

MAURICIO SPERANDIO

CLASSIFICAÇÃO DE CONJUNTOS  
CONSUMIDORES DE ENERGIA ELÉTRICA  
VIA MAPAS AUTO-ORGANIZÁVEIS  
E ESTATÍSTICA MULTIVARIADA

Florianópolis – SC

2004

**UNIVERSIDADE FEDERAL DE SANTA CATARINA**

**PROGRAMA DE PÓS-GRADUAÇÃO  
EM ENGENHARIA ELÉTRICA**

**CLASSIFICAÇÃO DE CONJUNTOS CONSUMIDORES DE  
ENERGIA ELÉTRICA VIA MAPAS AUTO-ORGANIZÁVEIS  
E ESTATÍSTICA MULTIVARIADA**

Dissertação submetida à  
Universidade Federal de Santa Catarina  
como parte dos requisitos para a  
obtenção do grau de Mestre em Engenharia Elétrica.

**MAURICIO SPERANDIO**

Florianópolis, Março de 2004.

# CLASSIFICAÇÃO DE CONJUNTOS CONSUMIDORES DE ENERGIA ELÉTRICA VIA MAPAS AUTO-ORGANIZÁVEIS E ESTATÍSTICA MULTIVARIADA

Mauricio Sperandio

‘Esta Dissertação foi julgada adequada para obtenção do Título de Mestre em Engenharia Elétrica, Área de Concentração em *Planejamento de Sistemas de Energia Elétrica*, e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Santa Catarina.’



---

Prof. Jorge Coelho, D.Sc.

Orientador



---

Prof. Jefferson L. Brum Marques, Ph.D.

Coordenador do Programa de Pós-Graduação em Engenharia Elétrica

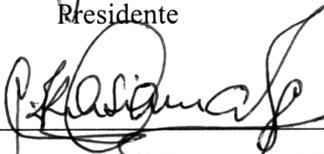
Banca Examinadora:



---

Prof. Jorge Coelho, D.Sc.

Presidente



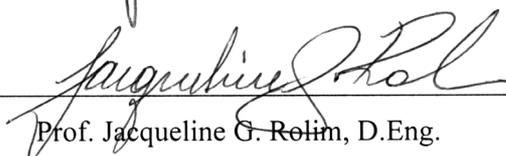
---

Prof. C. Celso de Brasil Camargo, D.Eng.



---

Prof. Edward Gauche, D.Eng.



---

Prof. Jacqueline G. Rolim, D.Eng.

Aos meus pais,  
Alarico e Noemi,  
pelo apoio absoluto.

## **Payada**

Jayme Caetano Braun

Hoje, tempo de mudar, meu coração continua

O mesmo tigre charrua das andanças do passado.

Sempre de pingo encilhado, bombeando pampa e coxilha...

A pátria é minha família! Não há Brasil sem Rio Grande

E nem tirano que mande na alma de um Farroupilha!

Aos pagos de Santa Maria e São Sepé.

## **AGRADECIMENTOS**

Ao time CELESC / UFSC: Prof. Coelho, Hermes, Sílvia, Edward e Ana Barbara, que fizeram parte deste trabalho com idéias, incentivos, revisões e discussões.

Aos demais professores do LabPlan, pela oportunidade de trabalhar nesse laboratório e pelos ensinamentos transmitidos.

Aos colegas do mestrado, Juliana, Paulo Henrique, Guto, Marcelo, Rody e José Luis, pelo companheirismo durante a fase dos créditos e pelas dicas durante a formulação desta dissertação, especialmente às relacionadas com a formatação de texto e figuras fornecidas pelo José.

À turma da velha guarda labplaniana, Agostini, Erlon, Piara, Adriano, Zucarato, Sica, Rafael, Fabiano, João Marco, Dotta, pelos conhecimentos compartilhados e pela parceria e camaradagem nos churrascos e happy hours.

Aos meus parentes aqui em Floripa, tio Nando e tia Lica, Fabiana e Rodrigo, Fernanda e Ricardo, pelo suporte desde minha chegada até hoje, e aos demais amigos e familiares que me incentivaram.

À Mãisa, pelo amor e carinho que venceu distâncias.

Resumo da Dissertação apresentada à UFSC como parte dos requisitos necessários para a obtenção do grau de Mestre em Engenharia Elétrica.

## **CLASSIFICAÇÃO DE CONJUNTOS CONSUMIDORES DE ENERGIA ELÉTRICA VIA MAPAS AUTO-ORGANIZÁVEIS E ESTATÍSTICA MULTIVARIADA**

**Mauricio Sperandio**

Março/2004

Orientador: Prof. Jorge Coelho, D.Sc.

Área de Concentração: Planejamento de Sistemas de Energia Elétrica

Palavras-chave: Metas de Qualidade, Métodos de Agrupamento, Conjuntos Consumidores de Energia Elétrica, Redes Neurais Artificiais.

Número de Páginas: 92 (Noventa e Duas)

É apresentada uma metodologia para formação de agrupamentos de conjuntos consumidores de energia elétrica, contemplando desde a seleção das variáveis a serem utilizadas no processo, até a classificação dos grupos de acordo com a qualidade do fornecimento de seus integrantes. A partir dos dados atuais de 260 conjuntos sob concessão das Centrais Elétricas de Santa Catarina (CELESC), contendo 18 variáveis que descrevem características do mercado consumidor e do sistema elétrico de fornecimento, são selecionadas as variáveis mais significativas por meio da análise estatística multivariada. Então, utilizando uma validação cruzada entre um tipo de rede neural auto-organizável e o tradicional método estatístico das k-médias é feito o agrupamento dos conjuntos consumidores. Os grupos obtidos são classificados de forma a sinalizar quando o nível de continuidade seja inferior ao nível de referência, possibilitando detectar incompatibilidades com a meta, ou com o sistema elétrico de fornecimento. Conhecendo melhor a situação de cada conjunto, a concessionária pode aplicar o controle de qualidade necessário, e rever as tarifas de maneira a remunerar os investimentos indispensáveis à prestação do serviço de acordo com a qualidade regulamentada, ou discutir junto ao órgão regulador a situação do conjunto em relação à meta determinada.

Abstract of Dissertation presented to UFSC as a partial fulfillment of the requirements for the degree of Master in Electrical Engineering.

## **ELECTRICAL CONSUMERS SETS CLASSIFICATION VIA SELF-ORGANIZING MAPS AND MULTIVARIATE STATISTICS**

**Mauricio Sperandio**

March/2004

Advisor: Prof. Jorge Coelho, D.Sc.

Area of Concentration: Electrical Energy Systems Planning

Keywords: Quality Standards, Clustering Methods, Electrical Consumers Sets, Neural Networks.

Number of Pages: 92 (Ninety Two)

This work presents a methodology for clustering electrical consumers sets, from variables selection to clusters classification according to the supply quality of their members. Its applied to a real database of 260 electrical consumers sets of a local utility (Santa Catarina's Electrical Center - CELESC), with 18 variables that describe the consumer market and the electrical supply system, the most significant variables are selected by means of multivariate statistics. Then, the clustering is made using a concurrent validation scheme between a self-organizing neural network and the k-means traditional method. The obtained clusters are arranged to reveal when interruption levels are over the references, allowing detection of standards or systems incompatibilities. With a good description of each set condition, the dealership can identify if investments in the system are really necessary to reach regulator's requirements, or may argue with the regulator about an inconsistent quality goal.

# SUMÁRIO

<b>SUMÁRIO.....</b>	<b>viii</b>
<b>LISTA DE TABELAS.....</b>	<b>x</b>
<b>LISTA DE FIGURAS.....</b>	<b>xi</b>
<b>1 INTRODUÇÃO .....</b>	<b>1</b>
1.1. MOTIVAÇÃO .....	1
1.2. OBJETIVOS.....	2
1.3. ESTRUTURA DO TRABALHO .....	2
<b>2 REGULAMENTAÇÃO DA QUALIDADE DE FORNECIMENTO.....</b>	<b>4</b>
2.1. INDICADORES DE CONTINUIDADE .....	5
2.2. AS PENALIDADES POR INFRAÇÃO .....	7
2.3. METODOLOGIA DA ANEEL PARA O AGRUPAMENTO DE CONJUNTOS.....	10
2.3.1. <i>Determinação das Metas de Qualidade</i> .....	12
<b>3 MÉTODOS DE AGRUPAMENTO.....</b>	<b>14</b>
3.1. K-MÉDIAS .....	15
3.1.1. <i>Algoritmo</i> .....	15
3.1.2. <i>Observações sobre o Processo</i> .....	16
3.2. MAPAS AUTO-ORGANIZÁVEIS.....	19
3.2.1. <i>História</i> .....	19
3.2.2. <i>O Algoritmo</i> .....	22
3.2.3. <i>Visualização dos Mapas</i> .....	31
3.2.4. <i>Aplicações do SOM</i> .....	39
3.2.5. <i>Comparação do SOM com outros Métodos de Classificação</i> .....	41
<b>4 METODOLOGIA PROPOSTA PARA A FORMAÇÃO DE AGRUPAMENTOS</b> <b>.....</b>	<b>44</b>
4.1. REVISÃO DOS CRITÉRIOS .....	45
4.2. SELEÇÃO DE VARIÁVEIS VIA ESTATÍSTICA MULTIVARIADA.....	49
4.2.1. <i>Análise das Variáveis de Mercado</i> .....	51
4.2.2. <i>Análise das Variáveis do Sistema Elétrico</i> .....	57
4.3. MÉTODO DE AGRUPAMENTO POR VALIDAÇÃO CRUZADA .....	61

4.3.1.	<i>Agrupamento da Base de Mercado</i> .....	65
4.3.2.	<i>Agrupamento da Base do Sistema Elétrico</i> .....	72
<b>5</b>	<b>CLASSIFICAÇÃO E ANÁLISE DOS GRUPOS FORMADOS</b> .....	<b>75</b>
5.1.	CLASSIFICAÇÃO DOS AGRUPAMENTOS PELA QUALIDADE .....	76
5.1.1.	<i>Ordenação da Base de Mercado</i> .....	76
5.1.2.	<i>Ordenação da Base do Sistema Elétrico</i> .....	77
5.2.	CRUZAMENTO DOS GRUPOS DE MERCADO E DE SISTEMA.....	78
5.3.	AVALIAÇÃO DOS CONJUNTOS CRÍTICOS .....	79
<b>6</b>	<b>CONCLUSÕES E SUGESTÕES</b> .....	<b>84</b>
6.1.	CONCLUSÕES.....	84
6.1.1.	<i>Sobre a Seleção das Variáveis</i> .....	84
6.1.2.	<i>Sobre o Método de Agrupamento</i> .....	85
6.1.3.	<i>Sobre a Classificação e Análise dos Grupos Obtidos</i> .....	85
6.2.	SUGESTÕES PARA TRABALHOS FUTUROS.....	86
	<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	<b>87</b>

## LISTA DE TABELAS

<b>Tabela 1.</b>	<i>Variação de configurações em 500 inícios diferentes.....</i>	<i>17</i>
<b>Tabela 2.</b>	<i>Índices de Estabilidade da Topologia para um mapa 7x7 classificando dados macroeconômicos de diferentes países.....</i>	<i>30</i>
<b>Tabela 3.</b>	<i>Descrição do Consumo por Classe e da Área do Município.....</i>	<i>51</i>
<b>Tabela 4.</b>	<i>Descrição do Número de Consumidores por Classe.....</i>	<i>51</i>
<b>Tabela 5.</b>	<i>Relação entre o Consumo e o Número de Consumidores por Classe.....</i>	<i>52</i>
<b>Tabela 6.</b>	<i>Matriz de Correlações entre as Variáveis de Mercado.....</i>	<i>53</i>
<b>Tabela 7.</b>	<i>Análise Fatorial das Variáveis de Mercado.....</i>	<i>54</i>
<b>Tabela 8.</b>	<i>Cargas Fatoriais para as Variáveis de Mercado.....</i>	<i>55</i>
<b>Tabela 9.</b>	<i>Análise Fatorial da Nova Base de Mercado.....</i>	<i>56</i>
<b>Tabela 10.</b>	<i>Cargas Fatoriais da Nova Base de Mercado.....</i>	<i>56</i>
<b>Tabela 11.</b>	<i>Descrição das Variáveis do Sistema Elétrico.....</i>	<i>57</i>
<b>Tabela 12.</b>	<i>Matriz de Correlações das Variáveis do Sistema.....</i>	<i>58</i>
<b>Tabela 13.</b>	<i>Análise Fatorial das Variáveis do Sistema.....</i>	<i>59</i>
<b>Tabela 14.</b>	<i>Cargas Fatoriais para as Variáveis do Sistema.....</i>	<i>59</i>
<b>Tabela 15.</b>	<i>Análise Fatorial da Nova Base do Sistema.....</i>	<i>60</i>
<b>Tabela 16.</b>	<i>Cargas Fatoriais da Nova Base do Sistema.....</i>	<i>60</i>
<b>Tabela 17.</b>	<i>Dados Sobre a Validação Cruzada de Mercado ao Alterar a Topologia do Mapa com 6 Grupos.....</i>	<i>70</i>
<b>Tabela 18.</b>	<i>Dados Sobre a Validação Cruzada do Sistema ao Alterar a Topologia do Mapa com 10 Grupos.....</i>	<i>73</i>
<b>Tabela 19.</b>	<i>Classificação dos Grupos de Mercado.....</i>	<i>76</i>
<b>Tabela 20.</b>	<i>Classificação dos Grupos de Sistema.....</i>	<i>77</i>
<b>Tabela 21.</b>	<i>Comparação das Classificações com o Grupo 5 da ANEEL.....</i>	<i>82</i>

## LISTA DE FIGURAS

<b>Figura 1.</b> Modelo de Yardstick Competition aplicado ao grupo 27 da ANEEL.....	13
<b>Figura 2.</b> Decaimento das horas a serem reduzidas do DEC anualmente para o competidor 33 do grupo. ....	13
<b>Figura 3.</b> Soluções possíveis para agrupar 4 dados equidistantes em 2 grupos (WISHART) .....	16
<b>Figura 4.</b> Rede Neural do tipo Perceptron.....	20
<b>Figura 5.</b> Rede Auto-Organizável. ....	21
<b>Figura 6.</b> Transformação $\Phi: \mathcal{E} \rightarrow \mathcal{S}$ . (adaptada de HAYKIN).....	22
<b>Figura 7.</b> Mapa ou Rede hexagonal com 16 modelos (4x4).....	23
<b>Figura 8.</b> Representação do processo competitivo, todos os modelos recebem a mesma informação (D1), mas apenas um é ativado (M6).....	25
<b>Figura 9.</b> Um exemplo de função de vizinhança gaussianiana. ....	26
<b>Figura 10.</b> Função de vizinhança aplicada em um mapa hexagonal 4x4.....	26
<b>Figura 11.</b> Agrupamento de dados pela densidade do mapa (adaptada de XUEGONG et al) .....	32
<b>Figura 12.</b> Matriz de distâncias para uma base de dados em forma de tetraedro (ULTSCH,1990).....	33
<b>Figura 13.</b> Condição inicial do mapa 5x5; (a) gráficos “pizza” mostrando a quantidade de cada cor nos modelos iniciais (RGB); (b) configuração inicial do mapa hexagonal com as composições de cores . ....	34
<b>Figura 14.</b> Condição final do mapa treinado; (a) gráficos “pizza” mostrando a quantidade de cada cor nos modelos organizados (RGB); (b) configuração final do mapa hexagonal com as cores somadas.....	34
<b>Figura 15.</b> As três fases do processo de auto-organização;(a) base de dados / configuração inicial; (b) mapa treinado; (c) distribuição de densidade.....	35
<b>Figura 16.</b> Quatro tipos de gráficos para visualizar a matriz de distâncias; (a) com as distâncias individuais em 5 tons de cinza; (b) com uma superfície colorida em 256 tons; (c) histograma com 7 níveis; (d) superfície colorida em 3 dimensões.....	36
<b>Figura 17.</b> Mapas componentes das variáveis R,G e B.....	37
<b>Figura 18.</b> Agrupamento dos dados RGB. ....	38
<b>Figura 19.</b> Mapa 10x10 treinado para determinar estados de operação de um sistema 9 barras, 3 geradores e 3 cargas.....	40

<b>Figura 20.</b>	<i>Anéis entrelaçados, e tentativa do k-médias em dividi-los em dois grupos</i>	42
<b>Figura 21.</b>	<i>Matriz de distâncias para uma rede 64x64 treinada com a base de dois anéis.</i>	43
<b>Figura 22.</b>	<i>Percentagem de Fatores Gerenciáveis e Não-Gerenciáveis no FEC de um Conjunto da AES Eletropaulo</i>	46
<b>Figura 23.</b>	<i>Histogramas de Frequências: Sem e Com Transformação Logarítmica</i>	47
<b>Figura 24.</b>	<i>Árvore de Associação entre as Variáveis de Mercado</i>	55
<b>Figura 25.</b>	<i>Representação do Sistema de Fornecimento de Energia Elétrica</i>	58
<b>Figura 26.</b>	<i>Árvore de Associação entre as Variáveis do Sistema</i>	60
<b>Figura 27.</b>	<i>Fluxograma do Procedimento de Validação Cruzada (SOM x K-médias)</i>	62
<b>Figura 28.</b>	<i>Matriz de Distâncias de Mercado: (a) U-mat com densidade; (b) Histogramas suavizados.</i>	65
<b>Figura 29.</b>	<i>Mapas Componentes da Base de Mercado.</i>	66
<b>Figura 30.</b>	<i>Mapa 8x7 Mostrando os 5 Grupos do k-médias Sobre os Neurônios Agrupados.</i>	67
<b>Figura 31.</b>	<i>Mapa 8x7 Mostrando 6 Grupos do k-médias Sobre os Neurônios Agrupados.</i>	68
<b>Figura 32.</b>	<i>Reorganização ao Treinar o Mapa de Mercado sem a Variável Área.</i>	69
<b>Figura 33.</b>	<i>Mapas Componentes do Treinamento Sem a Área.</i>	69
<b>Figura 34.</b>	<i>Mapa 8x9 Escolhido Pela Validação Cruzada para os Grupos de Mercado.</i>	71
<b>Figura 35.</b>	<i>Matriz de Distâncias do Sistema: (a) U-mat com densidade; (b) Histogramas suavizados.</i>	72
<b>Figura 36.</b>	<i>Mapa 11x10 Escolhido Pela Validação Cruzada para os Grupos de Sistema.</i>	73
<b>Figura 37.</b>	<i>Mapas Componentes 11x10 das Variáveis de Sistema.</i>	74
<b>Figura 38.</b>	<i>Histograma com a Frequência do Cruzamento de Grupos de Mercado e Sistema.</i>	78
<b>Figura 39.</b>	<i>Dispersão dos Conjuntos em Relação às Metas de DEC e FEC.</i>	80
<b>Figura 40.</b>	<i>Histograma dos Índices Relativos Maiores que 100.</i>	80
<b>Figura 41.</b>	<i>Histograma do Número de Violações por Combinação Mercado-Sistema.</i>	81

# - CAPÍTULO 1 -

## INTRODUÇÃO

### 1.1. Motivação

O Setor Elétrico Brasileiro está em fase de transformação. A estrutura de fornecimento de energia elétrica está sendo reavaliada e mudanças estão sendo propostas, dentre as quais, pode-se destacar uma mudança que deve afetar tanto a qualidade do fornecimento quanto à questão tarifária: *a revisão dos agrupamentos de conjuntos consumidores de energia elétrica.*

Conjuntos consumidores de energia elétrica são grandes áreas sob concessão das empresas de distribuição, normalmente caracterizados por municípios inteiros como no caso de Santa Catarina. Atualmente a qualidade do atendimento a esses conjuntos é medida apenas por dois índices : o FEC (Frequência Equivalente de Interrupção por Unidade Consumidora), e o DEC (Duração Equivalente de Interrupção por Unidade Consumidora), que basicamente determinam a quantidade e a duração das faltas de energia elétrica.

Com o intuito de melhorar a qualidade do fornecimento, o órgão regulador (Agência Nacional de Energia Elétrica - ANEEL) determinou metas a serem atingidas por esses índices comparando concessionárias brasileiras. Para tanto, tomou como base as informações de DEC e FEC relativas ao ano 2000, e cinco características (variáveis) de cada conjunto. Essas variáveis foram empregadas para agrupar conjuntos similares, e então foram comparados os índices dentro de cada grupo e traçada uma meta.

Contudo, a partir do segundo ano da determinação das metas, muitos conjuntos não foram capazes de cumprí-las, levando as concessionárias a questionar como foram determinadas. Não obstante, a ANEEL permite através das resoluções 024/00 e 075/03 que cada concessionária proponha novos critérios para o agrupamento dos conjuntos consumidores, visando evidenciar vantagens técnicas, econômicas e sociais.

Sendo assim, este trabalho visa investigar a formação de grupos de conjuntos consumidores, tanto na seleção das variáveis a serem utilizadas para o agrupamento dos dados, quanto os métodos para tal função.

## **1.2. Objetivos**

Selecionar de uma série de variáveis disponibilizadas pela empresa, uma base de dados mínima que melhor represente as diferenças e semelhanças entre os conjuntos.

Comparar dois métodos de agregação diferentes, sendo um, o utilizado pela ANEEL, e o outro, um moderno método de agregação de dados e extração de conhecimento que se baseia em conceitos de inteligência artificial.

Reclassificar os conjuntos de forma a obter grupos mais homogêneos, que reflitam melhor a situação dos dados agrupados, possibilitando à concessionária identificar se um conjunto foi mal classificado pela ANEEL, ou se realmente é necessário aplicar um controle de qualidade ou reforços no sistema de distribuição.

## **1.3. Estrutura do Trabalho**

Para alcançar os objetivos propostos, os próximos 5 capítulos irão apresentar o seguinte:

O Capítulo 2 irá mostrar a regulamentação atual que rege a formação dos conjuntos consumidores, o cálculo dos indicadores de qualidade e as penalidades previstas de acordo com a infração. Também é apresentada a metodologia utilizada pelo órgão regulador para a formação dos agrupamentos de conjuntos e determinação das metas.

No Capítulo 3 é feita uma revisão bibliográfica sobre os métodos de agregação que serão utilizados, com enfoque especial no método baseado em princípios de inteligência chamado de Mapa Auto-Organizável. É citada um pouco da história do desenvolvimento deste método, bem como seu algoritmo e técnicas de visualização inerentes a ele, explicadas por meio de um exemplo, e por fim um breve relato de algumas possibilidades de aplicações desse poderoso algoritmo e comparações com outros métodos.

O Capítulo 4, começa com uma revisão dos critérios adotados pela ANEEL em sua classificação dos conjuntos consumidores, apontando alguns problemas já apresentados na literatura por empresas de distribuição. Na seqüência, é mostrada a análise estatística multivariada realizada para definir as principais variáveis que devem compor a base de dados efetiva para o agrupamento. Agrupamento este, que é realizado por meio de uma validação cruzada entre os dois algoritmos apresentados no Capítulo 3, de forma a garantir que os resultados obtidos são bastante corretos.

Os grupos obtidos com a metodologia proposta, são avaliados no Capítulo 5, onde são detectados conjuntos críticos em que a empresa deve atuar de forma a sanar problemas com o sistema de fornecimento, ou deve discutir com o órgão regulador por causa de uma classificação errônea.

Por fim, são resumidas as contribuições da pesquisa e feitas algumas sugestões para trabalhos futuros no Capítulo 6.

## - CAPÍTULO 2 -

### REGULAMENTAÇÃO DA QUALIDADE DE FORNECIMENTO

O setor elétrico brasileiro vem passando por um processo de reestruturação, que foi iniciado pela separação das atividades de geração, transmissão e distribuição de energia elétrica dentro de uma mesma empresa. Tal ação foi seguida por um processo de privatização, face à ausência de recursos do governo para investir na expansão do sistema, e que atingiu grande parte das concessionárias de distribuição. A intenção era levar a competição na geração e comercialização de energia, e o livre acesso às redes de transmissão, podendo as distribuidoras escolher de quem comprar a eletricidade. Hoje em dia estes aspectos estão sendo revistos e muito ainda pode mudar.

Porém, o consumidor final de pequeno-médio porte (cativo) é sujeito à concessionária, que é um monopólio, e portanto precisa de regulação. Este papel é exercido pela Agência Nacional de Energia Elétrica (ANEEL), que tem dentre suas atribuições regimentais a competência de regular os serviços de energia elétrica, expedindo atos necessários ao cumprimento das normas estabelecidas pela legislação em vigor, e estimular a melhoria do serviço prestado.

A qualidade do fornecimento de energia elétrica no Brasil atualmente é apurada apenas com relação à continuidade de sua distribuição, nos aspectos de duração e frequência de faltas, devendo ser supervisionada e avaliada por meio de indicadores que expressem os valores vinculados a conjuntos de unidades consumidoras, expostos nas resoluções 024 de 2000 e 075 de 2003 .

Uma unidade consumidora é um conjunto de instalações e equipamentos elétricos caracterizados pelo recebimento de energia elétrica em um só ponto de entrega, com medição individualizada e vinculada a um único consumidor.

Um conjunto de unidades consumidoras abrange uma área geográfica contígua que é atendida pela concessionária. Geralmente é considerado o município como um todo.

## 2.1. Indicadores de Continuidade

As concessionárias devem apurar os indicadores de continuidade de duas formas distintas:

- considerando as interrupções com duração maior ou igual a 3 minutos; e
- considerando as interrupções com duração maior ou igual a 1 minuto.

Com exceção de faltas em instalações de unidades consumidoras que não provoquem interrupções em instalações de terceiros.

A contabilização das interrupções ocorridas no conjunto de unidades consumidoras para o cálculo dos indicadores deve ser feita entre o início e o fim de cada mês.

Os indicadores de continuidade de conjunto são:

- Duração Equivalente de Interrupção por Unidade Consumidora (DEC)
- Frequência Equivalente de Interrupção por Unidade Consumidora (FEC)

Para a apuração do DEC deve ser utilizada a seguinte fórmula:

$$DEC = \frac{\sum_{i=1}^k Ca(i) \times t(i)}{Cc} \quad (1)$$

Para o FEC a fórmula é:

$$FEC = \frac{\sum_{i=1}^k Ca(i)}{Cc} \quad (2)$$

onde:

- DEC = Expresso em horas e centésimos de hora;
- FEC = Expresso em número de interrupções e centésimos de número de interrupções;
- $Ca(i)$  = Número de unidades consumidoras interrompidas em um evento (i) no período de apuração;
- $t(i)$  = Duração de cada evento (i) no período de apuração;
- $i$  = Índice de eventos ocorridos no sistema que provocam interrupções em unidades consumidoras;
- $k$  = Número máximo de eventos no período considerado;
- $Cc$  = Número total de unidades consumidoras do conjunto considerado no final do período de apuração.

As concessionárias devem enviar para a ANEEL os indicadores DEC e FEC de todos os seus conjuntos atuais até o último dia útil do mês subsequente ao período de apuração. Esses valores também devem ser apresentados aos consumidores na fatura de energia elétrica em conjunto com os indicadores de continuidade individuais.

Os indicadores de continuidade individuais são:

- Duração de Interrupção por Unidade Consumidora (DIC)

$$DIC = \sum_{i=1}^k t(i) \quad (3)$$

- Duração Máxima de Interrupção Contínua por Unidade Consumidora (DMIC)
- Frequência de Interrupção por Unidade Consumidora (FIC)

$$FIC = n \quad (4)$$

onde:

- DIC = Expresso em horas e centésimos de hora;
- FIC = Expresso em número de interrupções e centésimos de número de interrupções;
- $i$  = Índice de interrupções da unidade consumidora no período de apuração, variando de 1 a  $n$ ;
- $n$  = Número de interrupções da unidade consumidora considerada no período de apuração;
- $t(i)$  = Duração de cada interrupção ( $i$ ) da unidade considerada no período de apuração.

Visando manter ou melhorar a continuidade da distribuição de energia elétrica foram estabelecidas pela ANEEL metas para os indicadores individuais e de conjunto. As metas para os indicadores individuais estão expostas nas resoluções 024/00 e 075/03 por meio de tabelas que relacionam o nível de tensão do consumidor com a faixa de variação das metas anuais dos indicadores de continuidade de conjunto. Já as metas para DEC e FEC foram determinadas com base no trabalho de TANURE (2000), e publicadas em

resoluções separadas para cada concessionária do Brasil, por exemplo na resolução nº 573 de 2000 para a CELESC.

## 2.2. As Penalidades por Infração

Em caso de violação das metas estabelecidas no período de apuração as concessionárias deverão sofrer penalidades de acordo com o fato gerador. Estas penalidades são classificadas em duas categorias, apresentadas a seguir.

### I - Violação de Padrão Indicador de Continuidade Individual

Penalidade: Compensação ao consumidor de valor a ser creditado na fatura de energia elétrica no mês subsequente à apuração.

