dois métodos propostos neste trabalho: o *Yardstick* e o Método para Regulação de

Desempenho para Não Piorar. Ambos os métodos apresentam valores próximos para um limite mínimo em que deve estar o valor de Disponibilidade e limites máximos para os valores de Tempo Médio de Reparo e Taxa de falhas.

Comparando os valores próximos que são gerados a partir da aplicação dos dois métodos, faz-se necessário escolher um deles como limite para aplicar à UG e definir se seu desempenho está aceitável ou não. Um deles será mais rigoroso e o outro será menos. Cabe ao órgão fiscalizador, no caso a ANEEL, julgar qual lhe interessa mais do ponto de vista fiscalizatório.

## 11. CONSIDERAÇÕES FINAIS

O tema deste trabalho está inserido dentro do contexto de planejamento e gestão energética, visto que permite, a partir de seus dados, analisar o desempenho da principal tecnologia fonte de energia elétrica do país: as unidades geradoras de usinas hidrelétricas. Diante de tamanha importância das usinas hidrelétricas, faz-se necessário fiscalizá-las e, para isso, é necessário ter um parâmetro passível de comparação entre uma usina e outra, que permita avaliar seu desempenho. Esse parâmetro pode ser os limites dos indicadores de desempenho redefinidos neste trabalho.

O resultado do trabalho pode ser considerado positivo do ponto de vista da Superintendência de Fiscalização dos Serviços de Geração da ANEEL, pois proporciona a eles limites para os indicadores de desempenho das unidades geradoras de usinas hidrelétricas com um respaldo metodológico. Além disso, a pesquisa bibliográfica apresentada ao longo deste documento permite que a superintendência altere esses limites de acordo com suas necessidades.

Este trabalho poderá embasar trabalhos futuros com tema semelhante, tanto a partir da bibliografia utilizada, quanto a partir da metodologia proposta e dos resultados alcançados. A partir dos resultados, é possível, então, fazer mudanças no processo de regulamentação com vista a otimizar a forma como se fiscaliza a geração de energia elétrica.

## 12. SUGESTÃO DE CONTINUIDADE

### 12.1. ANÁLISE POR USINA

Este trabalho usou como unidade de análise as UGs das usinas hidrelétricas, porém é possível extrapolar esse estudo para a própria usina, a partir dos dados das suas UGs. Todos os indicadores seriam recalculados para a usina em questão, ao somar o valor de cada um de cada UG e dividir pelo número de unidade geradora, ou seja, o indicador da usina seria a média do indicador de suas UGs. Então haveria o dado de Disponibilidade, Taxa de falhas e Tempo Médio de Reparo para cada empreendimento, não só para cada UG. A partir disso, é possível fazer uma regulamentação com base nos indicadores por usina hidrelétrica.

No mercado de geração de energia elétrica, é comum um mesmo agente de geração ter mais de empreendimento hidrelétrico. Portanto, é interessante também que haja uma análise dos indicadores de qualidade feita voltada para cada agente, no intuito de proporcionar a ele uma avaliação externa, a partir de dados internos, de como está sendo a gestão de sua empresa. Esses indicadores podem fazer com que o agente reflita sobre como está sendo a eficiência da manutenção de suas usinas, bem como o tempo que elas levam para ser reparadas e assim possa decidir se os procedimentos escolhidos são o melhor para atingir os objetivos de uso de seus empreendimentos.

### 12.2. OUTROS PARÂMETROS

Para a realização do trabalho, foram escolhidos para caracterizar as usinas hidrelétricas, tanto parâmetros de construção, como o tipo de tecnologia da turbina e sua potência, quanto parâmetros operacionais, como a região em que se localiza e data de entrada em operação da usina. Porém as características passíveis de influenciar os indicadores de desempenho de uma usina não se restringem às escolhidas neste estudo nem somente à características operacionais e construtivas.

Diante de tantos fatores que influenciam no desempenho de uma unidade geradora, é possível ainda identificar características das usinas ou das UGs que influenciem nos indicadores de qualidade temas de estudo. A partir da inserção de

outros parâmetros da usina, o processo de clusterização seria enriquecido e o de fiscalização, aprimorado.

