

**Aplicação de Técnicas de Agrupamento Sobre Séries de Potência Para Avaliação
de Confiabilidade em Sistemas Hidrotérmicos**

Marcia Endler Kranen

Universidade Federal do Rio de Janeiro
Curso de Mestrado

Adriano Joaquim de Oliveira Cruz
Ph.D.

Albert Cordeiro Geber de Melo
D.Sc.

Rio de Janeiro
2004

**Aplicação de Técnicas de Agrupamento Sobre Séries de Potência Para Avaliação
de Confiabilidade em Sistemas Hidrotérmicos**

Marcia Endler Kranen

Dissertação submetida ao corpo docente do IM/DCC E NCE/UFRJ Universidade Federal do Rio de Janeiro - UFRJ, como parte dos requisitos necessários à obtenção do grau de Mestre.

Aprovada por:

Prof. _____ - Orientador
Adriano Joaquim de Oliveira Cruz – Ph.D.

Prof. _____ - Co-orientador
Albert Cordeiro Geber de Melo – D.Sc.

Prof. _____
Antônio Carlos Gay Thomé – Ph.D.

Prof. _____
Maria Elvira Piñeiro Maceira – D.Sc.

Rio de Janeiro
2004

K49 Kranen, Márcia Endler.

Aplicação de técnicas de agrupamento sobre séries de Potência para avaliação de confiabilidade em sistemas hidrotérmicos / Márcia Endler Kranen. – Rio de Janeiro, 2004.

161f.; il.

Dissertação (Mestrado em Informática) – Universidade Federal do Rio de Janeiro, Instituto de Matemática, Núcleo de Computação Eletrônica, 2005.

Orientador: Adriano de Oliveira Cruz

1. Análise de Dados – Teses. 2. Análise de Confiabilidade Composta - Teses. 3. Validação de Agrupamentos – Teses. I. Adriano de Oliveira Cruz (Orient.). II. Universidade Federal do Rio de Janeiro. Instituto de Matemática. Núcleo de Computação Eletrônica. III. Título.

CDD

*A minha princesinha Andressa,
Por ser uma criança maravilhosa.*

*Ao Pablo,
Por todo o amor, dedicação e paciência.*

*Aos meus pais, Dóris e Harri,
Por terem sempre me apoiado.*

Agradecimentos

Ao orientador e professor Adriano Joaquim de Oliveira Cruz pela excelente orientação na aplicação das técnicas nebulosas, extensa dedicação e conhecimentos transmitidos durante esta dissertação.

Ao co-orientador Albert Cordeiro Geber de Melo, pelos valiosos ensinamentos sobre o sistema elétrico brasileiro e dos modelos utilizados neste trabalho.

Aos amigos do CEPTEL, em particular, José Francisco Moreira Pessanha, meu primeiro professor no mundo das técnicas de agrupamento, cujo incentivo e conhecimento na área foram fundamentais. À Débora Jardim, Raquel Marcato e Vitor Duarte, pelo apoio junto aos modelos energéticos. E a tantos outros que sempre estiveram por perto nos momentos de dúvidas e incertezas.

Aos colegas do Serpro, que na reta final me estenderam a mão de forma a ajudar que este trabalho fosse finalizado, em especial a Myrian Patrícia, pela compreensão em relação a minha demora em apresentar resultados.

Ao IM-NCE e ao Cepel sem os quais este trabalho não existiria.

Ao Pablo, por toda a compreensão, dedicação, incentivo, apoio, discussões, enfim por estar sempre ao meu lado, tanto nos momentos de vitória quanto nos momentos de puro cansaço. Principalmente, por disponibilizar todo o seu conhecimento em engenharia como apoio em minha busca por mais e mais alternativas de solução.

À minha filha Andressa, uma criança maravilhosa, que brigou muito por atenção, mas que do alto de sua sabedoria de criança também soube se interessar pelo trabalho que estava sendo feito, indicar pequenas correções neste texto e dizer “que desenho bonito!” e “o NH2 já terminou de funcionar”.

Aos meus pais, Dóris e Harri, pelo incentivo, apoio e compreensão nesta e em todas as etapas de minha vida.

Aos amigos, profissionais das mais diversas áreas, que mesmo sem entender o que estava sendo dito, por muitas vezes tiveram a paciência de me escutar em momentos que perceberem que eu só precisava falar. Em especial a Ana Carolina companheira de mestrado que passando pelas mesmas aflições sempre teve um sorriso e um momento de atenção.

A Deus que durante todo o meu caminho sempre foi um porto seguro.

Resumo

Kranen, Marcia Endler. **Aplicação de Técnicas de Agrupamento Sobre Séries de Potência Para Avaliação de Confiabilidade em Sistemas Hidrotérmicos**. Orientador: Adriano de Oliveira Cruz. Rio de Janeiro: UFRJ/IM-NCE, 2004. Dissertação (Mestrado em Informática).

O objetivo deste trabalho é propor uma metodologia para reduzir o esforço computacional na avaliação de confiabilidade, considerando a variação das potências disponíveis nas usinas hidrelétricas em função da aleatoriedade das vazões afluentes aos reservatórios. A metodologia proposta consiste na utilização de técnicas de agrupamento, baseadas em lógica nebulosa, aplicadas às séries de potência disponível obtidas a partir de estudos hidrotérmicos de operação do sistema elétrico brasileiro. A meta é obter o menor número de séries que reflita o comportamento randômico da hidrologia e que preserve a dependência espacial e temporal dos dados originais. A análise de confiabilidade baseada neste conjunto reduzido de cenários hidrológicos deve manter a precisão na estimativa dos índices de confiabilidade obtidos quando os estudos são realizados a partir dos dados originais.

Abstract

Kranen, Marcia Endler. **Aplicação de Técnicas de Agrupamento Sobre Séries de Potência Para Avaliação de Confiabilidade em Sistemas Hidrotérmicos**. Orientador: Adriano de Oliveira Cruz. Rio de Janeiro: UFRJ/IM-NCE, 2004. Dissertação (Mestrado em Informática).

The objective of this work is to propose an approach to reduce the computational effort of the system reliability studies, when evaluating energy supply availability. The proposed approach consists on the application of clustering techniques to the generating output capacity sequences, derived from the hydro scenarios obtained in the hydrothermal coordination studies of the Brazilian generating system. The goal is to obtain an equivalent model of the random hydrological behavior with a smaller number of generating output capacity sequences that still reflect the temporal and spatial correlation of the original data. The reliability analysis based on this reduced number of hydrological scenarios results in a smaller computational effort while keeping the accuracy on the estimation of the reliability indices.

Lista de Figuras

| | |
|---|----|
| FIGURA 1.1 – REPRESENTAÇÃO SIMPLIFICADA DE UM SISTEMA DE ENERGIA ELÉTRICA | 1 |
| FIGURA 1.2 - FLUXO DE INFORMAÇÃO ENTRE OS MODELOS. | 4 |
| FIGURA 1.3 - SÉRIE DE POTÊNCIA DISPONÍVEL DA USINA HIDRELÉTRICA DE CACONDE..... | 4 |
| FIGURA 1.4 - FLUXO DE INFORMAÇÃO ENTRE OS MODELOS COM PROCESSO DE AGRUPAMENTO. | 6 |
| FIGURA 2.1 – REPRESENTAÇÃO DOS SUBSISTEMAS DO SIN | 10 |
| FIGURA 2.2 – CADEIA DE MODELOS PARA O PLANEJAMENTO ENERGÉTICO DO SISTEMA BRASILEIRO | 12 |
| FIGURA 2.3 – NÍVEIS HIERÁRQUICOS PARA A ANÁLISE DE CONFIABILIDADE..... | 18 |
| FIGURA 2.4 – MODELO DE MARKOV A 2 ESTADOS..... | 21 |
| FIGURA 3.1 – NÍVEIS DE COMPLEXIDADE DA ANÁLISE DE DADOS | 28 |
| FIGURA 3.2 – PASSOS DA ANÁLISE DE AGRUPAMENTOS | 30 |
| FIGURA 4.1 – ESPAÇO DE DADOS DO PRIMEIRO CONJUNTO DE TESTES | 60 |
| FIGURA 4.2 – GRÁFICO DA MEDIDA ICC PARA $M = 1,25$ | 60 |
| FIGURA 4.3 – GRÁFICO DA MEDIDA ICC PARA $M = 1,50$ | 61 |
| FIGURA 4.4 – GRÁFICO DA MEDIDA ICC PARA $M = 1,75$ | 62 |
| FIGURA 4.5 – GRÁFICO DA MEDIDA ICC PARA $M = 2,00$ | 62 |
| FIGURA 4.6 – GRÁFICO DE COMPARAÇÃO DOS RESULTADOS DA ICC E ICC_{Novo} PARA $M = 1,25$ | 63 |
| FIGURA 4.7 – GRÁFICO DE COMPARAÇÃO DOS RESULTADOS DA ICC E ICC_{Novo} PARA $M = 1,50$ | 63 |
| FIGURA 4.8 – GRÁFICO DE COMPARAÇÃO DOS RESULTADOS DA ICC E ICC_{Novo} PARA $M = 1,75$ | 64 |
| FIGURA 4.9 – GRÁFICO DE COMPARAÇÃO DOS RESULTADOS DA ICC E ICC_{Novo} PARA $M = 2,00$ | 64 |
| FIGURA 4.10 – GRÁFICO COMPARAÇÃO ENTRE \sqrt{c} E $\sqrt{c \cdot \ln c}$ | 65 |
| FIGURA 4.11 – GRÁFICO DOS VALORES DE ICC DA USINA DE SERRA DA MESA PARA $M = 1,25$ | 66 |
| FIGURA 4.12 – GRÁFICO DOS VALORES DE ICC DA USINA DE SERRA DA MESA PARA $M = 1,50$ | 67 |
| FIGURA 4.13 – GRÁFICO DOS VALORES DE ICC DA USINA DE SERRA DA MESA PARA $M = 1,75$ | 67 |
| FIGURA 4.14 – GRÁFICO DOS VALORES DE ICC DA USINA DE SERRA DA MESA PARA $M = 2,00$ | 68 |
| FIGURA 4.15 – GRÁFICO DOS VALORES DE ICC DA USINA DE MASCARENHAS PARA $M = 1,25$ | 69 |
| FIGURA 4.16 – GRÁFICO DOS VALORES DE ICC DA USINA DE MASCARENHAS PARA $M = 1,50$ | 69 |
| FIGURA 4.17 – GRÁFICO DOS VALORES DE ICC DA USINA DE MASCARENHAS PARA $M = 1,75$ | 70 |
| FIGURA 4.18 – GRÁFICO DOS VALORES DE ICC DA USINA DE MASCARENHAS PARA $M = 2,00$ | 70 |
| FIGURA 4.19 – GRÁFICO DOS VALORES DE ICC DA USINA DE JUPIÁ PARA $M = 1,25$ | 71 |
| FIGURA 4.20 – GRÁFICO DOS VALORES DE ICC DA USINA DE JUPIÁ PARA $M = 1,50$ | 71 |
| FIGURA 4.21 – GRÁFICO DOS VALORES DE ICC DA USINA DE JUPIÁ PARA $M = 1,75$ | 72 |
| FIGURA 4.22 – GRÁFICO DOS VALORES DE ICC DA USINA DE JUPIÁ PARA $M = 2,00$ | 72 |
| FIGURA 4.23 – GRÁFICO DOS VALORES DE ICC DA USINA DE FUNIL PARA $M = 1,25$ | 73 |
| FIGURA 4.24 – GRÁFICO DOS VALORES DE ICC DA USINA DE FUNIL PARA $M = 1,50$ | 74 |
| FIGURA 4.25 – GRÁFICO DOS VALORES DE ICC DA USINA DE FUNIL PARA $M = 1,75$ | 74 |

| | |
|--|-----|
| FIGURA 4.26 – GRÁFICO DOS VALORES DE ICC DA USINA DE FUNIL PARA $M = 2,00$ | 75 |
| FIGURA 5.1 – DIAGRAMA UNIFILAR DO SISTEMA IEEE-RTS | 80 |
| FIGURA 5.2 – POTÊNCIAS MÁXIMAS NOS CASOS DO EXEMPLO | 82 |
| FIGURA 5.3 – LOLP DOS CASOS DE CONFIABILIDADE DO EXEMPLO | 83 |
| FIGURA 5.4 – EENS DOS CASOS DE CONFIABILIDADE DO EXEMPLO | 83 |
| FIGURA 5.5 – FLUXO DE INFORMAÇÕES ENTRE O NEWAVE, O SUISHI-O E O NH2 | 85 |
| FIGURA 5.6 – GRÁFICO DAS FAIXAS DE POTÊNCIA DA USINA DE JAGUARÁ | 88 |
| FIGURA 5.7 – POTÊNCIAS MÁXIMAS (MW) NOS CASOS DE CONFIABILIDADE | 92 |
| FIGURA 5.8 – LOLP DOS CASOS DE CONFIABILIDADE | 93 |
| FIGURA 5.9 – EENS DOS CASOS DE CONFIABILIDADE | 93 |
| FIGURA 5.10 – FLUXO DE INFORMAÇÕES ENTRE OS MODELOS ENERGÉTICOS E O NH2 COM A UTILIZAÇÃO DO PROCESSO DE AGRUPAMENTO. | 94 |
| FIGURA 6.1 – GRÁFICO DA POTÊNCIA DISPONÍVEL DA USINA DE FURNAS | 97 |
| FIGURA 6.2 – GRÁFICO DA POTÊNCIA DISPONÍVEL DA USINA DE JAGUARÁ | 97 |
| FIGURA 6.3 – POTÊNCIA DISPONÍVEL DA USINA DE EUCLIDES DA CUNHA | 98 |
| FIGURA 6.4 – GRÁFICO DA POTÊNCIA DISPONÍVEL NÃO NORMALIZADA DA USINA DE FUNIL | 105 |
| FIGURA 6.5 – GRÁFICO DA POTÊNCIA DISPONÍVEL NORMALIZADA PELO MÁXIMO DE CADA SÉRIE DA USINA DE FUNIL | 105 |
| FIGURA 6.6 – GRÁFICO DA POTÊNCIA DISPONÍVEL NORMALIZADA PELO MÁXIMO GLOBAL DO ESPAÇO DE DADOS DA USINA DE FUNIL | 106 |
| FIGURA 6.7 – REPRESENTAÇÃO DO ESPAÇO DE DADOS - AGRUPAMENTO POR MATRIZ ÚNICA | 108 |
| FIGURA 6.8 – GRÁFICO DE COMPARAÇÃO DAS ESTATÍSTICAS DA LOLP - <i>CASOS DO HISTÓRICO X AMU</i> . .. | 115 |
| FIGURA 6.9 – REPRESENTAÇÃO DO ESPAÇO DE DADOS UTILIZADO NO AGRUPAMENTO POR MAPEAMENTO DE RESULTADOS | 117 |
| FIGURA 6.10 – REPRESENTAÇÃO DO <i>MAPA DE RESULTADOS</i> | 121 |
| FIGURA 6.11 - EXEMPLO DE AMOSTRAS DO MR | 122 |
| FIGURA 6.12 – GRÁFICO DE COMPARAÇÃO DAS ESTATÍSTICAS DA LOLP - <i>CASOS DO HISTÓRICO X AMR</i> .. | 127 |
| FIGURA 6.13 – GRÁFICO DE COMPARAÇÃO DAS ESTATÍSTICAS DA LOLP - <i>CASOS DO HISTÓRICO X AMRM</i> | 134 |
| | |
| FIGURA A. 1 – GRÁFICO DAS USINAS COM COEFICIENTE DE VARIAÇÃO MÉDIO IGUAL A ZERO | 149 |

Lista de Tabelas

| | |
|---|-----|
| TABELA 4.1 – COMPARAÇÃO ENTRE \sqrt{c} E $\sqrt{c \cdot \ln c}$ | 65 |
| TABELA 4.2 – VALORES DA <i>ICC</i> PARA 3 AGRUPAMENTOS NA USINA DE SERRA DA MESA | 68 |
| TABELA 4.3 – VALORES DA <i>ICC</i> PARA 2 AGRUPAMENTOS NA USINA DE JUPIÁ | 73 |
| TABELA 4.4 – COMPARAÇÃO DO NÚMERO DE CASOS EM QUE NÃO FOI DETERMINADO UM MÁXIMO PARA A MEDIDA <i>ICC</i> E O VALOR DO COEFICIENTE DE NEBULIZAÇÃO <i>M</i> | 76 |
| TABELA 5.1 – LIMITES MÁXIMOS DE GERAÇÃO CONSIDERADOS NOS 6 CASOS DO EXEMPLO | 81 |
| TABELA 5.2 – RESULTADOS DA ANÁLISE DE CONFIABILIDADE DO CASO EXEMPLO | 81 |
| TABELA 5.3 - ESTATÍSTICAS DO CASO EXEMPLO | 82 |
| TABELA 5.4 – USINAS HIDRÁULICAS QUE COMPÕEM O SUBSISTEMA SUDESTE/CENTRO-OESTE | 86 |
| TABELA 5.5 – RESULTADOS DA ANÁLISE DE CONFIABILIDADE UTILIZANDO TODOS OS VALORES DAS POTÊNCIAS DISPONÍVEIS OBTIDAS A PARTIR DO HISTÓRICO (<i>CASOS DO HISTÓRICO</i>) | 90 |
| TABELA 5.6 – ESTATÍSTICAS DOS <i>CASOS DO HISTÓRICO</i> | 92 |
| TABELA 6.1 – USINAS QUE PARTICIPAM DO PROCESSO DE AGRUPAMENTO | 103 |
| TABELA 6.2 – RESULTADOS DO <i>AMU</i> | 111 |
| TABELA 6.3 – RESULTADOS DO NH2 – CASOS FORMADOS PELO <i>AMU</i> – <i>M</i> = 1,25 | 112 |
| TABELA 6.4 – RESULTADOS DO NH2 – CASOS FORMADOS PELO <i>AMU</i> – <i>M</i> = 1,50 | 113 |
| TABELA 6.5 – RESULTADOS DO NH2 – CASOS FORMADOS PELO <i>AMU</i> – <i>M</i> = 1,75 | 113 |
| TABELA 6.6 – RESULTADOS DO NH2 – CASOS FORMADOS PELO <i>AMU</i> – <i>M</i> = 2,00 | 113 |
| TABELA 6.7 – ESTATÍSTICAS DOS RESULTADOS DOS CASOS DO NH2 (<i>AMU</i>) PARA LOLP | 114 |
| TABELA 6.8 – RESULTADOS DO <i>AMR</i> | 124 |
| TABELA 6.9 – RESULTADOS DO NH2 – CASOS FORMADOS PELO <i>AMR</i> – <i>M</i> = 1,25 | 125 |
| TABELA 6.10 – RESULTADOS DO NH2 – CASOS FORMADOS PELO <i>AMR</i> – <i>M</i> = 1,50 | 126 |
| TABELA 6.11 – RESULTADOS DO NH2 – CASOS FORMADOS PELO <i>AMR</i> – <i>M</i> = 1,75 | 126 |
| TABELA 6.12 – ESTATÍSTICAS DOS RESULTADOS DOS CASOS DO NH2 (<i>AMR</i>) PARA LOLP | 126 |
| TABELA 6.13 – ESTATÍSTICAS DOS RESULTADOS DOS CASOS DO NH2 (<i>AMR</i>) PARA LOLP | 126 |
| TABELA 6.14 – RESULTADOS DO <i>AMRM</i> | 131 |
| TABELA 6.15 – RESULTADOS DO NH2 – CASOS FORMADOS PELO <i>AMRM</i> | 133 |
| TABELA 6.16 – ESTATÍSTICAS DOS RESULTADOS DOS CASOS DO NH2 (<i>AMRM</i>) PARA LOLP | 133 |
| TABELA 6.17 – RESULTADOS PARA A USINAS DE MASCARENHAS | 135 |
| | |
| TABELA A. 1 – USINAS COM COEFICIENTE DE VARIAÇÃO IGUAL A ZERO | 150 |
| TABELA A. 2 – USINAS COM COEFICIENTE DE VARIAÇÃO MÉDIO MAIOR QUE ZERO E MENOR QUE UM | 150 |
| | |
| TABELA B. 1 – USINAS QUE PARTICIPAM DO PROCESSO DE AGRUPAMENTO | 154 |

Lista de Siglas

| | |
|----------|--|
| ACP | – Análise de componentes principais |
| CCPE | – Comitê Coordenador do Planejamento da Expansão dos Sistemas Elétricos |
| CHEIAS | – Sistema para Estudos de Previsão de Cheias em Bacias Hidrográficas |
| CMO | – Custo Marginal de Operação |
| CONFINT | – Modelo de Confiabilidade em Sistemas Hidrotérmicos Interligados |
| CS | – Compacidade e Separação (Compactness and Separation) |
| CV | – Coeficiente de Variação |
| DECOMP | – Modelo para Determinação da Coordenação da Operação a Curto Prazo |
| DESSEM | – Modelo para Programação de Operação |
| DUSI | – Dados de Usina |
| EENS | – Valor esperado da energia não suprida (Expected energy not supplied) |
| EFLD | – Discriminante Linear de Fisher Estendido (Extended Fisher Linear Discriminant) |
| EPNS | – Valor esperado da potência não suprida (Expected power not supplied) |
| FCM | – Algoritmo de Agrupamento Nebuloso Fuzzy C-Means |
| FLD | – Discriminante Linear de Fisher (Fisher Linear Discriminant) |
| GEVAZP | – Modelo de Geração de Séries Sintéticas de Vazões e Energias |
| GG | – Algoritmo de Agrupamento Nebuloso Gustafson-Kessel |
| GK | – Algoritmo de Agrupamento Nebuloso Gath-Geva |
| GW | – Giga Watt |
| ICC | – Contraste Entre Classes (Inter Class Contrast) |
| IEEE | – The Institute of Electrical and Electronics Engineers |
| IEEE-RTS | – Sistema Teste de Confiabilidade do IEEE (Reliability Test System) |
| LOLD | – Valor esperado de duração de perda de carga (Loss of load duration) |
| LOLE | – Valor esperado de horas de déficit de potência (Loss of load expectation) |
| LOLF | – Valor esperado da frequência corte de carga (Loss of load frequency) |
| LOLP | – Loss of load probability - Probabilidade de perda de carga |
| MME | – Ministério de Minas e Energia |
| MVAR | – Mega Volt Ampere Reativo |
| MW | – Mega Watt |
| NEWAVE | – Modelo Estratégico de Geração Hidrotérmica a Subsistemas Equivalentes |
| ONS | – Operador Nacional do Sistema |

| | |
|-----------|--|
| PAM | – Partitioning Around Medoids |
| AMR | – Processo de Agrupamento por Mapeamento de Resultados |
| AMRM | – Processo de Agrupamento por Mapeamento de Resultados com Seleção do Coeficiente de Nebulização |
| AMU | – Processo de Agrupamento de Matriz Única |
| PAR(p) | – Modelos Auto-regressivos Mensais Periódicos de Ordem Variável |
| PARMA | – Modelos Auto-regressivos Média-móvel Periódicos |
| PDD | – Programação Dinâmica Dual Determinística |
| PDDE | – Programação Dinâmica Dual Estocástica |
| PMO | – Plano Mensal de Operação |
| PREVIVAZ | – Modelo de Previsão de Vazões Semanais |
| \bar{S} | – Silhueta Média Global (Overall Average Silhouette) |
| SIN | – Sistema Integrado Nacional |
| SUISHI-O | – Modelo de Simulação a Usinas Individualizadas de Subsistemas Hidrotérmicos Interligados |

Sumário

| | | |
|----------|--|-----------|
| 1 | INTRODUÇÃO | 1 |
| 1.1 | CONSIDERAÇÕES GERAIS | 1 |
| 1.2 | O SISTEMA ELÉTRICO BRASILEIRO | 3 |
| 1.3 | OBJETIVO | 5 |
| 1.4 | SOLUÇÃO PROPOSTA | 6 |
| 1.5 | ORGANIZAÇÃO DA DISSERTAÇÃO | 7 |
| 2 | SISTEMA ELÉTRICO | 9 |
| 2.1 | SISTEMA ELÉTRICO BRASILEIRO | 9 |
| 2.2 | PLANEJAMENTO DA OPERAÇÃO ENERGÉTICA | 10 |
| 2.2.1 | <i>NEWAVE</i> | 12 |
| 2.2.2 | <i>SUSHI-O</i> | 13 |
| 2.2.3 | <i>GEVAZP</i> | 14 |
| 2.2.4 | <i>DECOMP</i> | 14 |
| 2.2.5 | <i>CONFINT</i> | 15 |
| 2.2.6 | <i>PREVIVAZ</i> | 15 |
| 2.2.7 | <i>DESSEM</i> | 16 |
| 2.2.8 | <i>CHEIAS</i> | 16 |
| 2.3 | PLANEJAMENTO ELÉTRICO..... | 16 |
| 2.3.1 | <i>Análise de Confiabilidade</i> | 17 |
| 2.3.1.1 | Conceitos Básicos Sobre Análise de Confiabilidade Composta | 19 |
| 2.3.1.2 | Algoritmo Conceitual Para Avaliação de Confiabilidade | 21 |
| 2.3.2 | <i>Análise de Confiabilidade Aplicada ao Sistema Elétrico Brasileiro - O Sistema NH2</i> . | 23 |
| 3 | TÉCNICAS DE ANÁLISE E AGRUPAMENTO DE DADOS | 27 |
| 3.1 | ANÁLISE DE DADOS | 27 |
| 3.2 | ANÁLISE DE AGRUPAMENTOS | 29 |
| 3.2.1 | <i>Algoritmos de Agrupamento</i> | 31 |
| 3.2.1.1 | Algoritmo k-Means | 32 |
| 3.2.1.2 | Partitioning Around Medoids (PAM)..... | 36 |

| | | |
|----------|--|-----------|
| 3.2.1.3 | Algoritmo Fuzzy C-Means (FCM) | 38 |
| 3.2.1.4 | Algoritmo Gustafson-Kessel (GK) | 42 |
| 3.2.1.5 | Algoritmo Gath-Geva (GG)..... | 43 |
| 3.3 | ALGORITMOS SELECIONADOS | 45 |
| 3.4 | MEDIDAS DE VALIDAÇÃO DE AGRUPAMENTOS | 46 |
| 3.4.1 | <i>Compacidade e Separação (CS)</i> | 47 |
| 3.4.2 | <i>Discriminante Linear de Fisher (FLD)</i> | 47 |
| 3.4.3 | <i>Discriminante Linear de Fisher Estendido (EFLD)</i> | 49 |
| 3.4.4 | <i>Contraste entre Classes (ICC)</i> | 51 |
| 3.4.5 | <i>Silhueta Média Global (\bar{S})</i> | 53 |
| 3.4.6 | <i>Medidas de Validação Seleccionadas</i> | 56 |
| 4 | APLICAÇÃO DA MEDIDA ICC PARA DADOS MULTIDIMENSIONAIS | 57 |
| 4.1 | PRIMEIRO CONJUNTO DE TESTES | 59 |
| 4.2 | SEGUNDO CONJUNTO DE TESTES | 66 |
| 4.3 | RESULTADOS E CONCLUSÕES SOBRE OS TESTES | 75 |
| 5 | AVALIAÇÃO DE CONFIABILIDADE EM SISTEMAS HIDROTÉRMICOS | 78 |
| 5.1 | AVALIANDO A VARIAÇÃO HIDROLÓGICA – CASO EXEMPLO | 79 |
| 5.2 | AVALIANDO A VARIAÇÃO HIDROLÓGICA – CASO SUDESTE/CENTRO-OESTE | 84 |
| 6 | PROCESSOS DE AGRUPAMENTO | 95 |
| 6.1 | SELEÇÃO DE CARACTERÍSTICAS..... | 95 |
| 6.1.1 | <i>Aquisição de Dados</i> | 95 |
| 6.1.1.1 | Caracterização das Usinas | 96 |
| 6.1.1.2 | Métodos Estatísticos..... | 98 |
| 6.1.2 | <i>Pré-Processamento</i> | 104 |
| 6.2 | DEFINIÇÕES PRELIMINARES | 106 |
| 6.3 | AGRUPAMENTO DE MATRIZ ÚNICA..... | 107 |
| 6.3.1 | <i>Espaço de Dados</i> | 108 |
| 6.3.2 | <i>Algoritmo</i> | 109 |
| 6.3.3 | <i>Resultados e Conclusões Sobre o Processo de Agrupamento</i> | 110 |
| 6.3.4 | <i>Resultados da Análise de Confiabilidade</i> | 112 |

| | | |
|---------|--|-----|
| 6.4 | AGRUPAMENTO POR MAPEAMENTO DE RESULTADOS | 115 |
| 6.4.1 | <i>Primeira Etapa do Agrupamento Por Mapeamento de Resultados</i> | 116 |
| 6.4.1.1 | Espaço de Dados da Primeira Etapa | 117 |
| 6.4.1.2 | Algoritmo da Primeira Etapa | 117 |
| 6.4.2 | <i>Segunda Etapa do Agrupamento Por Mapeamento de Resultados</i> | 119 |
| 6.4.2.1 | Espaço de Dados Da Segunda Etapa - O Mapa de Resultados (MR) | 120 |
| 6.4.2.2 | Algoritmo Da Segunda Etapa | 122 |
| 6.4.3 | <i>Resultados e Conclusões Sobre o Processo de Agrupamento</i> | 124 |
| 6.4.4 | <i>Resultados da Análise de Confiabilidade</i> | 125 |
| 6.5 | AGRUPAMENTO POR MAPEAMENTO DE RESULTADOS COM SELEÇÃO DO COEFICIENTE DE NEBULIZAÇÃO | 128 |
| 6.5.1 | <i>Algoritmo Alterado da Primeira Etapa</i> | 128 |
| 6.5.2 | <i>Resultados e Conclusões Sobre o Processo de Agrupamento</i> | 130 |
| 6.5.3 | <i>Resultados e Conclusões Sobre a Análise de Confiabilidade</i> | 132 |
| 7 | CONCLUSÕES | 136 |
| 7.1 | SUGESTÕES PARA TRABALHOS FUTUROS | 138 |
| | REFERÊNCIAS BIBLIOGRÁFICAS | 140 |
| | GLOSSÁRIO | 145 |
| | APÊNDICE | 149 |
| | A - USINAS QUE NÃO PARTICIPAM DO PROCESSO DE AGRUPAMENTO | 149 |
| | B - USINAS QUE PARTICIPAM DO PROCESSO DE AGRUPAMENTO | 154 |

1 INTRODUÇÃO

1.1 CONSIDERAÇÕES GERAIS

O objetivo de um sistema elétrico de potência é produzir, transmitir e distribuir energia elétrica com segurança, com custo mínimo e com o maior grau de continuidade e qualidade do serviço. A fim de garantir que estas condições serão atendidas, torna-se necessário que o planejamento e a operação do sistema mantenham reservas adequadas de geração que minimizem o risco de falhas no suprimento de energia. (MELLO, 1994)

Em sua representação mais simplificada, um sistema de energia elétrica pode ser subdividido em meios de produção, meios de transporte e meios de consumo da energia elétrica, como mostra a Figura 1.1. (FORTUNATO, 1990).

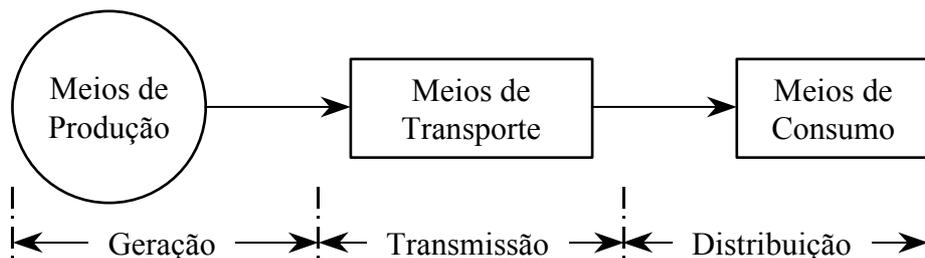


Figura 1.1 – Representação simplificada de um sistema de energia elétrica

Cada uma das partes da Figura 1.1 pode ser interpretada como um elo de uma cadeia de produção da energia elétrica.

O primeiro elo é formado pelos meios de produção, que correspondem aos diferentes tipos de equipamentos necessários para a geração de energia elétrica em escala industrial. Dentre estes equipamentos, são particularmente relevantes às usinas hidrelétricas, onde a energia potencial dos cursos d'água é transformada em energia elétrica, e as usinas termelétricas, onde a energia resulta da transformação da energia cinética de gases e vapores em expansão, aquecidos pela queima de combustíveis.

O segundo elo é representado pelos meios de transporte, que correspondem às linhas de transmissão e subestações, utilizadas para fazer a energia produzida nas usinas chegar aos consumidores.

No último elo encontram-se os meios de consumo de energia elétrica, que correspondem ao conjunto de cargas (equipamentos, instalações, etc.) dos diferentes tipos de consumidores e a rede de distribuição de energia.

Cada um dos elos desta cadeia possui um conjunto de padrões e especificações que influem na qualidade do produto final, ou seja, da energia utilizada pelos consumidores. Deve ser ressaltado que o conceito de qualidade do produto energia elétrica está usualmente associado à continuidade do suprimento e ao atendimento de padrões de regulação de frequência e tensão.

Sobre o ponto de vista do investimento, cada um dos elos possui necessidades específicas. Em contra ponto, embora suas finalidades sejam diferentes, não há como diferenciar os elos quanto a sua importância relativa na cadeia. São elementos em série e, portanto, a confiabilidade e a economicidade do todo é função da confiabilidade e da economicidade de cada um dos elos.

Uma das técnicas utilizadas para mensurar o risco de ocorrência de falhas no suprimento de energia em um sistema elétrico, e conseqüentemente, a qualidade do produto energia elétrica, é a análise de confiabilidade. Esta técnica possui como principal resultado um conjunto de índices de confiabilidade, calculados por meio de métodos estatísticos, que serão utilizados por especialistas como base para a verificação do estado do sistema. (MELO, 1990)

Cabe aqui definir e diferenciar uma configuração do sistema elétrico e do estado em que este se encontra. A configuração diz respeito ao conjunto completo de equipamentos e instalações existentes em um sistema elétrico, ou seja, o número total de geradores, circuitos, cargas, etc que existem fisicamente no sistema. Por sua vez, um estado do sistema representa quais dos equipamentos e instalações pertencentes a uma configuração que se encontram operando em um determinado instante. Por exemplo, uma configuração de um sistema elétrico pode possuir 10 geradores, e ter um estado definido com 9 geradores operando e 1 gerador parado por falha.

Ao avaliar a confiabilidade do sistema elétrico é importante considerar o comportamento estocástico das variáveis envolvidas no problema que definem os diversos estados do sistema, tais como saídas forçadas dos equipamentos de geração e de transmissão, assim como a variação da carga ao longo do tempo. Os índices de confiabilidade são obtidos analisando-se o desempenho de diversos estados do sistema obtidos pela combinação dos níveis da carga e da disponibilidade dos equipamentos, condicionadas a um cenário hidrológico específico.

A análise de confiabilidade quando efetuada desta forma considera que a capacidade de produção de todas as usinas é constante.

Entretanto, nas usinas hidrelétricas, a produção de energia depende do volume de água armazenado nos reservatórios das usinas. Por sua vez, o volume armazenado no reservatório depende das condições hidrológicas observadas durante um período de tempo. Portanto, pode-se afirmar que a potência disponível das usinas hidrelétricas é também função da hidrologia, que também é uma variável com comportamento aleatório cuja previsão é complexa.

Logo, em sistemas em que a capacidade de geração é predominantemente hidrelétrica, a hidrologia também desempenha um papel importante nos níveis da confiabilidade por causa da variação que esta impõem a capacidade de produção de energia do sistema. Esta capacidade de produção também é conhecida pelo termo potência disponível.

Uma forma de avaliar a confiabilidade deste tipo de sistema, levando-se em conta a variação hidrológica, é considerando diversos cenários hidrológicos refletidos nos valores de potência disponível para cada uma das usinas hidrelétricas existentes no sistema.

Como a hidrologia é uma variável estocástica de comportamento distinto em cada uma das cascatas nas quais as usinas se localizam, a quantidade de cenários que devem ser simulados pode atingir valores consideráveis. Além do número elevado de cenários, também deve ser levado em conta que para cada cenário simulado deve-se solucionar um conjunto de análises dos estados do sistema, que podem compreender a solução de um fluxo de potência não linear. Após a análise de todos os fatores envolvidos pode-se inferir que o esforço computacional associado a estes estudos torna-se extremamente elevado.

1.2 O SISTEMA ELÉTRICO BRASILEIRO

No caso do sistema elétrico brasileiro, seria possível realizar os estudos de análise de confiabilidade levando-se em conta a variação na potência disponível utilizando os modelos Newave (Modelo Estratégico de Geração Hidrotérmica a Subsistemas Equivalentes) e Suishi-O (Modelo de Simulação de Usinas Individualizadas de Subsistemas Hidrotérmicos Interligados) (AMADO 2000; MACEIRA, 1997, 1998, 2002), pertencentes à cadeia de modelos utilizados no planejamento energético do sistema, e o sistema NH2 (Sistema Computacional para Análise da Confiabilidade de Geração e Transmissão) (MELO, 1995, 1999), uma das ferramentas mais utilizadas para nos estudos de confiabilidade composta geração/transmissão. Todos os três

modelos foram desenvolvidos pelo Cepel (Centro de Pesquisas em Energia Elétrica) e serão apresentados com mais detalhes no capítulo 2.

O fluxo de informação entre estes 3 sistemas para realização da análise de confiabilidade levando em conta a variação na potência disponível imposta pela variação da hidrologia é apresentado de forma esquemática na Figura 1.2.

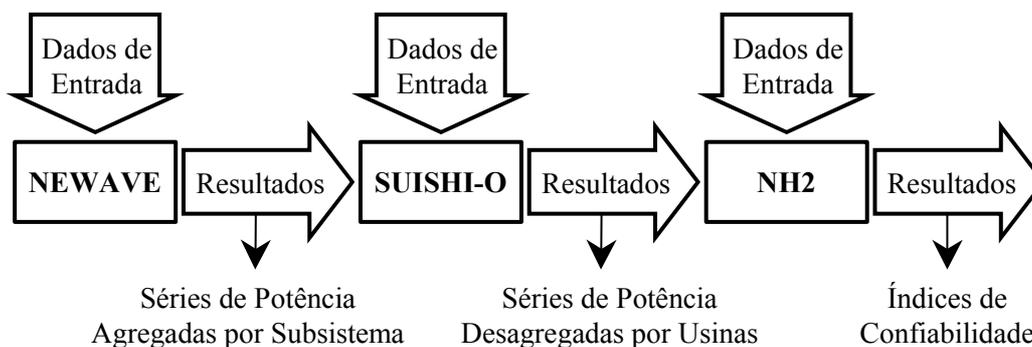


Figura 1.2 - Fluxo de informação entre os modelos.

Resumidamente, os modelos Newave e Suishi-O, utilizando os dados de uma configuração do sistema e de séries de vazão, históricas ou sintéticas, são capazes de calcular séries de potência disponível para cada uma das usinas hidrelétricas pertencentes ao sistema. Na Figura 1.3, são apresentadas as séries de potência disponível da usina hidrelétrica de Caconde, calculadas a partir das séries históricas de vazão.

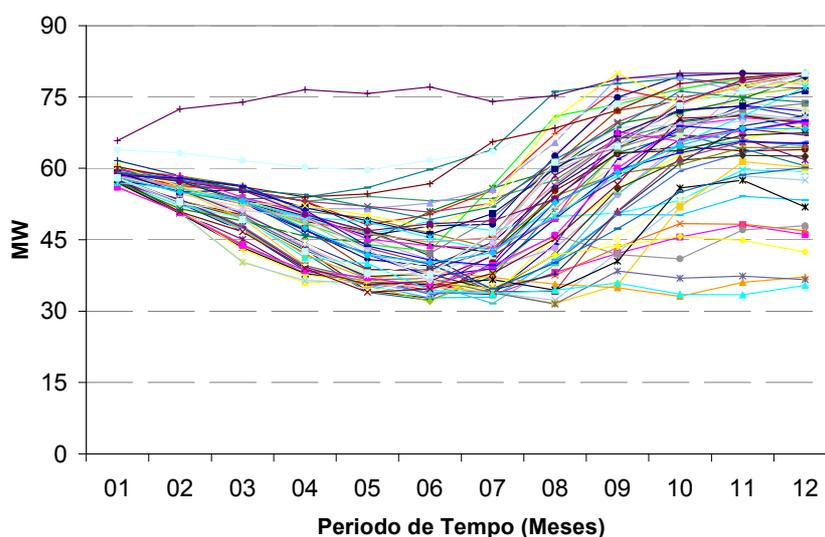


Figura 1.3 - Série de potência disponível da usina hidrelétrica de Caconde.

No sistema NH2 os cálculos de confiabilidade são realizados para dados de um determinado cenário hidrológico. Desta forma, para avaliar a confiabilidade do sistema deve ser

formado um estudo para cada um dos valores das séries de potência disponível de cada uma das usinas pertencentes ao caso. Por exemplo, no primeiro estudo formado, o valor da potência disponível de cada uma das usinas hidrelétricas refletiria o valor obtido no primeiro mês da primeira série de potência disponível, respectivamente.

Procedendo desta forma o número de estudos de confiabilidade formados seria equivalente ao produto do número de séries de potência disponível pelo número de meses do período de tempo que se deseja estudar.

De forma a obter resultados os mais acurados o possível, o número de séries potência disponível que devem ser avaliadas deve ser o maior possível, para que estas reflitam períodos úmidos e secos de grande duração. Com isso o número de casos a serem avaliados é muito grande.

Atualmente, a avaliação de confiabilidade levando em conta a variação hidrológica do sistema elétrico brasileiro não é realizada, pois o esforço computacional para realizar o estudo de confiabilidade deste sistema para um cenário hidrológico já implica em algumas horas de processamento, dependendo da configuração escolhida para análise. Se levarmos em conta que, para avaliar a variação hidrológica, devem ser realizadas algumas dezenas de estudos. O esforço computacional envolvido é muito grande.

1.3 OBJETIVO

O objetivo deste trabalho é propor um processo de agrupamento aplicado às séries de potência disponível das usinas hidrelétricas, visando reduzir o esforço computacional proveniente da avaliação dos diversos cenários hidrológicos nos estudos de confiabilidade. Este processo consiste na aplicação das técnicas agrupamento (*Cluster Techniques*), baseadas em lógica nebulosa (*Fuzzy Logic*) e em técnicas estatísticas.

A Figura 1.4 apresenta alteração do fluxo de informações entre os 3 modelos incluindo o processo de agrupamento, a ser definido. Do conjunto de resultados do SUSHI-O, o processo de agrupamento utilizada as séries de potência disponível de cada uma das usinas hidrelétricas e calcula um número menor de séries que refletem os centros dos agrupamentos.

Como o conjunto de series de potência resultante do processo de agrupamento é muito menor que o número inicial de séries de potência disponível, o número de casos de análise de confiabilidade que deve ser executado é muito menor, e assim o esforço computacional resultante é muito menor também.

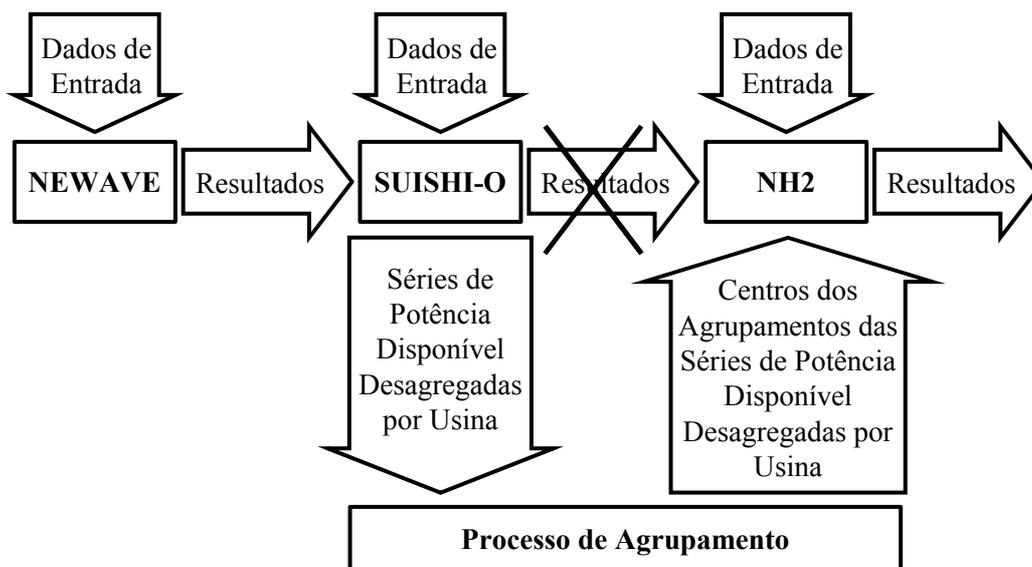


Figura 1.4 - Fluxo de informação entre os modelos com processo de agrupamento.

O processo de agrupamento proposto tem o compromisso de refletir o comportamento aleatório das séries hidrológicas em um número menor de séries de potência disponível mantendo a dependência temporal e espacial dos dados originais. A análise de confiabilidade baseada neste número reduzido de cenários hidrológicos resulta em um esforço computacional reduzido e mantém a exatidão na estimativa dos índices de confiabilidade.

Para avaliação do modelo proposto serão utilizados dados referentes ao sistema elétrico brasileiro, que possui aproximadamente 78% de sua capacidade de geração oriunda de usinas hidrelétricas. Grande parte destas usinas possui reservatórios com capacidade de regulação plurianual, distribuídos ao longo de cascatas localizadas em diversas bacias hidrográficas espalhadas pelo território nacional. (ANEEL, 2004)

1.4 SOLUÇÃO PROPOSTA

Neste trabalho, o espaço de dados que será utilizado pelo processo de agrupamento é composto por conjuntos de séries mensais de potência disponível para cada uma das usinas hidrelétricas do sistema elétrico brasileiro. Estas séries de potência foram obtidas a partir da utilização do modelo NEWAVE (Modelo Estratégico de Geração Hidrotérmica a Subsistemas Equivalentes) e do modelo SUIHI-O (Modelo de Simulação a Usinas Individualizadas de Subsistemas Hidrotérmicos Interligados), ambos desenvolvidos pelo Centro de Pesquisas em Energia Elétrica (CEPEL). (AMADO 2000; MACEIRA, 1997, 1998, 2002)

A análise de confiabilidade será realizada para o subsistema Sudeste/Centro-Oeste, ou seja, para uma fração do sistema elétrico brasileiro. Esta análise será executada com auxílio do programa NH2, também desenvolvido pelo CEPEL. Atualmente este programa é uma das ferramentas mais utilizadas para avaliação de confiabilidade composta geração/transmissão do sistema elétrico brasileiro. (MELO, 1995, 1999)

Nesta dissertação, a avaliação da confiabilidade levando-se em conta a variação hidrológica foi executada sob dois enfoques distintos e complementares.

No primeiro enfoque, que reflete o fluxo de informação apresentado na Figura 1.1, a avaliação da confiabilidade foi repetida para cada um dos valores da série de potência disponível obtidas a partir das séries históricas de vazão. Com auxílio destes resultados foi possível avaliar o esforço computacional necessário além da variação dos valores dos principais índices de confiabilidade.

No segundo enfoque, que reflete o fluxo de informação apresentado na Figura 1.4, a avaliação de confiabilidade foi realizada para as séries resultantes do processo de agrupamento proposto. Os resultados obtidos foram discutidos em termos da exatidão e variação dos índices da confiabilidade do sistema e em termos do esforço computacional resultante.

1.5 ORGANIZAÇÃO DA DISSERTAÇÃO

No Capítulo 2, é realizada uma revisão sobre o *Sistema Elétrico Brasileiro*, enfatizando as características os sistemas utilizados no processo de planejamento elétrico e energético. No item *Planejamento da Operação Energética* é feita uma breve descrição sobre os modelos pertencentes a cadeia de modelos energéticos, dentre estes os modelos NEWAVE e SUIISHI-O utilizados neste trabalho. No item seguinte, *Planejamento Elétrico*, é apresentado um resumo sobre análise de confiabilidade e sobre as características do sistema computacional NH2 para a avaliação de confiabilidade composta geração/transmissão.

No Capítulo 3, é realizada uma revisão bibliográfica sobre as *Técnicas de Análise e Agrupamento de Dados*. Neste capítulo estão descritos os algoritmos de agrupamento assim como as medidas de validação utilizadas para determinação do melhor número de agrupamentos.

No Capítulo 4, foram realizados testes com a medida de validação *ICC* de forma a verificar sua acuidade quando usada em espaços de dados multidimensionais.

O Capítulo 5 apresenta a complexidade e a necessidade em realizar a avaliação de confiabilidade de um sistema predominantemente hidrotérmico levando-se em conta a variação na potência disponível nas usinas hidrelétricas.

No Capítulo 6, são apresentados e discutidos segundo a visão da análise de dados, a utilização dos três processos de agrupamento propostos, o *Agrupamento de Matriz Única*, o *Agrupamento por Mapeamento de Resultados* e o *Agrupamento por Mapeamento de Resultados com Seleção do Coeficiente de Nebulização*, que visam permitir a solução do problema relativo ao tempo de processamento necessário para realizar a análise de confiabilidade associada a variação nas potências disponíveis das usinas hidrelétricas.

Finalmente, o Capítulo 7, apresenta os resultados deste trabalho analisando-os segundo a visão de compatibilidade dos resultados dos estudos de análise de confiabilidade realizados com base nos resultados dos processos de agrupamento. Neste capítulo também são feitas as sugestões de trabalhos futuros.

2 SISTEMA ELÉTRICO

Neste capítulo foi feita uma revisão bibliográfica de assuntos ligados ao planejamento da expansão e da operação de um sistema elétrico, mais especificamente sobre os modelos computacionais utilizados para realização destas etapas do planejamento no sistema elétrico brasileiro.

Por se tratar de um assunto vasto, nesta dissertação foram enfocados somente os aspectos necessários à compreensão do problema que está sendo discutido.

2.1 SISTEMA ELÉTRICO BRASILEIRO

O Brasil é um país de dimensões continentais, e seu sistema elétrico é subdividido em 5 subsistemas, Norte (N), Nordeste (NE), Sul (S), Sudeste/Centro-Oeste (SE) e Norte Isolado (NI).

Atualmente os quatro primeiros subsistemas compõem o Sistema Interligado Nacional (SIN), mostrado na Figura 2.1. O subsistema Norte Isolado, que não foi representado na figura, corresponde aos estados do Acre, Amazonas, Rondônia e Roraima. (ONS, 2002a)

O Sistema Elétrico Brasileiro possui uma capacidade instalada para produção de energia de aproximadamente 88.500 GW, onde 68.200 GW correspondem a usinas hidrelétrica, 18.200 GW correspondem a usinas termelétricas convencionais, ou seja, que utilizam combustíveis fósseis para produzir energia, e 2.100 GW correspondem a usinas que utilizam outras fontes de energia, tais como, fissão nuclear, ventos e solar (ANEEL, 2004). A simples análise destes números indica a importância da geração hidrelétrica para este sistema.

As usinas hidrelétricas que compõem este sistema são caracterizadas por reservatórios com capacidade de regulação plurianual, em geral localizadas ao longo de uma mesma bacia hidrográfica, formando cascatas. Desta forma, a operação de uma usina influencia na operação das usinas que se encontram à jusante. Além disto, a variabilidade do processo de aflúências é uma característica marcante de sistemas com predominância hidráulica (AMADO, 2000).

De forma a garantir que o sistema elétrico tenha capacidade de atender a demanda de energia existente e futura é necessário que seja realizado o planejamento da sua operação e

expansão. O planejamento possui dois enfoques complementares: o planejamento energético e o planejamento elétrico que serão detalhados a seguir.