No cálculo do valor da compensação serão utilizadas as seguintes fórmulas:

a) Para o DIC:

$$Penalidade = \left( \frac{DIC_v}{DIC_p} - 1 \right) DIC_p \times \frac{CM}{730} \times kei \quad (5)$$

b) Para o DMIC:

$$Penalidade = \left( \frac{DMIC_v}{DMIC_p} - 1 \right) DMIC_p \times \frac{CM}{730} \times kei \quad (6)$$

c) Para o FIC:

$$Penalidade = \left( \frac{FIC_v}{FIC_p} - 1 \right) DIC_p \times \frac{CM}{730} \times kei \quad (7)$$

onde:

DIC<sub>v</sub> = Duração de Interrupção por Unidade Consumidora verificada no período considerado, expresso em horas e centésimos de hora;

DIC<sub>p</sub> = Padrão de continuidade estabelecido no período considerado para o indicador de Duração de Interrupção por Unidade Consumidora, expresso em horas e centésimos de hora;

DMIC<sub>v</sub> = Duração Máxima de Interrupção Contínua, verificada no período considerado, expresso em horas e centésimos de hora;

DMICp = Padrão de continuidade estabelecido no período considerado para o indicador de Duração Máxima de Interrupção Contínua, expresso em horas;

FICv = Frequência de Interrupção por Unidade Consumidora verificada no período considerado, expresso em número de interrupções;

FICp = Padrão de continuidade estabelecido no período considerado para o indicador de Frequência de Interrupção por Unidade Consumidora, expresso em número de interrupções;

CM = Média aritmética dos valores líquidos das faturas de energia elétrica ou dos encargos de uso dos sistemas de distribuição correspondentes aos meses do período de apuração do indicador;

730 = Número médio de horas no mês; e

kei = Coeficiente de majoração, que variará de 10 a 50, para consumidor cativo, e cujo valor, fixado em 10 (dez), poderá ser alterado pela ANEEL a cada revisão periódica das tarifas.

## II - Violação de Padrão do Indicador de Continuidade de Conjunto:

Penalidade: Pagamento de multa à ANEEL conforme as disposições da Resolução nº-318, de 6 de outubro de 1998, ou de suas eventuais atualizações.

Alguns itens dessa resolução relativos aos indicadores de continuidade são:

- Não encaminhar a ANEEL, nos prazos estabelecidos e conforme previsto nos regulamentos específicos, indicadores utilizados para a apuração da qualidade do fornecimento de energia elétrica; (multa do Grupo I )
- Não manter registro atualizado dos dados utilizados para apuração dos índices de qualidade do fornecimento de energia elétrica, continuidade e conformidade, segundo definido nos regulamentos específicos, com a anotação, quando for o caso, das causas, dos períodos de duração e das providências adotadas para a solução do problema; (multa do Grupo II )
- Não cumprir as disposições legais e contratuais relativas aos níveis de qualidade dos serviços; (multa do Grupo III )

Os valores das multas serão determinados mediante aplicação, sobre o valor do faturamento, nos casos de concessionários, permissionários e autorizados de instalações e

serviços de energia elétrica, ou sobre o valor estimado da energia produzida, nos casos de autoprodução e produção independente, correspondente aos últimos doze meses anteriores à lavratura do Auto de Infração, dos seguintes percentuais:

- Grupo I: até 0,01% (um centésimo por cento) do faturamento;
- Grupo II: até 0,1% (um décimo por cento) do faturamento;
- Grupo III: até 1% (um por cento) do faturamento;

Sendo o valor do faturamento as receitas oriundas da venda de energia elétrica e prestação de serviços, deduzidos o ICMS e o ISS. Para a determinação do valor estimado da energia será considerada a quantidade produzida, multiplicada pelo valor de referência de geração, estabelecido pela ANEEL.

Ocorrendo a reincidência, dentro do período de doze meses subsequentes à lavratura do Auto de Infração, proceder-se-á da seguinte forma:

I - aplicação da multa correspondente ao Grupo I, para os casos punidos com advertência;

II - aplicação de acréscimo de cinquenta por cento, limitado ao percentual de dois por cento, para os casos punidos com multa.

Na hipótese da ocorrência concomitante de mais de uma infração serão aplicadas, simultânea e cumulativamente, as penalidades correspondentes a cada uma delas.

Interrupções associadas à situação de emergência ou de calamidade pública decretada por órgão competente, não serão consideradas para efeito de compensação quando da violação dos padrões de indicadores individuais;

Quando ocorrer violação de mais de um indicador de continuidade individual, no período de apuração, deverá ser considerado, para efeito de compensação, aquele indicador que apresentar o maior valor de compensação;

### **2.3. Metodologia da ANEEL para o Agrupamento de Conjuntos**

A seguir será apresentado um resumo da metodologia adotada pela ANEEL (TANURE, 2000) para a determinação das metas dos índices de continuidade de conjunto.

Com a intenção de introduzir instrumentos para induzir os agentes do setor elétrico a melhorarem seus padrões de desempenho foi criada uma metodologia para comparar as concessionárias de todo o Brasil. Além do setor ser monopolista, as empresas detentoras de concessões não enfrentavam nenhum tipo de contestação da qualidade de seu fornecimento, e a criação de mecanismos que permitissem uma verificação do seu padrão de desempenho relativo tornou-se fundamental para o estabelecimento de metas a serem perseguidas.

No caso de atividades onde existe concorrência estas metas surgem naturalmente pois a existência de concorrentes em um mesmo mercado obriga que as empresas assumam uma correção de suas práticas pela simples razão de que, se não o fizerem, serão substituídas por outras de melhor desempenho frente aos seus consumidores.

Assim, os segmentos monopolistas demandam um esforço adicional do órgão regulador, que deve criar mecanismos que emulem um ambiente competitivo entre os agentes que atuam nestes segmentos, visando estimular melhorias contínuas no desempenho dos mesmos.

A metodologia adotada visa definir metas diferenciadas considerando as particularidades de cada conjunto (atributos ou variáveis), mantendo a diretriz de que os padrões de atendimento ao consumidor serão continuamente melhorados. Esses atributos foram escolhidos respeitando-se as atuais limitações das empresas para informar os dados de seus sistemas de forma confiável. Como as diversas empresas apresentam diferentes níveis de conhecimento de seus sistemas, o grau de confiabilidade dos dados informados varia muito de empresa para empresa, o que implica em utilizar um conjunto mínimo de dados que possam ser enviados por todas elas.

Assim, foram considerados apenas cinco atributos básicos, dentre 4135 conjuntos de unidades consumidoras de 56 concessionárias distribuidoras de todo país.

1. Área de cada conjunto- Área - em km<sup>2</sup>;
2. Extensão da rede primária – RAP - em km;
3. Potência instalada- Potinst - em kVA;
4. Número de consumidores - Ncons;
5. Consumo médio mensal de cada conjunto – CMM - em MWh;

Para classificar essa quantidade de dados em grupos com características semelhantes (*clusters*) foi escolhida a técnica de agregação estatística das k-médias. Nesse método é necessário definir o número de grupos desejados *a priori*, e então o algoritmo encontra a melhor partição dos dados. Esse processo será melhor explicado no item 3.1.

Antes de agrupar os conjuntos foi realizada a normalização das variáveis para compatibilizar a variação e ordem de grandeza entre elas, pois a diferença de escala influencia significativamente os resultados, uma vez que a mesma pode ser interpretada como um peso ou ponderador diferenciado entre os atributos. Foi utilizada a normalização de variância, onde cada valor do atributo é reduzido da média e dividido pelo desvio padrão, como mostra a fórmula:

$$X(i)' = \frac{X(i) - \bar{X}}{\sigma(X)} \quad (8)$$

onde:

$X(i)'$  = Dado normalizado;

$X(i)$  = Dado original;

$\bar{X}$  = Média aritmética do atributo, dada por:

$$\bar{X} = \frac{\sum_{i=1}^N X(i)}{N} \quad (9)$$

$N$  = Número total de dados;

$\sigma(X)$  = Desvio Padrão do atributo, dado por:

$$\sigma(X) = \sqrt{\frac{\sum_{i=1}^N (X(i) - \bar{X})^2}{N}} \quad (10)$$

Após a normalização ainda foi aplicada uma transformação logarítmica para reduzir a dispersão dos dados, com o intuito de aumentar a proximidade entre eles para formar grupos maiores.

Para a definição do número final de agrupamentos formados foi realizado um estudo da dispersão dentro dos grupos\*. Essa dispersão é máxima para um grupo contendo todos os dados, e zero quando cada dado é definido como o próprio grupo. Assim, decidiu-se por um total de 30 grupos. O autor não utilizou um procedimento de otimização para minimizar as distâncias dentro do grupo e/ou maximizar a distância entre grupos.

### 2.3.1. Determinação das Metas de Qualidade

O passo seguinte foi então determinar as metas para cada um dos trinta grupos. O modelo de regulação de desempenho escolhido foi o *Yardstick Competition* devido à distância entre os indicadores de um mesmo grupo. Nesse modelo as empresas são estimuladas a adotarem um desempenho baseado em um comportamento padrão, fixado de acordo com análises de desempenho dos agentes que compõem o universo considerado e que se julga razoável atingir. Nesta abordagem, as empresas que se situam abaixo da referência tem uma margem de folga para seu desempenho. As empresas situadas acima desta marca devem buscar melhorar seus indicadores para aproximar-se do valor de referência. Para identificação desta marca de referência pode-se adotar muitas técnicas diferentes, dependendo de cada contexto em que se trabalha e do desempenho que se busca atingir. Nesse caso a ANEEL definiu o primeiro decil como referência. Decis são valores que dividem os dados em 10 partes iguais.

Para a transição até a meta a ser atingida adotou-se um período de duas revisões tarifárias médias, ou seja oito anos. Definiu-se uma taxa de melhoria constante a partir da fórmula:

$$M_x = \left( \frac{M_f}{I_{atual}} \right)^{\frac{x}{8}} \cdot I_{atual} \quad (11)$$

onde:  $M_x$  = Meta para o ano x (de 1 a 7);

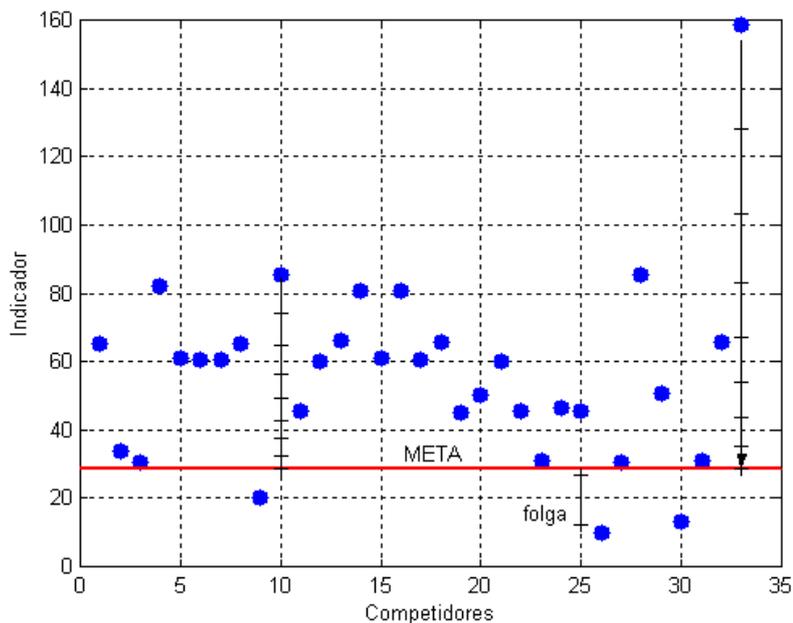
$M_f$  = Meta final a ser atingida no ano 8;

$I_{atual}$  = Indicador atual, DEC ou FEC.

---

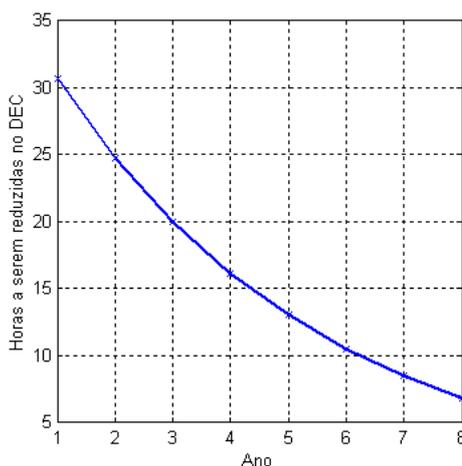
\* Isso será explicado no item 3.1.

Um exemplo de como os componentes de um grupo devem proceder com seus indicadores de qualidade pode ser visto na Figura 1 que mostra 33 conjuntos reunidos no grupo 27 da ANEEL.



**Figura 1.** Modelo de *Yardstick Competition* Aplicado ao Grupo 27 da ANEEL.

O tempo de transição envolve uma redução do tipo logaritmo inverso, que pode ser observada na Figura 1 para os competidores 10 e 33. Esse último é detalhado na Figura 2.



**Figura 2.** Decaimento das Horas a Serem Reduzidas do DEC Anualmente para o Competidor 33 do Grupo.

## - CAPÍTULO 3 -

### MÉTODOS DE AGRUPAMENTO

Métodos de agrupamento, ou *Cluster Analysis*, são ferramentas para análise exploratória de dados a fim de resolver problemas de classificação. O objetivo é associar argumentos variáveis (pessoas, coisas, eventos, etc) em grupos, ou *clusters*, de forma que o grau de similaridade seja grande entre membros de um mesmo grupo e pequeno entre grupos diferentes. Então cada conjunto descreve, em termos dos dados agrupados, a classe a qual seus membros pertencem, e abstrai uma ou mais características particulares que se destacam.

Portanto a análise de agrupamentos é uma ferramenta de extração de conhecimento, podendo revelar associações e estruturas em uma base de dados que não são evidentes a princípio, ainda que não sejam proveitosas após serem descobertas. Os resultados dessa análise podem contribuir para uma definição formal de um esquema de classificação, tal como taxonomia de animais, insetos ou plantas; ou sugerir modelos estatísticos para descrever populações; indicar regras para atribuir novos casos à classes de diagnósticos, entre outras funções de relacionamento e similaridade.

Muitos métodos de agregação precisam primeiramente ter estabelecida uma medida de proximidade ou similaridade entre os elementos que compõem a amostra em estudo para poder formar grupos. Essa medida é normalmente chamada de distância, e para determiná-la pode-se usar diversos critérios, porém a norma Euclidiana é a preferida e normalmente utilizada.

A distância entre dois elementos (vetores) em um espaço euclidiano de  $n$ -dimensões é dada pela seguinte equação :

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (12)$$

onde:

$d(x,y)$  = Distância do vetor  $x$  para o vetor  $y$ ;

$x_i, y_i$  = Elemento da dimensão ou atributo  $i$  dos vetores  $x$  e  $y$ ;

$n$  = Número total de dimensões ou atributos.

A seguir serão apresentados dois métodos de agregação que utilizam a métrica euclidiana, porém sob paradigmas diferentes. O primeiro é o tradicional algoritmo das  $k$ -médias, e o outro um modelo de Rede Neural Artificial (RNA) que utiliza princípios de Inteligência Artificial (IA) para realizar a classificação.

### 3.1. K-Médias

O objetivo deste método é construir um conjunto de  $k$  grupos a partir de uma massa de dados ( $m$ ), de forma que cada unidade pertença a apenas um dos grupos. É um método de agrupamento não-hierárquico, ou seja, não existe um encadeamento entre os dados agrupados.

#### 3.1.1. Algoritmo

O algoritmo das  $k$ -médias foi desenvolvido por J.B. MacQueen em 1967, e tem esse nome para indicar o processo de designar um grupo ( $k$ ) com o centróide mais próximo (média) a cada unidade de dado. É composto por 4 passos básicos:

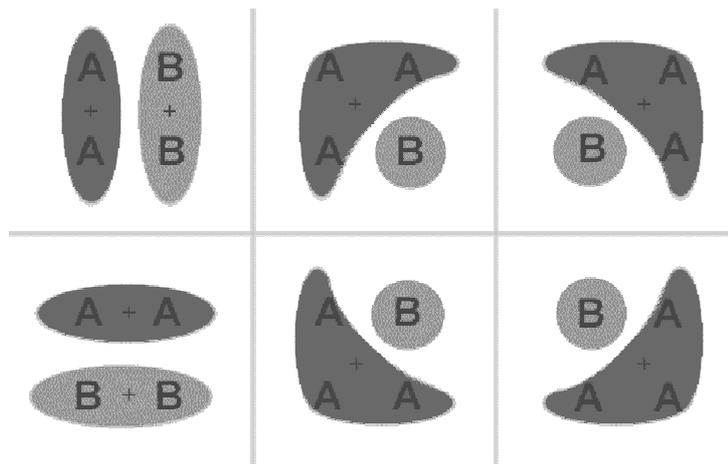
- 1- Defini-se o número de grupos a serem formados ( $k$ ). Então se escolhem  $k$  unidades de dados para serem os primeiros centróides.
- 2- Atribuir cada uma das ( $m - k$ ) unidades restantes a um grupo com o centróide mais próximo.
- 3- Recalculam-se os centróides com a média de cada grupo.

- 4- Repasse-se a base de dados procurando o novo centróide mais próximo, retorna ao passo 3 até que não haja mais alteração dos grupos.

É um processo bem simples que não exige cálculos complexos, mas dependendo da base de dados que será explorada o número de iterações pode ser elevado. Ainda assim alguns aspectos devem ser considerados para que o resultado não seja mal interpretado, ou o algoritmo não atinja um objetivo esperado.

### 3.1.2. Observações sobre o Processo

Dependendo da base de dados, o passo inicial pode influir no arranjo final dos grupos, pois o processo pode convergir em um ótimo local (ANDERBERG, 1973). Um exemplo simples pode ser visto na Figura 3 onde são mostradas as seis soluções encontradas para o caso de agrupar 4 dados equidistantes em dois grupos.



**Figura 3.** Soluções possíveis para agrupar 4 dados equidistantes em 2 grupos (WISHART,2003)

Os dados nos grupos formados pelas letras A e B estão mais próximos de seu centróide (+) do que do centróide do outro grupo nas seis configurações. No entanto a dispersão (soma das distâncias) é menor para as configurações 2-2 na esquerda, sendo uma solução ótima global, enquanto as outras quatro configurações (3-1) são ótimos locais.

Esse início pode ser feito de diferentes maneiras, algumas técnicas utilizadas são:

- Escolher os k dados iniciais da base;
- Sortear aleatoriamente k dados da base;
- Deliberadamente escolher dados da base;
- Gerar centróides com base nos autovetores da base de dados;
- Utilizar outros métodos para gerar uma partição inicial.

Uma forma de analisar a influência do passo inicial é a repetição do algoritmo com inícios diferentes, então se pode escolher a formação final que ocorrer maior número vezes. Um exemplo apresentado por WISHART (2003) com dados de 25 animais mamíferos para serem agrupados em 5 classes foi repetido 500 vezes com centróides iniciais diferentes é mostrado na Tabela 1.

**Tabela 1.** Variação de configurações em 500 inícios diferentes

<b>Solução</b>	<b>Ocorrências</b>	$\delta_t$	<b>Tamanho dos Grupos</b>
1	336	2.58086	8-5-6-4-2
2	49	2.80894	10-2-7-5-1
3	4	2.92945	10-4-7-2-2
4	101	2.94606	9-1-7-2-6
5	2	2.95241	6-4-11-2-2
6	3	3.38873	10-8-1-5-1
7	5	3.45002	11-1-6-6-1

Ainda existem técnicas de otimização que podem ser aplicadas durante o processo de procura dos centróides, buscando minimizar a distância dentro dos grupos e/ou maximizar a distância entre os centróides de cada grupo.

Porém a principal questão do método é a escolha do número de partições ideal para uma base de dados. Pois se os dados da base possuem mais de 3 dimensões não é possível visualizar uma estrutura espacial que possa indicar o número de grupos previamente. Sem um conhecimento *a priori* dos dados que passarão pelo processo de agrupamento é preciso executar o algoritmo algumas vezes variando o número de grupos  $k$ . Um modo de apreciar esses resultados e tomar uma decisão é avaliar a razão da dispersão dos dados dentro dos grupos sobre a dispersão total (TANURE, 2000).

$$\Delta_k = \frac{\delta_k}{\delta_t} \quad (13)$$

onde:

$\Delta_k$  = Razão de Dispersão para  $k$  grupos;

$\delta_k$  = Somatório da dispersão dos  $k$  grupos;

$\delta_t$  = Dispersão total da base de dados, é constante para uma mesma base.

A razão de dispersão reduz a medida que aumenta o número de grupos, chegando a zero quando cada elemento da base de dados é o seu próprio grupo. Porém deve-se considerar o compromisso entre a redução da dispersão e o número de grupos a serem formados, pois a partir de uma certa quantidade, grupos coerentes começam a ser decompostos, o que não traz informação relevante e torna mais difícil a análise dos resultados.

O critério de convergência para finalizar o algoritmo, quando uma iteração completa não altera algum centróide, pode falhar, e o procedimento pode ficar oscilando indefinidamente entre duas ou mais partições e nunca convergir. Testes mostram que o número de dados reagrupados a cada iteração não diminui necessariamente ao longo do tempo, e portanto esse tipo de regra de parada pode prevenir o algoritmo de chegar a um ótimo global (WISHART, 2003).

A convergência pode ser garantida ao se computar um teste de relocação pela dispersão, quando um dado é movido para outro grupo apenas se a troca reduz a dispersão em ambos os grupos. Como a dispersão é calculada pela soma das distâncias dos centróides, ela não pode reduzir indefinidamente, e portanto deve convergir. Dessa forma é mais provável que o processo chegue a um ótimo global.

## 3.2. Mapas Auto-Organizáveis

### 3.2.1. História

Os Mapas Auto-Organizáveis são uma categoria de Rede Neural Artificial (RNA), caracterizados por um processo de “aprendizagem” não supervisionado. As RNAs são uma tentativa de descrever alguns fenômenos biofísicos que acontecem nos neurônios biológicos reais, que se imagina serem os elementos básicos responsáveis pelo processamento de informações no cérebro. Não passam de simples modelos matemáticos (neurônio) que interagem de acordo com uma topologia definida (rede), nem perto do complexo sistema dinâmico formado pelas células neurais, pois essas são controladas por uma diversidade de sinais como pequenos campos elétricos, numerosos transmissores químicos e moléculas mensageiras. Dentre as pesquisas realizadas em RNA destacam-se três tipos de abordagem, modelos, paradigmas e métodos.

Os modelos consistem de um conjunto finito de variáveis e suas interações quantitativas que devem descrever, como estados ou sinais de um sistema real, e geralmente assumem supostas regras simples da natureza.

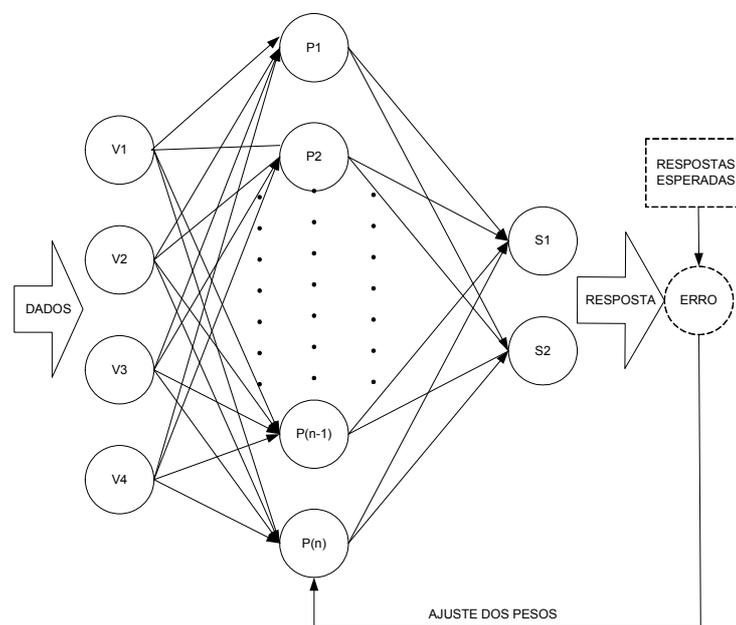
Um paradigma é uma forma tradicional ou aproximação simplificada de problemas complexos, e normalmente direciona a seleção de variáveis e processos a serem estudados, configuração de modelos e interpretação dos dados estudados.

Um método pode seguir modelos e paradigmas, ou ser desenvolvido sem nenhuma consideração a estes, e não há outra justificativa para um método senão a de que funciona com eficácia para certa aplicação.

Algumas visões teóricas sobre o pensamento que podem ser consideradas como modelagem já eram concebidas pelos filósofos gregos, especialmente Aristóteles. Outros que tiveram uma visão mecanicista do sistema nervoso foram os filósofos empíricos do século XVI, como Descartes.

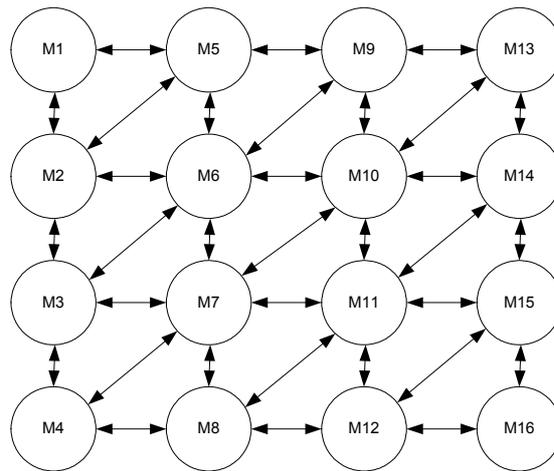
No início do século XX descobriu-se o funcionamento básico dos neurônios biológicos, como o disparo de impulsos elétricos através de suas densas conexões. Este princípio foi utilizado nos primeiros modelos matemáticos na década de 40, onde vários elementos interagem formando uma soma ponderada de sinais que produz uma resposta quando essa soma excede um valor limite. Esse paradigma foi primeiramente empregado em circuitos analógicos.

Então na década de 70, com o avanço da computação começou-se a explorar algoritmos que utilizavam tal paradigma em redes encadeadas, onde um estímulo passa por camadas de modelos até que a resposta seja conhecida (Figura 4). Os pesos de ponderação dos modelos são ajustados por retro-propagação, quando são apresentados estímulos (dados) com resposta conhecida à rede, e o erro é retornado para ser compensado, esta fase é conhecida por treinamento. Após a rede ter sido treinada com uma base de dados, ela passa a ter, por exemplo, a capacidade de interpolação das respostas esperadas. Esse método ficou conhecido como Perceptrons Multicamada (*Multi-Layer Perceptrons - MLP*), e é o tipo de RNA mais difundido.



**Figura 4.** Rede Neural do tipo Perceptron.

A partir dos anos 80 alguns estudos voltaram-se para redes de aprendizado competitivo, inspiradas na configuração celular do tecido cerebral, onde todos os neurônios disputam o mesmo estímulo e apenas o vencedor dispara a resposta. Nesse método não há um processo de retro-propagação, o que significa que o treinamento da rede não é supervisionado, sendo que os modelos são atualizados pela informação que cada um recebe. Esse tipo de RNA geralmente é formada por uma rede bidimensional com vários neurônios interligados (Figura 5), e por um processo de organização, que será descrito a seguir, dispondo os dados topologicamente, caracterizando um mapa que reconstrói a estrutura hierárquica dos dados.



**Figura 5.** Rede Auto-Organizável.

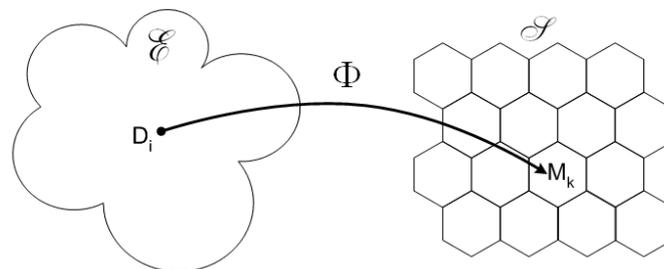
O algoritmo dos Mapas Auto-Organizáveis, ou SOM (*Self-Organizing Maps*) como são mais conhecidos, foi introduzido na literatura pelo professor Teuvo Kohonen em 1982, e sua monografia está na terceira edição (KOHONEN, 2001). O método foi rotulado inicialmente como uma heurística para classificação de dados, principalmente pela quantidade de parâmetros a serem ajustados no algoritmo, que o tornava surpreendentemente resistente a um estudo matemático completo e, até então, não havendo prova de sua convergência (COTTRELL *et al.*, 1995).

No início da década de 90 começaram a surgir trabalhos apresentando o SOM como um método eficiente de classificação e agrupamento, para reconhecimento de padrões e análise de dados (TÖRMÄ, 1994). Este método apresentou vantagens sobre os outros, especialmente em se tratando de dados com grandes dimensões e sem conhecimento a priori.

Hoje em dia, o SOM já é um método amplamente utilizado nas mais variadas áreas que exigem classificação de dados, tendo mais de 3500 artigos publicados até 1997 (KASKI *et al.*, 1998), o que levou a criação de eventos como o WSOM (*Workshop on Self-Organizing Maps*) dedicado exclusivamente às suas características e aplicações. A seguir serão apresentados alguns trabalhos com este algoritmo.

### 3.2.2. O Algoritmo

Em (HAYKIN, 2001) o algoritmo SOM é definido como uma transformação  $\Phi$  não-linear, que mapeia um espaço de entrada  $\mathcal{E} \in \mathbb{R}^n$  contínuo, cuja topologia é definida pela relação métrica dos vetores de dados  $D \in \mathcal{E}$ , para um espaço de saída  $\mathcal{S}$  discreto<sup>†</sup>, cuja topologia é definida por um conjunto de modelos  $M$  dispostos em uma grade bidimensional, como mostra a Figura 6.

