

UNIVERSIDADE FEDERAL DE GOIÁS
ESCOLA DE ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA E
DE COMPUTAÇÃO

*Modelo de Análise Espacial de Mercado de
Energia Elétrica Utilizando Técnicas de
Classificação*

João Caetano Neto

Orientador: Prof. Dr. Wagner da Silva Lima

Goiânia

2007

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

©Copyright 2007
João Caetano Neto

JOÃO CAETANO NETO

***Modelo de Análise Espacial de Mercado de
Energia Elétrica Utilizando Técnicas de
Classificação***

Dissertação apresentada ao Curso de Mestrado em Engenharia Elétrica e de Computação da Escola de Engenharia Elétrica e de Computação da Universidade Federal de Goiás, para obtenção do título de Mestre em Engenharia.

Área de Concentração: Engenharia de Computação

Orientador: Prof. Dr. Wagner da Silva Lima

Goiânia

2007

Universidade Federal de Goiás
Escola de Engenharia Elétrica
Coordenação do Programa de Pós-Graduação em
Engenharia Elétrica e de Computação - CPPEEC

FOLHA DE APROVAÇÃO

**”Modelo de Análise Espacial de Mercado de
Energia Elétrica Utilizando Técnicas de
Classificação”**

JOÃO CAETANO NETO

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:

Prof. Dr. Wagner da Silva Lima, Orientador - EEEEC/UFG

Prof. Dr. Carlos Márcio Vieira Tahan - EPUSP

Prof. Dr. Pedro José Abrão - CEFET/GO

Prof. Dr. Leonardo da Cunha Brito - EEEEC/UFG

Goiânia, 30 de Março de 2007

Dedicatória

Primeiramente a Deus por todo seu amor e graça;

À minha mãe pelo amor e pela luta para me proporcionar as oportunidades da vida;

À minha querida esposa por seu incentivo e apoio;

À minha filha Ana Clara, motivo de grandes alegrias e felicidade contínua.

Agradecimentos

Ao Prof. Wagner pela orientação e todo o apoio durante a execução deste trabalho;

À Companhia Energética de Goiás (CELG) por fornecer as informações e dados necessários para a realização deste trabalho.

Aos colegas da CELG por proporcionarem um ambiente propício para a realização desta pesquisa e por nunca negarem a gastar um pouco do seu tempo para contribuir com este projeto;

À minha família que sempre me incentivou nessa jornada e por várias vezes abriram mão da minha companhia;

À Deus por sempre iluminar meu caminho e abençoar minha vida.

*"Há homens que lutam um dia e são bons.
Há outros que lutam um ano e são melhores.
Há os que lutam muitos anos e são muito bons.
Porém, há os que lutam toda a vida.
Esses são os imprescindíveis."*

Bertolt Brecht

Resumo

As companhias de distribuição de energia elétrica convivem com uma realidade antagônica; um crescente aumento da demanda energética, devido ao constante crescimento econômico e populacional, e uma limitação em seus recursos financeiros para expansão de sua rede. Assim sendo, torna-se imprescindível um aprimoramento na atividade de planejamento do sistema de distribuição de energia procurando aperfeiçoar a aplicação dos recursos disponíveis.

Neste contexto se enquadra a aplicação do Sistema de Informação Geográfica aliado a técnicas de agrupamento e classificação com o objetivo de potencializar o processo de planejamento, proporcionando ao planejador uma visão mais completa do sistema elétrico.

Esta dissertação apresenta um sistema que faz uso do Sistema de Informação Geográfica combinado com a aplicação das técnicas de agrupamento e classificação K-Médias, Mapas Auto-Organizáveis e Fuzzy C-Médias, com o intuito de analisar a distribuição de carga da rede e de comparar o desempenho das técnicas utilizadas. Cada agrupamento realizado origina uma representação espacial (cenário). Este, juntamente com um índice de medida da performance do agrupamento (intra-grupo e inter-grupo) implementado neste trabalho, proporciona um ambiente favorável para a análise dos resultados.

Por meio dos estudos de caso realizados, utilizando-se os índices de medida intra-grupo e inter-grupo, constatou-se que o método Fuzzy C-Médias obteve o melhor resultado na classificação das curvas de treinamento. Apesar do método K-Médias possuir um tempo de processamento inferior, os valores apresentados não foram tão precisos como o método Fuzzy. Já os resultados do método Kohonen em anel (Mapas Auto-Organizáveis) não foram tão satisfatórios no tempo de processamento e nos índices de medida quanto os outros métodos.

O sistema desenvolvido possui ainda outras funcionalidades como histogramas de consumo, filtros e pesquisas espaciais, geração de cenários de carga, tudo sobre uma base de dados geográfica real de uma companhia elétrica, onde se podem visualizar espacialmente a rede de consumidores, transformadores, postes e demais componentes da rede de distribuição.

Palavras-Chave: Sistema de Informação Geográfica, Rede de Distribuição de Energia, Planejamento, Agrupamento, Classificação.

Abstract

The electrical distribution companies pass through an antagonistic reality - a growing increase of energetic demand, concerning to the continuous economic and population growth, and a financial resources restriction to invest in the expansion of their electrical network. If so, the activities improvement of energy distribution systems planning is indispensable to the efficient application of available resources.

The application of Geographic Information System associated to assortment and grouping methods fit in this context to potentialize the planning process, providing to the users a complete vision about the electrical system.

This paper presents a system that uses Geographic Information System associated to assortment and grouping methods K-means, Self-Organizing Maps and Fuzzy C-Means with the goal of analyzing the network load distribution and comparing these methods performance. Each accomplished grouping creates a spatial representation (scene). This scene amongst a grouping performance measurement index (intragroup and intergroup) implemented in this thesis, provide a favorable environment for outcome analysis.

Through the accomplished case studies using measurement index intragroup and intergroup, the Fuzzy C-Means method obtained the best outcome in training curve classification. In spite of having a lower processing time, the K-Means values were not precise enough as Fuzzy method. The Kohonen ring model (Self-Organizing Maps) was not so satisfying in processing time and measurement index requisites as the other methods.

This developed system possesses others functions like consumption histograms, filters and spatial researches, load scene generation; and all of this using the data base of a real electrical company, where the consumers network, transformers, electrical masts and others electrical distribution network components can be seen.

Keywords: Geographic Information System, Electrical Distribution Network, Planning, Grouping, Assortment.

Sumário

Lista de abreviaturas e siglas

Lista de símbolos

Lista de Figuras

Lista de Tabelas

1	Introdução	19
1.1	Justificativa	19
1.2	Objetivo	23
1.3	Estrutura do Trabalho	23
2	Planejamento da Distribuição de Energia Elétrica	25
2.1	Considerações Gerais	25
2.2	Critérios de Planejamento	26
2.3	Requisitos para o Planejamento da Expansão	27
2.4	Avaliação de Subestações	27
2.4.1	Previsão de carga inicial	27
2.4.1.1	Aquisição das informações de consumo	28
2.4.1.2	Estimativa de demanda por quadrícula	29
2.4.2	Ajuste da demanda	29
2.4.3	Avaliar demanda das áreas	29
2.4.4	Emissão e armazenamento dos dados	30

2.5	Planejamento da Expansão da Rede	30
2.5.1	Cenários	30
2.5.1.1	Previsão de carga inicial	30
2.5.1.2	Ajuste da demanda	31
2.5.2	Obras e alternativas	31
2.5.2.1	Avaliar desempenho da rede	31
2.5.2.2	Avaliação econômica das alternativas	32
2.5.2.3	Emissão de dados para relatório	32
2.5.2.4	Plano de obras	32
2.6	Requisitos de Caráter Geral	32
3	Técnicas de Agrupamento e Classificação	35
3.1	Agrupamento e Classificação	35
3.2	Redes Neurais Artificiais	36
3.2.1	Aprendizado	38
3.2.2	Redes auto-organizáveis	38
3.2.3	Redes SOM	39
3.2.3.1	Treinamento	40
3.2.3.2	Inicialização	42
3.2.3.3	Neurônio vencedor e a sua vizinhança	42
3.2.3.4	Atualização dos pesos	43
3.3	K-Médias	44
3.4	Lógica Nebulosa	46
3.5	Índice de Comparação de Métodos de Agrupamento e Classificação	51
4	Sistemas de Informações Geográficas	52
4.1	Definição	52

4.2	Estrutura geral	53
4.2.1	Tipos de dados em geoprocessamento	54
4.2.1.1	Mapas temáticos	55
4.2.1.2	Mapas cadastrais	55
4.2.1.3	Redes	57
4.2.1.4	Imagens	57
4.2.1.5	Modelos numéricos de terreno	58
4.3	Banco de Dados Geográfico	58
4.3.1	A arquitetura dual	59
4.3.2	A arquitetura em camadas	60
4.3.3	A arquitetura integrada	60
4.3.4	Modelo de dados espaciais	62
4.3.5	Consultas espaciais	62
5	Sistema Proposto	65
5.1	Descrição do Sistema	65
5.1.1	Ferramentas de navegação	65
5.1.2	Controle de camadas	66
5.1.3	Controle de escala	67
5.1.4	Controle de símbolos	68
5.1.5	Histograma de Consumo	70
5.2	Metodologia Geral proposta para o Agrupamento	70
5.2.1	Coleta de informações - etapa 1	71
5.2.2	Pré-carga de dados - etapa 2	71
5.2.3	Agrupamento e classificação - etapa 3	73
5.2.4	Visualização e análise dos resultados - etapa 4	76
5.2.4.1	Visualização das classes	77

5.2.4.2	Medidas de desempenho - índice intra-grupo (IAG)	82
5.2.4.3	Medidas de desempenho - índice inter-grupo (IEG)	82
5.2.4.4	Análise do crescimento horizontal	83
5.2.4.5	Análise de crescimento vertical	85
6	Estudo de Caso	87
6.1	Considerações Gerais	87
6.2	Metodologia de Análise dos Dados	87
6.3	Considerações sobre as Técnicas Empregadas	88
6.3.1	Mapa Auto-Organizável (SOM)	89
6.3.1.1	Conjunto de entrada	89
6.3.1.2	Normalização	90
6.3.1.3	Número de saídas	90
6.3.1.4	Inicialização dos pesos	90
6.3.1.5	Número de iterações	91
6.3.1.6	Taxa de aprendizagem e raio de vizinhança	91
6.3.2	K-Médias	91
6.3.3	Fuzzy C-Médias	92
6.4	Estudo de Caso 1 - Anápolis	92
6.4.1	Kohonen em anel	93
6.4.2	K-Médias	97
6.4.3	Fuzzy C-Médias	101
6.4.4	Distribuição das classificações	105
6.4.5	Medidas do agrupamento	106
6.5	Estudo de Caso 2 - Aparecida de Goiânia	107
6.5.1	Kohonen em anel	108
6.5.2	K-Médias	113

6.5.3	Fuzzy C-Médias	118
6.5.4	Distribuição das classificações	122
6.5.5	Medidas do agrupamento	124
6.6	Estudo de Caso 3 - Rio Verde	125
6.6.1	Considerações sobre o estudo de caso	127
6.6.2	Método K-Médias	127
6.6.3	Distribuição das classificações	132
6.6.4	Crescimento horizontal 2005/2006	133
6.6.5	Crescimento vertical 2005/2006	135
7	Conclusão	140
	Referências	142

Lista de abreviaturas e siglas

ANEEL	Agencia Nacional de Energia Elétrica
CAD	Computer Aided Design (desenho auxiliado por computador)
DEC	Duração Equivalente de Interrupção por Consumidor em horas e centésimos de horas
ESRI	”Environmental Systems Research Institute”
FEC	Frequência Equivalente de Interrupção por Consumidor em número de interrupções
GPS	Global Positioning System (Sistema de Posicionamento Global)
INMETRO	Instituto Nacional de Metrologia
IP	Iluminação Pública
RNA	Rede Neural Artificial
SDE	Spatial Database Engine
SGBD	Sistema de Gerenciamento de Banco de Dados
SIG	Sistema de Informação Geográfica
SOFM	Self-Organizing Feature Maps (Mapas de Características Auto-Organizáveis)
SOM	Self-Organizing Map (Mapa Auto-Organizável)
SQL	Structured Query Language (Linguagem de Consulta Estruturada)

Lista de símbolos

<i>kV</i>	Quilovolts
<i>kWh</i>	Quilowatt-hora
<i>kWh/Classe</i>	Quilowatt-hora por Classe

Lista de Figuras

3.1	Esquema de um neurônio artificial (Perceptron)	37
3.2	Organização em camadas (Multi-Layer Perceptron - MLP	37
3.3	Tipos de estímulos produzidos pela função chapéu mexicano	40
3.4	Formas de vizinhança de uma rede SOM	43
3.5	Ilustração do Algoritmo de K-Médias	46
3.6	Lógica de Boole, Lukasiewicz e a Lógica Nebulosa de Zadeh	47
4.1	Arquitetura de sistemas de informação geográfica	54
4.2	Representação vetorial e matricial de um mapa temático	55
4.3	Exemplo de mapa cadastral (países da América do Sul)	56
4.4	Exemplo de sobreposição de temas	56
4.5	Elementos de Rede	57
4.6	Imagem do Satélite IKONOS. Sydney, Austrália	58
4.7	Arquitetura dual	59
4.8	Arquitetura em Camadas	60
4.9	Arquitetura integrada	61
4.10	Arquitetura utilizada no projeto	61
4.11	Tipos espaciais primitivos	62
4.12	Pontos contidos em um polígono	63
4.13	Criação de um buffer ao redor de um rio e a identificação dos polígonos que tocam este buffer	64
5.1	Ferramentas de navegação do sistema GeoCom	65
5.2	Tela principal do sistema GeoCom	66

5.3	Exemplo de controle de escala	67
5.4	Exemplo de uma camada ativa fora de escala	68
5.5	Exemplo de controle de símbolos	69
5.6	Exemplo de controle de símbolos	69
5.7	Histograma de consumo	70
5.8	Exemplo de distribuição dos clientes nas quadrículas.	73
5.9	Diferença de clientes obtidos por transformador e por ponto de entrega	73
5.10	Configuração do treinamento pelo método Kohonen em Anel	74
5.11	configuração do treinamento pelo método K-Médias	75
5.12	Configuração do treinamento pelo método Fuzzy C-Médias	75
5.13	Tabela de saída do agrupamento pelo método Fuzzy C-Médias	76
5.14	Seleção dos critérios para visualização das classes	77
5.15	Visualização das classes geradas	77
5.16	Visualização das curvas de carga das quadrículas de mesma classe	78
5.17	Visualização de uma determinada curva de carga	79
5.18	Visualização de uma quadrícula selecionada na planilha.	79
5.19	Visualização da curva do centróide de um grupo	80
5.20	Visualização dos detalhes da quadrículas por Instalação	81
5.21	Visualização dos detalhes da quadrículas por Ponto de Entrega	81
5.22	Medida de desempenho IAG - exemplo do método Fuzzy C-Médias	82
5.23	Medida de desempenho IEG - exemplo do método Fuzzy C-Médias	83
5.24	Município de Marzagão - cenários 2005/2006	83
5.25	Tela de análise do crescimento horizontal	84
5.26	Crescimento Horizontal - Marzagão 2005/2006	84
5.27	Tela de análise do crescimento vertical	85
5.28	Crescimento Vertical - Marzagão 2005/2006	85