Figura 2.1 – Representação dos subsistemas do SIN

2.2 PLANEJAMENTO DA OPERAÇÃO ENERGÉTICA

O objetivo básico do planejamento da operação energética de um sistema hidrotérmico é determinar para os diferentes horizontes de estudo, do período de planejamento (curto, médio e longo prazo) as metas de geração para cada usina e as decisões de intercâmbio, de forma que o custo de operação ao longo do período de estudo seja minimizado e que a demanda prevista de energia elétrica seja atendida. Este custo de operação é composto pelo custo do combustível necessário para operação das usinas termelétricas somado ao custo atribuído às interrupções do fornecimento de energia, representado por uma função de penalização dos déficits de energia, também conhecido como custo de déficit. (MACEIRA, 1995, 1998).

O problema do planejamento da operação energética do sistema hidrelétrico brasileiro possui as seguintes características (JARDIM, 2002):

- *Dependência temporal*: os reservatórios das usinas hidrelétricas possuem capacidade de armazenamento limitada, portanto a decisão de geração em um determinado momento afeta o nível de armazenamento dos reservatórios ao final de um período que, conseqüentemente, afeta a capacidade de geração das usinas no futuro. Desta forma, surge a dependência temporal entre a decisão operativa de um determinado período e as decisões futuras.
- *Dependência espacial*: o parque hidrelétrico Brasileiro está disposto em diversas bacias hidrográficas, com várias usinas hidrelétricas dispostas em uma mesma cascata, onde a operação das usinas à montante afeta a operação das usinas à jusante.
- *Não determinístico*: as aflúências das bacias hidrográficas variam sazonal e regionalmente e podem possuir períodos secos de longa duração, o que cria incertezas quanto as aflúências futuras.

Portanto, devido à incerteza associada às aflúências futuras, ao número de aproveitamentos existentes e a inter-relação entre as decisões tomadas num determinado instante e suas conseqüências futuras, o problema do planejamento da operação energética do sistema hidrelétrico Brasileiro se torna um problema de grande porte, estocástico e de difícil solução. Para tratar todas estas complexidades do problema de operação, torna-se necessária à utilização de uma cadeia de modelos com diferentes graus de detalhe na representação do sistema (MACEIRA, 1995, 1998), como a apresenta na Figura 2.2 (RODRIGUEZ, 2001).

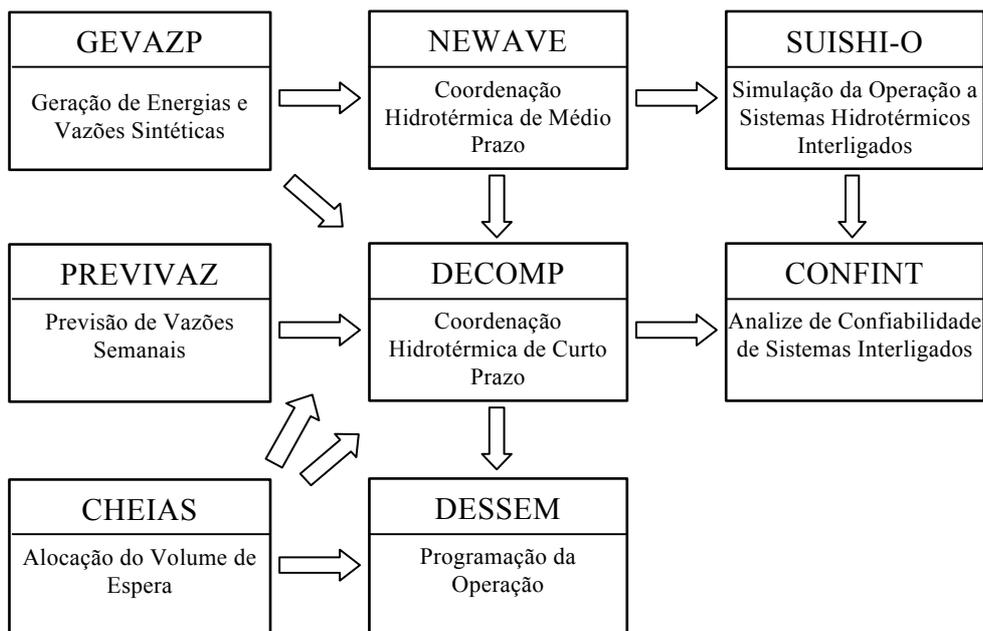


Figura 2.2 – Cadeia de modelos para o planejamento energético do sistema brasileiro

Os próximos itens descrevem cada um dos modelos apresentados na Figura 2.2, e foram elaborados com base nas seguintes referências: AMADO (2000), MACEIRA (1997, 1998, 2002), MARQUES (2000), ONS (2002b) e RODRIGUEZ (2001).

2.2.1 *NEWAVE*

O modelo *NEWAVE* (Modelo Estratégico de Geração Hidrotérmica a Subsistemas Equivalentes) é um modelo de planejamento da operação energética que define uma política ótima de operação para o sistema interligado, onde a cada mês do período de planejamento os recursos hídricos e térmicos são alocados de forma a minimizar o valor esperado do custo de operação ao longo de todo o período de planejamento. Este modelo tem capacidade para considerar vários subsistemas interligados, permite a representação estática ou dinâmica da configuração do sistema, representação do mercado em até três patamares de carga (por exemplo, pesada, média e leve), representação dos cortes no suprimento do mercado de energia elétrica em até quatro patamares de déficit, além da consideração de diversos cenários de energias afluentes, obtidos através de um modelo auto-regressivo periódico de ordem p , $PAR(p)$, que modela a afluência de um mês como sendo função de p afluências dos meses anteriores.

O modelo *NEWAVE* pode ser dividido em quatro módulos básicos, cujas características principais são descritas a seguir.

- Módulo de cálculo dos sistemas equivalentes – Calcula os subsistemas equivalentes de energia a partir de uma configuração do parque hidrotérmico (energias armazenáveis máximas, séries históricas de energias controláveis e energias fio d'água, energia de vazão mínima, energia evaporada, capacidade de turbinamento, etc). Atualmente o sistema elétrico brasileiro é dividido em 4 (quatro) subsistemas – Norte, Nordeste, Sul e Sudeste/Centro-Oeste.
- Módulo de cálculo do modelo estocástico de energias afluentes – Estima os parâmetros do modelo estocástico auto-regressivo periódico (PAR(p)) e gera séries sintéticas de energias afluentes, que são utilizadas no módulo de cálculo da política de operação hidrotérmica e na análise de desempenho no módulo de simulação da operação.
- Módulo de cálculo da política de operação hidrotérmica – O cálculo da política de operação mais econômica para os subsistemas equivalentes, que é refletido na função de custo futuro, leva em conta as incertezas nas afluências, os dados que descrevem o sistema elétrico existente e os dados do planejamento da expansão do sistema no horizonte de estudo. A função de custo futuro é construída utilizando-se técnicas de otimização baseadas na programação dinâmica dual estocástica. Esta função reflete as conseqüências futuras da utilização da água em cada período do horizonte de estudo.
- Módulo de simulação da operação – Simula a operação do sistema equivalente ao longo do período de planejamento, para cenários distintos de afluências futuras, falhas dos componentes e variações da demanda. Esta simulação visa avaliar a política ótima de operação apresentando como resultados para cada cenário de afluência e para cada mês, o montante total de geração hidrelétrica em cada subsistema, a geração das usinas térmicas e os intercâmbios entre os subsistemas. Indicadores estatísticos de desempenho do sistema também podem ser avaliados, tais como, a probabilidade de déficit de energia, o valor esperado do custo total de operação, o valor esperado da energia não suprida e os custos marginais de operação (CMO).

O modelo NEWAVE é utilizado pelo Operador Nacional do Sistema (ONS) no planejamento da operação de médio prazo (horizonte de 5 anos) e pelo Comitê Coordenador do Planejamento da Expansão dos Sistemas Elétricos (CCPE) do Ministério de Minas e Energia (MME) no planejamento decenal da expansão do parque gerador brasileiro (horizonte de 10 anos).

2.2.2 SUISHI-O

O modelo SUISHI-O (Modelo de Simulação a Usinas Individualizadas de Subsistemas Hidrotérmicos Interligados) simula a operação do sistema hidrotérmico interligado em base mensal,

considerando as usinas individualizadas, utilizando múltiplas séries hidrológicas, históricas ou sintéticas, de forma a obter índices probabilísticos de desempenho do sistema.

A cada mês, os intercâmbios entre subsistemas, as gerações térmicas e o bloco de geração hidráulica por subsistema são decididos considerando-se que a função de custo futuro produzida pelo modelo NEWAVE deve ser atendida. Estes valores são obtidos pela solução de um problema de otimização, que visa minimizar a soma dos custos atuais mais o valor presente esperado dos custos futuros de operação.

A meta de geração hidrelétrica é então desagregada por cada usina hidrelétrica através de regras heurísticas que representam as decisões operativas. Outros resultados do modelo são os índices de desempenho do sistema, tais como energias vertidas, energias armazenadas, geração média em cada usina hidráulica, déficit etc.

2.2.3 GEVAZP

O Modelo GEVAZP (Modelo de Geração de Séries Sintéticas de Vazões e Energias) tem por objetivo produzir cenários sintéticos de energias ou vazões médias mensais. O programa, a partir das séries históricas, utiliza modelos auto-regressivos mensais periódicos de ordem variável $PAR(p)$ para a geração de cenários sintéticos. É capaz de produzir um grande número de cenários, em paralelo, de vazões ou energias que são utilizadas pelo NEWAVE e pelo SUIISHI-O, ou ainda, cenários em forma de árvore que serão utilizados pelo modelo DECOMP. Este modelo estocástico provou ser adequado na reprodução de secas severas tão prováveis quanto as verificadas no registro histórico.

2.2.4 DECOMP

O modelo DECOMP (Modelo para Determinação da Coordenação da Operação a Curto Prazo) tem por objetivo determinar a estratégia de operação de curto prazo para sistemas hidrotérmicos interligados que minimiza o valor esperado do custo de operação para período de planejamento. O horizonte de planejamento é de alguns meses, onde o primeiro mês é discretizado em semanas. Nessas primeiras semanas considera-se que a afluência é determinística, obtida a partir de previsões fornecidas pelo modelo PREVIVAZ. Os meses seguintes são representados por cenários de afluências, apresentando uma estrutura em árvore, e são geradas pelo modelo GEVAZP.

O modelo representa as usinas individualmente além de representar as restrições físicas e operativas associadas ao problema (conservação da água, limites de turbinamento, defluência

mínima, armazenamento, atendimento à demanda, etc.). A metodologia empregada para a solução do problema é a Programação Dinâmica Dual Estocástica (PDDE). Integrado na cadeia de procedimentos do planejamento da operação, o modelo DECOMP possui interface com o modelo de médio prazo NEWAVE , o qual fornece a função de custo futuro a ser considerada no horizonte de planejamento do DECOMP.

2.2.5 CONFINT

O objetivo do modelo CONFINT (Confiabilidade em Sistemas Hidrotérmicos Interligados) é determinar os índices de confiabilidade, incluindo os índices de frequência e duração, para sistemas hidrotérmicos interligados, considerando também a variação na capacidade de produção das usinas hidrelétricas devido às condições hidrológicas.

Esse modelo representa o sistema interligado em áreas elétricas considerando as suas restrições de interligação e as saídas forçadas das unidades geradoras (térmicas e hidráulicas).

No modelo CONFINT, os estados do sistema são amostrados por simulação de Monte Carlo. Esse processo consiste em uma amostragem de diversos estados do sistema por sorteio da disponibilidade das unidades geradoras e interligações, série hidrológica e nível de carga. Como resultado é realizada a avaliação do atendimento à carga do sistema em cada um dos estados sorteados utilizando a modelagem de fluxo em rede. Cada estado pode ser associado à condição de “sucesso” quando a carga é atendida sem problemas, e “falha” quando ocorre algum déficit de potência.

2.2.6 PREVIVAZ

O modelo PREVIVAZ (Modelo de Previsão de Vazões Semanais) é um modelo univariado de previsão de vazões médias semanais. O programa, a partir da série histórica de vazões naturais semanais, combina várias alternativas de modelagem estocástica, incluindo a família dos modelos auto-regressivos média-móvel periódicos PARMA, com diferentes métodos de estimação de parâmetros e diversos tipos de transformações, resultando em 94 possibilidades de modelagens distintas.

Este modelo é processado para todos os postos-base, isto é, para os locais de aproveitamento que possuem série histórica de vazões naturais, sendo representativos em suas respectivas bacias. Ele é capaz de produzir previsões de vazões para um período de 6 (seis) semanas.

2.2.7 DESSEM

O Modelo DESSEM (Modelo para Programação de Operação) é um modelo de otimização para a programação da operação em um horizonte de curto prazo (de 7 a 13 dias), que tem como objetivo determinar a alocação da unidade e o despacho ótimo horário de geração para sistemas hidrotérmicos interligados que minimiza o custo total de operação no período de planejamento. Este modelo utiliza a técnica de programação dinâmica dual determinística (PDD) e decomposição pelo algoritmo de Benders.

O Modelo DESSEM considera uma representação detalhada do sistema. A representação da transmissão pode ser modelada por uma representação linearizada da rede elétrica ou apenas pelos limites de intercâmbio entre os subsistemas.

Ao final do período de estudo este modelo acopla-se à função de custo futuro produzida pelo modelo de planejamento de curto prazo, DECOMP.

2.2.8 CHEIAS

O modelo CHEIAS (Sistema para Estudos de Previsão de Cheias em Bacias Hidrográficas) é um sistema com um conjunto de funções de modelagem estocástica e geração de séries sintéticas diárias de aflúncias (DIANA), uma função para cálculo de curvas de volumes de espera para reservatórios equivalentes (CAEV), e outra para a desagregação destas em curvas individualizadas para cada reservatório do sistema (VESPOT). Possui também uma função que permite condicionar a geração de séries sintéticas de aflúncias às fases da ENSO (El Niño – Oscilação Sul). Este sistema gera informação que são utilizadas pelo NEWAVE, DECOMP, DESSEM e SUSHI-O.

2.3 PLANEJAMENTO ELÉTRICO

Um dos objetivos básicos do planejamento da expansão de um sistema elétrico de potência é determinar a seqüência de adições de equipamentos (geradores, circuitos, etc.) necessários para o suprimento da demanda prevista de forma econômica e confiável. Este objetivo é conflitante: um elevado grau de confiabilidade requer reforços no sistema, cujo custo tornaria as tarifas mais altas para os consumidores. A responsabilidade do planejador é atingir o ponto ótimo entre o custo e a confiabilidade, reconhecendo as incertezas com respeito às flutuações de carga e as indisponibilidades dos equipamentos. (PEREIRA, 1992)

O planejamento da expansão da capacidade de produção e da rede de transmissão de um sistema elétrico, bem como a inserção da análise de confiabilidade neste processo é um assunto vasto e complexo, fugindo ao escopo desta dissertação.

Assim, no que segue, será descrita a análise de confiabilidade de forma geral, e mais particularmente as técnicas para avaliação da confiabilidade de sistemas compostos geração/transmissão, por se tratarem de técnicas que serão utilizadas na avaliação dos resultados dos processos de agrupamento.

Os próximos itens deste capítulo foram elaborados com base nas seguintes referências: BILLINGTON (1984), LÍCIO (1997), MARCATO (1998), MELLO (1994), MELO (1990, 1995, 1999), ONS (2002b) e PEREIRA (1992).

2.3.1 ANÁLISE DE CONFIABILIDADE

A análise de confiabilidade é uma técnica utilizada em vários setores, como por exemplo na engenharia e em sistemas computacionais. Por isso, existem variações na definição do termo confiabilidade. Uma definição utilizada em sistemas elétricos é: “Confiabilidade é a probabilidade de um dispositivo desempenhar seu propósito adequadamente durante o tempo desejado, sob as condições operativas encontradas”. Nesta definição aparecem 4 parâmetros importantes: probabilidade, desempenho adequado, tempo e condições de operação.

A teoria da probabilidade fornece os princípios e as técnicas matemáticas para avaliar numericamente, isto é, medir, a confiabilidade. A probabilidade é o índice básico ou uma medida básica de adequação do sistema. Embora a probabilidade seja na maioria das vezes o índice mais significativo, existem muitos outros parâmetros calculados e usados, sendo que os mais apropriados são os índices que dependem do sistema que está sendo avaliado e de seus requisitos. Em razão disto, mais recentemente, em sistemas de energia elétrica tem sido utilizada a seguinte definição: “Confiabilidade é a capacidade de um dispositivo ou sistema desempenhar as suas funções adequadamente sob condições operativas pré-definidas”.

Um sistema de energia elétrica pode ser dividido em zonas funcionais de geração, transmissão e distribuição. Estas zonas podem ser combinadas para formar três níveis hierárquicos para análise de confiabilidade, conforme ilustrado na Figura 2.3.



Figura 2.3 – Níveis hierárquicos para a análise de confiabilidade

Nos métodos de *confiabilidade de geração*, ou primeiro nível hierárquico (NH1), toda a carga e geração são representados concentrados numa única barra, ignorando-se totalmente a rede de elétrica. A adequação de cada estado amostrado é avaliada pela comparação entre a capacidade total de geração e a carga total do sistema. Isto implica em menor esforço computacional para o cálculo dos índices de confiabilidade.

Com a expansão das redes elétricas e o aumento da interconexão dos sistemas de transmissão, tornou-se necessário o estudo e a representação das falhas compostas de geração e transmissão. Estes estudos estão considerados no segundo nível hierárquico (NH2), sendo também chamados de estudos de confiabilidade em sistemas compostos geração/transmissão, ou simplesmente, de métodos de *confiabilidade composta*. Nestes métodos o esforço computacional é maior, pois, devido à representação mais detalhada da rede elétrica, o número de variáveis aleatórias existente no sistema é maior. Além disso é necessário considerar o fluxo de potência do estado operativo da rede. É também necessário representar as medidas corretivas baseadas em otimização que são utilizadas para aliviar as violações operativas no sistema, o que eleva ainda mais o grau de complexidade do problema.

O terceiro nível hierárquico (NH3) compreende a análise conjunta de todas as três zonas funcionais do sistema de potência. Entretanto, devido à complexidade e as dimensões do problema, a avaliação de confiabilidade em sistemas de distribuição tem sido tradicionalmente desenvolvida separadamente dos sistemas de geração e transmissão.

Em sistemas como o brasileiro, onde há predominância de geração hidrelétrica, pode-se definir ainda o nível hierárquico zero (NH0), associado à confiabilidade energética do sistema, isto é, à variação da produção das usinas hidrelétricas devido às condições hidrológicas.

Em cada um dos níveis é calculado um conjunto de *índices de confiabilidade*. Neste trabalho estaremos interessados nos índices de confiabilidade composta. Neste caso, os índices calculados usualmente são:

- Probabilidade de perda de carga, representado em % (*LOLP – Loss of load probability*);
- Valor esperado da potência não suprida, representado em MW (*EPNS – Expected power not supplied*);
- Valor esperado de horas de déficit de potência, representado em horas (*LOLE – Loss of load expectation*),
- Valor esperado da energia não suprida, representado em MWh/período (*EENS – Expected energy not supplied*);
- Valor esperado da frequência corte de carga, representado em ocorrências/período (*LOLF – Loss of load frequency*),
- Valor esperado de duração de perda de carga, representado em horas (*LOLD – Loss of load duration*).

2.3.1.1 CONCEITOS BÁSICOS SOBRE ANÁLISE DE CONFIABILIDADE COMPOSTA

Um sistema de potência é composto por m elementos, tais como geradores, linhas de transmissão, transformadores e cargas. Cada elemento pode ser encontrado em algum estado dentro de um determinado conjunto de estados possíveis. Por exemplo, um gerador pode estar em dois estados: operando ou em falha; já os níveis de carga podem ser representados por um conjunto de níveis discretos de carga.

Seja um estado do sistema elétrico de potência representado por um vetor $\mathbf{x} = \{x_1, x_2, \dots, x_m\}$, onde x_i corresponde ao estado do i -ésimo elemento e m corresponde ao número total de elementos do sistema. O conjunto de todos os possíveis estados do sistema, denotado por \mathbf{X} , é obtido pela combinação de todos os estados possíveis dos elementos, e representa o *espaço de estados* do sistema.

Para cada estado do i -ésimo elemento é possível associar uma probabilidade de ocorrência $P(x_i)$.

Dado o estado de cada elemento, é possível calcular a probabilidade do vetor de estados \mathbf{x} , que corresponde a $P(\mathbf{x})$, e as taxas de transição entre os estados. Por exemplo, se as falhas dos

componentes do sistema são estatisticamente independentes, $P(\mathbf{x})$ corresponde ao produto das probabilidades associadas a cada estado dos componentes, como mostra a Equação 2.1.

$$P(\mathbf{x}) = \prod_{i=1}^m P(x_i) \quad \text{Equação 2.1}$$

A avaliação de índices probabilísticos é equivalente a calcular o valor esperado de uma dada função-teste, $F(\mathbf{x})$, apresentada na Equação 2.2.

$$E(F) = \sum_{\mathbf{x} \in X} F(\mathbf{x})P(\mathbf{x}) \quad \text{Equação 2.2}$$

Onde:

- \mathbf{x} – Vetor representando o estado do sistema, em que cada componente de \mathbf{x} representa um estado de um elemento do sistema;
- X – Espaço de estados, isto é, a série de todos os possíveis estados de \mathbf{x} resultantes de combinações de estados dos componentes;
- $P(\mathbf{x})$ – Probabilidade do estado \mathbf{x} ;
- $F(\mathbf{x})$ – Função Teste.

A função teste, $F(\mathbf{x})$, tem por objetivo verificar se o ponto de operação resultante da configuração específica de geradores e circuitos é adequada para atender ao nível de carga relacionado. Cada índice resulta da avaliação de uma função-teste específica.

O índice LOLP pode ser obtido como o valor esperado de uma função indicadora, definida pela Equação 2.3.

$$F(\mathbf{x}) = \begin{cases} 0, & \text{se } \mathbf{x} \text{ é um estado de sucesso} \\ 1, & \text{se } \mathbf{x} \text{ é um estado de falha} \end{cases} \quad \text{Equação 2.3}$$

A função-teste para o índice EPNS avalia o montante do corte de carga associado a cada estado do sistema como apresentado na Equação 2.4.

$$F(\mathbf{x}) = \begin{cases} 0, & \text{se } \mathbf{x} \text{ é um estado de sucesso} \\ V, & \text{se } \mathbf{x} \text{ é um estado de falha} \end{cases} \quad \text{Equação 2.4}$$

onde:

- V – Valor do corte de carga associado ao estado \mathbf{x} ;

A função-teste para o índice LOLF é dada pela Equação 2.5.

$$F(\mathbf{x}) = \begin{cases} 0, & \text{se } \mathbf{x} \text{ é um estado de sucesso} \\ T, & \text{se } \mathbf{x} \text{ é um estado de falha} \end{cases} \quad \text{Equação 2.5}$$

onde:

- T – Somatório das taxas de transição entre o estado \mathbf{x} e todos os estados de sucesso que podem ser alcançados a partir do estado \mathbf{x} em uma única transição;

Para a avaliação do índice LOLF é necessário que os estados dos componentes sejam representados por *modelos de Markov*. A Figura 2.4 ilustra o diagrama de Markov para um componente modelado a 2 estados.

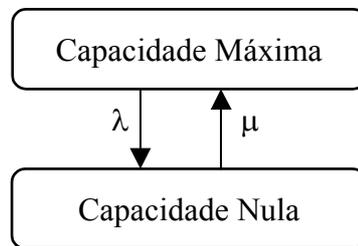


Figura 2.4 – Modelo de Markov a 2 estados

Onde:

- λ – Taxa de falha
 μ – Taxa de reparo

A LOLE e a EENS são obtidas multiplicando-se respectivamente o valor da LOLP e da EPNS pelo número de horas no período de estudo. Já a LODD é calculada pela divisão da LOLP pela LOLF.

2.3.1.2 ALGORITMO CONCEITUAL PARA AVALIAÇÃO DE CONFIABILIDADE

Um algoritmo conceitual para análise da confiabilidade composta pode ser definido da seguinte maneira:

- a) **Selecione um estado do sistema $\mathbf{x} \in X$** ; i.e., defina um nível de carga, disponibilidade dos equipamentos, condições operativas, etc;
- b) **Calcule $F(\mathbf{x})$ para o estado selecionado**; isto é, verifique se o estado \mathbf{x} é capaz de atender a carga sem violações operativas no sistema; se necessário, acione medidas

corretivas para solucionar o problema, tais como redespacho de geração, correção do perfil de tensões, e em última hipótese, corte de carga;

- c) **Atualize a estimativa de $E(F)$** baseado nos resultados obtidos no item (b), i.e. calcule os índices de confiabilidade;
- d) **Verifique a convergência do processo**, se a precisão das estimativas dos índices de confiabilidade é satisfatória; pare; caso contrário, retorne para (a).

É importante observar que o algoritmo anterior produz como resultado uma estimativa do valor de $E(F)$. Em outras palavras, se este algoritmo for utilizado para dois espaços de dados com diferentes números de amostras, o resultado final da estimativa poderia ser diferente. Como \mathbf{x} e $F(\mathbf{x})$ são variáveis aleatórias, pode-se inferir que o estimador também é uma variável aleatória. A incerteza associada a estimativa é dada pela variância do estimador, como mostra a Equação 2.6.

$$V(E(F)) = \frac{V(F)}{n} \quad \text{Equação 2.6}$$

Onde:

- $V(E(F))$ – Incerteza da estimativa $E(F)$
 $V(F)$ – Variância da função teste F
 n – Número de estados amostrados

Por sua vez a variância da função teste é dada pela Equação 2.7.

$$V(F) = \frac{1}{n} \sum_{i=1}^n [F(x(i)) - E(F)]^2 \quad \text{Equação 2.7}$$

Onde:

- $F(x(i))$ – Função teste relativa a amostra i

Através da Equação 2.6 pode-se definir os limites superior e inferior da estimativa.

Os programas de análise de confiabilidade para sistemas de grande porte diferem basicamente no método de seleção dos estados. Existem dois métodos básicos: a *Enumeração de Estados* e a *Simulação de Monte Carlo*.

O método de enumeração de estados tem como idéia básica a seleção de um conjunto de estados segundo níveis de profundidade de falha. Considere por exemplo uma rede com 40 circuitos, cada um deles representado por um modelo de *Markov* a dois estados. Essa rede apresentaria um total de 2^{40} estados possíveis, ou seja, mais de um trilhão de estados. A idéia

desse método é, então, enumerar um subconjunto dentro de todos os estados possíveis e calcular os limites mínimos e máximos das estimativas. Dessa maneira, se o subconjunto enumerado for escolhido adequadamente, ou seja, um conjunto menor de estados cuja contribuição for considerada relevante seja selecionado para enumeração, excelentes estimativas podem ser obtidas de modo desde que os estados não analisados tenham contribuição desprezível na obtenção dos índices calculados. Este método é mais utilizado em estudos de Confiabilidade de Transmissão, onde, usualmente, se limita a lista de contingências às saídas forçadas de primeira e segunda ordem, ou seja, às contingências simples e duplas. Isso ocorre devido ao fato de que, se o critério de seleção de estados se reduzir àqueles mais prováveis, desta forma, um grande número de contingências de transmissão não será considerado.

A Simulação de Monte Carlo consiste basicamente no sorteio dos estados dos componentes do sistema a partir de suas distribuições de probabilidades conjuntas. Quando a Análise de Confiabilidade é composta (geração/transmissão), esse método se mostra mais adequado devido às probabilidades de falha dos geradores serem relativamente bastante superiores às dos circuitos, por exemplo, para se dobrar a precisão na estimativa de um índice é necessário quadruplicar o tamanho da amostra. A maior limitação da simulação de Monte Carlo é a grande dependência entre o esforço computacional e a precisão desejada. Entretanto, sua maior vantagem é que o número de sorteios necessários para estimar os índices de confiabilidade não depende do número de componentes do sistema, isto é, não depende do tamanho do espaço de estados a ser analisado.

2.3.2 ANÁLISE DE CONFIABILIDADE APLICADA AO SISTEMA ELÉTRICO BRASILEIRO - O SISTEMA NH2

Nesta dissertação, os cálculos de confiabilidade serão realizados utilizando o sistema NH2. Atualmente este programa é utilizado para efetuar a análise de confiabilidade do sistema elétrico brasileiro.

O sistema computacional NH2 (Sistema Computacional para Análise da Confiabilidade de Geração e Transmissão), desenvolvido pelo CEPEL (Centro de Pesquisas de Energia Elétrica), é um sistema computacional para a análise probabilística do desempenho de sistemas de potência, incluindo a avaliação de confiabilidade de geração e transmissão de sistemas hidrotérmicos de grande porte. Neste sistema, podem ser diretamente modelados aspectos elétricos e indiretamente os aspectos energéticos do sistema. Diversas incertezas probabilísticas podem ser tratadas, incluindo as da rede de transmissão, parque gerador, cargas, etc. Em resumo, o NH2 oferece um conjunto de funções básicas voltadas à análise probabilística de

sistemas de potência sob o enfoque de adequação, podendo-se salientar, o cálculo do fluxo de potência determinístico e probabilístico, a análise de contingências, e a análise de confiabilidade.

Este sistema utiliza o conceito de cenário que se caracteriza por uma descrição completa do sistema de potência, das cargas e das condições hidrológicas, em um determinado instante no tempo, que determinam seu comportamento sob o ponto de vista da análise de confiabilidade, ou seja, sua utilização permite calcular os índices de confiabilidade do sistemas sob o enfoque estático, que permite a visualização de sua adequação.

Existem dois grandes agrupamentos de índices calculados pelo NH2 que são: os índices de problemas e os índices de corte de carga.

Os índices de problemas se referem a violações de restrições operativas para o despacho do caso em estudo. Para estes índices são calculadas as probabilidades de violação (SPP) de cada tipo de grandeza monitorada, como por exemplo, o isolamento de carga, o carregamento em circuitos ou a tensão nas barras.

Os índices de corte de carga traduzem a severidade dos problemas no sistema em termos do corte de carga necessário para trazer as grandezas violadas para dentro de seus limites, sendo os mais tradicionais pertencentes à análise de confiabilidade. No NH2 são calculados entre outros a LOLP, a LOLE, a EPNS, a EENS, a LOLF e a LOD, desagregados a nível de sistema, áreas e barras de carga.

Além dos índices tradicionais de confiabilidade e fluxo de potência probabilístico, o programa fornece também um conjunto de informações adicionais ao final de cada análise, tais como casos mais severos, estatísticas de violação por circuito e por barras, estatística das perdas por área etc.

A diferença entre as probabilidades de corte de carga e de problemas no sistema fornece uma avaliação da eficiência das medidas corretivas aplicadas ao sistema, com relação ao despacho especificado no caso base.

O NH2 também fornece distribuições de probabilidades de variáveis de interesse, tais como: carregamento em circuitos; tensão em barras; intercâmbio entre áreas; perdas em áreas; etc. O cálculo destas distribuições de probabilidades é usualmente efetuado em programas de fluxo de potência probabilístico. Observa-se que, por ser dotado de um modelo de medidas corretivas, o NH2 consegue calcular estas distribuições tanto antes quanto depois da atuação dos controles para eliminação das violações, sendo possível, por exemplo, a obtenção de distribuições de probabilidades de cortes de carga em barras.

O NH2 permite também avaliar o impacto sobre os valores médios e distribuições de probabilidades não apenas de incertezas nos patamares de carga, como também da curva de carga do sistema.

A contribuição relativa das falhas de geração, transmissão e compostas tem como resultado a análise de confiabilidade composta, onde é possível obter estimativas dos índices de confiabilidade de geração (representação em barra única) e de transmissão (sem as indisponibilidades do sistema de geração). A diferença entre o índice de confiabilidade do sistema global e os índices de geração e transmissão reflete a contribuição dos efeitos compostos nos índices finais, isto é, a parcela que não é calculada pelos modelos de confiabilidade de geração e de transmissão.

São também fornecidos como subprodutos do modelo de medidas corretivas, baseado em otimização, índices de sensibilidades que medem a variação do valor esperado da energia não suprida no sistema com relação a variações incrementais de demanda em barras e capacidades de transmissão. Estes índices, também denominados de custos marginais de potência, são importantes tanto a nível de planejamento, por permitir uma comparação coerente entre os reforços de geração e transmissão, como também a nível de operação, por estarem relacionados com os custos marginais do sistema.

Os principais dados necessários para utilização deste sistema são as características das barras do sistema, os dados dos circuitos e dos geradores. Além deste conjunto, dependendo do cálculo desejado, devem ser fornecidos mais dados, como por exemplo, informações sobre os níveis de tensão, os dados estocásticos dos circuitos e dos geradores, e listas de contingências dos circuitos e dos geradores.

Para execução do NH2 devem ser fornecidos vários dados sobre o sistema, também conhecido como caso base. Entre estes os principais são os dados das barras e dos circuitos. Normalmente estes dados são utilizados como referência para indicação da complexidade do sistema que está sendo analisado. Completando este conjunto de dados existe uma grande quantidade de dados que se destinam a execução das diferentes funcionalidades do programa, como por exemplo os dados sobre as usinas e de tensão, para a execução do fluxo de potência, as listas de contingências para execução da análise de confiabilidade por enumeração, ou os dados estocásticos para execução da análise de confiabilidade com simulação de Monte Carlo.

Na análise de confiabilidade composta de geração/transmissão, efetuada pelo NH2, cada estudo é efetuado utilizando simulação de Monte Carlo. Em cada estudo são sorteadas no mínimo 5 configurações prováveis de operação do sistema, onde, cada configuração é definida pelo sorteio do estado (em operação ou em reparo) de cada um dos equipamentos, geradores ou

linhas de transmissão. Para cada configuração é executado o fluxo de potência ótimo, para determinar se a configuração permite atender a demanda ou não. No caso de não atender, é calculado o corte de carga necessário para recompor o equilíbrio entre carga e geração.

O fluxo de potência ótimo é um problema de programação não linear de grande porte resolvido utilizando algoritmo de pontos interiores, ou seja, um algoritmo de alta complexidade e que isoladamente já implica em um grande esforço computacional.

Deve ser ressaltado que no NH2 o conceito de usina difere do conceito comum em que este termo está associado ao espaço físico onde se encontram instalados os geradores de energia. Este é o conceito utilizado nos estudos energéticos, onde a capacidade da usina está associada ao somatório das capacidades de todos geradores, independente de suas características próprias. Para este sistema uma usina é caracterizada por um conjunto de geradores de mesmas características ligados a uma mesma barra do sistema, ou seja, em uma usina representada no estudo energético pode ser representada mais do que uma usina no NH2.

3 TÉCNICAS DE ANÁLISE E AGRUPAMENTO DE DADOS

O objetivo deste capítulo é prover embasamento teórico sobre as técnicas de análise e agrupamento de dados. Serão descritos os algoritmos k-Means, PAM, FCM, GG e GK, que foram avaliados para formação do processo de agrupamento.

Como os algoritmos de agrupamento avaliados não possuem a capacidade de determinar o melhor número de agrupamentos a ser utilizado para um determinado espaço de dados, também serão apresentadas às definições de medidas de validação de agrupamentos. Dentre as medidas de validação apresentadas estão a Compacidade e Separação, o Discriminante Linear de Fisher e sua versão estendida, o Contrate Entre Classes e a Silhueta Média Global.

3.1 ANÁLISE DE DADOS

O termo *Dado* significa informação efetiva. Ele é originado do latim *Datum*, que significa ‘um fato único’. (DOWNING (1992), “*Data are factual information. Data is the plural of the word datum, witch means ‘a single fact’*”).

Segundo o dicionário de termos de informática da SUCESU-SC (2004), “*Dado é qualquer tipo de informação (em um processador de texto, programa de imagem, etc.) processada pelo computador*”.

Segundo HÖPPNER (1999), “*Um dado é uma informação arbitrária sobre um estado de um sistema*”. Esta informação pode ser obtida por meio de medições, comparações, índices entre outros. A sumarização dos vários estados possíveis de um sistema descreve um *espaço de estados* ou um *espaço de dados*, onde cada elemento descreve um estado particular do sistema.

A análise de dados tem por objetivo responder a perguntas sobre a informação contida em um espaço de dados, onde cada pergunta formulada tem a capacidade de indicar a forma da resposta que será obtida. A sumarização das várias respostas possíveis para um mesmo processo de análise descreve um *espaço de resultados*. O espaço de resultados deve ser constituído ao menos por dois resultados distintos, para que este agregue informação à análise. Se um espaço de resultados obtido após a análise de dados se refletir em um único resultado, isto significa que o espaço de dados já possui representação para toda a sua informação sem ambigüidade.

Ainda segundo HÖPPNER (1999), “a análise de dados pode ser dividida em quatro níveis de complexidade crescente”. Os níveis de complexidade encontram-se representados na Figura 3.1.

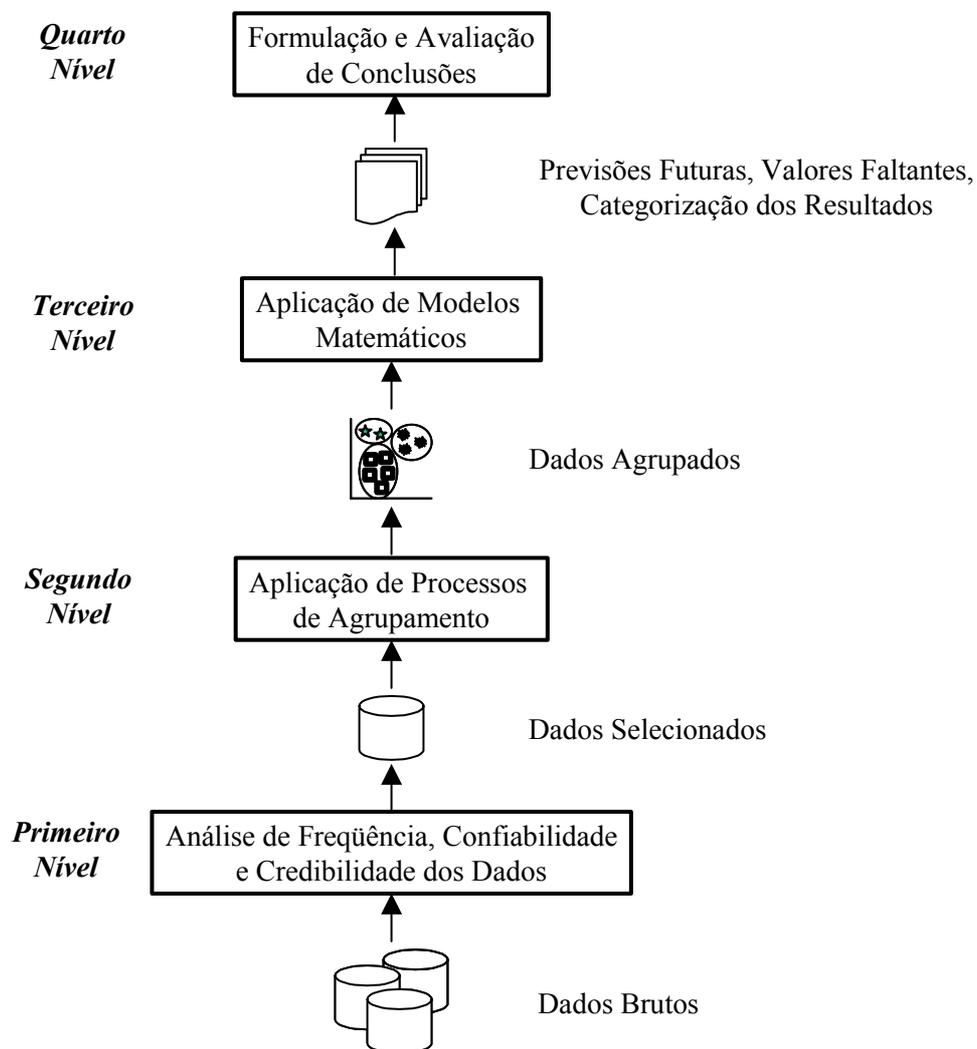


Figura 3.1 – Níveis de complexidade da análise de dados

O primeiro nível consiste na análise de frequência, confiabilidade e credibilidade dos dados. Esta análise leva a identificação de dados espúrios (*outliers*). Também faz parte deste nível a decisão em manter ou eliminar os dados espúrios identificados. No segundo nível os dados são agrupados e estruturados, ou seja, é neste nível que são utilizados os processos de agrupamento para reconhecimento de padrões e classificação. Estes dois primeiros níveis constituem a área de análise exploratória de dados onde a investigação é efetuada sem assumir modelos matemáticos para explicar as características dos dados.

No terceiro nível, os dados são analisados segundo modelos matemáticos. Neste nível é realizada a busca das relações entre os dados de forma quantitativa, isto é, as relações funcionais são determinadas e especificadas.

No quarto nível são formuladas e avaliadas conclusões sobre o espaço de resultados obtido. Entre as conclusões que podem ser obtidas estão as previsões futuras, os valores faltantes ou a categorização dos resultados. Estas conclusões devem conter o julgamento sobre o seu nível de confiabilidade e sobre a veracidade dos modelos obtidos. Além da verificação do modelo criado no terceiro nível, indicando assim se este deve ser reconstruído.

Nos próximos capítulos deste trabalho, a análise de dados estará sendo realizada sobre os dados das séries de potência disponível de usinas hidrelétricas, sem, no entanto, buscar-se um modelo matemático para representar seus resultados.

3.2 ANÁLISE DE AGRUPAMENTOS

A análise de agrupamentos (*clustering techniques*), é uma das ferramentas utilizadas nos processos de análise exploratória de dados para descobrir estruturas e/ou definir subconjuntos, de forma a auxiliar na identificação das informações neles contidas (HALKIDI, 2001; KLAWONN, 2003).

Os problemas de análise de agrupamentos dizem respeito a processos de particionamento de um espaço de dados, visando basicamente formar agrupamentos, cujas características devem ser: (HÖPPNER, 1999).

- Homogeneidade entre os dados pertencentes a um mesmo grupo, ou seja, os dados que pertencem a um mesmo grupo são os mais similares o possível;
- Heterogeneidade entre os agrupamentos, ou seja, os dados que pertencem a agrupamentos distintos devem ser o mais diferentes o possível.

A análise de agrupamentos é um processo não supervisionado, uma vez que não são fornecidos exemplos ou classes predefinidas para indicar o tipo de relações que se está tentando determinar (BERRY e LINOFF, *apud* HALKIDI, 2001). Em contraponto, as técnicas de classificação (*classification*) são procedimentos supervisionados que visam atribuir cada amostra do espaço de dados a uma categoria pré-definida (FAYYAD *et al*, *apud* HALKIDI, 2001). A análise de agrupamentos pode produzir as categorias iniciais a serem utilizadas nos processos de classificação.

A aplicação da análise de agrupamentos pode levar a definição de partições diferentes do espaço de dados, dependendo das características dos dados selecionados, do algoritmo utilizado e de outros critérios que devem ser adotados. A Figura 3.2 (HALKIDI, 2001) apresenta os passos básicos para avaliação de um problema de análise de agrupamentos, conforme descrito a seguir.

- **Passo 1 – Seleção de Características.**

Neste passo é selecionado o espaço de dados que será analisado pelo algoritmo de agrupamento. Esta etapa pode ser subdividida em: *aquisição de dados*, onde será realizada a seleção das observações (m linhas) e das variáveis (n colunas) que irão compor o espaço de dados (matriz $m \times n$); e *pré-processamento* dos dados, onde, caso necessário, é realizada a mudança de escala das variáveis, a normalização e a extração de características.

O espaço de dados resultante desta etapa deve conter o máximo de informação relativa ao problema que está sendo analisado.

- **Passo 2 – Seleção do Algoritmo de Agrupamento.**

A seleção do algoritmo de agrupamento é o passo de maior peso na definição de uma boa partição do espaço de dados. As principais características dos algoritmos de agrupamento são: a *medida de proximidade*, que quantifica a similaridade entre os dados; e o *critério de agrupamento*, que define se o algoritmo é apropriado para o tipo de dados e o tipo de agrupamentos que se deseja obter. Outras formas de caracterização são a *eficiência*, a *robustez* e o *grau de complexidade computacional*.

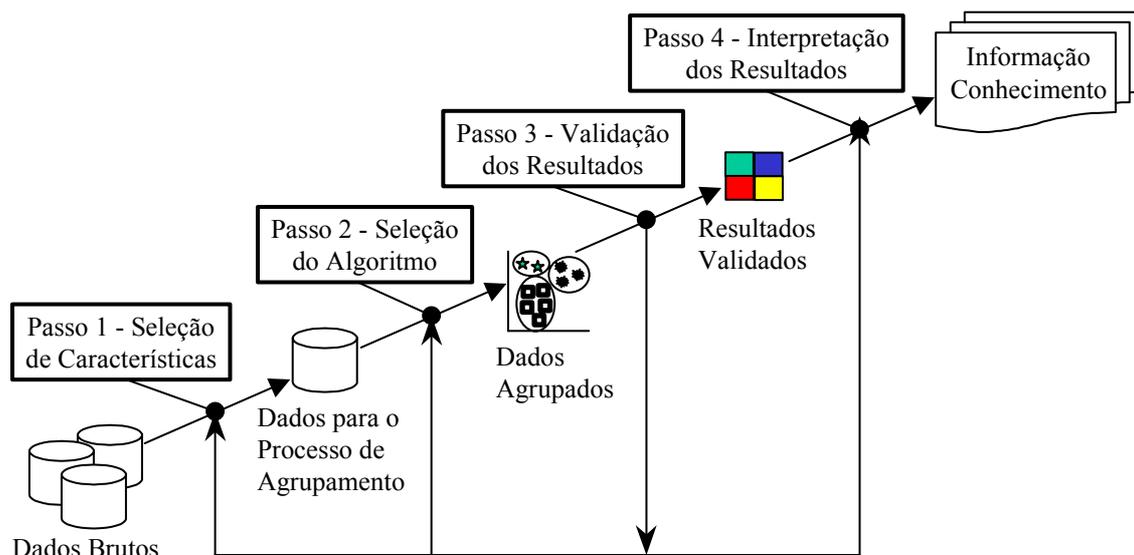


Figura 3.2 – Passos da análise de agrupamentos

- **Passo 3 – Validação dos Resultados.**

Existe um grande número de técnicas de avaliação das partições resultantes da aplicação dos algoritmos de agrupamento. Cada uma destas técnicas podem ser utilizadas para verificar se os agrupamentos obtidos são satisfatórios. Caso esta validação indique que a partição realizada não corresponde ao desejado, retorna-se a um dos passos anteriores, de forma a melhorar as características selecionadas para compor o espaço de dados ou a aplicar um algoritmo que seja mais adequado.

- **Passo 4 – Interpretação dos Resultados.**

Nesta etapa, os especialistas avaliam o espaço de resultados final, podendo ainda aplicar outras técnicas de análise para extrair informação e conhecimento, como por exemplo, a análise estatística.

3.2.1 ALGORITMOS DE AGRUPAMENTO

Atualmente, uma grande variedade de algoritmos de agrupamento pode ser encontrada na literatura, apresentando uma série de suposições sobre o espaço de dados e os tipos de agrupamentos que devem ser obtidos, como por exemplo, em: BARALDI (1999), DAVÉ (1997), GOWDA (1992), HALKIDI (2001), HÖPPNER (1999), JAIN (1999), KAUFMAN (1990) e VAN DER LAAN (2002).

De forma geral os algoritmos de agrupamento podem ser classificados de acordo com o tipo dos dados que estão sendo agrupados, o critério de agrupamento que define a similaridade entre os dados, ou pelos conceitos fundamentais que são utilizados no algoritmo.

Segundo o método empregado no algoritmo, tem-se a seguinte classificação (JAIN, 1999):

- Algoritmos de Particionamento: Estes algoritmos, também são conhecidos como não-hierárquicos, procuram determinar um número inteiro de partições do espaço de dados que maximize um critério predeterminado, por meio de um processo iterativo.
- Algoritmos Hierárquicos: Este tipo de algoritmo se subdivide em algoritmos aglomerativos, que partem de pequenos agrupamentos procurando por agrupamentos maiores, e algoritmos divisivos, que partem de um grande grupo visando obter agrupamentos menores. Seu resultado é uma árvore de agrupamentos chamada de dendograma, que mostra como os agrupamentos estão relacionados. A

seleção de um determinado nível do dendograma indica o número de agrupamentos que será adotado.

A classificação dos algoritmos também pode ser feita segundo o tipo de variáveis que compõem o espaço de dados, onde:

- Algoritmos Estatísticos: Baseados nos conceitos da análise estatística, estes algoritmos só têm capacidade para lidar com variáveis numéricas. Caso seja necessário utilizar outro tipo de dados, estes devem passar por um pré-processamento que permita sua transformação em variáveis numéricas. Esta restrição baseia-se no fato de que estes algoritmos utilizam medidas de distância para determinar os agrupamentos.
- Algoritmos Conceituais: São algoritmos que têm a capacidade de agrupar dados categóricos, criando agrupamentos baseados nos conceitos existentes no espaço de dados.

Os algoritmos podem ainda ser classificados segundo a forma com que estes lidam com as incertezas provenientes da sobreposição dos agrupamentos dentro do espaço de dados.

- Algoritmos de Agrupamento Rígidos (*Crisp Clustering*): Neste tipo de algoritmo um dado só pode pertencer a um determinado grupo, não permitindo desta forma que ocorra sobreposição dos agrupamentos.
- Algoritmos de Agrupamento Nebulosos (*Fuzzy Clustering*): Neste tipo de algoritmo são aplicados os conceitos de lógica nebulosa introduzidos por Zadeh, em 1965. Desta forma estes algoritmos consideram que um dado pode pertencer a mais de um grupo, segundo um grau de inclusão, permitindo assim que dados que apresentam sobreposição sejam tratados de forma mais realista.

A seguir serão apresentados os algoritmos de agrupamento rígidos *k-Means* e *Partitioning Around Medoids* (PAM), os algoritmos nebulosos *Fuzzy C-Means* (FCM), Gath-Geva (GG) e Gustafson-Kessel (GK) os quais foram avaliados durante o desenvolvimento desta dissertação.

3.2.1.1 ALGORITMO K-MEANS

O algoritmo k-Means foi definido por MacQueen em 1967 (*apud* HALKIDI, 2001a). Ele é um algoritmo de particionamento rígido do espaço de dados, e se tornou popular por não impor restrições ao conjunto de amostras, podendo ser aplicado a qualquer quantidade de dados. (ROSS, 1997 *apud* FRANCO, 2002)

No k-Means, o grau de pertinência de uma amostra a um agrupamento é binária, ou seja, os graus de pertinência de uma amostra i a um agrupamento j são: $u_{ij} = 0$, quando a amostra i não pertence ao agrupamento j , ou $u_{ij} = 1$, quando a amostra i pertence integralmente ao agrupamento j .

Por ser um algoritmo de particionamento rígido o k-Means obedece as seguintes propriedades para um conjunto de agrupamentos $\mathcal{A} = \{a_1, \dots, a_c\}$ de c partições.

$$\bigcup_{j=1}^c a_j = X \quad \text{Equação 3.1}$$

$$a_i \cap a_j = \phi \quad \forall i \neq j \quad \text{Equação 3.2}$$

$$\phi \subset a_j \subset X \quad \forall j \in \{1, 2, \dots, c\} \quad \text{Equação 3.3}$$

A Equação 3.1 indica que o espaço de dados é formado pela união de todos os agrupamentos formados pelo algoritmo. Por sua vez a Equação 3.2 indica que nenhuma amostra do espaço de dados pode pertencer a mais de um agrupamento e a Equação 3.3 mostra que nenhum agrupamento pode ficar vazio ou conter todos os elementos.