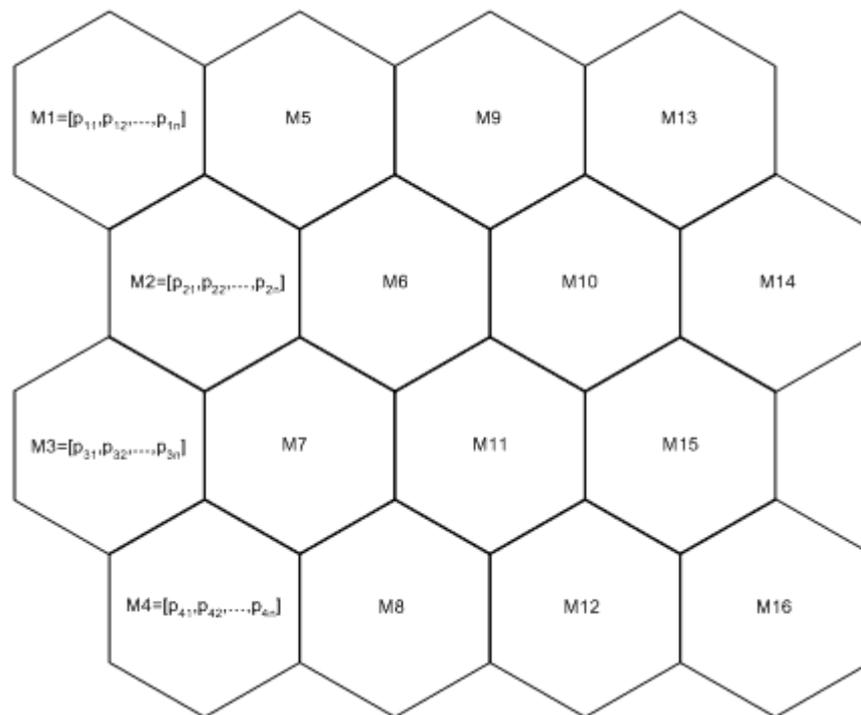


**Figura 6.** Transformação  $\Phi: \mathcal{E} \rightarrow \mathcal{S}$ . (adaptada de HAYKIN)

“O principal objetivo do Mapa Auto-Organizável é transformar um padrão de sinal incidente de dimensão arbitrária em um mapa discreto uni- ou bi-dimensional e realizar esta transformação adaptativamente de uma maneira topologicamente ordenada” (HAYKIN, 2001).

O padrão de sinal incidente é um dado de entrada, composto por um vetor de dimensão arbitrária “n” (variáveis). O mapa é formado por neurônios, contendo modelos que também são vetores de dimensão “n” (pesos sinápticos), conectados entre si como uma rede (Figura 7). O processo de ordenação acontece em três etapas e consiste em encontrar o modelo que melhor represente um dado de entrada na rede (vencedor) . As ligações servem para ajustar modelos de neurônios próximos ao vencedor (vizinhos) ao padrão de entrada. Assim, ao final de várias iterações o mapa apresenta os dados associados aos seus neurônios, de forma que padrões similares encontram-se em neurônios contíguos.

<sup>†</sup> Que não está em continuidade com a vizinhança; separado, distinto: pontos discretos.



**Figura 7.** Mapa ou Rede Hexagonal com 16 Modelos (4x4)

A Figura 7 apresenta um mapa bidimensional (4x4), com um total de 16 neurônios, onde cada um tem uma forma hexagonal que permite interagir com até 6 vizinhos. Por exemplo, o primeiro neurônio contém um modelo que é um vetor com “n” variáveis ( $M1=[p_{11}, p_{12}, \dots, p_{1n}]$ ) e se comunica com os modelos M2 e M5; já o modelo M6 tem como seus primeiros vizinhos os modelos M2, M7, M11, M10, M9 e M5.

Outra forma de rede normalmente utilizada é a retangular, onde um neurônio pode ter no máximo quatro vizinhos, o que diminui a cooperação entre neurônios.

Depois de definir o tamanho e o tipo da rede, deve-se proceder com a inicialização do mapa, que é o primeiro passo do algoritmo de auto-organização. Existem duas formas de montar a configuração inicial dos modelos:

- Aleatória
- Linear

Na forma aleatória os pesos sinápticos de cada modelo são distribuídos por sorteio, podendo ser qualquer valor pequeno o suficiente para estar dentro da variância dos dados de entrada. Essa é uma maneira de comprovar a auto-organização do mapa, mas exige muito mais processamento.

A inicialização linear tem a intenção de impor alguma ordem já na criação do mapa, mesmo que alguns modelos fiquem fora da distribuição de probabilidade da base de dados. O método proposto em (KOHONEN, 2001) consiste em determinar os dois autovetores da matriz de autocorrelação da base de dados que possui os maiores autovalores, e então expandir um subespaço linear de duas dimensões sobre os neurônios da grade. Assim, o mapa já começa de certa forma organizado, e consegue atingir uma configuração estável em muito menos iterações do que com uma inicialização aleatória.

Iniciado o mapa passa-se para a fase de treinamento, quando os pesos sinápticos serão ajustados aos padrões de entrada para formar o espaço de saída discreto. Esta fase está dividida em três partes:

- **O Processo Competitivo**

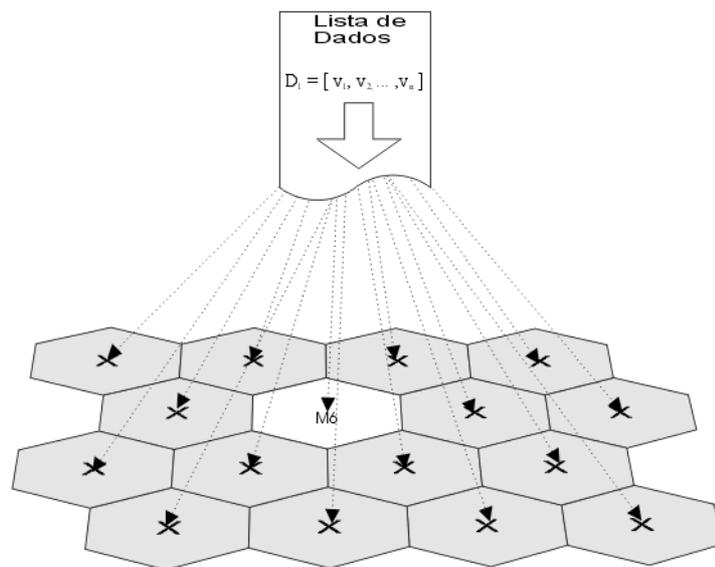
Sendo uma base de dados com “m” elementos de “n” variáveis ( $v$ ), um vetor de dado de entrada ( $t$ ) é caracterizado por:

$$\mathbf{D}_t = [v_{t1}, v_{t2}, \dots, v_{tm}], \quad t=1,2,\dots,m \quad (14)$$

Numa rede contendo “x” neurônios, o vetor do modelo de pesos sinápticos ( $p$ ) de cada neurônio ( $k$ ) tem a mesma dimensão que o espaço de entrada ( $n$ ):

$$\mathbf{M}(k) = [p_{k1}, p_{k2}, \dots, p_{kn}], \quad k=1,2,\dots,x \quad (15)$$

Busca-se o modelo que melhor represente o vetor de entrada comparando o produto interno  $\mathbf{M}(k) \cdot \mathbf{D}_t^T$  para  $k$  variando de 1 a “x”, e seleciona-se o maior. A posição do modelo vencedor na rede será o centro da função de vizinhança topológica, que definirá o processo cooperativo a seguir.



**Figura 8.** Representação do processo competitivo, todos os modelos recebem a mesma informação (D1), mas apenas um é ativado (M6).

- **O Processo Cooperativo:**

Após ser definido o centro, a função de vizinhança deve ajustar com maior ênfase os neurônios imediatamente vizinhos ao vencedor e decair com a distância lateral. Para isso essa função deve satisfazer o seguinte:

- ser simétrica em relação ao centro (foco)
- possuir amplitude máxima no centro, e decrescer com a distância lateral até zero, o que é uma condição necessária para a convergência.

Uma escolha típica para isso é a função gaussiana, que não varia com a translação:

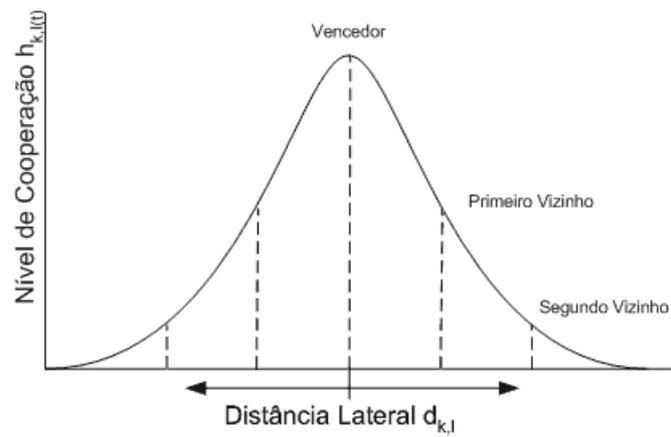
$$h_{k,l(t)}(i) = \exp\left(-\frac{d_{k,l}^2}{2.\sigma(i)^2}\right) \quad (16)$$

onde :

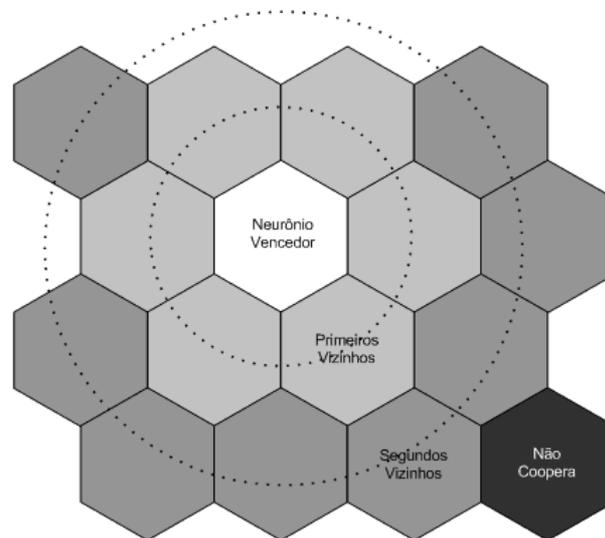
$h_{k,l(t)}(i)$  = função de vizinhança para o vizinho (k) em relação ao vencedor (l) para o dado (t) na iteração (i), determina o nível de cooperação;

$d_{k,l}$  = distância lateral entre o vizinho (k) e o neurônio vencedor (l);

$\sigma(i)$  = largura da vizinhança (abrangência) na iteração (i).



**Figura 9.** Um exemplo de função de vizinhança gaussiana.



**Figura 10.** Função de vizinhança aplicada em um mapa hexagonal 4x4.

A Figura 9 mostra uma típica curva gaussiana, onde a amplitude determina o nível de cooperação entre o vencedor e seus vizinhos, e a Figura 10 demonstra a abrangência da função de vizinhança em uma mapa 4x4 de topologia hexagonal.

Outra característica deste processo é que o tamanho da vizinhança topológica ( $\sigma$ ) deve diminuir ao longo das iterações. Uma escolha plausível para tal função é o decaimento exponencial, como em (17).

$$\sigma(i) = \sigma_0 \cdot \exp\left(-\frac{i}{\tau_1}\right) \quad (17)$$

onde:

$\sigma(i)$  = tamanho da vizinhança na iteração (i), arredondado para o inteiro superior;

$\sigma_0$  = vizinhança máxima inicial;

i = n° da iteração;

$\tau_1$  = constante de tempo.

- **O Processo Adaptativo:**

Esse é o último processo de uma iteração na formação auto-organizada de um mapa de características, quando o vetor de pesos sináptico  $\mathbf{M}_k$  é modificado em relação ao vetor de entrada  $\mathbf{D}_t$  através da seguinte expressão:

$$\mathbf{M}_k(i+1) = \mathbf{M}_k(i) + \alpha(i) \cdot h_{k,l(t)}(i) \cdot (\mathbf{D}_t - \mathbf{M}_k(i)) \quad (18)$$

onde :

$\mathbf{M}_k(i+1)$  = Modelo k para adaptado para a próxima iteração;

$\mathbf{M}_k(i)$  = Modelo k na iteração atual (i);

$\alpha(i)$  = taxa de aprendizagem na iteração (i);

$h_{k,l(t)}(i)$  = Nível de cooperação na iteração (i);

$\mathbf{D}_t$  = Dado de entrada (t).

A taxa de aprendizagem também deve diminuir ao longo das iterações, o que corresponde ao caso da aproximação estocástica, começando de um valor inicial  $\alpha_0$  e decrescendo gradualmente com o aumento das iterações (i). Aqui também pode ser aplicado o decaimento exponencial como mostra a equação a seguir.

$$\alpha(i) = \alpha_0 \cdot \exp\left(-\frac{i}{\tau_2}\right) \quad (19)$$

onde:

$\alpha(i)$  = taxa de aprendizagem na iteração (i);

$\alpha_0$  = taxa de aprendizagem na primeira iteração;

$\tau_2$  = constante de tempo.

- **Parâmetros de Treinamento:**

Os parâmetros apresentados nos três itens anteriores podem receber uma infinidade de valores para formarem o processo de auto-organização, porém quando é feita a inicialização aleatória do mapa é considerável dividir o processo de treinamento em duas fases. A primeira, chamada de ordenação, onde a função de vizinhança abrange praticamente todo o mapa e a taxa de aprendizagem é alta, permite uma aproximação mais rápida de neurônios semelhantes e já traça a característica topológica do mapa. E a segunda, definida como fase de convergência, assim como seu nome sugere, busca uma sintonia fina do mapa utilizando um raio de vizinhança e uma taxa de aprendizagem bem menores durante um número de iterações muito maior que o da fase de ordenação.

Esses são alguns valores de ordem comum para serem usados em tentativas iniciais:

- durante a fase de ordenação ( $i \approx 1000$ ) :

$$\alpha_0 = 0,1$$

$$\tau_2 = 1000$$

$$\sigma_0 = \text{raio da grade}$$

$$\tau_1 = 1000 / \log(\sigma_0)$$

- durante a fase de convergência ( $i \geq 500$  vezes o n° de neurônios da grade):

$$\alpha_0 \approx 0,01 \text{ mas } \neq 0$$

$$\tau_2 = 1000$$

$$\sigma_0 \approx \text{pequeno comparado ao tamanho da grade}$$

$$\tau_1 = 1000 / \log(\sigma_0)$$

- **Observações sobre a Convergência**

Quando o algoritmo do SOM foi apresentado como um processo de organização não-supervisionado, capaz de formar um mapa de características ordenado a partir de um estado inicial qualquer, poucos acreditaram em tal façanha. O principal motivo desse descrédito era o fato de não haver prova matemática dessa habilidade, e que tudo não passava de uma heurística aplicada em alguns casos.

Porém, com a difusão da aplicação desse algoritmo para resolver problemas de classificação e agrupamento nas mais diversas áreas, viu-se que com quase qualquer quantidade ou qualidade de variáveis envolvidas no processo o SOM era bem sucedido. Mas ainda restava a dúvida sobre a qualidade de tais resultados. Isso chamou a atenção de muitos matemáticos, e surgiram vários trabalhos na tentativa de provar a convergência do processo de auto-organização do SOM.

Marie Cottrell, professora titular da Panthion-Sorbonne de Paris, foi pioneira na busca por comprovar a eficiência do algoritmo SOM, sendo hoje um dos maiores nomes da área, e organizadora do WSOM'05. Ela apresentou em (COTTRELL *et al.*, 2001) índices estatísticos para avaliar a estabilidade de configurações dos mapas. Como a dispersão intra-classe, que é a soma das distâncias dos dados agrupados em um neurônio, para seu modelo de pesos sinápticos (centróide). Isso ajuda a definir o tamanho do mapa para uma base de dados. Outro índice proposto é uma medida da preservação da topologia, que indica a qualidade de um mapa quanto a elementos próximos na base de dados, estarem próximos no mapa. Um exemplo é mostrado na Tabela 2 onde foi classificada em um mapa 7x7 uma base de dados real contendo seis variáveis macroeconômicas medidas em 1984 em diferentes países .

Por se tratar de uma classificação de países é possível validar os resultados apresentados na Tabela 2, pois espera-se que países com situação semelhantes possuam um valor alto para o índice, enquanto países com situações muito diferentes recebem valores baixos.

Foram utilizadas diferentes técnicas de inicialização e sorteio dos dados apresentados ao mapa, como *Monte-Carlo* e *Bootstrap*, e concluiu-se que o SOM é pouco sensível a problemas de mínimo local ou convergência devido a configurações iniciais.

**Tabela 2.** Índices de Estabilidade da Topologia para um mapa 7x7 classificando dados macroeconômicos de diferentes países

<b>País</b>	<b>Índice de Estabilidade</b>
Grécia / França	0.18
Austrália / França	0.82
Grécia / Bélgica	0.21
Turquia / França	0.02
Singapura / EUA	0.49
Suécia / Japão	0.73
Grécia / Malta	1
Canadá / França	0.84
Suécia / França	0.97
EUA / Zimbábue	0
EUA / Finlândia	0.85
EUA / Austrália	0.68

Muitos trabalhos foram publicados sobre a convergência do processo de auto-organização, e apesar da variedade de técnicas e abordagens utilizadas, nenhum conseguiu definir uma prova completa, sem recorrer a aproximações ou alterar características originais do algoritmo. O grande problema reside nas não-linearidades e descontinuidades dos parâmetros envolvidos em um tempo discreto. A seguir é apresentado um resumo das principais contribuições nesse sentido.

COTTRELL *et al.* (1995) comprova através de técnicas de cadeias de Markov que existem conjuntos de absorção para o caso de auto-organização em uma dimensão, assegurando a sua convergência. Mas não traça nenhuma prova para o caso bi-dimensional.

Seguindo o trabalho de Cottrell, mas propondo uma modificação do algoritmo de treinamento, em (YIN e ALLISON, 1995) é apresentada uma variante que consegue atingir uma convergência “quase certa” do processo de auto-organização partindo de qualquer condição inicial

É provado em (ERWIN *et al.*, 1992a) que a dinâmica do treinamento não pode ser descrita como a otimização de uma única função de energia, mas sim pela minimização independente de gradientes de funções potenciais para cada neurônio, quando longe do equilíbrio.

ERWIN *et al.*(1992b) mostra que para evitar estados estacionários, quando o mapa não consegue se organizar, é preciso utilizar uma função de vizinhança convexa. Assim, utilizando uma função do tipo gaussiana, o único estado em que o mapa pode se tornar

estável é quando estiver organizado. Além disso, é comprovado que funções de vizinhança mais abrangentes levam a uma convergência mais rápida (LO *et al.*, 1991a e 1991b).

Uma extensa formulação matemática é apresentada por LO *et al.* em (1992 e 1993) demonstrando dois teoremas sobre a convergência de mapas auto-organizáveis, com base no Teorema de Gladyshev que usa o método de funções de Liapunov. Esses teoremas definem o que é um estado ordenado em um mapa bidimensional e comprovam uma convergência assintótica. Em (SPITZER e POLANI., 1998) também são definidos parâmetros de organização; um envolve estados sinérgicos do mapa e não necessita de uma interpretação geométrica, enquanto o outro, plota um gráfico de níveis de organização a partir da taxa de aprendizagem e da largura da função de vizinhança, mostrando combinações que podem gerar topologias distorcidas.

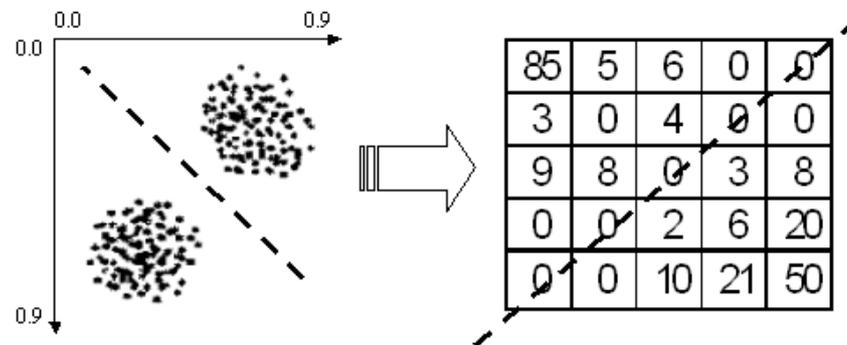
SADEGHI (2001) realizou um estudo sistemático das características do algoritmo de Kohonen sem se deter na dimensão do mapa ou na forma da função de vizinhança. O único artifício utilizado foi considerar uma taxa de aprendizagem constante, o que não degrada o desempenho do processo se for relativamente pequena. Assim, pode-se assumir a teoria de cadeias de Markov homogêneas e provar que independente da dimensão do mapa, pode-se atingir um estado estável a partir de qualquer estado inicial.

### 3.2.3. Visualização dos Mapas

A interpretação do resultado final do algoritmo de auto-organização é muito importante para a correta obtenção dos resultados, e para extrair o máximo de informações que esse método proporciona.

Como foi mostrado na Figura 6 no item 3.2.2, o mapa é um espaço de saída discreto de uma transformação injetora. Após o processo de organização, cada elemento da base de dados utilizada para treinar o mapa será representado por um, e apenas um, dos neurônios da rede. Porém, cada neurônio poderá representar qualquer quantidade de elementos, inclusive nenhum. Como o mapa representa a densidade de probabilidade de ocorrência dos elementos da base de dados de uma forma ordenada, quando existir uma distância muito grande entre grupos de elementos, ou elementos discrepantes isolados (*outliers*), se o mapa for relativamente grande, durante o processo de organização existirão alguns modelos que ficarão adaptados ao espaço que separa estes grupos, e portanto não conseguirão vencer nenhum dado de entrada.

A inspeção da densidade do mapa foi um dos primeiros métodos utilizados para agrupar neurônios a fim de definir grandes grupos de dados (XUEGONG *et al.*, 1993).



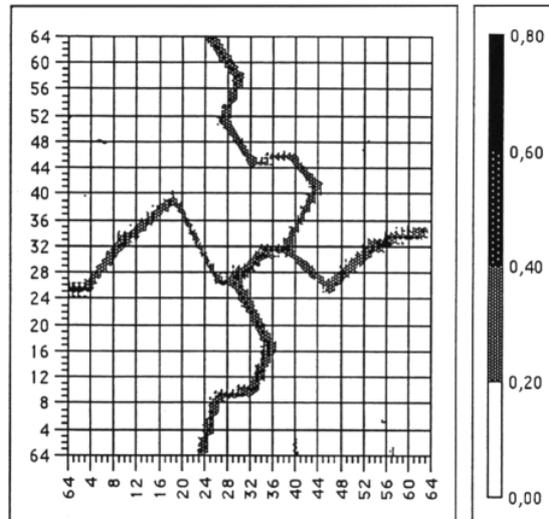
**Figura 11.** Agrupamento de dados pela densidade do mapa (adaptada de XUEGONG *et al.*)

Alguns autores como FLEXER (2001) consideram um comprometimento entre classificação e visualização de dados, considerando que o número de neurônios da rede deve ser igual ao número de grupos desejados para classificação, ou maior para extração de conhecimento. Também faz crítica ao mapa por não preservar a estrutura dos dados, já que é uma projeção em um espaço de duas dimensões. Porém, o fato de manter a estrutura dos dados nunca foi a intenção do algoritmo, e sim fazer uma organização hierárquica sobre um plano (mapa). Quanto à suposição de que o número de neurônios deve ser igual ao número de grupos desejados está completamente equivocada, pois isso tornaria o algoritmo dependente de uma decisão *a priori*, como no algoritmo das *k*-médias, e a grande vantagem do SOM é justamente a descoberta de padrões.

Para utilizar um Mapa Auto-Organizável tanto para agrupamento de dados, quanto para descoberta de padrões ou extração de conhecimento é preciso preparar no mínimo duas ferramentas, a matriz de distâncias e os mapas componentes.

A Matriz de Distâncias Unificada, ou U-Mat como é mais conhecida, foi apresentada por ULTSCH e SIEMON (1990) como uma forma de definir o agrupamento de neurônios no mapa. Essa matriz contém a distância euclidiana (1) entre neurônios vizinhos, medida entre modelos que fazem ligação. Ao plotar essas distâncias em escala de cor, ou em nível, poderá se observar regiões de mesma cor ou nível, agrupando neurônios, e faixas de cor diferente ou picos, representando uma fronteira entre grupos.

A Figura 12 a seguir mostra um exemplo de U-Mat para um mapa 64x64 usado para classificar uma base de dados gerada nos vértices de um tetraedro em um espaço tridimensional (ULTSCH e SIEMON, 1990).



**Figura 12.** Matriz de distâncias para uma base de dados em forma de tetraedro (ULTSCH e SIEMON, 1990).

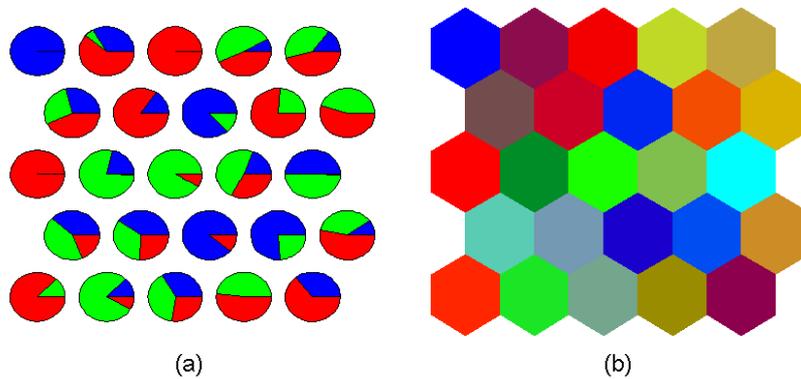
As regiões claras indicam uma distância euclidiana pequena (de 0 a 0,2) enquanto que as escuras são maiores ( $> 0,2$ ), formando quatro áreas, ou seja, quatro grupos de neurônios. Isso mostra a classificação correta de uma base de dados gerada propositalmente com quatro grupos, validando o uso da U-Mat em conjunto com o SOM para agrupamento de dados.

Uma forma de verificar o que levou à formação de cada agrupamento pela U-Mat é verificar os mapas componentes, que mostram a participação de uma variável em cada neurônio do mapa em escala de cores. Juha Vesanto, membro do Laboratório de Computação e Ciências da Informação da Universidade de Tecnologia de Helsinki na Finlândia fundado pelo professor Teuvo Kohonen, elaborou uma série de métodos para visualização e caracterização dos agrupamentos formados pelo SOM em (VESANTO,1999) e (VESANTO, 2002).

A seguir será apresentado um exemplo da ordenação realizada pelo algoritmo SOM e algumas formas de visualização para extração das características que levaram a tal organização.

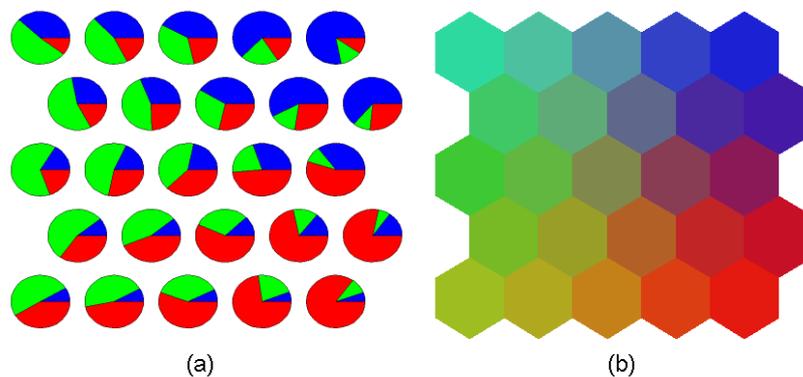
As cores exibidas na tela do computador são uma composição de três cores básicas, vermelho (R – *red*), verde (G – *green*) e azul (B – *blue*), cada uma podendo variar a proporção de 0 a 255, logo suas combinações resultam em mais de 16,5 milhões de cores.

Para mostrar o poder de organização do SOM, foi gerada uma seqüência aleatória de 25 vetores de 3 variáveis (R-G-B) como base de dados do espaço de entrada, e um mapa hexagonal 5x5 para o espaço de saída como mostra a Figura 13.



**Figura 13.** Condição inicial do mapa 5x5; (a) gráficos “pizza” mostrando a quantidade de cada cor nos modelos iniciais (RGB); (b) configuração inicial do mapa hexagonal com as composições de cores .

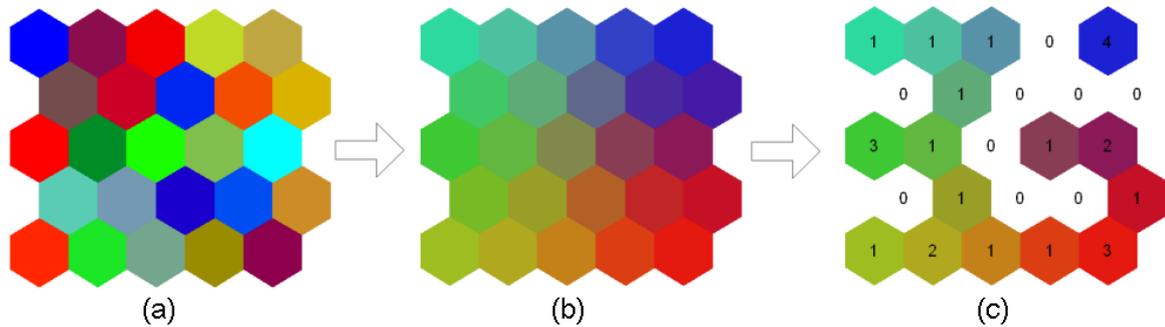
Partindo de uma configuração aleatória dos modelos, seguiu-se o processo de alto organização. O mapa final, treinado, ficou da seguinte forma:



**Figura 14.** Condição final do mapa treinado; (a) gráficos “pizza” mostrando a quantidade de cada cor nos modelos organizados (RGB); (b) configuração final do mapa hexagonal com as cores somadas.

Observando a Figura 14, em (b) pode-se notar a organização realizada, quando as cores mais puras estão nos extremos do mapa, e vão se misturando suavemente. No centro, tem-se uma proporção quase igual das três cores, vide (a).

A disposição dos valores iniciais sobre o mapa treinado, ou seja, a quantidade de dados de entrada vencidos por cada neurônio, representa a densidade do espaço de estados e sua organização hierárquica (Figura 15(c)) em apenas um plano. A posição de cada neurônio é a mesma em todos os mapas, os seus modelos é que são adaptados de maneira a formar um estado organizado do mapa.