6.1	Legenda de cores da classificação	88
6.2	Gráficos dos centróides do treinamento pelo método Kohonen em Anel - Anápolis	95
6.3	Tela do resultado da classificação pelo método Kohonen em Anel	96
6.4	Gráficos dos centróides do treinamento pelo método K-Médias - Anápolis . . .	99
6.5	Tela do resultado da classificação pelo método K-Médias	100
6.6	Gráficos dos centróides do treinamento pelo método Fuzzy C-Médias - Anápolis	103
6.7	Tela do resultado da classificação pelo método Fuzzy C-Médias	104
6.8	Gráficos dos centróides do treinamento pelo método Kohonen em Anel - Apa- recida de Goiás	112
6.9	Tela do resultado da classificação pelo método Kohonen em Anel	113
6.10	Gráficos dos centróides do treinamento pelo método K-Médias - Aparecida de Goiânia	116
6.11	Tela do resultado da classificação pelo método K-Médias	117
6.12	Gráficos dos centróides do treinamento do método Fuzzy C-Médias - Aparecida de Goiânia	121
6.13	Tela do resultado da classificação pelo método Fuzzy C-Médias	122
6.14	Gráficos dos centróides do treinamento pelo método K-Médias - Rio Verde . .	129
6.15	Tela do resultado da classificação pelo método K-Médias em Rio Verde - 2005 .	130
6.16	<i>Zoom</i> do resultado da classificação pelo método K-Médias em Rio Verde - 2005	131
6.17	<i>Zoom</i> do resultado da classificação pelo método K-Médias em Rio Verde - 2006	132
6.18	Município de Rio Verde - 2005/2006	134
6.19	Crescimento Horizontal de Rio Verde - 2005/2006	135
6.20	Crescimento Vertical de Rio Verde - 2005/2006	136
6.21	Consulta da quadrícula Qd1 - Cenário 2005	137
6.22	Consulta da quadrícula Qd1 - Cenário 2006	138
6.23	Consulta da quadrícula Qd2 - Cenário 2005	138
6.24	Consulta da quadrícula Qd2 - Cenário 2006	139

Lista de Tabelas

6.1	Consumo e número de consumidores por classe - Anápolis 2006	92
6.2	Relação de matrículas por grupo - Anápolis 2006	105
6.3	Tempo de processamento dos métodos - Anápolis	106
6.4	Resultados dos Índices Intra-Grupo (IAG) - Anápolis	107
6.5	Resultados dos Índices Inter-Grupo (IEG) - Anápolis	107
6.6	Consumo e número de consumidores por classe - Aparecida de Goiânia 2006	108
6.7	Relação de matrículas por grupo - Aparecida de Goiânia	123
6.8	Tempo de processamento dos métodos - Aparecida de Goiânia	124
6.9	Resultados dos Índices Intra-Grupo (IAG) - Aparecida de Goiânia	125
6.10	Resultados dos Índices Inter-Grupo (IEG) - Aparecida de Goiânia	125
6.11	Consumo e número de consumidores por classe - Rio Verde 2005	126
6.12	Consumo e número de consumidores por classe - Rio Verde 2006	126
6.13	Relação de matrículas por grupo - Rio Verde	133

1 Introdução

1.1 Justificativa

Com o processo de desverticalização ocorrido no Setor Elétrico Brasileiro na última década, os agentes tradicionais (geração, transmissão, distribuição e comercialização) foram redefinidos, passando a ter estruturas individualizadas (separação entre transporte, produção e comercialização de energia elétrica).

A atividade de planejamento, que é essencial a qualquer sistema, torna-se imprescindível à distribuição de energia elétrica, de forma a atender o crescimento de carga em níveis de qualidade de serviços compatíveis com suas características, procurando aperfeiçoar a aplicação dos recursos disponíveis que são relativamente escassos.

Os sistemas de análise de mercado utilizados hoje normalmente fornecem ao planejador apenas um conjunto de dados organizados em relatórios, planilhas e gráficos, o que não proporciona um ambiente de análise espacial para crítica dos valores fornecidos pela base de dados da empresa.

Com o estabelecimento das funções dos sistemas de informação geográfica (SIG), no final da década de 80, e o seu crescente aumento quanto a integração com os planejadores, no início da década de 90, houve a popularização das técnicas de análise espacial. Isto proporcionou ao planejador inserir e integrar, em uma única base de dados, informações descritivas e geométricas dos diferentes tipos de dados geográficos, tornando a representação do mundo real um objetivo factível.

O processo de planejamento da distribuição de energia elétrica será potencializado utilizando-se o SIG integrado com o agrupamento e classificação das curvas de variação de consumo por classe dos clientes (curva de carga), por proporcionar um ambiente que fornece apoio às decisões sobre a expansão da rede de distribuição e identificação de áreas com concentração de carga, já que o cliente não é analisado isoladamente, mas em conjunto com os demais clientes da sua área de localização.

A análise das curvas de carga é parte das atividades de planejamento da distribuição de energia elétrica a longo-prazo. Vários trabalhos propõem o uso de diversas técnicas e métodos para a análise de curvas de carga para fins de planejamento de redes de distribuição, tendo, em sua maioria, a finalidade de efetuar previsão de carga, como descrito a seguir.

Camastra e Colla (CAMASTRA; COLLA, 1997) propõem um modelo neural para o problema de previsão a curto prazo da carga. O modelo neural é composto de duas redes Mapas Auto-organizáveis (*self-organizing map* - SOM), uma sobre a outra, as quais são estruturadas baseadas em um modelo a ser atingido. Resultados satisfatórios foram obtidos na predição da carga elétrica de uma hora adiante utilizando-se uma análise de seis meses anteriores. Por fim o resultado do treinamento é comparado a uma previsão de carga que utiliza um *Multi-Layer Perceptron* (MLP).

Chow, Zhu e Tram (CHOW; ZHU; TRAM, 1998) desenvolveram uma metodologia de análise do crescimento urbano e de previsão de carga espacial aplicando o método de decisão multi-objetivo fuzzy. Nesta metodologia, o crescimento estimado da carga é obtido pelo cruzamento de informações geográficas com das decisões de uso do solo. A área a ser analisada é dividida em quadrículas (10 x 10 m) e parâmetros como minimização do custo de desenvolvimento, distância das quadrículas das estradas, distância das quadrículas de pólos urbanos e custo de desenvolvimento de um uso do solo (residencial, comercial e industrial) para outro.

Lotufo e Missuni (LOTUFO; MINUSSI, 1999) propuseram uma revisão dos últimos trabalhos sobre previsão de carga, classificando-os de acordo com os modelos e métodos apresentados, como estatísticos, sistemas inteligentes (redes neurais) e lógica fuzzy. Foram analisados trabalhos de previsão de carga de longo e médio prazo, de planejamento da operação, de operação em tempo real, dentre outros. Eles avaliaram as vantagens de cada modelo, inclusive de modelos híbridos, ou seja, modelos compostos por mais de um método.

Miranda e Monteiro (MIRANDA; MONTEIRO, 1999) propõem uma metodologia que usa um modelo de inferência *fuzzy* sobre um suporte SIG para capturar o comportamento dos fatores de influência no crescimento da carga e conseqüentemente traçar o potencial para o desenvolvimento. O crescimento dinâmico da carga é simulado sobre autômatos celulares estendidos, nos quais o potencial de desenvolvimento e a demanda de uma área, em cada estágio, dirigem o sistema nos estágios seguintes de desenvolvimento. O produto final é uma série de cenários simulados (mapas de potenciais do crescimento da carga), os quais são usados em estudos do planejamento da expansão.

YU *et al.* (2000) desenvolveram através de modelos e métodos, uma ferramenta de suporte à decisão no planejamento da rede de distribuição urbana, denominada CNP. O CNP foi de-

envolvido em AutoCad e consiste em sete subsistemas: Suporte a Previsão de Carga (LFS), Suporte ao Planejamento de Subestações de Distribuição de Média Voltagem (MDSPS), Suporte ao Planejamento de Subestações de Distribuição de Alta Voltagem (HDSPS), Suporte ao Planejamento da Rede de Distribuição de Média Voltagem (MDNPS), Suporte ao Planejamento da Rede (sub-transmissão) de Distribuição de Alta Voltagem (HDNPS), Suporte ao Planejamento de Otimização de Fontes Reativas (RPS) e Suporte de Avaliação da Confiabilidade (RES).

Dorado e Cidrás (DORADO; MIGUEZ; CIDRÁS, 2001) propõem algoritmos, utilizando programação dinâmica, para planejamento da rede de distribuição rural de baixa voltagem.

Boulaxis e Papadopoulos (BOULAXIS; PAPAPOPOULOS, 2002) propõem, para o planejamento da distribuição, um algoritmo para a otimização do problema de roteamento de alimentadores utilizando técnicas de programação dinâmica e SIG. Todas as práticas, tais como parâmetros de custo (investimento, perda de linhas, confiabilidade), restrições técnicas (queda de tensão e limites térmicos) e restrições de roteamento físico (obstáculos, caminhos de alto custo, seções de linha existentes), são consideradas.

Alves (ALVES, 2002) apresenta técnicas que visam otimizar a localização de subestações de distribuição de energia elétrica. Foi proposto a resolução do problema de localização de subestações na Região Metropolitana de Curitiba e adjacências, usando métodos da Pesquisa Operacional, dentre os quais, destacam-se os Algoritmos Genéticos. Fatores como disponibilidade do terreno ideal, incluindo seu custo e outros, foram considerados, de forma a tornar certos resultados mais atrativos do que apenas aqueles com a "melhor" solução em termos matemáticos. A localização "ideal" é uma questão de grande importância no Planejamento da Distribuição, que, dentre outros, visa a economia de investimentos.

Franz Henry (ZAMORA, 2003) desenvolveu uma metodologia de previsão espacial de carga voltada ao planejamento de distribuição de energia elétrica, onde são usadas as técnicas de inteligência artificial, notadamente as Redes Neurais Artificiais (RNA) e Lógica Nebulosa (LN). Seu trabalho propõe sistematizar a previsão de carga, considerando o planejamento urbano e o uso do solo. O ambiente foi desenvolvido sobre a plataforma ArcView 8.3 da ESRI, melhorando o processo de planejamento devido ao suporte gráfico para a visualização dos cenários da previsão.

Maurício Sperandio (SPERANDIO, 2004) apresentou 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. Para isso, utilizou-se de uma validação cruzada entre um tipo de rede neural auto-organizável e o tradicional método estatístico das K-Médias.

Os grupos obtidos foram classificados de forma a sinalizar quando o nível de continuidade for inferior ao nível de referência, possibilitando detectar incompatibilidades com a meta, ou com o sistema elétrico de fornecimento.

Hwang e Kim (HWANG; KIM, 2004) desenvolveram um algoritmo para previsão de carga espacial a longo prazo sobre uma área de distribuição de energia. Foi construído um modelo de previsão de quatro níveis introduzindo uma estrutura espacial hierárquica: nação, região, área e subestação. O modelo de previsão é obtido pela união da estrutura espacial com o ano de horizonte da demanda a ser avaliado. Eles consideraram também o problema causado pela parada de uma subestação, aplicando um algoritmo de transferência de carga.

Brandão (BANDRÃO, 2004) implementou um programa matemático que utiliza-se de um modelo matemático convexo para a análise e otimização da rede de distribuição de energia elétrica. É apresentado neste trabalho o caminho que a energia elétrica percorre desde a sua geração até o final de seu trajeto, bem como um histórico da evolução da energia elétrica no Paraná e na cidade de Campo Mourão. Como produto final, é disponibilizado o resultado da otimização da rede para a data atual, bem como ao final dos próximos dez anos, tendo sido utilizado um estudo de previsão para a demanda futura durante este período. Este planejamento envolve a otimização da configuração radial e posicionamento das subestações de média e baixa voltagem, utilizando funções de custo. Neste trabalho são analisadas a queda de tensão, perda de carga na linha e cargas determinísticas considerando os tipos diferentes de clientes (mono, bi e trifásicos).

Assim, como foi mostrado, a grande maioria dos trabalhos propõem utilizar as curvas de carga para previsão de carga, não dando ênfase aos métodos e utilizações dos agrupamentos. Isso motivou o desenvolvimento deste trabalho, que utiliza os sistemas de informação geográfico e técnicas de agrupamento e classificação para disponibilizar informações com maior qualidade e dar suporte a tomada de decisão para o planejamento da expansão das redes de distribuição de energia elétrica de média tensão.

Este trabalho também objetiva comparar o desempenho das três técnicas de agrupamento e classificação utilizadas, Kohonen em anel, K-Médias e Fuzzy C-Médias, tendo como variáveis de entrada as curvas de carga dos clientes, curvas estas que trazem informações do consumo em *kWh*/classe relativas a área analisada.

Portanto, um ambiente que forneça, além dos dados dos clientes, informações sobre sua localização, curvas de carga, criação de cenários, histogramas de carga, agrupamento de classificação das curvas de carga e a análise do crescimento horizontal e vertical da rede de distribuição é parte integrante do processo de previsão de carga futura. Assim sendo, este trabalho contribui

para o desenvolvimento de uma ferramenta de planejamento mais abrangente, por proporcionar uma melhor análise da situação passada e atual da rede de distribuição de energia elétrica.

1.2 Objetivo

Este trabalho tem como objetivo criar um programa, denominado GeoCom, que utilizando das funções do Sistema de Informação Geográfica (SIG) e das técnicas de classificação Kohonen em anel, K-Médias e Fuzzy C-Médias, cria um ambiente favorável à análise comercial de empresas de distribuição de energia elétrica que possuam um levantamento georeferenciado de sua rede de distribuição.

Será realizada também uma comparação entre os métodos de agrupamento e classificação utilizados com o intuito de se identificar o método que obtém o melhor resultado sobre os estudos de caso realizados.

O programa possibilitará ao planejador identificar áreas de concentração de carga bem como o perfil dos consumidores presentes nestas áreas. Através do uso dos métodos de classificação estas áreas serão classificadas quanto a sua curva de carga e cada área receberá uma cor característica do seu grupo. Isso possibilitará a comparação das áreas levando o planejador a conclusões mais seguras com a utilização do SIG.

1.3 Estrutura do Trabalho

A seguir é apresentada a estrutura da dissertação.

No Capítulo 2 são apresentadas as informações básicas necessárias para o desenvolvimento do planejamento de distribuição de energia. Ele inclui ainda sugestões de requisitos, funcionalidades e procedimentos com vistas ao desenvolvimento de um ambiente de planejamento de distribuição de energia elétrica.

No Capítulo 3 são apresentados os conceitos das técnicas utilizadas para o agrupamento e classificação das curvas de carga.