O algoritmo k-Means pode ser visto como uma estratégia para minimização da função objetivo apresentada na Equação 3.4, aplicado a um espaço de dados $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ de n amostras, de p dimensões, ou seja, onde cada elemento \mathbf{x}_j possui p coordenadas ou características.

$$\min J(\mathbf{U}, \mathbf{M}) = \sum_{i=1}^n \sum_{j=1}^c (u_{ij}) d^2(\mathbf{x}_i, \mathbf{m}_j) \quad \text{Equação 3.4}$$

Onde:

- J – Função objetivo
- \mathbf{U} – Matriz de pertinências rígida
- u_{ij} – Grau de pertinência rígida da amostra i no agrupamento j
- \mathbf{M} – Conjunto dos centros dos agrupamentos
- \mathbf{m}_j – Centro do agrupamento j
- \mathbf{x}_i – Amostra i
- c – Número de agrupamentos
- n – Número de amostras no espaço de dados
- $d(\mathbf{x}_i, \mathbf{m}_j)$ – Distância entre a amostra i e o centro do agrupamento j

Sendo a minimização da função objetivo sujeita as seguintes restrições:

$$0 < \sum_{i=1}^n u_{ij} < n \quad \forall j \in \{1, 2, \dots, c\} \quad \text{Equação 3.5}$$

$$\sum_{j=1}^c u_{ij} = 1 \quad \forall i \in \{1, 2, \dots, n\} \quad \text{Equação 3.6}$$

A Equação 3.5 indica que nenhum agrupamento pode permanecer vazio e que o número máximo de elementos em um agrupamento só pode ser $n-1$. Já a Equação 3.6, associada ao fato de que u_{ij} só pode assumir os valores 0 e 1, garante que uma amostra só pode pertencer a um único agrupamento.

Na Equação 3.4, o termo $d(\mathbf{x}_i, \mathbf{m}_j)$ representa a distância entre o centro do agrupamento j e a amostra i , que é calculada utilizando-se a *distância Euclidiana*, dada pela Equação 3.7.

$$d(\mathbf{x}_i, \mathbf{m}_j) = \|\mathbf{x}_i - \mathbf{m}_j\|^2 \quad \text{Equação 3.7}$$

Uma vez que o particionamento é realizado de acordo com a disposição das amostras no *Espaço Euclidiano*, cada um dos c agrupamentos assume a forma de uma hiperesfera no \mathfrak{R}^p , todas com aproximadamente o mesmo volume. Sendo que cada agrupamento é representada por um centro pertencente ao conjunto $\mathbf{M} = \{\mathbf{m}_1, \dots, \mathbf{m}_c\}$.

O k-Means é um algoritmo apropriado para ser aplicado quando a distribuição das amostras no espaço de dados pode ser aproximada por hiperesferas, sendo intrinsecamente dependente da disposição das amostras no espaço. Ou seja, quanto mais próximo do formato de hiperesferas é a distribuição das amostras no espaço de dados, mais coesos serão os agrupamentos gerados. Para distribuições com outros formatos, os agrupamentos formados podem contêm amostras que deveriam pertencer a outro(s) agrupamento(s), comprometendo a qualidade dos agrupamentos determinados.

Mais algumas observações devem ser discutidas com relação ao número de agrupamentos c e ao fator de parada do algoritmo ξ .

O valor do número de agrupamentos c deve variar entre $2 \leq c < n$, dado que para $c = 1$, todos os pontos pertencem à mesma categoria e para $c = n$, cada ponto pertence à sua própria categoria. Nestes dois casos, não existe de fato um problema de análise de agrupamentos.

No Algoritmo k-Means, é o fator de parada $\xi \geq 0$ que determina o seu término. Quando a matriz \mathbf{U} tem uma variação menor que ξ entre duas iterações, o deslocamento dos centros dos

agrupamentos no espaço é muito pequeno, não havendo uma mudança significativa no formato das partições. Isto indica que o algoritmo se estabilizou e uma solução eficiente foi encontrada.

Como o k-Means é um método rígido, o que implica que sua matriz U é rígida. Neste método o critério de parada mais utilizado é a diferença entre a matriz U em duas iterações, sendo que esta diferença deve ser igual a 0. Este critério de parada indica que os centros não foram deslocados ou não tiveram um deslocamento significativo.

A seguir é apresentado o algoritmo de processamento do k-Means:

Passo 1. Para um espaço de dados $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, fixar:

O número de agrupamentos $c \in [2, n)$

O número máximo de iterações T

O fator de parada ξ

Passo 2. Iniciar aleatoriamente a matriz $U^{(0)}$, obedecendo as restrições apresentada na Equação 3.5 e na Equação 3.6

Passo 3. Para $\tau = 0, 1, 2, \dots, T$

Passo 3.a. Incrementar τ

Passo 3.b. Calcular os centros dos c agrupamentos na iteração τ , segundo a Equação 3.8.

$$\mathbf{m}_j^{(\tau)} = \frac{\sum_{i=1}^n (u_{ij}^{(\tau-1)}) \mathbf{x}_i}{\sum_{i=1}^n (u_{ij}^{(\tau-1)})} \quad \text{Equação 3.8}$$

Passo 3.c. Calcular as distâncias entre cada amostra \mathbf{x}_i e cada centro \mathbf{m}_j , segundo a Equação 3.14 ou Equação 3.15.

Passo 3.d. Atualizar a matriz de pertinências $U^{(\tau)}$ para os novos centros dos agrupamentos.

Passo 3.e. Se $\tau = T$

Termina o processamento

Senão

Compara $U^{(\tau)}$ e $U^{(\tau-1)}$

Se $\|U^{(\tau)} - U^{(\tau-1)}\| \leq \xi$

Termina o processamento

Senão

Retorna ao Passo 3.a.

3.2.1.2 PARTITIONING AROUND MEDOIDS (PAM)

Definido por Kauffman e Rousseeuw (1990), o algoritmo *Partitioning Around Medoids*, também conhecido como *PAM*, efetua a partição do espaço de dados em c agrupamentos, definidos a priori, procurando entre as n amostras as que melhor representem cada um dos agrupamentos. Os autores denominam *medoids* cada uma das amostras escolhidas como representativas dos agrupamentos. Na literatura encontra-se também a denominação centrótipos (*centrotypes*), que será a denominação adotada neste trabalho.

Para realizar o particionamento do espaço de dados, o PAM realiza uma busca exaustiva pela troca de cada um dos c centrótipos previamente selecionados por uma das $(n - c)$ amostras não selecionadas, de forma a minimizar as dissimilaridades entre os c centrótipos e os membros dos agrupamentos.

Este algoritmo exige um pré-processamento dos dados, de forma a obter uma matriz de dissimilaridades entre as n amostras. Cada elemento desta matriz é determinado através do cálculo da distância entre duas amostras distintas. Desta forma, a matriz de dissimilaridades é uma matriz quadrada ($n \times n$) e simétrica, onde a diagonal principal é composta de zeros, pois a dissimilaridade entre uma amostra e ela mesma não existe.

O fato de realizar o particionamento do espaço de dados tomando como base uma matriz de dissimilaridades pré-calculada torna este algoritmo mais flexível com relação aos dados, pois, para cada tipo de dado, pode ser utilizada uma forma distinta de cálculo da distância entre as amostras.

Outra característica deste algoritmo é ser independente da ordem com que os dados são apresentados, pois na matriz de dissimilaridades as distâncias entre as amostras são calculadas duas a duas para todo o espaço de dados.

Por outro lado, a desvantagem deste algoritmo está na realização da busca exaustiva das partições, pois para um espaço de dados extenso, este processo pode se tornar muito custoso em termos de tempo de processamento. Para este tipo de espaço de dados o cálculo da matriz de dissimilaridade também pode se tornar um processo custoso.

O objetivo deste algoritmo é, definir um conjunto de centrótipos $\mathbf{M} = \{\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_c\}$ a partir do espaço de dados $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, onde $\mathbf{M} \subset \mathbf{X}$, que minimize a função objetivo apresentada na Equação 3.9.

$$F(\mathbf{M}) = \sum_{i=1}^n \min_{\substack{j=1, \dots, c \\ \mathbf{x}_i \neq \mathbf{m}_j}} D(\mathbf{x}_i, \mathbf{m}_j) \quad \text{Equação 3.9}$$

Onde:

| | |
|---------------------------------|--|
| F | – Função objetivo |
| \mathbf{M} | – Conjunto dos centrótipos |
| $D(\mathbf{x}_i, \mathbf{m}_j)$ | – Dissimilaridade entre a amostra \mathbf{x}_i e o centrótipo \mathbf{m}_j |
| \mathbf{x}_i | – Amostra i |
| \mathbf{m}_j | – Centrótipo j |
| c | – Número de agrupamentos |
| n | – Número de amostras |

O algoritmo do PAM é composto de duas fases, onde, a primeira fase é chamada de Construção (*Build*) e a segunda fase é chamada de Permuta (*Swap*).

Na fase de Construção, são selecionados os centrótipos iniciais, como apresentado no algoritmo abaixo:

Passo 1. Para um espaço de dados $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ fixar:

O número de agrupamentos $c \in [2, n)$

Calcular a matriz de dissimilaridades D , utilizando a forma de cálculo da distância que mais se ajuste ao tipo de dados que está sendo analisado.

Passo 2. Determinar a amostra mais central dos dados. Esta amostra é representada pela amostra que possui a menor soma das dissimilaridades em relação as demais amostras. Esta será a primeira amostra selecionada como centrótipo.

Passo 3. Para cada uma das demais amostras \mathbf{x}_g , onde \mathbf{x}_g não é um centrótipo

Passo 3.1. Será selecionada como centrótipo a amostra que minimizar a função objetivo da Equação 3.9.

Passo 3.2. Se já foram selecionados c centrótipos

Termina o processamento

Senão

Retorna ao Passo 3.

Na fase de Permuta, a minimização da função objetivo da Equação 3.9 é analisada após a permuta de cada um dos centrótipos selecionados por todas as amostras não selecionadas. Caso o resultado da Equação 3.9 após a permuta seja menor que o resultado obtido com o centrótipo anteriormente selecionado, este é substituído pela nova amostra. Este processo é realizado até que a função objetivo não seja mais alterada. O algoritmo desta fase é apresentado abaixo:

Passo 1. Para um espaço de dados $X = \{x_1, x_2, \dots, x_n\}$ e um conjunto inicial de centrótipos

$$M = \{m_1, m_2, \dots, m_c\}.$$

Passo 2. Para $i = 1, 2, \dots, n$

Passo 2.1. Selecionar x_i onde $x_i \notin \{m_1, \dots, m_c\}$

Passo 2.2. Para $j = 1, 2, \dots, c$

Passo 2.2.1. Selecionar m_j , onde $m_j \in \{m_1, \dots, m_c\}$.

Passo 2.2.2. Trocar x_i e m_j e calcular o valor da função objetivo da Equação 3.9.

Passo 2.2.3. Se o valor da função objetivo diminuir

Efetivar a troca x_i e m_j .

Retornar ao Passo 2.2.

Senão

Se todas as amostras $x_i \in X$ e todos os centrótipos $m_j \in M$ foram selecionados.

Terminar o processo.

Senão

Retornar ao Passo 2.2.

3.2.1.3 ALGORITMO FUZZY C-MEANS (FCM)

O algoritmo de agrupamento nebuloso *Fuzzy C-Means*, também conhecido como *FCM*, foi definido por Bezdek em 1981. Ele é uma extensão do algoritmo de agrupamento rígido *Hard C-Means*, ou *HCM*, definido por Ruspini em 1969 (*apud* HÄRDLE, 2004). A principal contribuição de Bezdek foi a introdução do coeficiente de nebulização m (*fuzzifier*) na função objetivo do algoritmo.

No HCM, assim como no k-Means, o grau de pertinência de uma amostra a um agrupamento é binária, ou seja, os graus de pertinência de uma amostra i a um agrupamento j são: $u_{ij} = 0$, quando a amostra i não pertence ao agrupamento j , ou $u_{ij} = 1$, quando a amostra i pertence integralmente ao agrupamento j .

O FCM possibilita que o grau de pertinência de cada amostra i a um agrupamento j pertença ao intervalo $0 \leq u_{ij} \leq 1$, desta forma cada amostra pode pertencer a todos os agrupamentos, com graus de pertinência distintos, sendo que usualmente uma amostra i é alocada ao agrupamento j onde o seu grau de pertinência u_{ij} é máximo.

Este algoritmo pode ser visto como uma estratégia para minimização da função objetivo apresentada na Equação 3.10, aplicado a um espaço de dados $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ de n amostras, de p dimensões, ou seja, cada elemento \mathbf{x}_i possui p coordenadas ou características.

$$\min J(\mathbf{U}, \mathbf{M}) = \sum_{i=1}^n \sum_{j=1}^c (u_{ij})^m d^2(\mathbf{x}_i, \mathbf{m}_j) \quad \text{Equação 3.10}$$

Onde:

| | |
|---------------------------------|---|
| J | – Função objetivo |
| \mathbf{U} | – Matriz de pertinências |
| u_{ij} | – Grau de pertinência da amostra i no agrupamento j |
| \mathbf{M} | – Conjunto dos centros dos agrupamentos |
| \mathbf{m}_j | – Centro do agrupamento j |
| \mathbf{x}_i | – Amostra i |
| c | – Número de agrupamentos |
| n | – Número de amostras no espaço de dados |
| m | – Coeficiente de nebulização |
| $d(\mathbf{x}_i, \mathbf{m}_j)$ | – Distância entre a amostra i e o centro do agrupamento j |

Sendo a Equação 3.10 sujeita as seguintes restrições:

$$0 < \sum_{i=1}^n u_{ij} < n \quad \forall j \in \{1, 2, \dots, c\} \quad \text{Equação 3.11}$$

$$\sum_{i=1}^n \sum_{j=1}^c u_{ij} = n \quad \text{Equação 3.12}$$

$$\sum_{j=1}^c u_{ij} = 1 \quad \forall i \in \{1, 2, \dots, n\} \quad \text{Equação 3.13}$$

A restrição apresentada na Equação 3.11 visa garantir que nenhum agrupamento poderá ficar vazio, ou seja, que o resultado do algoritmo será o particionamento do espaço de dados em c agrupamentos. Já a Equação 3.12 visa garantir que todas as amostras são completamente consideradas nos agrupamentos. Por sua vez, a Equação 3.13 visa garantir que o somatório dos graus de pertinência de cada uma das amostras em todos os agrupamentos equivale à unidade.

Algumas observações devem ser feitas a respeito do número de agrupamentos c , da distância $d(\mathbf{x}_i, \mathbf{m}_j)$ e da constante de nebulização m apresentadas na Equação 3.10.

O FCM efetua o particionamento do espaço de dados em um número inteiro fixo de agrupamentos c definido à priori, onde: $c \in [2, n)$. Os extremos deste intervalo, $c = 1$ e $c = n$,

não podem ser incluídos, pois, para $c = 1$, o espaço de dados se refletiria em um único agrupamento, ou seja, não teria partições; e para $c = n$ cada agrupamento possuiria somente uma amostra, que coincidiria com o centro do agrupamento. Além disso, para $c = n$, o grau de pertinência de cada amostra em seu agrupamento seria 1, o que não é permitido pelas restrições da Equação 3.10.

Não foram encontradas na literatura informações suficientes sobre como definir o número máximo de agrupamentos ($c_{máx}$) a ser adotado. Segundo Pal e Bezdek (*apud* FRANCO, 2000), “existe um pequeno direcionamento na literatura a respeito $c_{máx}$. Uma regra simples que muitos investigadores utilizam é $c_{máx} \leq \sqrt{n}$ ”.

Na Equação 3.10, o termo $d(\mathbf{x}_i, \mathbf{m}_j)$ representa a distância entre o centro do agrupamento j e a amostra i , que é calculada utilizando-se a *distância Euclidiana*, dada pela Equação 3.14.

$$d(\mathbf{x}_i, \mathbf{m}_j) = \|\mathbf{x}_i - \mathbf{m}_j\|^2 \quad \text{Equação 3.14}$$

Portanto, neste algoritmo o particionamento é realizado de acordo com a disposição das amostras no *Espaço Euclidiano*, onde cada um dos c agrupamentos assume a forma de uma hipersfera no \mathfrak{R}^p , todas com aproximadamente o mesmo volume.

Como \mathbf{x}_i e \mathbf{m}_j são vetores compostos de p dimensões, a Equação 3.14 pode ser reescrita em função de suas dimensões resultando na Equação 3.15.

$$d(\mathbf{x}_i, \mathbf{m}_j) = \left(\sum_{k=1}^p (x_{i,k} - m_{j,k})^2 \right)^{1/2} \quad \text{Equação 3.15}$$

Ainda referindo-se à Equação 3.10, deve ser salientado que a constante m confere ao algoritmo uma medida de nebulosidade dos agrupamentos, ou seja, o quanto os agrupamentos podem se sobrepor. Esta constante deve ser definida no intervalo $1 < m < \infty$. Sendo que quando $m \rightarrow 1$, a matriz de pertinências \mathbf{U} tende a se igualar com a matriz de pertinências rígida. Valores baixos de m devem ser escolhidos quando o espaço de dados possui agrupamentos bem definidos e separados, sendo que neste caso os métodos de agrupamento rígidos podem ser uma boa escolha. Já quando $m \rightarrow \infty$, os graus de pertinência de cada amostra tendem a se igualar, tendendo a assumir o valor $1/c$. Quanto maior o valor de m , mais nebulosa será a partição obtida.

Segundo HÖPPNER (1999), “é necessário experiência para tomar esta decisão (escolher o valor de m). Além do mais, um espaço de dados pode conter agrupamentos com diferentes separabilidades. Um escolha comum da constante de nebulização é $m = 2$ ”.

Segundo FRANCO (2000), “não existe uma regra para definir o valor de m a ser usado, dependendo apenas de quão nebuloso ou rígido deseja-se que o particionamento do sistema seja. Os valores de m mais comumente usados são 1,25 e 2”.

A seguir é apresentado o algoritmo de processamento do FCM:

Passo 1. Para um espaço de dados $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, fixar:

O número de agrupamentos $c \in [2, n)$

O valor para a constante de nebulização $m \in (1, \infty)$

O número máximo de iterações T

O fator de parada ξ

Passo 2. Iniciar aleatoriamente a matriz $\mathbf{U}^{(0)}$, obedecendo as restrições apresentadas na Equação 3.11, Equação 3.12 e Equação 3.13.

Passo 3. Para $\tau = 0, 1, 2, \dots, T$

Passo 3.a. Incrementar τ

Passo 3.b. Calcular os centros dos c agrupamentos na iteração τ , segundo a Equação 3.16.

$$\mathbf{m}_j^{(\tau)} = \frac{\sum_{i=1}^n (u_{ij}^{(\tau-1)})^m \mathbf{x}_i}{\sum_{i=1}^n (u_{ij}^{(\tau-1)})^m} \quad \text{Equação 3.16}$$

Passo 3.c. Calcular as distâncias entre cada amostra \mathbf{x}_i e cada centro \mathbf{m}_j , segundo a Equação 3.14 ou Equação 3.15.

Passo 3.d. Atualizar a matriz de pertinências $\mathbf{U}^{(\tau)}$ para os novos centros dos agrupamentos, segundo a Equação 3.17.

$$u_{ij}^{(\tau)} = \frac{1}{\sum_{k=1, k \neq j}^c \left(\frac{\|\mathbf{x}_i - \mathbf{v}_j^{(\tau)}\|}{\|\mathbf{x}_i - \mathbf{v}_k^{(\tau)}\|} \right)^{\frac{2}{m-1}}} \quad \text{Equação 3.17}$$

Passo 3.e. Se $\tau = T$

Termina o processamento

Senão

Comparar $\mathbf{U}^{(\tau)}$ e $\mathbf{U}^{(\tau-1)}$

Se $\|\mathbf{U}^{(\tau)} - \mathbf{U}^{(\tau-1)}\| \leq \xi$

Termina o processamento

Senão

Retorna ao Passo 3.a.

O FCM se tornou um algoritmo bastante difundido principalmente por apresentar poucas restrições ao espaço de dados. Outras características que o tornaram popular são: sua baixa complexidade e sua implementação simples, além do baixo esforço computacional. (HÖPPNER, 1999)

Sua grande desvantagem está na utilização da *distância Euclidiana* como medida de dissimilaridade entre as amostras. A utilização desta distância torna o algoritmo suscetível a erros na presença de dados espúrios, pois amostras afastadas do conjunto podem distorcer o resultado do particionamento.

3.2.1.4 ALGORITMO GUSTAFSON-KESSEL (GK)

Este algoritmo, também conhecido como GK, foi criado por Gustafson e Kessel. Ele é um algoritmo de agrupamento nebuloso, análogo ao FCM, diferindo deste apenas na métrica da distância utilizada.

No GK a *distância Euclidiana* empregada no FCM é substituída pela *distância de Mahalanobis* dada pela Equação 3.18.

$$d^2(\mathbf{x}_i, \mathbf{m}_j) = (\mathbf{x}_i - \mathbf{m}_j)^T \mathbf{A}_j (\mathbf{x}_i - \mathbf{m}_j) \quad \text{Equação 3.18}$$

Onde:

- $d(\mathbf{x}_i, \mathbf{m}_j)$ – Distância entre a amostra i e o centro do agrupamento j
- \mathbf{m}_j – Centro do agrupamento j
- \mathbf{x}_i – Amostra i
- \mathbf{A}_j – Matriz de norma j

As matrizes de norma (*norm matrices*) \mathbf{A}_j são dadas pela Equação 3.19.

$$\mathbf{A}_j = \sqrt[p]{\det(\mathbf{S}_j)} \cdot \mathbf{S}_j^{-1} \quad \text{Equação 3.19}$$

Onde a matriz de covariância nebulosa \mathbf{S}_j é dada pela Equação 3.20.

$$\mathbf{S}_j = \frac{\sum_{i=1}^n \mu_{ij}^m (\mathbf{x}_i - \mathbf{m}_j)(\mathbf{x}_i - \mathbf{m}_j)^T}{\sum_{i=1}^n \mu_{ij}^m} \quad \text{Equação 3.20}$$

Neste caso, os agrupamentos assumem a forma hiperelipsóides no espaço \mathcal{R}^p tendo aproximadamente o mesmo tamanho e o mesmo número de elementos. (GUSTAFSON, 1979)

Este algoritmo oferece restrições ao conjunto de amostras utilizado, pois o cálculo da distância de Mahalanobis envolve a inversão da matriz de covariância fuzzy \mathbf{S}_j , dada pela Equação 3.20.

Neste caso, para que \mathbf{S}_j não seja singular, o número de amostras n deve ser maior que o número de dimensões dos dados mais 1, ou seja, maior que $p+1$.

Por outro lado, $p+1$ amostras não fornecem uma boa estimativa de \mathbf{S}_j . Como \mathbf{S}_j é matriz simétrica de p^2 elementos, ela contém $p(p-1)/2$ elementos independentes. Assim, estimativas válidas da matriz de covariância nebulosa \mathbf{S}_j só são obtidas utilizando-se um espaço de dados onde o número de amostras é maior que $p(p-1)/2$. (VIVARELLI, 1999)

O método GK é um algoritmo que tem um bom desempenho quando a distribuição das amostras no espaço de dados se aproxima de hiperelipsóides com o mesmo número de pontos. Quando a distribuição não apresenta esta disposição ou quando os hiperelipsóides não contêm o mesmo número de amostras, o método converge lentamente, gerando categorias que pouco têm haver com as classes do problema. Mesmo quando obtém bons resultados, sua convergência é mais lenta que a do método FCM. (FRANCO, 2002)

Uma desvantagem do método é a imposição de restrições ao conjunto de dados, não podendo ser aplicado a dados com muitas características quando a quantidade de amostras é pequena.

3.2.1.5 ALGORITMO GATH-GEVA (GG)

Este algoritmo foi definido Gath e Geva, e também é conhecido como método GG ou método de decomposição de mistura Gaussiana (GMD - Gaussian Mixture Decomposition), também foi criado tendo como base o método FCM. (GATH, 1989)

A métrica de distância que substitui a distância Euclidiana neste algoritmo é a Distância de Gauss, que é calculada como mostra a Equação 3.21.

$$d^2(\mathbf{x}_i, \mathbf{m}_j) = \frac{\sqrt{\det(S_j)}}{P_j} \cdot \exp\left(\frac{1}{2}(\mathbf{x}_i - \mathbf{m}_j)^T A_j (\mathbf{x}_i - \mathbf{m}_j)\right) \quad \text{Equação 3.21}$$

Onde:

- $d(\mathbf{x}_i, \mathbf{m}_j)$ – Distância entre a amostra i e o centro do agrupamento j
- \mathbf{m}_j – Centro do agrupamento j
- \mathbf{x}_i – Amostra i
- A_j – Matriz de norma j
- S_j – Matriz de Covariância Nebulosa j
- P_j – Probabilidade a priori j

A Distância de Gauss é calculada a partir da probabilidade a priori P_j de uma amostra pertencer a uma distribuição normal (agrupamento), ou seja, P_j pode ser interpretado como um parâmetro de tamanho do cluster, que é dada pela Equação 3.22.

$$P_j = \frac{\sum_{i=1}^n \mu_{ij}^m}{\sum_{i=1}^n \sum_{j=1}^c \mu_{ij}^m} \quad \text{Equação 3.22}$$

Neste caso as matrizes de covariância nebulosa S_j também são calculadas como mostra a Equação 3.20, e as matrizes de norma (*norm matrices*) são dadas pela Equação 3.23.

$$A_j = S_j^{-1} \quad \text{Equação 3.23}$$

Os agrupamentos gerados já não têm mais uma forma geométrica definida, sendo estruturas no hiperespaço de tamanhos variados.

O algoritmo GG assume que as amostras são uma mistura de c distribuições Gaussianas normais, sendo c o número de agrupamentos.

Por causa da inclusão da probabilidade a priori no cálculo da distância, os agrupamentos que possuem mais elementos têm maior probabilidade de conter uma amostra que está sendo analisada, o que leva à tendência de um agrupamento ser maior que outro.

Assim como o método GK, o GG também impõem restrições ao conjunto de amostras, pois a distância de Gauss também faz a inversão da matriz de covariância nebulosa S_j .

A presença do termo exponencial no cálculo da distância de Gauss reduz a capacidade do GG de evitar mínimos locais, pois quando a matriz de pertinências é inicializada

aleatoriamente, os centros gerados ficam geralmente muito próximos entre si e do centro global dos dados.

Para evitar este problema, é interessante inicializar o GG sempre pelos centros ao invés de pela matriz de pertinências e certificar-se de que estes centros tenham sido obtidos a partir da execução de outros algoritmos de agrupamento, como o FCM, de forma a obter agrupamentos melhores.

O método GG é um método útil para aprimorar as categorias geradas pelo método FCM. Neste caso, o método é interessante porque consegue aproximar as categorias produzindo formas mais precisas que as hiperesferas obtidas pelo FCM.

Também se deve observar que o método GG também impõe restrições ao conjunto de amostras iguais as impostas pelo GK, por efetuar a inversão da matriz de covariância nebulosa.

3.3 ALGORITMOS SELECIONADOS

Neste trabalho foram avaliados 5 algoritmos de agrupamento distintos, dois rígidos, o k-Means e o PAM, e 3 algoritmos nebulosos, o FCM, o GG e o GK.

Neste trabalho serão necessários dois algoritmos de agrupamento com características distintas.

O primeiro tipo de algoritmo deve trabalhar com dados numéricos, sendo que todos os algoritmos testados possuem esta capacidade. Um vez que o espaço de dados é composto por amostras com a característica de grande sobreposição, decidiu-se utilizar algoritmos de agrupamento nebulosos.

Dentre os algoritmos nebulosos testados foi selecionado o FCM por ser o algoritmo de agrupamento nebuloso menos restritivo em relação ao espaço de dados. Também deve ser ressaltado que o espaço de dados utilizado no processo de agrupamento será composto por 65 amostras de 12 posições, como será discutido com mais detalhes no capítulo 6 (Processos de Agrupamento), o que torna impeditiva a utilização dos algoritmos GG e GK, devido as restrições que estes impõem ao espaço de dados apresentadas anteriormente.

Outra motivação para utilização do FCM foi que este é o algoritmo de agrupamento nebuloso mais popular existindo ampla literatura sobre sua aplicação em diversas áreas de conhecimento.

O segundo tipo de algoritmo deve ser capaz de trabalhar com dados nominais, selecionando entre estes um elemento característico.

Dentre os algoritmos avaliados neste trabalho, o PAM foi selecionado por efetuar a partição do espaço de dados utilizando uma matriz de dissimilaridade, o que o torna independente do tipo de dados existente no espaço de dados.

Ainda seria possível utilizar algoritmos de agrupamento baseados em redes neurais, tais como *Self Organizing Maps* (SOM), também conhecidos como *Mapas de Kohonen*, *Adaptive Resonance Theory* (ART) ou *Learning Vector Quantization* (LVQ), ou ainda em algoritmos evolucionários (*evolutionary algorithms*), tais como Algoritmos Genéticos (*Genetic Algorithms* - GA), Estratégias Evolucionárias (*Evolution Strategies* - ES) ou Programação Evolucionária (*Evolutionary Programming* - EP). (JAIN,1999)

Estas técnicas ficarão indicadas como trabalhos futuros para ambas as etapas do processo de agrupamento.

3.4 MEDIDAS DE VALIDAÇÃO DE AGRUPAMENTOS

Quando se trabalha com espaços de dados de duas ou três dimensões é possível visualizar graficamente a distribuição das amostras, o que possibilita a determinação prévia do número de agrupamentos que se está buscando. Mas mesmo nestes tipos de espaços de dados podem ocorrer casos em que as amostras são muito sobrepostas, dificultando a visualização dos possíveis particionamentos. Quando se analisa espaço de dados com um número maior de dimensões, o processo de visualização se torna ainda mais difícil. (HALKIDI, 2001a)

A forma encontrada para verificação do melhor número de agrupamentos que reflete as características dos dados é a utilização de medidas de validação. As medidas de validação são aplicadas às partições de um espaço de dados produzidas por um algoritmo de agrupamento, a fim de estimar a qualidade dos particionamentos produzidos pelo algoritmo quando o número de agrupamentos e suas condições iniciais são alterados (FRANCO, 2000).

Existe uma vasta literatura sobre medidas de validação, onde podemos citar alguns exemplos: FRANCO (2000), HALKIDI (2001a, 2001b), KAUFFMAN (1999) e PAKHIRA (2003).

3.4.1 COMPACIDADE E SEPARAÇÃO (CS)

A Compacidade e Separação – *CS* (*Compactness and Separation*) é uma medida de validação completa, pois avalia tanto a compacidade dos agrupamentos gerados, ou seja, se cada agrupamento formado possui amostras que encontram-se próximas, como a qualidade da separação entre estes, ou seja, se os centros dos agrupamentos encontram-se o mais afastados o possível. (XIE, 1991 *apud* FRANCO)

Esta medida de validação corresponde a minimizar a função objetivo J , que é a finalidade de vários algoritmos de agrupamento, tais como o k-Means, o FCM, o GG e o GK. A medida CS é dada pela Equação 3.40.

$$CS = \frac{J}{n \cdot (d_{\min})^2} \quad \text{Equação 3.24}$$

Onde:

- J – Função objetivo a ser minimizada. A mesma utilizada no algoritmo de agrupamento.
- d_{\min} – Medida de distância utilizada no algoritmo de agrupamento.
- n – Número de amostras.

Por se tratar de um processo de minimização, tem-se que quanto menor o valor de CS tanto melhor pode ser considerado o resultado do algoritmo de agrupamento.

Como mostrado na Equação 3.24, a CS se altera dependendo do algoritmo de agrupamento que será validado, uma vez que ela está baseada na função objetivo do algoritmo.

Devido a sua implementação, a CS só é uma medida inócua quando o número de agrupamentos c se aproxima do número de amostras n . Mas isto não chega a ser um problema, pois, normalmente, procura-se por um número de agrupamentos bem menor que o número de amostras.

3.4.2 DISCRIMINANTE LINEAR DE FISHER (FLD)

O Discriminante Linear de Fisher (*Fisher Linear Discriminant– FLD*) (BISHOP, 1995; DUDA, 1973) é uma técnica supervisionada que mede a separação e a compacidade dos agrupamentos e do espaço de dados através de *Matrizes de Espalhamento*.

A Matriz de Espalhamento entre Classes – \mathbf{S}_B (*Between-Class Scatter Matrix*) mede a qualidade da separação entre as categorias. Para um conjunto de n amostras e c agrupamentos, ela é dada pela Equação 3.25.

$$\mathbf{S}_B = \sum_{j=1}^c (\mathbf{m}_j - \mathbf{m})(\mathbf{m}_j - \mathbf{m})^T \quad \text{Equação 3.25}$$

Onde:

- \mathbf{m} – Média do espaço de dados.
- \mathbf{m}_j – Média do agrupamento j

A média do espaço de dados \mathbf{m} pode ser obtida pela Equação 3.26.

$$\mathbf{m} = \sum_{i=1}^n \mathbf{x}_i \quad \text{Equação 3.26}$$

Onde:

- \mathbf{x}_i – Amostra i

E a média de cada um dos agrupamentos \mathbf{m}_i é dada pela Equação 3.27.

$$\mathbf{m}_i = \frac{1}{n_i} \sum_{i=1}^{n_i} \mathbf{x}_i \quad \text{Equação 3.27}$$

Onde:

- n_i – Número de amostras do agrupamento i

A Matriz de Espalhamento Interno – \mathbf{S}_W (*Within-Class Scatter Matrix*) mede a compacidade das categorias. Ela é obtida pela soma dos espalhamentos internos de cada agrupamento. Sua formulação é dada pela Equação 3.28.

$$\mathbf{S}_W = \sum_{j=1}^c \sum_{i=1}^n (\mathbf{x}_i - \mathbf{m}_j)(\mathbf{x}_i - \mathbf{m}_j)^T \quad \text{Equação 3.28}$$

A Matriz de Espalhamento Total do Sistema – \mathbf{S}_T é obtida pela soma do espalhamento interno das categorias e do espalhamento entre categorias, como mostra a Equação 3.29.

$$\mathbf{S}_T = \mathbf{S}_W + \mathbf{S}_B \quad \text{Equação 3.29}$$

Em um espaço de dados com particionamento ótimo, a separação entre os agrupamentos gerados deve ser máxima e o espalhamento das amostras de cada agrupamento deve ser mínimo. Neste caso, como o espalhamento total do sistema é independente das partições, observa-se que: se \mathbf{S}_B tende a se aproximar de \mathbf{S}_T , a matriz \mathbf{S}_W tende à matriz nula.

De forma a avaliar a qualidade da separação e da compacidade, Fisher definiu o *critério* J , que deve ser maximizado. Este critério é obtido através da razão entre os determinantes de \mathbf{S}_B e \mathbf{S}_W , como mostra a Equação 3.30.

$$J = \frac{|\mathbf{S}_B|}{|\mathbf{S}_W|} \quad \text{Equação 3.30}$$

O *critério* J pode ser re-escrito, sem perda de precisão pela Equação 3.31. A alteração feita visa diminuir o alto custo de computacional imposto pelo cálculo dos determinantes das matrizes.

$$J = \frac{\text{traço}(\mathbf{S}_B)}{\text{traço}(\mathbf{S}_W)} \quad \text{Equação 3.31}$$

Os valores deste critério podem variar no intervalo $0 \leq J \leq \infty$. O FLD é uma medida estritamente rígida cuja desvantagem é a perda de precisão quando a sobreposição das amostras aumenta.

3.4.3 DISCRIMINANTE LINEAR DE FISHER ESTENDIDO (EFLD)

O Discriminante Linear de Fisher Estendido, ou *EFLD* (Extended Fisher Linear Discriminant) foi proposto por FRANCO em 2002. Esta medida de validação foi criada a partir do FLD visando permitir a sua utilização para validação dos resultados de algoritmos nebulosos.

Esta medida de validação trabalha com o mesmo conceito de *Matrizes de Espalhamento* que o *FLD* sendo que neste caso todas são consideradas como estendidas, ou seja, são utilizadas a *Matriz Estendida de Espalhamento Entre Classes* (\mathbf{S}_{Be}), a *Matriz Estendida de Espalhamento Interno* (\mathbf{S}_{We}) e a *Matriz Estendida de Espalhamento Total* (\mathbf{S}_{Te}).

A matriz \mathbf{S}_{Be} é obtida a partir da Equação 3.32.

$$\mathbf{S}_{Be} = \sum_{i=1}^n \sum_{j=1}^c \mu_{ij} (\mathbf{m}_{ej} - \mathbf{m})(\mathbf{m}_{ej} - \mathbf{m})^T \quad \text{Equação 3.32}$$

Onde:

\mathbf{m} – Centro do espaço de dados

- \mathbf{m}_{ej} – Centro do agrupamento j
- μ_{ij} – Grau de pertinência da amostra i no agrupamento j

Antes de calcular o \mathbf{S}_{Be} , é necessário calcular o centro do espaço de dados \mathbf{m} e o centro de cada um dos agrupamentos \mathbf{m}_{ej} , que são dados respectivamente pelas Equação 3.33 e Equação 3.34.

$$\mathbf{m} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \quad \text{Equação 3.33}$$

$$\mathbf{m}_{ej} = \frac{\sum_{i=1}^n \mu_{ij} \mathbf{x}_i}{\mu_j} \quad \text{Equação 3.34}$$

Onde μ_j é o somatório dos graus de inclusão de todas as amostras no agrupamento j que é dados pela Equação 3.35.

$$\mu_j = \sum_{i=1}^n \mu_{ij} \quad \text{Equação 3.35}$$

Por sua vez a matriz \mathbf{S}_{We} é obtida a partir da Equação 3.36.

$$\mathbf{S}_{We} = \sum_{i=1}^n \sum_{j=1}^c \mu_{ij} (\mathbf{x}_j - \mathbf{m}_{ej})(\mathbf{x}_j - \mathbf{m}_{ej})^T \quad \text{Equação 3.36}$$

Assim como no FLD matriz estendida de espalhamento total \mathbf{S}_{Te} é obtida pela soma das matrizes \mathbf{S}_{Be} e \mathbf{S}_{We} , pois, segundo FRANCO: “*O espalhamento total do sistema é independente da natureza das partições, sejam estas nebulosas ou rígidas*”.

Assim, o *critério* J_e , descrito por FRANCO, que é a extensão do *critério* J , descrito por Fisher, é calculado a partir da razão dos determinantes das matrizes de espalhamento estendidas \mathbf{S}_{Be} e \mathbf{S}_{We} . Seu valor deve ser maximizado a fim de validar a qualidade da separação e da compacidade das categorias analisadas.

Também de forma análoga ao *critério* J , o *critério* J_e também pode ser expresso pela razão dos traços das matrizes de espalhamentos \mathbf{S}_{Be} e \mathbf{S}_{We} .

O cálculo deste *critério* pode ser otimizado observando-se que as matrizes de espalhamento são geradas pelo produto de um vetor coluna pelo seu transposto, portanto seus

traços correspondem ao quadrado do módulo de seu vetor gerador. Assim, o critério J_e do EFLD pode ser reescrito como a Equação 3.37

$$J_e = \frac{S_{Be}}{S_T - S_{Be}} \quad \text{Equação 3.37}$$

Onde, S_{Be} e S_T são os valores dos traços das respectivas matrizes dados pelas Equação 3.38 e Equação 3.39.

$$s_{Be} = \text{traço}(\mathbf{S}_{Be}) = \sum_{i=1}^n \sum_{j=1}^c \mu_{ij} \|\mathbf{m}_{ej} - \mathbf{m}\|^2 \quad \text{Equação 3.38}$$

$$s_T = \text{traço}(\mathbf{S}_T) = \sum_{i=1}^n \mu_{ij} \|\mathbf{x}_i - \mathbf{m}\|^2 \quad \text{Equação 3.39}$$

A Equação 3.37 é mais rápida de calcular, dado que o termo s_T precisa ser calculado apenas uma vez para o conjunto de amostras e que o termo s_{Be} é mais rápido de calcular do que s_{We} .

Esta medida de validação manteve a característica negativa relatada para o FLD, pois não apresenta bons resultados quando o espaço de dados apresenta muita sobreposição nas amostras.

3.4.4 CONTRASTE ENTRE CLASSES (ICC)

O Contraste Entre Classes, ou simplesmente *ICC*, do inglês *Inter Class Contrast*, também foi definido por FRANCO em 2002, sendo representada pela Equação 3.40.

$$ICC = \frac{S_{Be}}{n} \cdot d_{\min} \cdot \sqrt{c} \quad \text{Equação 3.40}$$

Onde:

- ICC – Contraste Entre Classes
- S_{Be} – Traço da matriz estendida de espalhamento entre classes
- d_{\min} – Distância mínima entre centros dos agrupamentos
- c – Número de agrupamentos
- n – Número de amostras

Na equação acima, cada termo tem uma finalidade:

- S_{Be} tem por finalidade estimar a qualidade da alocação dos centros dos agrupamentos, onde centros mal alocados produzem valores baixos para este termo.
- d_{min} tem por finalidade evitar que o máximo da ICC seja atingido para um número de agrupamentos c maior que o valor ideal.
- \sqrt{c} tem como finalidade garantir que a medida ICC cresça juntamente com o número de agrupamentos, alcançando seu máximo para valores próximos ao valor ótimo de c .
- $\frac{1}{n}$ é um fator de escala, usado para compensar a influência do número de pontos sobre o termo S_{Be} .

Segundo a autora, “a ICC foi criada para ser capaz de avaliar um espaço particionado por uma ferramenta de categorização nebulosa ou rígida, levando em conta a separação entre as categorias geradas e a compacidade destas. Ela também foi moldada para detectar centros alocados muito próximos, o que compromete uma boa categorização”.

Cada um dos termos existentes na Equação 3.40, possui uma formulação própria. Inicialmente, tem-se o traço da matriz estendida de espalhamento entre classes (S_{Be}) que é representado pela Equação 3.41.

$$S_{Be} = \sum_{i=1}^n \sum_{j=1}^c u_{ij} \|\mathbf{m}_j - \mathbf{m}\| \quad \text{Equação 3.41}$$

Onde:

- u_{ij} – Grau de pertinência da amostra i no agrupamento j
- \mathbf{m}_j – Centro do agrupamento j
- \mathbf{m} – Centro do conjunto de dados

O centro do conjunto de dados (\mathbf{m}), apresentado na Equação 3.41, deve ser calculado em relação ao espaço de dados $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, como mostra a Equação 3.42.

$$\mathbf{m} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \quad \text{Equação 3.42}$$

Onde:

- \mathbf{x}_i – Amostra i

E finalmente, a distância mínima entre centros dos agrupamentos (d_{\min}) é dada pela Equação 3.43.

$$d_{\min} = \min_{1 \leq j \leq c} [\min_{j+1 \leq k \leq c-1} \| \mathbf{m}_j - \mathbf{m}_k \|] \quad \text{Equação 3.43}$$

Onde:

\mathbf{m}_k – Centro do agrupamento k , onde $k \neq j$

Para seleccionar o melhor número de agrupamentos, o algoritmo de agrupamento deve ser executado para vários valores de c , sendo a *ICC* calculada para cada um dos resultados. O melhor número de agrupamentos no intervalo $[2, c]$ é indicado pelo valor máximo da *ICC* calculada para cada um dos valores de c .

3.4.5 SILHUETA MÉDIA GLOBAL (\bar{S})

A Silhueta Média Global (*Overall Average Silhouette*) é uma medida de validação, definida especificamente para o algoritmo PAM, que visa medir o quão distintos são os agrupamentos. Na realidade o cálculo para obtenção desta medida é efetuado segundo uma série de passos, como mostrado a seguir. Neste algoritmo cada passo intermediário possui alguma interpretação sobre a qualidade relativa da amostra nos agrupamentos.

Na primeira etapa, para cada amostra \mathbf{x}_i pertencente ao agrupamento A , calcula-se a distância da amostra \mathbf{x}_i a todas as demais amostras \mathbf{x}_j pertencentes a A , como mostra a Equação 3.44.

$$a(\mathbf{x}_i) = \frac{1}{|A|-1} \left(\sum_{j=1, j \neq i}^{|A|-1} D(\mathbf{x}_i, \mathbf{x}_j) \right) \quad \text{Equação 3.44}$$

Onde:

$a(\mathbf{x}_i)$ – Dissimilaridade média da amostra \mathbf{x}_i em relação a todas as demais amostras do agrupamento A .

$|A|$ – Número de amostras pertencentes ao agrupamento A .

$D(\mathbf{x}_i, \mathbf{x}_j)$ – Dissimilaridade entre as amostras \mathbf{x}_i e \mathbf{x}_j , onde $\mathbf{x}_i \in A, \mathbf{x}_j \in A$ e $\mathbf{x}_i \neq \mathbf{x}_j$

Para cada agrupamento C diferente de A , calcula-se a distância entre a amostra \mathbf{x}_i , pertencente a A , e todas as amostras \mathbf{x}_k pertencentes ao agrupamento C , conforme a Equação 3.45.

$$D(\mathbf{x}_i, C) = \frac{1}{|C|} \sum_{k=1}^{|C|} D(\mathbf{x}_i, \mathbf{x}_k) \quad \text{Equação 3.45}$$

Onde:

- $D(\mathbf{x}_i, C)$ – Dissimilaridade média da amostra de \mathbf{x}_i em relação a todas as amostras pertencentes ao agrupamento C .
- $|C|$ – Número de amostras no agrupamento de C .
- $D(\mathbf{x}_i, \mathbf{x}_k)$ – Dissimilaridade entre as amostras \mathbf{x}_i e \mathbf{x}_k , onde $\mathbf{x}_i \in A, \mathbf{x}_k \in C$ e $\mathbf{x}_i \neq \mathbf{x}_k$

Assim que a dissimilaridade média da amostra de \mathbf{x}_i for calculada em relação a todos os agrupamentos $C \neq A$, aplica-se a Equação 3.46, de forma a obter a dissimilaridade mínima da amostra \mathbf{x}_i .

$$b(\mathbf{x}_i) = \min_{C \neq A} D(\mathbf{x}_i, C) \quad \text{Equação 3.46}$$

Onde:

- $b(\mathbf{x}_i)$ – Dissimilaridade mínima da amostra de \mathbf{x}_i em relação a todos os demais agrupamentos C .

O agrupamento B que possui $D(\mathbf{x}_i, C) = b(\mathbf{x}_i)$ é denominado o agrupamento vizinho da amostra \mathbf{x}_i , ou seja, o segundo melhor agrupamento onde a amostra \mathbf{x}_i poderia ser alocada.

Finalmente pode-se calcular o valor da silhueta (*silhouette*) da amostra \mathbf{x}_i , segundo a Equação 3.47.

$$s(\mathbf{x}_i) = \frac{b(\mathbf{x}_i) - a(\mathbf{x}_i)}{\max\{a(\mathbf{x}_i), b(\mathbf{x}_i)\}} \quad \text{Equação 3.47}$$

Onde:

- $s(\mathbf{x}_i)$ – Silhueta da amostra \mathbf{x}_i .

Em NAGPAUL (2004), encontra-se uma indicação de que este primeiro resultado permite verificar a qualidade da alocação da amostra \mathbf{x}_i em relação aos agrupamentos A e B , sendo que:

- $s(\mathbf{x}_i) \approx 1$ indica que a amostra \mathbf{x}_i está bem alocada no agrupamento A .
- $s(\mathbf{x}_i) \approx 0$ indica que a amostra \mathbf{x}_i está entre os agrupamentos A e B .
- $s(\mathbf{x}_i) \approx -1$ indica que a amostra \mathbf{x}_i está mal alocada no agrupamento A .

Terminado o processamento das silhuetas de cada uma das amostras \mathbf{x}_i pertencentes ao espaço de dados, ainda existem mais dois indicadores que devem ser calculados. O primeiro indicador obtido é a Silhueta Média (*Average Silhouette*), $S(A)$, de cada um dos agrupamentos a partir da Equação 3.48. Este primeiro indicador permite verificar a qualidade dos agrupamentos entre si.

$$S(A) = \frac{1}{|A|} \sum_{i=1}^{|A|} s(\mathbf{x}_i) \quad \text{Equação 3.48}$$

Onde:

$S(A)$ – Silhueta Média do agrupamento A

O segundo indicador é a Silhueta Média Global (*Overall Average Silhouette*), \bar{S} , apresentado na Equação 3.49. Este indicador avalia o resultado do algoritmo de agrupamento em relação à totalidade do espaço de dados.

$$\bar{S} = \frac{1}{n} \sum_{i=1}^n s(\mathbf{x}_i) \quad \text{Equação 3.49}$$

Onde:

\bar{S} – Silhueta Média Global do espaço de dados
 n – Número de amostras no espaço de dados

Para determinar o melhor número de agrupamentos, o algoritmo do PAM deve ser executado sucessivamente para diferentes número de agrupamentos ($c = 1, 2, 3, \dots, c_{máx}$) e o valor da Silhueta Média Global deve ser calculado para cada um dos números de agrupamentos. O melhor número de agrupamentos no intervalo $[2, c]$ é indicado pelo valor máximo da *ICC* calculada para cada um dos valores de c .

Ainda segundo NAGPAUL (2004), “este coeficiente é uma medida adimensional do grau de estruturação do agrupamento que foi revelado pelo algoritmo”. Segundo a sua interpretação a Silhueta Média Global pode ser interpretada como:

- $\bar{S} > 0,7$ - Indica que a estrutura encontrada é boa.
- $0,5 > \bar{S} \geq 0,7$ - Indica que a estrutura encontrada é razoável.
- $0,25 > \bar{S} \geq 0,5$ - Indica que a estrutura encontrada é fraca ou é artificial. Outros métodos de análise de dados deveriam ser utilizados.
- $\bar{S} \leq 0,25$ - Indica que a estrutura encontrada não é substancial.

3.4.6 MEDIDAS DE VALIDAÇÃO SELECIONADAS

Neste trabalho serão apresentadas as medidas de validação utilizadas para verificação da qualidade dos agrupamentos obtidos após a aplicação do FCM e do PAM. No primeiro caso, será utilizado Contraste Entre Classes (Inter Class Contrast), ou ICC, definido por FRANCO (2000) e no segundo, a Silhueta Média Global (Overall Average Silhouette), ou \bar{S} , definida por KAUFFMAN (1999).

A ICC foi selecionada, pois no trabalho apresentado por FRANCO, esta medida foi comparada a várias outras medidas e apresentou os melhores resultados para os conjuntos de testes avaliados.

A \bar{S} foi selecionada por ter sido definida especialmente para trabalhar com o PAM e também por apresentar uma possibilidade de representação gráfica.

4 APLICAÇÃO DA MEDIDA *ICC* PARA DADOS MULTIDIMENSIONAIS

Em FRANCO (2000), foram descritas e comparadas várias medidas de validação (*validity measures*) de agrupamentos, entre elas o Coeficiente de Partição - *F* (*Partition Coeficiente*), a Entropia de Partição - *H* (*Partition Entropy*), a Compacidade e Separação - *CS* (*Compactness and Separation*) e o Discriminante Linear de Fisher - *FLD* (*Fisher Linear Discriminant*). Neste trabalho também foi definida uma nova medida de validação de agrupamentos o Contrates Entre Classes - *ICC* (*Inter Class Contrast*). Esta medida foi comparada e validada em relação as demais, tendo apresentado bons resultados para vários espaços de dados bidimensionais.

Em OLIVEIRA (2002), foi feita uma sugestão de alteração desta medida após a realização de testes em espaços de dados bidimensionais que apresentavam como característica a colinearidade dos centros dos agrupamentos, segundo a análise realizada neste tipo de espaço de dados a *ICC* não alcança bons resultados para este tipo de espaços de dados. A medida proposta somente alterava a *ICC* (Equação 3.40) incluindo mais um termo constante como apresentado na Equação 4.1.

$$ICC_{novo} = \frac{S_{Be}}{n} \cdot d_{\min} \cdot \sqrt{c \cdot \ln c} \quad \text{Equação 4.1}$$

Neste trabalho de dissertação, o número de dimensões dos espaços de dados é sempre maior que 2, como será descrito nos próximos itens. Uma vez que já haviam sido detectados casos de espaços de dados bidimensionais em que a medida não se comportava como esperado foram realizados testes visando verificar a sua acuidade para espaços de dados com um número maior de dimensões.