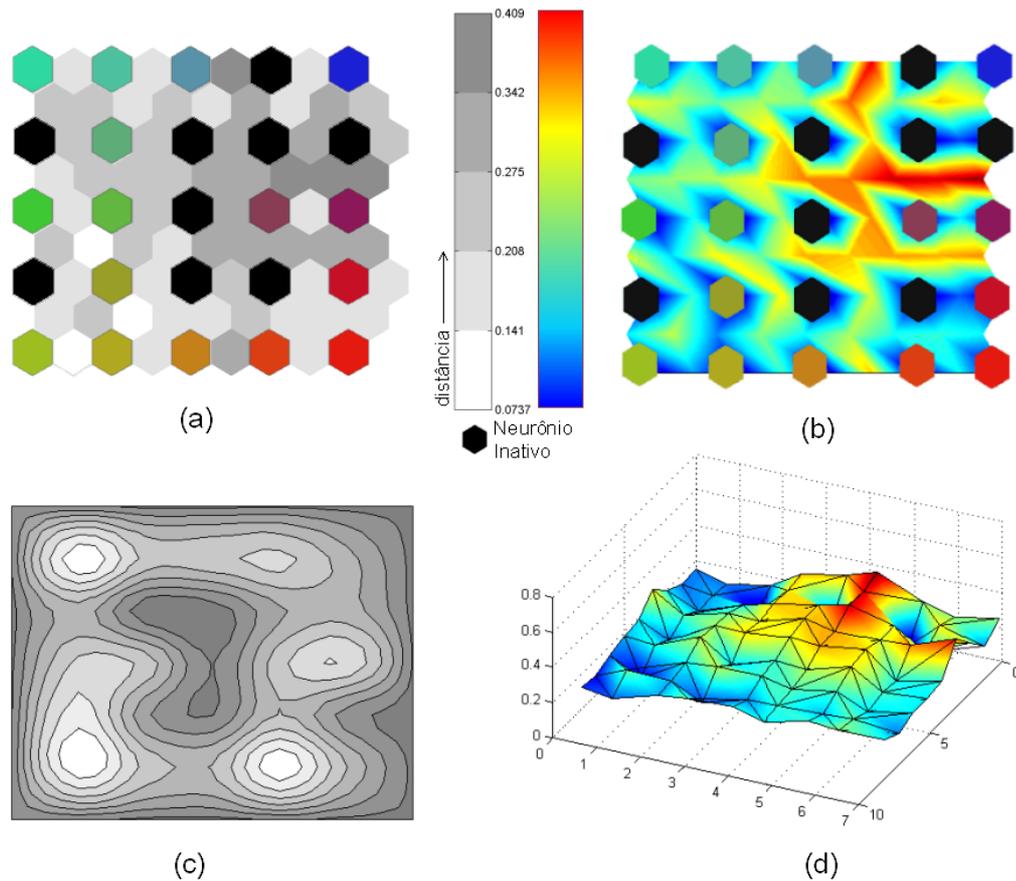
**Figura 15.** As três fases do processo de auto-organização;(a) base de dados / configuração inicial; (b) mapa treinado; (c) distribuição de densidade.

Fazendo uma analogia, observando as três fases da Figura 15, (c) surge como se fossem peneiradas as cores originais de (a) através do mapa (b). Observe o neurônio isolado no canto superior direito de (c) contendo 4 elementos, e agora procure por 4 tons de azul similares em (a). Tente fazer o mesmo com os tons de (b) que ficaram vazios em (c).

Porém o fato de neurônios serem vizinhos, não implica que os dados alocados neles estejam tão próximos. Para verificar isso é necessário observar a matriz das distâncias entre neurônios (U-Mat). Essa matriz tem a dimensões  $2*N_l-1$  por  $2*N_c-1$ , onde  $N_l$  e  $N_c$  são o número de neurônios em linha e em coluna, respectivamente. Para o exemplo RGB em questão, a matriz U é a seguinte:

$$\mathbf{U} = \begin{bmatrix} 0.1997 & 0.1543 & 0.1919 & 0.1929 & 0.2116 & 0.3663 & 0.2647 & 0.1654 & 0.2467 \\ 0.2451 & 0.2280 & 0.1908 & 0.2179 & 0.2053 & 0.3258 & 0.2036 & 0.2830 & 0.2467 \\ 0.2132 & 0.1753 & 0.2203 & 0.2788 & 0.2948 & 0.2681 & 0.2755 & 0.0737 & 0.2467 \\ 0.1987 & 0.2132 & 0.2228 & 0.2551 & 0.3108 & 0.3144 & 0.3870 & 0.3777 & 0.4089 \\ 0.1987 & 0.1599 & 0.2160 & 0.2188 & 0.2746 & 0.3028 & 0.3013 & 0.1436 & 0.3050 \\ 0.2409 & 0.1366 & 0.2592 & 0.1908 & 0.2941 & 0.2792 & 0.2998 & 0.2922 & 0.3050 \\ 0.1706 & 0.1706 & 0.1986 & 0.2619 & 0.2501 & 0.2382 & 0.2045 & 0.0832 & 0.1525 \\ 0.1457 & 0.2307 & 0.1043 & 0.2064 & 0.1547 & 0.2199 & 0.1632 & 0.1707 & 0.1525 \\ 0.1280 & 0.1103 & 0.1424 & 0.1744 & 0.1904 & 0.2794 & 0.1915 & 0.1505 & 0.1525 \end{bmatrix} \quad (20)$$

Os elementos da matriz  $U$  em (9) que estão em tom mais claro representam a mediana das distâncias entre neurônios vizinhos (preto). Dependendo do tamanho do mapa, inspecionar essas distâncias não é uma tarefa simples. Então para facilitar a identificação da proximidade entre neurônios podem ser usadas diferentes maneiras de visualizar a  $U$ -Mat. Quatro tipos de gráficos com escala de cores, níveis e relevo são mostrados na Figura 16.



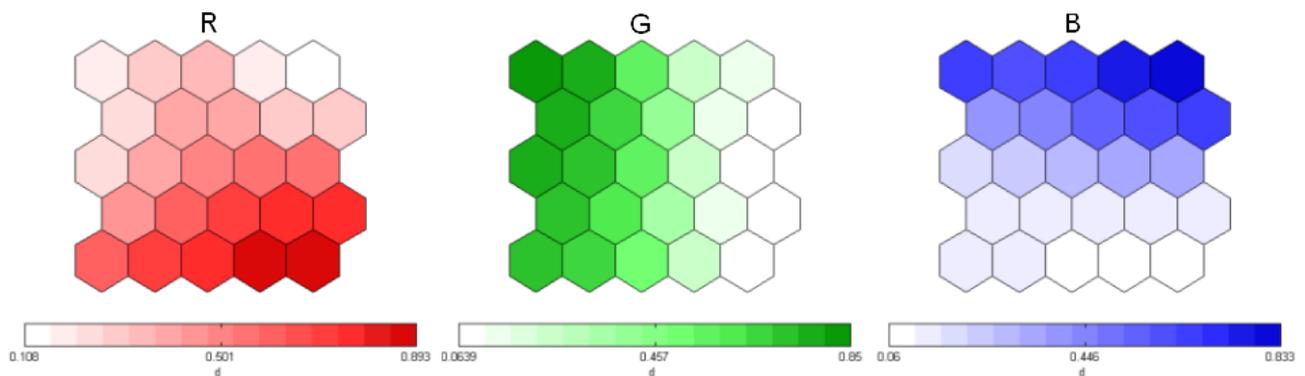
**Figura 16.** Quatro tipos de gráficos para visualizar a matriz de distâncias; (a) com as distâncias individuais em 5 tons de cinza; (b) com uma superfície colorida em 256 tons; (c) histograma com 7 níveis; (d) superfície colorida em 3 dimensões.

A Figura 16 mostra em (a) e (b) como o neurônio no canto superior direito (hexágono azul) está longe dos demais, e também que há uma distância significativa entre os dois neurônios na região centro-direita (hexágonos bordô/vinho) para os mais abaixo (hexágonos vermelhos). Os tons com maior participação do verde estão mais próximos, mas há uma pequena fronteira entre a segunda e terceira linha.

A Figura 16(c) mostra um método de visualização de agrupamentos proposto por PAMPALK *et al.*(2002), que utiliza histogramas de frequência suavizados para estimar a densidade de probabilidade e determinar os centros dos agrupamentos de neurônios. Nesse caso, utilizando sete níveis de contorno detecta-se a presença de cinco regiões de agrupamento.

Para visualizar o relevo em três dimensões Figura 16(d) é necessário plotar as coordenadas da rede em um plano (eixos  $x / y$ ) e utiliza as distâncias da U-Mat como altura (eixo  $z$ ), assim como fronteiras geográficas, picos separam neurônios, e vales ou planícies os mantêm próximos.

Além dos gráficos “pizza”, pode-se explorar individualmente a distribuição das variáveis através dos mapas componentes. Esses mapas mostram em escala de cores a quantidade da variável em questão em cada neurônio do mapa. Como a posição dos neurônios é a mesma em todos os mapas, pode-se procurar por correlações entre variáveis. Mapas componentes similares indicam variáveis com correlação forte.

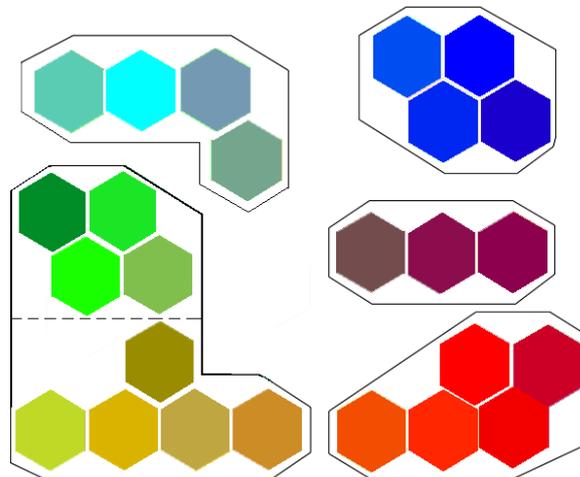


**Figura 17.** Mapas componentes das variáveis R,G e B.

Na Figura 17 pode-se observar a distribuição de cada variável sobre o mapa. Por exemplo, a componente R (vermelho) domina a região inferior direita do mapa, praticamente oposta ao G (verde) que se concentra na parte superior esquerda, enquanto o B (azul) tem maior influência nos neurônios do canto superior direito. Note que a intensidade de cada variável segue um gradiente, e que nos neurônios onde há a maior intensidade de uma cor quase não há participação das outras.

Na análise dos mapas componentes está a propriedade de abstração<sup>‡</sup> do SOM. Para a busca de correlações entre variáveis se analisam os mapas inteiros, e ao se comparar as intensidades em cada neurônio pode-se encontrar características dos dados ali alocados. Os gráficos “pizza” de cada neurônio resumem os mapas componentes, o que pode ser observado comparando a Figura 14(a) com a Figura 17.

Somando todas essas análises anteriores pode-se definir os grupos de dados. Portanto, temos os dados agrupados em um mesmo neurônio, mais os grupos de neurônios com base na distância de seus modelos e/ou nas variáveis componentes. A Figura 18, a seguir, mostra o arranjo dos grupos com os dados RGB usados para treinar o mapa. Pela matriz de distâncias poderiam ser cinco grupos, mas pela análise dos mapas componentes ainda é possível partir o grupo maior à esquerda em dois (linha tracejada), separando os elementos de verde mais puro dos com maior participação de vermelho.



**Figura 18.** Agrupamento dos dados RGB.

---

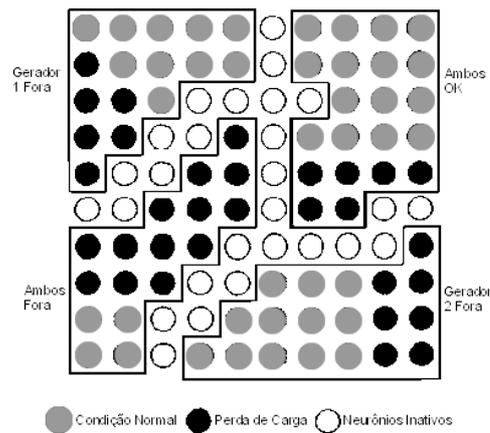
<sup>‡</sup> Ato de separar um ou mais elementos de uma totalidade complexa

### 3.2.4. Aplicações do SOM

Os Mapas Auto-Organizáveis (SOM) se tornaram um poderoso método para visualização de dados com grandes dimensões, convertendo complexas relações estatísticas não-lineares entre esses dados em simples relações geométricas, ou topológicas, em uma superfície de pequena dimensão. Conseqüentemente existe uma compressão de informação, mas sem perder a métrica ou hierarquia da relação entre os elementos expostos no mapa, e podendo também abstrair características intrínsecas da relação das variáveis envolvidas. Estes aspectos, visualização e abstração, são necessários em uma quantidade de aplicações, especialmente na engenharia, como análise de processos, reconhecimento, controle e comunicações (KOHONEN, 1996).

Um dos primeiros trabalhos utilizando o SOM na área de Sistemas de Potência foi a classificação de curvas de carga diárias. HSU *et al.* (1991) fizeram a identificação de dias típicos (24h) por dados históricos, e posteriormente usaram o mapa treinado para previsão da carga ao longo do dia com 5h de antecedência para a Companhia de Eletricidade de Taiwan. Os autores destacam a facilidade e agilidade do algoritmo SOM em agrupar as curvas e detectar dias atípicos, facilitando a decisão dos operadores. CHANTELOU *et al.* (1996) também utilizaram o SOM para agrupar 2665 curvas de carga diárias de 205 clientes da EDF a fim de tarifação diferenciada, é destacada a capacidade de classificação hierárquica e visualização dos grupos em um mapa 10x10, que mostra 100 grupos de curva em apenas uma folha de papel A4.

LUO *et al.* (1999) utilizaram o SOM para verificar estados de instabilidade e perda de carga em Sistemas de Potência. Foi apresentado um sistema teste de 9 barras, 3 geradores e 3 cargas. O vetor de entrada era composto de 8 variáveis, as potências ativas e reativas nas 3 cargas e duas potências geradas, assumindo que a terceira era a barra de folga e estava sempre disponível. Então foram gerados 208 estados diferentes, variando a carga e a disponibilidade dos geradores, para treinar um mapa 10x10. Com o mapa treinado, foram utilizados os modelos de cada neurônio como parâmetros para um Fluxo de Potência Ótimo, a fim de detectar estados de perda de carga. Assim, os neurônios foram classificados em estados de condição de operação normal, e com perda de carga. Ao visualizar a saída do mapa, pode-se notar que a topologia é ordenada, e estados de operação normal são sucedidos por perda de carga, como mostra a Figura 19.



**Figura 19.** Mapa 10x10 treinado para determinar estados de operação de um sistema 9 barras, 3 geradores e 3 cargas.

Mais interessante seria observar os mapas componentes relativos à Figura 19, que mostrariam o porquê da transição de uma condição para outra, mas esses não foram apresentados no artigo.

Dois trabalhos brasileiros utilizando SOM e variáveis climáticas são apresentados em (LIMA *et al.*, 1999, e COELHO *et al.*, 2002). No primeiro foram classificadas curvas de cargas, a fim de separar perfis de carga associados a variáveis climáticas como nebulosidade, temperatura e insolação. Foram determinadas variáveis exógenas do modelo de previsão tradicional, e definido um perfil mais adequado para cada trecho de curva de carga. No segundo trabalho, o SOM é aplicado a duas bases de dados separadas, uma de variáveis climáticas como chuva, vento e temperatura, e outra de variáveis técnicas como DEC, FEC, TMA e número de grupos de atendimento. É feito o agrupamento de dados em cada base, e posteriormente o cruzamento de grupos mostra a correlação entre o clima e ocorrências na rede elétrica.

Outras aplicações, fora da engenharia elétrica são, classificar sinais sonoros em ambientes com ruído e interferência (EMANIAN *et al.*, 2000) e identificar interlocutores em uma conversa, nesse caso, VOITOVETSKY *et al.* (1997) utilizou mapas diferentes para serem treinados com a voz de cada interlocutor, obtendo mais de 90% acerto.

Os Mapas Auto-Organizáveis tornaram-se um dos métodos mais eficientes para análise de dados sobre expressão gênica (AZUAJE, 2002). O SOM não é capaz apenas de classificar categorias de genes ou células, mas também de um processo de descobrimento e identificação de subtipos relevantes de classes.

Outras áreas de pesquisa que o SOM tem participado significativamente são Finanças, Economia e Marketing. Algumas aplicações citadas por DEBOEK (1999) são:

- análise financeira inicial
- previsão financeira
- análise de oportunidade de investimentos
- alocação estratégica de portfólios
- previsão de indicadores macro-econômicos
- mapeamento sócio-econômico
- classificação de consumidores
- segmentação de mercados

Esse trabalho apresenta um exemplo em que 50 fundos de investimentos, considerados como sendo do mesmo tipo por uma agência financeira, foram classificados pelo algoritmo SOM em três grupos diferentes. Detalhes mostrando o motivo da diferença entre os três grupos são comentados.

Outro estudo similar feito por CARDON *et al.*(2002) classifica fundos de investimento por SOM. Foram analisadas 33 características de 5822 fundos através de um mapa 10x10, que depois de treinado ainda passou seus modelos pelo algoritmo de Ward para agrupar neurônios, resultando em 20 grupos. Esses grupos foram então comparados com uma classificação padrão, e o método proposto conseguiu reduzir a dispersão intra-grupo em quase 50%.

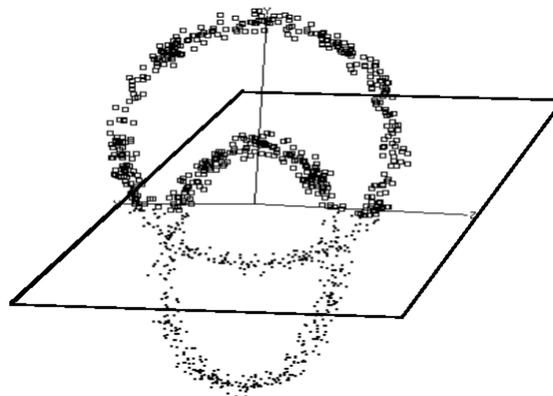
Por fim, vale citar outro trabalho de ULTSCH *et al.*(1994) que apresenta um método para integração dos Mapas Auto-Organizáveis com Sistemas Especialistas e/ou algoritmos de Lógica Difusa, onde a matriz de distâncias gerada pelo SOM é utilizada para extração de regras, ou conhecimento, de uma base de dados. São citados exemplos de diagnósticos clínicos como de anemia e dosagem de insulina para diabéticos, e controle de processos industriais como destilação ou controle de qualidade.

### **3.2.5. Comparação do SOM com outros Métodos de Classificação**

Com a crescente utilização de novas técnicas como redes neurais e lógica difusa para classificação de dados surgiu a necessidade de avaliá-las em relação às técnicas estatísticas tradicionais.

Um dos primeiros trabalhos nesse sentido foi realizado por TÖRMA (1993), comparando o algoritmo SOM com o k-médias tradicional e um k-médias adaptado com lógica difusa. Foram testados três tipos diferentes de base de dados, uma com distribuição normal em 3-dimensões, outra não-normal com dois conjuntos distintos em 2-dimensões, e uma fotografia de satélite. No primeiro caso, o SOM e o k-médias tiveram um resultado muito próximo, porém o k-médias apresentou um valor de dispersão intra-grupo um pouco menor, enquanto o algoritmo difuso obteve uma dispersão maior. Para a base com dois conjuntos, o algoritmo difuso identificou quase 100% dos dados corretamente, enquanto o SOM e o k-médias tiveram um desempenho significativamente pior. No caso da imagem, o SOM obteve o melhor resultado, mas o algoritmo de lógica difusa também conseguiu formar grupos concisos, como regiões de florestas e gramado, enquanto o k-médias não foi capaz de encontrar tal similaridade, misturando porções de imagens muito diferentes em um mesmo grupo. Por fim, o autor concluiu que a escolha do algoritmo depende muito da estrutura da base de dados, mas caso não haja um conhecimento dessa estrutura, o SOM é a melhor escolha, por se adaptar bem aos diferentes casos.

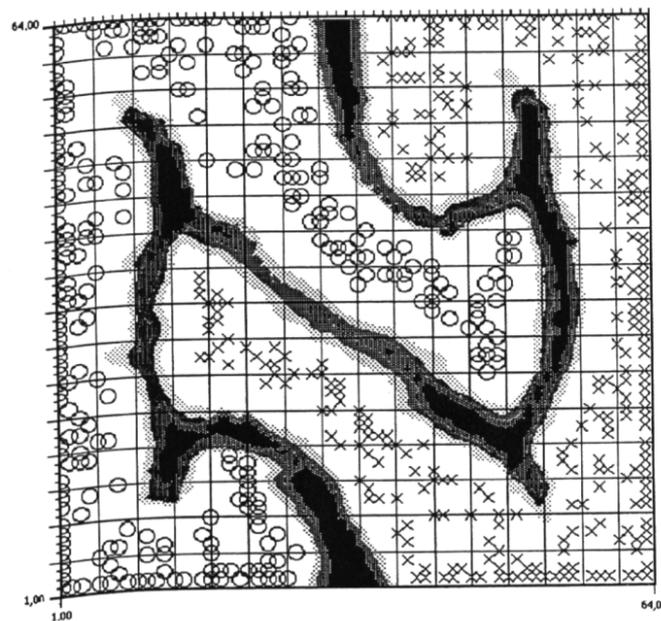
Outra comparação entre o SOM e o k-médias foi realizada por ULTSCH (1996), utilizando uma base de dados composta por duas nuvens de pontos em forma de anéis entrelaçados, mas não se tocando, em um espaço de 3-dimensões (Figura 20).



**Figura 20.** Anéis entrelaçados, e tentativa do k-médias em dividí-los em dois grupos

A Figura 20 mostra um resultado do algoritmo das k-médias ao dividir os anéis de pontos em dois grupos, como se passasse um hiperplano entre eles. Outras tentativas com maior número de grupos também não foram corretas em separar os anéis, o que expõe uma deficiência do método para classificar dados com uma distribuição não-linear.

Para o método SOM, foi utilizada uma rede 64x64 e o resultado da matriz de distâncias (U-Mat) é exibido na Figura 21.



**Figura 21.** Matriz de distâncias para uma rede 64x64 treinada com a base de dois anéis.

A Figura 21 revela uma forte divisão, a área escura que representa uma grande distância entre neurônios, entre os pontos de um anel e outro (O e X), mostrando que o SOM conseguiu obter uma separação correta dos grupos pré-definidos.

Apesar desse exemplo mostrar uma superioridade do algoritmo SOM em lidar com esse tipo de estrutura de dados, não quer dizer que seja melhor que o método das k-médias, pois podem existir casos em que aconteça exatamente o contrário. Mas o importante é ter em mente que diferenças existem, e que é preciso investigar as vantagens de cada algoritmo para cada caso.

Outro exemplo de comparação entre redes neurais e estatística é apresentado em (SHUMANN, 1997), onde é estudada uma base de dados real extraída de características geológicas. Nos testes, a Rede Neural obteve uma leve vantagem sobre métodos de Análise Discriminante paramétricos e não-paramétricos.

Uma boa referência para se estudar os diferentes tipos de algoritmos para classificação de dados, tanto estatísticos quando os que utilizam princípios de inteligência artificial, é o livro “Machine Learnig, Neural and Statistical Classification” (MICHIE *et al.*, 1994).

## - CAPÍTULO 4 -

### **METODOLOGIA PROPOSTA APLICADA NA FORMAÇÃO DE AGRUPAMENTOS**

No capítulo anterior foram apresentados dois métodos diferentes de agrupamento de dados, o algoritmo das “k-médias” e os Mapas Auto-Organizáveis (SOM), assim como comparações entre eles. Neste capítulo será apresentada uma metodologia que utiliza ambos os métodos em conjunto, para validação concorrente dos agrupamentos realizados.

No caso específico do agrupamento de conjuntos consumidores de energia elétrica, é proposto um novo paradigma, que envolve desde a mudança das variáveis envolvidas, que serão escolhidas por análise estatística multivariada, até a forma como são aplicadas na agregação. Possibilitando ao final do processo, a avaliação dos resultados obtidos e a identificação de incompatibilidades com as metas definidas pelo órgão regulador.

#### 4.1. Revisão dos Critérios

Até o ano de 1999 boa parte das concessionárias brasileiras não aplicava uma política de monitoramento da qualidade do fornecimento de energia elétrica, apesar da Portaria DNAEE 046 estabelecer padrões a serem cumpridos desde 1978. Quando a ANEEL decidiu implementar a gestão dos valores de DEC e FEC, por meio de metas de desempenho com base na análise comparativa das empresas de distribuição, poucas dispunham de dados suficientes sobre todos os seus conjuntos.

Os critérios adotados para determinar as metas de qualidade a serem atingidas pelos conjuntos consumidores foram descritos no Capítulo 2. Como essas metas foram determinadas agrupando mais de 5000 conjuntos de todo o Brasil em 30 grupos, a partir de 5 variáveis escolhidas, algumas das quais tiveram de ser estimadas em muitos conjuntos por causa da falta de gerenciamento desses dados, surgiram grupos que não ficaram homogêneos o suficiente, resultando em incompatibilidades.

Citando TANURE (2000):

“Para a aplicação da metodologia é fundamental que os atributos sejam corretamente determinados, uma vez que os mesmos são a base de toda classificação a ser feita. A implementação deste método deverá passar por um processo de depuração a cada ciclo de definição das metas, uma vez que a maioria das distribuidoras ainda não dispõem de sistemas de gerência de redes e de dados confiáveis.”

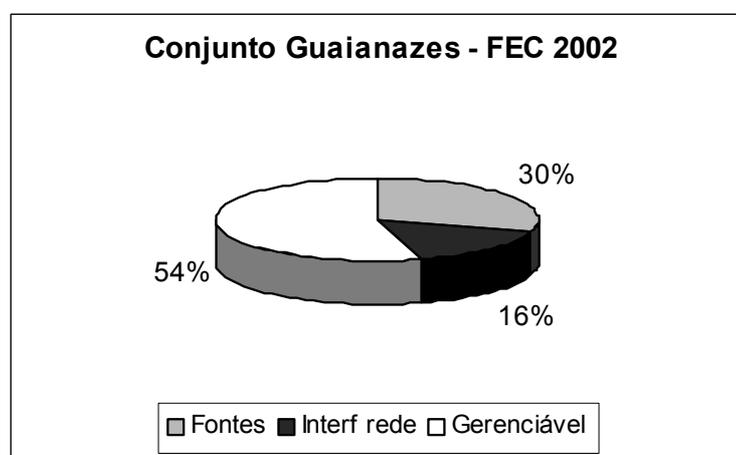
“A correta determinação destes atributos é uma das partes mais sensíveis da metodologia proposta. Assim, é conveniente que nos primeiros ciclos de aplicação da mesma os valores encontrados com o método sejam adotados apenas como valores de referência. Os valores definitivos a serem atribuídos às empresas devem ser validados junto às concessionárias visando incorporar realidades que a simples classificação não consegue traduzir.”

No balanço realizado em 2002, em média 22% dos conjuntos brasileiros transgrediram as metas determinadas para aquele ano. No caso da CELESC, esse valor foi em torno de 28%.

Ciente das penalidades previstas, apresentadas no Capítulo 2, e estando na metade do ciclo (ano de 2004) para se atingir as metas finais, as concessionárias estão buscando compreender melhor a metodologia adotada pela ANEEL para validarem ou contestarem a

classificação atual, desenvolvendo ferramentas para o gerenciamento das metas de conjuntos consumidores sob sua concessão.

Alguns trabalhos já apresentados destacam a pouca informação utilizada para realizar a classificação, as cinco variáveis escolhidas não são suficientes para definir de forma clara a situação dos conjuntos. NETO *et al.*(2003) destaca conjuntos semelhantes da região da grande São Paulo, atendidos pela AES Eletropaulo, com metas muito distintas. Mostra também a influência de fatores não-gerenciáveis nos índices DEC e FEC (Figura 22), dentre alguns estão os elevados índices de vandalismo, ligações ilegais e enrolamento de pipas na rede aérea em conjuntos suburbanos com crescimento desordenado e favelas.



**Figura 22.** Percentagem de Fatores Gerenciáveis e Não-Gerenciáveis no FEC de um Conjunto da AES Eletropaulo

Como esses fatores não foram levados em conta na classificação, conjuntos com essas características foram prejudicados na determinação de metas, sendo hoje, difícil de adequar seus sistemas de forma a alcançar tais metas. O que a Eletropaulo está propondo à ANEEL é uma relocação de metas, diminuindo a meta de conjuntos que possuem uma boa folga para aumentar onde não há condições de atingi-la., mantendo a meta global da empresa.

Está claro que a escolha das características (variáveis) a serem levadas em conta na classificação é crucial para o processo, e determina o destino final que é a definição das metas para os agrupamentos de conjuntos consumidores.