No Capítulo 4 são descritos os fundamentos teóricos dos Sistemas de informação Geográfica e dos Bancos de dados Espaciais para um melhor entendimento do sistema proposto por este trabalho.

No Capítulo 5 é apresentado em detalhes o sistema proposto denominado GeoCom, bem como a metodologia adotada para o agrupamento e classificação das curvas de carga.

O capítulo 6 contém os estudos de caso realizados e a análise dos dados obtidos. Essa análise é feita tanto através da visualização gráfica, quanto pelo cálculo de alguns índices, visando a comparação das técnicas utilizadas.

No Capítulo 7 são apresentadas as conclusões do trabalho e sugestões de trabalhos futuros.

2 Planejamento da Distribuição de Energia Elétrica

2.1 Considerações Gerais

Este capítulo apresenta as informações básicas necessárias para o desenvolvimento do planejamento de distribuição de energia. Ele inclui sugestões de requisitos, funcionalidades e procedimentos, com vistas ao desenvolvimento de um ambiente que permita criar planos de médio e longo prazo, avaliar oportunidades de implantação de projetos e de melhorar a utilização dos sistemas de distribuição (LIBANIO; LIMA, 2004).

O planejamento dos sistemas de distribuição de energia elétrica é um processo de estudo e análise onde as concessionárias de energia buscam atender suas obrigações de entrega de energia de forma confiável e economicamente viável.

Como a energia passou a ser considerada um produto, o planejamento torna-se imprescindível para a saúde técnica e econômica da concessionária, onde o objetivo é minimizar os custos de entrega além de prestar um bom serviço a seus consumidores. Através da evolução dos computadores houve uma evolução nos modelos de planejamento de sistemas de distribuição, permitindo uma forte interação entre os engenheiros de planejamento e os modelos de estudo e a análise de redes elétricas.

Para o setor de planejamento, podem-se destacar os seguintes objetivos básicos (GUEMBAROVSKI, 1999):

- Elaborar os estudos de previsão de cargas e o diagnóstico a nível de alimentador, subestação e região elétrica, a curto e a médio prazo, do sistema de distribuição;
- Elaborar estudos específicos para a consolidação de critérios para o planejamento de distribuição;
- Pesquisar, desenvolver e manter os sistemas computacionais para o planejamento de dis-

tribuição;

- Manter e atualizar o banco de dados do planejamento da distribuição;
- Elaborar os estudos de proteção, regulação de tensão e compensação de reativos;
- Programar a aquisição dos equipamentos e acessórios para atender o programa de obras.

Há diferentes categorias de planejamento da expansão sendo que uma empresa de distribuição deve conciliar diferentes categorias em sua área de planejamento:

- Visão de longo prazo: São estabelecidas as tendências de evolução da rede bem como os montantes de investimento para assegurar determinados níveis de qualidade de serviço.
- Visão de médio prazo: As regiões elétricas são estudadas pela empresa, a qual determina um plano de obras mais detalhado em função do crescimento espacial da carga.
- Visão de curto prazo: Os estudos são mais específicos, aplicados à instalação de dispositivos para suporte reativo, melhoria da qualidade de serviço, sem a necessidade de obras de vulto, como é o caso de manobras para minimização de perdas, melhoria dos níveis de tensão e instalação de dispositivos de proteção e seccionamento.

2.2 Critérios de Planejamento

Conforme estabelecido pela ANEEL (ANEEL, 2006), as empresas distribuidoras devem definir critérios na formulação das alternativas técnicas para a expansão dos sistemas de distribuição, sendo estes:

- Critérios de segurança da distribuição: Adoção de critérios e normas de engenharia para os projetos das obras programadas, incluindo aquelas referentes à restauração do sistema existente, que assegurem as necessárias condições de segurança dos seus sistemas de distribuição, de forma a evitar e minimizar acidentes com pessoal (interno e o público externo) e instalações próprias e de terceiros.
- Carregamento das redes de distribuição: Definição de faixas ideais de carregamento para os diversos componentes do sistema de distribuição com base em aspectos de economia, operação, segurança e qualidade desejada para o sistema elétrico, além da própria dinâmica do crescimento da carga.

- Faixas de tensão admissíveis para operação normal: Atendimento às faixas de tensão admissíveis e regulamentadas pela ANEEL.
- Faixas de tensão admissíveis para operação na ocorrência de contingências: Fixação das faixas de tensão admissíveis quando da ocorrência de contingências.
- Qualidade do fornecimento: Fixação de critérios mínimos de qualidade de fornecimento de energia elétrica, que, em última instância, traduz-se na qualidade da tensão fornecida.
- Confiabilidade: Fixação de índices de confiabilidade regulamentados pela ANEEL, que devem ser verificados através da projeção anual para cada configuração do sistema de distribuição ao longo do período de planejamento.
- Requisitos de planejamento estabelecidos em outros dispositivos legais: Utilização de critérios de planejamento que sejam estabelecidos por outros dispositivos legais (padronização de equipamentos pelo INMETRO, por exemplo).

2.3 Requisitos para o Planejamento da Expansão

Primeiramente, deve ser definida a periodicidade dos estudos de previsão do mercado de energia elétrica, sendo que a ANEEL sugere que seja anual. Estes estudos resultam em previsões de consumo e demanda referentes a área de abrangência, a qual pode ser toda a área de concessão, um ou mais municípios, uma ou mais subestações/alimentadores ou ainda uma região específica da rede.

Para o planejamento da expansão da distribuição, toda a rede de distribuição existente, com todos os seus equipamentos instalados dentro da área escolhida deverão ser utilizados.

2.4 Avaliação de Subestações

Avaliar a melhor localização, quando na ampliação ou definição de novas subestações, avaliando as áreas de atendimento das subestações existentes é parte do planejamento da expansão da distribuição.

2.4.1 Previsão de carga inicial

Para a área de abrangência escolhida obtém-se a carga da mesma, a qual, para previsão do consumo e da demanda de energia elétrica, deve subdividir as unidades consumidoras nos

seguintes segmentos: residencial, industrial, comercial e serviços (inclui outras atividades), rural, poder público, iluminação pública, serviços públicos e consumo próprio.

- Consumo residencial: A previsão do consumo para o mercado residencial mostra-se importante, pois serve como indicador da evolução de outras classes como: comercial, serviços e poderes públicos devido a sua correlação com as mesmas. Além disso, esta classe possui uma relativa participação na estrutura do mercado global.
- Consumo industrial: A previsão de consumo para esta classe deve considerar dois segmentos:
 - Consumidores industriais especiais: constituído pelas unidades consumidoras dos subgrupos tarifários A2 (ligados em tensão de fornecimento de 88 a 138 kV), A3 (ligados em 69 kV), A3a (ligados em tensão de 30 a 44 kV) e A4 (ligados em tensão de 2,3 a 25 kV).
 - Consumidores industriais tradicionais: constituído pelas unidades consumidoras industriais do subgrupo tarifário B (fornecimento em baixa tensão).
- Consumo comercial, serviços e outras atividades: Caracteriza-se por unidades consumidoras que possuem suas atividades voltadas para a prestação de serviços em geral ou voltadas ao comércio. Esta classe depende, para sua projeção, do grau de urbanização dos municípios, da estrutura do setor terciário, do regime de trabalho, dentre outros fatores.
- Consumo rural: Formada pelas subclasses: Agropecuária, Cooperativa de Eletrificação Rural, Indústria Rural, Coletividade Rural, Serviço Público de Irrigação Rural e Escola Agrotécnica. Esta classe utiliza da tendência histórica, agregada a planos e programas de irrigação para compor a previsão do consumo quando do consumo associado à irrigação.
- Consumo do poder público, iluminação pública, serviço público e consumo próprio: Constituído das seguintes unidades: administrativas do poder público (federal, estadual ou distrital e municipal) não enquadradas em outras categorias, iluminação pública, serviços de transporte (tração elétrica urbana ou ferroviária), água, esgoto e saneamento e consumo próprio (próprio, canteiro de obras e interno).

2.4.1.1 Aquisição das informações de consumo

Os dados de consumo e classificação dos clientes devem ser obtidos do sistema comercial da empresa, sendo estes relativos aos últimos quatro anos, conforme resolução da ANEEL (ANEEL,

2006). Estes dados são vinculados ao consumo mensal de energia dos clientes, onde, para os consumidores do grupo A, deve também ser verificada a demanda.

2.4.1.2 Estimativa de demanda por quadrícula

O consumo de energia por classe de consumidores é armazenado em quadrículas (dimensões de 500 x 500 m). Para cada quadrícula, calcula-se a curva de demanda estimada e o desvio padrão, atual e futuro (horizonte de até 10 anos), a partir da energia e das curvas de carga. A obtenção da taxa de crescimento se dará a partir da média anual dos últimos quatro anos (AN-EEL, 2006).

2.4.2 Ajuste da demanda

O ajuste da demanda é um processo originado da comparação entre os dados disponibilizados pelo setor de mercado e os dados levantados por quadrícula no planejamento. A partir desta comparação, há a definição de um fator de ajuste a ser aplicado na demanda e desvio padrão, de cada quadrícula, obtendo as cargas ajustadas.

O processo de ajuste da demanda deve seguir a definição da área de abrangência, possibilitando assim, estimar a curva de demanda e de desvio padrão por município, subestação/alimentadores, região específica ou por toda a área de concessão.

Quando da opção por subestação, é importante proporcionar a previsão de demanda de áreas não contíguas atendidas pela mesma subestação.

O processo de carga das quadrículas deve ser dinâmico, possibilitando a alteração das cargas das quadrículas, baseado nas informações da área de planejamento como, por exemplo, a entrada de novas cargas ou alteração áreas que estão sendo loteadas.

2.4.3 Avaliar demanda das áreas

O processo de planejamento das subestações é cíclico, podendo-se estimar para cada subestação a área de atendimento de acordo com a potência instalada e verificar ainda a necessidade de ampliação ou criação de subestações para os anos futuros. Para essas avaliações faz-se necessário os seguintes passos:

- Divisão em áreas e alternativas: Dada a área a ser estudada deve-se dividi-la levando em consideração as subestações que a atenderão, sendo que esta área pode ser reduzida ou

aumentada no decorrer do estudo. Ano a ano serão definidas as áreas de atendimento das subestações existentes e futuras, baseando-se no cálculo da curva de demanda e o desvio padrão de cada subestação. Diferentes alternativas poderão ser alcançadas dada a solução escolhida, sendo estas alternativas armazenadas e apresentadas em relatórios.

- Escolha do ponto ideal de subestação: As cargas são armazenadas em quadrículas (dimensões de 500 x 500 m). Assim, o ponto central de gravidade das cargas de um grupo de quadrículas será o baricentro da figura formada por esse grupo.

2.4.4 Emissão e armazenamento dos dados

Para as alternativas propostas por cada estudo, deve ser elaborado um relatório final, onde devem ser apresentadas tabelas contendo as estimativas de demanda (curvas de demanda e desvio padrão) e mapas das áreas de atendimento. Estes mapas apresentarão quadrículas coloridas de acordo com a demanda, áreas de influência de cada subestação (existentes e futuras), municípios e regionais, alguns elementos da urbanização, bem como a própria rede de distribuição. Estes mapas e tabelas que compõem o relatório final deverão ser armazenados e disponibilizados para outros estudos.

2.5 Planejamento da Expansão da Rede

O processo de avaliação das redes de distribuição, identificando necessidades de reforços, ampliação ou construção de novos alimentadores é foco do planejamento da distribuição, objetivando assegurar a qualidade do fornecimento de energia aos clientes dentro de critérios pré-estabelecidos.

2.5.1 Cenários

Os planejamentos efetuados são gravados para possibilitar uma comparação futura, criando-se cenários para cada planejamento efetuado.

2.5.1.1 Previsão de carga inicial

A previsão de carga no planejamento da rede se faz igualmente ao processo de avaliação de subestação. Entretanto, inclui-se a opção de utilização de uma previsão realizada anteriormente, já que o conceito de cenários é utilizado.

2.5.1.2 Ajuste da demanda

O ajuste da demanda inicia-se com a obtenção das demandas ao longo da rede, onde com base na curva de carga dos consumidores obtém-se a carga total da instalação através do fluxo de carga, considerando a curva de carga e desvio padrão. Posteriormente, utilizando o fluxo de carga, obtém-se a curva do alimentador. Conseqüentemente a curva da subestação será a soma das curvas obtidas de cada um de seus alimentadores.

O processo de ajuste da demanda segue o processo aplicado na avaliação de subestações, devendo porém, ser prevista a inserção de cargas pontuais na rede. Cada previsão de carga deverá ser armazenada em cenários e disponibilizada para o uso em outras previsões.

2.5.2 Obras e alternativas

2.5.2.1 Avaliar desempenho da rede

Anualmente, deve-se verificar as condições da rede sob estudo, avaliando problemas de carregamento da rede e transformadores, perdas e regulação de tensão, para cada alimentador.

- **Análise da rede existente:** O fluxo de carga é utilizado nesta avaliação e a execução do fluxo deve ser disponibilizada nos horários de interesse do estudo. Deverão ser identificados os locais onde ocorreram ultrapassagens dos valores permitidos (adequado, tolerável e crítico). A ANEEL fixa limites máximo e mínimo de tensão no ponto de entrega, sendo assim importante a regulação de tensão.
- **Definição de obras:** Os problemas identificados pela análise da rede deverão ser avaliados objetivando a definição das obras necessárias para solucioná-los, através de readequação, inclusão de novos elementos e reconfiguração da rede. Portanto as obras aplicadas podem ser tanto uma alteração física da rede, como recondutoramento, transferência de carga entre alimentadores, instalação de equipamentos como reguladores de tensão, banco de capacitores, dentro outros. Assim, para cada obra definida, um novo processamento se faz necessário de forma a comprovar a eficácia da solução.

As obras são definidas individualmente, possuindo uma identificação e uma descrição detalhada, podendo ser utilizadas na simulação de qualquer ano futuro e em qualquer alternativa prevista.

- **Análise de contingência de atendimento e atendimento à novas cargas:** Poderão ser simuladas e avaliadas as contingências e atendimentos a novas cargas pontuais como se

fossem uma obra específica ou uma alternativa a partir de uma alternativa existente. A princípio, as contingências a serem simuladas deverão ser perda parcial ou total de trecho de rede do alimentador e perda da subestação, devendo sua carga ser totalmente assumida por alimentadores de outra subestação. Os estudos de casos de liberação de carga serão atendidos pela inserção de cargas pontuais. Deverão ser adotados limites menos exigentes quando do atendimento de contingência.

- Criação das alternativas: A composição das obras darão origem as alternativas, onde cada obra relacionada poderá compor uma ou mais alternativas de solução para os problemas identificados.

2.5.2.2 Avaliação econômica das alternativas

Deverão ser custeadas as alternativas validadas tecnicamente com o objetivo de identificar a alternativa mais econômica.