O principal resultado obtido a partir dos testes foi a verificação de que a *ICC* pode ser utilizada, não somente para determinar o melhor número de agrupamentos para um espaço de dados, mas também para determinar o melhor valor do coeficiente de nebulização a ser utilizado no FCM.

Foram realizados dois conjuntos de testes da medida *ICC* utilizando espaços de dados multidimensionais.

O primeiro conjunto de testes foi realizado utilizando um espaço de dados com 360 dimensões, igual ao que será utilizado no *Agrupamento de Matriz Única* e que será descrito no item 6.3.1.

O segundo conjunto de testes foi realizado utilizando vários espaços de dados de 12 dimensões, representativos das séries de potência disponível das usinas hidrelétricas, onde os espaços de dados selecionados são representantes de quatro casos distintos de resposta da *ICC*. Este espaço de dados será descrito no item 6.4.1.1 e será utilizado tanto no *Agrupamento por Mapeamento de Resultados* quanto no *Agrupamento por Mapeamento de Resultados com Seleção do Coeficiente de Nebulização*.

Em ambos os testes os espaços de dados utilizados foram previamente normalizados, conforme será discutido no item 6.1.2 e as definições apresentadas no item 6.2. O algoritmo utilizado para a realização dos testes é apresentado a seguir:

Passo 1. Fixar:

O número máximo de agrupamentos $c_{máx} = 20$.

O número de rodadas $r_{máx} = 20$.

O valor da constante de nebulização m .

Passo 2. Determinar o centro do espaço de dados X'_{usina}

Passo 3. Para $c = 2, 3, \dots, c_{máx}$

Passo 3.1. Para $r = 1, 2, \dots, r_{máx}$

Passo 3.1.1. Executar o FCM

Passo 3.1.2. Se o valor da função objetivo $J(r)$ for menor que o valor de $J(r-1)$

Armazenar a matriz de pertinências $U(c) = U(r)$

Armazenar a matriz de centros $M(c) = M(r)$

Passo 3.1.3. Retornar ao Passo 3.1.

Passo 3.2. Calcular a *ICC*(c), como apresentado no item 3.4.4.

Passo 3.3. Retornar ao Passo 3.

Passo 4. Desenhar o gráfico dos resultados da *ICC*

No item 6.2 (*Definições Preliminares*), será indicado que o número máximo de agrupamentos a ser utilizado nos processos de agrupamento $c_{máx} = 10$, mas para a realização dos testes foi detectado que seria necessário ampliar o número máximo de agrupamentos de forma a permitir a melhor visualização do comportamento da medida *ICC*. Foram realizados testes com o valor de $c_{máx} = 15, 20, 25$ e 30 , sendo que no algoritmo anterior foi adotado o número máximo de agrupamentos $c_{máx} = 20$, pois, com 15 agrupamentos ainda não era possível uma boa

visualização do comportamento da métrica. Este número de 20 agrupamentos foi selecionado por representar aproximadamente um terço do número de amostras no espaço de dados, o dobro do número máximo de agrupamentos que será utilizado nos processos de agrupamento, além de já permitir uma boa visualização do comportamento da *ICC* para os diversos espaços de dados.

4.1 PRIMEIRO CONJUNTO DE TESTES

Neste conjunto de testes foi utilizado um espaço de dados normalizado de 360 dimensões, igual ao espaço de dados que será utilizado no *Agrupamento de Matriz Única* apresentado no item 6.3. Ele é constituído por uma matriz de 65 linhas de 360 dimensões, onde, o critério para formação desta matriz foi: as primeiras 12 colunas representam os dados das séries de potência referentes a primeira usina, as colunas 13 a 24 representam os dados referentes a segunda usina e assim sucessivamente até que todas as usinas tenham sido representadas.

Por causa do elevado número de dimensões foi necessário dividir o gráfico em 4 partes para melhorar a sua visualização. Os gráficos representativos do espaço de dados são apresentados na Figura 4.1 a), b), c) e d).

Pode ser observado a partir das figuras que compõem o espaço de dados que a cada 12 dimensões as características de dispersão dos dados mudam.

Por exemplo, na Figura 4.1 a), o intervalo 37 a 48, representa um intervalo do espaço de dados em que a dispersão é grande. Este intervalo apresenta os dados da usina de Caconde cujo máximo do $CV = 22,20$ e a média $CV = 15,56$.

Este intervalo é seguido pelo intervalo 49 a 60, onde são representados os dados de Euclides da Cunha (máximo $CV = 3,21$; média $CV = 1,66$), em que a dispersão diminui, que por sua vez é seguido pelo intervalo 61 a 72, representando os dados de Marimondo (máximo $CV = 23,14$; média $CV = 15,97$), onde a dispersão dos dados torna a aumentar.

Este tipo de comportamento continua ao longo de todo o espaço de dados e é consequência da forma como este foi montado.

O gráfico da Figura 4.2, representa os resultados da medida *ICC* obtidos após a utilização do FCM com coeficiente de nebulização m igual a 1,25. Neste gráfico podemos visualizar o primeiro problema na utilização desta medida para um espaço de dados com um número elevado de dimensões, ou seja, não foi possível determinar um máximo no intervalo para o qual a medida foi calculada.

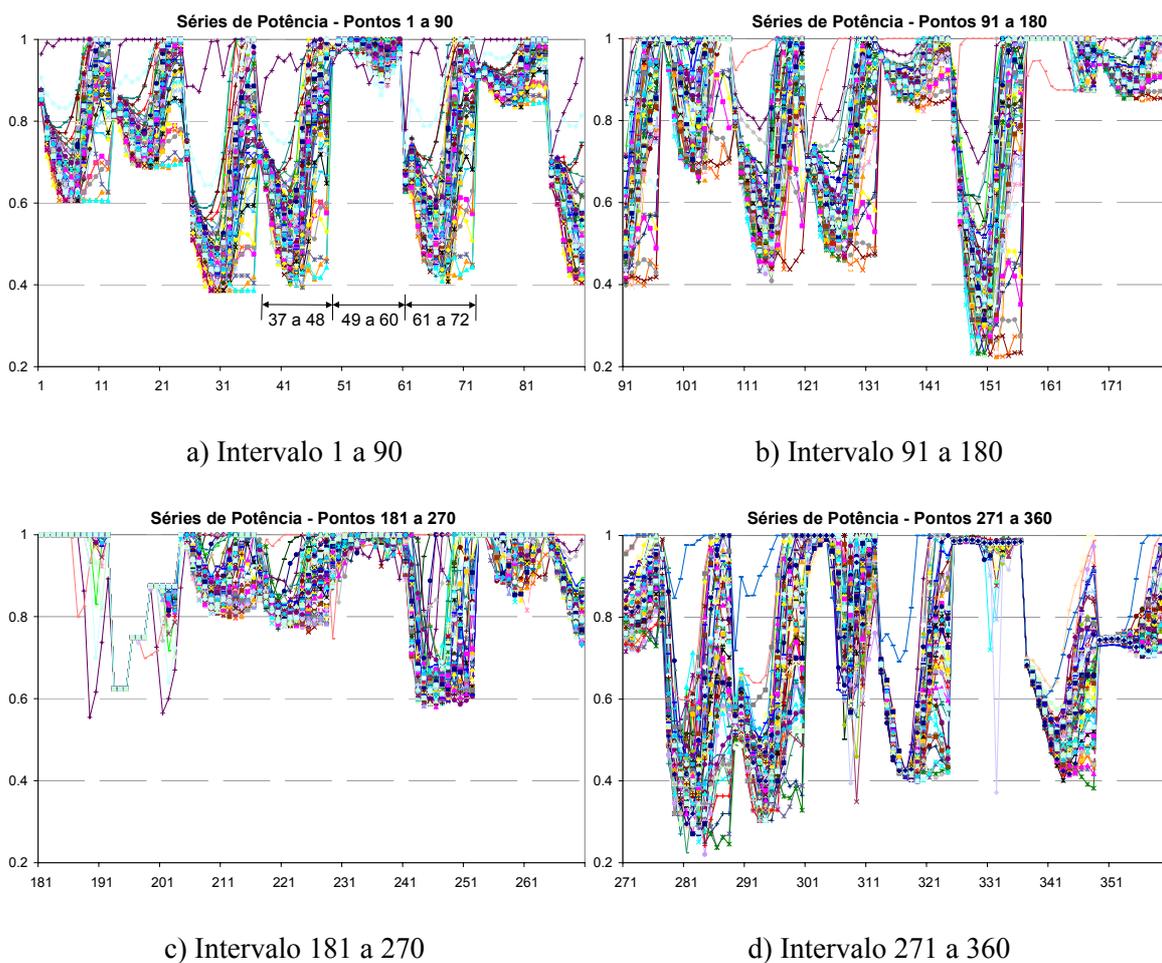


Figura 4.1 – Espaço de dados do primeiro conjunto de testes

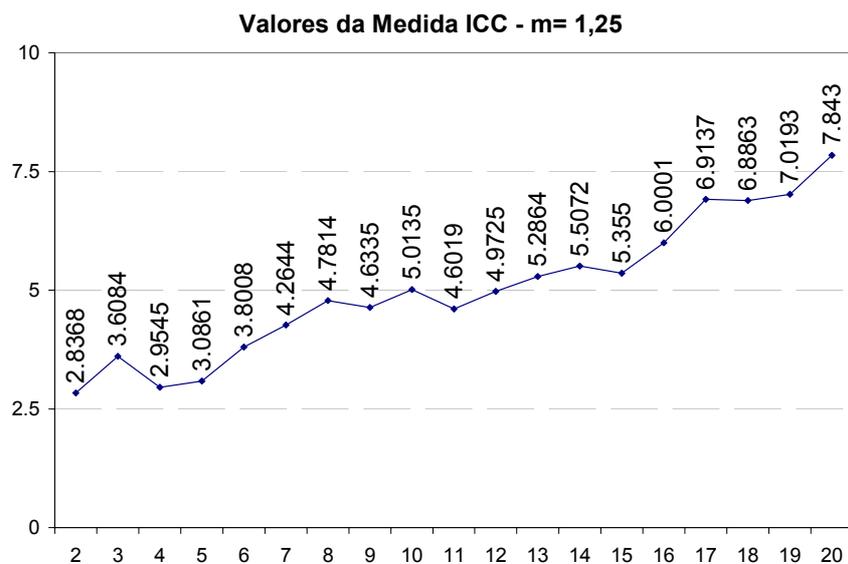


Figura 4.2 – Gráfico da medida ICC para $m = 1,25$

O gráfico da Figura 4.3, representa os resultados da medida *ICC* onde o FCM foi utilizado com coeficiente de nebulização *m* igual a 1,50. Neste caso a *ICC* apresentou um máximo no intervalo que indica que o número de agrupamentos ideal seria igual a 3.

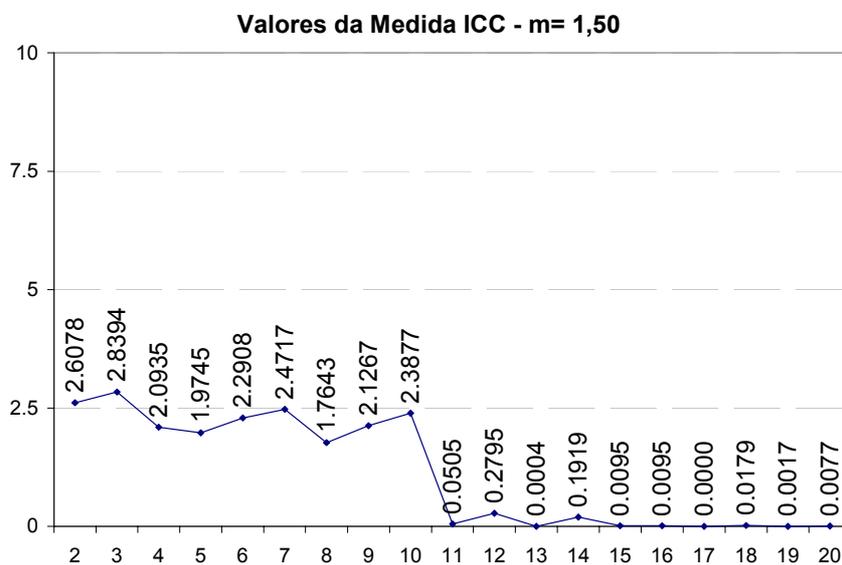


Figura 4.3 – Gráfico da medida *ICC* para *m* = 1,50

Neste gráfico pode ser verificado também que a partir de 11 agrupamentos os resultados da medida tenderem a zero. A explicação para este fenômeno está no fato de que para números elevados de agrupamentos os centros definidos pelo FCM são muito próximos, o que torna o termo d_{min} da *ICC* muito pequeno.

Os gráficos apresentados nas Figura 4.4 e Figura 4.5, onde o FCM foi utilizado com coeficiente de nebulização *m* igual a 1,75 e 2,00, também apresentam como característica o fato da medida *ICC* tender para zero, sendo que nestes dois casos, este fenômeno inicia muito mais rápido, ou seja, para 4 agrupamentos.

Estes dois últimos resultados indicaram que pela característica de sobreposição existente no espaço de dados a utilização do FCM com valores do coeficiente de nebulização mais altos define centros ainda mais próximos. Este fenômeno também fica evidenciado pela diminuição do valor do máximo da *ICC* obtido em cada um dos testes.

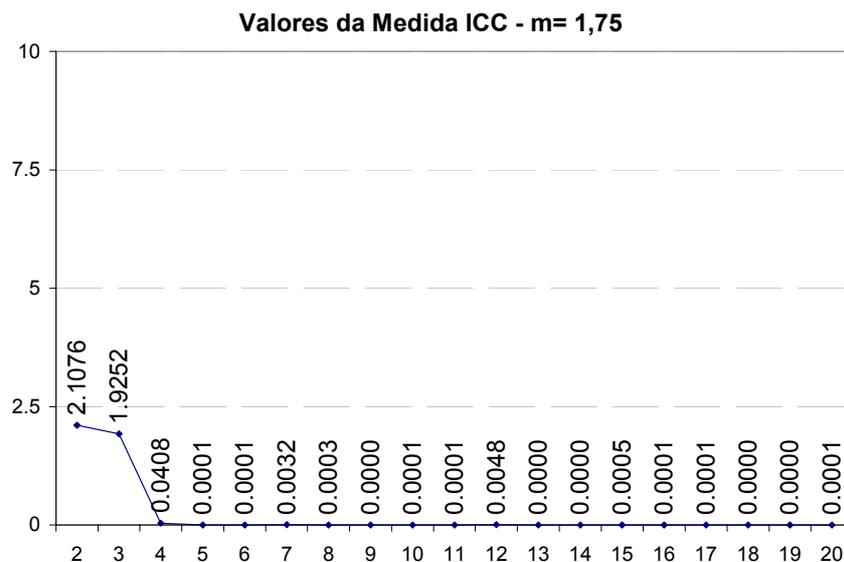


Figura 4.4 – Gráfico da medida ICC para $m = 1,75$

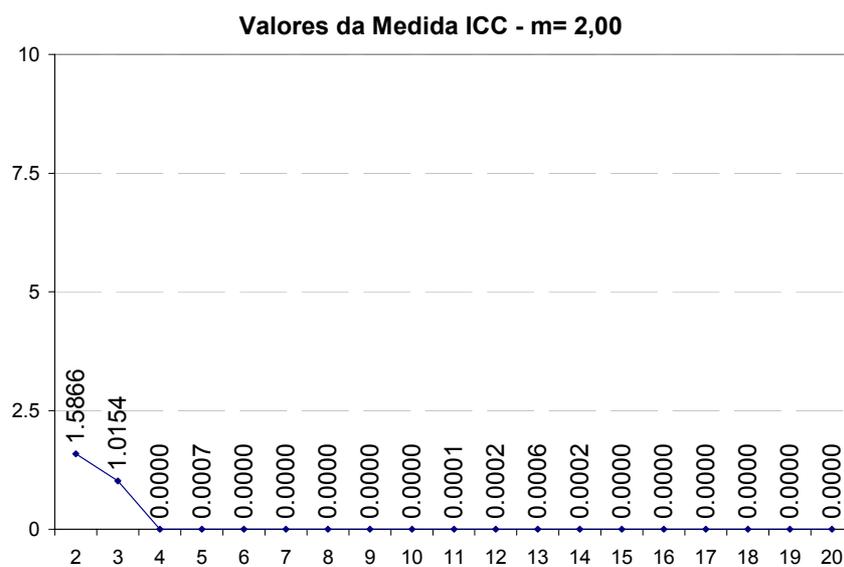


Figura 4.5 – Gráfico da medida ICC para $m = 2,00$

Dando continuidade a este conjunto de testes foram avaliados os resultados obtidos substituindo-se a ICC pela ICC_{Novo} cujos gráficos comparativos são apresentados na Figura 4.6, Figura 4.7, Figura 4.8 e Figura 4.9.

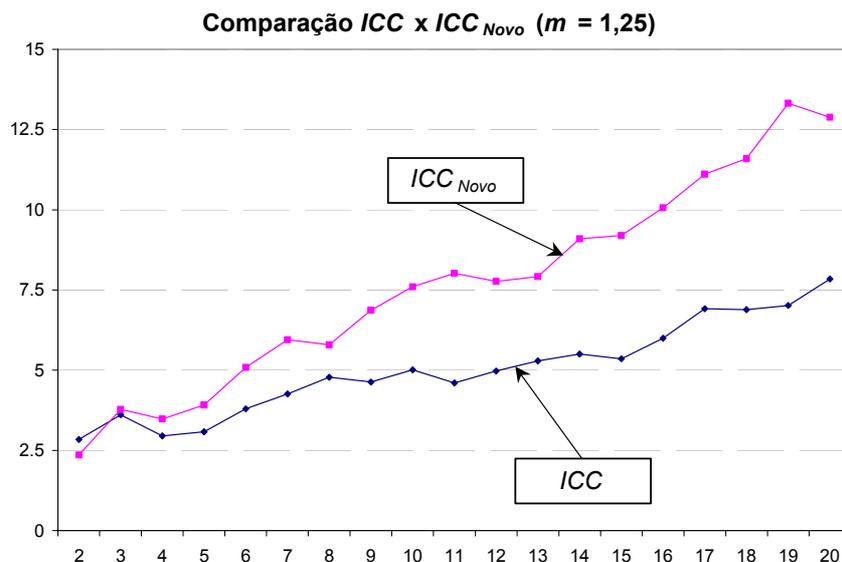


Figura 4.6 – Gráfico de comparação dos resultados da ICC e ICC_{Novo} para $m = 1,25$

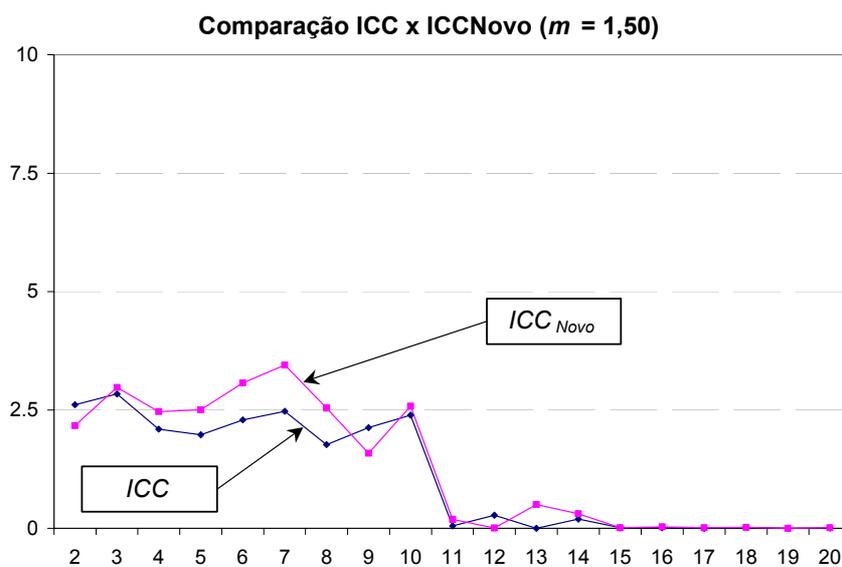


Figura 4.7 – Gráfico de comparação dos resultados da ICC e ICC_{Novo} para $m = 1,50$

Estes gráficos indicam que a alteração feita na medida não afetou significativamente o comportamento dos gráficos. Sendo que para os resultados obtidos com m igual a 1,25, indicam que a alteração torna ainda mais difícil a determinação de um máximo no intervalo.

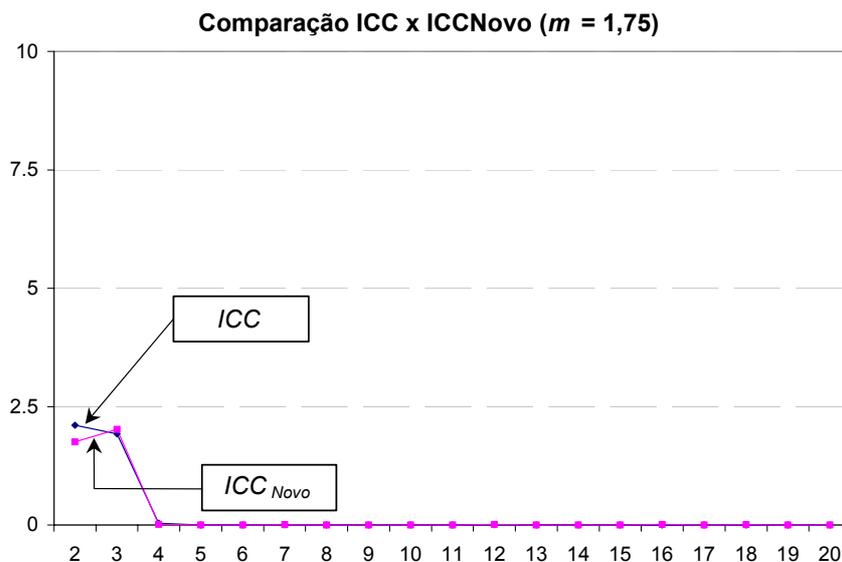


Figura 4.8 – Gráfico de comparação dos resultados da ICC e $ICCNovo$ para $m = 1,75$

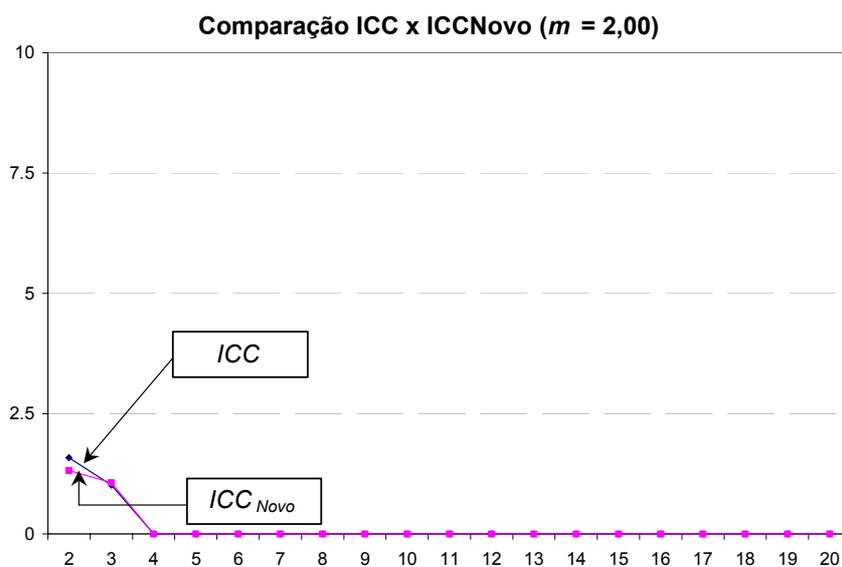


Figura 4.9 – Gráfico de comparação dos resultados da ICC e $ICCNovo$ para $m = 2,00$

O efeito da alteração pode ser visualizado na Tabela 4.1 e no gráfico da Figura 4.10. Comparativamente, o efeito do termo $\sqrt{c \cdot \ln c}$ é aumentar o resultado da medida para valores mais altos do número de agrupamentos c , ou seja, em casos onde a ICC determina vários valores de máximo com valores próximos, como no gráfico da Figura 4.7, a $ICCNovo$ privilegiaria o máximo associado ao maior número de agrupamentos. No gráfico citado, o efeito de privilegiar máximos obtidos para valores de c mais altos fica evidente, pois: o valor da ICC para $c = 2$

diminui por efeito do termo incluído, para $c = 3$ o valor quase não se altera enquanto que para $c = 7$ o valor aumenta significativamente.

Tabela 4.1 – Comparação entre \sqrt{c} e $\sqrt{c \cdot \ln c}$

| Número de Agrupamentos (c) | \sqrt{c} | $\sqrt{c \cdot \ln c}$ |
|--------------------------------|------------|------------------------|
| 2 | 1,41421 | 0,83255 |
| 3 | 1,73205 | 1,28371 |
| 4 | 2,00000 | 1,66511 |
| 5 | 2,23607 | 2,00589 |
| 6 | 2,44949 | 2,31846 |
| 7 | 2,64575 | 2,60973 |
| 8 | 2,82843 | 2,88405 |
| 9 | 3,00000 | 3,14444 |
| 10 | 3,16228 | 3,39307 |

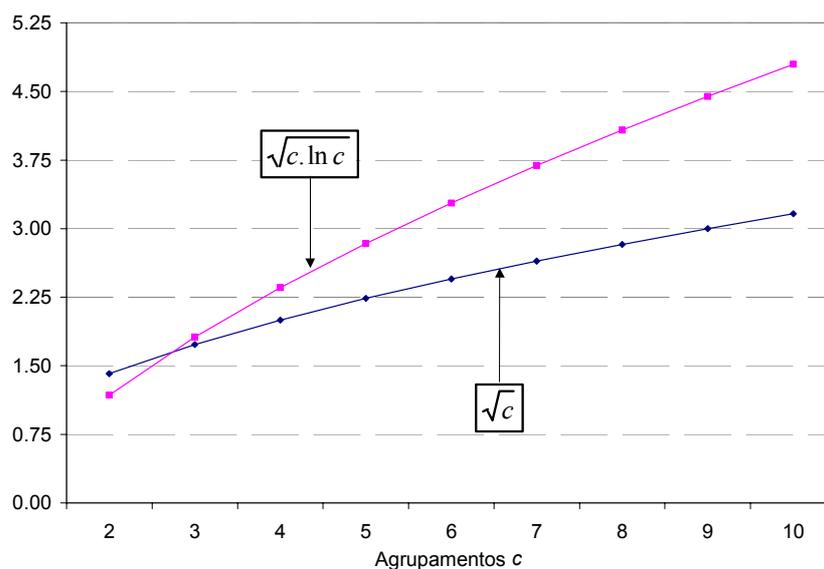


Figura 4.10 – Gráfico comparação entre \sqrt{c} e $\sqrt{c \cdot \ln c}$

No segundo conjunto de testes a medida ICC_{novo} não foi mais avaliada, pois já foi verificado que a alteração proposta, em relação ao ICC , não acrescentou nenhuma informação relevante a determinação do melhor número de agrupamentos, especialmente no caso em que a medida original não obteve bons resultados.

4.2 SEGUNDO CONJUNTO DE TESTES

Neste conjunto de testes foram utilizados espaços de dados compostos de 12 dimensões cada. Os gráficos representativos destes espaços de dados são apresentados no Apêndice B, onde cada usina foi representada por sua série de potência disponível normalizada. Do conjunto de 30 usinas existentes neste anexo, foram selecionadas as usinas de Serra da Mesa, Mascarenhas, Jupia e Funil, cada uma delas por apresentar um comportamento específico em relação a *ICC*.

A primeira usina selecionada para o teste da *ICC* foi a usina de Serra da Mesa, cujo máximo do coeficiente de variação é 23,04 e a média do coeficiente de variação é 11,42. Esta usina foi escolhida por ser uma representante das usinas nas quais a medida *ICC* apresentou resultados com a definição de um máximo no intervalo em que a medida foi calculada, independentemente do valor do coeficiente de nebulização utilizado no FCM, como pode ser verificado na Figura 4.11, Figura 4.12, Figura 4.13 e Figura 4.14.

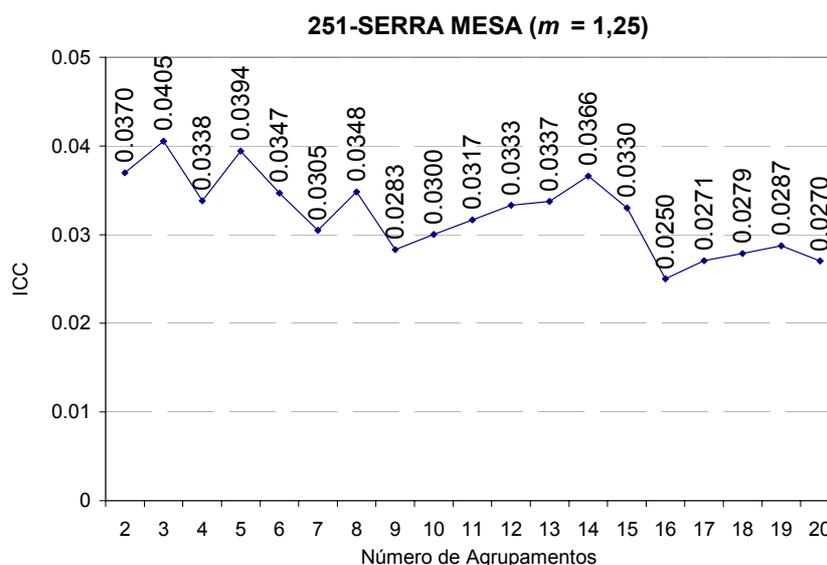


Figura 4.11 – Gráfico dos valores de *ICC* da usina de Serra da Mesa para $m = 1,25$

Observando estes gráficos, é possível verificar que para todos os casos, o número de agrupamentos em que ocorreu o máximo da medida permaneceu igual, em 3 agrupamentos, independente do valor do coeficiente de nebulização m utilizado.

Também pode ser verificado que o resultado da *ICC* é muito próximo em todos os casos, como mostra a Tabela 4.2.

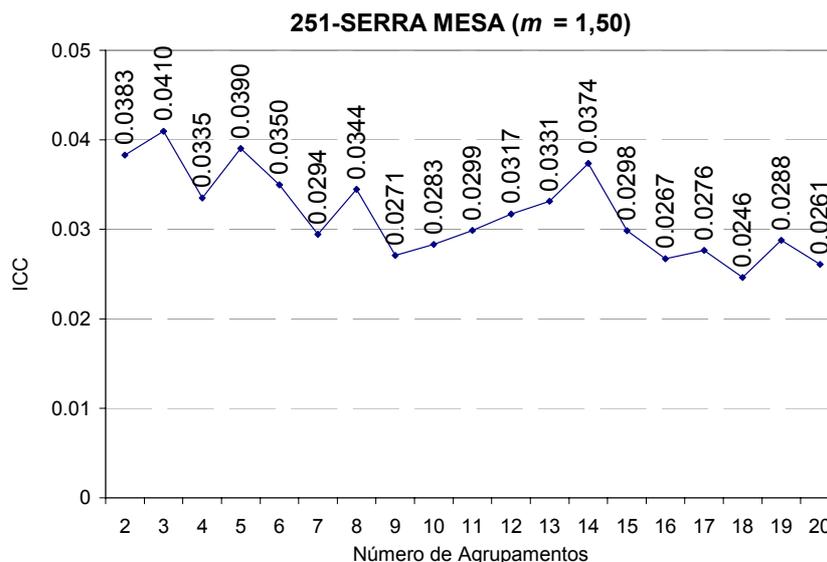


Figura 4.12 – Gráfico dos valores de ICC da usina de Serra da Mesa para $m = 1,50$

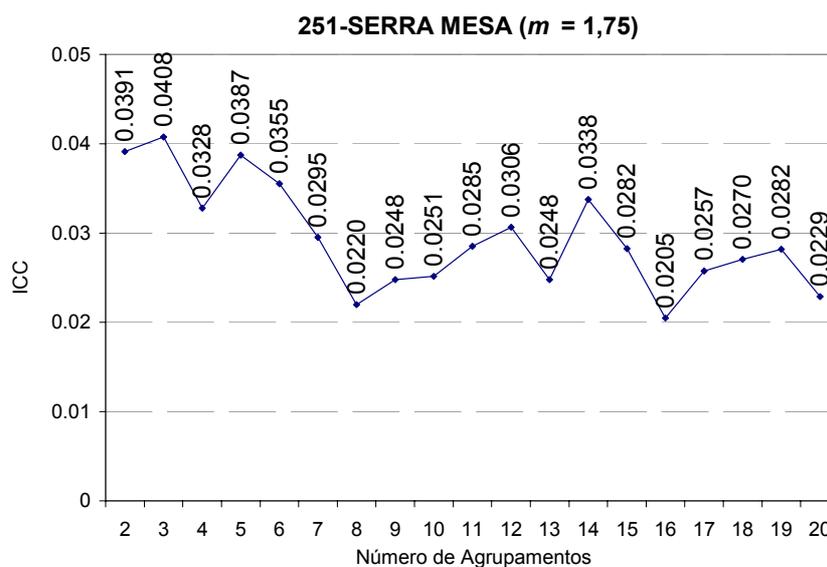


Figura 4.13 – Gráfico dos valores de ICC da usina de Serra da Mesa para $m = 1,75$

Outra característica interessante é que, apesar do valor da ICC variar pouco, pode ser observado a partir dos gráficos da ICC desta usina que conforme o valor do coeficiente de nebulização aumenta, mais definido fica o valor do máximo da ICC . Este comportamento da ICC indica que além de definir o melhor número de agrupamentos a medida também pode ser utilizada para indicar o melhor coeficiente de nebulização m para as características do espaço de dados.

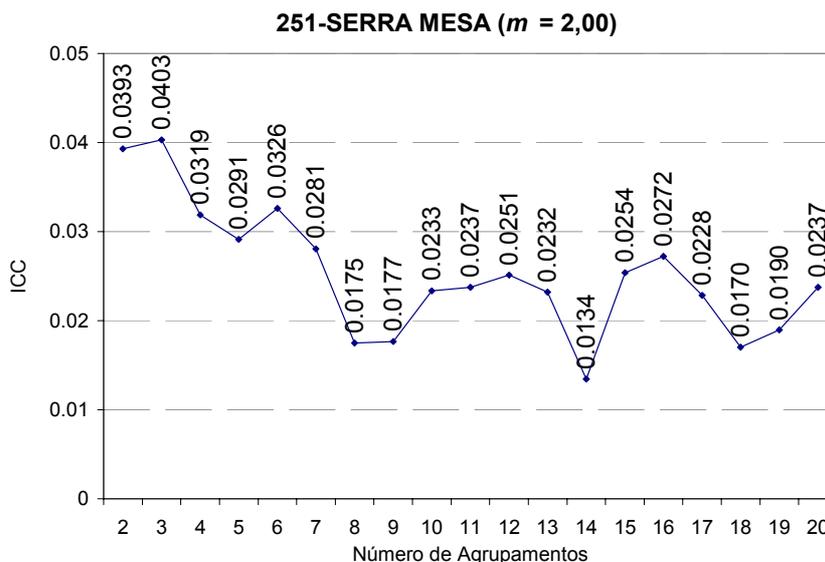


Figura 4.14 – Gráfico dos valores de ICC da usina de Serra da Mesa para $m = 2,00$

Tabela 4.2 – Valores da ICC para 3 agrupamentos na usina de Serra da Mesa

| Coefficiente de Nebulização (m) | ICC |
|---|-------------------------|
| 1,25 | 0,0405 |
| 1,50 | 0,0410 |
| 1,75 | 0,0408 |
| 2,00 | 0,0403 |

A segunda usina testada foi a usina de Mascarenhas, cujo máximo do coeficiente de variação é 17,30 e a média do coeficiente de variação é 7,90. Esta usina foi escolhida por ser uma representante das usinas nas quais a medida ICC apresentou resultados em que não há a definição clara de um máximo no intervalo, para 3 dos 4 valores do coeficiente de nebulização utilizados no FCM, como pode ser verificado na Figura 4.15, Figura 4.16, Figura 4.17 e Figura 4.18.

É possível verificar que para o espaço de dados desta usina, quando o processo de agrupamento foi executado com os três primeiros valores do coeficiente de nebulização não foi possível definir claramente o valor do máximo da ICC , sendo que quanto maior foi o valor do coeficiente de nebulização utilizado, mais difícil foi determinar o valor máximo.

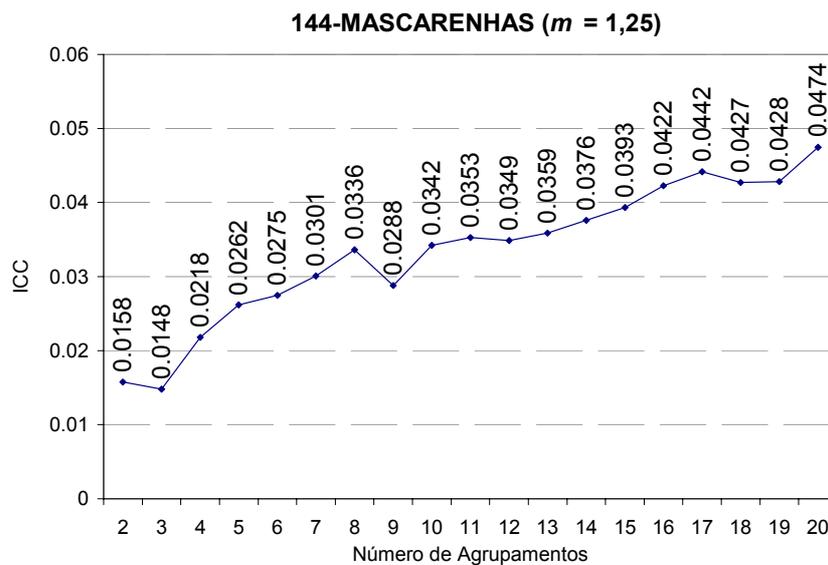


Figura 4.15 – Gráfico dos valores de ICC da usina de Mascarenhas para $m = 1,25$

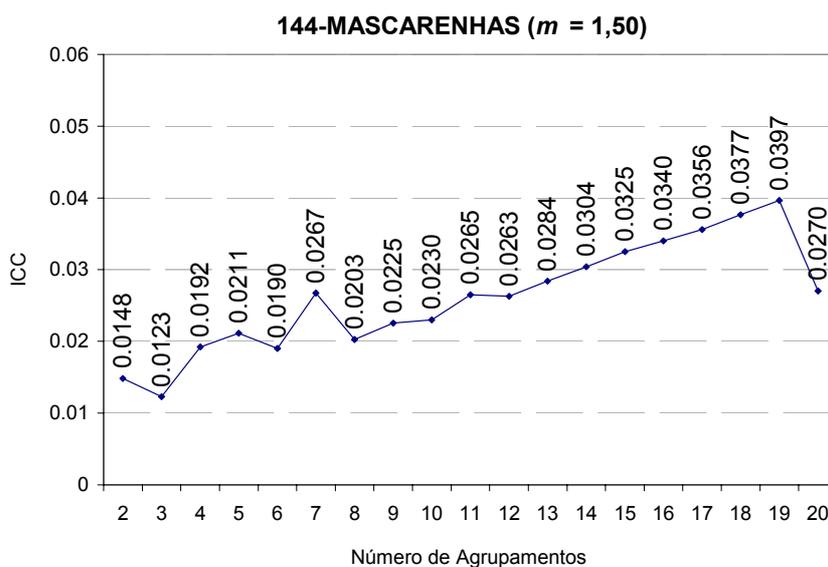


Figura 4.16 – Gráfico dos valores de ICC da usina de Mascarenhas para $m = 1,50$

Particularmente, para $m = 2,00$ (Figura 4.18), o gráfico se comporta como os gráficos da Figura 4.4 e da Figura 4.5, ou seja, fica definido um valor máximo para a ICC cujo valor é baixo. Além disto, a queda dos valores da ICC quando o número de agrupamentos é igual a 5 indica que os centros dos agrupamentos são muito próximos.

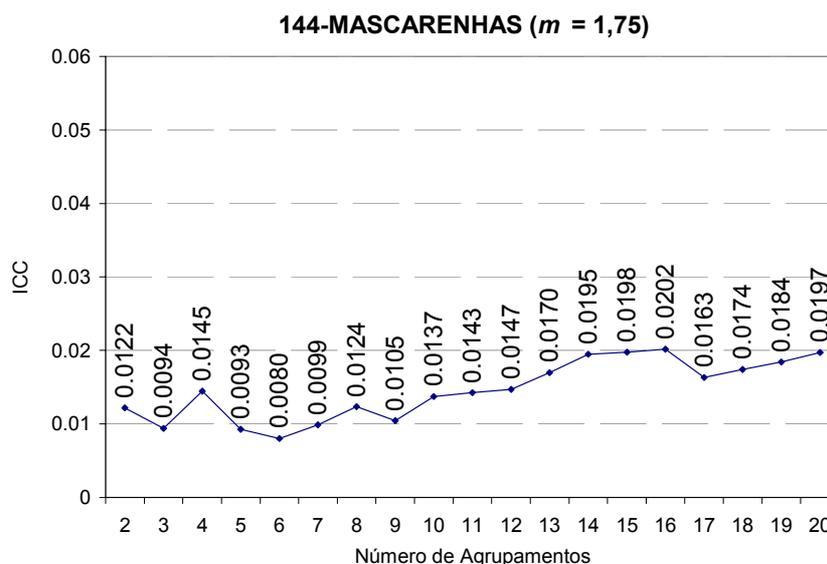


Figura 4.17 – Gráfico dos valores de ICC da usina de Mascarenhas para $m = 1,75$

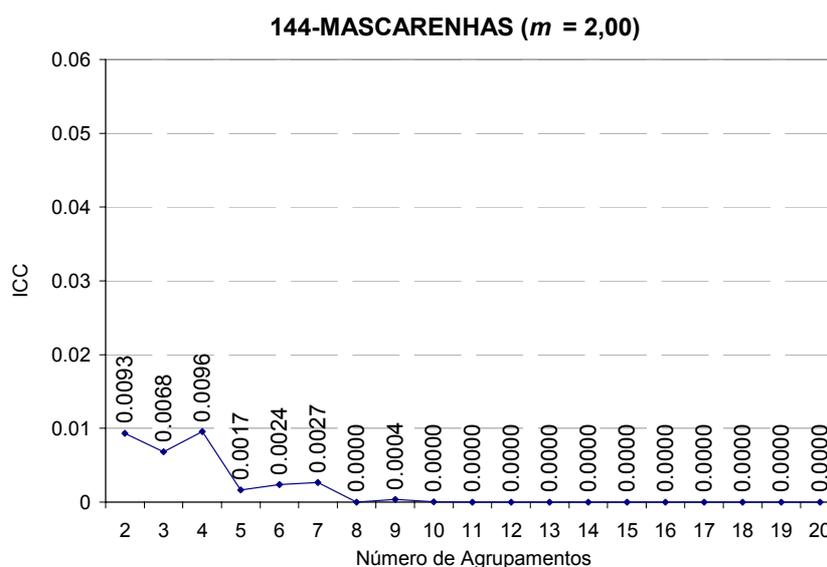


Figura 4.18 – Gráfico dos valores de ICC da usina de Mascarenhas para $m = 2,00$

Analisando sobre outro enfoque, o comportamento da *ICC* observado para este espaço de dados quando agrupado utilizando diferentes valores do coeficiente de nebulização, sinaliza ainda mais fortemente que esta medida pode ser utilizada para indicar o melhor coeficiente de nebulização m para as características do espaço de dados, além de definir o melhor número de agrupamentos.

A terceira usina selecionada para os testes foi a usina de Jupiá, cujo máximo do coeficiente de variação é 6,44 e a média do coeficiente de variação é 1,95. Esta usina foi

selecionada por apresentar valores sempre decrescentes no valor da medida *ICC*, sendo que estes valores são muito pequenos, na casa dos milésimos.

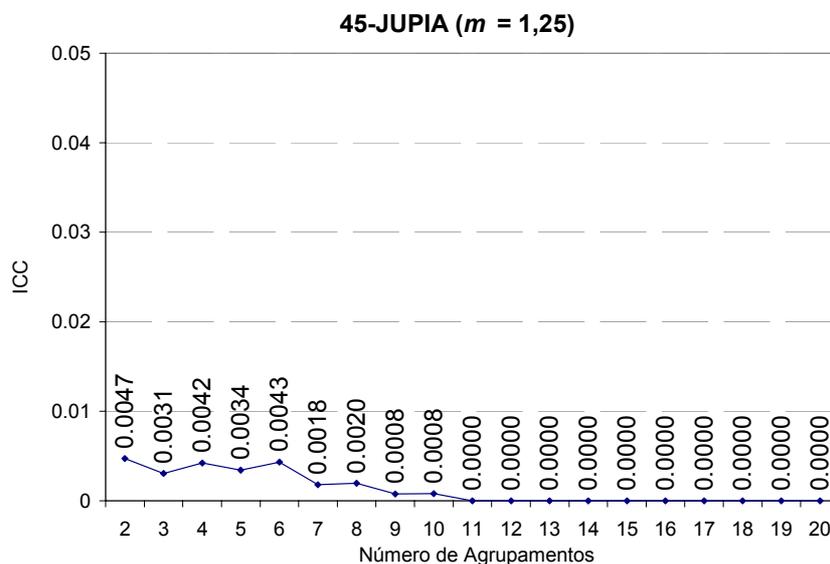


Figura 4.19 – Gráfico dos valores de *ICC* da usina de Jupiá para $m = 1,25$

Apesar dos valores muito baixos da *ICC*, característicos de um espaço de dados com muita sobreposição, os gráficos apresentados na Figura 4.19, Figura 4.20, Figura 4.21 e Figura 4.22, indicam que a medida obteve êxito na determinação dos agrupamentos de um máximo para $c = 2$.

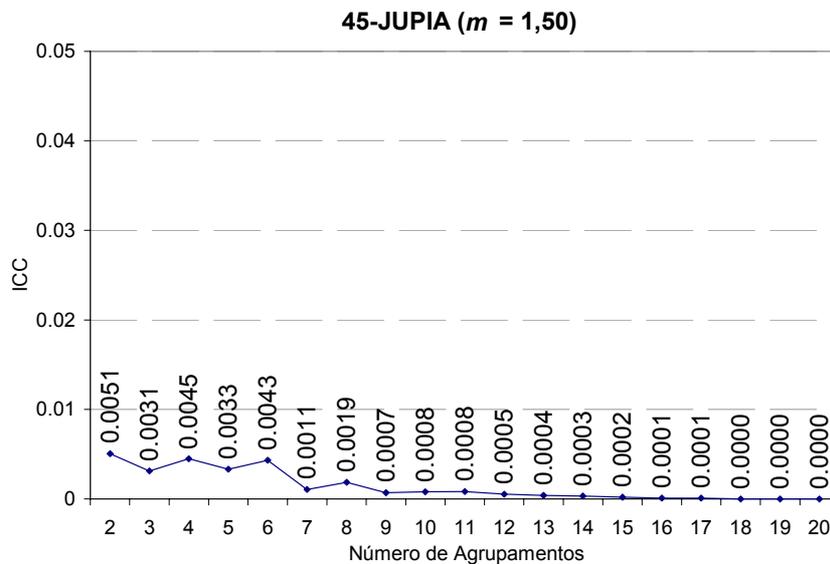


Figura 4.20 – Gráfico dos valores de *ICC* da usina de Jupiá para $m = 1,50$

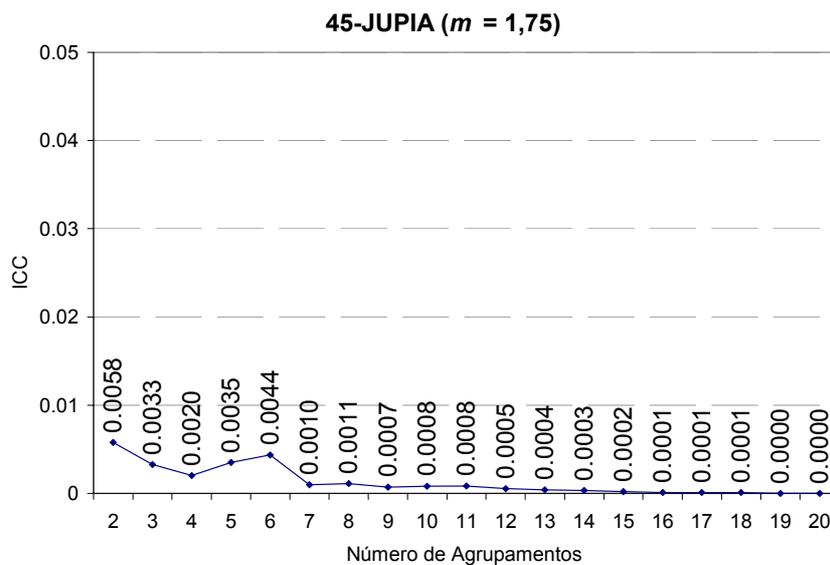


Figura 4.21 – Gráfico dos valores de ICC da usina de Jupiá para $m = 1,75$

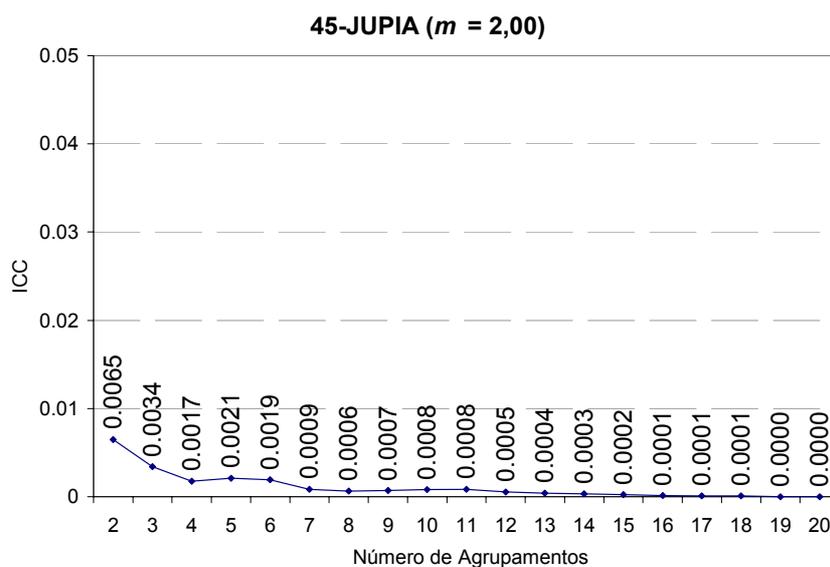


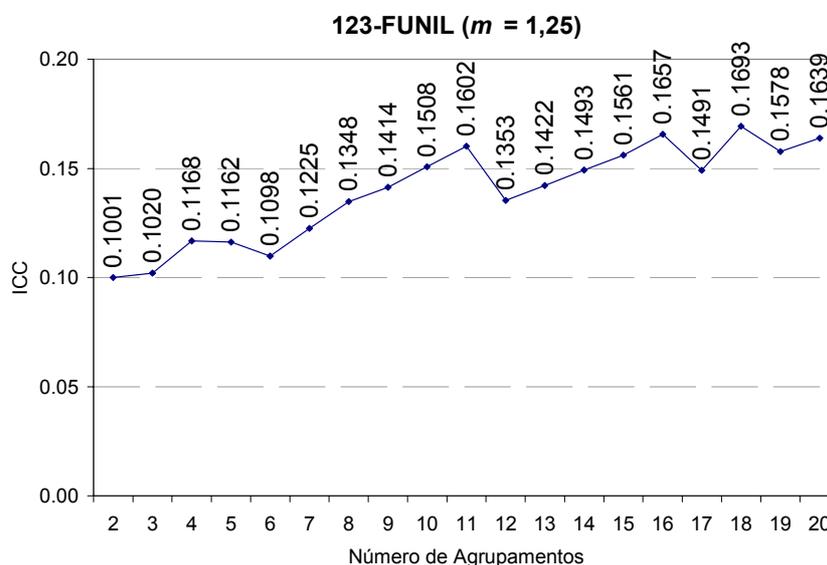
Figura 4.22 – Gráfico dos valores de ICC da usina de Jupiá para $m = 2,00$

A Tabela 4.3 apresenta os valores obtidos da ICC para a usina de Jupiá, para $c = 2$ e para cada um dos valores do coeficiente de nebulização utilizados. Nesta tabela podemos verificar que há a indicação de que o melhor valor do coeficiente de nebulização a ser utilizado para este espaço de dados é 2,00, pois obteve o maior valor da ICC entre os 4 valores apresentados.