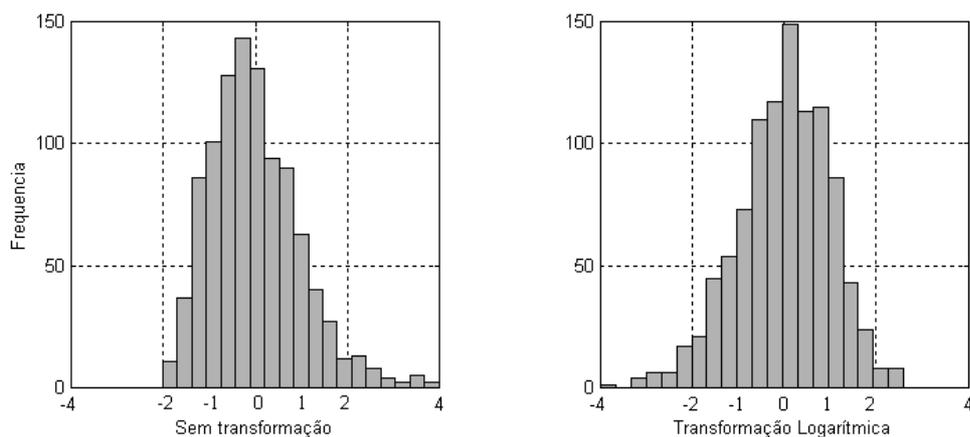
Outro fator importante, e que não foi abordado pelo órgão regulador, é a própria definição de conjunto consumidor. Algumas concessionárias, como a CELESC, utilizam o

município como sendo um conjunto consumidor, outras adotam regionais, e algumas particionam municípios em vários conjuntos menores, como faz a Eletropaulo.

Apenas esse fato, de não existir um padrão na definição do que deve ser um conjunto consumidor, já é suficiente para gerar uma classificação discutível ao final do processo. Pois, é possível que uma mesma meta seja atribuída a um município ou região (que engloba áreas urbanas, industriais e rurais), e a um conjunto definido por um grande bairro de uma metrópole.

Uma investigação inicial sobre a influência das variáveis no processo de agrupamento foi publicada em (SPERANDIO *et al.*, 2003). Nesse trabalho foi realizado o agrupamento de 260 conjuntos sob concessão da CELESC utilizando 20 variáveis e a técnica dos Mapas Auto-Organizáveis. É demonstrada a vantagem do emprego dos mapas para busca de correlação entre variáveis, definição de agrupamentos e extração de características intrínsecas dos dados organizados.

Outro detalhe dos critérios adotados pela ANEEL foi o uso de uma transformação logarítmica no pré-processamento dos dados. Essa transformação foi utilizada com o intuito de diminuir a dispersão dos dados, e torná-los “melhor” distribuídos para facilitar a classificação. Porém esse fato faz com que aconteça uma classificação errônea, ainda mais se tratando da determinação de metas, pois os valores são aproximados para o centro da distribuição de freqüências (Figura 23).



**Figura 23.** Histogramas de Frequências: Sem e Com Transformação Logarítmica.

Pode ser visto na esquerda da Figura 23, uma distribuição de freqüência de uma certa variável, e a direita a mesma variável após sofrer a transformação logarítmica. Note o

aumento da concentração de dados no centro (entre  $-1$  e  $0$  e  $0$  e  $1$ ), a ausência dos valores próximos a  $4$ , e o surgimento de valores entre  $-4$  e  $-2$ .

Os valores dos índices de qualidade (DEC e FEC) não sofrem tal transformação, mas são utilizados para determinar a meta para os conjuntos agrupados por variáveis transformadas. Perceba que no caso da variável apresentada na Figura 23 estar inversamente correlacionada com os indicadores, ou seja, quanto maior a variável menor o DEC e/ou o FEC, conjuntos que antes da transformação eram distintos passam a ter valores muito próximos, embora ainda com índices diferentes.

Além das variáveis adotadas para a classificação serem poucas para representar a diversidade de fatores que influenciam na qualidade do fornecimento de energia elétrica, tomá-las juntas dificulta a interpretação dos resultados.

Uma forma de abordar o problema sob um novo paradigma foi apresentada por QUEIROZ (2002) da CELESC. Com base no que é feito pelo órgão regulador equivalente à ANEEL na Argentina, o ENRE (Ente Nacional Regulador de la Eletricidad), que separa as características em variáveis de Mercado e do Sistema Elétrico.

As variáveis de Mercado, definem o tipo e a quantidade de consumidores de energia elétrica, como o consumo e o número de consumidores por classe. Já as características do Sistema Elétrico mostram a estrutura do fornecimento, e devem refletir a qualidade desse sistema.

Ao cruzar os resultados de agrupamentos de Mercado com os de Sistema Elétrico pode-se identificar peculiaridades importantes dos conjuntos, como grandes mercados consumidores com deficiências no sistema de fornecimento.

## 4.2. Seleção de Variáveis via Estatística Multivariada

O estudo para seleção de variáveis parte da disponibilidade dessas variáveis para todos os conjuntos que serão analisados. Para caracterizar o Mercado foram disponibilizados 11 atributos para os 260 municípios sob concessão da CELESC, que são o consumo e a quantidade de consumidores de 5 classes mais a área do município:

- Consumo Residencial (kWh/ano);
- Consumo Industrial (kWh/ano);
- Consumo Rural (kWh/ano);
- Consumo Comercial (kWh/ano);
- Consumo Público (kWh/ano);
- Área do Município (km<sup>2</sup>);
- Número de Consumidores Residenciais ;
- Número de Consumidores Industriais;
- Número de Consumidores Rurais;
- Número de Consumidores Comerciais;
- Número de Consumidores Públicos.

Para formar a base de dados de variáveis do Sistema Elétrico, dispomos de:

- Número de Subestações (SEs);
- Número de Alimentadores;
- Número de Transformadores;
- Capacidade de Manobra entre Alimentadores;
- Capacidade Instalada (kVA);
- Distância da Sede do Município à SE (km);
- Extensão da Rede Aérea Primária (km).

O estudo das variáveis parte de uma análise descritiva dos dados, onde busca-se compreender a variação, ou dispersão dos dados. Primeiramente determina-se o valor médio, que nada mais é do que a média aritmética dos dados para uma determinada variável. Porém a média é muito influenciada pela amplitude determinada pelo valor máximo, que às vezes pode ser um dado discrepante, que não representa a tendência da maioria.

Para evitar a má interpretação deve-se avaliar outras medidas de tendência central, como a mediana e os quartis, e uma medida de dispersão em relação ao valor médio, como o desvio padrão.

Os quartis, são valores que colocando os dados em ordem crescente, dividem-os em quatro partes com quantidades iguais. Sendo o primeiro quartil o que contém 25% dos dados, o segundo engloba 50%, dividindo os dados em duas partes e por isso também é chamado de mediana, e por fim o terceiro, ou superior, que representa a quantidade de 75% (SPIEGEL, 1985). Verificando os quartis se pode imaginar a distribuição de frequência dos dados. É importante notar a diferença entre a Média e a Mediana, a primeira está vinculada a amplitude de variação dos dados, enquanto a segunda à quantidade desses.

O desvio padrão mostra o quanto se pode esperar de variação em torno do valor médio, para uma distribuição normal, três vezes esse valor mais ou menos o valor médio envolvem 99% dos dados da amostra.

Como o objetivo é realizar o agrupamento de dados e entender os motivos que levaram a formação de cada grupo, utilizar uma grande quantidade de variáveis em conjunto dificulta a análise dos resultados finais. Além disso, um excesso de variáveis pode conter muita informação redundante, exigindo um trabalho computacional elevado, sem trazer benefícios para a formação de grupos distintos.

O primeiro passo para uma seleção de um conjunto de variáveis representativo, ou um “constructo teórico”<sup>§</sup>, é a avaliação da matriz de correlações entre as variáveis em questão. O coeficiente de correlação mede o grau em que existe uma associação linear entre duas variáveis. Uma correlação positiva reflete uma tendência de que um valor alto em uma variável esteja associado a um valor alto em uma segunda variável. Uma correlação negativa reflete uma associação de um valor alto na primeira variável com um valor baixo na segunda. Portanto, variáveis com forte correlação positiva trazem informação redundante.

Porém, essa análise não é suficiente para retirar uma variável da base de dados, é preciso ainda descobrir dentre as variáveis correlacionadas qual traz a melhor contribuição para a variabilidade dos dados, ou seja, qual delas poderá explicar melhor a diferença entre dados em grupos diferentes.

Para isso é utilizada a técnica de Análise Fatorial, que serve para apoiar a decisão de reduzir o número de variáveis de um conjunto, procurando reter o máximo de informação possível, e tornar as variáveis remanescentes mais significativas e fáceis de serem trabalhadas.

---

<sup>§</sup> Aquilo que é elaborado ou criado com base em dados simples, um conceito.

A extração de componentes principais é um método para transformar as variáveis originais em variáveis novas, não correlacionadas, chamadas de fatores. Cada fator é uma combinação linear das variáveis originais (AAKER,2001).

Uma medida da quantidade de informação transferida para cada fator é a sua variância, e os fatores são arranjados em ordem decrescente em relação a sua variância. Assim, o fator mais informativo é o primeiro.

Outra informação dada por esse método é a carga fatorial, que mostra a associação entre um fator e as variáveis originais, oferecendo uma indicação das variáveis mais representativas de um conjunto.

Com essas informações pode-se então selecionar uma, ou mais variáveis de entrada para representar cada fator. Essa seleção deve ser baseada em suas cargas fatoriais e em um julgamento sobre a utilidade e validade dessas variáveis.

#### 4.2.1. Análise das Variáveis de Mercado

Utilizando a premissa de separar as variáveis que caracterizam o Mercado consumidor das características do sistema de fornecimento de energia elétrica, para posterior cruzamento entre os grupos, facilitando a identificação de conjuntos que estão sendo mal atendidos, ou que realmente foram erroneamente classificados durante o processo de determinação das metas pelo órgão regulador.

Parte-se então da análise descritiva do conjunto de variáveis que definem as características de mercado, resumidas nas tabelas a seguir:

**Tabela 3.** Descrição do Consumo por Classe e da Área do Município (kWh/ano)

Consumox1000	Média	Quartil 25%	Mediana	Quartil 75%	Amplitude	Desvio Padrão
Residencial	11.560,56	700,21	1.834,14	6.854,52	384.961,53	37.393,39
Industrial	22.556,68	296,96	1.471,28	12.286,55	1.239.991,20	89.460,13
Comercial	6.698,84	263,41	852,18	2.682,11	268.353,18	24.740,49
Rural	3.105,30	1.029,55	1.982,83	3.325,21	44.347,69	4.526,45
Público	3.378,59	365,35	825,71	1.968,29	138.807,84	10.994,69
Área	340,54	135,05	238,80	400,75	2614,8	335,11

**Tabela 4.** Descrição do Número de Consumidores por Classe (unidade)

Número de Consumidores	Média	Quartil 25%	Mediana	Quartil 75%	Amplitude	Desvio Padrão
Residencial	5.427	383	1.123	3.914	147.484	15.051
Industrial	212	19	44	163	5.253	523
Comercial	558	63	146	346	17.363	1.584
Rural	773	353	600	1.029	4.853	654
Público	67	29	42	63	1.049	95

A primeira informação das tabelas indica a classe de consumo, com exceção da área, depois a média e os quartis, mostrando a distribuição central dos dados, e a amplitude e o desvio padrão que dão a idéia da dispersão em relação ao centro, e podem indicar a presença de dados discrepantes.

Na Tabela 3, percebe-se que a média é muito superior ao quartil 75%, a não ser para o consumo rural, indicando a disparidade dos dados acima desse quartil. O que é confirmado pela grande amplitude e o alto desvio padrão. Mesmo multiplicando o desvio por três e somando à média, ainda existe uma distância muito grande para a amplitude. Pode-se afirmar que já existe um grupo distinto de grandes consumidores, que envolve menos de 25% dos municípios analisados.

Outra informação interessante é que o consumo rural predomina no quartil 25%, apontando que existem municípios com predominância dessa classe. Já a mediana mostra uma maior homogeneidade, indicando que pelo menos metade dos municípios não possuem um perfil de consumo diferenciado. Essa característica diferenciada é mais evidente na classe industrial, o que é de se esperar, pois indústrias eletrointensivas consomem mais energia do que muitas cidades de pequeno porte, e é notório que o estado de Santa Catarina possui algumas desse tipo. Isso pode ser constatado comparando a relação entre o consumo e o número de consumidores (Tabela 5). A variável área não apresenta características discrepantes, tendo um perfil de distribuição quase normal.

Na Tabela 4 é evidente a diferença entre o número de consumidores da classe residencial para as outras, mas ainda se percebe uma discrepância dos valores centrais para a amplitude.

**Tabela 5.** Relação entre o Consumo e o Número de Consumidores por Classe (kWh/consumidor)

Índices Consumo / Número	Média	Quartil 25%	Mediana	Quartil 75%	Amplitude	Desvio Padrão
Residencial	1,80	1,61	1,77	2,01	1,85	0,32
Industrial	112,18	9,99	30,85	114,46	3269,08	323,17
Comercial	6,72	4,08	5,61	8,08	35,55	4,30
Rural	19,49	2,36	2,95	4,01	3155,13	198,49
Público	28,69	11,77	21,00	38,12	155,12	24,75

Interessante notar como a discrepância desaparece quando se observa o índice consumo/número da classe residencial. A amplitude, diferença entre o maior e o menor valor, é de apenas 1,85 e o desvio padrão de 0,32 ! Isso mostra que o consumo residencial é homogêneo, e que em cidades grandes o consumidor residencial gasta, em média, quase a mesma quantidade de energia que o de uma cidade de pequeno porte.

Já na classe industrial ocorre justamente o contrário, a amplitude do índice é de mais de 3000, mostrando a diferença de indústrias de grande porte (eletrointensivas) para os demais tipos. Algo semelhante aparece na classe rural, porém discutindo a respeito com técnicos da concessionária se descobriu que essa discrepância é devido à cooperativas de eletrificação rural, que são consideradas pela empresa como um único consumidor, mas na verdade fornecem energia para um grande número de seus associados.

Essa análise prévia serve apenas para se conhecer as variáveis individualmente, e a não ser que se tenha um conhecimento *a priori*, não indicam relações ou dependências entre si. Mas como se sabe que as variáveis que definem as características de mercado, em sua maioria, tratam de qualidades iguais, como o consumo e o número de consumidores, espera-se que existam informações redundantes. Para verificar esse fato procedeu-se o estudo de correlação e análise fatorial.

**Tabela 6.** Matriz de Correlações entre as Variáveis de Mercado

	Área	C.Res.	C.Ind.	C.Com.	C.Rur.	C.Pub.	N.Res.	N.Ind.	N.Com.	N.Rur.	N.Pub.
Área	1,00	0,16	0,22	0,14	0,11	0,19	0,19	0,16	0,16	0,50	0,30
C.Res.	0,16	1,00	0,68	<b>0,98</b>	0,08	<b>0,96</b>	0,99	<b>0,91</b>	<b>0,99</b>	0,03	<b>0,93</b>
C.Ind.	0,22	0,68	1,00	0,61	0,14	0,55	0,67	<b>0,82</b>	0,60	0,18	0,57
C.Com.	0,14	<b>0,98</b>	0,61	1,00	0,08	<b>0,96</b>	<b>0,96</b>	<b>0,86</b>	<b>0,98</b>	0,03	<b>0,93</b>
C.Rur.	0,11	0,08	0,14	0,08	1,00	0,09	0,09	0,10	0,12	0,39	0,18
C.Pub.	0,19	<b>0,96</b>	0,55	<b>0,96</b>	0,09	1,00	<b>0,96</b>	<b>0,80</b>	<b>0,97</b>	0,08	<b>0,95</b>
N.Res.	0,19	<b>0,99</b>	0,67	<b>0,96</b>	0,09	<b>0,96</b>	1,00	<b>0,91</b>	<b>0,98</b>	0,05	<b>0,94</b>
N.Ind.	0,16	<b>0,91</b>	<b>0,82</b>	<b>0,86</b>	0,10	<b>0,80</b>	<b>0,91</b>	1,00	<b>0,86</b>	0,07	<b>0,82</b>
N.Com.	0,16	<b>0,99</b>	0,60	<b>0,98</b>	0,12	<b>0,97</b>	<b>0,98</b>	<b>0,86</b>	1,00	0,07	<b>0,95</b>
N.Rur.	0,50	0,03	0,18	0,03	0,39	0,08	0,05	0,07	0,07	1,00	0,22
N.Pub.	0,30	<b>0,93</b>	0,57	<b>0,93</b>	0,18	<b>0,95</b>	<b>0,94</b>	<b>0,82</b>	<b>0,95</b>	0,22	1,00

Verificando a Tabela 6 se pode notar muitas variáveis com uma forte correlação positiva ( $> 0,8$  em negrito). Lembrando que a matriz de correlações é simétrica e possui os elementos da diagonal iguais a 1, pois é o cruzamento da mesma variável. Isso já é um indicativo de informações redundantes. Por exemplo, o consumo residencial está fortemente ligado aos consumos comercial e público, e claro, ao número de consumidores residenciais, e das outras classes citadas. Porém surge um dado que chama a atenção, a forte correlação entre o consumo residencial e o número de consumidores industriais (0,91), mas que não chega a 0,7 com o consumo dessa classe. Pode-se dizer que nas cidades com maior consumo residencial, as grandes cidades, está o maior número de indústrias, mas não necessariamente estão localizadas as maiores indústrias.

Já as variáveis da classe rural são as que possuem a menor dependência, mesmo entre o consumo e o número de consumidores (0,39), esta tendo até maior correlação com

a área do município (0,5). Contudo, como dito anteriormente, as cooperativas de eletrificação rural devem interferir nesse valor.

A análise de correlações indica que as variáveis das classes residencial, comercial e público estão fornecendo basicamente o mesmo tipo de informação, que é o tamanho da zona urbana do município. Então para avaliar a contribuição de cada uma na variabilidade dos dados, procede-se com a análise fatorial e extração de componentes principais para redução das variáveis envolvidas.

**Tabela 7.** Análise Fatorial das Variáveis de Mercado

<b>Fator</b>	<b>Autovalor</b>	<b>% de Variância</b>	<b>% Acumulada</b>
1	7,15	65,03	65,03
2	1,67	15,18	80,21
3	0,90	8,15	88,36
4	0,67	6,07	94,43
5	0,41	3,71	98,14
6	0,09	0,83	98,97
7	0,05	0,45	99,43
8	0,03	0,30	99,72
9	0,02	0,20	99,92
10	0,01	0,06	99,98
11	0,00	0,02	100,00

A Tabela 7 mostra o total de 11 fatores extraídos da base de variáveis que representam o mercado consumidor de energia elétrica, o autovalor ligado a esse fator e a percentagem da variância dos dados explicada por ele, assim como a percentagem acumulada dessa variância, que deve ser de 100% com todos os fatores.

O primeiro fator se mostra muito significativo, explicando 65% da variância dos dados, e a contribuição dos próximos quatro fatores já explica 98%. Isso indica que se pode reduzir a dimensão do problema utilizando os cinco primeiros fatores sem perda de informação.

Para descobrir quais variáveis contribuem para cada fator deve-se analisar as cargas fatoriais, e assim, escolher a variável mais correlacionada ao fator, ou uma combinação de variáveis.

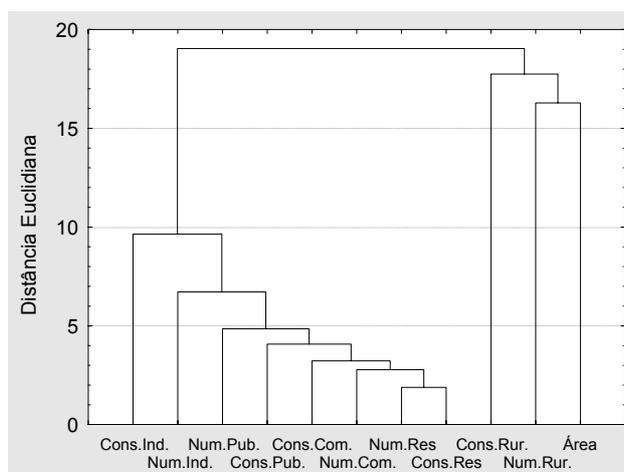
A Tabela 8, a seguir, apresenta quais variáveis estão ligadas a quais fatores. Nesse caso, o Fator 1, que explica 65% da variância dos dados, está vinculado a uma série de variáveis, valores destacados, que são justamente aquelas variáveis fortemente correlacionadas apresentadas na Tabela 6, e que representam característica urbanas dos municípios estudados.

**Tabela 8.** Cargas Fatoriais para as Variáveis de Mercado

	Fator 1	Fator 2	Fator 3	Fator 4	Fator 5	Fator 6	Fator 7	Fator 8	Fator 9	Fator 10	Fator 11
<b>Área</b>	0,111	0,961	0,026	0,068	0,241	0,002	0,003	0,001	0,001	0,000	0,000
<b>C.Res</b>	0,953	0,040	0,010	0,286	-0,016	0,028	-0,062	0,013	-0,014	0,013	-0,045
<b>C.Ind.</b>	0,443	0,088	0,060	0,886	0,083	-0,021	0,003	-0,002	0,001	0,000	0,000
<b>C.Com.</b>	0,964	0,022	0,020	0,199	-0,010	0,002	-0,041	-0,164	-0,028	0,000	0,001
<b>C.Rur.</b>	0,058	0,026	0,980	0,044	0,185	0,002	0,002	0,000	0,000	0,000	0,000
<b>C.Pub.</b>	0,976	0,072	0,026	0,116	0,026	-0,068	0,000	0,038	0,147	-0,003	0,002
<b>N.Res</b>	0,948	0,075	0,022	0,280	-0,014	0,070	-0,032	0,073	-0,038	0,061	0,012
<b>N.Ind.</b>	0,777	0,031	0,027	0,540	0,009	0,322	0,002	0,001	-0,009	0,001	0,000
<b>N.Com.</b>	0,976	0,038	0,053	0,180	0,015	0,013	-0,019	0,014	-0,069	-0,059	0,025
<b>N.Rur.</b>	0,020	0,260	0,212	0,062	0,940	0,000	0,004	0,001	0,001	0,000	0,000
<b>N.Pub.</b>	0,942	0,154	0,088	0,139	0,147	0,008	0,198	0,022	0,001	-0,001	0,000

O segundo fator pode ser representado apenas pela variável Área, assim como os fatores 3, 4 e 5 podem ser representados diretamente pelas variáveis de Consumo Rural, Consumo Industrial e Número de Consumidores Rurais, respectivamente. Os demais não têm representatividade para explicar a variação dos dados.

Identificadas as características que trazem mais informação, resta agora escolher o conjunto reduzido de variáveis para formar a base dados que vai ser utilizada para realizar o agrupamento dos municípios estudados.

**Figura 24.** Árvore de Associação entre as Variáveis de Mercado

Como o primeiro fator está vinculado a seis variáveis deve-se decidir qual delas serão utilizadas na base final. Para apoiar essa decisão é verificada a distância entre elas, mostrada por um dendrograma, ou árvore de associação (Figura 24). Essa figura mostra a grande distância da classe rural para as demais, que com exceção da classe industrial, são associadas com uma distância menor que 5.

Em vez de escolher apenas uma das seis variáveis para representar o primeiro fator, optou-se por somá-las, e como não se pode somar variáveis que representam coisas diferentes, como o consumo com o número de consumidores, foram adicionados os consumos residencial, comercial e público, formando uma nova variável chamada de consumo urbano.

Como o quinto fator, representado pelo número de consumidores rurais, possui pouca representatividade (3,7%), e os demais indicadores do número de consumidores das outras classes estão muito atrelados, decidiu-se somar esses valores para formar o número de consumidores total do município.

Portando a nova base de dados de Mercado será composta por cinco variáveis, o Consumo Urbano, o Consumo Industrial, o Consumo Rural, a Área e o Número de Consumidores Total.

Uma nova análise fatorial foi realizada para verificar a contribuição de cada variável dessa nova base.

**Tabela 9.** Análise Fatorial da Nova Base de Mercado

Fator	Autovalor	% de	%
1	2,64	52,82	52,82
2	1,03	20,59	73,41
3	0,89	17,72	91,14
4	0,43	8,64	99,77
5	0,01	0,23	100,00

**Tabela 10.** Cargas Fatoriais da Nova Base de Mercado

	Fator 1	Fator 2	Fator 3	Fator 4	Fator 5
<b>Área</b>	0,087	0,049	0,992	0,081	0,001
<b>C.Ind</b>	0,427	0,070	0,103	0,896	0,000
<b>C.Rural</b>	0,041	0,997	0,048	0,054	0,000
<b>C.Urban</b>	0,965	0,026	0,054	0,245	-0,072
<b>N.Total</b>	0,947	0,048	0,101	0,289	0,078

Verifica-se pela Tabela 9 e pela Tabela 10 que as variáveis somadas, consumo urbano e número de consumidores total, ainda estão fortemente atreladas, e trazem informação redundante, que é o tamanho da zona urbana do município. Porém optou-se por manter as duas na base de dados pelo fato de comparar o consumo com as outras duas variáveis, e número de consumidores total ser utilizado no cálculo do DEC e FEC, facilitando a análise posterior dos agrupamentos críticos.

#### 4.2.2. Análise das Variáveis do Sistema Elétrico

A análise das variáveis que caracterizam o sistema de fornecimento de energia elétrica dos municípios estudados segue a mesma forma utilizada para as variáveis de mercado, a partir de uma base com sete variáveis, busca-se extrair as mais representativas por meio da análise estatística multivariada, que é constituída pela busca de correlações, análise fatorial e de distância entre variáveis.

Primeiramente deve-se estudar as características de cada variável, para se ter uma idéia do que elas representam, e de como se comporta a sua variação.

**Tabela 11.** Descrição das Variáveis do Sistema Elétrico

Variável	Média	Quartil 25%	Mediana	Quartil 75%	Amplitude	Desvio Padrão
Dist.M-SE	17,72	5,06	15,25	27,00	60,0	14,02
Num.SE	0,42	0,00	0,00	1,00	6,0	0,79
Num.Al.	3,21	1,00	2,00	4,00	40,0	4,66
Manobra	-	-	-	-	1,0	-
RAP_KM	222,73	101,22	162,24	279,78	1124,7	193,11
Num.Trf	485,50	197,50	323,50	536,50	4867,0	580,11
P_KVA	25609,15	4736,50	9354,50	19681,00	529168,0	59064,41

Da Tabela 11 pode-se extrair uma série de informações de cada variável. A distância do município à subestação parece ter uma distribuição regular, sem a presença de dados discrepantes ou concentração em algum valor específico. Já o número de subestações revela que pelo menos 50% dos municípios estudados não possuem SE própria, assim como o número de alimentadores, que é de no máximo 2 para metade dos dados. A capacidade de manobra entre alimentadores é uma variável binária, ou seja, é 0 ou 1, que representa a possibilidade dos alimentadores que abastecem o município transferirem carga para outros. É uma variável importante pois reflete diretamente a qualidade do fornecimento em relação ao DEC. A extensão da rede aérea primária está vinculada ao tamanho da zona urbana da cidade, e possui uma concentração em valores relativamente pequenos, 75% dos dados possuem uma rede menor do que 280km enquanto a amplitude é de 1124km. Fato semelhante acontece com o número de transformadores e com a potência instalada. Isso mostra que o estado de Santa Catarina é composto, em sua maioria, por municípios de pequeno porte.

Tabela 12. Matriz de Correlações das Variáveis do Sistema

	Dist.M-SE	Num.SE	Num.Al.	Manobra	RAP_KM	Num.Trf	P_KVA
Dist.M-SE	1,00	-0,52	-0,33	-0,42	-0,33	-0,37	-0,35
Num.SE	-0,52	1,00	<b>0,83</b>	0,33	0,58	0,79	<b>0,83</b>
Num.Al.	-0,33	<b>0,83</b>	1,00	0,27	0,57	<b>0,87</b>	<b>0,94</b>
Manobra	-0,42	0,33	0,27	1,00	0,27	0,28	0,22
RAP_KM	-0,33	0,58	0,57	0,27	1,00	<b>0,84</b>	0,60
Num.Trf	-0,37	0,79	<b>0,87</b>	0,28	<b>0,84</b>	1,00	<b>0,91</b>
P_KVA	-0,35	<b>0,83</b>	<b>0,94</b>	0,22	0,60	<b>0,91</b>	1,00

Diferente do caso das variáveis de mercado, a matriz de correlações da Tabela 12 apresenta alguns coeficientes negativos. Isso indica que com o aumento do valor de uma variável, a outra tende a diminuir, como acontece com a distância do município à SE que diminui com o aumento do número de SEs. Os valores de correlação forte estão em negrito na tabela, e pode-se destacar a associação entre o número de alimentadores e a potência instalada. Poderia-se esperar que 5 variáveis estivessem bem atreladas, número de SEs, número de alimentadores, extensão da rede aérea primária, número de transformadores e a potência instalada, devido ao encadeamento físico delas (Figura 25), porém a correlação não é direta entre todas elas.

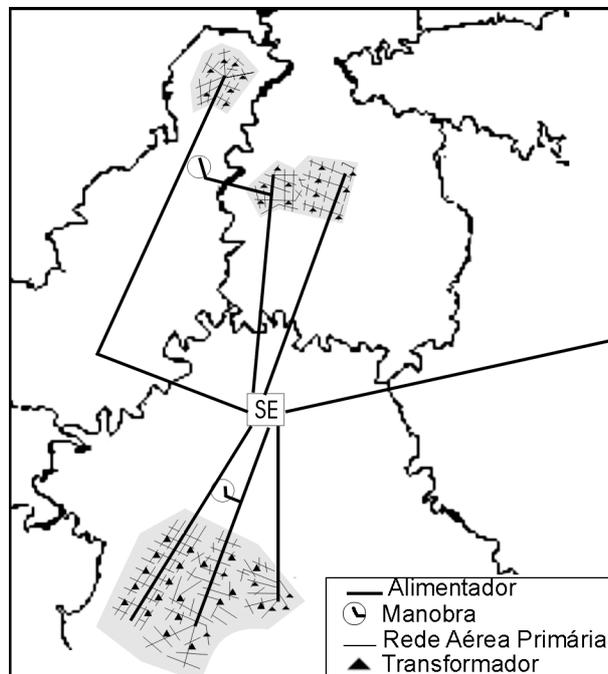


Figura 25. Representação do Sistema de Fornecimento de Energia Elétrica

Seguindo o procedimento para a escolha das variáveis que ficarão na base de dados do sistema elétrico, realizou-se a análise fatorial.