2.5.2.3 Emissão de dados para relatório

Para cada estudo realizado deve ser elaborado um relatório final contendo o mapa da área de atendimento e as tabelas com as estimativas de demanda (curvas de demanda e desvio padrão). Os mapas gerados deverão, além de identificar cada alimentador de forma diferenciada, identificar a urbanização e as soluções inseridas, ano a ano.

2.5.2.4 Plano de obras

As melhores alternativas definidas nos estudos deverão compor o Plano de Obras da empresa.

2.6 Requisitos de Caráter Geral

Para a obtenção de um ambiente de análise mais robusto, seguem mais algumas sugestões de caráter geral.

- Telas amigáveis e manual do planejador: Com o intuito de facilitar o trabalho do planejador, as telas do sistema de planejamento devem ser de fácil manuseio e dispostas em

uma seqüência natural da realização dos estudos. Também é importante o desenvolvimento de um manual do sistema de forma a proporcionar um melhor entendimento das funcionalidades do sistema e sua seqüência de uso.

- Cálculos de fluxo para rede radial e rede em anel: Conforme a configuração da rede deve ser utilizada uma metodologia de cálculo específica.
- Sazonalidade: O sistema deve permitir a avaliação das sazonalidades das cargas ao longo dos meses do ano, principalmente as relacionadas com a irrigação. Para isso, ao se criar uma alternativa ou obra, deve-se possibilitar o acréscimo ou decréscimo na carga de uma determinada área, instalação transformadora, alimentador ou subestação.
- Relatórios de saída com informação das unidades: É necessário que fique bem explícito as unidades dos dados nos relatórios de saída, bem como na entrada dos dados no sistema.
- Redes monofásicas: Devem ser consideradas as cargas trifásicas e monofásicas para o cálculo do fluxo de carga na rede primária e secundária, sendo as informações das fases previamente modeladas e cadastradas no sistema.
- Otimização do tempo de processamento: A duração dos processos deve ser analisada levando-se em conta a freqüência de execução dos processos. Portanto o cálculo do fluxo de carga de um grupo de alimentadores seria executado em tempo real pelo planejador, enquanto que para uma previsão de carga aceitam-se tempos de processamento mais longos, já que são realizadas com pouca freqüência e seus resultados disponibilizados para vários planejadores.
- Escolha do horário do processamento: Para a obtenção de curvas, deve-se realizar o processamento de todos os 96 horários disponíveis. Já para o cálculo de fluxo de carga, por exemplo, deve-se permitir a escolha dos horários de processamento, bem como se utilizar no processamento a demanda média mais ou menos o desvio padrão.
- Aproveitamento de serviços: Uma determinada obra definida deve ser disponibilizada a outros planejadores em outras alternativas e até mesmo em outros cenários. O mesmo deve ocorrer com uma alternativa definida, podendo esta ser utilizada por outro planejador no mesmo cenário ou em cenários diferentes. Para tal é preciso que fique sempre explícito em que consiste cada obra e quais as obras presentes em uma alternativa identificando em qual ano foram inseridas.
- Demanda dos clientes faturados no grupo A: Para este tipo de cliente, a demanda lida deve ser utilizada somente para averiguação de sua curva de carga, mas caso não se tenha

a curva de carga a demanda pode ser utilizada conforme critério definido previamente.

- **Perdas:** O sistema deve permitir a obtenção das perdas na rede primária, a medida que estão sendo criadas as alternativas, já que estas são importantes na avaliação dos benefícios a serem auferidos com a criação de uma nova subestação, alimentador, capacitor, recondutoramento, dentre outros procedimentos.
- **Limite econômico de condutores:** Deverá ser considerado o limite econômico de condutores, além de seu limite térmico.
- **DEC e FEC:** O sistema deve permitir estimar as melhorias nos indicadores de DEC e FEC dos alimentadores quando da realização de obras.
- **Iluminação pública:** Para a iluminação pública, deverá ser desenvolvida uma modelagem diferenciada, tendo como curva de carga dois patamares em relação a potência instalada, ligada e desligada. As cargas com comando centralizado devem ser consideradas como tal, inclusive as atendidas através de rede diretamente lançada ao solo.

3 Técnicas de Agrupamento e Classificação

3.1 Agrupamento e Classificação

Para a análise de carga espacial da rede de distribuição, faz-se primordial realizar o agrupamento de suas áreas. Para esta tarefa pode ser empregada a clusterização (formação de agrupamentos), processo que pelo qual procuram-se classificar objetos em categorias.

A clusterização tem por finalidade reunir os padrões em vários grupos, sendo que os dados a serem analisados terão um grau de similaridade alto com o grupo de sua classificação e tão diferentes quanto possível, dos elementos dos outros grupos.

Há diferentes abordagens para o tratamento do problema de clusterização, dentre elas, a abordagem convencional, onde cada elemento a ser classificado pertencerá totalmente a uma determinada categoria, e a abordagem nebulosa, onde um elemento pode ser classificado em várias categorias, possuindo diferentes graus de associação a cada uma delas.

Na clusterização são identificadas as classes de cada um dos elementos através de suas descrições, como um vetor de pares (atributo, valor do atributo). Baseando-se em certas suposições e/ou critérios, os algoritmos de clusterização tentam particionar o conjunto de objetos de entrada, o que conseqüentemente pode produzir ou não saídas significativas e úteis destes dados de entrada.

Neste trabalho foram utilizadas a abordagem convencional, através das técnicas de agrupamento e classificação K-Médias (KMD) e Mapas Auto-Organizáveis (SOM), e a abordagem nebulosa, mais flexível, através do método Fuzzy C-Médias.

3.2 Redes Neurais Artificiais

O final da década de 80 marcou o ressurgimento da área de Redes Neurais Artificiais (RNAs), também conhecida como redes conexionistas ou sistema de processamento paralelo e distribuído (CARVALHO; BRAGA; LUDERMIR, 1998). Esta área é caracterizada por sistemas que relembram a estrutura do cérebro humano, onde, por não ser baseada em regras ou programas, se constitui uma alternativa à computação algorítmica convencional.

As RNAs são sistemas paralelos distribuídos compostos por unidades de processamento simples (nós) que computam determinadas funções matemáticas, normalmente não lineares. Estas unidades são dispostas em uma ou mais camadas, interligadas por um grande número de conexões, geralmente unidirecionais. Na maioria dos modelos estas conexões estão associadas a pesos, os quais armazenam o conhecimento representado no modelo e servem para ponderar a entrada recebida por cada neurônio da rede.

A operação de uma unidade de processamento, proposta por McCulloch e Pitts em 1943 pode ser resumida da seguinte maneira (ver Figura 3.1):

- Sinais são apresentados à entrada do neurônio (x_i);
- Cada sinal é multiplicado por um número, ou peso, que indica a sua influência na saída da unidade (w_{ij});
- É feita a soma ponderada de todos os sinais de entrada que produz um nível de atividade (z);
- Se este nível de atividade exceder um certo limite, limiar (l) do neurônio, para ativação do neurônio ($Y_i = g(z_i)$), ele se ativará, caso contrário ele ficará inativo.

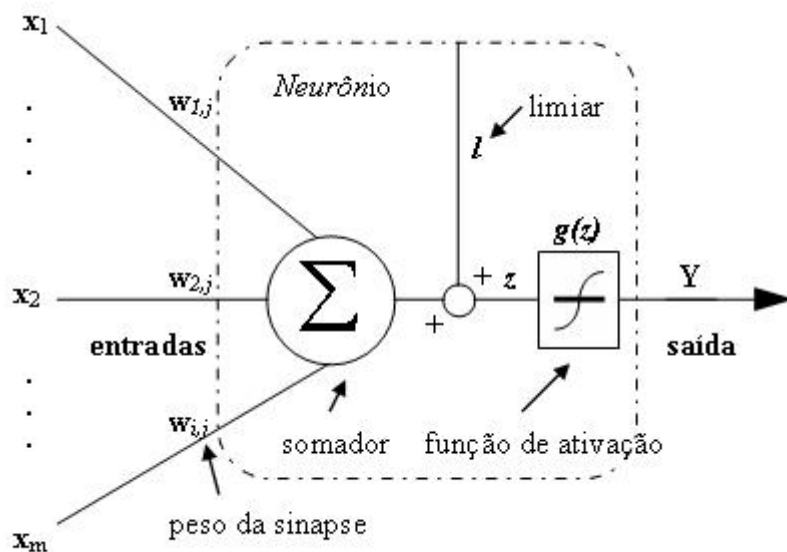


Figura 3.1: Esquema de um neurônio artificial (Perceptron) (ZANON; LIMA, 2003)

As RNAs são tipicamente organizadas em camadas, sendo que em redes multi-camadas o processamento realizado por cada nó é definido pela combinação dos processamentos realizados pelos nós da camada anterior que estão conectados a ele (ver Figura 3.2).

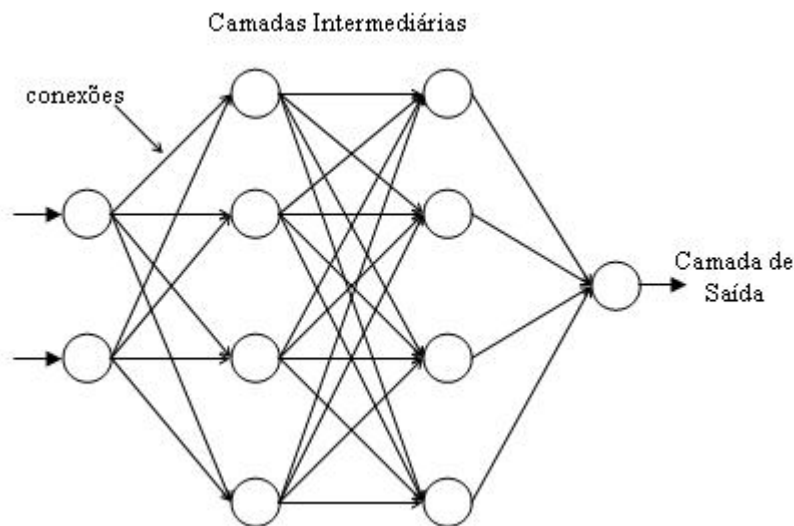


Figura 3.2: Organização em camadas (Multi-Layer Perceptron - MLP) (ZANON; LIMA, 2003)

O procedimento usual na solução de problemas em RNAs passa inicialmente por uma fase de aprendizagem, onde um conjunto de exemplos é apresentado para a rede, a qual extrai automaticamente as características necessárias para representar a informação fornecida, sendo utilizadas posteriormente para gerar respostas para o problema. A capacidade de aprender através

de exemplos e de generalizar a informação aprendida são, sem dúvida, os atrativos principais da solução de problemas através de RNAs (CARVALHO; BRAGA; LUDERMIR, 1998).

3.2.1 Aprendizado

Aprendizagem é o processo pelo qual os parâmetros de uma rede neural são ajustados através de uma forma continuada de estímulo pelo ambiente no qual a rede está operando, sendo o tipo específico de aprendizagem realizada definido pela maneira particular como ocorrem os ajustes realizados nos parâmetros (CARVALHO; BRAGA; LUDERMIR, 1998).

Os métodos de treinamento podem ser agrupados em dois paradigmas principais:

- **Aprendizado supervisionado:** Com o objetivo de ajustar os parâmetros da rede, encontrando uma ligação entre os pares de entrada e saída fornecidos, é introduzida a ação de um supervisor. Este supervisor avalia o comportamento da rede visando direcionar o processo de treinamento.
 - **Aprendizado por reforço:** caso particular de aprendizado supervisionado, onde o aprendizado é obtido através de um processo de triagem e erro desenvolvido para maximizar o índice de desempenho escalar chamado sinal de reforço.
- **Aprendizado não-supervisionado:** Não há a existência de um supervisor, sendo o conjunto de treinamento composto somente por padrões de entrada. Assim este aprendizado apenas se torna possível quando há redundância nos dados de entrada.
 - **Aprendizado por Competição:** caso particular de aprendizado não supervisionado, onde dado um padrão de entrada há uma competição entre os nós de saída para serem ativados. Neste treinamento o nó de saída mais forte fica ainda mais forte e seu efeito inibidor sobre as outras unidades de saída torna-se dominante.

3.2.2 Redes auto-organizáveis

As redes auto-Organizáveis são uma classe de redes neurais artificiais (RNAs), as quais possuem capacidade de aprender através de exemplos, sem a necessidade de ser explicitamente programada.

Esta classe de RNA é treinada sem a necessidade de um supervisor externo ou punição/recompensa, sendo um conjunto de padrões de entrada a única informação fornecida para o treinamento. Desta forma, estas redes definem seus parâmetros por si próprias, sem auxílio externo.

Há uma amplo leque de aplicações potenciais para estas redes, sendo a principal aplicação em problemas de reconhecimento de padrões, onde os padrões que compartilham características comuns devem ser agrupados, com cada grupo de padrões representando uma e apenas uma classe. Para realizar este agrupamento é necessário que a rede encontre características ou padrões significativos nos dados de entrada, sendo isto apenas possível se houver redundância nos dados de entrada.

Existem vários modelos de redes neurais artificiais auto-organizáveis, cada um com suas peculiaridades e propósitos. Estes modelos podem ser divididos em dois grandes grupos, baseando-se no algoritmo de treinamento:

- **Aprendizado competitivo:** os nós da rede competem entre si pelo direito de atualizar seus pesos. São mais utilizados em problemas de classificação, de extração de características, compressão de dados e formação de agrupamentos.
- **Aprendizado Hebbiano:** baseado no princípio para a atualização dos pesos proposto pelo neurofisiologista Hebb em seu livro *the Organization of Behaviour*. São utilizados em extração de características, análise de dados e memória associativa.

Neste trabalho foi escolhido a utilização dos mapas auto-organizáveis de Kohonen , também conhecidos como redes de Kohonen ou redes SOM, as quais utilizam aprendizado competitivo e assumem uma estrutura topológica entre seus nós.

3.2.3 Redes SOM

As redes mapas auto-organizáveis, SOM, também chamadas de mapas de características auto-organizáveis, SOFM (Self-Organizing Feature Maps), foram desenvolvidas por Teuvo Kohonen na década de 80.

Nestas redes, os nós estão espacialmente ordenados dentro de áreas, onde os nós topologicamente próximos tendem a responder a padrões ou estímulos semelhantes. Essa ordenação topológica é resultado do uso de feedback lateral entre os nós, geralmente modelado por uma função conhecida como "chapéu mexicano".

Na função chapéu mexicano, cada nó influencia o estado de ativação de seus nós vizinhos de três formas possíveis (ver Figura 3.3):

- **Excitatória:** se os vizinhos estão próximos a ele (por exemplo, dentro de uma área R1)

- Inibitória: se os vizinhos esta fora da área anterior, mas dentro de uma segunda área (por exemplo, dentro de uma área R2, mas fora da área R1, onde $R2 > R1$)
- Levemente Excitatória: se os vizinhos estão fora das áreas anteriores, mas dentro de uma terceira área (por exemplo, dentro de uma área R3, mas fora dos raios R1 e R2, onde $R3 > R2 > R1$).