Tabela 4.3 – Valores da ICC para 2 agrupamentos na usina de Jupιά

| Coefficiente de Nebulização (m) | ICC |
|-------------------------------------|--------|
| 1,25 | 0,0047 |
| 1,50 | 0,0051 |
| 1,75 | 0,0058 |
| 2,00 | 0,0065 |

Finalmente, a última usina selecionada foi a usina de Funil, cujo máximo do coeficiente de variação é 27,25 e a média do coeficiente de variação é 20,35. Esta usina foi selecionada por apresentar características de comportamento distintas da medida ICC para cada um dos valores do coeficiente de nebulização utilizados.

Figura 4.23 – Gráfico dos valores de ICC da usina de Funil para $m = 1,25$

Para m igual a 1,25 e 1,50, os resultados da medida cresceram conforme o número de agrupamentos aumentou. Sendo que para m igual a 1,50, aparentemente fica definido um valor máximo da medida em $c = 18$, mas não se tem garantia de que o fenômeno de crescimento continuaria se a análise houvesse sido feita para um número máximo de agrupamentos maior.

Para m igual a 1,75, fica bem definido o valor máximo da medida em 10 agrupamentos, o que poderia indicar que este é o melhor valor do coeficiente de nebulização para o espaço de dados desta usina.

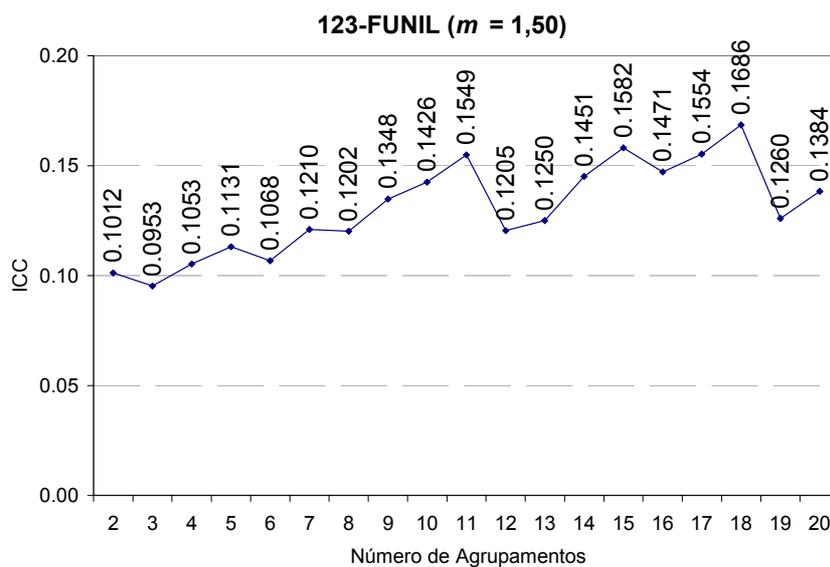


Figura 4.24 – Gráfico dos valores de ICC da usina de Funil para $m = 1,50$

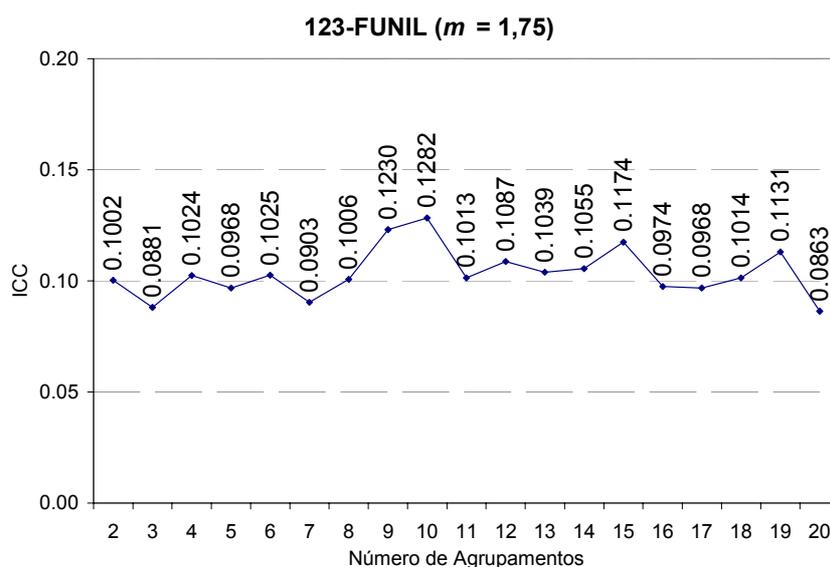


Figura 4.25 – Gráfico dos valores de ICC da usina de Funil para $m = 1,75$

Por fim, no teste que utilizou m igual a 2,00, foi encontrada uma característica muito interessante, foram definidos dois máximos, o primeiro em 10 agrupamentos e o segundo em 12 agrupamentos, com o mesmo valor da ICC. Estes valores de máximo encontram-se ressaltados na Figura 4.26.

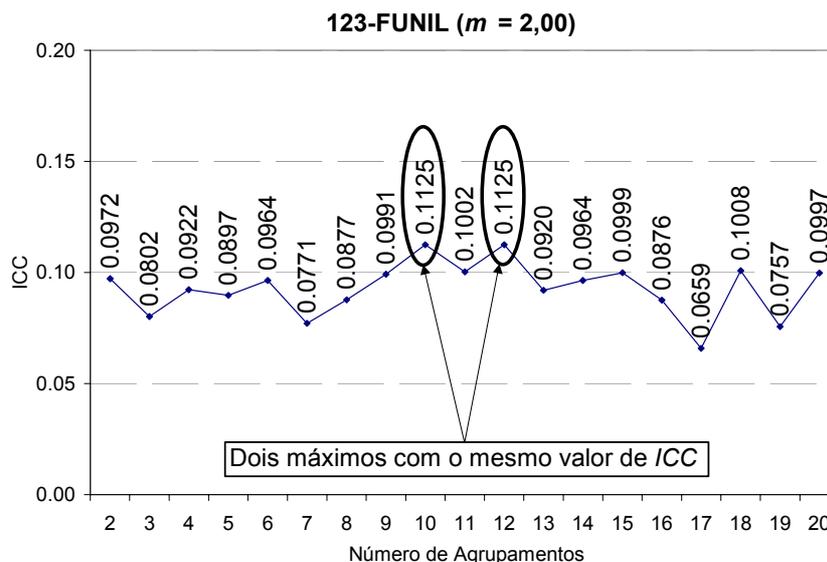


Figura 4.26 – Gráfico dos valores de ICC da usina de Funil para $m = 2,00$

A coincidência de valores não caracteriza um erro na determinação dos valores da ICC. O que fica caracterizado é que a distância mínima e o espalhamento dos centros em ambos os casos é bastante semelhante. Este fenômeno indica que é necessário determinar, além do número máximo de agrupamentos, qual deve ser o critério de desempate neste tipo de situação, ou seja, qual o número de agrupamentos que deve ser utilizado.

4.3 RESULTADOS E CONCLUSÕES SOBRE OS TESTES

Este item tem por objetivo reunir e discutir as conclusões apresentadas ao longo dos dois testes realizados com a medida ICC.

Para o primeiro conjunto de testes, no qual o espaço de dados possuía 360 dimensões, a resposta obtida da ICC quando o coeficiente de nebulização utilizado no FCM foi igual a 1,25 indicava que a medida apresentava problemas na definição de um máximo, visto que o valor da ICC foi crescente em todo o intervalo de agrupamentos testado.

Por outro lado, também foi verificado que para outros valores do coeficiente de nebulização esta característica não se manteve, ou seja, foi determinado um número de agrupamentos onde ocorreu um valor máximo para ICC.

Este tipo de comportamento foi verificado novamente nos resultados obtidos no segundo conjunto de testes, o que intensificou a sinalização de que a medida ICC pode ser

também utilizada para a determinação do melhor valor do coeficiente de nebulização para o espaço de dados.

Uma característica comum a todos os resultados obtidos para a medida *ICC* neste segundo conjunto de testes foi que os valores obtidos são muito pequenos, na casa dos centésimos ou milésimos.

Existem dois fatores que influenciaram para obtenção de resultados com valores tão baixos: o primeiro está ligado ao fato de se estar utilizando valores normalizados no intervalo $[0, 1]$ e uma dimensionalidade mais baixa que no espaço de dados utilizado no primeiro conjunto de testes; o segundo está ligado as características de sobreposição dos dados, ou seja, quanto menor o valor do coeficiente de variação determinado para o espaço de dados tanto mais baixos foram os resultados de *ICC*.

No primeiro conjunto de testes foram avaliadas 4 situações distintas, uma para cada valor do coeficiente de nebulização. Já no segundo conjunto de teste foram avaliadas 4 situações para cada uma das 30 usinas selecionadas para participar dos processos de agrupamento. Ou seja, somando-se as situações avaliadas nos dois conjuntos de testes tem-se um total de 124 casos distintos em que a medida foi avaliada. Deste universo foi possível verificar que em 29 casos, ou seja, em 23,4 % dos casos, não foi possível determinar um máximo para a medida.

Na Tabela 4.4 foi apresentada a contagem do número de casos em que a medida *ICC* não determinou um valor máximo em relação ao valor do coeficiente de nebulização utilizado e os números identificadores destas usinas no modelo SUSHI-O.

Nesta tabela pode ser verificado que a maior concentração do número de casos em que não ficou definido um máximo para a medida *ICC* estava associada aos valores mais baixos do coeficiente de nebulização.

Tabela 4.4 – Comparação do número de casos em que não foi determinado um máximo para a medida *ICC* e o valor do coeficiente de nebulização *m*

| Coeficiente de Nebulização <i>m</i> | Número de Casos | Número das Usinas no SUSHI-O |
|--|------------------------|---|
| 1,25 | 15 | 1, 14, 15, 17, 18, 30, 33, 47, 49, 51, 61, 120, 122, 123, 144 |
| 1,50 | 10 | 1, 14, 15, 30, 33, 51, 120, 122, 123, 144 |
| 1,75 | 3 | 15, 120, 144 |
| 2,00 | 1 | 120 |
| Total | 29 | |

Com base nos resultados obtidos, chegou-se a conclusão que a medida *ICC* pode ser utilizada na validação dos espaços de dados testados, pois mesmo nos casos em que a característica de crescimento do seu valor foi detectada a utilização de um valor máximo para o número de agrupamentos compatível com o número de amostras no espaço de dados faz com que o número de agrupamentos seja determinado.

O principal resultado obtido a partir dos testes foi a verificação de que a *ICC* pode ser utilizada, não somente para determinar o melhor número de agrupamentos para um espaço de dados, mas também para determinar o melhor valor do coeficiente de nebulização a ser utilizado no FCM. Este resultado pode ser validado pela análise da terceira coluna da Tabela 4.4, que apresenta os números das usinas no modelo SUIISHI-O. Esta coluna apresenta uma lista de números que diminui conforme o valor do coeficiente de nebulização m aumenta, mas, por exemplo, os valores apresentados na lista para m igual a 1,50 faziam parte da lista associada ao coeficiente de nebulização anterior, ou seja, m igual a 1,25.

5 AVALIAÇÃO DE CONFIABILIDADE EM SISTEMAS HIDROTÉRMICOS

Neste capítulo são avaliados dois casos de análise de confiabilidade composta avaliando a variação hidrológica, um caso exemplo baseado no sistema IEEE-RTS e o segundo baseado na configuração do subsistema Sudeste/Centro-Oeste do sistema elétrico brasileiro. Em ambos os casos o intuito foi verificar o impacto sobre os índices de confiabilidade decorrente da variação de potência disponível, sendo que no segundo procurou-se determinar este impacto utilizando um caso envolvendo um sistema de grande porte, também avaliando o tempo de processamento necessário para executar esta avaliação.

Em estudos de confiabilidade, as usinas que compõem o sistema elétrico são representadas por conjuntos de unidades geradoras com características semelhantes, ou seja, uma usina é um conjunto de equipamentos que apresentam valores semelhantes para: taxa média de falha, expressa em falhas/ano; tempo médio de reparo, expresso em horas; potências ativas máxima e mínima, expressas em MW; e potências reativas máxima e mínima, expressas em Mvar.

Os limites de geração de potência ativa e reativa para estas unidades são geralmente representados por valores constantes. Esta forma de representação é adequada para sistemas predominantemente térmicos, onde estes limites são função da potência nominal dos geradores.

Entretanto, a capacidade de geração das usinas hidrelétricas não é função somente de suas características próprias, mas também do nível de armazenamento em seus reservatórios. Como o nível de armazenamento dos reservatórios das usinas hidrelétricas é função das condições hidrológicas e da afluência dos rios que compõem as bacias em que a usina encontra-se instalada, a capacidade máxima de geração destas usinas apresenta variabilidade ao longo do tempo. Conseqüentemente, em sistemas com predominância hidrelétrica, os estudos de confiabilidade devem levar em conta a variação do limite máximo da potência ativa das usinas hidrelétricas em função das condições hidrológicas.

No NH2 os dados das usinas são indicados no código DUSI, portanto, para realizar a representação da variação da potência disponível, devem ser criados vários estudos variando o valor da potência máxima dos geradores das usinas hidrelétricas neste código.

5.1 AVALIANDO A VARIAÇÃO HIDROLÓGICA – CASO EXEMPLO

Este caso exemplo foi elaborado com o intuito de verificar o impacto sobre os índices de confiabilidade decorrente da variação de potência disponível gerada pela variação hidrológica em sistemas onde a capacidade de geração é predominantemente hidrelétrica. Para tanto, será analisado o sistema IEEE-RTS. Este caso é largamente utilizado na literatura como exemplo para estudos de confiabilidade composta, como em GRIGG (1999), PEREIRA (1992), PINHEIRO (1998), MELLO (1994) e MARCATO (1998).

O sistema IEEE-RTS é composto por 24 barras e 38 circuitos (GRIGG, 1999). Neste sistema, existem 10 barras de geração, cujo somatório dos limites máximos de geração é de 5621 MW. A Figura 5.1 apresenta o seu diagrama unifilar.

De forma a simular que este sistema é predominantemente hidráulico, foi suposto que as usinas conectadas às barras 13, 15, 16, 18, 22 e 23, são usinas hidrelétricas. Desta forma, 3845 MW, ou seja, 68,4% da capacidade instalada, provêm de usinas hidrelétricas.

Os dados do sistema foram tratados de forma a criar 6 situações distintas em relação ao valor da potência disponível de cada uma das usinas consideradas como hidrelétricas. A forma de tratamento é descrita a seguir.

- Caso 1 – Todos os limites de geração foram mantidos em seus valores máximos.
- Caso 2 – Todos os limites de geração de todas as usinas foram decrescidos em 10%.
- Caso 3 – Todos os limites de geração de todas as usinas foram decrescidos em 20%.
- Caso 4 – Os limites de geração das usinas 13 e 15 foram mantidos em seus valores máximos, das usinas 16 e 18 foram decrescidos em 10%, e das usinas 22 e 23 foram decrescidos em 20%.
- Caso 5 – Os limites de geração das usinas 16 e 18 foram mantidos em seus valores máximos, das usinas 22 e 23 foram decrescidos em 10%, e das usinas 13 e 15 foram decrescidos em 20%.
- Caso 6 – Os limites de geração das usinas 22 e 23 foram mantidos em seus valores máximos, das usinas 13 e 15 foram decrescidos em 10%, e das usinas 16 e 18 foram decrescidos em 20%.

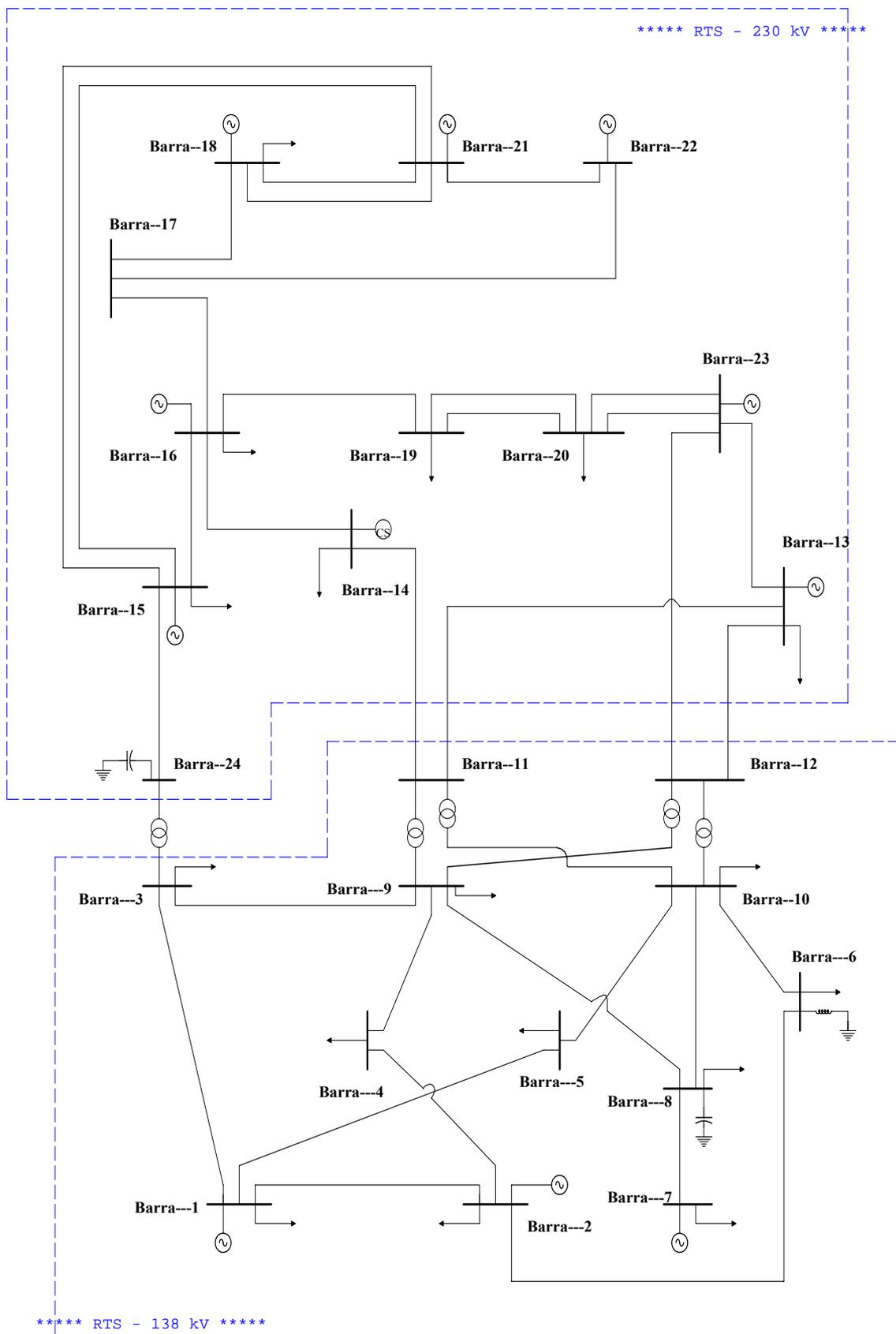


Figura 5.1 – Diagrama unifilar do sistema IEEE-RTS

Os valores totais das potências máximas de cada usina em cada um dos casos formados são apresentados na Tabela 5.1. Como já foi descrito no item 2.4.2, onde é feita uma descrição do sistema NH2, este sistema é capaz de calcular os índices de confiabilidade, por simulação de Monte Carlo. Neste trabalho foi utilizado como critério de parada a convergência do coeficiente de variação da LOLP igual a 5% e da EPNS igual a 5%. Os resultados obtidos encontram-se representados na Tabela 5.2.

Tabela 5.1 – Limites máximos de geração considerados nos 6 casos do exemplo

| | Caso 1 | Caso 2 | Caso 3 | Caso 4 | Caso 5 | Caso 6 |
|-----------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| Usina | Potência (MW) |
| Barra 13 | 788,0 | 709,2 | 630,4 | 788,0 | 630,4 | 709,2 |
| Barra 15 | 392,0 | 343,8 | 305,6 | 392,0 | 305,6 | 343,8 |
| Barra 16 | 310,0 | 279,0 | 248,0 | 279,0 | 310,0 | 248,0 |
| Barra 18 | 800,0 | 720,0 | 640,0 | 720,0 | 800,0 | 640,0 |
| Barra 22 | 400,0 | 360,0 | 320,0 | 320,0 | 360,0 | 400,0 |
| Barra 23 | 1165,0 | 1048,5 | 932,0 | 932,0 | 1048,5 | 1165,0 |
| Total | 3845,0 | 3460,5 | 3076,0 | 3431,0 | 3454,5 | 3506,0 |

Tabela 5.2 – Resultados da análise de confiabilidade do caso exemplo

| C | Pot. Máx. (MW) | LOLP | | EPNS (MW) | | LOLE (h/ano) | EENS (MWh/ano) | Monte Carlo¹ | | |
|----------|-----------------------|--------------|---------------|------------------|---------------|---------------------|-----------------------|--------------------------------|--------------------------|--------------------------|
| | | Valor | CV (%) | Valor | CV (%) | | | Sim.² | Sist.³ | Corte⁴ |
| 1 | 5621,0 | 0,0200 | 3,53 | 2,47 | 5,00 | 175,52 | 21676,10 | 39328 | 10600 | 788 |
| 2 | 5236,5 | 0,0799 | 3,48 | 12,20 | 5,00 | 699,80 | 106893,87 | 9501 | 2799 | 759 |
| 3 | 4852,0 | 0,3164 | 3,44 | 67,57 | 5,00 | 2771,36 | 591888,81 | 1827 | 983 | 578 |
| 4 | 5197,0 | 0,0866 | 3,55 | 14,18 | 5,00 | 758,42 | 124177,10 | 8374 | 2504 | 725 |
| 5 | 5230,5 | 0,1082 | 3,33 | 16,16 | 5,00 | 947,73 | 141559,58 | 7450 | 2814 | 806 |
| 6 | 5282,0 | 0,0574 | 3,57 | 8,81 | 4,99 | 503,03 | 77202,36 | 12904 | 3653 | 741 |

Da forma como este exemplo está sendo trabalhado, os casos simularam, a partir da situação normal, representada pelo caso 1, a ocorrência de um período seco refletido na

¹ Resultados provenientes da simulação de Monte Carlo para determinação dos índices de confiabilidade.

² Número de eventos simulados e solucionados

³ Número de eventos críticos com problemas no sistema

⁴ Número de eventos críticos com corte de carga

diminuição da potência disponível de todas as usinas hidrelétricas, representada nos casos 2 e 3, e a variação das condições hidrológicas sobre conjuntos de usinas distintos, representada nos casos 4, 5 e 6.

Como já foi descrito no item 2.3.2, onde é feita uma descrição do sistema NH2, este sistema é capaz de calcular os índices de confiabilidade, por simulação de Monte Carlo, com critério de parada pela convergência do coeficiente de variação da LOLP igual a 5% e da EPNS igual a 5%. Os resultados obtidos encontram-se representados na Tabela 5.2.

Na Tabela 5.3 são apresentadas as estatísticas relativas ao caso exemplo. Esta tabela foi montada com base nos resultados apresentados na Tabela 5.2, pode ser verificado que a alteração nos limites máximos de geração das usinas hidrelétricas pode acarretar em índices de confiabilidade bastante diferenciados. Estas estatísticas se resumem ao máximo, mínimo e média aritmética, por se tratar de um exemplo fictício.

Tabela 5.3 - Estatísticas do caso exemplo

| | Potência Máxima (MW) | LOLP | EPNS (MW) | LOLE (h/Ano) | EENS (MWh/Ano) |
|---------------|----------------------|----------|-----------|--------------|----------------|
| Máximo | 5620,9 | 0,316365 | 67,57 | 2771,36 | 591888,80 |
| Média | 5236,5 | 0,111413 | 20,23 | 975,98 | 177233,00 |
| Mínimo | 4852,0 | 0,020036 | 2,47 | 175,52 | 21676,10 |

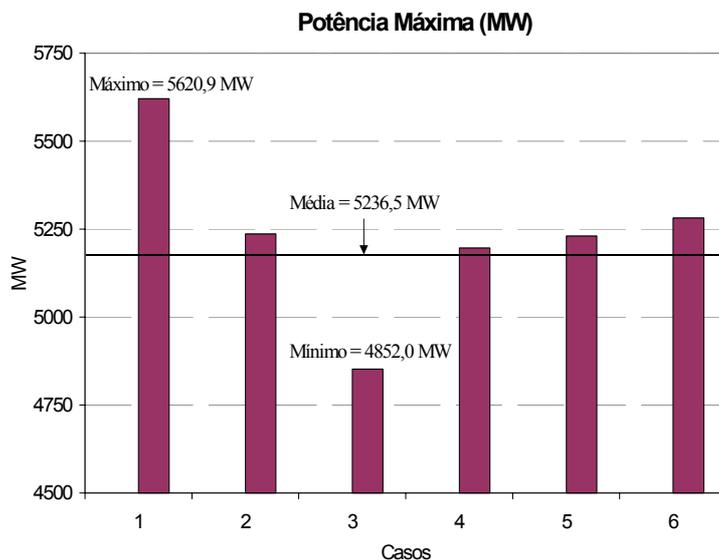


Figura 5.2 – Potências máximas nos casos do exemplo

A Figura 5.2, Figura 5.3 e Figura 5.4 apresentam os gráficos que nos permitem visualizar o efeito da variação da potência disponível sobre os principais índices de confiabilidade calculados pelo NH2.

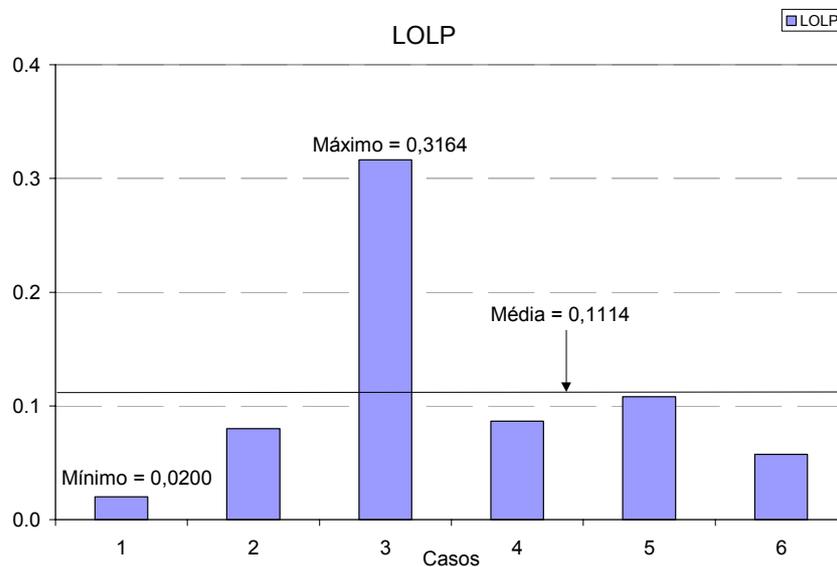


Figura 5.3 – LOLP dos casos de confiabilidade do exemplo

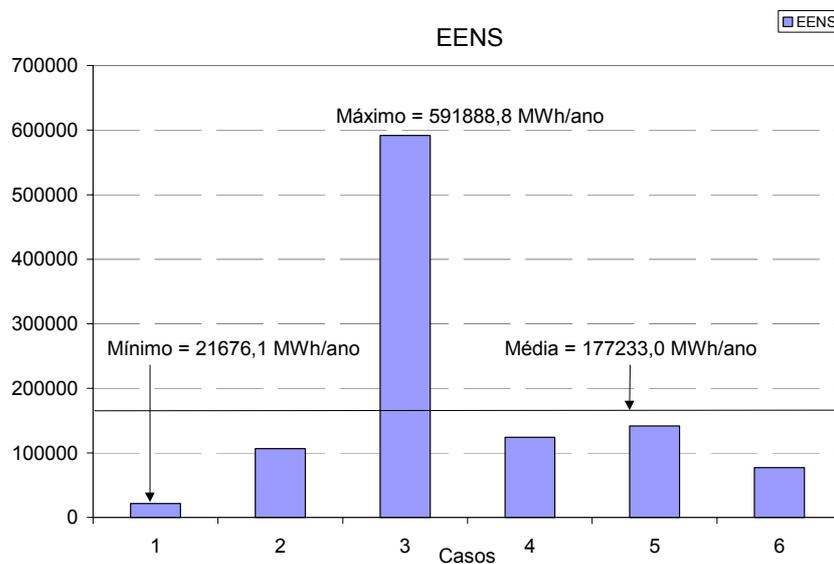


Figura 5.4 – EENS dos casos de confiabilidade do exemplo

Deve ser ressaltado, no entanto, que a variação hidrológica analisada neste caso exemplo foi criada artificialmente a partir dos dados existentes para o sistema RTS, portanto, por ser fictícia, é pouco provável de ser verificada em casos reais.

5.2 AVALIANDO A VARIAÇÃO HIDROLÓGICA – CASO SUDESTE/CENTRO-OESTE

O exemplo anterior mostrou que a variação da potência disponível das usinas hidrelétricas pode influenciar nos índices de confiabilidade de um sistema elétrico de forma significativa, provando ser importante a análise desta variação.

Neste item, foi avaliada a variação hidrológica para um caso do sistema elétrico brasileiro, mais especificamente o caso do subsistema Sudeste/Centro-Oeste, utilizando para tanto as séries de potência disponível calculadas pelos modelos energéticos a partir das séries históricas de vazão.

O objetivo principal foi determinar o impacto da variação hidrológica sobre os índices de confiabilidade de um caso envolvendo um sistema de grande porte. Esta avaliação também levou em conta a complexidade e o esforço computacional necessário para efetuar esta análise.

Nesta dissertação, o estudo energético foi realizado pelo NEWAVE utilizando os dados relativos ao Plano Mensal de Operação, PMO, de Junho de 2001, com o período de estudo de 5 anos. O modelo NEWAVE foi executado exatamente como este é utilizado pelo ONS para realizar o planejamento da operação, ou seja a simulação da operação foi realizada utilizando 2000 séries sintéticas de vazão.

Para o SUISEI-O, foram utilizados os resultados obtidos pelo NEWAVE, mas neste caso a simulação da operação foi realizada utilizando as séries históricas de vazão.

Por se tratar de dados de 2001 do sistema elétrico brasileiro, encontravam-se representadas, nos dados de entrada de ambos os modelos, 94 usinas hidrelétricas para as quais estavam disponíveis 65 anos de séries históricas de vazões.

Este cenário foi escolhido por marcar o início do período de racionamento, em que os reservatórios estavam com seus níveis de armazenamento baixos e enquanto que os estudos de previsão de carga do sistema permaneciam indicando uma tendência de crescimento.

O estudo de confiabilidade composta foi realizado utilizando dados do subsistema Sudeste/Centro-Oeste, com representação da interligação deste com o subsistema Sul, no NH2. Este caso possui 2293 barras e 3554 circuitos, onde se encontram representadas 56 das 94 usinas hidrelétricas representadas no estudo energético.

O fluxo de informações entre o NEWAVE, o SUIISHI-O e o NH2 é representado na Figura 5.5. Nesta figura foram representados os dados mais relevantes para solução deste problema.

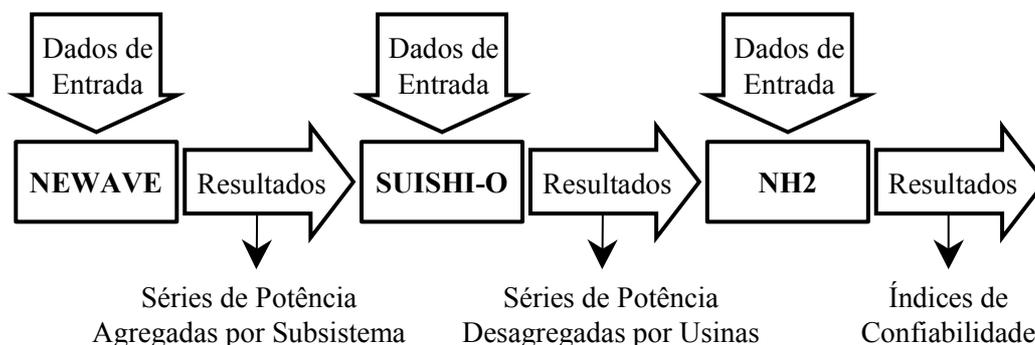


Figura 5.5 – Fluxo de informações entre o NEWAVE, o SUIISHI-O e o NH2.

De forma resumida, o fluxo apresentado na Figura 5.5 indica que o NEWAVE utiliza como dados de entrada a configuração do sistema no mês/ano inicial do estudo, os dados relativos ao planejamento da expansão do sistema relativos ao período de estudo e as séries de vazão, para calcular a política operação do sistema (PMO) agregada por subsistema para o período de estudo.

Os resultados do NEWAVE são repassados ao SUIISHI-O que efetua a sua desagregação, calculando a potência disponível para cada usina do sistema em cada mês do período de estudo.

Levando-se em conta que existem 65 séries históricas de vazão, ao finalizar o estudo energético obteremos 65 séries de potência disponível para cada uma das usinas representadas no caso.

A interpretação destas séries de potência disponível é: caso uma série histórica de vazão de uma determinada usina se repita no período de estudo, a potência disponível da usina será a calculada pelo modelo.

As usinas que compõem este subsistema e que participam do caso que será analisado encontram-se listadas na Tabela 5.4.

Tabela 5.4 – Usinas hidráulicas que compõem o subsistema Sudeste/Centro-Oeste

| Núm. ⁵ | Nome | Núm. | Nome | Núm. | Nome |
|-------------------|-----------------------|------|-------------------|------|-----------------|
| 1 | Camargos ⁶ | 2 | Itutinga | 6 | Furnas |
| 7 | Mascarenhas de Moraes | 8 | Estreito | 9 | Jaguará |
| 10 | Igarapava | 11 | Volta Grande | 12 | Porto Colômbia |
| 14 | Cancode | 15 | Euclides da Cunha | 16 | A. S. Oliveira |
| 17 | Marimbondo | 18 | Água Vermelha | 24 | Emborcação |
| 25 | Nova Ponte | 26 | Miranda | 30 | Corumbá I |
| 31 | Itumbiara | 32 | Cachoeira Dourada | 33 | São Simão |
| 37 | Barra Bonita | 38 | A. S. Lima | 39 | Ibitinga |
| 40 | Promissão | 42 | Nova Avanhandava | 44 | Ilha Solteira |
| 45 | Jupia | 46 | Porto Primavera | 47 | A. A. Laydner |
| 49 | Chavantes | 50 | L. N. Garcez | 51 | Canoas II |
| 52 | Canoas I | 61 | Capivara | 62 | Taquaruçu |
| 63 | Rosana | 66 | Itaipú | 119 | Henry Borden |
| 120 | Jaguarí | 121 | Paraibuna | 122 | Santa Branca |
| 123 | Funil | 127 | Sobragi | 130 | Ilha dos Pombos |
| 131 | Nilo Peçanha | 132 | Fontes | 133 | Perreira Passos |
| 134 | Salto | 144 | Mascarenhas | 156 | Três Marias |
| 192 | Guilman Amorim | 193 | Sá Carvalho | 217 | Rosal |
| 251 | Serra da Mesa | 278 | Manso | | |

O NH2 não permite a análise ao longo do tempo, pois seu processamento é realizado utilizando as características do sistema em um instante do tempo. Por isto, deve ser selecionado um mês/ano da série de potência disponível para formar os casos do NH2, ou seja, foram obtidos 65 casos distintos para o NH2, cada um refletindo uma configuração de variação hidrológica no código DUSI (Dados de Usinas) associados a um determinado período do horizonte de planejamento.

Apesar de na descrição anterior o acoplamento entre os três sistemas parecer simples, na realidade existem alguns problemas.

⁵ Número identificados da usina nos dados do modelo SUIISHI-O.

⁶ No estudo de confiabilidade composta as usinas de Camargos e Itutinga são consideradas como uma única usina, pois se encontram conectadas à mesma barra do sistema.

O primeiro problema está diretamente ligado à diferença da definição de uma usina nos modelos energéticos NEWAVE e SUIISHI-O, e no NH2. Os resultados dos modelos energéticos são fornecidos em valores totais para uma instalação física, a qual é a sua representação de uma usina hidrelétrica. No NH2, esta mesma instalação física pode encontrar-se representada em vários registros de usina no código DUSI, pois para este sistema uma usina é representada por um grupo de equipamentos com características iguais ligados a uma determinada barra do sistema. Ou seja, para poder formar os casos deste sistema, valores resultantes do sistema energético de cada usina devem ser repartidos entre os geradores representados no código DUSI.

Este problema foi solucionado por meio de uma tabela de ligação entre os números das usinas no SUIISHI-O e os seus respectivos números das barras no NH2, ou seja, os números que identificam as usinas no SUIISHI-O e o número das barras do NH2 que representam estas usinas foram associados.

O segundo problema encontrado também diz respeito a forma de representação da usina. Nos modelos energéticos a potência disponível das usinas é calculada em valores médios mensais que podem variar em uma faixa contínua entre zero e o seu máximo de potência instalada. No NH2 isto não é permitido pois os equipamentos possuem valores mínimos para geração.

Por exemplo, na configuração que está sendo utilizada neste trabalho, a usina de Jaguará é constituída por 4 máquinas com capacidade de geração entre 80 e 100 MW. Para os modelos energéticos esta usina possui a capacidade de geração entre 0 e 400 MW. Mas, como pode ser verificado no gráfico da Figura 5.6, os limites impostos pelas máquinas da usina criam intervalos de potência em que a usina pode operar, representadas pelas faixas escuras da figura, e intervalos onde não é possível a usina operar, como por exemplo entre 0 e 79 MW ou entre 101 e 159 MW. Assim, o modelo energético poderia indicar que em um determinado ponto do período de estudo a potência disponível usina seria 220 MW, o que não é possível, pois este valor encontra-se em uma faixa que a usina não pode operar.

De forma a evitar este problema, foi considerado no código DUSI do NH2 que todas as usinas possuem potência mínima igual a zero, desta forma as faixas de operação se tornaram sobrepostas.

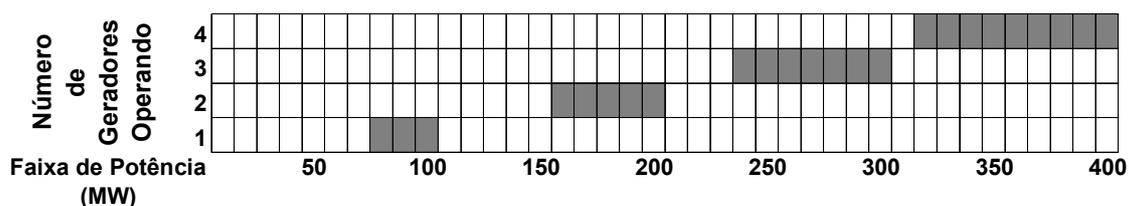


Figura 5.6 – Gráfico das faixas de potência da usina de Jaguará

De posse desta premissa, e da tabela de ligação entre o SUIISHI-O e o NH2, foi possível realizar a repartição dos valores de potência disponível calculados no primeiro de forma a montar os casos a ser utilizados pelo segundo. Esta repartição foi realizada segundo o algoritmo abaixo:

Passo 1. Selecionar as usinas do SUIISHI-O que participam do caso do NH2.

Passo 2. Relacionar estas usinas com as barras em que estas se encontram conectadas, ordenando-as em forma decrescente de potência máxima.

Passo 3. Para cada usina do SUIISHI-O.

Passo 3.1. Selecionar o valor da potência disponível (PotDisp).

Passo 3.2. Para cada barra do NH2 relacionar a usina do SUIISHI-O.

Passo 3.2.1. Se a potência disponível (PotDisp) for maior ou igual a produto do número de geradores pela potência máxima ($\text{NumMaqs} * \text{PotMax}$) existente na barra

O número de geradores (NumMaqs) e a potência máxima no registro da barra são mantidos.

A potência disponível é decrescida

$$(\text{PotDisp} = \text{PotDisp} - (\text{NumMaqs} * \text{PotMax})).$$

Retorna ao Passo 3.2.

Senão

Procura pelo número inteiro de geradores (NumMaqsAlt) capaz de produzir a potência disponível indicada (PotDisp).

Atribui a potência máxima do registro o resultado da divisão da potência disponível pelo número de geradores encontrado

$$(\text{PotMax} = \text{PotDisp} / \text{NumMaqsAlt}).$$

Retorna ao Passo 3.

O terceiro problema diz respeito a não existência uma base única de dados sobre o Sistema Elétrico Brasileiro. O NEWAVE, o SUIISHI-O e o NH2 trabalham com arquivos textos

e os dados existentes podem ser adquiridos de fontes diferentes. Isto acarreta em diferenças nos arquivos, como a diferença de nomenclatura das usinas, por exemplo, no SUIISHI-O a usina número 38 é denominada A. S. Lima, enquanto que a mesma usina no NH2 é denominada Bariri, para citar um caso de fácil solução.

Em alguns casos os valores das capacidades de geração não são iguais o que torna ainda mais difícil a determinação de quais valores utilizar.

No presente trabalho foram adotados os nomes de usinas utilizados no sistema SUIISHI-O e os valores das potências máximas dos geradores ligados as barras do NH2 foram alterados de forma a tornarem-se compatíveis com os valores indicados nos dados de entrada do NEWAVE.

Adotadas estas premissas, foi realizado um estudo de confiabilidade onde caso analisado representa todas as usinas do subsistema Sudeste/Centro-Oeste no código DUSI com seus valores máximos de potência para todas as usinas nele representadas. Os resultados da análise de confiabilidade para este caso são:

- Potência Máxima = 60356,4 MW
- LOLP = 0,212545 com coeficiente de variação = 4,213 %
- EPNS = 2,0062 MW com coeficiente de variação = 14,889 %
- LOLE = 1861,90 h/ano
- EENS = 17574,72 MWh/ano
- Tempo de Processamento = 3 hs e 53 min.

Dando prosseguimento a análise, selecionou-se o mês de março de 2002 para formar os 65 casos representativos da variação na potência disponível. Os resultados destes casos são apresentados na Tabela 5.5, e serão referenciados a partir deste ponto como *Casos do Histórico*.

Todos os estudos de confiabilidade realizados neste item foram efetuados utilizando o NH2, efetuando a análise de confiabilidade composta utilizando simulação de Monte Carlo, com critério de parada pela convergência o coeficiente de variação da LOLP igual a 5% e da EPNS igual a 15%.

Para fins de comparação, buscou-se no conjunto de resultados os valores máximos e mínimos dos índices de confiabilidade. Como as séries históricas de vazão utilizadas no estudo energético são equiprováveis, e conseqüentemente as séries de potência disponível calculadas

também o são, é possível calcular o valor médio dos índices de confiabilidade utilizando a média aritmética. Estes indicadores são apresentados na Tabela 5.6.

Tabela 5.5 – Resultados da análise de confiabilidade utilizando todos os valores das potências disponíveis obtidas a partir do histórico (Casos do Histórico)

| Série | Pot. Máx. (MW) | LOLP | | EPNS (MW) | | LOLE (h/ano) | EENS (MWh/ano) | Monte Carlo ⁷ | | |
|-------|----------------|--------|--------|-----------|--------|--------------|----------------|--------------------------|--------------------|--------------------|
| | | Valor | CV (%) | Valor | CV (%) | | | Sim. ⁸ | Sist. ⁹ | Cort ¹⁰ |
| 1 | 58956,1 | 0,2224 | 5,00 | 1,7225 | 12,24 | 1948,59 | 15088,98 | 2743 | 1898 | 518 |
| 2 | 56662,9 | 0,2333 | 4,99 | 1,8939 | 12,22 | 2043,60 | 16590,96 | 1578 | 1263 | 341 |
| 3 | 58280,5 | 0,2290 | 5,00 | 1,7887 | 11,49 | 2006,29 | 15669,05 | 2602 | 2025 | 553 |
| 4 | 57037,7 | 0,2241 | 3,87 | 1,7061 | 14,99 | 1962,72 | 14945,59 | 2486 | 1920 | 557 |
| 5 | 57757,1 | 0,1958 | 4,17 | 1,7786 | 15,00 | 1715,53 | 15580,40 | 2594 | 1848 | 508 |
| 6 | 58764,5 | 0,1981 | 5,00 | 1,6221 | 11,50 | 1735,33 | 14209,53 | 1538 | 1183 | 337 |
| 7 | 58609,1 | 0,2244 | 5,00 | 1,9666 | 12,30 | 1965,71 | 17227,31 | 1542 | 1192 | 344 |
| 8 | 58513,7 | 0,1873 | 4,23 | 1,6884 | 14,97 | 1640,42 | 14790,21 | 2638 | 1873 | 494 |
| 9 | 57423,5 | 0,2101 | 5,00 | 2,0729 | 14,12 | 1840,34 | 18158,33 | 1617 | 1278 | 343 |
| 10 | 58441,4 | 0,2256 | 5,00 | 1,3563 | 7,63 | 1976,16 | 11881,11 | 2598 | 1992 | 573 |
| 11 | 58774,4 | 0,2097 | 4,19 | 1,9250 | 15,00 | 1836,90 | 16862,92 | 2332 | 1831 | 489 |
| 12 | 57854,1 | 0,1927 | 5,00 | 1,5420 | 11,93 | 1688,15 | 13508,25 | 2536 | 1980 | 558 |
| 13 | 56855,6 | 0,1892 | 3,65 | 2,1469 | 15,00 | 1657,70 | 18806,43 | 3520 | 2529 | 668 |
| 14 | 60044,9 | 0,1971 | 5,00 | 1,6611 | 11,32 | 1726,90 | 14551,52 | 1546 | 1167 | 336 |
| 15 | 59134,3 | 0,2030 | 5,00 | 1,6667 | 11,44 | 1778,12 | 14600,17 | 2745 | 2012 | 543 |
| 16 | 59148,6 | 0,2096 | 5,00 | 1,7177 | 11,90 | 1836,25 | 15047,10 | 2424 | 1737 | 474 |
| 17 | 58299,3 | 0,1983 | 5,00 | 1,7325 | 11,91 | 1736,77 | 15177,01 | 1669 | 1242 | 342 |
| 18 | 57673,9 | 0,2626 | 5,00 | 2,1065 | 11,93 | 2300,00 | 18453,29 | 2414 | 1849 | 512 |
| 19 | 59260,0 | 0,1865 | 5,00 | 1,5842 | 11,17 | 1633,38 | 13877,85 | 2709 | 1944 | 505 |
| 20 | 56988,9 | 0,2192 | 4,99 | 1,9418 | 14,77 | 1920,38 | 17009,98 | 1481 | 1147 | 251 |
| 21 | 54195,1 | 0,2450 | 3,66 | 2,8296 | 14,99 | 2146,30 | 24787,28 | 2506 | 2137 | 614 |
| 22 | 55902,6 | 0,2149 | 4,16 | 2,1806 | 14,98 | 1882,28 | 19102,37 | 2313 | 1750 | 497 |
| 23 | 54749,6 | 0,2363 | 3,19 | 2,5379 | 15,00 | 2069,64 | 22232,02 | 3530 | 2987 | 834 |

⁷ Resultados provenientes da simulação de Monte Carlo para determinação dos índices de confiabilidade.