**Tabela 13.** Análise Fatorial das Variáveis do Sistema

Fator	Autovalor	% de Variância	% Acumulada
1	4,48	64,03	64,03
2	1,11	15,87	79,90
3	0,60	8,61	88,51
4	0,55	7,85	96,36
5	0,18	2,52	98,88
6	0,06	0,86	99,74
7	0,02	0,26	100,00

A Tabela 13 mostra que com quatro fatores já é possível explicar 96% da variância dos dados, indicando uma redução do número de variáveis sem perda de informação. Para saber quais variáveis estão envolvidas diretamente com os fatores verifica-se a carga fatorial.

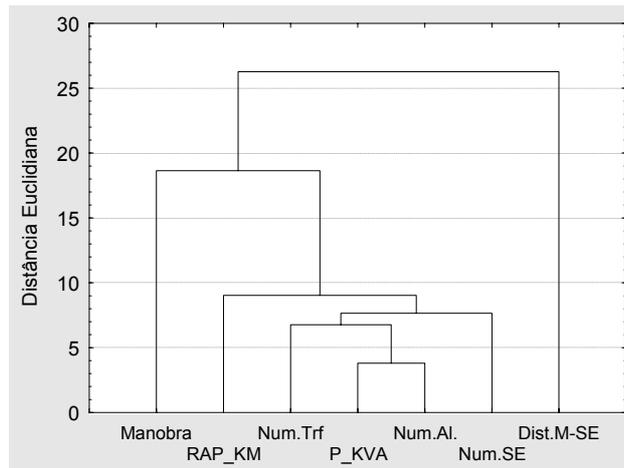
**Tabela 14.** Cargas Fatoriais para as Variáveis do Sistema

	Fator 1	Fator 2	Fator 3	Fator 4	Fator 5	Fator 6	Fator 7
<b>Dist.M-SE</b>	-0,195	-0,210	-0,949	-0,122	-0,047	0,001	0,003
<b>Num.SE</b>	0,789	0,148	0,309	0,241	0,450	-0,003	0,002
<b>Num.AI.</b>	0,943	0,105	0,146	0,247	0,050	-0,115	0,041
<b>Manobra</b>	0,112	0,969	0,196	0,102	0,028	-0,001	-0,003
<b>Num.Trf</b>	0,778	0,115	0,134	0,582	-0,013	0,015	-0,157
<b>P_KVA</b>	0,953	0,069	0,121	0,249	-0,011	0,105	0,001
<b>RAP_KM</b>	0,367	0,113	0,129	0,912	0,062	-0,003	0,021

Na Tabela 14 comprova-se a forte participação de cinco variáveis nos quatro primeiros fatores. O Fator 1, tem uma participação de outras variáveis, mas as que se destacam são os números de alimentadores e a potência instalada, enquanto os fatores 2,3 e 4 possuem participação quase que exclusiva das variáveis manobra, distância do município à SE e extensão da rede aérea primária, respectivamente.

Com isso, três variáveis já estão definidas que ficarão na base de dados, mas ainda resta a informação redundante do número de alimentadores e da potência. Para escolher uma delas pode-se recorrer a árvore de associação entre as variáveis.

O dendrograma da Figura 26 não ajuda a escolher uma das variáveis, pois as duas são as primeiras a se ligarem, e ambas possuem uma correlação com as outras variáveis escolhidas muito semelhantes, o que torna indiferente a sua escolha.



**Figura 26.** Árvore de Associação entre as Variáveis do Sistema

Contudo, a informação de potência já está de certa forma embutida na base de Mercado pelas variáveis de consumo, e os dados sobre o número de alimentadores são mais representativos para avaliar a qualidade do sistema de fornecimento.

Definida a nova base fez-se uma reavaliação dos fatores.

**Tabela 15.** Análise Fatorial da Nova Base do Sistema

Fator	Autovalor	% de Variância	% Acumulada
1	2,10	52,55	52,55
2	0,90	22,49	75,05
3	0,57	14,32	89,37
4	0,43	10,63	100,00

**Tabela 16.** Cargas Fatoriais da Nova Base do Sistema

	Fator 1	Fator 2	Fator 3	Fator 4
<b>Dist.M-SE</b>	-0,135	-0,206	-0,959	-0,139
<b>Num.Al.</b>	0,940	0,114	0,141	0,288
<b>Manobra</b>	0,107	0,968	0,201	0,105
<b>RAP_KM</b>	0,288	0,113	0,146	0,940

A Tabela 15 e a Tabela 16 mostram que as quatro variáveis escolhidas representam informações bem distintas, e que a retirada de uma delas acarreta a perda de, no mínimo, 10% da variância dos dados.

### 4.3. Método de Agrupamento por Validação Cruzada

Definidas as variáveis que irão participar de ambas as bases de dados, de Mercado e de Sistema, a próxima fase desta metodologia é o agrupamento desses dados. Com isso, serão formados grupos de municípios com características semelhantes, que posteriormente serão classificados de acordo com a qualidade de fornecimento.

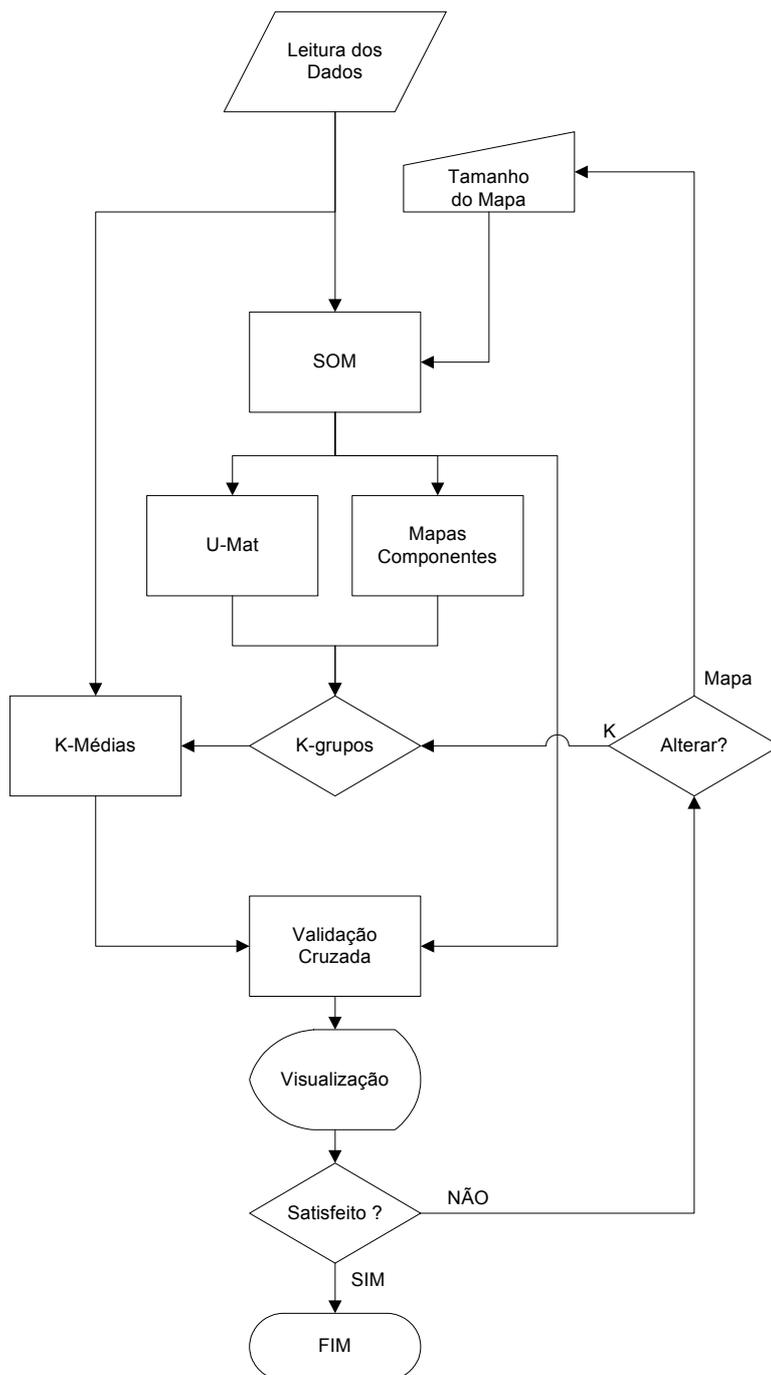
Para certificar que os grupos encontrados são realmente coerentes na sua formação, é realizada uma validação cruzada entre dois métodos de agrupamentos diferentes, o tradicional método estatístico das k-médias, e um algoritmo que utiliza paradigmas da inteligência artificial chamado de Mapa Auto-Organizável, mais conhecido pela sigla do seu nome em inglês (SOM – *Self-Organizing Map*), apresentados no Capítulo 3.

Como também foi apresentado no capítulo anterior, muitos autores costumam comparar o desempenho desses dois métodos, evidenciando as vantagens e desvantagens de cada um. Porém não utilizam ambos em conjunto, para justamente explorar as melhores características de cada um. Um trabalho que trouxe a idéia de realizar uma validação concorrente entre o SOM e o k-médias foi apresentado em (COELHO *et al.*, 2003), mas os métodos não trocavam informação, e o objetivo era o de realizar previsões com um mapa treinado a partir de dados históricos.

No método proposto, ambos os procedimentos avaliam a mesma base de dados em paralelo, e a troca de informações entre os algoritmos se dá primeiramente por meio da escolha do número de grupos a serem formados pelo k-médias, a partir da matriz de distâncias do mapa auto-organizável já treinado, e depois os dados são rotulados com os grupos definidos pelo k-médias e impressos no neurônio vencedor do mapa. A validação é feita verificando o grupo definido pelo k-médias que predomina em cada neurônio. Como o mapa realiza uma ordenação topológica, onde neurônios vizinhos possuem um alto grau de semelhança, não são aceitas configurações de grupos definidos pelo k-médias que estejam dispersos pelo mapa. No caso de não se estar satisfeito com a discordância entre os algoritmos, pode-se alterar o número de grupos, ou o tamanho do mapa, buscando minimizar as discordâncias e os erros de dispersão intra-grupo e topológica.

Portanto, ao final do procedimento de validação cruzada, tem-se a definição dos agrupamentos consolidada por dois algoritmos, e além disso, a vantagem de visualização que o SOM oferece, mostrando a organização dos grupos em um mapa, facilitando a identificação de conjuntos situados nas fronteiras, bem como os mapas componentes, que mostram a participação de cada variável na formação do mapa.

Um fluxograma desse procedimento é apresentado a seguir, mostrando a ordem em que ocorre cada etapa. É um processo de interação com o usuário, que deve decidir o tamanho do mapa e o número dos grupos ( $k$ ) fundamentado em ferramentas como a matriz de distâncias, mapas componentes, disposição dos grupos no mapa, valores dos erros de dispersão intra-grupo e topológica, e a discordância entre os agrupamentos apresentados pelos dois métodos.



**Figura 27.** Fluxograma do Procedimento de Validação Cruzada (SOM x K-médias)

A Figura 27 apresenta o fluxo dos dados no processo de validação cruzada, onde aparecem quatro etapas em que ocorrem intervenções do usuário. A primeira, é a escolha do tamanho do mapa, depois, estudando a U-Mat e os Mapas Componentes, deve-se determinar um número de grupos para o k-médias, ao final da validação, o usuário avalia os erros do cruzamento dos algoritmos e decide se quer mudar algum parâmetro, se sim, qual deles, o número k ou o tamanho do mapa.

A estrutura dos algoritmos das k-médias e do SOM foram implementados para Matlab© por um grupo do Centro de Pesquisas em Redes Neurais da Universidade de Helsinki, e disponibilizados como software livre (*SOM Toolbox*), com ampla documentação (VESANTO *et al.*, 1999). Esse *toolbox* é uma biblioteca de códigos que exploram a flexibilidade da série de funções prontas do Matlab, especialmente as gráficas, para treinamento e visualização dos mapas e outros dados que se pode extrair deste poderoso método.

Portanto todo o processo apresentado no fluxograma da Figura 27 roda no ambiente do Matlab, no qual foram desenvolvidos a interface de dados, o algoritmo para minimizar o erro do k-médias, o processo de validação cruzada, e também foram feitas adaptações para visualizar os dados do k-médias sobre o mapa.

Para iniciar o processo de agrupamento é necessário definir um tamanho para o mapa auto-organizável, sendo que essa decisão poderá influenciar na escolha do número de grupos para o k-médias. Como a intenção é de apenas realizar agrupamentos, uma opção é começar com um mapa pequeno, com aproximadamente um neurônio para cada quatro dados de entrada, que tende a formar grupos concisos. Todavia, mapas pequenos podem ter uma maior dispersão. Então, posterior a validação, pode-se estudar mapas maiores, com mais neurônios para cada dado, que por um lado, formam mapas esparsos, com muitos neurônios vazios, o que pode dificultar a associação de neurônios para formar grupos maiores, mas revelam relações entre variáveis que não são evidentes na base de dados. Mapas grandes são necessários quando se deseja utilizá-los para previsão.

Uma boa escolha inicial depende de certa forma das características da base de dados. Por exemplo, se os dados tendem a se concentrarem em algum valor, um mapa pequeno é suficiente para agrupá-los. Se possuírem uma grande dispersão em relação ao valor médio, um mapa grande os caracteriza melhor. Essa estimativa inicial deve se basear na análise descritiva das variáveis da base de dados a ser classificada.

No tocante ao processo de validação cruzada, cabe realçar que os dois algoritmos são independentes, e a única informação que é trocada entre eles depende da decisão do usuário ao definir o número de grupos do k-médias e o tamanho do mapa auto-organizável. Mesmo avaliando a tabela de erros fornecida durante o processo, deve-se estar ciente de que ao aumentar o número de grupos, reduz-se a dispersão, pois menos dados serão alocados para cada grupo, e ao se aumentar o tamanho do mapa, reduz-se a discordância com o k-médias, pois com mais neurônios haverá mais espaço para alocar dados, diminuindo os conflitos de fronteira, que exceto poucas exceções, são o principal motivo de divergências entre os dois métodos de agrupamento.

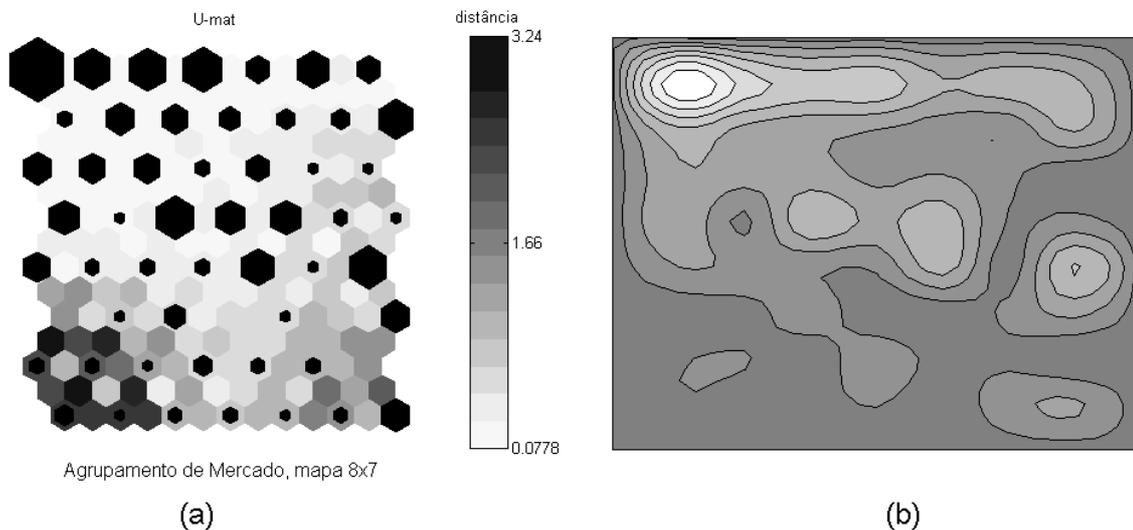
Mesmo com mapas grandes, tais exceções parecem ser decorrentes de uma maior sensibilidade do SOM à variabilidade dos dados, enquanto o k-médias está atrelado a um número de grupos definido pelo usuário.

O SOM pode indicar uma nova partição dos dados quando grupos definidos pelo k-médias ficam desconexos na topologia do mapa, ou seja, surgem neurônios vazios entre dados definidos como sendo de um mesmo grupo pelo outro algoritmo. Outra forma de aumentar a partição de grupos é verificando os mapas componentes, que apresentam a densidade de cada variável, e portanto, pode-se esperar que exista um grupo que tenha uma grande participação de uma variável, outro com uma participação média, outro em que tal variável não seja representativa, e etc.

### 4.3.1. Agrupamento da Base de Mercado

Seguindo a metodologia apresentada, foi realizado o estudo para determinar o agrupamento da base de dados com as variáveis que definem as características do mercado consumidor de energia elétrica.

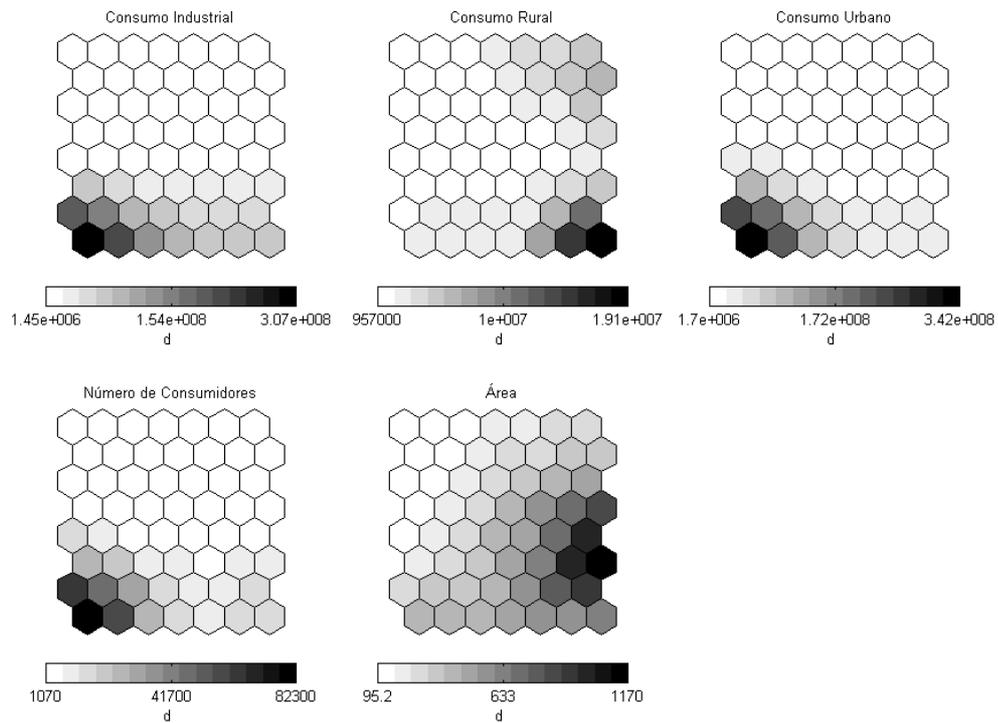
Como a base é composta por 260 dados, referentes aos municípios sob concessão da CELESC, o tamanho inicial escolhido para o mapa foi de 8x7, e as figuras da matriz de distâncias e dos mapas componentes são exibidas a seguir.



**Figura 28.** Matriz de Distâncias de Mercado: (a) U-mat com densidade; (b) Histogramas suavizados.

A Figura 28 mostra dois gráficos obtidos a partir da matriz de distâncias entre os neurônios, o primeiro (U-mat) mostra as distâncias em escala de cinza, onde vê-se uma grande área clara na parte superior esquerda, indicando um grupo de neurônios próximos, que estão com o tamanho proporcional a quantidade de dados que possuem, portanto, também se tem uma idéia da densidade do mapa. O segundo gráfico, mostra histogramas suavizados, como curvas de nível, delimitando regiões com uma concentração maior de dados. Analisando esses gráficos, pode-se tomar como cinco o valor inicial para o k-médias. Quatro grupos estão bem definidos na Figura 28(b), e o quinto é por causa da grande distância entre os neurônios do canto inferior esquerdo em (a).

Para tirar mais conclusões sobre este mapa também se deve verificar os mapas componentes.



**Figura 29.** Mapas Componentes da Base de Mercado.

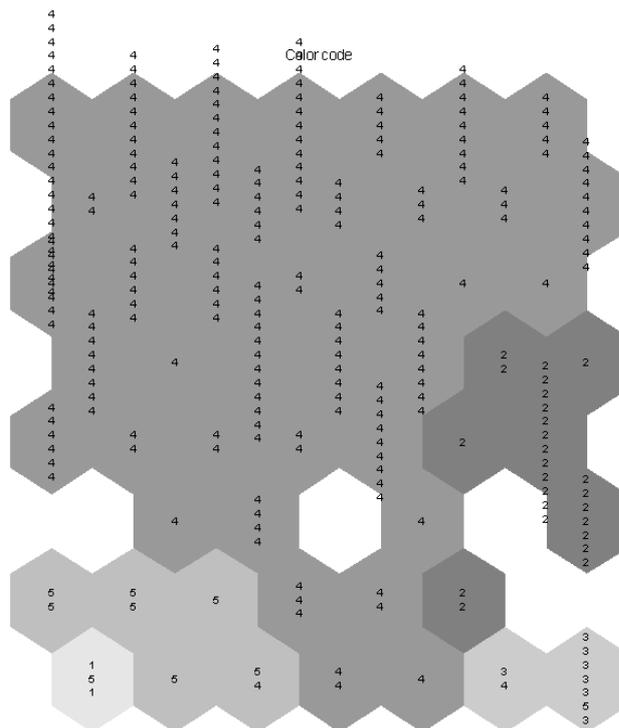
Os mapas componentes da Figura 29 revelam a quantidade das variáveis em cada neurônio. Pode-se perceber uma certa semelhança entre a distribuição das variáveis de consumo urbano e industrial e do número de consumidores, que ocupam uma grande região do mapa com valores baixos (neurônios brancos). Também é interessante notar a oposição do consumo rural e da área para as outras variáveis, o que já define que deve existir um grupo de característica tipicamente urbana, e outro com característica tipicamente rural. Outra constatação que se tem por esses mapas, é de que municípios com maior área tendem a ter um maior consumo rural, mas os maiores consumidores rurais não são os municípios de maior área. Isso é uma observação um tanto óbvia, e que já poderia ser constatada na análise estatística, mas serve para lembrar o porquê do SOM ser uma ferramenta para extração de conhecimento, facilitando e agilizando a análise de uma base de dados.

Então, corroborando com os gráficos da matriz de distâncias, pode se escolher cinco grupos para a validação cruzada, um devido a grande região do mapa com baixos valores em todas as variáveis, um que concentra os municípios de maior área, outro com os

maiores consumidores rurais, outro com os maiores urbanos/industriais, e por fim um intermediário.

Com o número de grupos para o k-médias definido, o algoritmo busca a configuração com o menor erro de dispersão intra-grupo total. Esse erro é calculado somando-se as distâncias de cada dado do grupo para seu centróide, que é a média dos valores dos dados agrupados.

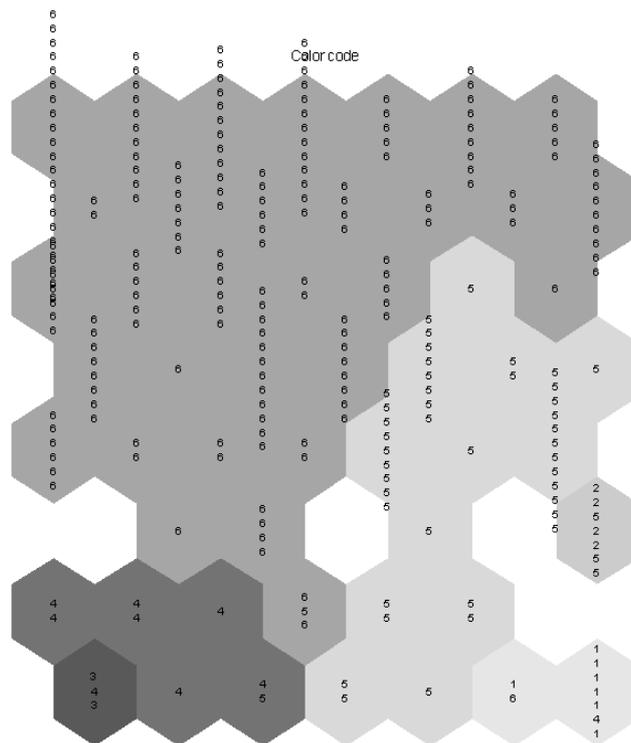
O resultado obtido pelo k-médias é então impresso sobre o mapa treinado.



**Figura 30.** Mapa 8x7 Mostrando os 5 Grupos do k-médias Sobre os Neurônios Agrupados.

O resultado da validação cruzada para 5 grupos mostrado na Figura 30 foi de 5 dados discordantes, com um erro de dispersão dentro dos neurônios de 0,474, e erro topológico de 0,05. Esse erro topológico mede a porcentagem de vezes que os dados caíram em neurônios não vizinhos durante o treinamento.

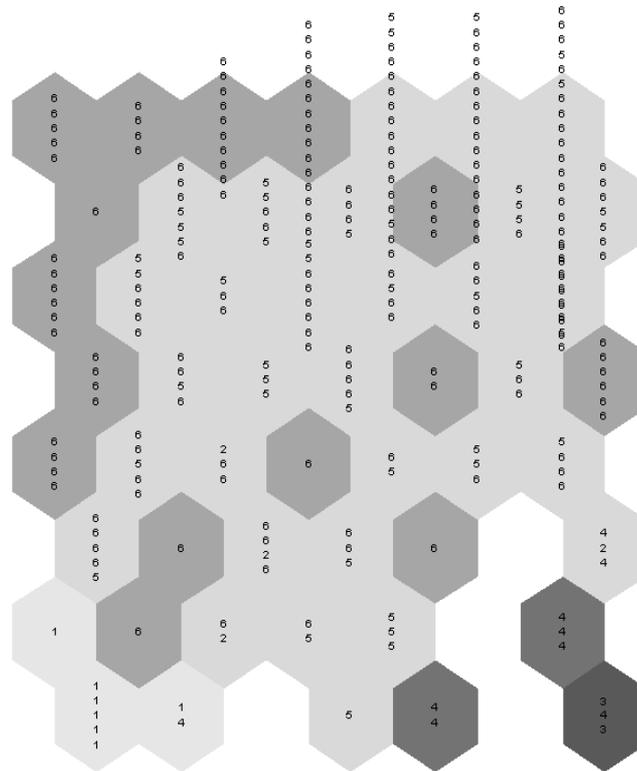
Outra informação da Figura 30, é que cinco grupos parece ser pouco para definir a variabilidade dos dados, pois o grupo maior (4) está abrangendo uma região muito grande do mapa, que engloba dados com a influência de variáveis muito diferentes, como toda a extensão vertical, onde na parte superior estão os neurônios com menores valores em todas as variáveis, e na inferior há um nível intermediário (vide Figura 29).



**Figura 31.** Mapa 8x7 Mostrando 6 Grupos do k-médias Sobre os Neurônios Agrupados.

A Figura 31 mostra um novo agrupamento, agora com 6 grupos, que basicamente criou um grupo com um nível intermediário da variável área. A discordância aumentou para 9, mas os erros continuam os mesmos, pois o mapa não foi alterado. Essa nova divisão chamou a atenção para o fato de dois grupos estarem sendo definidos principalmente pela variável área, e que o município de Lages está agrupado com outros conhecidamente de menor porte como Água Doce e São Joaquim. Isso levou a uma investigação da influência da variável Área do Município no agrupamento de mercado, pois é um dado que não está vinculado ao consumo de energia, e a classificação de Lages com esses outros municípios deverá acarretar em metas de qualidade mais rigorosas, que dificilmente poderão ser cumpridas nesses municípios de menor porte.

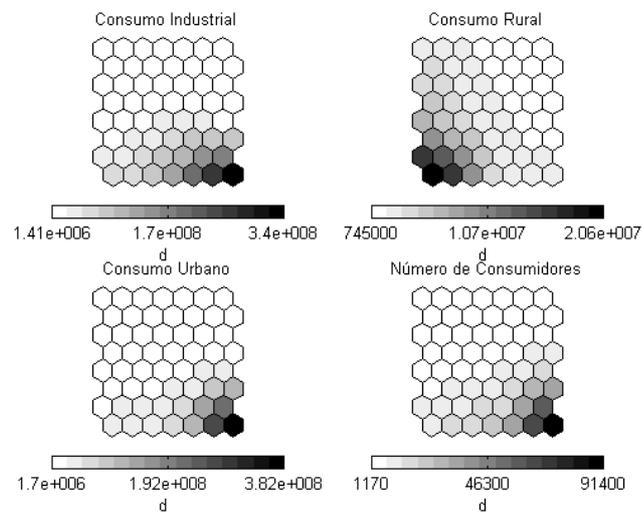
Então, o mapa foi treinado novamente, mas sem a variável Área, e a validação foi feita com o agrupamento anterior do k-médias, que continha a Área, justamente para avaliar a reorganização dos dados sem a influência dessa variável. E dessa vez, Lages fica junto de Palhoça e Balneário Camboriú, e próxima de São José e Criciúma, o que é mais coerente do ponto de vista do mercado consumidor de energia elétrica.