### 12.3. TÉCNICAS HIERÁRQUICAS AGLOMERATIVAS E MEDIDAS DE SIMILARIDADE

Outra possibilidade de trabalho futuro a ser realizado a partir deste é a análise de outras técnicas hierárquicas aglomerativas, como citadas no item 5.4.2.1. A técnica escolhida para realizar este trabalho foi o método de ligação simples, porém há também o de ligação completa, o das médias das distâncias e o do centróide. A forma como essas outras metodologias lidam com os dados pode modificar a visualização dos *clusters* e, portanto, sua compreensão.

Com o mesmo intuito de visualizar de forma diferente os dados a partir das técnicas hierárquicas, é possível também trabalhar com outras medidas de similaridade, como citadas no item 5.2.4.

### 12.4. ANÁLISE PARA OUTROS EMPREENDIMENTOS DE GERAÇÃO DE ENERGIA

Com base na metodologia mostrada neste trabalho, é possível abranger também outras formas de geração de energia elétrica, como, por exemplo, UHEs Tipo 2 e 3 PCHs, usinas termelétricas, usinas eólicas e empreendimentos de geração solar, pois todos são passíveis de fiscalização e podem melhorar seu desempenho, caso esse seja insatisfatório.

### 12.5. REALIZAÇÃO DE CONSULTA PÚBLICA PARA CONTRIBUIÇÃO DOS AGENTES DE GERAÇÃO

Como continuação do trabalho, é interessante realizar uma Consulta Pública, a fim de coletar opiniões dos agentes sobre outras possíveis metodologias utilizadas tanto para a clusterização e quanto para a redefinição dos valores limites dos indicadores trabalhados. Além disso, a Consulta Pública serviria para os agentes opinarem quanto aos parâmetros das usinas hidrelétricas e das unidades geradoras que podem contribuir para a análise realizada neste trabalho. Diante da questão de

escolha dos percentis para a uma melhor clusterização no método para avaliação de desempenho para não piorar, é interessante ter também o ponto de vista do agente, visto que é ele quem administra as UGs.

## BIBLIOGRAFIA

- ANEEL. **Atlas de Energia Elétrica do Brasil**. Agência Nacional de Energia Elétrica. [S.l.]. 2008. Parte II, Cap.3.
- ANEEL. Agência Nacional de Energia Elétrica. **Capacidade de Geração do Brasil**, 2014. Disponível em:  
<<http://www.aneel.gov.br/aplicacoes/capacidadebrasil/capacidadebrasil.cfm>>. Acesso em: 05 jun. 2014.
- ANEEL. Agência Nacional de Energia Elétrica. **Resolução Normativa Nº 67**, de 08 de junho de 2004.
- BASTISTA, B. D. D. O. **DISTRIBUIÇÃO EXATA DA MIDRANGE ESTUDENTIZADA EXTERNAMENTE DA NORMAL E DESENVOLVIMENTO DE UMA BIBLIOTECA R UTILIZANDO QUADRATURA GAUSSIANA**. Universidade Federal de Lavras. Lavras, p. 13. 2013.
- DI CARLANTONIO, L. M. **Novas Metodologias para Clusterização de Dados**. Universidade Federal do Rio de Janeiro (UFRJ). Rio de Janeiro. 2001.
- ELETROBRÁS. **Manual de Microcentrais Hidrelétricas**. [S.l.]: Ministério das Minas e Energia, 1985.
- FERNANDES, C.; COSTA, R.; SILVA, P. **Elaboração de indicadores referentes aos processos de fiscalização do parque gerador elétrico brasileiro**. Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio). Rio de Janeiro. 2013.
- FONSECA, F. C. S.; BELTRAME, W. A. R. **Aplicações Práticas dos Algoritmos de Clusterização K-means e Bisecting K-means**. Universidade Federal do Espírito Santo (UFES). Vitória - ES.
- GADELHA, D. **Scilab 5.X**. Universidade Federal do Ceará. [S.l.].
- GUIMARÃES, M. B. **Reflexos do Planejamento da Manutenção na Produção e Rentabilidade em Usinas de Geração de Energia Elétrica**. Universidade Tecnológica Federal do Paraná. Ponta Grossa. 2007.
- GUIMARÃES, M. B. et al. Programações de Manutenções Preventivas em Unidades Geradoras: os custos, benefícios e penalidades conforme a Nova Regulamentação do Setor Elétrico Brasileiro. **VII Seminário Paranaense de Manutenção - ABRAMAN**, Curitiba, 2006.
- HADDAD, J. et al. **Indicadores Nacionais e Internacionais**. Escola Federal de Engenharia de Itajubá. Itajubá. 2000.
- HAIR, J. F. et al. **Análise Multivariada de Dados**. 5ª. ed. Porto Alegre: Bookman, 2005.



- JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. Data clustering: a review. **Journal ACM Computing Surveys (CSUR)**, New York, EUA, v. 31, n. 3, p. 263-323, Setembro 1999.
- KATO, S. Teste de Hipótese. **PUC-FAMAT-Dpto. de Estatística**. Disponível em: [http://www.pucrs.br/famat/sergio/Estatistica\\_Basica\\_T126/Teste\\_de\\_hipotese.pdf](http://www.pucrs.br/famat/sergio/Estatistica_Basica_T126/Teste_de_hipotese.pdf). Acesso em: 27 Outubro 2013.
- KOÇOUSKI, A. R. **Desempenho dos Serviços de Geração de Energia Elétrica Brasileira no Sistema Isolado: reflexões e proposições metodológicas com enfoque estratégico**. Fundação Getúlio Vargas - FGV. Brasília. 2011.
- LINDEN, R. Técnicas de Agrupamento. **Revista de Sistemas de Informação da FSMA**, Visconde de Araújo, v. 4, p. 18-36, 2009.
- MINGOTI, S. A. **ANÁLISE DE DADOS ATRAVÉS DE MÉTODOS DE ESTATÍSTICA MULTIVARIADA: UMA ABORDAGEM PRÁTICA**. 1ª. ed. Belo Horizonte: UFMG, 2013.
- MOSCATO; VON ZUBEN. Uma Visão Geral de Clusterização de Dados. **DCA/FEEC/Unicamp**. Disponível em: [ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/ia368\\_02/topico5\\_02.pdf](ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/ia368_02/topico5_02.pdf). Acesso em: 27 out. 2013.
- ONS. **Procedimentos de Rede**. Operador Nacional do Sistema Elétrico. [S.l.], p. Submódulos 20.1, 26.1, 25.8. 2011.
- ONS. **Operador Nacional do Sistema Elétrico**, 2013. Disponível em: <http://www.ons.org.br/>. Acesso em: 31 out. 2013.
- ONS. Operador Nacional do Sistema Elétrico. **O que é o SIN - Sistema Interligado Nacional**, 2013. Disponível em: [http://www.ons.org.br/conheca\\_sistema/o\\_que\\_e\\_sin.aspx](http://www.ons.org.br/conheca_sistema/o_que_e_sin.aspx). Acesso em: 06 de Novembro de 2013.
- PRATI, R. C. Data Preparation. **Data Mining**. Disponível em: <http://professor.ufabc.edu.br/~ronaldo.prati/DataMining/DataPreparation.pdf>. Acesso em: 15 maio 2014.
- SNEATH, P. H. A. The application of computer to taxonomy. **Journal of General Microbiology**, p. 201 - 226, 1957.
- SÓ MATEMÁTICA. Estatística Básica - Moda, Mediana e Comparação entre Mediana e Média. **Só Matemática**. Disponível em: <http://www.somatematica.com.br/estat/basica/pagina6.php>. Acesso em: 03 Julho 2014.
- SRC. **Resolução nº 574 de 20 de agosto de 2013**. Agência Nacional de Energia Elétrica. Brasília. 2013.

**SRD. Resultados da segunda parte da Audiência Pública nº 046/2010 relativa ao aprimoramento da metodologia utilizada pela ANEEL para o estabelecimento dos limites dos indicadores de continuidade coletivos e a criação do indicador de desempenho global de contin.** Agência Nacional de Energia Elétrica. Brasília. 2011.

**TANURE, J. E. P. S. ANÁLISE COMPARATIVA DE EMPRESAS DE DISTRIBUIÇÃO PARA O ESTABELECIMENTO DE METAS DE DESEMPENHO PARA INDICADORES DE CONTINUIDADE DO SERVIÇO DE DISTRIBUIÇÃO.** Itajubá: ESCOLA FEDERAL DE ENGENHARIA DE ITAJUBÁ, 2000.

**WEBER, T. S. Tolerância a falhas: conceitos e exemplos.** UFRGS. [S.l.]. 2003.