⁸ Número de eventos simulados e solucionados

⁹ Número de eventos críticos com problemas no sistema

¹⁰ Número de eventos críticos com corte de carga

| S é r i e | Pot. Máx. (MW) | LOLP | | EPNS (MW) | | LOLE (h/ano) | EENS (MWh/ano) | Monte Carlo ⁷ | | |
|-----------------------|----------------------|--------|-----------|-----------|-----------|-----------------|-------------------|--------------------------|--------------------|--------------------|
| | | Valor | CV (%) | Valor | CV (%) | | | Sim. ⁸ | Sist. ⁹ | Cort ¹⁰ |
| 24 | 56314,0 | 0,2180 | 5,00 | 1,7583 | 12,39 | 1909,52 | 15402,81 | 1684 | 1448 | 379 |
| 25 | 58029,9 | 0,2238 | 5,00 | 1,5640 | 10,81 | 1960,24 | 13700,30 | 1506 | 1171 | 337 |
| 26 | 58770,8 | 0,1949 | 4,20 | 1,8068 | 14,91 | 1707,43 | 15827,55 | 2555 | 1793 | 498 |
| 27 | 58409,7 | 0,2260 | 4,99 | 1,5741 | 11,49 | 1979,40 | 13788,80 | 2499 | 1885 | 525 |
| 28 | 57736,2 | 0,2197 | 4,99 | 1,8265 | 12,39 | 1924,16 | 16000,17 | 2460 | 1828 | 508 |
| 29 | 59546,9 | 0,1993 | 5,00 | 1,5744 | 11,87 | 1745,91 | 13791,88 | 1620 | 1178 | 340 |
| 30 | 57912,7 | 0,2308 | 5,00 | 1,9434 | 12,13 | 2021,54 | 17024,57 | 2639 | 1885 | 527 |
| 31 | 59227,2 | 0,1873 | 5,00 | 1,5632 | 10,97 | 1640,73 | 13693,68 | 1721 | 1245 | 345 |
| 32 | 57519,5 | 0,2060 | 4,99 | 1,2886 | 8,02 | 1804,14 | 11288,37 | 2486 | 1827 | 538 |
| 33 | 59676,7 | 0,1829 | 5,00 | 1,3861 | 10,38 | 1601,93 | 12142,50 | 1507 | 1110 | 339 |
| 34 | 59058,5 | 0,2028 | 4,22 | 2,0691 | 15,00 | 1776,81 | 18125,66 | 2401 | 1752 | 487 |
| 35 | 59456,1 | 0,2005 | 5,00 | 1,8173 | 12,02 | 1756,08 | 15919,53 | 1707 | 1248 | 342 |
| 36 | 57421,4 | 0,2231 | 4,00 | 2,1081 | 14,99 | 1953,96 | 18467,27 | 2403 | 1892 | 536 |
| 37 | 56535,6 | 0,2090 | 5,00 | 1,4004 | 11,34 | 1830,85 | 12267,76 | 2524 | 1908 | 541 |
| 38 | 59209,7 | 0,2090 | 5,00 | 1,7794 | 11,94 | 1830,77 | 15587,84 | 1403 | 1068 | 332 |
| 39 | 54127,9 | 0,2249 | 3,33 | 3,9813 | 14,99 | 1970,19 | 34876,43 | 3517 | 2984 | 791 |
| 40 | 58170,5 | 0,2088 | 5,00 | 1,4692 | 11,34 | 1828,78 | 12869,93 | 1643 | 1286 | 343 |
| 41 | 58876,4 | 0,2211 | 4,07 | 2,0851 | 15,00 | 1936,87 | 18265,87 | 2284 | 1736 | 505 |
| 42 | 59018,0 | 0,2090 | 4,18 | 2,0725 | 14,99 | 1830,67 | 18154,95 | 2316 | 1782 | 484 |
| 43 | 57367,7 | 0,2158 | 3,90 | 1,9246 | 14,99 | 1890,62 | 16859,20 | 2692 | 1999 | 581 |
| 44 | 58109,5 | 0,2031 | 5,00 | 1,8551 | 11,74 | 1778,81 | 16250,83 | 1583 | 1182 | 342 |
| 45 | 59429,4 | 0,2121 | 4,08 | 2,1019 | 14,96 | 1858,41 | 18413,02 | 2371 | 1813 | 503 |
| 46 | 58832,6 | 0,2221 | 4,99 | 1,6863 | 11,45 | 1945,38 | 14772,06 | 2673 | 1895 | 553 |
| 47 | 58704,1 | 0,2342 | 3,95 | 2,0252 | 14,98 | 2051,20 | 17740,75 | 2272 | 1790 | 532 |
| 48 | 60148,1 | 0,2173 | 5,00 | 1,4914 | 10,56 | 1903,37 | 13064,76 | 1658 | 1203 | 341 |
| 49 | 59573,1 | 0,2073 | 5,00 | 1,4802 | 10,26 | 1815,90 | 12966,45 | 1575 | 1149 | 339 |
| 50 | 59900,3 | 0,1983 | 5,00 | 1,1517 | 7,13 | 1736,68 | 10089,30 | 1586 | 1221 | 340 |
| 51 | 60010,6 | 0,2247 | 4,99 | 1,7922 | 10,86 | 1968,67 | 15700,07 | 2562 | 1872 | 511 |
| 52 | 59550,7 | 0,2160 | 5,00 | 1,8897 | 14,20 | 1891,81 | 16553,75 | 1579 | 1258 | 341 |
| 53 | 59073,1 | 0,2115 | 4,19 | 2,1938 | 14,99 | 1852,93 | 19218,06 | 2274 | 1773 | 481 |
| 54 | 57143,0 | 0,2206 | 3,91 | 1,6569 | 14,97 | 1932,35 | 14514,30 | 2516 | 1922 | 555 |
| 55 | 57857,6 | 0,1937 | 4,04 | 1,6324 | 14,99 | 1696,41 | 14299,95 | 2742 | 2109 | 531 |
| 56 | 58463,7 | 0,2027 | 5,00 | 1,6576 | 11,97 | 1775,96 | 14520,55 | 1682 | 1222 | 341 |
| 57 | 58596,5 | 0,2002 | 5,00 | 1,6733 | 11,54 | 1754,03 | 14657,86 | 1777 | 1374 | 343 |

| Série | Pot. Máx. (MW) | LOLP | | EPNS (MW) | | LOLE (h/ano) | EENS (MWh/ano) | Monte Carlo ⁷ | | |
|-------|----------------|--------|--------|-----------|--------|--------------|----------------|--------------------------|--------------------|--------------------|
| | | Valor | CV (%) | Valor | CV (%) | | | Sim. ⁸ | Sist. ⁹ | Cort ¹⁰ |
| 58 | 59604,7 | 0,2162 | 5,00 | 1,5977 | 11,97 | 1893,60 | 13996,16 | 1417 | 1093 | 332 |
| 59 | 58586,6 | 0,1992 | 4,99 | 1,4679 | 12,43 | 1745,04 | 12858,53 | 1669 | 1242 | 342 |
| 60 | 59313,1 | 0,2230 | 5,00 | 1,7023 | 11,65 | 1953,09 | 14911,74 | 2486 | 1920 | 557 |
| 61 | 59362,6 | 0,2125 | 4,23 | 2,1033 | 15,00 | 1861,65 | 18425,16 | 2221 | 1689 | 472 |
| 62 | 59172,7 | 0,2125 | 4,99 | 1,6998 | 11,42 | 1861,84 | 14890,12 | 2231 | 1678 | 482 |
| 63 | 59033,5 | 0,1976 | 5,00 | 1,6465 | 11,52 | 1730,75 | 14423,67 | 1660 | 1232 | 350 |
| 64 | 57971,1 | 0,2400 | 4,99 | 1,6354 | 12,03 | 2102,15 | 14325,79 | 1421 | 1173 | 341 |
| 65 | 59622,5 | 0,2057 | 5,00 | 1,7407 | 11,89 | 1802,12 | 15248,38 | 1643 | 1286 | 343 |

Tabela 5.6 – Estatísticas dos casos do histórico

| | Potência Máxima (MW) | LOLP | EPNS (MW) | LOLE (h/Ano) | EENS (MWh/Ano) |
|-------------------------------------|----------------------|----------|-----------|--------------|----------------|
| Máximo | 60148,1 | 0,262557 | 3,9813 | 2300,00 | 34876,43 |
| Mínimo | 54127,9 | 0,182869 | 1,1517 | 1601,93 | 10089,30 |
| Média | 58304,2 | 0,212259 | 1,8160 | 1859,39 | 15908,52 |
| Desvio Padrão | 1310,7 | 0,0154 | 0,3931 | 134,85 | 3443,32 |
| Coefficiente de Variação (%) | 2,25 | 7,25 | 21,64 | 7,25 | 21,64 |

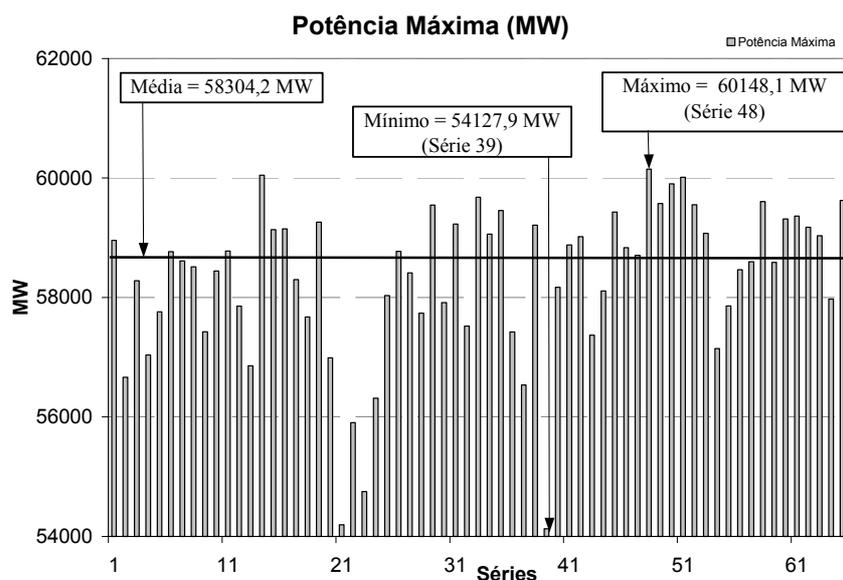


Figura 5.7 – Potências máximas (MW) nos casos de confiabilidade

Os gráficos apresentados nas Figura 5.7, Figura 5.8 e Figura 5.9, permitem a comparação e a visualização dos resultados da potência máxima em cada caso e as suas respectivas LOLP e EENS.

O tempo de processamento total dos 65 casos foi de aproximadamente 300hs, processando em um microcomputador AMD Athlon de 1,9 GHz com 256 MB de memória RAM. Este tempo de processamento torna a análise utilizando todos os resultados das séries históricas inviável.

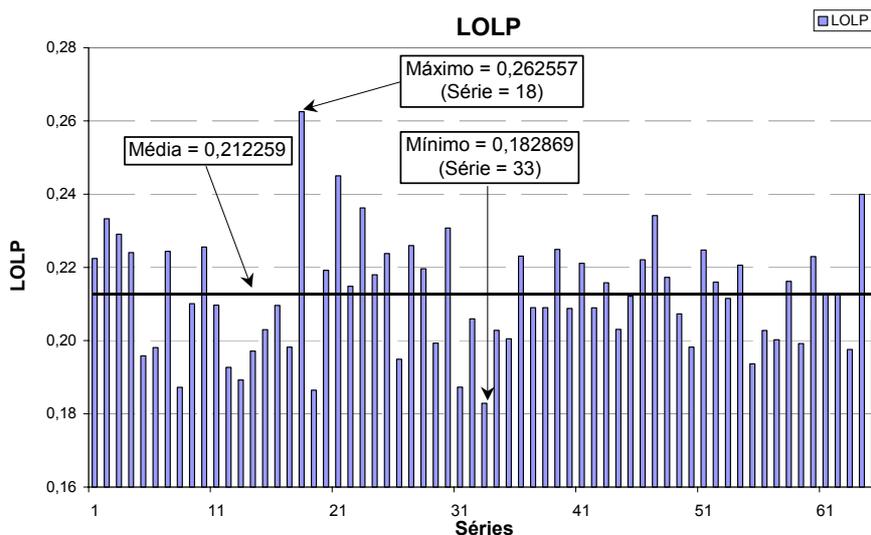


Figura 5.8 – LOLP dos casos de confiabilidade

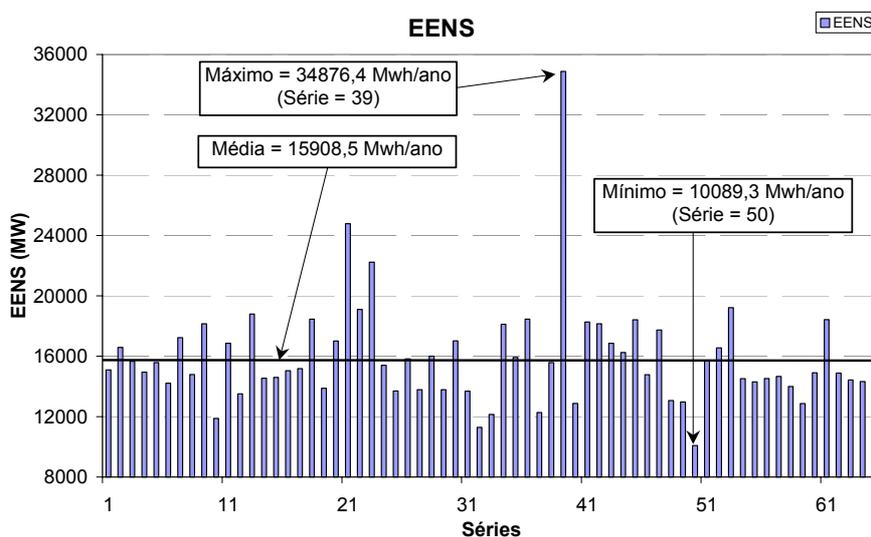


Figura 5.9 – EENS dos Casos de confiabilidade

A análise de somente um mês do período histórico, é mais apropriada para os estudos do planejamento da expansão. Do ponto de vista deste tipo de estudo, deseja-se determinar se a quantidade de energia produzida pelas usinas existentes e pelas que estão entrando em operação no sistema é suficiente para atender a carga em um determinado instante.

Do ponto de vista do planejamento da operação, coordenado pelo ONS (Operador Nacional do Sistema), seria bastante útil se verificar a evolução dos índices de confiabilidade do sistema elétrico interligado ao longo de um período de tempo, normalmente 3, 6 ou 12 meses a frente. Isto significaria multiplicar o esforço computacional citado anteriormente pelo número de meses do período de estudo desejado.

O objetivo desta dissertação é propor a utilização de um processo de agrupamento, a ser inserido no fluxo de informações entre os sistemas apresentado na Figura 5.5 que será transformado no fluxo apresentado na Figura 5.10.

No próximo capítulo, serão propostos 3 processos de agrupamentos nos quais, as séries de potência disponível serão tratadas de forma a obter-se um número menor de séries, representadas pelos centros dos agrupamentos. Este conjunto menor de séries deve refletir as características dos dados originais, mantendo a dependência temporal e espacial existente.

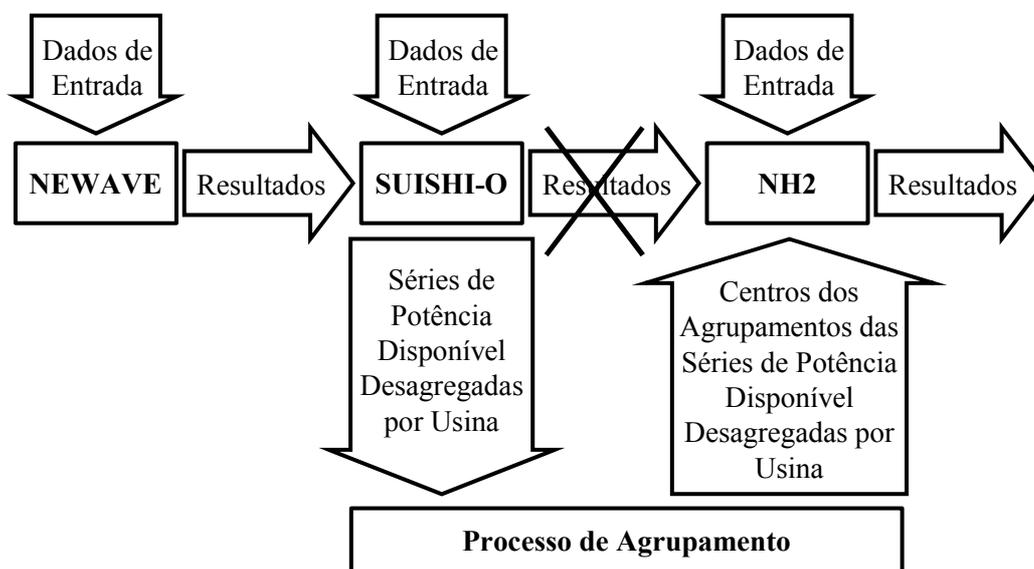


Figura 5.10 – Fluxo de informações entre os modelos energéticos e o NH2 com a utilização do processo de agrupamento.

6 PROCESSOS DE AGRUPAMENTO

Neste capítulo, serão propostos e avaliados 3 processos de agrupamento: o *Agrupamento de Matriz Única*, o *Agrupamento por Mapeamento de Resultados* e o *Agrupamento por Mapeamento de Resultados com Seleção do Coeficiente de Nebulização*, que serão aplicados sobre espaços de dados formados pelas séries de potência disponível das usinas hidrelétricas do sistema elétrico brasileiro de forma a reduzir o número de casos de confiabilidade composta necessários para avaliação da variação hidrológica. As séries resultantes da aplicação dos processos devem manter as características de dependência temporal e espacial existentes nos espaços de dados originais, assim como refletir a dispersão dos dados.

Os próximos itens deste capítulo descrevem respectivamente os métodos utilizados para a seleção dos dados que fizeram parte da composição do espaço de dados utilizados pelos processos, as premissas utilizadas para execução dos algoritmos de agrupamento, e por fim os três processos de agrupamento serão avaliados.

6.1 SELEÇÃO DE CARACTERÍSTICAS

A seleção de características tem por finalidade definir quais características de um espaço de dados são relevantes para um determinado estudo de análise de agrupamentos.

Neste trabalho a seleção de características foi composta por duas etapas: a aquisição de dados e o pré-processamento.

6.1.1 AQUISIÇÃO DE DADOS

Em estudos de análise de agrupamentos esta etapa é responsável por separar quais são as características que devem compor o espaço de dados que será analisado pelo algoritmo de agrupamento.

No presente trabalho, cada processo de agrupamento utilizou um espaço de dados próprio formado pela composição das séries de potência disponível das usinas hidrelétricas. Portanto, as características que devem ser avaliadas são representadas por estas séries.

6.1.1.1 CARACTERIZAÇÃO DAS USINAS

As usinas hidrelétricas que compõem um sistema elétrico são classificadas em dois tipos (FORTUNATO, 1990):

- *Usinas com reservatório*: Este tipo de usina possui reservatórios de acumulação, ou seja, com capacidade de armazenamento para regularizar as vazões de um mês, um ano ou mesmo de vários anos.
- *Usinas fio d'água*: São usinas que possuem reservatórios de compensação, ou seja, possuem volume suficiente somente para regularização de descargas semanais ou diárias.

O estudo energético relativo ao sistema elétrico Brasileiro que está sendo utilizado neste trabalho (PMO, Junho/2001) possuía a representação de 94 usinas.

Uma vez que as avaliações de confiabilidade foram realizadas para o subsistema Sudeste/Centro-Oeste, a seleção de características foi realizada para as 56 usinas pertencentes a este subsistema, já citadas na Tabela 5.4. Como no NH2 as usinas de Carmargos e Itutinga são consideradas em conjunto, neste trabalho o total de usinas consideradas é 55. Neste conjunto de usinas existem usinas dos dois tipos.

As usinas de Camargos, Furnas e Água Vermelha são exemplos de usinas com reservatório. Este tipo de usina normalmente apresenta variação nos valores das séries de potência disponível. Esta variação está diretamente associada ao nível de armazenamento de seus reservatórios, como mostrado no gráfico representativo da usina de Furnas, apresentado na Figura 6.1.

Por sua vez as usinas de Jaguará, Porto Colômbia e Itaipu, são exemplos de usinas fio d'água. Neste tipo de usina a potência disponível normalmente varia muito pouco, como mostra o gráfico representativo da usina de Jaguará, apresentado na Figura 6.2.

Usinas que possuem séries de potência disponível como as apresentadas na Figura 6.2, ou seja, com pouca dispersão nos dados, não são apropriadas para a aplicação de análise de agrupamentos, uma vez que a principal característica deste tipo de análise é buscar partições diferentes do espaço de dados, dependendo das características dos dados selecionados.

A utilização de um algoritmo de agrupamento em um espaço de dados cujos dados são completamente sobrepostos não permite a separação dos agrupamentos e conseqüentemente resultam em centros muito próximos, ou até, sobrepostos. Isto acontece por que os algoritmos de agrupamento tratam os dados das séries de potência disponível como pontos

multidimensionais e utilizam para aferir sua dissimilaridade medidas de distância, como por exemplo, a distância euclidiana utilizada no FCM.

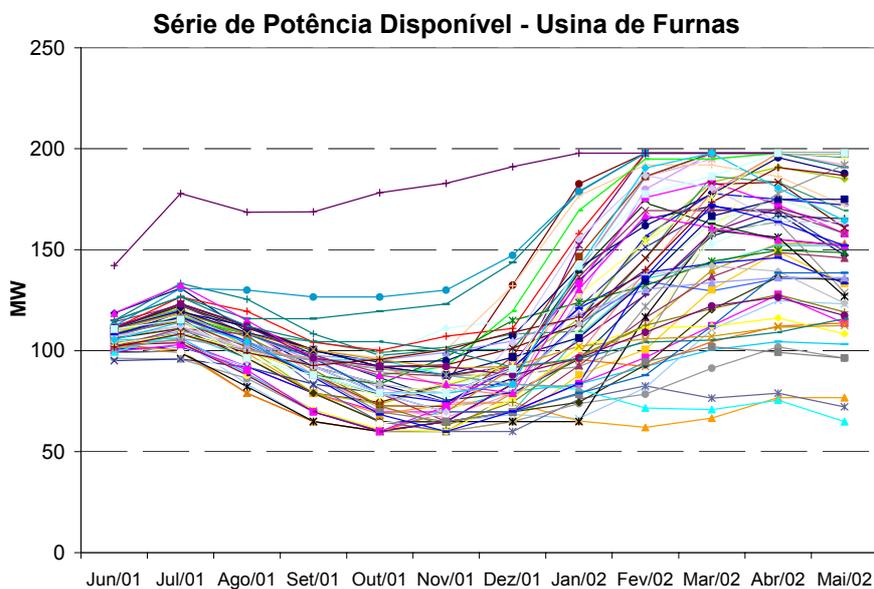


Figura 6.1 – Gráfico da potência disponível da usina de Furnas

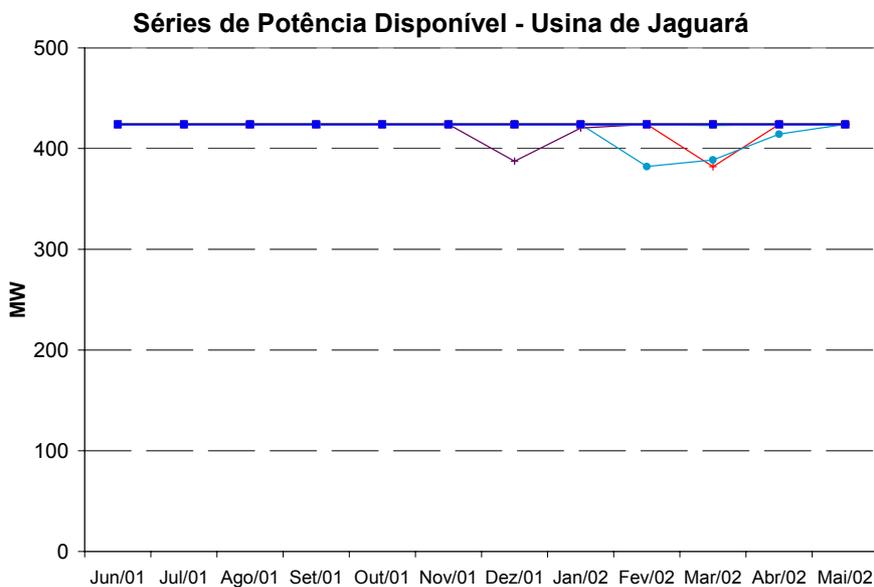


Figura 6.2 – Gráfico da potência disponível da usina de Jaguará

Apesar da pouca dispersão dos dados ser uma característica das usinas fio d'água, a utilização desta classificação não nos permite indicar se uma usina deve ou não participar do processo de agrupamento. Isto pode ser exemplificado verificando o gráfico da potência disponível da usina de Euclides da Cunha apresentado na Figura 6.3. Apesar desta usina ser

classificada como fio d'água, a distribuição das séries de potência no gráfico não permite avaliar se esta usina possui variação nos dados suficiente para formar agrupamentos com informação relevante ou não.

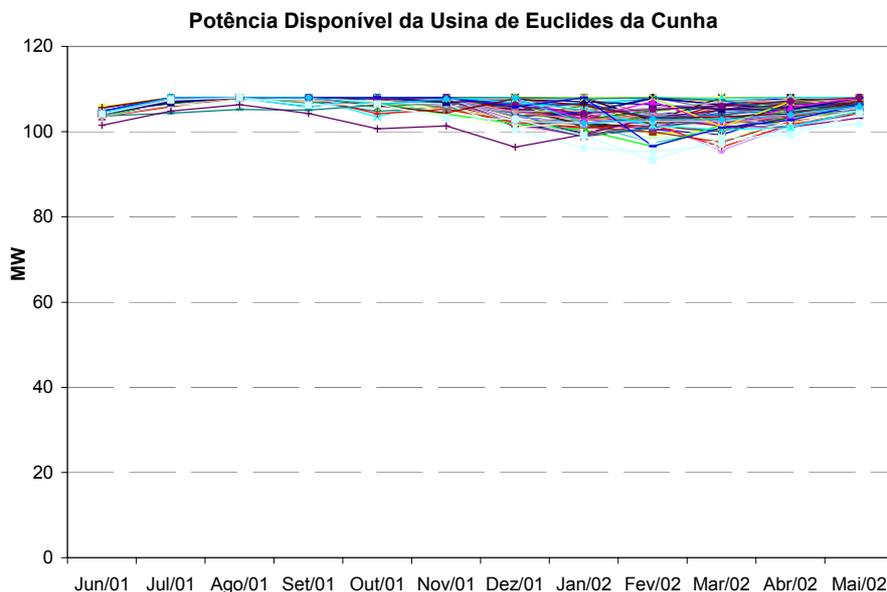


Figura 6.3 – Potência disponível da usina de Euclides da Cunha

Por outro lado não podemos afastar a hipótese de que em alguns casos as usinas com reservatório também podem se comportar como usinas fio d'água, ou seja, possuir variações pequenas nas séries de potência disponível em um determinado período de tempo.

Uma vez que a classificação como usinas com reservatório e usinas fio d'água não indica com precisão o tipo de variabilidade encontrada nas séries de potência, optou-se por fazer a verificação da variabilidade dos dados das séries de potência disponível das usinas utilizando técnicas estatísticas, como será mostrado no próximo item.

6.1.1.2 MÉTODOS ESTATÍSTICOS

O primeiro método estatístico de análise multivariada avaliado, foi a *Análise de Componentes Principais*.

Segundo BARROSO (2003), “*Análise de Componentes Principais (ACP)* é uma técnica estatística que transforma linearmente um conjunto de p variáveis em um conjunto com um número menor (k) de variáveis não correlacionadas, que explica uma parcela substancial das informações do conjunto original”.

Por tanto, as n amostras de $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$, sendo $\mathbf{x}_i = (x_1, x_2, \dots, x_p)$, são transformadas em $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$, sendo $\mathbf{x}_i = (y_1, y_2, \dots, y_k)$, onde $k < p$, denominadas componentes principais, de modo que y_1 é aquela que explica a maior parcela da variabilidade total dos dados, y_2 explica a segunda maior parcela e assim por diante.

Um a interpretação importante das componentes é que a primeira componente principal se reflete em um eixo no espaço. Quando cada amostra é projetada sobre este eixo, é formada uma nova variável, cuja variância é máxima dentre toda as possibilidades do primeiro eixo.

A segunda componente representa outro eixo no espaço, perpendicular ao primeiro. Assim como para a primeira componente, projetar as amostras sobre este eixo cria outra variável, e a variância desta é máxima em relação a todas as possibilidades deste segundo eixo.

O conjunto completo das componentes principais tem o mesmo tamanho do conjunto original. Mas, é comum que a variância das primeiras componentes somem entorno de 80% da variância total do espaço de dados original.

Os principais objetivos da análise de componentes principais são: reduzir a dimensão do espaço de dados; obter combinações interpretáveis do espaço de dados; e a descrição e entendimento da estrutura de correlação das amostras.

É uma prática comum no ambiente de pesquisa, utilizar gráficos obtidos a partir das duas ou três primeiras componentes, de forma a ampliar a compreensão, das características do espaço de dados original cuja dimensão p não permite a visualização gráfica.

No presente trabalho, a aplicação deste tipo de análise visou a reduzir a dimensão do espaço de dados, de forma a melhorar o resultado obtido pelo FCM e pela ICC.

Para realizar a ACP foi utilizada a função *princomp* implementada no *MatLab 6* (MATLAB, 2004), cuja assinatura é apresentada a seguir.

$$[pcs, newdata, variances, t2] = princomp (data)$$

Onde a esquerda do sinal de igualdade, entre colchetes, encontram-se os resultados que serão fornecidos pela função, enquanto que a direita da função é passado o parâmetro de entrada da função, entre parenteses. A seguir são descritos o parâmetro de entrada assim como cada um dos resultados.

- *data* é o parâmetro de entrada da função, ou seja, é um espaço de dados que será fornecido, em forma de matriz $n \times p$, para ser tratado pela ACP.
- *pcs* são as componentes principais propriamente ditas. Este resultado é representado por uma matriz quadrada $p \times p$, onde p representa o número de dimensões dos dados de

entrada. Elas representam a combinação linear dos dados originais geradores do novo espaço de dados (*newdata*). A análise deste resultado pode indicar quais são as dimensões do espaço de dados original que são mais representativas.

- *newdata* representa os dados transformados para o novo sistema de coordenadas definido pela ACP. Este resultado possui o mesmo tamanho da matriz original de dados, ou seja, é uma matriz $n \times p$.
- *variance* é um vetor de p coordenadas contendo a variância explicada por cada uma das colunas do novo conjunto de dados (*newdata*). A partir deste vetor podem ser calculados o percentual de explicação de cada uma das variáveis, como mostra a Equação 6.1, assim como o percentual de explicação acumulado por um conjunto de componentes, como mostra a Equação 6.2.

$$Var(\mathbf{y}_i)\% = 100 \frac{Var(\mathbf{y}_i)}{\sum_{i=1}^p Var(\mathbf{y}_i)} \quad \text{Equação 6.1}$$

Onde:

- $Var(\mathbf{y}_i)\%$ – Percentual de explicação da componente principal \mathbf{y}_i
- $Var(\mathbf{y}_i)$ – Variância da componente principal \mathbf{y}_i
- p – Número de dimensões do espaço de dados

$$Var(\mathbf{Y}_k)\% = \sum_{i=1}^k Var(\mathbf{y}_i) \quad \text{Equação 6.2}$$

Onde: $Var(\mathbf{Y}_k)\%$

- $Var(\mathbf{Y}_k)\%$ – Percentual de explicação acumulado por um conjunto de k componentes principais.
- k – Número de componentes principais selecionado.

- t^2 (*Hotelling's T²*) este resultado representa uma medida estatística da distância multivariada de cada uma das observações em relação ao centro do espaço de dados. Ele é uma forma analítica que permite descobrir os pontos mais extremos dos dados.

Na metodologia utilizada, inicialmente aplicou-se a ACP ao espaço de dados de cada uma das usinas, ou seja, a espaços de dados constituídos por 12 dimensões.

O número de componentes principais foi escolhido quando o percentual de explicação obtido pelas k componentes escolhidas fosse maior ou igual a 75%. Este percentual normalmente com duas ou três dimensões.

Após a determinação do espaço de dados reduzido foram utilizados o FCM e a ICC buscando obter o melhor número de agrupamentos. Após a etapa de determinação do número de agrupamentos, o FCM foi novamente utilizado desta vez com o espaço de dados original de forma a criar os agrupamentos e os centros que seriam utilizados posteriormente para realização da análise de confiabilidade.

Todavia, foi observado nos testes efetuados que os agrupamentos, centros e resultados da ICC obtidos, utilizando ou não a ACP, foram muito semelhantes.

Entretanto, a principal motivação para deixar de utilizar foi o tempo de processamento maior quando a ACP estava sendo utilizada, quando comparada ao tempo de processamento sem a utilização da ACP. O motivo para que este tempo fosse maior é que nesta metodologia é necessário executar o FCM em duas etapas.

O segundo método estatístico utilizado foi o cálculo do *Coefficiente de Variação (CV)* do espaço de dados $X = \{x_1, \dots, x_n\}$, dado pela Equação 6.3 (MEYER, 2000).

$$CV = \frac{\bar{x}}{S} \quad \text{Equação 6.3}$$

Onde:

- \bar{x} – Média da amostra
- S – Desvio padrão da amostra

Tanto a média quanto o desvio padrão da amostra são vetores de p dimensões, dados respectivamente pelas Equação 6.4 e Equação 6.5.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad \text{Equação 6.4}$$

Onde:

- \bar{x} – Média da amostra
- x_i – Amostra i
- n – Número de amostras

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad \text{Equação 6.5}$$

Onde:

S – Desvio padrão da amostra

Portanto, como resultado da Equação 6.3, tem-se um vetor $\mathbf{CV} = \{cv_1, cv_2, \dots, cv_p\}$.

Após o cálculo do \mathbf{CV} , procurou-se determinar quais seriam as usinas que deveriam participar do processo de agrupamento a partir do valores do coeficiente de variação calculados. Foi determinado, um ponto de corte em que este valor deveria ser maior que 5, como mostra a Equação 6.6.

$$\max_{1 \leq j \leq p} (cv_j) > 5 \quad \text{Equação 6.6}$$

Onde:

p – Número de dimensões da amostra

Este limite para o máximo do coeficiente de variação foi escolhido por se tratar do valor padrão utilizado como critério de convergência da simulação de Monte Carlo na análise de confiabilidade composta no sistema NH2. A convergência neste caso é calculada em relação ao coeficiente de variação dos índices de confiabilidade LOLP e EPNS.

Os resultados obtidos para este ponto de corte não foram satisfatórios pois usinas como Euclides da Cunha (máximo CV = 3,21), Água Vermelha (máximo CV = 4,82), São Simão (máximo CV = 4,72) e Canoas II (máximo CV = 4,65), não eram selecionadas para participar do processo de agrupamento.

A análise dos gráficos destas usinas, apresentados no Apêndice B, permite verificar que suas séries de potência disponível apresentam dispersão nos dados suficiente para formação de agrupamentos. O que foi verificado pela aplicação do FCM sobre os dados da usina.

Por outro lado ao reduzir-se a valor de corte para 1, algumas usinas, como Pereira Passos (máximo CV = 2,21), A. S. Lima (máximo CV = 2,08), Jaguará (máximo CV = 1,60) e Ilha dos Pombos (máximo CV = 1,55), que apresentam alguns pontos de valores mais baixos eram selecionadas para participar do processo de agrupamento, o que também não é desejável.

Como pode ser observado no gráfico de Jaraguá apresentado na Figura 6.2, o máximo do coeficiente de variação pode ser muito afetado por estes pontos de valores mais baixos. Isto ocorre por que este índice é calculado em relação a somente uma dimensão do vetor. Este foi o principal motivo para decidir-se por abandonar este indicador.

Finalmente, foi decidido que a participação no processo de agrupamento seria determinada pela média dos coeficientes de variação, como mostrado na Equação 6.7.

$$\frac{\sum_{j=1}^p cv_j}{p} > 1 \quad \text{Equação 6.7}$$

Por tratar-se de um valor médio de todas as p dimensões do coeficiente de variação, a influência dos dados destoantes sobre uma das dimensão cv_j é minimizada.

O valor do ponto de corte maior que 1 foi determinado pela observação dos resultados obtidos para a média do coeficiente e pelas características dos gráficos das potências disponíveis das usinas.

Tabela 6.1 – Usinas que participam do Processo de Agrupamento

| Núm. ¹¹ | Nome | Núm. | Nome | Núm. | Nome |
|--------------------|-----------------------|------|-----------------|------|--------------|
| 1 | Camargos+Itutinga | 31 | Itumbiara | 61 | Capivara |
| 6 | Furnas | 33 | Sã Simão | 120 | Jaguari |
| 7 | Mascarenhas de Moraes | 37 | Barra Bonita | 121 | Paraibuna |
| 14 | Caconde | 39 | Ibitinga | 122 | Santa Branca |
| 15 | Euclides da Cunha | 40 | Promissão | 123 | Funil |
| 17 | Marimbondo | 45 | Jupiaá | 144 | Mascarenhas |
| 18 | Água Vermelha | 46 | Porto Primavera | 156 | Três Marias |
| 24 | Emborcação | 47 | A. A. Laydner | 217 | Rosal |
| 25 | Nova Ponte | 49 | Xavantes | 251 | Serra Mesa |
| 30 | Corumbá I | 51 | Canoas II | 278 | Manso |

¹¹ Número da usina no Suishi-O

Como resultado da análise do valor médio do coeficiente de variação, foram separadas 25 usinas para não participar do processo de agrupamento, cujos dados e gráficos da potência disponível normalizadas encontram-se representados no apêndice A, e 30 usinas para participar do processo de agrupamento, estas representadas no apêndice B.

A Tabela 6.1, apresenta a lista das usinas que foram selecionadas para participar dos processos de agrupamento. Nesta tabela as usinas foram apresentadas na ordem em que foram incluídas nos espaços de dados utilizados nos processos de agrupamento, ou seja, em ordem crescente do número que as identifica no SUSHI-O¹¹.

6.1.2 PRÉ-PROCESSAMENTO

As séries de potência disponível de cada uma das usinas do sistema possuem dados que são apresentados em valores reais e variam de acordo com a capacidade máxima de geração dos equipamentos de cada uma das usinas, ou seja, apesar de serem expressas na mesma unidade (MW) seus valores são muito dispares. Por exemplo, na usina de Furnas o máximo da potência disponível é 1312 MW enquanto que na usina de Mascarenhas de Moraes, que se localiza na mesma cascata a jusante daquela, o máximo da potência disponível é 478 MW, ou seja, na primeira o máximo da potência disponível é, aproximadamente, 3 vezes maior que na segunda.

De forma a permitir que os dados de todas as usinas envolvidas no problema possam ser analisados em conjunto deve ser efetuada a normalização dos dados de suas séries de potência disponível. Este processo visa tornar os valores adimensionais com variação limitada e semelhante.

Existem várias formas de normalização dos dados, dentre elas as mais simples são a divisão dos dados por um valor constante obtido a partir do próprio espaço de dados. Dentre os valores usualmente utilizados estão o máximo, a média e o desvio padrão.

Cada um destes valores pode ser obtido a partir de cada uma das séries, ou seja, cada série é normalizada por um valor distinto, ou por um valor obtido do espaço de dados como um todo.

Neste trabalho os valores foram normalizados efetuando a divisão pelo máximo, desta forma todos os espaços de dados passaram a estar contidos em um intervalo adimensional [0, 1].

Para verificar o efeito da normalização do espaço de dados foi selecionada a usina de Funil, para a qual foram desenhados 3 gráficos. A Figura 6.4 apresenta o gráfico do espaço de dados original, por sua vez a Figura 6.5 apresenta o gráfico do espaço de dados normalizado

pelo máximo de cada série, e finalmente a Figura 6.6 apresenta o gráfico do espaço de dados normalizado pelo máximo global.

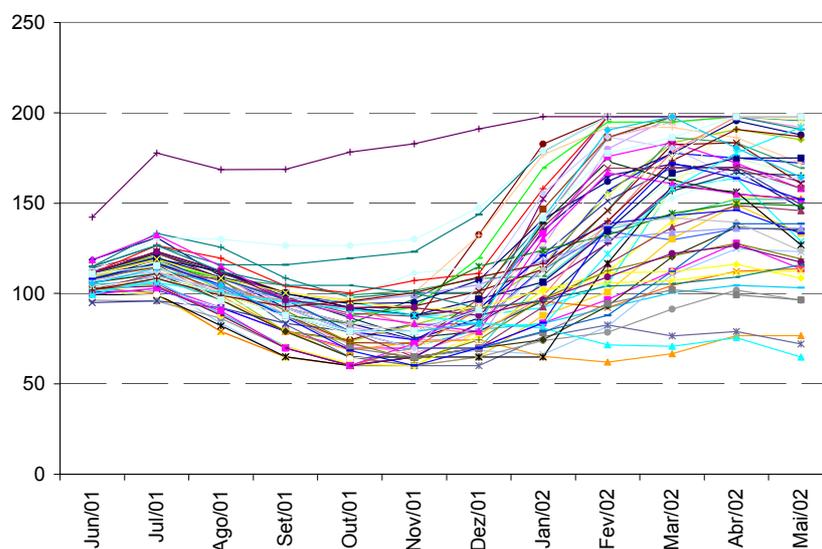


Figura 6.4 – Gráfico da potência disponível não normalizada da usina de Funil

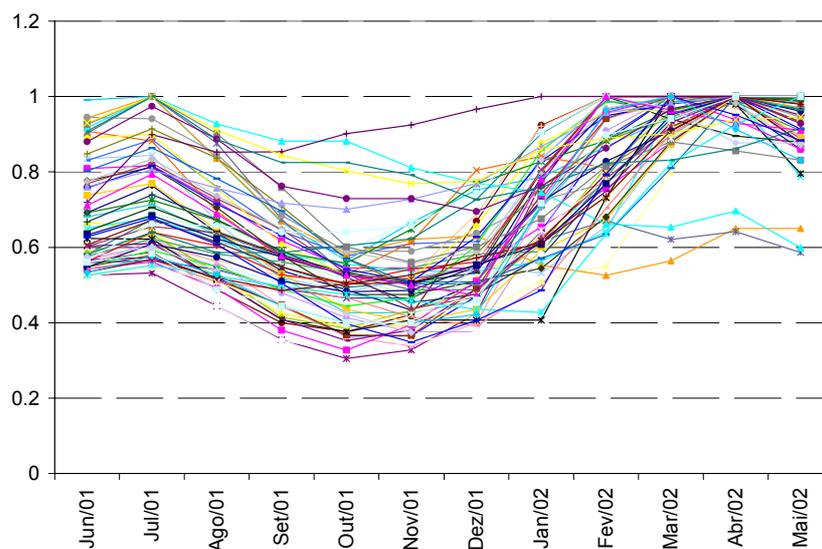


Figura 6.5 – Gráfico da potência disponível normalizada pelo máximo de cada série da usina de Funil

Uma simples análise visual permite verificar que a normalização do espaço de dados pelo máximo de cada série (Figura 6.5) altera a forma do gráfico do espaço de dados original (Figura 6.4). Isto não é desejável pois influencia nos resultados dos algoritmos de agrupamento.

Por sua vez, comparando o gráfico do espaço de dados original (Figura 6.4) com o gráfico do espaço de dados normalizado pelo máximo global (Figura 6.6), podemos verificar que a forma é mantida.

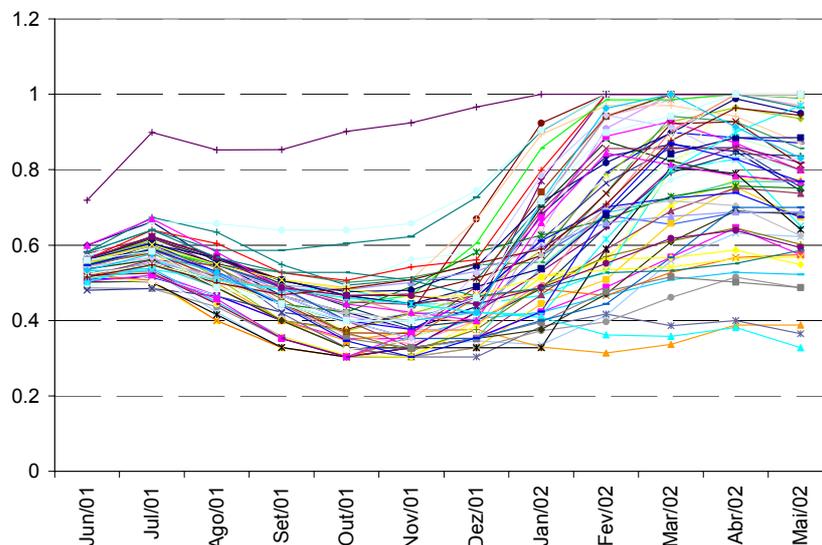


Figura 6.6– Gráfico da potência disponível normalizada pelo máximo global do espaço de dados da usina de Funil

Neste trabalho, a forma de normalização selecionada foi a divisão dos dados das séries de potência disponível de cada usina que participa do processo de agrupamento pelo valor máximo obtido a partir do seu espaço de dados.

6.2 DEFINIÇÕES PRELIMINARES

Como já foi citado no Capítulo 3, existe uma indicação na literatura no sentido de utilizar como número máximo de agrupamentos o valor $c_{m\acute{a}x} = \sqrt{n}$, onde n é o número de amostras existente no espaço de dados.

No presente trabalho, o espaço de dados é composto por 65 amostras, representadas pelas séries de potência disponível, para cada uma das 30 usinas selecionadas para participar dos processos de agrupamento. Portanto, observando a indicação anterior deveria ser utilizado o número máximo de agrupamentos $c_{m\acute{a}x}$ igual a 8. Como esta indicação trata-se somente de um direcionamento, decidiu-se por utilizar o número máximo de agrupamentos igual a 10, por ser um número mais intuitivo.

Para execução do FCM é necessário também indicar a priori um valor para o coeficiente de nebulização (m). Neste trabalho, visando enriquecer os resultados, decidiu-se por avaliar a utilização de 4 valores distintos do coeficiente de nebulização, $m = 1,25$, $m = 1,5$, $m = 1,75$ e $m = 2,0$. O primeiro e o último por serem valores indicados na literatura como valores comuns, e os demais por dividirem o intervalo existente entre o primeiro e o segundo em partes iguais.

Uma consideração adicional deve ser feita em relação a matriz de pertinências inicial $U^{(0)}$. No FCM esta matriz é iniciada aleatoriamente, esta característica torna necessário que o algoritmo seja executado várias vezes de modo a selecionar dentre as diferentes execuções a melhor função objetivo, ou seja, dentro de um número máximo de rodadas $r_{máx}$ estipulado, devem ser selecionados os resultados da execução do FCM que obtiver o menor valor da função objeto.

Foram realizados vários testes buscando determinar o melhor número de execuções do FCM de forma que o resultado de sua função objetivo fosse minimizada, avaliando-se também o compromisso entre o número de vezes que a função objetivo melhorou e o acréscimo no tempo de processamento pela inclusão de mais uma execução do algoritmo. Como resultado dos testes determinou-se que o número de execuções do algoritmo $r_{máx}$ seria igual a 20.

Neste trabalho, foi utilizada a implementação do algoritmo do FCM encontrada no MatLab release 12 e a implementação do PAM, desenvolvida em Fortran 77, que foi disponibilizada pelos autores em conjunto com o livro que apresenta o algoritmo e a medida de validação Silhueta Média Global \bar{S} . (KAUFFMAN, 1990)

6.3 AGRUPAMENTO DE MATRIZ ÚNICA

O primeiro processo de agrupamento definido neste trabalho de dissertação foi o *Agrupamento de Matriz Única*. Neste processo, todos os dados de todas as usinas foram dispostos seqüencialmente, lado a lado, e estes dados passaram pelo processo de agrupamento em uma única execução do FCM. A motivação para a denominação deste processo foi a utilização de uma única matriz com os dados de todas as usinas.

A matriz de dados é a forma matemática assumida pelo espaço de dados para apresentação para o FCM. O critério para formação desta matriz foi: as primeiras 12 colunas representam os dados das séries de potência referentes a primeira usina, as colunas 13 a 24 representam os dados da segunda usina e assim sucessivamente até que todas as usinas foram representadas. O fato de todos os dados, de todas as usinas que participam do processo de

agrupamento, serem processados em uma única execução do algoritmo de agrupamento garante que as dependências temporal e espacial existentes nos dados sejam mantidas.

Da forma como foi definido, este processo deve ser executado separadamente para cada um dos valores do coeficiente de nebulização que deseja-se avaliar, por tanto, neste trabalho sua execução foi repetida para cada um dos 4 valores do coeficiente de nebulização indicados no item 6.2.

6.3.1 ESPAÇO DE DADOS

O espaço de dados de cada usina é representado pelos dados de sua série de potência disponível obtidas a partir dos modelos energéticos NEWAVE e SUIISHI-O. Matematicamente, as séries de potência disponível de cada uma das usinas, pode ser representada por uma matriz onde cada linha representa um registro da série e cada coluna um mês, portanto, neste trabalho cada usina foi representada por uma matriz de 65 x 12.

Por sua vez, o espaço de dados utilizado pelo *Agrupamento de Matriz Única*, também pode ser representado matematicamente por uma matriz, sendo que, para a composição desta matriz os dados de cada uma das usinas foram colocados uns ao lado dos outros, como mostra a Figura 6.7.

| S e r i e s | Usina 1 | | | | Usina 2 | | | | | Usina 30 | | | |
|----------------------------|--------------------|--------------------|-----|---------------------|--------------------|--------------------|-----|---------------------|--|---------------------|---------------------|-----|----------------------|
| | Jun/01 | Jul/01 | ... | Mai/02 | Jun/01 | Jul/01 | ... | Mai/02 | | Jun/01 | Jul/01 | ... | Mai/02 |
| 1 | u1 _{1,1} | u1 _{1,2} | ... | u1 _{1,12} | u2 _{1,1} | u2 _{1,2} | ... | u2 _{1,12} | | u30 _{1,1} | u30 _{1,2} | ... | u30 _{1,12} |
| 2 | u1 _{2,1} | u1 _{2,2} | ... | u1 _{2,12} | u2 _{2,1} | u2 _{2,2} | ... | u2 _{2,12} | | u30 _{2,1} | u30 _{2,2} | ... | u30 _{2,12} |
| | | | | | | | | | | | | | |
| 65 | u1 _{65,1} | u1 _{65,2} | ... | u1 _{65,12} | u2 _{65,1} | u2 _{65,2} | ... | u2 _{65,12} | | u30 _{65,1} | u30 _{65,2} | ... | u30 _{65,12} |

Figura 6.7 – Representação do espaço de dados - Agrupamento por Matriz Única

Na Figura 6.7 cada célula foi representada por um mnemônico, $uN_{s,m}$, para o valor real localizado na célula, onde:

- uN – Indica que o dado existente na célula pertence a uma série de potência disponível da usina N
- s – Indica a qual série pertence o valor
- m – Indica a qual mês pertence o valor

Um exemplo de interpretação para estes mnemônicos pode ser a célula com a representação $u17_{32,8}$, indica o valor do oitavo mês da série número 32 da décima sétima usina representada no conjunto de dados.

Como já foi determinado no item 6.1.1, 30 usinas participaram do processo de agrupamento, cada uma destas usinas representadas por uma matriz 65×12 , portanto a matriz utilizada tem tamanho 65×360 .

6.3.2 ALGORITMO

O algoritmo para a realização do processo de agrupamento é descrito a seguir:

/ Inicialização de variáveis, como indicado no item 6.2 **/**

Passo 1. Fixar:

O número de usinas que participam do processo, $n_{usinas} = 30$.

O número máximo de agrupamentos, $c_{máx} = 10$.

O número de rodadas do FCM, $r_{máx} = 20$.

A constante de nebulização, m .

O número de séries de potência disponível de cada usina, $n_{séries} = 65$.

O número de dimensões do espaço de dados de cada usina, $p = 12$.

/ Laço de formação do espaço de dados **/**

/ Resultados: matrizes X_{usina} e X'_{usina} , de tamanho $n_{séries} \times (p \cdot n_{usinas})$ **/**

Passo 2. Para $usina = 1, 2, \dots, n_{usinas}$

Passo 2.1. Utilizar o espaço de dados X_{usina} para formar o espaço de dados X , como apresentado no item 6.3.1.

Passo 2.2. Normalizar o espaço de dados X_{usina} , como apresentado no item 6.1.2, obtendo X'_{usina}

Passo 2.3. Utilizar o espaço de dados X'_{usina} para formar o espaço de dados X' , como apresentado no item 6.3.1.

Passo 2.4. Retornar ao Passo 2.

/ Laço para execução do FCM e cálculo da ICC **/**

Passo 3. Determinar o centro \mathbf{m} do espaço de dados X' , como mostrado na Equação 6.8.

$$\mathbf{m} = \frac{1}{n_{séries}} \sum_{i=1}^{n_{séries}} \mathbf{x}_i \quad \text{Equação 6.8}$$

Passo 4. Para $c = 2, 3, \dots, c_{\text{máx}}$

Passo 4.1. Inicia o valor da função objetivo $J^{(0)}$ com um valor muito alto

Passo 4.2. Para $r = 1, 2, \dots, r_{\text{máx}}$

Passo 4.2.1. Executar o FCM, como apresentado no item 3.2.1.3

Passo 4.2.2. Se o valor da função objetivo $J^{(r)}$ for menor que o valor de $J^{(r-1)}$

Armazenar a matriz de pertinências $\mathbf{U}^{(c)} = \mathbf{U}^{(r)}$,

onde $\mathbf{U}^{(c)}$ e $\mathbf{U}^{(r)}$ são matrizes de tamanho $c \times n_{\text{séries}}$

Armazenar a matriz de centros normalizados $\mathbf{M}^{(c)} = \mathbf{M}^{(r)}$,

onde $\mathbf{M}^{(c)} = \mathbf{M}^{(r)}$ são matrizes de tamanho $c \times (p \cdot n_{\text{usinas}})$.

Passo 4.2.3. Retornar ao Passo 4.2.

Passo 4.3. Calcular a $ICC(c)$, como apresentado no item 3.4.4.

Passo 4.4. Retornar ao Passo 4.

Passo 5. Determinar o número de agrupamentos c_{sel} onde ocorre o valor máximo da $ICC(c)$.

Passo 6. Utilizar a matriz de pertinências $\mathbf{U}^{(c_{\text{sel}})}$ e o espaço de dados \mathbf{X} , para obter a matriz de centros não normalizados dos agrupamentos $\mathbf{M}^{(c_{\text{sel}})}$, de tamanho $c \times (p \cdot n_{\text{usinas}})$.

/ Laço para formar as matrizes de centros por usina **/**

Passo 7. Para $usina = 1, 2, \dots, n_{\text{usinas}}$

Passo 7.1. Subdividir a matriz $\mathbf{M}^{(c_{\text{sel}})}$, separando p colunas relativas a $usina$, para formar uma matriz de centros \mathbf{M}_{usina} para cada usina.