**Figura 32.** Reorganização ao Treinar o Mapa de Mercado sem a Variável Área.

A discordância aumentou para 43, devido ao espalhamento dos municípios atrelados a grandes áreas, os erros de dispersão e topológico foram de 0,285 e 0,058, respectivamente.

Na Figura 32, o segundo tom de cinza mais claro mostra neurônios com elementos que antes eram atrelado aos grupos com influência da área (2 e 5), que se dispersaram por quase todo o mapa treinado apenas com as variáveis de consumo.



**Figura 33.** Mapas Componentes do Treinamento Sem a Área.

Decidiu-se então excluir a variável Área do Município da base de dados utilizada para realizar os agrupamentos de Mercado.

Feito isso, seguiu-se o procedimento de validação cruzada, dessa vez buscando explorar a topologia do mapa auto-organizável.

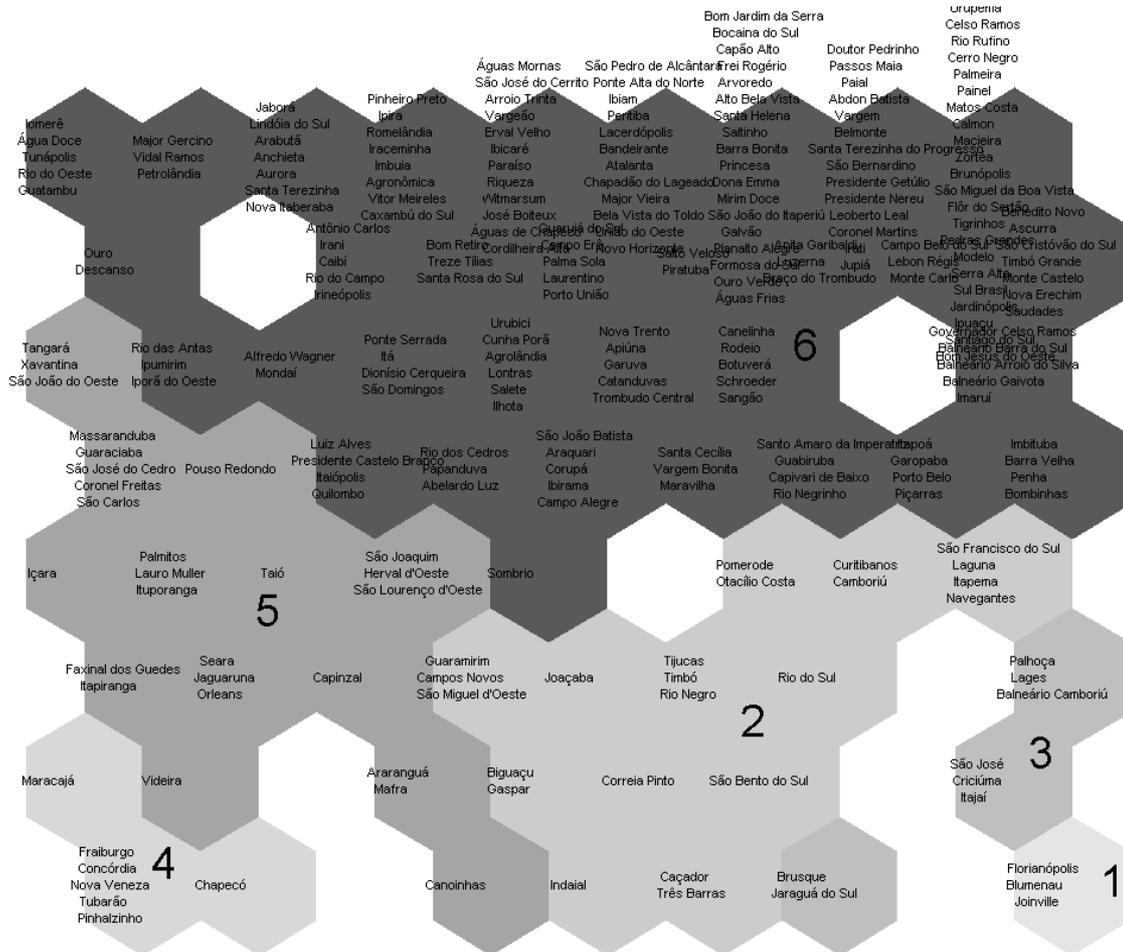
**Tabela 17.** Dados Sobre a Validação Cruzada de Mercado ao Alterar a Topologia do Mapa com 6 Grupos.

Tamanho do Mapa	Erro de Dispersão	Erro Topográfico	Discordância	Conexo
8x7	0,285	0,058	4	sim
8x8	0,271	0,050	5	sim
8x9	0,265	0,048	2	sim
9x8	0,265	0,035	6	sim
9x9	0,255	0,038	7	não
9x10	0,243	0,046	4	não
10x10	0,227	0,050	3	não
10x11	0,223	0,065	4	não

A Tabela 17 mostra alguns valores de erro da topologia do mapa auto-organizável, onde pode-se perceber que a medida que se aumenta o tamanho do mapa o erro de dispersão diminui, isso porque com mais neurônios, há mais modelos para se adaptarem aos dados, e conseqüentemente haverá menos dados agrupados em um mesmo neurônio. O erro topográfico indica a proporção dos dados em que o primeiro e segundo neurônio vencedor não foram adjacentes, mede a preservação da topologia. A discordância é a quantidade de dados rotulados pelo k-médias que caíram em neurônios definidos como de outro grupo pelo SOM, que na grande maioria das vezes estão em neurônios que fazem fronteira entre grupos. A conexão indica se os neurônios de um mesmo grupo estão próximos no mapa, grupos desconexos podem indicar uma nova partição dos dados.

Um mapa de tamanho 8x9, total de 72 neurônios, foi o escolhido para melhor representar esse agrupamento de mercado com 6 partições. Essa escolha baseou-se principalmente na discordância de apenas 2 elementos, sendo que 1 deles é Chapecó, que foi classificada em um grupo do k-médias que ficou desconexo, mas isso aconteceu em todas as outras topologias.

Mapas maiores que o 9x9 começaram a ficar desconexos, surgindo neurônios vazios separando grupos. Mas isso é devido à presença de municípios com valores discrepantes dos dados de consumo, não valendo definir um número maior de grupos.



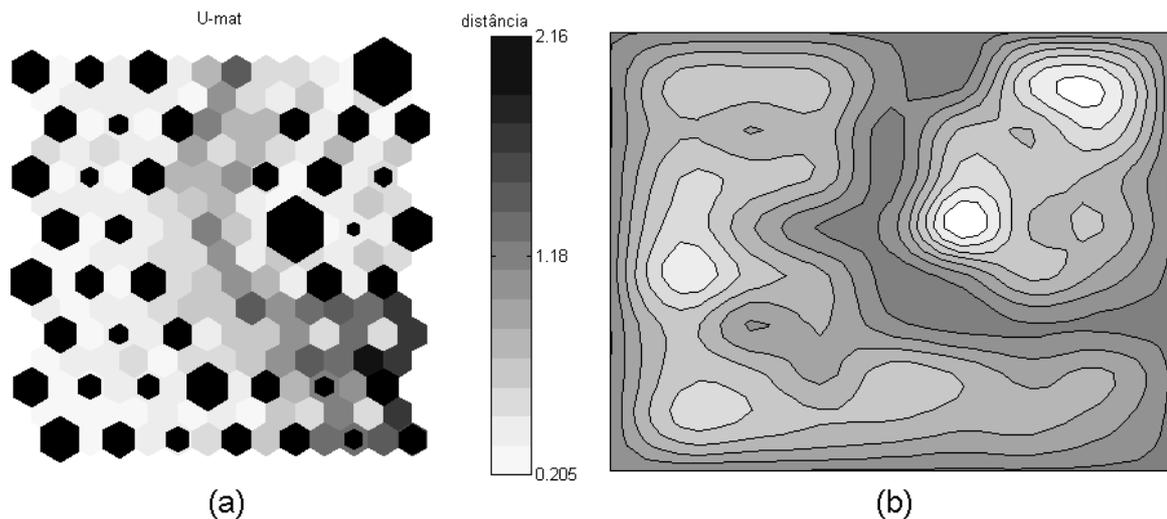
**Figura 34.** Mapa 8x9 Escolhido Pela Validação Cruzada para os Grupos de Mercado.

A configuração apresentada na Figura 34 mostrou-se a melhor para compreender a relação entre os dados agrupados. Na parte inferior esquerda tem-se o grupo dos grandes consumidores rurais; seguidos por um grupo de consumo rural médio com até um pouco de consumo industrial, no caso de Canoinhas, Araranguá e Mafra; na parte centro-inferior estão os municípios que tem características intermediárias de todos os consumos; e na direita os maiores consumidores urbanos e/ou industriais; e o grupo maior na parte superior do mapa é formado pelos menores consumidores de todas as classes.

Cabe lembrar que o grupo formado apenas por Florianópolis, Blumenau e Joinville tem um consumo aproximadamente 5 vezes maior do que os membros do grupo adjacente, que tem São José, Criciúma, Lages, etc.

### 4.3.2. Agrupamento da Base do Sistema Elétrico

Seguindo o mesmo processo realizado para agrupar a base de mercado, agora com a base contendo as quatro variáveis escolhidas para representar o sistema elétrico de fornecimento.



**Figura 35.** Matriz de Distâncias do Sistema: (a) U-mat com densidade; (b) Histogramas suavizados.

Avaliando os gráficos apresentados na Figura 35, logo percebe-se uma grande divisão entre dois grandes grupos, e também uma maior quantidade de pequenos grupos do que no caso do mercado. Parece haver também uma distribuição mais uniforme dos dados entre os neurônios, isso se deve ao fato de que nessa base não existem valores muito discrepantes.

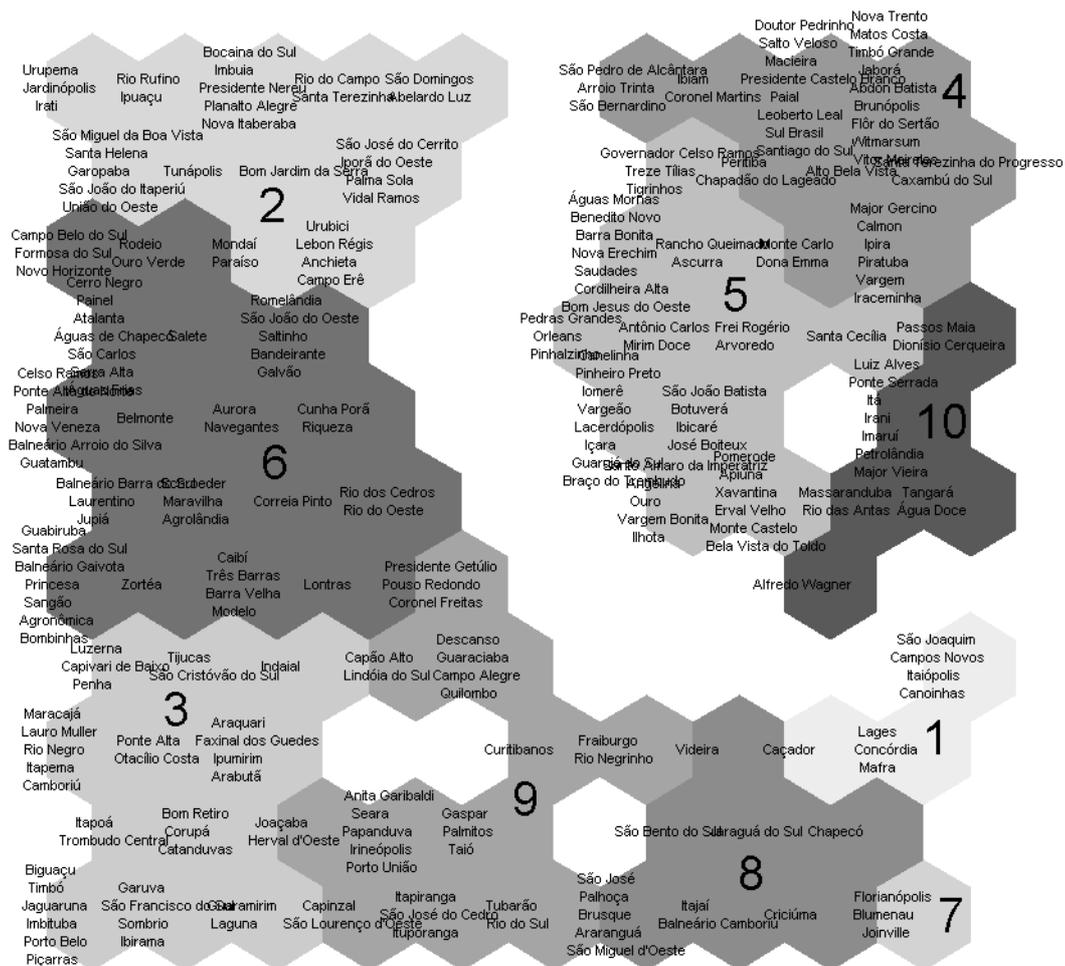
A estimativa inicial para o número de grupos a serem formados pelo k-médias foi de 8, porém ao longo do processo se resolveu aumentar para 10, acima disso, os novos grupos continham apenas 2 ou 3 elementos.

Os valores dos erros ao mexer na topologia do mapa são apresentados na Tabela 18, a seguir, onde está destacado o tamanho do mapa escolhido para representar os agrupamentos dos municípios pelo sistema elétrico.

**Tabela 18.** Dados Sobre a Validação Cruzada do Sistema ao Alterar a Topologia do Mapa com 10 Grupos.

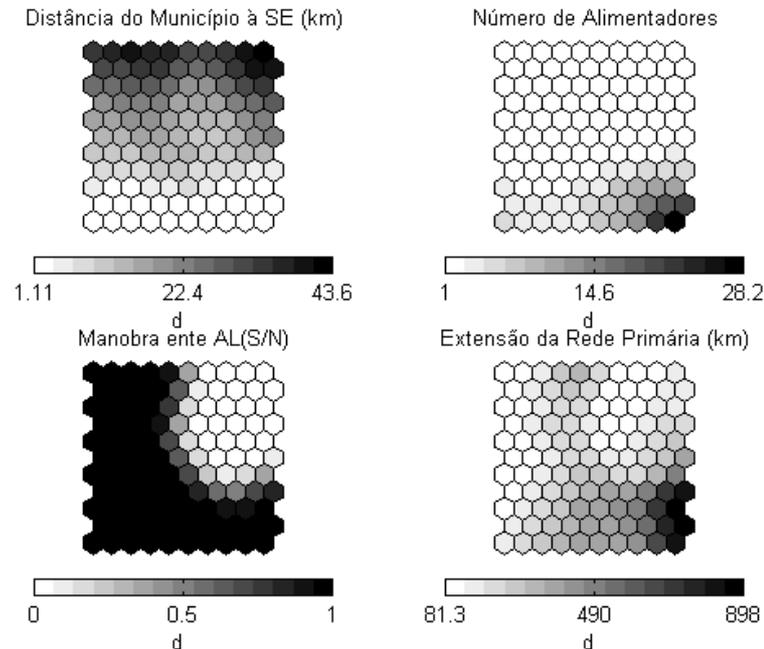
Tamanho do Mapa	Erro de Dispersão	Erro Topográfico	Discordância	Conexo
8x7	0,476	0,031	13	sim
8x8	0,434	0,027	9	não
8x9	0,420	0,015	10	não
9x8	0,412	0,027	14	não
10x8	0,373	0,023	16	sim
8x10	0,393	0,012	10	não
10x9	0,353	0,019	9	não
10x10	0,341	0,019	6	sim
11x10	0,319	0,038	4	sim
11x11	0,314	0,038	5	sim

Dos 4 municípios discordantes, 2 não ficavam na fronteira de seus grupos, eram esses Tubarão e Coronel Freitas, e foram agrupados de acordo com o grupo de seus neurônios vencedores, ambos no 9.



**Figura 36.** Mapa 11x10 Escolhido Pela Validação Cruzada para os Grupos de Sistema.

Estando definidos os grupos, usam-se novamente os mapas componentes para explicar a variável que contribuí com maior força para as formações.



**Figura 37.** Mapas Componentes 11x10 das Variáveis de Sistema.

A Figura 37 mostra a distribuição de intensidade de cada uma das 4 variáveis utilizadas no treinamento do mapa auto-organizável, onde se destaca o mapa de manobra entre alimentadores, que explica a divisão dos dois grupos maiores apontados na Figura 35(b). Pode-se ver também que na região dos municípios sem manobra (branco), estão também aqueles com maior distância para a SE, e ainda possuem apenas 1 ou 2 alimentadores. Isso se explica pela pequena extensão da rede aérea primeira, ou seja, são municípios de pequeno porte. Conseqüentemente esses municípios devem ser os que possuem os piores indicadores de qualidade.

Do mesmo modo, municípios situados na parte superior esquerda do mapa, estão muito distantes da SE ( $\approx 40\text{km}$ ), e ainda que possuam capacidade de manobra, devem ter um FEC muito alto.

Os municípios com maior sistema de fornecimento estão situados na região inferior direita do mapa, que também devem ser os maiores consumidores. Isso será verificado ao cruzar os grupos de mercado com os de sistema, a seguir.

## - CAPÍTULO 5 -

### CLASSIFICAÇÃO E ANÁLISE DOS GRUPOS FORMADOS

Neste capítulo será feita uma classificação dos agrupamentos, formados a partir das características do Mercado Consumidor e pelas características representativas do Sistema Elétrico de Fornecimento, de acordo com os indicadores de qualidade dos municípios sob concessão da CELESC.

Após essa classificação pela qualidade das duas bases de dados, é apresentado o cruzamento obtido entre elas, gerando novos subgrupos, de onde podem ser observados os diferentes sistemas de fornecimento que abastecem municípios considerados como tendo um mesmo perfil de consumo de energia elétrica.

Em seguida, serão averiguadas as ocorrências de conjuntos que não conseguiram cumprir as metas determinadas pelo órgão regulador, e assim, poderão ser identificados casos em que houve incompatibilidade da meta com a realidade do conjunto, ou se de fato, houve problemas por parte da concessionária.

## 5.1. Classificação dos Agrupamentos pela Qualidade

### 5.1.1. Ordenação da Base de Mercado

Com os grupos obtidos no item 4.3.1, foram calculadas as médias dos índices de qualidade (DEC e FEC) para cada um, bem como o desvio padrão. Então, os grupos foram classificados em ordem decrescente de qualidade e rotulados em ordem alfabética, da menor para maior média.

**Tabela 19.** Classificação dos Grupos de Mercado.

Grupo	Quantidade	Média DEC	Desvio DEC	Média FEC	Desvio FEC	Qualidade
1	3	12,55	2,35	11,72	2,14	A
3	8	13,53	5,89	12,30	4,33	B
4	7	17,83	12,59	15,28	4,39	C
2	23	20,22	7,47	16,92	6,72	D
5	27	26,12	10,18	23,03	13,77	E
6	192	39,91	23,17	29,55	16,86	F

A Tabela 19 mostra a classificação final dos seis grupos de Mercado, na primeira coluna está o número atribuído ao grupo pelo algoritmo das k-médias, e validado em conjunto com um mapa auto-organizável de dimensão 8x9, na seqüência aparecem a média e o desvio dos indicadores de qualidade, e por último, a letra que vai passar a identificar esse grupo.

A diferença entre os indicadores dos grupos A e B é pequena, apesar do desvio ser um pouco maior no B, mas vale lembrar que o perfil de consumo dos municípios agrupados em A é muito superior. Já o desvio no DEC do grupo F, e a grande quantidade de dados agrupados, parecem indicar que poderia existir um grupo a mais, mas isso será detectado ao cruzar estes grupos com os de Sistema Elétrico.

O que a grande quantidade de elementos no grupo F demonstra é que 73% dos municípios estudados, que compõem o estado de Santa Catarina, caracterizam-se como pequenos consumidores de energia elétrica, e tem uma qualidade de fornecimento pior do que conjuntos com predominância do consumo rural como o 4 e o 5.

### 5.1.2. Ordenação da Base do Sistema Elétrico

O mesmo procedimento de classificação foi aplicado para os dez grupos relativos às características do Sistema Elétrico. A ordenação final de A a J é a seguinte:

**Tabela 20.** Classificação dos Grupos de Sistema.

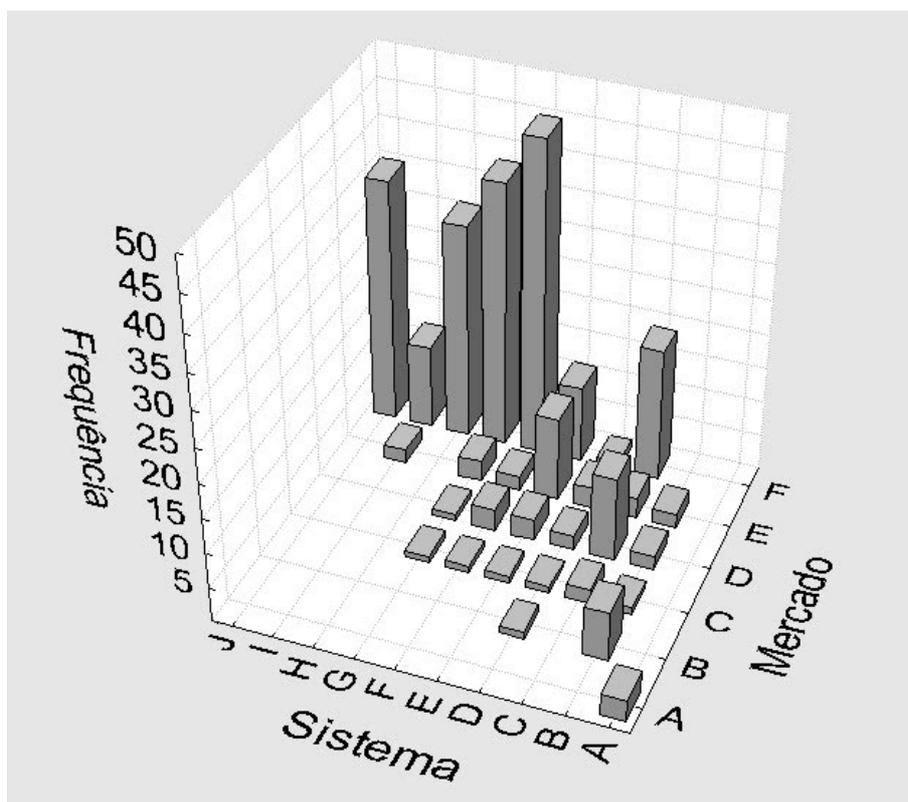
Grupo	Quantidade	Média DEC	Desvio DEC	Média FEC	Desvio FEC	Qualidade
7	3	12,55	2,35	11,72	2,14	A
8	12	13,28	4,67	12,52	3,46	B
3	36	22,32	12,13	17,06	7,62	C
1	8	28,78	11,53	20,48	8,33	D
9	27	30,37	17,01	22,18	8,70	E
6	51	32,02	14,97	24,69	12,80	F
5	43	36,19	17,61	28,81	14,39	G
2	31	39,97	22,86	28,94	10,73	H
10	14	42,92	12,12	31,33	6,35	I
4	35	57,82	32,91	43,69	26,63	J

Os grupos D,E e F têm valores médios dos indicadores um tanto próximos, mas desvios diferentes, o D tem os menores desvios, o E tem um desvio no DEC acentuado, e o F um desvio do FEC maior que dos outros dois. Ao buscar as características que os definem nos mapas das Figuras 36 e 37, vê-se que o grupo 6 (F) contém municípios mais distantes da SE, o que de certa forma explica os maiores índices de FEC. A diferença entre os grupos 1 e 9 (D e E) está principalmente na extensão da rede aérea primária, que é maior no grupo 1. Interessante notar que esse grupo de cidades relativamente grandes como Lages e Concórdia tem uma qualidade média pior (D) do que o grupo 3 (C), que é formado por cidades como Catanduvas, Porto Belo e Timbó. Isso é explicado pelo fato dessas cidades menores abastecerem também cidades vizinhas, possuindo um sistema robusto para seu fornecimento próprio.

Os grupos com piores indicadores, são os que contêm municípios sem capacidade de manobra entre alimentadores (5, 10 e 4), e de maior distância da SE (2 e 4), o que realça a importância da adoção dessas duas variáveis para caracterizar a qualidade do sistema de fornecimento, e que foram deixadas de lado nos critérios adotados pelo órgão regulador quando foram determinadas as metas de desempenho.

## 5.2. Cruzamento dos Grupos de Mercado e de Sistema

Agrupados de acordo com suas características de consumo e de sistema elétrico, e ordenados de acordo com os indicadores médios de cada grupo, as bases de dados foram cruzadas para verificar a quantidade de subgrupos de Mercado-Sistema que existem.



**Figura 38.** Histograma com a Frequência do Cruzamento de Grupos de Mercado e Sistema.

A Figura 38 trás uma série de informações a respeito das características dos conjuntos consumidores estudados. Nela se vê um plano cruzando os grupos classificados de Mercado e de Sistema, e barras mostrando a quantidade de conjuntos com as mesmas definições. O número total de combinações (Mercado-Sistema) foi de 30, por coincidência é igual à quantidade definida pela ANEEL, mas não foi feito nenhum esforço para se obter esse número.

Interessante notar a variabilidade de grupos de Sistema nos Mercados de C à F, isso mostra a heterogeneidade do fornecimento para conjuntos com um perfil de consumo semelhante. Note o caso dos conjuntos com a combinação (F-C), são aqueles pequenos municípios que servem como centro de distribuição para municípios vizinhos citados anteriormente.

Conjuntos com mercado C,D e E, e com sistemas F ou G, estão com sistemas, a princípio, incompatíveis com seu perfil de consumo, assim como o B-D. Pode-se observar que a maior frequência está em grupos com qualidades equivalentes, como (A-A, B-B, D-C, E-E e F-F), com exceção dos 7 conjuntos com mercado C, que estão cada um com um sistema de qualidade diferente.

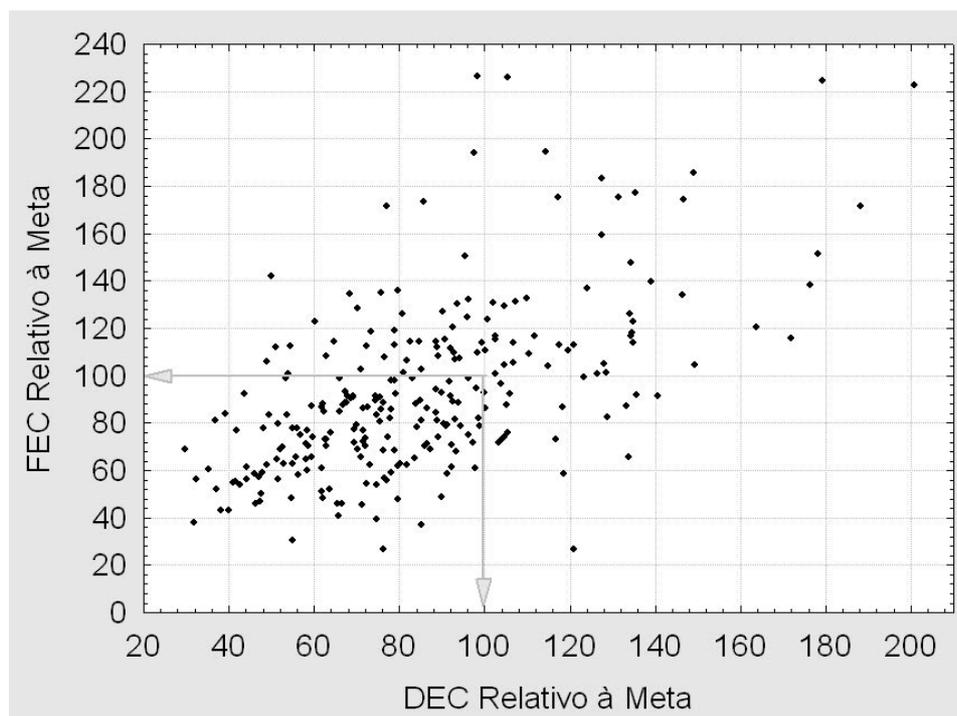
Mas na verdade, o sistema passa a ser um problema para a concessionária quando suas metas de qualidade são violadas, portanto, deve-se fazer uma busca por sistemas críticos, e então identificar onde a concessionária deve atuar, se em reforços e gerenciamento do sistema, ou negociando metas incompatíveis com o órgão regulador.

### **5.3. Avaliação dos Conjuntos Críticos**

No ano de 2003, a CELESC teve 42% dos seus conjuntos consumidores que violaram a meta de DEC ou FEC, ou ambas. É um valor muito alto, e assim sendo, a concessionária deve agir de forma a sanar os problemas. Para isso deve identificar qual a dificuldade em cumprir a meta, falhas no sistema ou metas incompatíveis devido à uma má classificação do conjunto feita pelo regulador.

Para identificar primeiramente o quão crítica é a situação de cada conjunto, e se ter uma idéia da distância de cada um em relação às metas, foi feita a divisão do indicador ocorrido em 2003 pela meta para esse ano, por conjunto. O resultado dos índices relativos foi plotado em um gráfico para facilitar e agilizar a compreensão da situação dos conjuntos críticos sob concessão da CELESC.