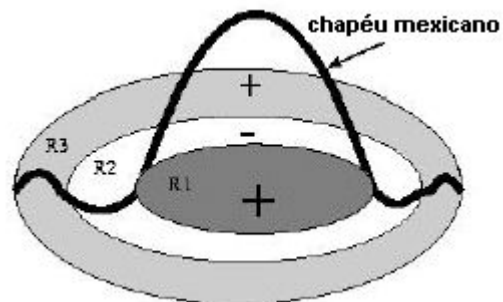


Figura 3.3: Tipos de estímulos produzidos pela função chapéu mexicano (CARVALHO; BRAGA; LUDERMIR, 1998)

Os Mapas Auto-Organizáveis funcionam basicamente da seguinte forma: quando um padrão de entrada ρ é apresentado, a rede procura a unidade mais parecida com ρ . Durante o seu treinamento, a rede aumenta a semelhança do neurônio escolhido e de seus vizinhos ao padrão ρ . Desta forma, a rede constrói um mapa topológico onde neurônios que estão topologicamente próximos respondem de forma semelhante ao padrão de entrada semelhante, criando assim uma região denominada bolha de atividade em torno do neurônio de maior ativação (VASCONCELOS, 2000).

A rede SOM é uma rede biologicamente plausível, ou seja, apresenta respaldo biológico. Assim como no córtex cerebral, neurônios ativos apresentam conexões mais fortes com neurônios que estão fisicamente mais próximos (CARVALHO; BRAGA; LUDERMIR, 1998).

Segundo este conceito, quando um nó vence uma competição, produz a maior saída para uma dada entrada, não apenas ele, mas também os nós localizados na sua vizinhança têm seus pesos ajustados.

3.2.3.1 Treinamento

O mapeamento dos padrões de entrada nos neurônios da rede SOM é realizado durante a fase de treinamento, a qual de acordo com (HAYKIN, 2001) possui três etapas básicas:

1. **Competição:** Para cada padrão de entrada, os neurônios do mapa calculam seus respectivos valores de uma função discriminante. Esta função fornece a base para a competição entre os neurônios. O neurônio com o maior valor da função discriminante é declarado vencedor da competição;
2. **Cooperação:** O neurônio vencedor determina a localização espacial de uma vizinhança topológica de neurônios excitados que cooperarão entre si;
3. **Adaptação sináptica:** Os neurônios excitados aumentam seus valores individuais da função discriminante em relação ao padrão de entrada através de ajustes adequados aplicados a seus pesos sinápticos. Os ajustes feitos são tais que a resposta do neurônio vencedor à aplicação subsequente de um padrão de entrada similar é melhorada.

Através de um processo cíclico de comparação de padrões de entrada o mapa de Kohonen é auto-organizado, onde os vetores de pesos são armazenados em cada neurônio. O treinamento desse modelo, mostrado abaixo, baseia-se simplesmente na procura do neurônio cujos pesos são mais próximos de um determinado padrão de entrada (com a menor distância Euclidiana) e no aumento da similaridade entre eles (padrão de entrada e pesos do neurônio vencedor) (BATISTA, 2004).

1. Inicialize o mapa μ

Sejam $w_{ji}(t)$ ($0 \leq j \leq n - 1$) o peso da entrada j ao neurônio i no tempo t e k o número de épocas de treinamento. Inicialize os vetores de pesos das n entradas com valores aleatórios. Inicialize o raio da vizinhança em torno de cada neurônio i com um valor alto.

2. Apresente o padrão de entrada X ao mapa μ

$$X = x_0(t), x_1(t), \dots, x_{n-1}(t)$$

3. Calcule a distância d_i entre a entrada X e cada neurônio i do mapa μ

$$d_i = \sum_{i=1}^n \|x_i - w_{ji}\|$$

4. Selecione a menor distância

O neurônio vencedor i^* é aquele com menor d_i .

5. Atualize os pesos para o neurônio i e para sua vizinhança $h_{ji}(t)$

$$w_{ji}(t + 1) = w_{ji}(t) + \alpha(t)h_{ji}(t)(x_j - w_{ji}(t))$$

onde $\alpha(t)$ é a taxa de aprendizagem ($0 < \alpha(t) < 1$), $w_{ji}(t)$ é o vetor peso do neurônio j no tempo t , $w_{ji}(t + 1)$ é o vetor de peso atualizado no tempo $t + 1$ e $h_{ji}(t)$ é a função de vizinhança no tempo t .

6. Continue a partir do passo 2

Até atingir o número de épocas de treinamento k .

O treinamento da rede SOM pode, resumidamente, ser dividido em apenas duas fases:

- Fase de ordenação: ocorre a ordenação topológica dos vetores de pesos, com pesos iniciais aleatórios. Durante esta fase a taxa de aprendizado, o raio de vizinhança e os pesos sofrem grande mudança. Busca-se agrupar os nós do mapa topológico em conjuntos, de modo a refletir a distribuição dos padrões de entrada.
- Fase de convergência: ajuste fino do mapa, onde é utilizada uma taxa de aprendizado baixa e o raio da vizinhança envolve um ou nenhum vizinho. Busca aprimorar o agrupamento realizado na fase de ordenação.

Após o treinamento, a rede SOM reúne os padrões de entrada em agrupamentos, sendo possível rotular os nós de saída para indicar as classes que representam.

3.2.3.2 Inicialização

A inicialização é o primeiro passo no treinamento da rede SOM. Os pesos iniciais de uma rede do tipo SOM são definidos aleatoriamente. Entretanto, isso pode levar a rede a não convergir ou apresentar ciclos muito lentos, já que algum peso pode receber um valor que está longe dos padrões de entrada. Diversas alternativas foram propostas para resolver ou minimizar este problema (CARVALHO; BRAGA; LUDERMIR, 1998):

- Utilização de vetores de pesos iniciais iguais;
- Utilização de um limiar para cada nó (consciência);
- Modificação dinâmica de raio de vizinhança.
- Utilização dos vetores de entrada para atribuir os pesos iniciais.

3.2.3.3 Neurônio vencedor e a sua vizinhança

O estado de ativação de um nó é determinado pela distância entre seu peso e o vetor de saída. A função de ativação mais usada na rede SOM é a baseada na medida de distância Euclidiana (equação (3.2)).

$$y_j = \sum_{i=1}^n \|x_i - w_{ji}\| \quad (3.1)$$

Assim, para identificar o nó que mais se aproxima do vetor de entrada x pode-se aplicar o índice $i(x)$ utilizando a condição,

$$i(x) = \arg \min \|x - w_j\|, j \in [1, k] \quad (3.2)$$

Sendo w_j é o vetor de pesos.

Após identificar o neurônio vencedor i , é necessário identificar os seus vizinhos. A vizinhança define a influência do nó vencedor, determinando quantos nós em torno do vencedor terão seus pesos ajustados. Na Figura 3.4 são representadas diferentes topologias, onde estão representados os polígonos que definem os limite do conjunto de vizinhos em volta do neurônio vencedor. A alternativa de modificação dinâmica do raio de vizinhança é utilizada por Kohonen e reduz a vizinhança dos nós vencedores durante o treinamento.

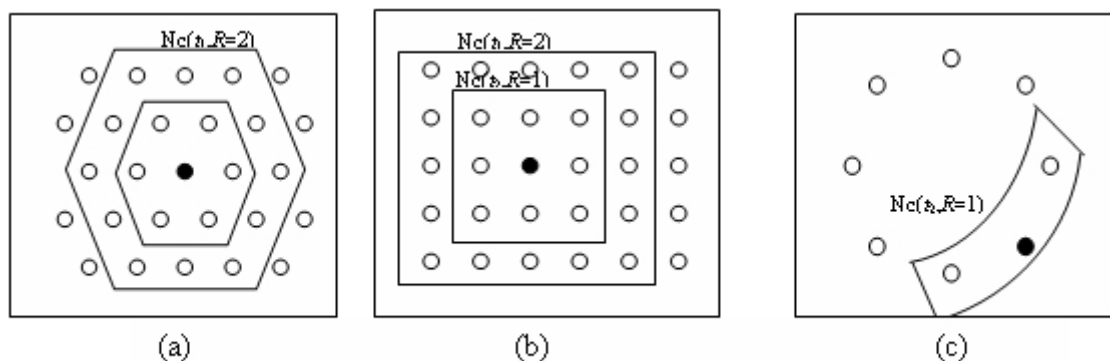


Figura 3.4: Formas de vizinhança de uma rede SOM (ZANON; LIMA, 2003)

Neste trabalho foi utilizada a topologia em anel (c), a qual demanda menos tempo de processamento e apresenta um resultado satisfatório para as curvas de carga utilizadas (ZANON; LIMA, 2003).

3.2.3.4 Atualização dos pesos

A equação (3.3) (CARVALHO; BRAGA; LUDERMIR, 1998) ilustra como os pesos da unidade vencedora e daquelas situadas na sua vizinhança são atualizados:

$$w_{ji}(t+1) = w_{ji}(t) + \alpha(t)h_{ji}(t)(x_j - w_{ji}(t)) \quad (3.3)$$

onde,

$\alpha(t)$ é a taxa de aprendizagem ($0 < \alpha(t) < 1$), $w_{ji}(t)$ é o vetor peso do neurônio j no tempo t , $w_{ji}(t+1)$ é o vetor de peso atualizado no tempo $t+1$ e $h_{ji}(t)$ é a Função de vizinhança no tempo t .

Esta atualização dos pesos tende a aproximar o vetor peso w do vetor de entrada X , já que no treinamento há a apresentação repetida dos dados de entrada.

O parâmetro da taxa de aprendizagem (equação 3.4) deve começar em um valor inicial α_0 e então decrescer gradualmente com o aumento do tempo t . Não é importante uma forma precisa da função do tempo, podendo ser exponencial, linear ou inversamente proporcional a t .

$$\alpha(t) = \alpha_0 \exp\left(\frac{-t}{\tau_2}\right), t = 0, 1, 2, \dots \quad (3.4)$$

Onde τ_2 é uma constante de tempo do algoritmo SOM.

Apesar de não serem ótimas, as fórmulas de decaimento exponencial apresentadas para as funções de vizinhança e da taxa de aprendizagem são normalmente adequadas para a formação da rede SOM de uma maneira auto-organizada.

3.3 K-Médias

O K-Médias é um tipo de algoritmo de agrupamento, que pode ser utilizado para a classificação não-supervisionada. Nesta técnica os grupos baseiam-se em sua proximidade em relação a um centróide, o qual é obtido das médias das características que descrevem os padrões de entrada.

O algoritmo K-Médias é um processo iterativo que tenta minimizar o erro quadrático médio atribuindo cada padrão de entrada ao grupo de centróide mais próximo.

Abaixo são apresentados os passos envolvidos no algoritmo K-Médias:

1. Defini-se o número de *clusters* (agrupamentos).
2. Inicializa-se os *clusters* determinando os k centróides iniciais dos grupos através de:
 - Tomando-se simplesmente as k primeiras entradas de dados ou;

- Um conjunto de centróides em posições arbitrárias coincidentes com amostras dos dados.
3. Iteração nas amostras; para cada amostra:
 - Procura-se o centróide mais próximo.
 - Atribui-se a amostra ao *cluster* correspondente.
 - Recalcula-se o centróide para esse *cluster*.
 4. Volta-se ao passo 3 até um critério de convergência ser cumprido.

Como critério de convergência pode ser empregado um número máximo de iterações ou executar o algoritmo até que os centros não se movam mais, ou apresentam uma mudança muito pequena. No último caso, deve ser definido um erro mínimo no início da execução.

Dentre as técnicas utilizadas neste trabalho esta é a que possui o algoritmo mais simples, porém possui duas grandes desvantagens. Primeiro, ele requerer o valor de k como parâmetro, sendo que normalmente esta informação é desconhecida. A outra desvantagem é que ele é muito sensível ao particionamento inicial, já que este é fruto de uma escolha aleatória de padrões de entrada para representar inicialmente cada grupo.

Este algoritmo realiza uma busca de um ponto de máximo, o qual é atingido quando não há mais mudança na escolha do centróide para cada entrada, dado um critério de parada a partir da solução inicial. Não existem garantias de que o algoritmo encontre o máximo global, sendo possível que o algoritmo apresente diferentes soluções quando executado mais de uma vez.

A Figura 3.5 ilustra os passos descritos no algoritmo K-Médias.

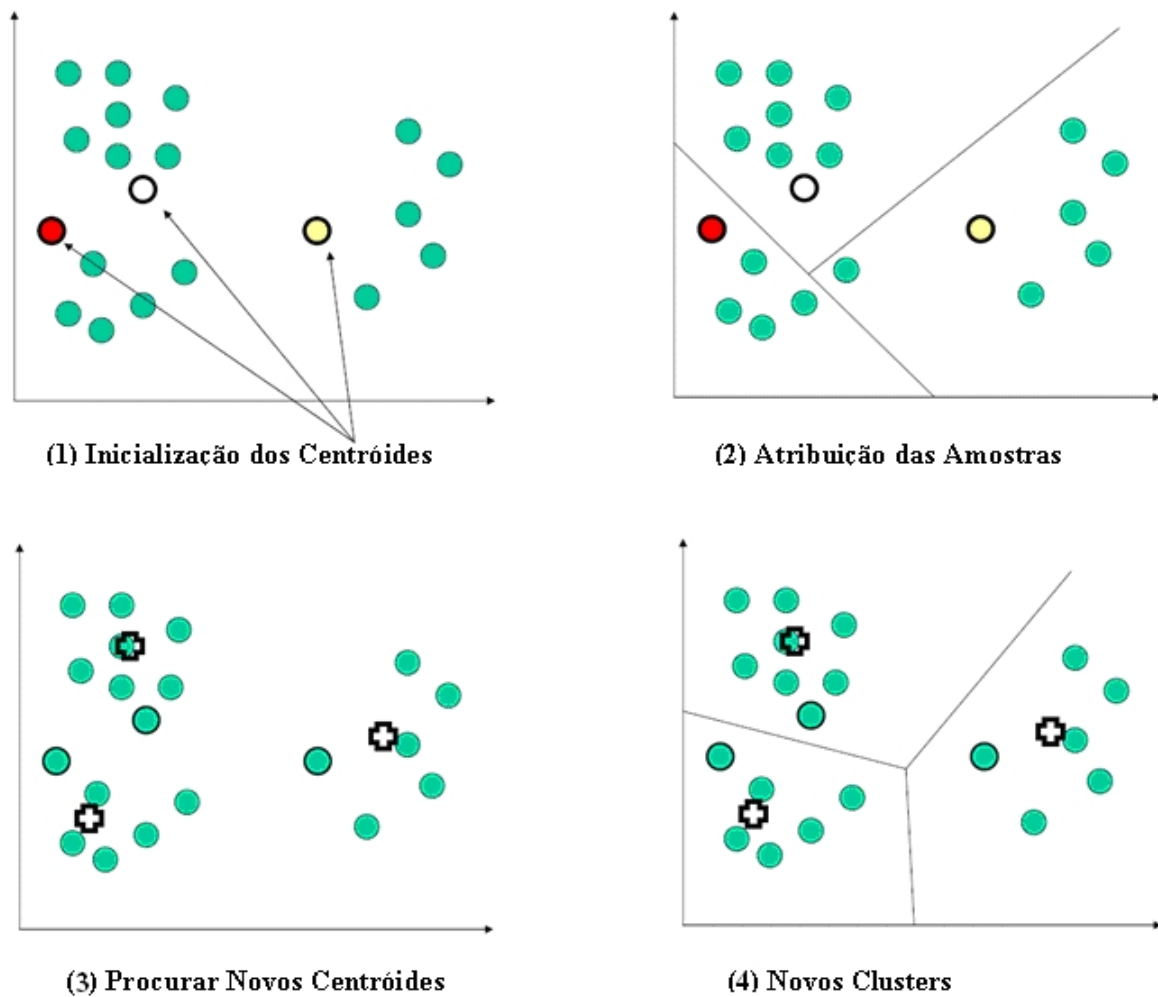


Figura 3.5: Ilustração do Algoritmo de K-Médias

Originalmente este algoritmo possui uma restrição onde cada grupo deve ter pelo menos um padrão de entrada atribuído a ele. Com o objetivo de possibilitar a definição do número de grupos durante o treinamento, como ocorre na rede SOM, o algoritmo foi alterado, podendo haver grupos que não representem nenhum padrão.