Passo 7.2. Voltar ao Passo 7.

6.3.3 RESULTADOS E CONCLUSÕES SOBRE O PROCESSO DE AGRUPAMENTO

O principal resultado deste processo de agrupamento são as n_{usinas} matrizes \mathbf{M}_{usina} de centros dos agrupamentos, todas de tamanho $c_{\text{sel}} \times p$, onde c_{sel} representa o número de agrupamentos determinados pelo processo de agrupamento e p representa o número de meses de cada uma das séries representadas por estes centros.

Os resultados referentes ao *Agrupamento de Matriz Única*, que a partir deste ponto será referenciado como *AMU*, encontram-se listados na Tabela 6.2, onde encontram-se relacionados o coeficiente de nebulização utilizado no FCM, o número de agrupamentos determinado pela *ICC* assim como o seu valor, e o tempo total de processamento, que inclui a execução do processo e a criação dos casos de confiabilidade composta baseados nos centros determinados.

O tempo total para execução deste processo para os quatro valores do coeficiente de nebulização foi de 22min e 55seg. Este tempo total de processamento representa 8,5 % do tempo médio de processamento de um caso do NH2, que é de 4hs e 30min. Portanto, o acréscimo no tempo de processamento da avaliação de confiabilidade, seria muito pouco afetada pela inclusão do processo de agrupamento no fluxo de informação entre os sistemas Newave, Suishi-O e NH2.

Tabela 6.2 – Resultados do *AMU*

| Coeficiente de Nebulização (<i>m</i>) | Número de Agrupamentos (<i>c</i>) | <i>ICC</i> | Tempo de Processamento | Tempo de Processamento Relativo (%)¹² |
|--|--|-------------------|-------------------------------|---|
| 1,25 | 10 | 5,0135 | 5m 35s | 2,1 |
| 1,50 | 3 | 2,8394 | 5m 48s | 2,2 |
| 1,75 | 2 | 2,1076 | 5m 59s | 2,2 |
| 2,00 | 2 | 1,5866 | 5m 33s | 2,0 |
| Tempo de Processamento Total | | | 22m 55s | 8,5 |

Uma consideração importante deve ser feita. Por avaliar separadamente cada um dos valores do coeficiente de nebulização o *AMU*, poderia ter definido um número de centros entre 8, relativos ao número mínimo de agrupamentos que poderia ser definido por cada uma das 4 execuções do processo, e 40 centros, relativos ao número máximo de agrupamentos. Este intervalo seria diretamente refletido no número total de casos de confiabilidade composta que poderiam ser definidos pelo processo.

Analisando o *AMU* em relação ao número mínimo de casos de confiabilidade composta que este pode definir, a redução neste número de casos de 65 (analisados no item 5.2) para 8, resultaria em uma redução de 87% no tempo de processamento, ou seja, um percentual interessante. Por outro lado, realizando a mesma comparação em relação ao número máximo de casos de confiabilidade, a redução seria de 65 para 40 casos, resultado em uma redução de 38,5%, ou seja, um percentual pequeno de redução, pois a execução de 40 casos de confiabilidade ainda incorre em um tempo de processamento muito alto (aproximadamente 180hs).

Analisando as considerações anteriores, conclui-se que os resultados relativos a redução do tempo de processamento da análise de confiabilidade deste processo é muito dependente da

¹² Tempo de processamento do processo de agrupamento dividido pelo tempo médio de processamento de um caso de confiabilidade composta (utilizando o programa NH2)

forma com que o processo for montado, ou seja, da quantidade de coeficientes de nebulização selecionados para avaliação e do número máximo de agrupamentos a ser considerado.

6.3.4 RESULTADOS DA ANÁLISE DE CONFIABILIDADE

Os casos do NH2 foram formados utilizando o algoritmo apresentado no item 5.2, a partir do valor obtido da décima coluna da matriz de centros $M^{(cseI)}$, ou seja, o valor referente ao mês de março de 2002.

As tabelas a seguir (Tabela 6.3, Tabela 6.4, Tabela 6.5, Tabela 6.6) apresentam os resultados obtidos após a execução dos casos de confiabilidade composta. Cada uma delas reflete os principais índices de confiabilidade determinados pelo NH2 para os casos formados após a execução do processo de agrupamento. Cada tabela esta associada a um dos valores do coeficiente de nebulização avaliados.

Tabela 6.3 – Resultados do NH2 – Casos formados pelo AMU – $m = 1,25$

| 13 A g r | 14 n. c. a. | Pot. Máx. (MW) | LOLP | | EPNS (MW) | | LOLE (h/ano) | EENS (MWh/ ano) | Monte Carlo ¹⁵ | | |
|-------------------|----------------------|----------------------|--------|-----------|-----------|-----------|-----------------|-----------------------|---------------------------|---------------------|--------------------|
| | | | Valor | CV (%) | Valor | CV (%) | | | Sim. ¹⁶ | Sist. ¹⁷ | Cort ¹⁸ |
| 1 | 7 | 58872,8 | 0,2040 | 5,00 | 1,63 | 11,73 | 1786,7 | 14305,1 | 1667 | 1231 | 340 |
| 2 | 11 | 57846,8 | 0,2187 | 5,00 | 1,48 | 11,28 | 1915,7 | 12985,2 | 1541 | 1169 | 337 |
| 3 | 10 | 57115,9 | 0,2261 | 3,79 | 1,56 | 14,99 | 1980,3 | 13678,1 | 2548 | 1959 | 576 |
| 4 | 3 | 54171,5 | 0,2320 | 2,99 | 4,61 | 15,00 | 2032,8 | 40412,8 | 4206 | 3579 | 976 |
| 5 | 4 | 58567,9 | 0,2198 | 4,11 | 1,97 | 14,99 | 1925,6 | 17213,5 | 3311 | 1700 | 508 |
| 6 | 1 | 59473,2 | 0,2147 | 4,09 | 2,10 | 14,99 | 1880,9 | 18375,5 | 2324 | 1778 | 499 |
| 7 | 6 | 56013,8 | 0,2319 | 3,78 | 1,78 | 14,99 | 2031,3 | 15575,0 | 2497 | 1981 | 579 |
| 8 | 5 | 59039,4 | 0,2153 | 4,99 | 1,56 | 10,97 | 1885,9 | 13702,5 | 1570 | 1182 | 338 |
| 9 | 11 | 58213,9 | 0,2101 | 5,00 | 1,48 | 11,09 | 1840,8 | 12960,4 | 1681 | 1212 | 340 |
| 10 | 7 | 58737,3 | 0,1923 | 4,10 | 1,64 | 14,98 | 1684,6 | 14371,3 | 2740 | 1950 | 520 |

¹³ Número de agrupamentos

¹⁴ Número de curvas no agrupamento

¹⁵ Resultados provenientes da simulação de Monte Carlo para determinação dos índices de confiabilidade.

¹⁶ Número de eventos simulados e solucionados

¹⁷ Número de eventos críticos com problemas no sistema

¹⁸ Número de eventos críticos com corte de carga

Tabela 6.4 – Resultados do NH2 – Casos formados pelo AMU – $m = 1,50$

| A g r | n. c. a. | Pot. Máx. (MW) | LOLP | | EPNS (MW) | | LOLE (h/ano) | EENS (MWh/ ano) | Monte Carlo | | |
|-------------|----------------|----------------------|--------|-----------|-----------|-----------|-----------------|-----------------------|-------------|-------|------|
| | | | Valor | CV (%) | Valor | CV (%) | | | Sim. | Sist. | Cort |
| 1 | 13 | 56167,7 | 0,2182 | 3,92 | 1,64 | 15,00 | 1911,3 | 14332,7 | 2585 | 1942 | 564 |
| 2 | 16 | 58774,5 | 0,2115 | 4,99 | 1,74 | 12,53 | 1853,1 | 15210,4 | 1612 | 1170 | 341 |
| 3 | 36 | 58093,8 | 0,2070 | 5,00 | 1,52 | 11,18 | 1813,0 | 13294,0 | 1667 | 1268 | 345 |

Tabela 6.5 – Resultados do NH2 – Casos formados pelo AMU – $m = 1,75$

| A g r | n. c. a. | Pot. Máx. (MW) | LOLP | | EPNS (MW) | | LOLE (h/ano) | EENS (MWh/ ano) | Monte Carlo | | |
|-------------|----------------|----------------------|--------|-----------|-----------|-----------|-----------------|-----------------------|-------------|-------|------|
| | | | Valor | CV (%) | Valor | CV (%) | | | Sim. | Sist. | Cort |
| 1 | 23 | 57098,3 | 0,2172 | 4,06 | 1,89 | 14,98 | 1902,6 | 16564,7 | 2408 | 1741 | 523 |
| 2 | 42 | 58424,1 | 0,2042 | 5,00 | 1,34 | 9,94 | 1788,6 | 11733,0 | 1675 | 1232 | 342 |

Tabela 6.6 – Resultados do NH2 – Casos formados pelo AMU – $m = 2,00$

| A g r | n. c. a. | Pot. Máx. (MW) | LOLP | | EPNS (MW) | | LOLE (h/ano) | EENS (MWh/ ano) | Monte Carlo | | |
|-------------|----------------|----------------------|--------|-----------|-----------|-----------|-----------------|-----------------------|-------------|-------|------|
| | | | Valor | CV (%) | Valor | CV (%) | | | Sim. | Sist. | Cort |
| 1 | 42 | 58314,0 | 0,1973 | 4,12 | 1,82 | 14,97 | 1728,6 | 15922,4 | 2615 | 1856 | 516 |
| 2 | 23 | 57217,8 | 0,2162 | 4,07 | 1,88 | 14,95 | 1893,7 | 16495,2 | 2461 | 1873 | 532 |

O tempo total para execução dos 17 casos, formados a partir dos resultados do processo de agrupamento, foi de aproximadamente 80hs. Comparativamente ao tempo de processamento dos *casos do históricos*, determinado no item 5.2, houve uma redução de aproximadamente 220hs.

A Tabela 6.7 apresenta as estatísticas em relação aos valores da LOLP apresentados anteriormente. Além do máximo, do mínimo, da média aritmética, do desvio padrão e do coeficiente de variação, também foi avaliado o erro médio absoluto da média ponderada em relação ao valor da média aritmética obtida para os *casos do histórico*. Este erro foi calculado segundo a Equação 6.9.

Tabela 6.7 – Estatísticas dos resultados dos casos do NH2 (AMU) para LOLP

| Coefficiente de Nebulização (m) | 1,25 | 1,50 | 1,75 | 2,00 |
|-------------------------------------|--------|--------|--------|--------|
| Máximo | 0,2320 | 0,2182 | 0,2172 | 0,2162 |
| Mínimo | 0,1923 | 0,2070 | 0,2042 | 0,1973 |
| Média Ponderada | 0,2155 | 0,2103 | 0,2088 | 0,2040 |
| Erro Médio Absoluto | 0,0033 | 0,0019 | 0,0035 | 0,0083 |
| Desvio Padrão | 0,0116 | 0,0044 | 0,0062 | 0,0090 |
| Coefficiente de Variação (%) | 5,3824 | 2,0716 | 2,9804 | 4,4183 |

$$\text{erro médio absoluto} = \frac{|\bar{x}_{ch} - \bar{x}_m|}{\bar{x}_m} \quad \text{Equação 6.9}$$

Onde:

- \bar{x}_{ch} – Média aritmética dos *casos do histórico*
- \bar{x}_m – Média ponderada para um valor de m

Com o auxílio do erro médio absoluto determinado para cada um dos conjuntos de resultados foi possível verificar que em todos os casos a média da LOLP foi preservada pelos casos determinados pelo AMU. Esta afirmação baseia-se no fato de que em todos os casos o erro médio absoluto determinado é menor que 1%. Sendo que, a média obtida para o coeficiente de nebulização $m = 1,50$ foi o que mais se aproximou foi da média obtida no item 5.2.

Os valores de desvio padrão e coeficiente de variação obtidos também indicam que os resultados obtidos são compatíveis com os resultados obtidos no item 5.2.

Com o objetivo de auxiliar a visualização dos resultados, foi criado o gráfico da Figura 6.8. Procurou-se representar para fins de comparação a dispersão das estatísticas dos *casos do histórico* e das estatísticas do AMU para cada um dos valores do coeficiente de nebulização. Neste gráfico os valores das LOLP foram representados no eixo das ordenadas e os diferentes casos foram representados no eixo das abscissas.

Com o auxílio deste gráfico pode ser verificado que os valores das médias obtidas para os casos do AMU se mantêm sempre próximos ao valor da média dos *casos do histórico*, sendo que a que mais se aproxima é a média do caso onde $m = 1,50$.

Por outro lado, pode ser observado que a dispersão dos resultados (distância entre o máximo e o mínimo) só é significativa para o casos em que $m = 1,25$, nos demais casos a

dispersão é muito pequena. Se aliarmos esta observação ao fato que no *Primeiro Conjunto de Testes* do item 4.1 o comportamento da medida para $m = 1,25$ foi pouco satisfatório, por apresentar o comportamento de crescimento a medida que o número de agrupamentos aumenta. Por causa deste comportamento, o número de agrupamentos escolhido pelo valor do número máximo de agrupamentos, ou seja, foi selecionado c igual a 10, o que se refletiu em um número maior de casos do NH2 do que para os outros valores do coeficiente de nebulização, o que explica a maior dispersão neste caso.

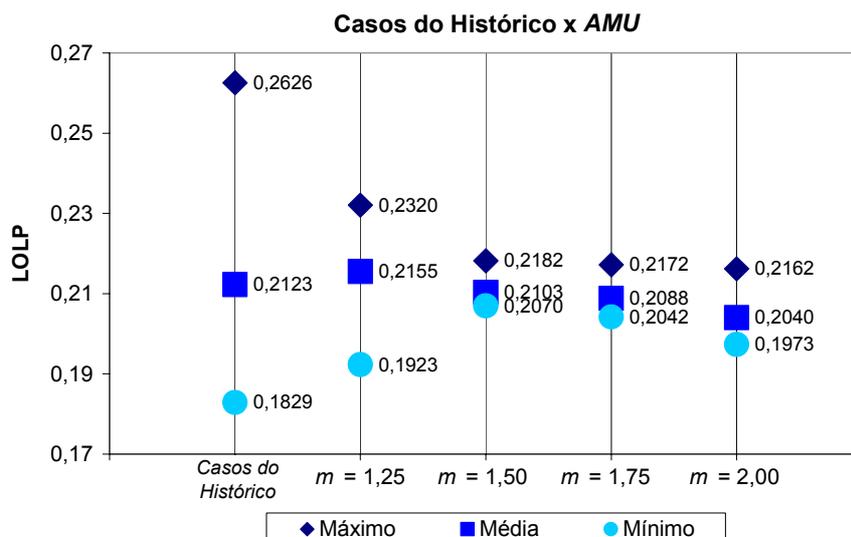


Figura 6.8 – Gráfico de comparação das estatísticas da LOLP - casos do histórico x AMU.

6.4 AGRUPAMENTO POR MAPEAMENTO DE RESULTADOS

O segundo processo de agrupamento definido foi o *Agrupamento por Mapeamento de Resultados*. Este processo apresenta características um pouco mais complexas que o *Agrupamento de Matriz Única*, pois o tratamento da dependência temporal e espacial são feitos separadamente em duas etapas consecutivas.

A motivação para escolha da sua denominação provem do fato de que as duas etapas do processo são interligadas por um *Mapa de Resultados (MR)*.

Nos próximos itens serão descritos com detalhes a finalidade, o espaço de dados, o algoritmo e as características de cada uma das duas etapas deste processo. E finalizando, encontram-se os resultados obtidos após a utilização dos centros determinados por este processo para formar os casos de confiabilidade composta.

6.4.1 PRIMEIRA ETAPA DO AGRUPAMENTO POR MAPEAMENTO DE RESULTADOS

A primeira etapa do *Agrupamento por Mapeamento de Resultados* é responsável por tratar a dependência temporal existente entre as amostras que compõem o espaço de dados de cada uma das usinas que participam do processo de agrupamento, em outras palavras, esta etapa é responsável por formar agrupamentos cuja característica dos centros determinados é manter a dependência temporal das séries de potência disponível de cada uma das usinas.

Como as usinas serão tratadas separadamente, o algoritmo da primeira etapa será composto por um grande laço de processamento onde o espaço de dados de cada usina será agrupado utilizando o FCM, que será executado para todos os números de agrupamento c contidos no intervalo $[2, c_{máx}]$, sendo também calculada a medida de validação ICC para cada um dos resultados. O número de agrupamentos em que ocorrer o valor máximo da ICC indicará qual foi o melhor resultado obtido pelo FCM. Ou seja, a análise dos diversos valores da ICC permitirá determinar qual foi o número de agrupamentos que melhor representou os dados das séries de potência disponível de cada usina que está sendo avaliada.

Avaliando sobre outro ângulo, ao trabalhar com os espaços de dados das usinas separadamente, o *Agrupamento por Mapeamento de Resultados* permite o FCM produza agrupamentos que refletem as características específicas dos espaço de dados de cada uma das usinas. Por exemplo, uma usina em que as amostras encontram-se muito sobrepostas por ter um número de agrupamentos menor que uma usina em que as amostras encontram-se mais distribuídas.

Por causa desta característica, este processo permite uma representação melhor dos espaços de dados das usinas do que o *Agrupamento de Matriz Única*, onde todas as usinas eram representadas pelo mesmo número de agrupamentos.

Uma vez que, na primeira etapa, os dados das usinas serão tratados separadamente, foi necessário criar uma forma de refletir o acoplamento espacial existente entre as diferentes usinas. A forma encontrada foi criar o *Mapa de Resultados (MR)*, que será formado, iterativamente, nesta etapa e servirá como espaço de dados para a segunda etapa do processo.

As formação do *Mapa de Resultados* será apresentada no Passo 4.8. do algoritmo apresentado no item 6.4.1.2 e suas características serão descritas com detalhes no item 6.4.2.1.

6.4.1.1 ESPAÇO DE DADOS DA PRIMEIRA ETAPA

O espaço de dados utilizado na primeira etapa deste processo também é formado a partir do espaço de dados de cada uma das usinas selecionadas no item 6.1.1. Sendo que este espaço de dados é formado colocando as séries de potência disponível de cada uma das usinas uma após a outra.

No item 6.1.1, foi determinado que 30 usinas participariam do processo de agrupamento, portanto a matriz que representa o espaço de dados utilizado na primeira etapa deste processo foi composto por 30 matrizes de tamanho 65 x 12 usinas resultando em uma matriz de 1950 x 12.

Na Figura 6.9 representa o esquema que exemplifica este espaço de dados. A interpretação desta figura é semelhante a apresentada para a Figura 6.7.

| | Séries | Jun/01 | Jul/01 | | Mai/02 |
|-----------------|--------|---------------------|---------------------|--|----------------------|
| Usina 1 | 1 | u1 _{1,1} | u1 _{1,2} | | u1 _{1,12} |
| | 2 | u1 _{2,1} | u1 _{2,2} | | u1 _{2,12} |
| | | | | | |
| | 65 | u1 _{65,1} | u1 _{65,2} | | u1 _{65,12} |
| Usina 2 | 1 | u2 _{1,1} | u2 _{1,2} | | u2 _{1,12} |
| | 2 | u2 _{2,1} | u2 _{2,2} | | u2 _{2,12} |
| | | | | | |
| | 65 | u2 _{65,1} | u2 _{65,2} | | u2 _{65,12} |
| | | | | | |
| Usina 30 | 1 | u30 _{1,1} | u30 _{1,2} | | u30 _{1,12} |
| | 2 | u30 _{2,1} | u30 _{2,2} | | u30 _{2,12} |
| | | | | | |
| | 65 | u30 _{65,1} | u30 _{65,2} | | u30 _{65,12} |

Figura 6.9 – Representação do espaço de dados utilizado no Agrupamento por Mapeamento de Resultados

6.4.1.2 ALGORITMO DA PRIMEIRA ETAPA

O algoritmo desta etapa é descrito a seguir:

/ Inicialização de variáveis, como indicado no item 6.2 **/**

Passo 1. Fixar:

O número de usinas que participam do processo, $n_{usinas} = 30$.

O número máximo de agrupamentos, $c_{m\acute{a}x} = 10$.

O número de rodadas do FCM, $r_{m\acute{a}x} = 20$.

A constante de nebulização, m .

O número de séries de potência disponível de cada usina, $n_{s\acute{e}ries} = 65$.

O número de dimensões do espaço de dados de cada usina, $p = 12$.

/ Criação do Mapa de Resultados (MR) **/**

Passo 2. Criar uma matriz **MR** de tamanho $n_{s\acute{e}ries} \times n_{usinas}$, inicializada com zeros.

/ Laço de formação do espaço de dados **/**

/ Resultados: matrizes X_{usina} e X'_{usina} , de tamanho $(n_{usinas} \cdot n_{s\acute{e}ries}) \times p$ **/**

Passo 3. Para $usina = 1, 2, \dots, n_{usinas}$

Passo 3.1. Utilizar o espaço de dados X_{usina} para formar o espaço de dados X , como apresentado no item 6.4.1.1.

Passo 3.2. Normalizar o espaço de dados X_{usina} , como apresentado no item 6.1.2, obtendo.

Passo 3.3. Utilizar o espaço de dados X'_{usina} para formar o espaço de dados X' , como apresentado no item 6.4.1.1.

Passo 3.4. Retornar ao Passo 3.

/ Laço para execução do FCM e cálculo da ICC **/**

Passo 4. Para $usina = 1, 2, \dots, n_{usinas}$

Passo 4.1. Separar o espaço de dados relativo a $usina$, $X_{usina} \in X$, e $X'_{usina} \in X'_{usina}$

Passo 4.3. Determinar o centro \mathbf{m} do espaço de dados X'_{usina} , como mostrado na Equação 6.10.

$$\mathbf{m} = \frac{1}{n_{s\acute{e}ries}} \sum_{i=1}^{n_{s\acute{e}ries}} \mathbf{x}_i \quad \text{Equação 6.10}$$

Passo 4.4. Para $c = 2, 3, \dots, c_{m\acute{a}x}$

Passo 4.4.1. Inicia $J^{(0)}$ com um valor muito alto.

Passo 4.4.2. Para $r = 1, 2, \dots, r_{m\acute{a}x}$

Passo 4.4.2.1. Executar o FCM, como apresentado no item 3.2.1.3.

Passo 4.4.2.2. Se o valor da função objetivo $J^{(r)}$ for menor que o valor de $J^{(r-1)}$

Armazenar a matriz de pertinências

$U(usina)^{(c)} = U(usina)^{(r)}$, onde $U(usina)^{(c)}$ e

$U(usina)^{(r)}$ são matrizes de tamanho $c \times n_{séries}$

Armazenar a matriz de centros normalizados

$M'(usina)^{(c)} = M'(usina)^{(r)}$, onde $M'(usina)^{(c)}$ e

$M'(usina)^{(r)}$ são matrizes de tamanho $c \times p$.

Passo 4.4.2.3. Retornar ao Passo 4.4.2.

Passo 4.4.3. Calcular a $ICC(c)$, como apresentado no item 3.4.4.

Passo 4.4.4. Retornar ao Passo 4.4.

Passo 4.5. Determinar o número de agrupamentos c_{sel} onde ocorre o valor máximo da $ICC(c)$.

Passo 4.7. Utilizar a matriz de pertinências $U^{(c_{sel})}$ e o espaço de dados X_{usina} , para obter os centros não normalizados dos agrupamentos, M_{usina} .

/ Laço para formação do Mapa de Resultados, apresentado na Figura 6.10 **/**

Passo 4.8. Para $série = 1, 2, \dots, n_{séries}$

Passo 4.8.1. Utilizando a matriz de pertinência $U^{(c_{sel})}$, determinar a qual agrupamento, c_{MR} , pertence a *série*, através da Equação 6.11.

$c_{MR} = i$, sendo i é a posição

onde ocorre $\max_{i=1}^{c_{sel}} [U^{(c_{sel})}(i, série)]$ Equação 6.11

Passo 4.8.2. Armazenar o valor c_{MR} na célula da matriz **MR**, como mostra a Equação 6.12.

$MR(série, usina) = c_{MR}$ Equação 6.12

Passo 4.8.3. Retornar ao passo 4.8.

Passo 4.9. Retornar ao Passo 4.

6.4.2 SEGUNDA ETAPA DO AGRUPAMENTO POR MAPEAMENTO DE RESULTADOS

A segunda etapa do *Agrupamento por Mapeamento de Resultados* é responsável por tratar a dependência espacial existente entre as usinas que participam do processo de agrupamento.

Para criar esta etapa o algoritmo de particionamento selecionado foi o PAM, que foi utilizado em conjunto com a Silhueta Média Global (\bar{S}), para determinação do melhor número

de agrupamentos. A finalidade desta etapa é realizar o particionamento do *Mapa de Resultados (MR)*, mantendo a dependência espacial existente entre as usinas.

A principal motivação para se escolher o PAM como algoritmo de agrupamento nesta etapa do processo foi sua característica de determinar um centrótipo, representado por uma das amostras existentes no espaço de dados. Outra característica do PAM necessária para esta etapa é sua versatilidade em utilizar tipos de dados diferentes dos dados numéricos.

Os espaço de dados desta etapa, o *Mapa de Resultados (MR)*, assim como o seu algoritmo será descrito a seguir. Ao final deste item serão apresentados os resultados obtidos pelo processo de agrupamento e pela análise de confiabilidade efetuada com base nos centros definidos.

6.4.2.1 ESPAÇO DE DADOS DA SEGUNDA ETAPA - O *MAPA DE RESULTADOS (MR)*

O *Mapa de Resultados (MR)*, formado pelo laço iniciado no Passo 4.8. do algoritmo apresentado no item 6.4.1.2, representa o espaço de dados que realiza a ligação entre a primeira e a segunda etapa do *Agrupamento por Mapeamento de Resultados*, sendo responsável por refletir o acoplamento espacial existente entre as usinas que participam do processo de agrupamento.

O **MR** também pode ser representado matematicamente por uma matriz, de tamanho $n_{séries} \times n_{usinas}$, na qual cada célula possui a representação de qual agrupamento uma da série de potência disponível de uma usina foi associada.

Uma possível representação desta matriz **MR** é apresentada na Figura 6.10. Nesta figura, cada célula possui um mnemônico, $cK_{uN,s}$, para o valor que deve ser utilizado na célula, onde:

- | | |
|----|---|
| CK | – Indica o número do agrupamento determinado pelo FCM em que a série encontra-se alocada. |
| uN | – Indica que a usina N que está representada |
| s | – Indica a série que está representada |

Por exemplo, uma célula com a representação $c7_{23,38}$, indica a série número 38 da vigésima terceira usina representada no conjunto de dados encontra-se alocada no agrupamento 7 desta usina.

| Série | Usina 1 | Usina 2 | | Usina 30 |
|-------|--------------|--------------|--|---------------|
| 1 | $c1_{u1,1}$ | $c1_{u2,1}$ | | $c5_{u30,1}$ |
| 2 | $c1_{u1,2}$ | $c1_{u2,2}$ | | $c5_{u30,2}$ |
| 3 | $c2_{u1,3}$ | $c3_{u2,3}$ | | $c1_{u30,3}$ |
| | | | | |
| 65 | $c2_{u1,65}$ | $c1_{u2,65}$ | | $c3_{u30,65}$ |

Figura 6.10 – Representação do *Mapa de Resultados*

Devido a sua característica de formação, o **MR** é um espaço de dados formado por valores nominais, ou seja, cada valor $cK_{uN,s}$ representa uma característica de uma dimensão do espaço de dados, cuja interpretação é o agrupamento ao qual uma determinada série de uma usina está alocada. Por se tratar de uma matriz de dados nominais, não é possível determinar uma distância entre as amostras, sendo necessário utilizar uma medida de dissimilaridade específica para dados nominais. Neste trabalho foi utilizada a medida de dissimilaridade apresentada na Equação 6.13.

$$D(\mathbf{x}_i, \mathbf{x}_j) = \frac{p_{MR} - m_{MR}}{p_{MR}} \quad \text{Equação 6.13}$$

Onde:

- $D(\mathbf{x}_i, \mathbf{x}_j)$ – Dissimilaridade entre a amostra \mathbf{x}_i e a amostra \mathbf{x}_j , onde $i \neq j$
- \mathbf{x}_i – Amostra i
- \mathbf{x}_j – Amostra j
- p_{MR} – Número total de dimensões das amostras
- m_{MR} – Número de dimensões das amostras \mathbf{x}_i e \mathbf{x}_j que tem o mesmo valor, ou seja, que são iguais

Esta medida de dissimilaridade visa determinar quão diferentes são duas amostras, comparando para isso as suas dimensões segundo a característica indicada. No caso em que todas as dimensões que compõem as duas amostras tem mesmo valor, ou seja as amostras tem todas as características iguais, o resultado da medida de dissimilaridade assumirá o valor 0, por outro lado quando todas as dimensões que compõem duas amostras tem valores diferentes o resultado será 1.

A Figura 6.11 apresenta um exemplo de duas amostra de um **MR**, neste exemplo as dimensões que tem valores iguais encontram-se em **negrito/sublinhado**, e a dissimilaridade será dada pelo cálculo abaixo.

| | Usina 1 | Usina 2 | Usina 3 | Usina 4 | Usina 5 | Usina 5 | Usina 7 | Usina 8 | Usina 9 | Usina 10 | Usina 11 | Usina 12 | Usina 13 | Usina 14 | Usina 15 | Usina 16 | Usina 17 | Usina 18 | Usina 19 | Usina 20 | Usina 21 | Usina 22 | Usina 23 | Usina 24 | Usina 25 | Usina 26 | Usina 27 | Usina 28 | Usina 29 | Usina 30 |
|-------|---------|---------|---------|---------|---------|---------|---------|---------|---------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| x_i | c1 | c1 | c1 | c1 | c2 | c2 | c3 | c3 | c3 | c4 | c5 | c6 | c2 | c2 | c2 | c3 | c3 | c7 | c7 | c7 | c8 | c8 | c8 | c9 | c7 | c4 | c1 | c1 | c1 | c1 |
| x_j | c1 | c1 | c1 | c3 | c3 | c2 | c3 | c2 | c3 | c4 | c5 | c6 | c1 | c2 | c2 | c3 | c3 | c1 | c2 | c3 | c4 | c5 | c6 | c9 | c7 | c1 | c2 | c3 | c1 | c1 |

Figura 6.11 - Exemplo de amostras do MR

$$D = \frac{30 - 17}{30} = 0,433 \quad \text{onde } p_{MR} = 30 \text{ e } m_{MR} = 17$$

6.4.2.2 ALGORITMO DA SEGUNDA ETAPA

O algoritmo desta etapa é descrito a seguir:

/ Inicialização de variáveis para a etapa **/**

Passo 1. Fixar:

O número de linhas do MR, $l_{MR} = 65$.

O número de colunas do MR, $p_{MR} = 30$.

O número máximo de agrupamentos $c_{máx} = 10$.

Iniciar, com zeros, a matriz de dissimilaridades **D** de tamanho $l_{MR} \times l_{MR}$

/ Laço para determinação da matriz de Dissimilaridades baseada no Mapa de Resultados obtido da primeira etapa do algoritmo **/**

/ Resultado: a matriz D, uma matriz, de tamanho $l_{MR} \times l_{MR}$, triangular inferior com a diagonal principal zerada **/**

Passo 2. Para $i = 1, 2, \dots, (l_{MR} - 1)$

Passo 2.1. Selecionar uma amostra $x_i = \mathbf{MR}(i)$

Passo 2.2. Para $j = (i + 1), \dots, l_{MR}$

Passo 2.2.1. Selecionar uma amostra $x_j = \mathbf{MR}(j)$

Passo 2.2.2. Para $k = 1, 2, \dots, p_{MR}$

Passo 2.2.2.1. Se a dimensão k de x_i e x_j são iguais, $x_i(k) = x_j(k)$

$$m_{MR} = m_{MR} + 1$$

Senão

Retornar ao Passo 2.2.

Passo 2.2.3. Calcular dissimilaridade $D(x_i, x_j)$ segundo a Equação 6.13.

Passo 2.2.4. Armazenar o resultado de $D(x_i, x_j)$ na matriz de dissimilaridades

\mathbf{D} , na posição $\mathbf{D}(i, k)$.

Passo 2.2.5. Retornar ao Passo 2.2.

Passo 2.3. Retornar ao Passo 2.

/ Laço para execução do PAM e cálculo da \bar{S} */**

Passo 3. Para $c = 2, 3, \dots, c_{máx}$

Passo 3.1. Executar o PAM, como apresentado no item 3.2.1.2.

Passo 3.2. Calcular \bar{S} , como apresentado no item 3.4.5.

Passo 3.3. Armazenar a $\bar{S}^{(c)}$

Passo 3.4. Armazenar o vetor de centrótipos $\mathbf{C}^{(c)}$,
onde $\mathbf{C}^{(c)}$ é um vetor de tamanho c , cujas posições possuem os
números dos centrótipos selecionados.

Passo 3.5. Armazenar o número de séries associadas a cada centrótipos $\mathbf{ns}^{(c)}$.

Passo 3.6. Retornar ao Passo 3.

Passo 4. Determinar o número de agrupamentos c_{sel} onde ocorre o máximo da \bar{S} .

/ Laço para obter os valores que serão utilizados nos casos de confiabilidade */**

Passo 5. Criar uma matriz \mathbf{R} , de tamanho $c_{sel} \times n_{usinas}$ para armazenar os valores que serão utilizados para formar os casos de confiabilidade. Inicializada com zeros.

Passo 6. Selecionar o vetor de centrótipos $\mathbf{C}^{(c_{sel})}$.

Passo 7. Para $i = 1, 2, \dots, c_{sel}$

Passo 7.1. Armazenar $n_{centrótipo} = \mathbf{C}(i)^{(c_{sel})}$

Passo 7.2. Armazenar $\mathbf{MR}_{centrótipo} = \mathbf{MR}(n_{centrótipo})$,
onde $\mathbf{MR}_{centrótipo}$ é um vetor de n_{usinas} posições

Passo 7.3. Para $usina = 1, 2, \dots, n_{usinas}$

Passo 7.3.1. Armazenar $c_{MR} = \mathbf{MR}_{centrótipo}(usina)$,
onde c_{MR} indica o número do agrupamento da $usina$ que
deve ser utilizado.

Passo 7.3.2. Selecionar $\mathbf{m}_{usina} = \mathbf{M}_{usina}(c_{MR})$, onde \mathbf{m}_{usina} é o vetor com os dados
do centro de agrupamento indicado por c_{MR} .

Passo 7.3.3. Selecionar a partir de \mathbf{m}_{usina} o valor que será utilizado para montar
o arquivo do caso de confiabilidade e armazenar na matriz de
resultados, $\mathbf{R}(i, usina)$

Passo 7.3.4. Retornar ao Passo 7.3.

Passo 7.4. Retornar ao Passo 7.

6.4.3 RESULTADOS E CONCLUSÕES SOBRE O PROCESSO DE AGRUPAMENTO

Neste processo de agrupamento tem-se como resultado a matriz **R** de tamanho $c_{sel} \times n_{usinas}$, que recebe os valores que devem ser utilizados para formar os casos de confiabilidade composta.

Neste trabalho, os valores que encontram-se armazenados na matriz **R** são referente a décima coluna dos centros, determinados pelo FCM na primeira etapa para cada uma das usinas, ou seja, o mês de março/2002. Esta matriz de resultados foi utilizada como dado de entrada para o algoritmo para formação dos casos do NH2 descrito no item 5.2.

Os resultados referentes ao *Agrupamento por Mapeamento de Resultados*, que a partir deste ponto será referenciado como *AMR*, encontram-se listados na Tabela 6.8, onde encontram-se relacionados o coeficiente de nebulização utilizado no FCM, o número de agrupamentos final determinado pela Silhueta Média Global \bar{S} , assim como o seu valor calculado, e o tempo total de processamento, que inclui a execução do processo e a criação dos casos de confiabilidade composta baseados nos centros determinados.

Tabela 6.8 – Resultados do AMR

| Coeficiente de Nebulização (<i>m</i>) | Número de Agrupamentos (<i>c</i>) | \bar{S} | Tempo de Processamento | Tempo de Processamento Relativo (%)¹⁹ |
|--|--|-----------|-------------------------------|---|
| 1,25 | 3 | 0,2400 | 4m 45s | 1,8 |
| 1,50 | 4 | 0,2358 | 5m 32s | 2,0 |
| 1,75 | 4 | 0,2704 | 7m 24s | 2,7 |
| 2,00 | 3 | 0,2841 | 5m 27s | 2,0 |
| Tempo Total de Processamento | | | 23m 8s | 8,5 |

O tempo total para execução do processo para os quatro valores do coeficiente de nebulização foi de 23min e 8seg. Da mesma forma que no *AMU*, teste tempo de processamento é muito menor que o tempo médio de processamento de um caso do NH2, que é de 4hs e 30 min. Portanto, assim como no *AMU*, o tempo de processamento da avaliação de confiabilidade,

¹⁹ Tempo de processamento do processo de agrupamento dividido pelo tempo médio de processamento de um caso de confiabilidade composta (utilizando o programa NH2)

seria muito pouco afetado pela inclusão do processo de agrupamento no fluxo de informação entre os sistemas Newave, Suishi-O e NH2.

Pelos mesmos motivos apresentados no item 6.3.3, no *AMR* os resultados relativos a redução do tempo de processamento da análise de confiabilidade também são muito dependente da forma com que o processo for montado, ou seja, da quantidade de coeficientes de nebulização selecionados para avaliação e do número máximo de agrupamentos a ser considerado.

6.4.4 RESULTADOS DA ANÁLISE DE CONFIABILIDADE

As tabelas a seguir (Tabela 6.9, Tabela 6.10, Tabela 6.11, Tabela 6.12) apresentam os resultados obtidos após a execução dos casos de confiabilidade composta no NH2. Nesta tabela encontram-se listados os principais resultados obtidos para cada um dos centros dos agrupamentos determinados pelo processo agrupados pelo coeficiente de nebulização utilizado. O tempo total para execução dos casos do NH2 foi de aproximadamente 63hs. Comparando ao tempo de processamento dos *casos do histórico*, determinado no item 5.2, houve uma redução de aproximadamente 237hs.

A Tabela 6.13 apresenta as estatísticas da LOLP relativas aos resultados obtidos após a execução do NH2 utilizando os centros determinados pelo *AMR*.

Tabela 6.9 – Resultados do NH2 – Casos formados pelo *AMR* – $m = 1,25$

| A g r a. | 21 n. c. a. | Pot. Máx. (MW) | LOLP | | EPNS (MW) | | LOLE (h/ano) | EENS (MWh/ ano) | Monte Carlo ²² | | |
|-------------------|----------------------|----------------------|--------|-----------|-----------|-----------|-----------------|-----------------------|---------------------------|---------------------|--------------------|
| | | | Valor | CV (%) | Valor | CV (%) | | | Sim. ²³ | Sist. ²⁴ | Cort ²⁵ |
| 1 | 14 | 59199,8 | 0,2195 | 4,06 | 2,08 | 15,00 | 1922,74 | 18180,46 | 2319 | 1751 | 509 |
| 2 | 13 | 55321,7 | 0,2090 | 4,05 | 1,72 | 14,94 | 1830,60 | 15047,26 | 2608 | 1926 | 545 |
| 3 | 46 | 58058,8 | 0,2194 | 5,00 | 1,53 | 11,72 | 1922,38 | 13400,09 | 1563 | 1207 | 343 |

²⁰ Número de agrupamentos

²¹ Número de curvas no agrupamento

²² Resultados provenientes da simulação de Monte Carlo para determinação dos índices de confiabilidade.

²³ Número de eventos simulados e solucionados

²⁴ Número de eventos críticos com problemas no sistema

²⁵ Número de eventos críticos com corte de carga

Tabela 6.10 – Resultados do NH2 – Casos formados pelo AMR – $m = 1,50$

| A g r | n. c. a. | Pot. Máx. (MW) | LOLP | | EPNS (MW) | | LOLE (h/ano) | EENS (MWh/ ano) | Monte Carlo | | |
|-------------|----------------|----------------------|--------|-----------|-----------|-----------|-----------------|-----------------------|-------------|-------|------|
| | | | Valor | CV (%) | Valor | CV (%) | | | Sim. | Sist. | Cort |
| 1 | 18 | 57500,6 | 0,2237 | 4,99 | 1,55 | 11,63 | 1959,92 | 13620,41 | 1542 | 1211 | 345 |
| 2 | 28 | 58346,9 | 0,1907 | 4,16 | 1,73 | 14,88 | 1670,60 | 15194,98 | 2665 | 1897 | 509 |
| 3 | 9 | 59155,9 | 0,2066 | 5,00 | 1,59 | 11,39 | 1809,83 | 13908,51 | 1636 | 1219 | 338 |
| 4 | 10 | 55456,7 | 0,2462 | 3,76 | 1,80 | 14,95 | 2157,10 | 15760,50 | 2396 | 1958 | 590 |

Tabela 6.11 – Resultados do NH2 – Casos formados pelo AMR – $m = 1,75$

| A g r | n. c. a. | Pot. Máx. (MW) | LOLP | | EPNS (MW) | | LOLE (h/ano) | EENS (MWh/ ano) | Monte Carlo | | |
|-------------|----------------|----------------------|--------|-----------|-----------|-----------|-----------------|-----------------------|-------------|-------|------|
| | | | Valor | CV (%) | Valor | CV (%) | | | Sim. | Sist. | Cort |
| 1 | 16 | 59164,0 | 0,1988 | 4,99 | 1,52 | 11,32 | 1741,84 | 13273,49 | 1725 | 1256 | 343 |
| 2 | 10 | 55655,7 | 0,2162 | 3,88 | 1,56 | 15,00 | 1893,50 | 13637,68 | 2563 | 2044 | 554 |
| 3 | 13 | 57340,9 | 0,2063 | 4,16 | 1,70 | 15,00 | 1807,33 | 14915,94 | 2438 | 1755 | 503 |
| 4 | 26 | 58140,1 | 0,2086 | 5,00 | 1,54 | 12,24 | 1827,24 | 13529,70 | 1630 | 1238 | 340 |

Tabela 6.12 – Estatísticas dos resultados dos casos do NH2 (AMR) para LOLP

| A g r | n. c. a. | Pot. Máx. (MW) | LOLP | | EPNS (MW) | | LOLE (h/ano) | EENS (MWh/ ano) | Monte Carlo | | |
|-------------|----------------|----------------------|--------|-----------|-----------|-----------|-----------------|-----------------------|-------------|-------|------|
| | | | Valor | CV (%) | Valor | CV (%) | | | Sim. | Sist. | Cort |
| 1 | 35 | 58194,0 | 0,2049 | 5,00 | 1,40 | 9,65 | 1795,08 | 12284,62 | 1708 | 1181 | 350 |
| 2 | 16 | 59016,0 | 0,2176 | 5,00 | 1,66 | 11,71 | 1906,32 | 14577,54 | 1544 | 1137 | 336 |
| 3 | 14 | 55953,7 | 0,2056 | 4,06 | 1,67 | 14,98 | 1800,74 | 14665,35 | 2588 | 1925 | 532 |

Tabela 6.13 – Estatísticas dos resultados dos casos do NH2 (AMR) para LOLP

| Coefficiente de Nebulização (m) | 1,25 | 1,50 | 1,75 | 2,00 |
|-------------------------------------|----------|----------|----------|----------|
| Máximo | 0,232049 | 0,218182 | 0,217193 | 0,216172 |
| Mínimo | 0,192308 | 0,206959 | 0,204179 | 0,197323 |
| Média Ponderada | 0,005333 | 0,210599 | 0,206899 | 0,208183 |
| Erro Médio Absoluto | 0,004252 | 0,001660 | 0,005360 | 0,004076 |
| Desvio Padrão (%) | 1,9541 | 2,0394 | 0,5511 | 0,5397 |
| Coefficiente de Variação (%) | 1,0519 | 9,6837 | 2,6636 | 2,5922 |

Avaliando o erro médio absoluto foi possível verificar que em todos os casos a média da LOLP foi preservada pelos casos determinados pelo *AMR*. Esta afirmação baseia-se no fato de que em todos os casos o erro médio absoluto determinado é menor que 1%. Sendo que, a média obtida para o coeficiente de nebulização $m = 2,00$ foi o que mais se aproximou da média obtida no item 5.2.

Para visualizar dos resultados obtidos, foi usado o gráfico da Figura 6.12 que apresenta as mesmas características do gráfico da Figura 6.8.

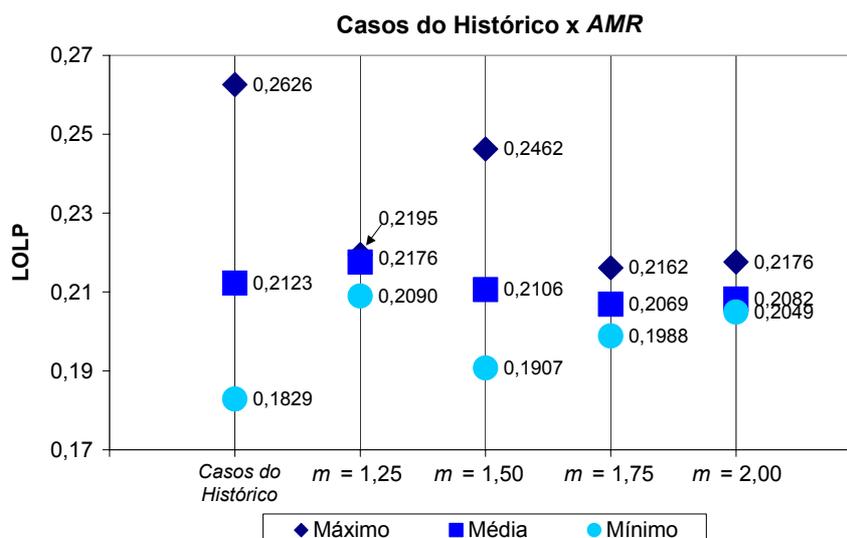


Figura 6.12 – Gráfico de comparação das estatísticas da LOLP - casos do histórico x AMR

Neste gráfico pode ser verificado que os valores das médias obtidas para os casos do *AMR* se mantém sempre próximos ao valor da média dos *casos do histórico*, sendo que a que mais se aproxima é a média do caso onde $m = 2,00$.

Por outro lado, pode ser observado que a dispersão dos resultados (distância entre o máximo e o mínimo) só é significativa para os casos em que $m = 1,50$, nos demais casos a dispersão é muito pequena.

Uma grande diferença entre este método e o *AMU* está no fato de que no *AMR* o número de agrupamentos para cada uma das usinas reflete as características do seu espaço de dados. No *AMU*, como os dados foram todos apresentados para o algoritmo em uma única matriz, todas as usinas são obrigadas a apresentar um mesmo número de agrupamentos independente das características de dispersão e sobreposição existentes em seus dados.

6.5 AGRUPAMENTO POR MAPEAMENTO DE RESULTADOS COM SELEÇÃO DO COEFICIENTE DE NEBULIZAÇÃO

O objetivo deste último processo é utilizar a medida *ICC* não só para determinar o melhor número de agrupamentos c para o espaço de dados de cada uma das usinas, mas também para indicação do melhor valor do coeficiente de nebulização m que deve ser utilizado em cada um dos casos, assim como foi mostrado no item 4.3.

Este último processo foi elaborado a partir do *AMR* difere deste somente no algoritmo da primeira etapa, descrito no item 6.4.1.2. Para tanto, foi incluído mais um laço no fluxo do algoritmo da primeira etapa que efetua a troca dos valores do coeficiente de nebulização.

A inclusão da seleção do coeficiente de nebulização (m) a motivou a denominação deste algoritmo.

6.5.1 ALGORITMO ALTERADO DA PRIMEIRA ETAPA

Como foi citado anteriormente, este processo de agrupamento difere do *AMR* somente no algoritmo da primeira etapa que será apresentado a seguir:

/ Inicialização de variáveis, como indicado no item 6.2 **/**

Passo 1. Fixar:

O número de usinas que participam do processo, $n_{usinas} = 30$.

O número máximo de agrupamentos, $c_{máx} = 10$.

O número de rodadas do FCM, $r_{máx} = 20$.

O número de séries de potência disponível de cada usina, $n_{séries} = 65$.

O número de dimensões do espaço de dados de cada usina, $p = 12$.

/ Criação do Mapa de Resultados (MR) **/**

Passo 2. Criar uma matriz **MR** de tamanho $n_{séries} \times n_{usinas}$, inicializada com zeros.

/ Laço de formação do espaço de dados **/**

/ Resultados: matrizes X_{usina} e X'_{usinas} de tamanho $(n_{usinas} \cdot n_{séries}) \times p$ **/**

Passo 3. Para $usina = 1, 2, \dots, n_{usinas}$

Passo 3.1. Utilizar o espaço de dados X_{usina} para formar o espaço de dados X , como apresentado no item 6.4.1.1.

Passo 3.2. Normalizar o espaço de dados X_{usina} , como apresentado no item 6.1.2, obtendo.

Passo 3.3. Utilizar o espaço de dados X'_{usina} para formar o espaço de dados X' , como apresentado no item 6.4.1.1.

Passo 3.4. Retornar ao Passo 3.

// Laço para execução do FCM e cálculo da ICC /**/**

Passo 4. Para $usina = 1, 2, \dots, n_{usinas}$

Passo 4.1. Separar o espaço de dados relativo a $usina$, $X_{usina} \in X$, e $X'_{usina} \in X'_{usina}$

Passo 4.2. Determinar o centro \mathbf{m} do espaço de dados X'_{usina} , como mostrado na Equação 6.10.

$$\mathbf{m} = \frac{1}{n_{séries}} \sum_{i=1}^{n_{séries}} \mathbf{x}_i \quad \text{Equação 6.14}$$

// Início da alteração em relação ao AMR /**/**

// Inclusão do laço para escolha do coeficiente de nebulização m /**/**

Passo 4.3. Fazer $m = 1$

Passo 4.4. Para $m < 2,00$ **/**/ Permite utilizar $m = 1,25; 1,50; 1,75$ e $2,00$ /**/**

Passo 4.4.1. $m = m + 0,25$

Passo 4.4.2. Para $c = 2, 3, \dots, c_{máx}$

Passo 4.4.2.1. Inicia $J^{(0)}$ com um valor muito alto.

Passo 4.4.2.2. Para $r = 1, 2, \dots, r_{máx}$

Passo 4.4.2.2.1. Executar o FCM, como apresentado no item 3.2.1.3.

Passo 4.4.2.2.2. Se o valor da função objetivo $J^{(r)}$ for menor que o valor de $J^{(r-1)}$

Armazenar a matriz de pertinências

$\mathbf{U}(usina)^{(c)} = \mathbf{U}(usina)^{(r)}$, onde $\mathbf{U}(usina)^{(c)}$ e

$\mathbf{U}(usina)^{(r)}$ são matrizes de tamanho $c \times n_{séries}$

Armazenar a matriz de centros normalizados

$\mathbf{M}'(usina)^{(c)} = \mathbf{M}'(usina)^{(r)}$,

onde $\mathbf{M}'(usina)^{(c)}$ e $\mathbf{M}'(usina)^{(r)}$ são matrizes de tamanho $c \times p$.

Passo 4.4.2.2.3. Retornar ao Passo 4.4.2.

Passo 4.4.2.3. Calcular a $ICC(c)$, como apresentado no item 3.4.4.

Passo 4.4.2.4. Retornar ao Passo 4.4.2.

Passo 4.4.3. Determinar o número de agrupamentos c_{sel} onde ocorre o valor máximo da $ICC(c)$.