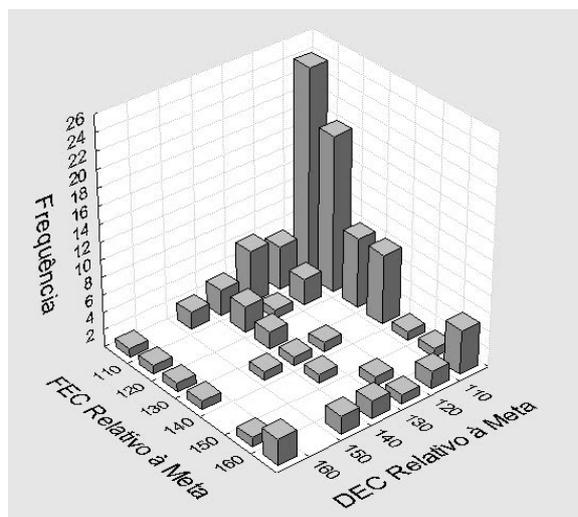
Neste gráfico (Figura 39), os pontos compreendidos entre os valores de DEC e FEC, relativos às metas, menores que 100 representam conjuntos que estão de acordo, qualquer ponto fora da área assinalada pelas flechas indica um conjunto que está violando alguma meta, e quanto mais acima de 100 for qualquer uma de suas coordenadas, mais difícil será de cumprir a meta para o próximo ano.



**Figura 39.** Dispersão dos Conjuntos em Relação às Metas de DEC e FEC.

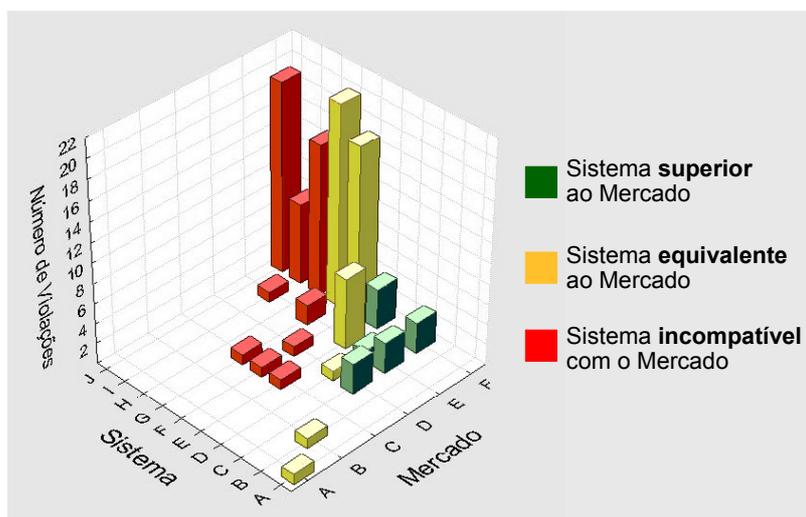
A Figura 39 mostra que a situação dos conjuntos da CELESC não é nada confortável, pois existem muitos com algum indicador de qualidade com valores de mais de 20% acima das metas. Porém, também existe uma boa quantidade de conjuntos com valores abaixo de 80% da meta, que não serão problema para os próximos anos.

O maior problema é o FEC, que registra 91 conjuntos com indicadores acima de suas metas, e ainda, os 66 conjuntos que ultrapassaram suas metas de DEC também estão bem distantes do valor adequado (Figura 40).



**Figura 40.** Histograma dos Índices Relativos Maiores que 100.

Resta agora investigar quais as combinações de Mercado-Sistema que violaram as metas, para então identificar ações a serem tomadas pela empresa distribuidora de energia elétrica. O gráfico a seguir extrai da Figura 38 apenas os conjuntos que não cumpriram suas metas de qualidade, identificando casos onde o sistema possui (ex: C-D, C-E, C-F)



**Figura 41.** Histograma do Número de Violações por Combinação Mercado-Sistema.

Observando a Figura 41, de imediato pode-se notar dois conjuntos considerados de melhor qualidade que violaram suas metas, o primeiro (A-A) é Florianópolis, que sofreu um grave acidente no sistema de fornecimento da ilha, que ficou 55h sem energia, sendo um caso excepcional e fora de questão. O outro (B-B), é Palhoça, que ultrapassou a meta de DEC em apenas 0,04 horas, e como tem um sistema de porte adequado ao seu mercado, pode estar tanto com a meta um tanto estreita por ter sido agrupado com conjuntos A-A pela ANEEL, quanto com um mal gerenciamento da operação e manutenção do sistema.

Na linha do mercado C, os conjuntos que ultrapassaram as metas foram justamente aqueles com sistemas E, F e G, que como dito anteriormente, não possuem um padrão adequado para atender o nível exigido pelo seu mercado consumidor. Por outro lado, os conjuntos com sistema C, e mercado D, E e F, podem ter sido classificados erroneamente pela ANEEL, pois possuem um sistema com um padrão de qualidade superior ao padrão requerido pelo seu mercado consumidor. Também ocorreram muitas violações em conjuntos com as configurações E-E, F-F e F-G, que possuem sistemas com padrões de qualidade equivalentes entre mercado e sistema, neste caso pode-se verificar se estes conjuntos foram classificados em grupos de qualidade muito superior, ou seja, possuem

meta incompatível, ou caso contrário, a concessionária deve investigar as causas das interrupções, e agir para prevenção de novas infrações de metas.

A maior ocorrência de violações está justamente nos municípios de menor porte, que com sistemas de G a J, somam 61 casos. São conjuntos críticos que têm tanto sistemas fracos, quanto possivelmente metas estreitas, pois boa parte ultrapassou a meta em mais de 40%. Nestes casos, a concessionária deve realizar estudos técnicos de priorização (cálculo do ITP) para determinar obras de expansão, como alimentadores e subestações.

**Tabela 21.** Comparação das Classificações com o Grupo 5 da ANEEL.

Grupo ANEEL	Classificação		DEC Relativo	FEC Relativo	Município
	Mercado	Sistema			
5	B	B	52,00	68,94	Brusque
5	B	B	47,97	59,17	Jaraguá do Sul
5	C	B	79,00	68,50	Chapecó
5	C	C	41,06	55,07	Tubarão
5	C	D	75,33	80,82	Concórdia
5	C	E	81,00	101,27	Fraiburgo
5	D	B	76,21	68,40	São Bento do Sul
5	D	B	58,61	70,44	São Miguel d'Oeste
5	D	C	84,19	78,40	Biguaçu
5	D	C	75,50	90,90	Camboriú
5	D	C	78,00	59,30	Guaramirim
5	D	C	71,36	45,29	Otacílio Costa
5	D	C	44,23	56,06	Rio do Sul
5	D	D	77,96	97,86	Caçador
5	D	D	124,03	137,00	Campos Novos
5	E	B	49,35	83,33	Araranguá
5	E	B	67,91	91,44	Videira
5	E	C	100,67	124,04	Faxinal dos Guedes
5	E	D	111,81	116,79	Canoinhas
5	E	D	74,58	53,81	São Joaquim
5	E	E	118,26	86,94	Capinzal
5	F	C	61,91	61,10	Luzerna
5	F	D	98,69	82,15	Itaiópolis
5	F	E	85,22	37,17	Anita Garibaldi
5	F	E	76,48	56,89	Rio Negrinho

A Tabela 21 mostra todos os municípios sob concessão da CELESC pertencentes ao grupo 5 da ANEEL (2002), isso quer dizer que ao final de 2008 todos eles devem estar compartilhando a mesma meta para os indicadores de qualidade. Nessa mesma tabela, está a classificação realizada pela metodologia proposta, e os indicadores relativos à meta de cada município, valores acima de 100 (destacados) indicam violação da meta.

O que se pode perceber, é que o grupo 5 da ANEEL engloba uma quantidade bem variada de municípios segundo a classificação proposta, e que os conjuntos mais críticos são justamente os menos qualificados.

Com exceção de Fraiburgo, que possui um sistema inadequado para sua classe de consumo, os municípios com indicadores relativos maiores, ou próximos de 100 podem ser considerados como tendo uma meta incompatível. Os quatro últimos conjuntos da Tabela 21, apesar de serem municípios pequenos (Mercado F), possuem um sistema conveniente, que ainda lhes permite cumprir as metas desse grupo. Essa mesma análise pode ser aplicada a todos os grupos definidos pela ANEEL, bem como ser estendida a outras concessionárias de distribuição brasileiras.

Portanto, a metodologia proposta para classificação dos conjuntos consumidores de energia elétrica pode indicar onde a concessionária deve agir, se no reforço, expansão ou manutenção do sistema, ou se negociando com o órgão regulador metas mais adequadas à realidade de conjuntos que estão deslocados em grupos definidos anteriormente.

Para sanar os problemas de conjuntos com sistemas inadequados, ou mesmo que possuam esquema de fornecimento considerado compatível mas que não atingiram o objetivo de ficar abaixo das metas, cabe a concessionária investigar as causas das interrupções, para determinar as ações a serem tomadas.

Este trabalho mostrou a importância da escolha das variáveis na formação final dos agrupamentos, como um atributo que não reflete características relacionadas ao consumo/fornecimento de energia pode interferir prejudicialmente na determinação das metas de qualidade, e por outro lado, como a escolha correta das variáveis referentes à constituição do sistema de fornecimento podem estar correlacionadas aos indicadores, tornando a classificação homogênea para a determinação de metas factíveis.

Contudo, sabe-se que ainda existem muitas outras variáveis exógenas à constituição do sistema que não foram consideradas, mas que têm grande influência na qualidade do fornecimento, especialmente no FEC, como índices pluviométricos, níveis cerâmicos, velocidade do vento, vegetação, etc.

## - CAPÍTULO 6 -

### CONCLUSÕES E SUGESTÕES

#### 6.1. Conclusões

##### 6.1.1. Sobre a Seleção das Variáveis

As variáveis, ou atributos, a serem consideradas na realização dos agrupamentos são de suma importância, ainda mais se tratando da determinação de metas a serem cumpridas pelos grupos, que se forem ultrapassadas constituem uma infração sujeita a penalidades.

O paradigma apresentado, de separar as variáveis nas que caracterizam o Mercado Consumidor das que representam o Sistema Elétrico de Fornecimento, mostrou-se uma decisão essencial para facilitar a compreensão sobre a situação dos conjuntos consumidores.

A base de dados fornecida pela CELESC possibilitou um estudo mais detalhado das características de cada conjunto. Pôde-se explorar a participação das variáveis no agrupamento dos dados, e se constatar que o uso de informação redundante reduz a influência de cada variável, mascarando a variabilidade de dados importantes.

Para formar o constructo teórico, ou a base de dados mínima que explica as principais diferenças entre os conjuntos, foi utilizada a análise estatística multivariada, que serviu muito bem ao propósito de eliminar variáveis que traziam informação redundante, reduzindo a quantidade de 10 para 5 variáveis na base de Mercado, inclusive foram criadas novas variáveis a partir da soma de outras, e de 7 para 4 na base do Sistema.

Na base de Mercado, ainda foi retirada a variável Área do Município, devido à contribuição negativa para o agrupamento, pois relacionava conjuntos com perfis de consumo distintos, prejudicando a definição de grupos homogêneos quanto ao consumo de energia elétrica.

### 6.1.2. Sobre o Método de Agrupamento

O esquema desenvolvido de validação cruzada entre o método estatístico das k-médias e o mapa auto-organizável (SOM) mostrou-se uma ferramenta flexível e confiável, pois explora as melhores características de ambos os algoritmos.

Aliando o agrupamento rígido e consagrado do k-médias, com a capacidade de extração de conhecimento e visualização do SOM, pôde-se buscar uma configuração dos grupos que melhor representasse as características dos dados agrupados, permitindo a identificação das fronteiras, assim como dos elementos mais suscetíveis a mudar de grupo. Isso é uma característica importante, pois possibilita uma reclassificação imediata, sem precisar rodar o algoritmo novamente, caso seja desejado.

O esquema apresentado também serve para mostrar que o SOM é uma excelente ferramenta de agrupamento, e que especulações sobre ser um método heurístico ou sua falta de embasamento matemático são infundadas, sendo basicamente uma resistência ao paradigma do processo competitivo não-supervisionado, ou simples falta de conhecimento para utilização correta do método.

O uso do algoritmo das k-médias isolado, exige cuidados quanto à sua inicialização e escolha do número de agrupamentos, para não ficar preso a soluções sub-ótimas ou a um número pouco expressivo, ou excessivo de grupos. Esses problemas são perfeitamente resolvidos com a combinação com o SOM, que mostra quando grupos definidos pelo k-médias ficam desconexos ou abrangem regiões que são influenciadas de maneiras distintas pelas variáveis analisadas.

### 6.1.3. Sobre a Classificação e Análise dos Grupos Obtidos

Definidos os grupos de ambas as bases, foi realizada uma classificação de acordo com o índice médio dos indicadores de qualidade de cada grupo, estabelecendo assim uma ordem qualitativa para os grupos, o que foi importante para uma perfeita compreensão dos gráficos obtidos a partir do cruzamento entre as bases.

Apesar da base de Mercado parecer ter poucos grupos em relação à de Sistema, quando o cruzamento foi realizado, o resultado obtido foi muito satisfatório.

Ao realizar o cruzamento entre os grupos de Mercado e do Sistema, foi possível encontrar as configurações de Consumo/Fornecimento dos conjuntos sob concessão da CELESC. Dos 6 grupos de Mercado com os 10 do Sistema surgiram 30 combinações diferentes, que foi igual ao número de grupos definidos pela ANEEL, sem ter sido esse o objetivo.

O histograma mostrando tal cruzamento (Figura 38) já indicava possíveis configurações com um sistema deficitário para seu mercado consumidor que poderiam acarretar em violações das metas, sendo essa hipótese comprovada com o histograma da Figura 41. Nesses casos a empresa deve agir de modo a fornecer um sistema mais adequado às exigências do conjunto consumidor.

Por outro lado, conjuntos com sistemas apropriados ao seu mercado, mas que transgrediram suas metas, podem estar erroneamente classificados em algum grupo da ANEEL, isso porque tal classificação considerou poucos atributos, sendo um inconsistente com as características elétricas, que foi a Área, como foi mostrado.

Logo, a abordagem proposta cumpriu seus objetivos, consolidando uma metodologia para classificação e análise de conjuntos consumidores de energia elétrica, que engloba desde a seleção das variáveis adequadas para realizar o agrupamento dos conjuntos, validado por dois algoritmos diferentes, até a identificação de conjuntos críticos, sinalizando para a concessionária onde deve realizar investimentos para compatibilizar o sistema de fornecimento com o mercado consumidor.

Com isso a empresa pode rever as tarifas de maneira a remunerar os investimentos indispensáveis à prestação do serviço de acordo com a qualidade regulamentada.

## **6.2. Sugestões para Trabalhos Futuros**

- Implementar um software que incorpore o modelo completo da metodologia proposta, desde a seleção das variáveis até a identificação dos conjuntos críticos;
- Avaliar outras variáveis, como as características geo-climáticas;
- Investigar um sistema de penalização, que sinalize às empresas a vantagem de investimentos em obras que eliminem as violações observadas, quando não for o caso de metas incompatíveis.

## REFERÊNCIAS BIBLIOGRÁFICAS

- AAKER, D.A., KUMAR, V., DAY, G., 2001, **Pesquisa de Marketing.**, São Paulo, Ed. Atlas.
- ANDERBERG, M.R., 1973, **Cluster Analysis for Applications**, Academic Press.
- ANEEL, 2000, Resolução 024.
- ANEEL, 2003, Resolução 075.
- AZUAJE, F., 2002: Discovering Genomic Expression Patterns with Self-Organizing Neural Networks, In: **Understanding and Using Microarray Analysis Techniques: A Practical Guide**, Berrar D, Dubitzky W and Granzow M, editors, Springer Verlag, London.
- BARALDI, A., SCHENATO, L.: Soft-to-Hard Model Transition in Clustering: a review, *Tech. Report 010*, Int. Comput. Sci. Inst., Berkeley, USA,
- BILLINTON, R., PAN, Z.: Incorporating Reliability Index Probability Distributions in Performance Based Regulation, *Proc. of the Canadian Conference on Electrical & Computer Engineering*, pp: 12 – 17, Canada, 2002.
- CARDON, P., LENDASSE, A., WERTZ, V., de BODT, E., VERLEYSSEN, M., 2002: Classification de Fonds Communs d'Investissement par Cartes Auto-Organisées, *Proc. of Connectionist Approaches in Economics and Management Sciences (ACSEG)*, pp. 201-212, Boulogne-sur-Mer, France.
- CHANG, M., YU, H.J., HEH, J.S., 1998: Evolutionary Self-Organizing Map, *Proc. IEEE World Congress on Computational Intelligence, Int'l Joint Conf. on Neural Networks (IJCNN)*, vol. 1, pp: 680 – 685, Anchorage, USA.
- CHANTELOU, D., HÉBRAIL, G., MULLER, C., 1996: Visualizing 2665 Electric Power Load Curves on a Single A4 Sheet of Paper, *Proc. International Conference on Intelligent Systems Applications to Power Systems (ISAP)*, pp:126–132, Orlando, USA.
- COELHO, J., GAUCHE, E., QUEIROZ, H., NASSAR, S.M., RICARDO, V.W., LIMA, M., 2002: Influence of Weather Variables in Continuity Levels of Electrical Power Supply – An Analysis through Artificial Neural Networks, *Simpósio de Especialistas em Planejamento da Operação e Expansão Elétrica (SEPOPE)*, Curitiba, Brasil.
- COELHO, J., OLGUÍN, G., 2000: Propuesta de Reglamentación de la Calidad de Servicio en Sistema de Distribución, *IV Cong. Latino Americano de Generación y Transporte de Energía Eléctrica*, Viña del Mar, Chile.

- COTTRELL, M., de BODT, E., VERLEYSEN, M., 2001: A Statistical Tool to Assess the Reliability of Self-Organizing Maps; *Proc. of Workshop on Self-Organizing Maps (WSOM)*, Lincoln, United Kingdom.
- COTTRELL, M., FORT, J.C., PAGÈS, G., 1995: Two or Three Things that we Know about the Kohonen Algorithm, *Proc. European Symp. on Artificial Neural Networks (ESANN)*, Brussels, Belgium.
- DEBOECK, J.G., 1998: Financial Applications of Self-Organizing Maps, *American Heuristics Electronic Newsletter*.
- DE OLIVEIRA, M., CAMARGO, I., FIGUEIREDO, F., TANURE, J.E., 1999: Power Quality Standards in Brazil During Electricity Industry Restructuring Program, *IEEE Power Engineering Society Summer Meeting*, vol.1, pp:18–22, Vancouver, Canada.
- DNAEE, 1978, Portaria nº 46.
- FLEXER, A., 2001: On the Use of Self-Organizing Maps for Clustering and Visualization, *Intelligent Data Analysis* (5), pp: 373–384.
- EMANIAN, V., KAVEH, M., TEWFIK, A., 2000: Robust Clustering of Acoustic Emission Signals Using the Kohonen Network, *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp: 3891–3894, Istanbul, Turkey.
- ERWIN, E., OBERMAYER, K., SCHULTEN, K., 1992a: Self-Organizing Maps: Ordering, Convergence Properties and Energy Functions, *Biological Cybernetics*, 67(1):47-55.
- ERWIN, E., OBERMAYER, K., SCHULTEN, K., 1992b: Self-Organizing Maps: Stationary States, Metastability and Convergence Rate, *Biological Cybernetics*, 67(1):35-45.
- HAYKIN, S., 2001, **Redes Neurais : Princípios e Práticas**, Bookman, 2<sup>a</sup> ed.
- HSU, Y.Y., YANG, C.C., 1991: Design of Artificial Neural Networks for Short-Term Load Forecasting: Self-Organizing Feature Maps for Day Type Identification, *IEE*.
- KAGAN, N., de OLIVEIRA, C.C.B., 2000: Planejamento de Longo Prazo Considerando Indicadores de Qualidade Através de Ações de Manutenção e Operação e de Investimentos de Expansão do Sistema, *III Seminário Internacional de Construção e Manutenção de Sistemas de Distribuição e Subtransmissão de Energia Elétrica (CIERTEC)*, Florianópolis, Brasil.
- KASKI, S., KANGAS, J., KOHONEN, T., 1998: Bibliography of Self-Organizing Map (SOM) Papers: 1981 – 1997, *Neural Computing Surveys*, vol.1, pp:102-350.

- KOHONEN, T., 1990: The Self-Organizing Map, *Proc. of the IEEE*, vol.78, no.9, pp:1464–1480.
- KOHONEN, T., SIMULLA, O., 1996: Engineering Applications of the SOM, *Proc. of the IEEE*, vol.84, n.10.
- KOHONEN, T., 1999: Analysis of Processes and Large Data Sets by a Self-Organizing Method, *Proc. of the IEEE*.
- KOHONEN, T., 2001, **Self-Organizing Maps**; Springer-Verlag, 3<sup>a</sup> ed.
- LIMA, W.S., de CASTRO, L.N., OHISHI, T., 1999: Mapas Auto-Organizáveis Não-Paramétricos para Análise da Influência Climática em Curvas de Carga, *XV Seminário Nacional de Produção e Transmissão de Energia Elétrica (SNPTEE)*, Foz do Iguaçu, Brasil.
- LO, Z.P., BAVARIAN, B., 1991a: Improved Rate of Convergence in Kohonen Neural Network, *Proc. of International Joint Conference on Neural Networks (IJCNN)*, vol.II, pp: 201 – 206 , Piscataway, USA.
- LO, Z.P., FUJITA, M., BAVARIAN, B., 1991b: Analysis of Neighborhood Interaction in Kohonen Neural Networks, *Proc. Fifth International Parallel Processing Symposium*, pp: 246 – 249, Anaheim, USA.
- LO, Z.P., FUJITA, M.; BAVARIAN, B., 1992: Two Theorems for the Kohonen Mapping Neural Network, *International Joint Conference on Neural Networks (IJCNN)*, vol.4 , pp: 755 – 760, Baltimore, USA.
- LO, Z.P., YU, Y., BAVARIAN, B., 1993: Analysis of the Convergence Properties of Topology Preserving Neural Networks. *IEEE Trans. on Neural Networks*, 4(2):207 – 220.
- LUO, X.; SINGH, C.; PATTON, A.D., 1999: Loss-of-load state identification using self-organizing map, *Power Engineering Society Summer Meeting IEEE*, vol.2, pp:670 - 675, Edmonton, Canada.
- MICHIE, D., SPIEGELHALTER, D.J., TAYLOR, C.C., 1994, Machine Learning, Neural and Statistical Classification, (freely available on the web).
- NETO, A.S., LONGUE, C.A., SANTOS, E.C., 2003: Metodologia para Reavaliação dos Limites de Indicadores de Performance dos Conjuntos Elétricos – ANEEL, *Anais do V Seminário Brasileiro sobre Qualidade da Energia Elétrica (SBQEE)*, vol.2, pp: 463 – 467, Aracaju, Brasil.

- OBU-CANN, K., FUJIMURA, K., and TOKUTAKA, H., 1999: Clustering by SOM (self-organising maps), MST (minimal spanning tree) and MCP (modified counter propagation). *Proc. 6th International Conference on Neural Information Processing (ICONIP)*, vol.3, pp: 986 – 991, Perth, Australia.
- OBU-CANN, K., FUJIMURA, K., TOKUTAKA, H., *et al.*, 2001: Data Mining of Power Transformer Database Using Self-Organizing Maps, *Proc. International Conferences on Info-tech and Info-net (ICII)*, vol.4, pp: 44–49 , Beijing, China.
- PAMPALK, E., RAUBER, A., MERKL, D., 2002: Using Smoothed Data Histograms for Cluster Visualization in Self-Organizing Maps, *Proc. of the International Conference on Artificial Neural Networks*, Springer Lecture Notes in Computer Science, Madrid, Spain.
- PAL, N.R., BEZDEK, J.C., TSAO, E.C.K., 1993: Generalized Clustering Networks and Kohonen's Self-Organizing Scheme. *IEEE Transactions on Neural Networks*, 4(4):549 – 557.
- QUEIROZ H.L., 2002: Metodologia para Definição de Agrupamento de Consumidores e os Requisitos de Qualidade do Fornecimento, *XV Seminário Nacional de Distribuição de Energia Elétrica (SENDI)*, Salvador, Brasil.
- QUEIROZ, H.L., BORBA, C.M., 2001: Metodologia para Avaliação dos Conjuntos de Consumidores por Meio dos Indicadores Técnicos de Qualidade, *Seminario Internacional de Planeamiento y Calidad de Sistemas de Distribución Eléctrica*, Puerto Iguazú, Argentina.
- ROMÁN, J., GÓMEZ, T., MUÑOZ, A., PECO, J., 1998: Regulation of Distribution Network Business, *IEEE Trans. on Power Delivery*.
- SADEGHI, A.A., 2001: Convergence in Distribution of the Multi-Dimensional Kohonen Algorithm. *Journal of Applied Probability*, 38(1):136 – 151.
- SESTER, M., 2002: Typifizierung mittels Kohonen Merkmalskarten, *Mitteilungen des Bundesamtes für Kartographie und Geoinformation*, Bande 22, pp: 161-169, Frankfurt, Deutschland.
- SCHUMANN, A., 1997: Neural Networks versus Statistics: a Comparing Study of their Classification Performance on Well Log Data, *Proc. of International Association for Mathematical Geology (IAMG)*, part 1, pp: 237 – 241, Barcelona, Spain.

- SPERANDIO, M., COELHO, J., QUEIROZ, H.L., 2003: Identificação de Agrupamentos de Consumidores de Energia Elétrica através de Mapas Auto-Organizáveis, *Anais do V Seminário Brasileiro sobre Qualidade da Energia Elétrica (SBQEE)*, vol.2, pp: 439-443, Aracaju, Brasil.
- SPERANDIO, M., COELHO, J., QUEIROZ, H.L., *et al.*, 2004: Revisão dos Critérios para Agrupamentos de Conjuntos Consumidores de Energia Elétrica, *IX Simpósio de Especialistas em Planejamento da Operação e Expansão Elétrica(SEPOPE)*, SP-34, Rio de Janeiro, Brasil.
- SPIEGEL, M.R., 1985, **Estatística**, McGraw-Hill, 2<sup>a</sup> Ed.
- SPITZER, A., POLANI, D., 1998: Order Parameters for Self-Organizing Maps, *Proc. of the 8th Int. Conf. on Artificial Neural Networks (ICANN)*, vol. 2, pp: 517 – 522 , Skövde, Sweden.
- TANURE, J.E.P.S. e de CARVALHO, E.B., 2000: “Regulação por Comparação de Desempenho para o Estabelecimento de Metas de Continuidade de Fornecimento”, *Anais do XIV Seminário Nacional de Distribuição de Energia Elétrica*, pp: 1 – 15 , Foz do Iguaçu, Brasil.
- TANURE, J.E.P.S., *Análise Comparativa de Empresas de Distribuição para o Estabelecimento de Metas de Desempenho para Indicadores de Continuidade do Serviço de Distribuição*. Itajubá, 2000. Dissertação (Mestre em Ciências em Engenharia Elétrica) – Escola Federal de Engenharia de Itajubá.
- TÖRMÄ, M., 1993: Comparison Between Three Different Clustering Algorithms, *Photogrammetric Journal of Finland*, vol. 13, No. 2, pp: 85 – 95 , Espoo, Finland.
- TÖRMÄ, M., 1994: Kohonen Self-Organizing Feature Map and its Use in Clustering, *Proc. of the International Society for Optical Engineering(SPIE)*, vol.2357, pp: 830 – 835, Ukraine.
- ULTSCH, A., SIEMON, H.P., 1990: Kohonen’s Self-Organizing Feature Maps for Exploratory Data Analysis, *Proc. Intern. Neural Networks*, pp: 305 – 308, Paris, France.
- ULTSCH, A., KORUS, D., GUIMARAES G., 1994: Integration von Neuronalen Netzen mit Wissensbasierten Systemen; *in Proc. Cottbuser Workshop Anwendungen Neuronaler Netze (CoWANN)*, Cottbuser, Deutschland.

- ULTSCH, A., 1996: Self Organizing Neural Networks Perform Different from Statistical k-means Clustering . In: *M. van der Meer, R. Schmidt, G. Wolf, (Eds.): BMBF Statusseminar "Künstliche Intelligenz, Neuroinformatik und Intelligente Systeme"*, pp: 433 – 443, München, Deutschland.
- UTSUGI, A., 1996: Topology Selection for Self-Organizing Maps, *Network Computation in Neural Systems*, 7(4):727 – 740 .
- VERLEYSSEN, M., FRANÇOIS, D., SIMON, G., WERTZ, V., 2003: On the effects of dimensionality on data analysis with neural networks. *IWANN'03 Artificial And Neural Networks Workshop*, pp: 105-112, Menorca, Spain.
- VESANTO, J., HIMBERG, J., ALHONIEMI, E., PARHANKANGAS, J., 1999: Self-Organizing Map in Matlab: the SOM Toolbox, *In Proc. of the Matlab DSP Conference*, pp: 35–40, Espoo, Finland.
- VESANTO, J., 1999: SOM Based Data Visualization Methods, *In Intelligent Data Analysis*, Vol.3, n.2, Elsevier Science, pp. 111-126.
- VESANTO, J., 2002: Data Exploration Process Based on the Self-Organizing Map, Espoo, Finland, Dissertation for the degree of Doctor of Technology, Helsinki University of Technology.
- VOITOVETSKY, I., GUTERMAN, H., COHEN, A., 1997: Unsupervised Speaker Classification using Self-Organizing Maps (SOM), *in VII IEEE Workshop on Neural Networks for Signal Processing*, pp:578–587, Florida, USA.
- WISHART, D., 2003, [http://www.clustan.com/k-means\\_critique.html](http://www.clustan.com/k-means_critique.html)
- XUEGONG ZHANG, YANDA LI, 1993: Self-Organizing Map as a New Method for Clustering and Data Analysis, *Proc. of Int. Joint Conf. on Neural Networks (IJCNN)*, vol. 3 , pp: 2448–2451.
- YIN, H., ALLINSON, N.M., 1995: On the Distribution and Convergence of Feature Space in Self-Organizing Maps. *Neural Computation*, 7(6):1178 – 1187.