3.4 Lógica Nebulosa

A Lógica Nebulosa (*Fuzzy Logic*), surgiu dos estudos desenvolvidos por Lotfi A. Zadeh nos anos setenta (ZADEH, 2006), o qual criou os sistemas nebulosos (*fuzzy logic systems*), fundamentando-se basicamente na representação e manipulação da imprecisão e da incerteza em informações (Figura 3.6).

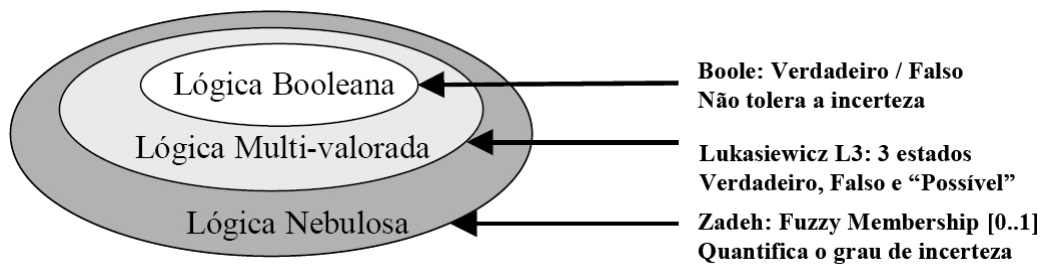


Figura 3.6: Lógica de Boole, Lukasiewicz e a Lógica Nebulosa de Zadeh (BITTENCOURT; OSÓRIO, 2002)

A lógica nebulosa possibilita que seja abordado de forma mais adequada a representação e manipulação de conhecimentos na área de Inteligência Artificial (IA), onde expressões que usualmente não podem ser tratadas pelos sistemas da lógica clássica, tais como "quase", "muito" e "pouco" são atendidas nos sistemas nebulosos.

A teoria nebulosa é uma extensão da teoria clássica, onde funções bivalentes da primeira são substituídas por funções de pertinência nebulosa; Esta função de pertinência indica o quanto um elemento pertence a um dado conjunto.

Na teoria clássica dos conjuntos, o conceito de pertinência de um elemento a um conjunto é bivalorada. Os elementos de um subconjunto A em um determinado conjunto X simplesmente pertencem ou não pertencem àquele conjunto. Isto pode ser expresso pela função característica f_A (ROSS, 1995):

$$f_A(x) = \begin{cases} 1 & \text{se e somente se } x \in A \\ 0 & \text{se e somente se } x \notin A \end{cases}$$

Propondo uma caracterização mais ampla, a teoria de conjuntos nebulosos generaliza a função característica de modo que ela possa assumir um número infinito de valores diferentes no intervalo $[0,1]$. Assim, um conjunto *fuzzy* A em X é um conjunto de pares ordenados,

$$A = \{\mu_A(x)/x\} \quad x \in X \quad (3.5)$$

Onde $\mu_A(x)$ é a função de pertinência de x em A e é definida como o mapeamento de X no intervalo fechado $[0,1]$:

$$\mu_A(x) : X \longrightarrow [0, 1] \quad (3.6)$$

Assim, as técnicas de agrupamento *fuzzy* conseguem levar em consideração situações genéricas onde os elementos próximos ao contorno dos grupos podem assumir, ao mesmo tempo, as características de diversos grupos distintos, pertencendo em parte a cada um deles.

No agrupamento *fuzzy* um elemento pode pertencer a vários grupos ou classes, com diferentes valores de pertinência, já no agrupamento com a abordagem convencional, cada elemento pertence totalmente a um único grupo, como os dois modelos de agrupamento citados anteriormente (ROSS, 1995).

O algoritmo de agrupamento de dados nebuloso Fuzzy C-Médias baseia-se em um modelo não linear de otimização que agrupa dados com características similares de acordo com um procedimento iterativo de minimização de uma função objetivo que representa um critério de partição, ponderado pelos graus de pertinência dos dados aos respectivos grupos (BEZDEK, 1981).

Neste método, dado um conjunto de dados $X = x_1, x_2, \dots, x_n$ onde x_k , em geral, é um vetor de características $x_k = [x_{k1}, x_{k2}, \dots, x_{kp}] \in R^p$ para todo $k \in 1, 2, \dots, n$ sendo R^p o espaço p -dimensional, o problema de agrupamento *fuzzy* é encontrar uma pseudopartição *fuzzy* que representa a estrutura dos dados da melhor forma possível. Uma pseudopartição *fuzzy* de X é uma família de c subconjuntos *fuzzy* de X , denotada por $P = A_1, A_2, \dots, A_c$ que satisfaz as equações 3.7 e 3.8.

$$\sum_{i=1}^c A_i(x_k) = 1 \quad (3.7)$$

para todo $k \in 1, 2, \dots, n$ sendo que n representa o número de elementos do conjunto X . Ou seja, a soma dos graus de pertinência de um elemento em todas as famílias deve ser igual a um.

$$0 < \sum_{k=1}^n A_i(x_k) < n \quad (3.8)$$

para todo $i \in 1, 2, \dots, c$ sendo que c representa o número de classes. Ou seja, a soma dos graus de pertinência de todos os elementos de uma família deve ser menor que o número de elementos existentes no conjunto universo X .

O método Fuzzy C-Médias pode ser equacionado por meio de um algoritmo iterativo, baseado na minimização de um índice de desempenho, que indica a adequabilidade da pseudopartição gerada. O desempenho do algoritmo é influenciado pela escolha do número de classes c , dos centros de *cluster* iniciais, da ordem na qual os vetores são processados, da medida de

distância, do critério de parada e pelas propriedades geométricas dos dados (YONAMINE et al., 2002).

Portanto, os parâmetros de entrada do algoritmo são: a instanciação inicial dos graus de pertinência dos objetos a cada uma das classes e dos centros de *cluster* dessas classes; o número de *clusters* c ; o índice de fuzzificação, que é a medida de distância $m \in (1, \infty)$, a qual define a distância permitida entre os pontos e os centros de *cluster*; e o critério de parada $\varepsilon > 0$.

A seguir, são descritos os passos do algoritmo (YONAMINE et al., 2002):

1. Considere $t = 0$ representando a iteração 0. Defina $P^{(0)}$ como uma pseudopartição $P = A_1, A_2, \dots, A_c$, atribuindo os graus de pertinência iniciais dos elementos às classes dessa pseudopartição, os quais podem ser informados ou calculados de forma aleatória.
2. No caso de $t = 0$, forneça os c centros de *cluster* $V_1^{(t)}, \dots, V_c^{(t)}$ iniciais. Para $t > 0$, calcule os c centros de *cluster* $V_1^{(t)}, \dots, V_c^{(t)}$ por (3.9) para $P^{(t)}$ e o valor de m fornecido.

$$V_i = \frac{\sum_{k=1}^n [A_i(x_k)]^m x_k}{\sum_{k=1}^n [A_i(x_k)]^m} \quad (3.9)$$

A variável real $m > 1$ é chamada de índice de fuzzificação e é usada para definir a distância permitida entre os pontos e o centro que está sendo calculado. Quanto maior o valor de m , mais elementos do conjunto são considerados como pertencentes a uma pseudopartição. Esse parâmetro é escolhido de acordo com o problema considerado. Não existe base teórica para uma escolha ótima do valor de m .

O vetor V_i calculado por (3.9), visto como o centro do *cluster* A_i , é a média ponderada dos dados em A_i . O peso do dado x_k é a m -ésima potência do seu grau de pertinência ao conjunto fuzzy A_i .

3. Atualize $P^{(t)}$ para $P^{(t+1)}$ usando o seguinte procedimento: para cada $x_k \in X$ e para todo $i \in 1, 2, \dots, c$, se $\|x_k - V_i^{(t)}\|^2 > 0$, calcule o grau de pertinência do elemento x_k à classe A_i , pela fórmula (4), sendo as classes A_i selecionadas em uma ordem pré-definida ou aleatoriamente.

$$A_i^{(t+1)} = \left[\sum_{j=1}^c \left(\frac{\|x_k - V_i^{(t)}\|^2}{\|x_k - V_j^{(t)}\|^2} \right)^{1/m-1} \right]^{-1} \quad (3.10)$$

Assume-se que $\|\Delta\|$ é alguma norma indicada para o produto interno no espaço R^p e $\|x_k - V_i^{(t)}\|^2$ representa a distância entre x_k e V_i .

Por meio da fórmula (3.10), a distância de cada elemento x_k ao centro de V_i da pseudopartição atual A_i é comparada à distância do mesmo elemento ao centro de *cluster* V_j de todas as demais pseudopartições A_j , de modo a atribuir um grau de pertinência ao elemento x_k no *cluster* A_i que seja proporcional ao seu grau de pertinência aos demais *clusters* A_j .

Quando $\|x_k - V_i^{(t)}\|^2 = 0$ para algum $i \in I \subseteq 1, 2, \dots, c$, o que indica que x_k corresponde ao centro de *cluster* V_i , defina $A_i^{(t+1)}(x_k)$ como um número real não negativo que satisfaz (3.11) e defina $A_i^{(t+1)}(x_k) = 0$ para $i \in 1, 2, \dots, c - I$.

$$\sum_{i \in I} A_i^{(t+1)}(x_k) = 1 \quad (3.11)$$

Em casos como este, quando o elemento coincide com o centro de *cluster* de uma classe A_i , o ideal seria que o algoritmo atribuísse a tal elemento o grau de pertinência máximo (3.7) àquela classe e o grau de pertinência mínimo (0) às demais classes. No entanto, o teste realizado pela fórmula (3.11) se torna necessário, uma vez que esse elemento pode ter sido associado a outras classes com um grau de pertinência maior que zero, se A_i não for a primeira classe selecionada pelo algoritmo. Nota-se, portanto, a influência da ordem na qual as pseudopartições são escolhidas.

4. Como critério de parada, compare $P^{(t)}$ e $P^{(t+1)}$. Se $|P^{(t)} - P^{(t+1)}| \leq \varepsilon$, então pare; caso contrário, faça $t = t + 1$ e retorne ao Passo 2.

Neste teste, $|P^{(t)} - P^{(t+1)}|$ denota a distância entre $P^{(t+1)}$ e $P^{(t)}$, dada pela fórmula (3.12):

$$|P^{(t)} - P^{(t+1)}| = \max_{i \in N_c, k \in N_n} |A_i^{(t+1)}(x_k) - A_i^{(t)}(x_k)| \quad (3.12)$$

Essa fórmula verifica se a máxima diferença entre o grau de pertinência de qualquer elemento x_k em uma dada classe A_i na iteração atual e na iteração anterior é menor que o erro definido ε , sendo isso feito para todas as classes. Em caso positivo, o algoritmo deve parar e retornar os graus de pertinências atuais de todos os elementos em todas as classes como resultado. Note que, quanto menor o valor de ε , maior o número de passos e, conseqüentemente, mais refinada é a pseudopartição final obtida.

3.5 Índice de Comparação de Métodos de Agrupamento e Classificação

Para comparação dos métodos utilizados neste trabalho foram utilizados dois índices para fazer uma comparação dos resultados do agrupamento e da classificação (ZANON; LIMA, 2003):

1. Intra-grupo (IAG);
2. Inter-grupo (IEG).

O índice Intra-grupo (equação 3.13) é uma função de custo baseada no erro quadrático médio e indica o quanto os padrões de entrada que estão em um mesmo grupo são similares.

$$IAG = \sum_{i=1}^D \sum_{x \in D_i} \|x - m_i\|^2 \quad (3.13)$$

Onde x é o número de amostras de D_i e m_i é a média dessas amostras. Assim, para um dado *cluster* D_i , o vetor m_i é a melhor representante das amostras em D_i de forma a minimizar a soma dos comprimentos quadrados de cada x em relação a m_i .

Assim o IAG mede o erro quadrático total em representar as amostras x_1, x_2, \dots, x_n por D agrupamentos com o centro em m_1, m_2, \dots, m_n . Quanto menor for IAG, melhor é o agrupamento ou classificação dos padrões, ou seja, os padrões que estão em um mesmo grupo possuem uma semelhança maior.

O índice inter-grupo (IEG) determina a relação entre os grupos, ou seja, o quão distantes eles estão. Seu cálculo é feito da seguinte forma:

$$IEG = \sum_{j=1}^D \|X_j - X\|^2 \quad (3.14)$$

Onde X_j é a curva do centróide do grupo j e X é a média global das curvas de carga do conjunto de treinamento.

Diferente do IAG este índice não permite afirmar que quanto menor o valor, melhor é o resultado, mas apenas que se ele tiver um valor alto os grupos estão mais separados ou se ele tiver um valor baixo significa que há uma semelhança maior entre os padrões de grupos diferentes.

4 *Sistemas de Informações Geográficas*

Esta capítulo apresenta os fundamentos teóricos dos SIGs para um melhor entendimento do sistema proposto por este trabalho.

4.1 Definição

O termo Sistema de Informação Geográfica (SIG) é aplicado para sistemas que realizam o tratamento computacional de dados geográficos. Estes sistemas possuem ampla gama de aplicações, incluindo temas como cartografia, agricultura, floresta, redes de concessionárias (telefonia, água e energia) e cadastro urbano (FREDERICK, 1999), podendo-se destacar três grandes maneiras de se utilizar um SIG:

- Como suporte para análise espacial de fenômenos;
- Como ferramenta para produção de mapas;
- Como um banco de dados geográfico, possibilitando o armazenamento, recuperação e manutenção das informações espaciais.