Passo 4.4.4. Armazenar $ICC_{sel}(m)$, onde ICC_{sel} é um vetor onde são armazenados os valores de ICC selecionado para cada um dos valores de m

Passo 4.4.5. Armazenar $C(m) = c'_{sel}$, onde $C(m)$ é um vetor onde são armazenados os números de agrupamentos em que ocorreram os valores máximos de ICC_{sel}

Passo 4.4.6. Retornar ao Passo 4.4.

Passo 4.5. Determinar m_{sel} onde ocorre o máximo de ICC_{sel}

Passo 4.5. Determinar o número de agrupamentos a ser utilizado para a *usina*,

$$c_{sel} = C(m_{sel})$$

Passo 4.7. Utilizar a matriz de pertinências $U^{(c_{sel})}$ e o espaço de dados X_{usina} , para obter os centros não normalizados dos agrupamentos, M_{usina} .

/** Fim da alteração em relação ao AMR **/

/** Laço para formação do Mapa de Resultados, apresentado na Figura 6.10 **/

Passo 4.8. Para $série = 1, 2, \dots, n_{séries}$

Passo 4.8.1. Utilizando a matriz de pertinência $U^{(c_{sel})}$, determinar a qual agrupamento, c_{MR} , pertence a *série*, através da Equação 6.11.

$$c_{MR} = i, \text{ sendo } i \text{ é a posição}$$

$$\text{onde ocorre } \max_{i=1}^{c_{sel}} [U^{(c_{sel})}(i, série)] \quad \text{Equação 6.15}$$

Passo 4.8.2. Armazenar o valor c_{MR} na célula da matriz MR , como mostra a Equação 6.12.

$$MR(série, usina) = c_{MR} \quad \text{Equação 6.16}$$

Passo 4.8.3. Retornar ao passo 4.8.

Passo 4.9. Retornar ao Passo 4.

6.5.2 RESULTADOS E CONCLUSÕES SOBRE O PROCESSO DE AGRUPAMENTO

Assim como no *AMR*, o principal resultado deste processo é a matriz R de tamanho c_{sel} x n_{usinas} , determinado na segunda etapa do algoritmo cujo algoritmo de processamento é igual ao apresentado no item 6.4.2.2. Esta matriz recebe os valores que devem ser utilizados para formar

os casos de confiabilidade composta, sendo que estes se referem a décima coluna dos centros, determinados pelo FCM na primeira etapa para cada uma das usinas, ou seja, o mês de março/2002. Esta matriz de resultados foi utilizada como dado de entrada para o algoritmo para formação dos casos do NH2 descrito no item 5.2.

Os resultados referentes ao *Agrupamento por Mapeamento de Resultados com Seleção do Coeficiente de Nebulização*, que a partir deste ponto será referenciado como *AMRM*, encontram-se listados na Tabela 6.14, onde encontram-se relacionados o número de agrupamentos determinado pela \bar{S} assim como o seu valor, e o tempo total de processamento, que inclui a execução do processo e a criação dos casos de confiabilidade composta baseados nos centros determinados.

Tabela 6.14 – Resultados do *AMRM*

| Número de Agrupamentos (<i>c</i>) | \bar{S} | Tempo de Processamento | Tempo de Processamento Relativo (%)²⁶ |
|--|-----------|-------------------------------|---|
| 7 | 0,2208 | 20m 43s | 7,7 |

Como este processo já avalia os diferentes valores do coeficiente de nebulização durante o processamento, ele resulta somente em um conjunto de centros. Neste caso o tempo total para execução do processo, apresentado na Tabela 6.14, foi de 20min e 43seg. Este tempo de processamento é compatível com o tempo total de processamento obtido nos outros dois processos, por tanto também não é significativo em relação ao tempo de processamento necessário para execução de um caso do NH2.

O *AMRM* apresenta uma vantagem em relação ao *AMU* e ao *AMR*. Neste o número mínimo e máximo de casos de confiabilidade composta que podem ser definidos dependente somente do número mínimo e máximo de agrupamentos que serão considerados. Por exemplo, nesta dissertação o *AMRM* só pode definir resultados no intervalo [2, 10].

A dependência em relação ao número de coeficientes de nebulização avaliados fica restrita a variação no tempo de processamento do próprio *AMRM*. Foram realizados testes com coeficientes efetuando a seleção de 10 valores de coeficiente de nebulização o que aumentou o tempo de processamento do *AMRM* para 1h e 10min. Deve ser ressaltado que este tempo de

²⁶ Tempo de processamento do processo de agrupamento dividido pelo tempo médio de processamento de um caso de confiabilidade composta (utilizando o programa NH2)

processamento resulta ainda em 25,9% de tempo relativo, ou seja, ainda bastante menor que o tempo de processamento de um caso do NH2.

Por tanto, o *AMRM* se apresenta como o processo com melhor desempenho entre os 3 definidos nesta dissertação, pois mantém o esforço computacional do processo em si e reduz o intervalo do número de centros que este pode definir. Esta redução se dá devido ao fato que no *AMU* e no *AMR* o intervalo do número de centros que podem ser definidos [$ncentros_{min}$, $ncentros_{máx}$] tem seus limites dados pela Equação 6.17. Por sua vez, no *AMRM* este intervalo é simplesmente por [c_{min} , $c_{máx}$].

$$\begin{aligned} ncentros_{min} &= nm \cdot c_{min} \\ ncentros_{máx} &= nm \cdot c_{máx} \end{aligned} \quad \text{Equação 6.17}$$

Onde:

- $ncentros_{min}$ – Número mínimo de centros que podem ser definidos
- $ncentros_{máx}$ – Número máximo de centros que podem ser definidos
- nm – Número de coeficientes de nebulização avaliados
- c_{min} – Número mínimo de agrupamentos
- $c_{máx}$ – Número máximo de agrupamentos

Analisando estes valores tendo como base os valores utilizados para $c_{min} = 2$ e $c_{máx} = 10$, como foi utilizado neste trabalho, a redução neste número de casos de 65 (analisados no item 5.2) para 2, resultaria em uma redução de 97% no tempo de processamento, ou seja, um percentual muito alto. Por outro lado, a redução de 65 para 10 casos, resultaria em uma redução de 84,6%, que também se trata de um percentual muito interessante.

6.5.3 RESULTADOS E CONCLUSÕES SOBRE A ANÁLISE DE CONFIABILIDADE

Os resultados obtidos após a execução dos casos de confiabilidade composta são apresentados na Tabela 6.15.

Neste caso o tempo total para execução dos casos do NH2 foi de aproximadamente 35h, ou seja, houve uma redução de aproximadamente 265h, que representa o melhor resultado em termos de esforço computacional.

O gráfico da Figura 6.13, também apresenta para fins de comparação a dispersão das estatísticas dos *casos do histórico* e das estatísticas do *AMRM*.

Tabela 6.15 – Resultados do NH2 – Casos formados pelo *AMRM*

| Agrupamento | n. c. a. | Pot. Máx. (MW) | LOLP | | EPNS (MW) | | LOLE (h/ano) | EENS (MWh/ano) | Monte Carlo ²⁹ | | |
|-------------|----------|----------------|--------|--------|-----------|--------|--------------|----------------|---------------------------|---------------------|--------------------|
| | | | Valor | CV (%) | Valor | CV (%) | | | Sim. ³⁰ | Sist. ³¹ | Cort ³² |
| 1 | 10 | 59028,1 | 0,2220 | 4,21 | 2,23 | 14,94 | 1944,40 | 19561,36 | 2673 | 1895 | 553 |
| 2 | 10 | 55472,8 | 0,2254 | 3,91 | 1,74 | 14,92 | 1974,49 | 15262,50 | 2598 | 1992 | 573 |
| 3 | 8 | 58287,4 | 0,1898 | 4,14 | 1,64 | 14,99 | 1662,43 | 14400,17 | 3520 | 2529 | 668 |
| 4 | 12 | 57872,2 | 0,2118 | 4,99 | 1,46 | 11,27 | 1855,06 | 12773,13 | 2274 | 1773 | 481 |
| 5 | 6 | 59172,2 | 0,2055 | 4,16 | 1,90 | 14,99 | 1800,20 | 16611,80 | 2588 | 1925 | 532 |
| 6 | 11 | 57184,9 | 0,2198 | 3,94 | 1,62 | 14,99 | 1925,08 | 14233,70 | 2662 | 1885 | 545 |
| 7 | 8 | 58480,9 | 0,2053 | 4,02 | 1,68 | 14,99 | 1798,81 | 14673,79 | 2583 | 1920 | 523 |

Tabela 6.16 – Estatísticas dos resultados dos casos do NH2 (*AMRM*) para LOLP

| | LOLP |
|-------------------------------------|----------|
| Máximo | 0,225398 |
| Mínimo | 0,189776 |
| Média Ponderada | 0,211358 |
| Erro Médio Absoluto | 0,212709 |
| Desvio Padrão (%) | 0,1374 |
| Coefficiente de Variação (%) | 0,6460 |

A Tabela 6.16 apresenta as estatísticas da LOLP relativas aos resultados obtidos após a execução do NH2 utilizando os centros determinados pelo *AMRM*.

Com o auxílio deste gráfico pode ser verificado que o valor da média obtida para o *AMRM* também se mantém sempre próxima ao valor da média dos *casos do histórico*.

²⁷ Número de agrupamentos

²⁸ Número de curvas no agrupamento

²⁹ Resultados provenientes da simulação de Monte Carlo para determinação dos índices de confiabilidade.

³⁰ Número de eventos simulados e solucionados

³¹ Número de eventos críticos com problemas no sistema

³² Número de eventos críticos com corte de carga

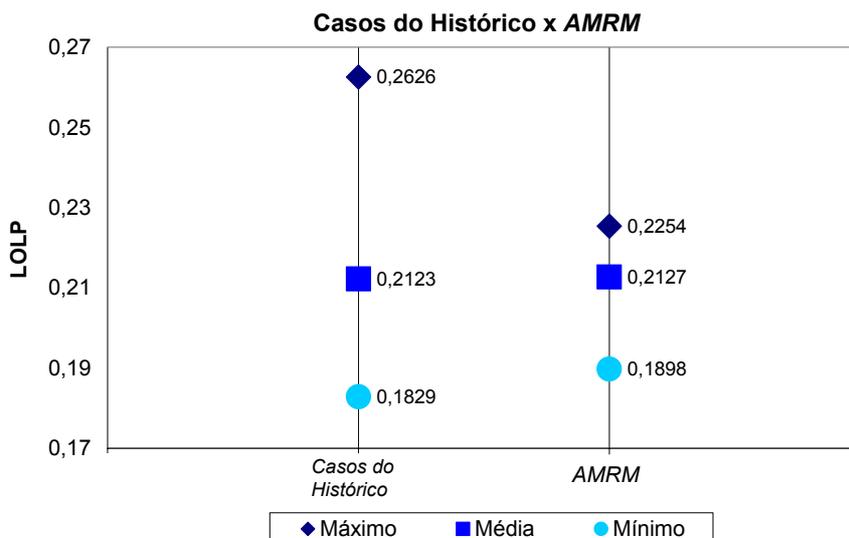


Figura 6.13 – Gráfico de comparação das estatísticas da LOLP - casos do histórico x AMRM.

Por outro lado, pode ser observado que a dispersão dos resultados (distância entre o máximo e o mínimo) não foi à esperada. Neste caso esta pouca dispersão pode estar ligada a falhas na determinação do melhor valor do coeficiente de nebulização, uma vez que esta determinação foi realizada simplesmente pela determinação do valor máximo obtido para a *ICC* para cada um dos valores de m , sem a prévia avaliação da acuidade da métrica.

O tipo de falha pelo qual este critério de seleção do melhor valor do coeficiente de nebulização passa pode ser exemplificado pela análise realizada para a usina de Mascarenhas no item 4.2. Ao analisar os gráficos sob o ponto de vista do ponto de corte c_{max} igual a 10 agrupamentos, seria obtidos os resultados listados na Tabela 6.17. Como pode ser verificado o critério adotado indica que o melhor valor de m é igual a 1,25, mas a análise dos gráficos apresentados no item 4.2 indica que a medida *ICC* só apresenta bons resultados para m igual a 2,00.

A constatação da existência desta falha não invalida a estruturação do processo somente indica que o critério de seleção é que deve ser mais analisado.

A principal diferença entre este método e o *AMR* está no fato de que no *AMRM*, a *ICC* também foi utilizada para determinar o coeficiente de nebulização para cada espaço de dados, ou seja, foi dado mais um grau de liberdade para criação dos agrupamentos na etapa de tratamento da dependência temporal.

Deve ser lembrado que no *AMR*, a *ICC* é utilizada somente para determinar o melhor número de agrupamentos para cada um dos espaços de dados.

Tabela 6.17 – Resultados para a usinas de Mascarenhas

| <i>m</i> | Número de Agrupamentos | <i>ICC</i> |
|-------------|------------------------|------------|
| 1,25 | 10 | 0,0342 |
| 1,50 | 7 | 0,0267 |
| 1,75 | 4 | 0,0145 |
| 2,00 | 4 | 0,096 |

7 CONCLUSÕES

Este item tem por objetivo reunir e discutir as conclusões apresentadas ao longo desta dissertação, para tanto estas serão compiladas conforme a sua seqüência de apresentação.

Os primeiros resultados obtidos foram discutidos no item 4.3 e estavam relacionados aos testes de validação da medida *ICC* para dados multidimensionais.

Para o primeiro conjunto de testes realizado, foi verificado que a resposta obtida da *ICC* quando o coeficiente de nebulização utilizado no FCM foi igual a 1,25 indicava que a medida apresentava problemas na definição de um máximo no intervalo em que foi calculada, visto que o valor da *ICC* foi crescente em todo o intervalo de agrupamentos testado.

Porém, para os demais valores do coeficiente de nebulização, foi verificado que esta característica não se manteve, ou seja, quando o FCM foi utilizado com os outros valores do coeficiente de nebulização houve a determinação de um número de agrupamentos onde ocorreu um valor máximo para *ICC*.

Este tipo de comportamento, onde os resultados da *ICC* quando o FCM é executado com determinados valores de coeficiente de nebulização não permite a determinação de um máximo da medida enquanto que, quando os outros valores do coeficiente são utilizados o máximo é determinado, também foi verificado no segundo conjunto de testes, o que intensificou a sinalização de que a medida *ICC* pode ser também utilizada para a determinação do melhor valor do coeficiente de nebulização para o espaço de dados, sendo para isso necessário executar o algoritmo de agrupamento para diferentes valores do coeficiente de nebulização.

O principal resultado obtido a partir dos testes foi à verificação de que a *ICC* pode ser utilizada, não somente para determinar o melhor número de agrupamentos para um espaço de dados, mas também para determinar o melhor valor do coeficiente de nebulização a ser utilizado no FCM.

No Capítulo 5, foram discutidos dois casos, o primeiro aplicado sobre os dados do sistema de teste de confiabilidade IEEE-RTS e o segundo aplicado sobre os dados de uma configuração real do subsistema Sudeste/Centro-Oeste. Nestes dois casos, foi avaliada e validada a necessidade de se realizar a análise de confiabilidade para sistemas elétricos com predominância de geração hidrelétrica, tendo sido obtidos os resultados dos *casos do histórico* (casos do NH2 formados a partir das séries de potência disponível obtidas dos modelos

energéticos utilizando as séries históricas de vazão), que foram utilizados como base de comparação para os processos de agrupamentos definidos no Capítulo 6.

A principal conclusão do Capítulo 5 foi que a variação da hidrologia influencia a potência disponível das usinas, devendo ser considerada durante a realização de estudos de confiabilidade composta em sistemas predominantemente hidrelétricos. Esta conclusão deve-se ao fato de que a potência disponível é, em geral, inferior à capacidade nominal de geração das usinas, o que torna necessário que o cálculo dos índices de confiabilidade sejam calculados de forma aderente à condição hidrológica do sistema.

Também foi possível verificar que este é um processo extremamente custoso em termos de processamento, pois o número de casos que devem ser analisados é muito grande e a solução de cada um destes casos envolve a solução de problemas complexos, tais como, o fluxo de potência ótimo de um sistema elétrico de grande porte.

Da forma como foi tratada nesta dissertação, o tempo de processamento, da avaliação de confiabilidade composta baseada nas séries de potência disponível do sistema Sudeste/Centro-Oeste, foi de aproximadamente 300hs.

No Capítulo 6 foram apresentados 3 processos para efetuar o agrupamento das séries de potência disponível: o *Agrupamento de Matriz Única (AMU)*, o *Agrupamento por Mapeamento de Resultados (AMR)* e o *Agrupamento por Mapeamento de Resultados com Seleção do Coeficiente de Nebulização (AMRM)*.

Em todos os 3 processos foi verificado que a aplicação de técnicas de agrupamento sobre as séries de potências disponível calculadas pelos modelos energéticos, permite manter a acuidade em relação a média da LOLP obtida para os *casos do histórico*, sendo que após a utilização dos processos o número de estudos efetuando foi menor. Sendo que o *AMU*, indicou que seria necessário realizar 17 casos, o *AMR* indicou 14 casos e o *AMRM* indicou 7 casos.

O primeiro processo tem por característica impor que todas as usinas apresentem o mesmo padrão de agrupamento, o mesmo número de centros para todos os espaços de dados representativos das diferentes usinas.

Por sua vez o segundo processo passou a permitir que cada usina fosse representada por um número de agrupamentos distinto, respeitando assim as características dos espaços de dados de cada uma das usinas.

Finalmente, no terceiro processo foi acrescentada a escolha do melhor valor para o coeficiente de nebulização a ser utilizado para cada um dos espaços de dados.

Os resultados observados para o *Agrupamento por Mapeamento de Resultados com Seleção do Coeficiente de Nebulização* apresentaram pouca dispersão que pode estar ligada a falhas na determinação do melhor valor do coeficiente de nebulização, uma vez que esta determinação foi realizada simplesmente pela determinação do valor máximo obtido para a ICC para cada um dos valores de m , sem a prévia avaliação da acuidade da métrica.

A constatação da existência desta falha não invalida a estruturação do processo somente indica que o critério de seleção deve ser melhor analisado de forma a não permitir a seleção de coeficientes de nebulização em que a medida ICC não apresenta bons resultados.

O tempo de processamento dos 3 processos de agrupamento se mostrou bastante compatível sendo no computo geral, ou seja, a soma do tempo de processamento de cada um dos processos com o tempo de processamento necessário para a execução dos casos do NH2 resultantes de cada um, o *Agrupamento por Mapeamento de Resultados com Seleção do Coeficiente de Nebulização* foi o que apresentou o menor tempo de processamento pois definiu um número menor de casos.

7.1 SUGESTÕES PARA TRABALHOS FUTUROS

A seguir são indicadas algumas linhas de pesquisa que poderiam ser realizadas no intuito de ampliar os resultados do presente trabalho.

- Testar os processos com a utilização de resultados obtidos dos modelos energéticos quando estes utilizam séries sintéticas de vazão. A utilização deste tipo de séries pode aumentar o número de amostras existentes no espaço de dados para 2000 amostras.
- A utilização dos resultados obtidos dos modelos energéticos quando estes utilizam séries sintéticas de vazão permitiriam utilizar outros algoritmos de agrupamento em substituição ao FCM, tais como Gustafson-Kessel (GK) ou Gath-Geva (GG), que exigem um espaço de dados maior, no processamento da etapa que trata a dependência temporal dos dados do *Agrupamento por Mapeamento de Resultados* e no *Agrupamento por Mapeamento de Resultados com Seleção do Coeficiente de Nebulização*.
- Melhorar o algoritmo de distribuição da potência disponível, calculada pelo modelo SUIISHI-O, na formação dos casos do NH2 de forma utilizar a potência mínima das usinas nos dados, retirando assim a simplificação adotada neste trabalho.

- Na etapa de tratamento da dependência espacial do *Agrupamento por Mapeamento de Resultados* e do *Agrupamento por Mapeamento de Resultados com Seleção do Coeficiente de Nebulização*, utilizar outras medidas de dissimilaridade para dados nominais.
- Ainda com respeito à etapa citada no item anterior, verificar a possibilidade de utilizar outros algoritmos de agrupamento capazes de tratar dados nominais, sendo bastante interessante a utilização de um algoritmo nebuloso, uma vez que o mapa resultante da primeira etapa também possui a característica de grande sobreposição nos dados.
- Com relação a medida *ICC*, deve ser ampliado o escopo dos testes realizados, para verificação mais precisa de como esta medida pode ser utilizada não somente com a finalidade de determinar o melhor número de agrupamentos para um determinado espaço de dados, mas também para determinar o melhor valor de coeficiente de nebulização a ser utilizado.

REFERÊNCIAS BIBLIOGRÁFICAS

- AMADO, S.M., KLIGERMAN, A.S., MACEIRA, M.E.P., RAYMOND, M.P., ROBITAILLE, A., **Making Hydro More Competitive Through Integrated Technologies**. Hydro 2000, Bern, Switzerland, October, 2000.p. 375-383.
- ANEEL, Agência Nacional de Energia Elétrica, BIG – Banco de Dados de Informação de Geração. Disponível em: www.aneel.gov.br/15.htm. Acesso em: 06/09/2004.
- BARALDI, A., BLONDA, P., **A Survey of Fuzzy Clustering Algorithms for Pattern Recognition - Part I**. IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics, Vol. 29, No. 6, December 1999, p. 778-785.
- BARROSO, L. P., ARTES, R., **Análise Multivariada**, 48ª Reunião da RBRAS e 10º SEAGRO, Julho de 2003.
- BEZDEK, J. C., EHRLICH, R., FULL, W., **FCM: Fuzzy C-Means Algorithm**, Computer and Geoscience, Vol. 10, 1984, p. 191-203.
- BILLINGTON, R., ALLAN, R. N., **Reliability Evaluation of Power Systems**, Pitman Advanced Publishing Program, 1984.
- DAVÉ, R.N., KRISHNAPURAM, R., **Robust Clustering Methods: A Unified View**. IEEE Transactions on Fuzzy Systems, Vol. 5, No. 2, May 1997, p. 270-293.
- DOWNING, D., COVINGTON, M., **Dictionary of Computer Terms**. Nove York: Barron's Business Guide, Terceira Edição, 1992. 364 p.
- FORTUNATO, L.A.M., ARARIPE NETO, T.A., ALBUQUERQUE, J.C.R., PEREIRA, M.V.F., **Introdução ao Planejamento da Expansão e Operação de Sistemas de Produção de Energia Elétrica**. Niterói: Universidade Federal Fluminense, EDUFF, 1990. 232 p.
- FRANCO, C.R., **Categorização Nebulosa e Validação de Categorias Aplicadas a Problemas de Reconhecimento de Padrões**. Dissertação (Mestrado) – Universidade Federal do Rio de Janeiro, IM-NCE, 2002.
- GRIGG, C., *et al*, **The IEEE Reliability Test System - 1996, A report prepared by the Reliability Test System Task Force of the Application of Probability Methods Subcommittee**. IEEE Transactions on Power Systems, Vol. 14, No. 3, August 1999.

- GOWDA, K.C., DIDAY, E., **Symbolic Clustering Using a New Similarity Measure**. IEEE Transactions on Systems, Man, and Cybernetics, Vol. 22, No. 2, March/April 1992, p. 368-378.
- GUIANET, O Guia do Brasil, **Mapa Energético do Brasil**. Disponível em: <http://www.guianet.com.br/brasil/mapaenergia.htm>. Acesso em: 01/07/2004.
- GUSTAFSON, D. E., KESSEL, W. C., **Fuzzy Clustering With a Fuzzy Covariance Matrix**. In: IEEE Conference on Decision and Control, Jan. 1979. Proceedings. P.761-766.
- HALKIDI, M., BATISTAKIS, Y., VARGIANNIS, M., **On Clustering Validation Techniques**. Journal of Intelligent Information Systems, Kluwer Academic Publishers, vol. 17, 2/3, Holanda, 2001a, p. 107-145.
- HALKIDI, M., , VARGIANNIS, M., **Clustering Validity Assessment: Finding the optimal partitioning of a data set**. In: IEEE International Conference on Data Mining (ICDM 2001) Proceedings, San Jose, California, USA, 2001b, p. 187-194.
- HÄRDLE, W., MORI, Y., VIEU, P., **Statistical Case Studies**. MD Tech, 2004. SOFYAN, H., cap 19: **Fuzzy Clustering**, March 20, 2003. Disponível em: <http://www.quantlet.com/mdstat/scripts/xcs/pdf/xcpspdf.pdf>. Acesso em: 22/10/2004.
- HÖPPNER, F., KLAWONN, F., KRUSE, R., RUNKLER, T., **Fuzzy Cluster Analysis – Methods for Classification, Data Analysis and Image Recognition**. Chichester: John Wiley & Sons, LTD, 1999, 289 p.
- JAIN, A. K., MURTY, M. N., FLYN, P. J., **Data Clustering: A Review**. ACM Computing Surveys, Vol. 31, p. 264 - 323, 1999.
- JARDIM, D.L.D.D., **Modelo de Geração de Séries Sintéticas de Vazões Utilizando Técnicas de Agregação**, 2002, 83p. Dissertação (Mestrado), COPPE/UFRJ, Rio de Janeiro.
- KAUFFMAN, L., ROUSSEEUW, P. J., **Finding Groups in Data: An Introduction to Cluster Analysis**. New York: Willey Series, LTD, 1990. 368 p.
- KLAWONN, F., HÖPPNER, F., **What is Fuzzy About Fuzzy Clustering? Understanding and Improving the Concept of the Fuzzifier**. Advances in Intelligent Data Analysis, p. 254-264, 2003.
- LÍCIO, R.C., MELO, A.C.G., ARAÚJO, J.L., **Análise de Sensibilidade de Índices de Confiabilidade Multi-Área com Respeito às Taxas de Falha e Reparo dos Equipamentos**. XIV SNPTEE, Belém, Pará, Brasil, 1997.

- MACEIRA, M.E., MERCIO, C.B., GORENSTIN, B., CUNHA, S.H., SUANNO, C., SACRAMENTO, M.C., KLIGERMAN, A., **Aplicação do Modelo NEWAVE na Avaliação Energética da Interligação dos Sistemas Norte/Nordeste e Sul/Sudeste Brasileiros**. SEPOPE, Salvador, Bahia, Brasil, 1998.
- MACEIRA, M. E. P., MERCIO, C. M. V. B., **Stochastic Sreamflow Model for Hydroelectric Systems**, 5th International Conference PMAPS - Probabilistic Methods Applied to Power Systems, Vancouver, Canada, 1997.
- MACEIRA, M. E., SUANNO, C., COSTA, J., KLIGERMAN, A., GORENSTIN, B., PERREIRA, M., **Uso do Modelo Estratégico de Geração Hidrotérmica a Subsistemas Equivalentes – NEWAVE – Na Avaliação da representação Explícita do Intercâmbio**. XIII SNPTEE, Florianópolis, Santa Catarina, Brasil, 1995.
- MACEIRA, M.E., TERRY, L.A., COSTA, F.S., DAMÁZIO, J.M., MELO, A.C.G., **Chain of Optimization Models for Setting the Energy Dispatch and Spot Price in the Brazilian System**. 14° PSCC, Sevilha, Espanha, 2002.
- MARCATO, A.L.M., **Avaliação dos Requisitos e Comercialização da Reserva de Potência em Sistemas Competitivos**. Dissertação (Mestrado), Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, PUC-Rio, 1998.
- MARQUES DE AZEVEDO FILHO, J., **Imperativos da Centralização e Coordenação da Operação Energética no Âmbito da Reforma Institucional do Setor Elétrico Brasileiro**. Dissertação (Mestrado), Departamento Planejamento Energético, COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2000.
- MELLO, J.C.O., **Determinação do Valor da Confiabilidade em Sistemas Compostos de Geração e Transmissão**. Dissertação (Doutorado), Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, PUC-Rio, 1994.
- MELO, A.C.G., **Avaliação dos Índices de Frequência e Duração no Cálculo da Confiabilidade Composta de Sistemas de Sistemas de Geração e Transmissão de Grande Porte**. Dissertação (Doutorado), Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, PUC-Rio, 1990.
- MELO, A.C.G., *et al.* **Probabilistic Adequacy Evaluation of Large Scale Power Systems - A Brazilian Case**. IEEE Stockholm Power Tech, Stockholm, Sweden, 1995.

- MELO, A.C.G., *et al.* **Avaliação Probabilística na Análise de Desempenho de Sistemas de Potência – Novos Desenvolvimento no Programa NH2.** XV SNPTEE, Foz do Iguaçu, 1999.
- MEYER, P. L., **Probabilidade, Aplicações à Estatística.** Rio de Janeiro: Editora LTC - Ao Livro Técnico e Científico Ltda. Segunda Edição, 2000. 426 p.
- NAGPAUL, P.S., *Guide to Advanced Data Analysis, Using IDANS Software.* New Delhi, India. Disponível em: www.unesco.org/webworld/idams/advguide/Chapt7-1-1.htm. Acesso em : 22/10/2004
- OLIVEIRA, S. C., **Uma Avaliação das Medidas de Validação ICC e CS nos Processos de Clusterização de Dados.** (não publicado) Pós-graduação em Informática NCE/IM, Dezembro de 2002.
- ONS, Operador Nacional do Sistema, *Planejamento Anual da Operação Energética*, 2002a. Disponível em: http://www.ons.org.br/ons/educacional/index_mapas.htm. Acesso em: 22/10/2004
- ONS, Operador Nacional do Sistema, *Procedimentos de Rede, Submódulo 18.2 - Modelos Computacionais*, 2002b. Disponível em: <http://www.ons.org.br/ons/procedimentos/index.htm>. Acesso em: 22/10/2004
- PAKHIRA, M. K., BANDYOPADHYAY, S., MAULIK, U., **Validity Index for Crisp and Fuzzy Clusters.** Pattern Recognition, Vol. 37, p. 487-501, 2004.
- PEREIRA, M.V.F., BALU, N.J. **Composite Generation and Transmission Reliability Evaluation.** In: Proceedings of the IEEE, vol. 80, no. 4, April, 1992.
- PINHEIRO, J. M. S., *et al.*, **Probing the New IEEE Reliability Test System (RTS-96): HL II Assessment.** In: IEEE Transactions on Power Systems, vol. 13, no. 1, February, 1998, p.171.
- RODRIGUES, M.A.M, *et al.*, **Sistema de Encadeamento de Modelos Energéticos.** In: XVI SNPTEE, Grupo VII, Planejamento de Sistemas Elétricos, Outubro, 2001.
- SILVA, A. M. L., MELO, A. C. G., CUNHA, S. H. F., **Frequency and Duration Method for Reliability Evaluation of Large-Scale Hydrothermal Generating Systems.** IEE Proceedings-C, Vol. 138, No. 1, January 1991.
- SUCESU-SC, Sociedade dos Usuários de Informática e Telecomunicações de Santa Catarina. *Dicionário.* Disponível em: http://www.sc.sucesu.org.br/images_cli/glossario.htm. Acesso em: 22/10/2004.

- VAN DER LAAN, M.J., POLLARD, K.S., Brian, J., **A New Partitioning Around Medoids Algorithm**. U. C. Berkeley Division of Biostatistics Working Paper Series, 2002. Disponível em <http://www.bepress.com/ucbbiostat/paper105>. Acesso em: 22/10/2004.
- VIVARELLI, F., WILLIAMS, C. K., **Discovering Hidden Features With Gaussian Process Regression**. In: Advances in Neural Information Processing Systems 11, eds. M. J. Kearns, S. A. Solla and D. A. Cohn. MIT Press, 1999.

GLOSSÁRIO

Ano seco – Ano baseado em critérios estatísticos, em que o curso de água tem afluições inferiores a média.

Ano úmido – Ano baseado em critérios estatísticos, em que o curso de água tem afluições superiores a média.

Bacias hidrográficas – Parte da superfície terrestre que contribui para alimentar um rio ou lago.

Carga – Potência demandada por um consumidor ou unidade consumidora, classificada de acordo com certas características de energia elétrica.

Cenário – Modelo científico que permite aos pesquisadores considerar elementos de um sistema como se realmente funcionassem da maneira descrita, não testando as hipóteses, mas permitindo o exame dos possíveis resultados, caso as hipóteses fossem verdadeiras.

Confiabilidade – Probabilidade de um sistema ou componente realizar suas funções previstas de forma contínua, adequada e segura, por um período de tempo preestabelecido, sob condições operativas predefinidas.

Consumidor – Pessoa física ou jurídica que ajusta, com o concessionário, autorizado ou permissionário, o fornecimento de energia elétrica.

Contingência – Perda de equipamentos principais ou linhas de transmissão, que provoca ou não violação dos limites operativos ou corte de carga.

Contingência múltipla – Perda de mais de um elemento do sistema elétrico, que altera as condições operativas do sistema.

Contingência simples – Perda de um único elemento do sistema elétrico, que altera as condições operativas do sistema.

Energia – Propriedade de um sistema que lhe permite realizar trabalho. Em eletricidade, mede-se em Watt-hora (Wh) ou seus múltiplos, KWh, MWh, GWh, TWh.

Energia Armazenada – Valoração energética do volume armazenado em um reservatório, para cujo cálculo se considera a produtividade do reservatório e das demais usinas hidrelétricas a jusante, descontando-se o volume morto.

Equipamento – Unidade funcional, completa e distinta que exerce uma ou mais funções elétricas relacionadas com geração, transmissão, conversão, distribuição ou utilização de energia elétrica.

Estudos Elétricos – Análise e descrição do comportamento do sistema elétrico visando a identificar seu desempenho diante de quaisquer tipos de condições operativas e propor soluções para garantir a qualidade do suprimento aos consumidores.

Falhas – Efeito ou consequência de uma ocorrência acidental em uma instalação ou equipamento que acarreta sua indisponibilidade operativa em condições não programadas, impedindo-o de funcionar, e, portanto, de desempenhar suas funções em caráter permanente ou em caráter temporário.

Geração – Conversão de uma forma qualquer de energia em energia elétrica.

Hidrologia – (1) Ciência que trata das águas da terra, sua ocorrência, circulação e distribuição, suas propriedades químicas e físicas e sua reação com meio ambiente, incluindo sua relação com os seres vivos. (2) Ciência que estuda as variações dos recursos hídricos naturais da terra em função das diferentes fases do ciclo hidrológico. (3) Interação de um composto com a água, chegando à sua decomposição em constituintes ácidos e básicos. Reação inversa à formação de sais.

Intercâmbio – Fluxo de energia elétrica entre os sistemas elétricos de dois ou mais concessionários.

Jusante (a jusante) – Localização inferior, ou seja, em cotas mais baixas. No caso de águas correntes (rios, córregos, arroios) são os pontos situados em direção à sua foz. Na direção da corrente, rio abaixo.

Linha de Transmissão – Componente do sistema elétrico destinado à transmissão de energia elétrica, que interliga duas subestações.

Montante (a montante) – Localização superior, ou seja, em cotas mais elevadas. No caso de águas correntes (rios, córregos, arroios), são os pontos situados em águas correntes (rios, córregos, arroios), são os pontos situados em direção à nascente. Na direção oposta à corrente, rio acima.

Planejamento da operação – Processo cujo objeto é a análise das condições futuras de atendimento ao mercado consumidor, com base no conhecimento específico requerido e na natureza das variáveis analisadas. Para tal processo, elaboram-se estudos especiais, analisa-se a proteção e o controle do SIN, bem como o desenvolvimento das atividades

de hidrologia operacional. O planejamento da operação compreende a análise energética, elétrica e hidrológica da operação futura em diferentes horizontes - plurianual, anual, mensal, semanal e diário.

Potência – Energia produzida ou consumida por unidade de tempo, num sistema gerador ou absorvedor de energia. Em eletricidade, a potência é medida em Watt ou seus múltiplos (KW, MW).

Potência disponível – Máxima potência injetada pela usina no sistema, calculada pela potência nominal instalada menos o consumo próprio e os fornecimentos feitos diretamente de sua subestação ou através de instalações de uso exclusivo de consumidores.

Potência nominal – Valor da potência de placa de um equipamento. Para unidades geradoras prevalece a potência homologada pela ANEEL.

Premissa – Proposição considerada e aceita como base referencial em estudos e planos.

Programa Mensal de Operação (PMO) – Documento que estabelece metas de geração de energia elétrica e as diretrizes elétricas e energéticas a serem alcançadas no mês em questão.

Qualidade – Desempenho do sistema em termos de continuidade de serviço e conformidade na forma de onda da tensão, incluindo os efeitos de flutuação de tensão, desequilíbrio, distorção harmônica e variação de tensão de curta duração.

Reservatórios – Recipiente delimitado por certa área da superfície terrestre e por uma ou mais barragens, no qual se armazena determinada quantidade de água.

Risco hidrológico – Risco decorrente da aleatoriedade das afluições.

Série Histórica – Conjunto de valores ordenados cronologicamente em intervalos constantes, representativos de um parâmetro ou grandeza física.

Série sintética de vazões – Série de vazões num local obtida por processos matemáticos. São muito utilizadas as séries geradas por modelos estocásticos.

Sistema elétrico – Conjunto de instalações para geração, transmissão e/ou distribuição de energia elétrica.

Sistema equivalente – Conjunto de instalações para geração, transmissão e/ou distribuição de energia elétrica.

Sistema interligado – Os sistemas de geração, transmissão e distribuição de propriedade das diversas empresas, com uso compartilhado por essas empresas, por onde transitam energias de diversas fontes e destinos, sujeitos aos Procedimentos de Rede do ONS.

Subsistema – Parte de um sistema elétrico.

Transmissão – Transporte de energia elétrica entre subestações.

Usina hidrelétrica – Usina elétrica na qual a energia elétrica é obtida por conversão da energia gravitacional da água.

Usina termelétrica – Usina elétrica na qual a energia elétrica é obtida por conversão de energia térmica.

Volume armazenado – Volume de água armazenado em um reservatório, geralmente expresso em km^3 ou m^3 .

APÊNDICE

A - USINAS QUE NÃO PARTICIPAM DO PROCESSO DE AGRUPAMENTO

Neste apêndice são apresentados os gráficos da potência disponível normalizada pelo valor máximo das usinas que não participam do processo de agrupamento por apresentar a média dos coeficientes de variação menor ou igual a 1%.

Neste conjunto de usinas existem um grupo, cujos nomes e valor de potência disponível são apresentados na Tabela A. 1, que apresentam como característica comum tem o a média do coeficiente de variação igual a zero. Os gráficos dos dados normalizados destas usinas todas são representadas pelo mesmo gráfico, apresentado na Figura A. 1.

Os gráficos das outras usinas, ou seja, das usinas apresentam média do coeficiente de variação maior que zero e menor ou igual a 1%, são apresentados na Tabela A. 1, assim como seus dados de potência disponível máxima e mínima, em ordem crescente de coeficiente de variação e, caso existam mais de uma usina dentro com o mesmo valor de coeficiente de variação, esta são apresentadas em ordem alfabética.

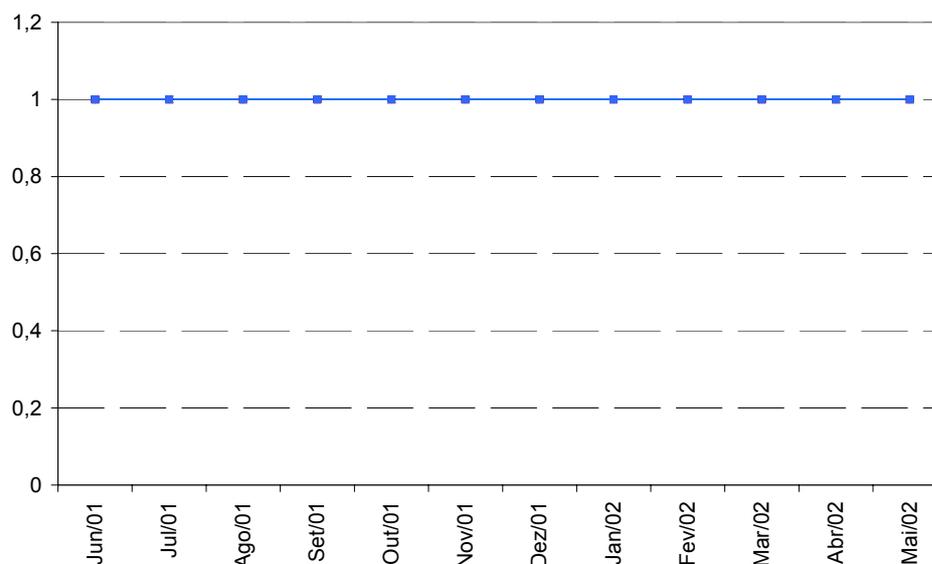


Figura A. 1 – Gráfico das usinas com coeficiente de variação médio igual a zero

Tabela A. 1 – Usinas com coeficiente de variação igual a zero

| Número da Usina | Nome da Usina | Potência Disponível | Número da Usina | Nome da Usina | Potência Disponível |
|-----------------|----------------|---------------------|-----------------|----------------|---------------------|
| 10 | Igarapava | 210,0 | 119 | Henry Borden | 837,3 |
| 12 | Porto Colômbia | 328,0 | 127 | Sobragi | 54,5 |
| 16 | A. S. Oliveira | 32,0 | 131 | Nilo Pecanha | 378,7 |
| 42 | N. Avanhandava | 347,4 | 132 | Fontes | 132,0 |
| 52 | Canoas I | 82,5 | 192 | Guilman Amorim | 138,9 |
| 63 | Rosana | 372,0 | 193 | Sá Carvalho | 77,7 |
| 66 | Itaipú | 12586,7 | | | |

Tabela A. 2 – Usinas com coeficiente de variação médio maior que zero e menor que um

| Dados da Usina | | Gráfico da Potência Disponível Normalizada |
|----------------------|-------------------------------|--|
| Usina: | 32 – Cachoeira Dourada | |
| Máximo: | 658,0 MW | |
| Mínimo: | 651,0 MW | |
| Máximo do CV: | 0,13 | |
| Média do CV: | 0,01 | |
| Usina: | 8 – Estreito | |
| Máximo: | 1104,0 MW | |
| Mínimo: | 1094,0 MW | |
| Máximo do CV: | 0,11 | |
| Média do CV: | 0,02 | |

| Dados da Usina | | Gráfico da Potência Disponível Normalizada |
|----------------------|---------------------------|--|
| Usina: | 44 – Ilha Solteira | |
| Máximo: | 4251,5 MW | |
| Mínimo: | 4200,0 MW | |
| Máximo do CV: | 0,15 | |
| Média do CV: | 0,02 | |
| Usina: | 50 – L. N. Garcez | |
| Máximo: | 72,0 MW | |
| Mínimo: | 68,0 MW | |
| Máximo do CV: | 0,62 | |
| Média do CV: | 0,05 | |
| Usina: | 62 – Taquaruçú | |
| Máximo: | 555,0 MW | |
| Mínimo: | 524,0 MW | |
| Máximo do CV: | 0,69 | |
| Média do CV: | 0,06 | |
| Usina: | 11 – Volta Grande | |
| Máximo: | 380,0 MW | |
| Mínimo: | 368,0 MW | |
| Máximo do CV: | 0,49 | |
| Média do CV: | 0,10 | |

| Dados da Usina | | Gráfico da Potência Disponível Normalizada |
|----------------|------------------------------|--|
| Usina: | 9 – Jaguará | |
| Máximo: | 424,0MW | |
| Mínimo: | 382,0MW | |
| Máximo do CV: | 1,60 | |
| Média do CV: | 0,36 | |
| Usina: | 130 – Ilha dos Pombos | |
| Máximo: | 183,0 MW | |
| Mínimo: | 160,0 MW | |
| Máximo do CV: | 1,55 | |
| Média do CV: | 0,42 | |
| Usina: | 38 – A. S. Lima | |
| Máximo: | 144,0 MW | |
| Mínimo: | 124,0 MW | |
| Máximo do CV: | 2,08 | |
| Média do CV: | 0,49 | |
| Usina: | 26 – Miranda | |
| Máximo: | 408,0 MW | |
| Mínimo: | 403,0 MW | |
| Máximo do CV: | 1,08 | |
| Média do CV: | 0,70 | |

| Dados da Usina | | Gráfico da Potência Disponível Normalizada |
|----------------------|-----------------------------|--|
| Usina: | 133 – Pereira Passos | |
| Máximo: | 100,0 MW | |
| Mínimo: | 93,0 MW | |
| Máximo do CV: | 2,21 | |
| Média do CV: | 0,77 | |
| Usina: | 134 – Salto Grande | |
| Máximo: | 102,0 MW | |
| Mínimo: | 98,0 MW | |
| Máximo do CV: | 1,65 | |
| Média do CV: | 0,86 | |

B - USINAS QUE PARTICIPAM DO PROCESSO DE AGRUPAMENTO

Na Tabela B. 1 são apresentados os gráficos da potência disponível normalizada pelo valor máximo das usinas que participam do processo de agrupamento, ou seja, que tem a média dos coeficientes de variação maior que 1%.

Os gráficos são apresentados na mesma ordem em que as usinas são tratadas pelo processo de agrupamento, ou seja, em ordem crescente do número da usina utilizado nos dados do SUSHI-O.

Tabela B. 1 – Usinas que participam do processo de agrupamento

| Dados da Usina | | Gráfico da Potência Disponível Normalizada |
|----------------------|--------------------------------|--|
| Usina: | 1 – Camargos e Itutinga | |
| Máximo: | 100 MW | |
| Mínimo: | 61 MW | |
| Máximo do CV: | 12,99 | |
| Média do CV: | 9,50 | |
| Usina: | 6 – Furnas | |
| Máximo: | 1312 MW | |
| Mínimo: | 899 MW | |
| Máximo do CV: | 10,30 | |
| Média do CV: | 7,41 | |

| Dados da Usina | | Gráfico da Potência Disponível Normalizada |
|----------------------|----------------------------------|--|
| Usina: | 7 – Marcarenhas de Moares | |
| Máximo: | 478 MW | |
| Mínimo: | 185 MW | |
| Máximo do CV: | 23,85 | |
| Média do CV: | 17,36 | |
| Usina: | 14 – Caconde | |
| Máximo: | 80 MW | |
| Mínimo: | 31 MW | |
| Máximo do CV: | 22,20 | |
| Média do CV: | 15,56 | |
| Usina: | 15 – Euclides da Cunha | |
| Máximo: | 108 MW | |
| Mínimo: | 93 MW | |
| Máximo do CV: | 3,21 | |
| Média do CV: | 1,66 | |
| Usina: | 17 – Marimondo | |
| Máximo: | 1488 MW | |
| Mínimo: | 607 MW | |
| Máximo do CV: | 23,14 | |
| Média do CV: | 15,97 | |

| Dados da Usina | | Gráfico da Potência Disponível Normalizada |
|----------------------|---------------------------|--|
| Usina: | 18 – Água Vermelha | |
| Máximo: | 1396 MW | |
| Mínimo: | 1165 MW | |
| Máximo do CV: | 4,82 | |
| Média do CV: | 3,34 | |
| Usina: | 24 – Emborcação | |
| Máximo: | 1192 MW | |
| Mínimo: | 464 MW | |
| Máximo do CV: | 24,00 | |
| Média do CV: | 16,91 | |
| Usina: | 25 – Nova Ponte | |
| Máximo: | 510 MW | |
| Mínimo: | 332 MW | |
| Máximo do CV: | 14,19 | |
| Média do CV: | 8,65 | |
| Usina: | 30 – Corumbá I | |
| Máximo: | 375 MW | |
| Mínimo: | 153 MW | |
| Máximo do CV: | 23,26 | |
| Média do CV: | 14,24 | |

| Dados da Usina | | Gráfico da Potência Disponível Normalizada |
|----------------------|--------------------------|--|
| Usina: | 31 – Itumbiara | |
| Máximo: | 2280 MW | |
| Mínimo: | 992 MW | |
| Máximo do CV: | 20,60 | |
| Média do CV: | 14,52 | |
| Usina: | 33 – São Simão | |
| Máximo: | 1710 MW | |
| Mínimo: | 1407 MW | |
| Máximo do CV: | 4,72 | |
| Média do CV: | 3,27 | |
| Usina: | 37 – Barra Bonita | |
| Máximo: | 140 MW | |
| Mínimo: | 31 MW | |
| Máximo do CV: | 36,51 | |
| Média do CV: | 25,39 | |
| Usina: | 39 – Ibitinga | |
| Máximo: | 132 MW | |
| Mínimo: | 115 MW | |
| Máximo do CV: | 5,39 | |
| Média do CV: | 2,38 | |

| Dados da Usina | | Gráfico da Potência Disponível Normalizada |
|----------------------|-----------------------------|--|
| Usina: | 40 – Promissão | |
| Máximo: | 264 MW | |
| Mínimo: | 224 MW | |
| Máximo do CV: | 5,12 | |
| Média do CV: | 3,36 | |
| Usina: | 45 – Jupia | |
| Máximo: | 1551 MW | |
| Mínimo: | 861 MW | |
| Máximo do CV: | 6,44 | |
| Média do CV: | 1,95 | |
| Usina: | 46 – Porto Primavera | |
| Máximo: | 1210 MW | |
| Mínimo: | 684 MW | |
| Máximo do CV: | 6,08 | |
| Média do CV: | 2,05 | |
| Usina: | 47 – A. A. Laydner | |
| Máximo: | 98 MW | |
| Mínimo: | 78 MW | |
| Máximo do CV: | 7,89 | |
| Média do CV: | 5,75 | |

| Dados da Usina | | Gráfico da Potência Disponível Normalizada |
|----------------|-----------------------|--|
| Usina: | 49 – Xavantes | |
| Máximo: | 416 MW | |
| Mínimo: | 317 MW | |
| Máximo do CV: | 8,45 | |
| Média do CV: | 5,90 | |
| Usina: | 51 – Canoas II | |
| Máximo: | 72 MW | |
| Mínimo: | 54 MW | |
| Máximo do CV: | 4,65 | |
| Média do CV: | 1,65 | |
| Usina: | 61 – Capivara | |
| Máximo: | 640MW | |
| Mínimo: | 371MW | |
| Máximo do CV: | 21,15 | |
| Média do CV: | 15,28 | |
| Usina: | 120 – Jaguari | |
| Máximo: | 28MW | |
| Mínimo: | 23MW | |
| Máximo do CV: | 5,13 | |
| Média do CV: | 3,08 | |

| Dados da Usina | | Gráfico da Potência Disponível Normalizada |
|----------------|---------------------------|--|
| Usina: | 121 – Paraibuna | |
| Máximo: | 85 MW | |
| Mínimo: | 61 MW | |
| Máximo do CV: | 7,94 | |
| Média do CV: | 5,79 | |
| Usina: | 122 – Santa Branca | |
| Máximo: | 45 MW | |
| Mínimo: | 10 MW | |
| Máximo do CV: | 36,51 | |
| Média do CV: | 25,86 | |
| Usina: | 123 – Funil | |
| Máximo: | 198 MW | |
| Mínimo: | 60 MW | |
| Máximo do CV: | 27,25 | |
| Média do CV: | 20,35 | |
| Usina: | 144 – Mascarenhas | |
| Máximo: | 131 MW | |
| Mínimo: | 46 MW | |
| Máximo do CV: | 17,30 | |
| Média do CV: | 7,90 | |

| Dados da Usina | | Gráfico da Potência Disponível Normalizada |
|----------------------|----------------------------|--|
| Usina: | 156 – Três Marias | |
| Máximo: | 396 MW | |
| Mínimo: | 158 MW | |
| Máximo do CV: | 26,93 | |
| Média do CV: | 15,20 | |
| Usina: | 217 – Rosal | |
| Máximo: | 54 MW | |
| Mínimo: | 20 MW | |
| Máximo do CV: | 8,39 | |
| Média do CV: | 1,39 | |
| Usina: | 251 – Serra da Mesa | |
| Máximo: | 1275 MW | |
| Mínimo: | 487 MW | |
| Máximo do CV: | 23,04 | |
| Média do CV: | 11,42 | |
| Usina: | 278 – Manso | |
| Máximo: | 210 MW | |
| Mínimo: | 148 MW | |
| Máximo do CV: | 10,17 | |
| Média do CV: | 4,18 | |