Como os SIGs são uma tecnologia relativamente recente que obteve um crescimento muito rápido, sua definição não é uma tarefa fácil, já que grupos heterogêneos formulam conceitos diferentes sobre SIG. Entretanto, seguem abaixo algumas definições sobre SIG (CASANOVA et al., 2005).

- Um caso especial de sistemas de informações, no qual o banco de dados consiste em informações sobre características distribuídas espacialmente, atividades ou eventos, os quais são definidos no espaço como pontos, linhas ou áreas. Os SIGs manipulam os dados acerca destes pontos, linhas e áreas para estabelecer perguntas *ad hoc* e análises.

- Um sistema que contém dados espacialmente referenciados que podem ser analisados e convertidos em informações para uso em um conjunto específico de finalidades. A característica principal de um SIG é analisar dados para gerar novas informações.
- um SIG é qualquer sistema de gerenciamento de informações capaz de: coletar, armazenar e recuperar informações baseadas nas suas localizações espaciais; identificar locais dentro de um ambiente que tenha sido selecionado a partir de determinados critérios; explorar relações entre os dados de um certo ambiente; analisar os dados espaciais para subsidiar os critérios de formulação de decisões; facilitar a exportação de modelos analíticos capazes de avaliar alternativas de impactos no meio ambiente; exibir e selecionar áreas, tanto graficamente como numericamente, antes e depois das análises.

Com base nas definições apresentadas é possível indicar as principais características dos SIGs (CASANOVA et al., 2005):

- Inserir e integrar, numa única base de dados, informações espaciais provenientes de meio físico-biótico, de dados censitários, de cadastros urbano e rural, e outras fontes de dados como imagens de satélite e GPS.
- Oferecer mecanismos para combinar as várias informações, através de algoritmos de manipulação e análise, bem como para consultar, recuperar e visualizar o conteúdo de base de dados geográfica.

4.2 Estrutura geral

Os SIGs possuem uma grande interação com várias tecnologias. Assim, sua aplicabilidade é vasta, pois retratam o mundo real de uma forma dinâmica e consistente, sendo que para cada sistema são implementados mecanismo distintos de manipulação dos dados, baseados nos objetivos e necessidades específicos de cada projeto.

Os sistemas que utilizam-se do SIG, numa visão abrangente, devem ter presentes os seguintes componentes:

- Interface com o usuário;
- Entrada e integração de dados;
- Funções de processamento gráfico e de imagens;

- Visualização e plotagem;
- Armazenamento e recuperação de dados (banco de dados geográfico)

A Figura 4.1 representa o relacionamento dos componentes citados anteriormente. Os sistemas desenvolvidos implementam de forma distinta estes componentes, em função de seus objetivos e necessidades. Entretanto, todos os componentes apresentados estão presentes num SIG. A utilização dos SIGs facilita a integração de dados coletados de fontes heterogêneas, de forma transparente ao usuário final (INPE, 1996).

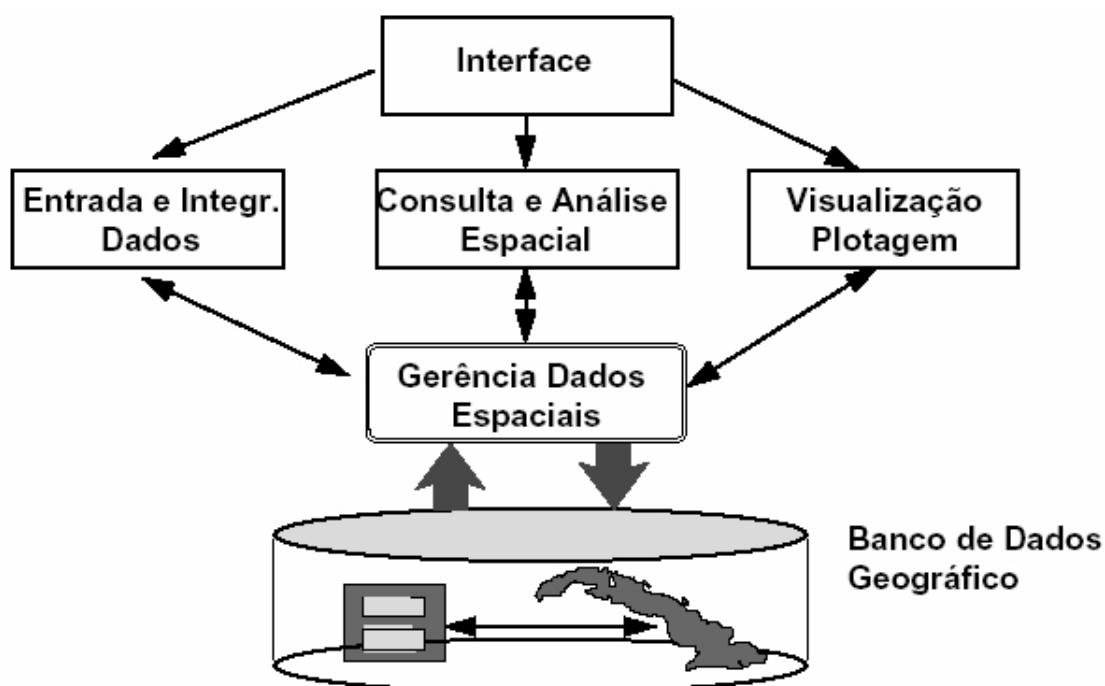


Figura 4.1: Arquitetura de sistemas de informação geográfica (INPE, 1996)

Assim, desenvolver um SIG significa desenvolver um sistema mais amplo, pois tratamos com uma variedade de estruturas de dados e algoritmos capazes de representar as concepções do mundo real.

4.2.1 Tipos de dados em geoprocessamento

Um sistema de geoprocessamento possui dados que estão georeferenciados, ou seja, localizados na superfície terrestre numa projeção cartográfica, armazenando a geometria e os atributos destes dados.

Uma característica básica e geral num SIG é a sua capacidade de tratar as relações espaciais entre objetos geográficos, diferenciando-o de sistemas de Projeto Auxiliado por Computador

(CAD) (TSUTOMU, 1997). O conceito de topologia foi criado para denotar a estrutura de relacionamentos espaciais (vizinhança, proximidade e pertinência) que podem ser estabelecidos entre objetos geográficos.

4.2.1.1 Mapas temáticos

Os mapas temáticos possuem dados obtidos a partir de levantamento de campo, os quais dão origem a um ou mais polígonos formando assim regiões geográficas. São exemplos, o uso do solo e a aptidão agrícola de uma região.

Na topologia do tipo arco-nó-região, os arcos que se conectam entre si através de nós (pontos inicial e final) e os arcos que circundam uma área definem um polígono (região), o qual é utilizado para permitir uma representação e análise mais aprimorada do espaço geográfico.

Um mapa temático também pode ser armazenado no formato matricial (*raster*), sendo a área do mapa dividida em células de tamanho fixo, as quais armazenam um valor correspondente ao tema mais freqüente naquela localização espacial.

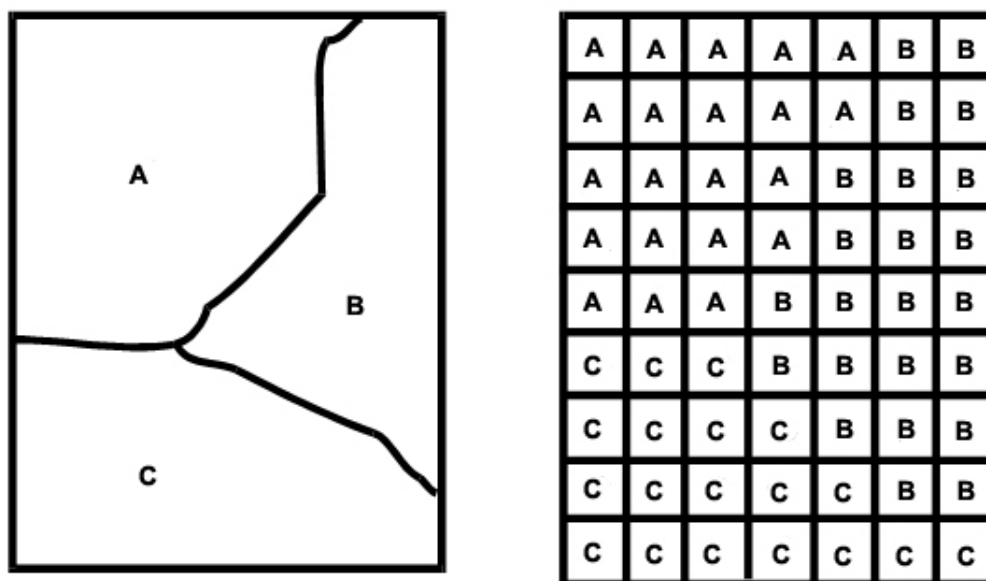


Figura 4.2: Representação vetorial e matricial de um mapa temático (SPRING, 2006)

4.2.1.2 Mapas cadastrais

Em um mapa cadastral cada um de seus elementos é um objeto geográfico. Estes objetos possuem atributos, topologia associada e podem ter representações gráficas diferentes em mapas

de escala distintas. Sua parte gráfica é armazenada em forma de coordenadas vetoriais, não sendo usual sua representação na forma matricial.

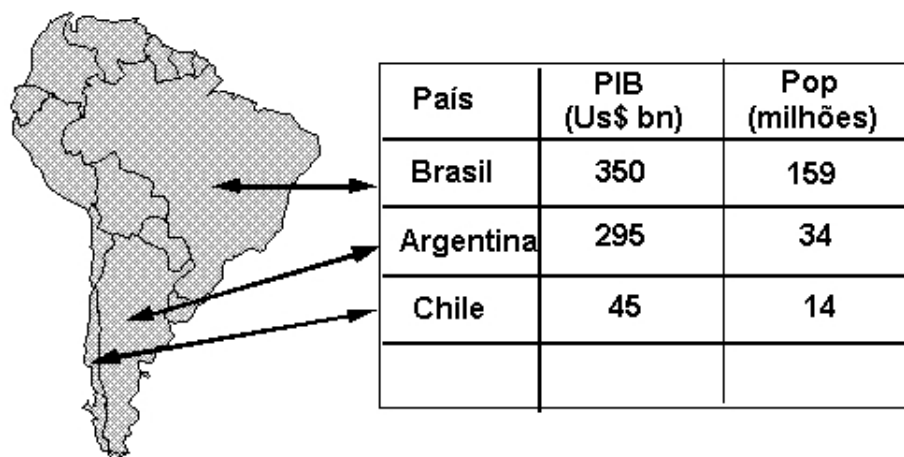


Figura 4.3: Exemplo de mapa cadastral (países da América do Sul) (SPRING, 2006)

Normalmente, para se analisar o comportamento das feições geográficas são usados mapas, onde a idéia de sobreposição de mapas é implementada. Esta sobreposição emprega o conceito de sobreposição de temas ou camadas, sendo criados diversos temas de dados georreferenciados, para um mesmo espaço geográfico (Figura 4.4) . Isso permite a combinação de dois ou mais temas para o processamento de determinada consulta.

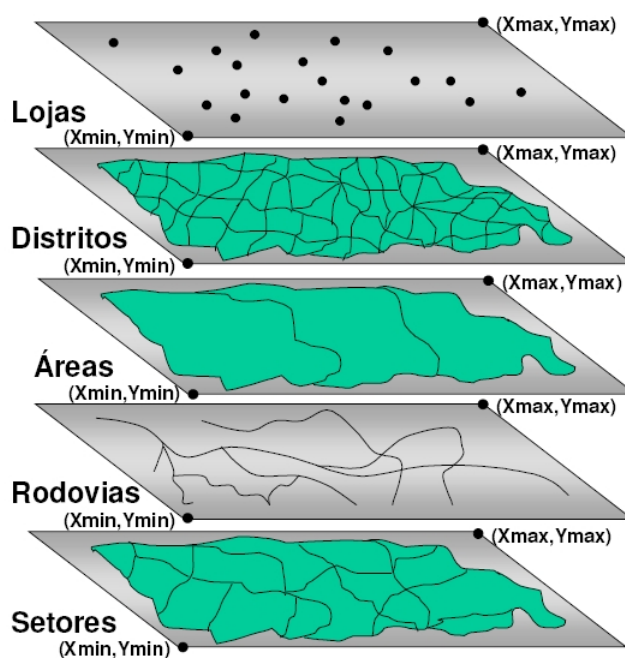


Figura 4.4: Exemplo de sobreposição de temas (FIDALGO, 2005)

4.2.1.3 Redes

Os mapas de redes armazenam suas informações em coordenadas vetoriais utilizando a topologia arco-nó, onde os arcos tem um sentido de fluxo e os nós possuem atributos vinculados (figura 4.5). Suas informações são associadas a serviços de utilidade pública, como luz, telefone, água ou malha viária, sendo fundamental sua ligação com um banco de dados. As redes são resultado direto da intervenção humana no meio ambiente.

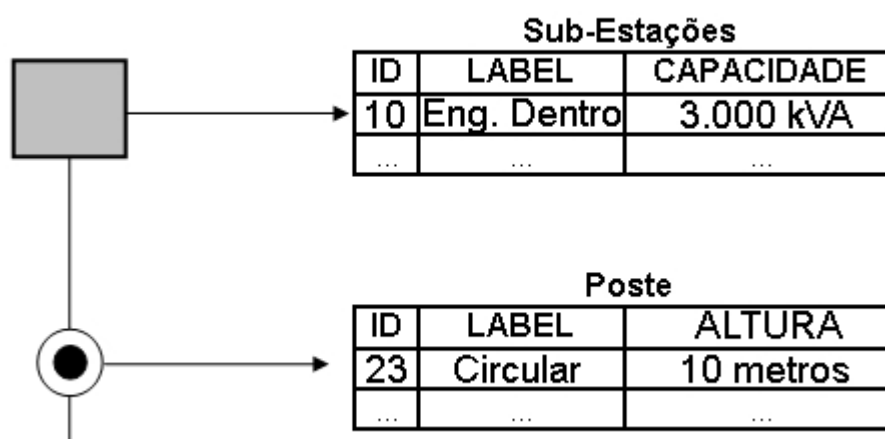


Figura 4.5: Elementos de Rede (SPRING, 2006)

4.2.1.4 Imagens

As imagens representam formas indiretas de captura de informação espacial, sendo seus dados armazenados como matrizes, onde cada elemento de imagem (pixel) tem um valor proporcional à reflectância do solo para a área imageada (TSUTOMU, 1997). São normalmente obtidas através de fotografias aéreas ou satélites (figura 4.6).

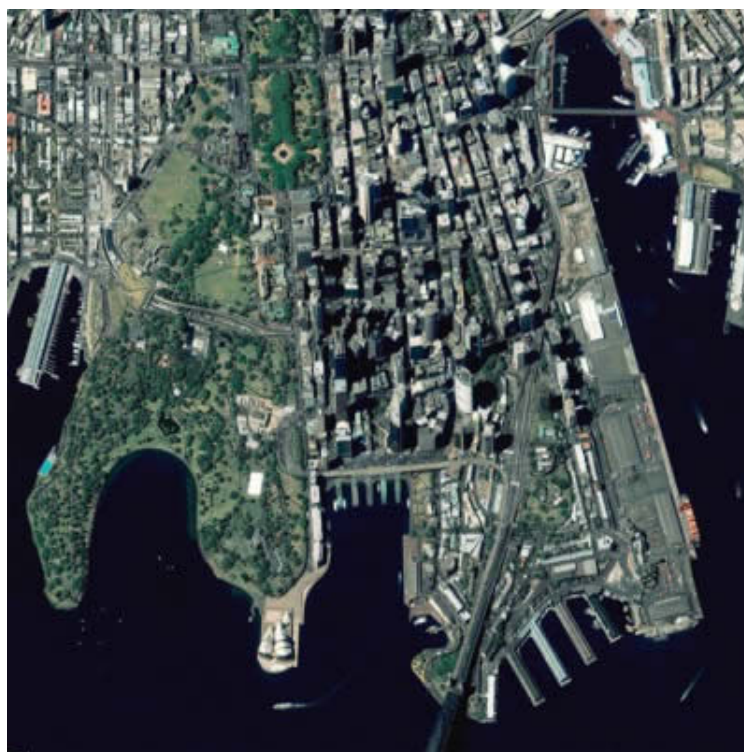


Figura 4.6: Imagem do Satélite IKONOS. Sydney, Austrália (EMBRAPA, 2006)

4.2.1.5 Modelos numéricos de terreno

Estes modelos são utilizados para representar uma grandeza que varia continuamente com o tempo. Comumente associados à altimetria, podem ser utilizados para modelar outros fenômenos de variação contínua, como o teor de minerais do solo, por exemplo. Suas amostras podem ser representadas de duas formas (SPRING, 2006):

- Grades regulares: matriz de elementos com espaçamento fixo, onde é associado o valor estimado da grandeza na posição geográfica de cada ponto da grade.
- Malhas triangulares: a grade é formada por conexão entre amostras do fenômeno, utilizando a triangulação de Delaunay (sujeita a restrições). A grade triangular é uma estrutura topológica vetorial do tipo arco-nó, formando recortes triangulares do espaço.

4.3 Banco de Dados Geográfico

A base de dados em um sistema SIG provê uma visão abstrata dos dados, deixando os detalhes de como os dados são armazenados e mantidos transparentes para o usuário.

Existem basicamente três principais formas de integração entre os SIGs e os Sistemas de Gerenciamento de Banco de Dados (SGBDs), as arquiteturas dual, em camadas e a integrada.

4.3.1 A arquitetura dual

Armazena as componentes espaciais dos objetos separadamente, ou seja, os dados alfanuméricos são armazenados em um SGBD convencional enquanto a componente espacial é armazenada em arquivos com formato proprietário. A Figura 4.7 mostra um desenho espacial dos países da América do Sul e uma tabela com os dados do PIB destes países. Observe que o desenho e os dados são representados separadamente, ilustrando a arquitetura dual.

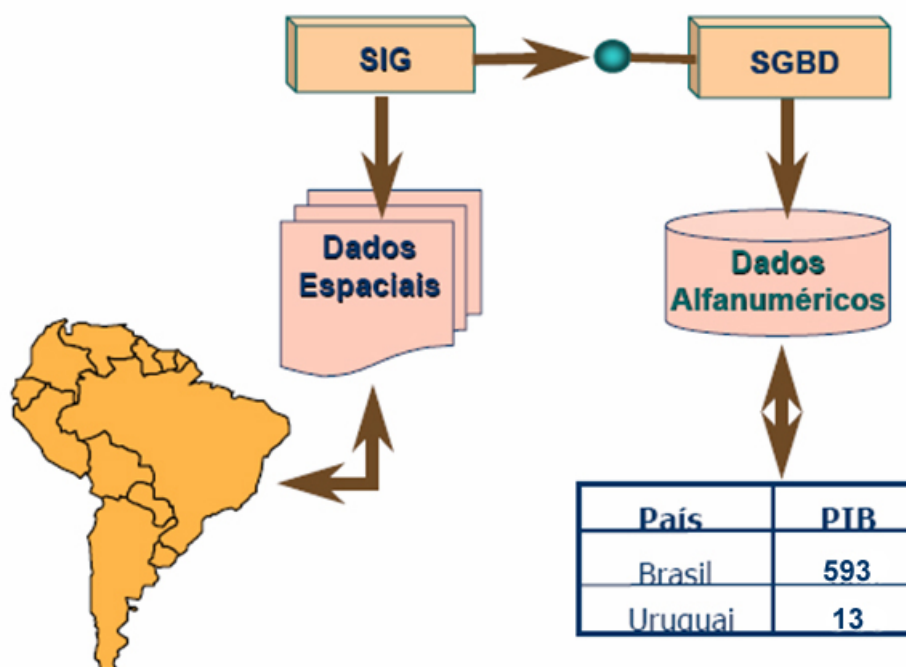


Figura 4.7: Arquitetura dual (QUEIROZ; CÂMARA, 2006) - PIB (IBGE/PAISES, 2007)

Alguns problemas são encontrados nessa arquitetura como (CASANOVA et al., 2005):

- Dificuldade em manter a integridade entre a componente espacial e a componente alfanumérica;
- Dificuldade no controle e manipulação das componentes espaciais;
- Separação entre o processamento da parte convencional, realizado pelo SGBD, e o processamento da parte espacial, realizado pelo aplicativo utilizando os arquivos proprietários;

- Dificuldade de interoperabilidade, já que cada sistema trabalha com arquivos com formato proprietário.

4.3.2 A arquitetura em camadas

Nesta arquitetura o banco de dados tem a função de dar suporte para campos longos, possibilitando o armazenamento de qualquer tipo de dado, sejam dados geográficos, numéricos ou pictóricos. Apesar do sistema de suporte espacial continuar fora do banco de dados (Figura 4.8), esta arquitetura concentra os objetos a serem armazenados, evitando os problemas de gerência de transações e de controle de integridade e de concorrência, presentes na arquitetura dual, permitindo assim acessos e atualizações multi-usuário.

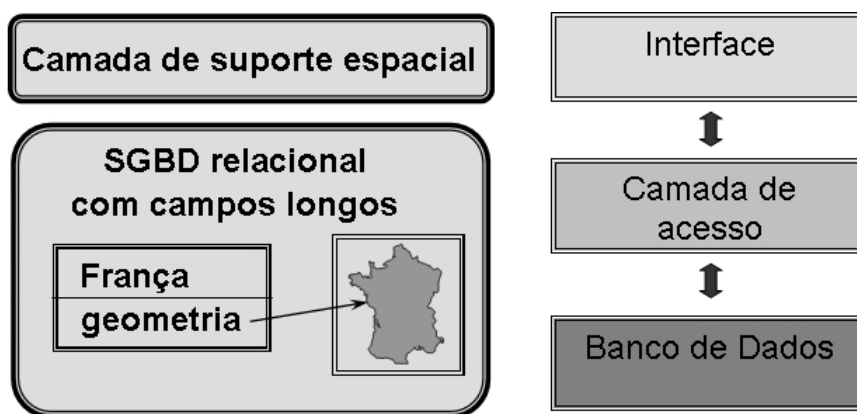


Figura 4.8: Arquitetura em Camadas (RIEDEL, 2006)

4.3.3 A arquitetura integrada

Consiste em armazenar tanto a componente espacial como os dados alfanuméricos (Figura 4.9), definindo tipos de dados geométricos e utilizando-se da extensão de linguagem SQL para atributos espaciais. Isso logicamente facilita a manipulação dos objetos espaciais, por evitar os problemas da arquitetura dual e utilizar-se plenamente dos recursos de um SGBD.

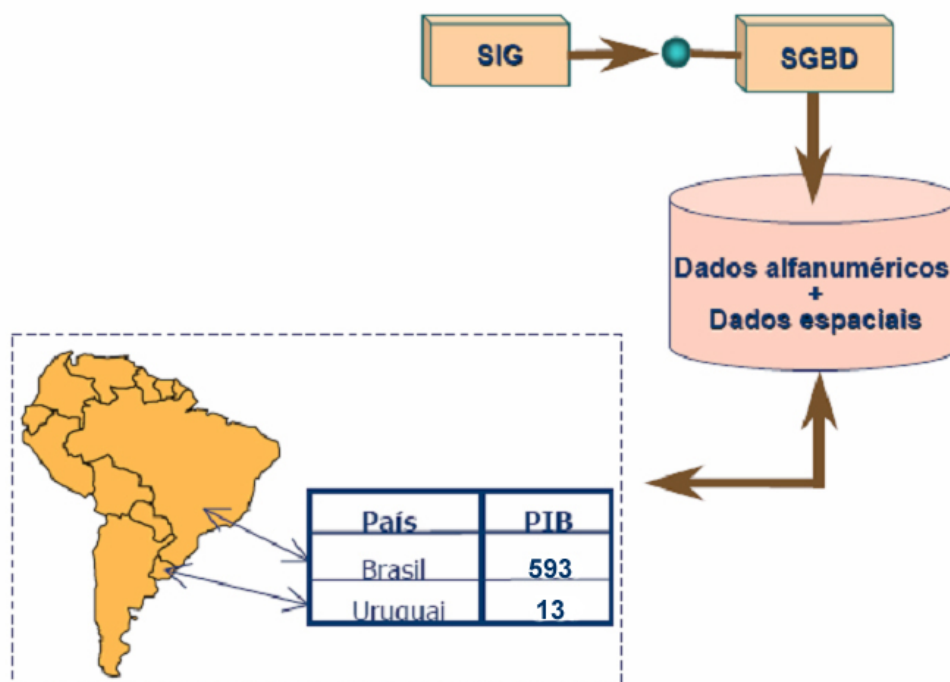


Figura 4.9: Arquitetura integrada (QUEIROZ; CÂMARA, 2006) - PIB (IBGE/PAÍSES, 2007)

Este trabalho utiliza-se da arquitetura em camadas para a implementação do sistema proposto. O SGBD utilizado é o Oracle 9i integrado com o SDE (Spatial Database Engine) da ESRI (ESRI, 2006); esse conjunto proporciona o ambiente de armazenamento e manipulação dos dados espaciais. O SDE juntamente com o componente MapObjects provê acesso a base espacial via aplicação para a execução de operações espaciais (ESRI, 2004). A Figura 4.10 resume a arquitetura da tecnologia utilizada neste trabalho.

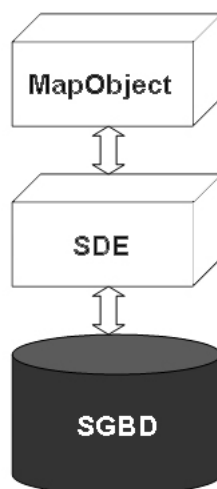


Figura 4.10: Arquitetura utilizada no projeto

A camada intermediária (SDE) se comunica tanto com o SGBD (Oracle 9i) quanto com a camada superior (MapObjects). A camada composta pela biblioteca MapObjects é utilizada pelo desenvolvedor para consultar e manipular dados espaciais passando pela camada SDE. Consultas e manipulação de dados comuns (não espaciais) são feitas diretamente no banco de dados, não passando portanto pelas camadas MapObjects e SDE.

4.3.4 Modelo de dados espaciais

Este modelo consiste em uma estrutura hierárquica de elementos, geometrias e planos de informação (layers), onde estes planos são formados por conjuntos de geometrias, as quais são formadas por conjuntos de elementos.

Cada elemento é associado a um tipo espacial primitivo (Figura 4.11).

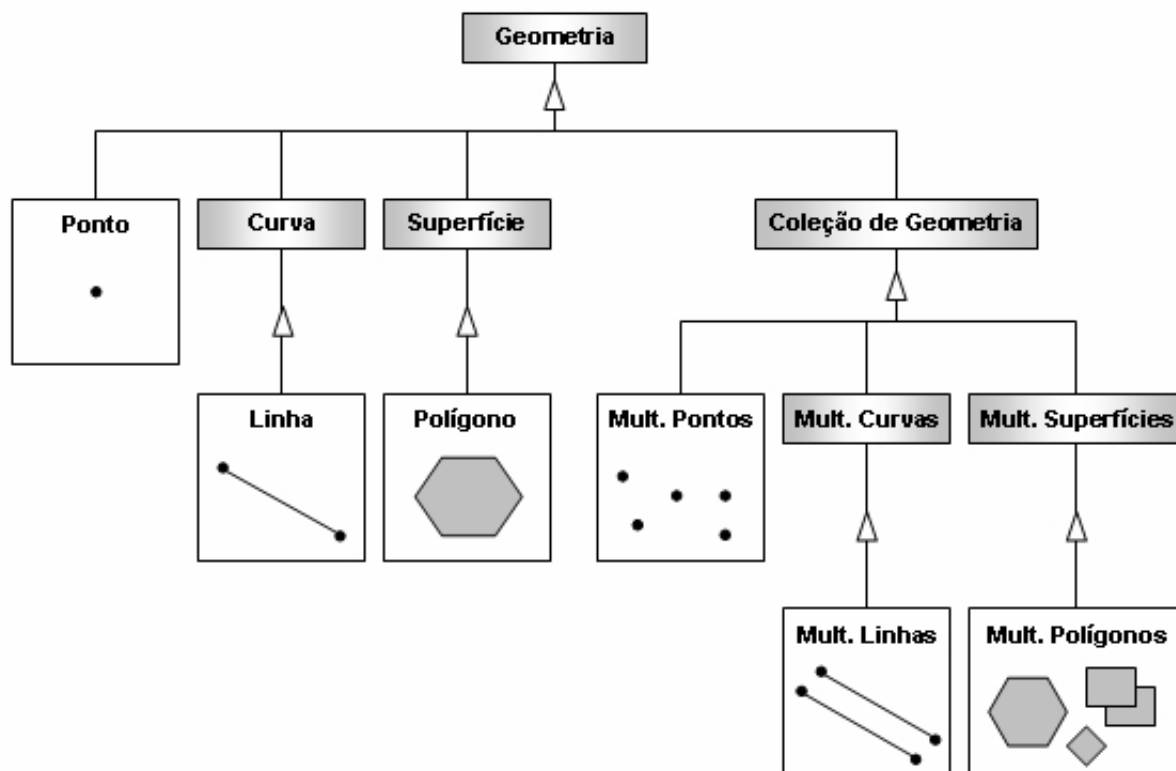


Figura 4.11: Tipos espaciais primitivos (CAVALCANTI, 2005)

4.3.5 Consultas espaciais

A existência de uma base de dados espacial possibilita a criação de consultas sem a necessidade de um relacionamento entre as tabelas. As consultas são realizadas baseando-se no relacionamento espacial entre as geometrias.

A Figura 4.12 representa uma consulta que retorna os pontos localizados em um determinado polígono (polígono em negrito). Para a execução desta consulta, não há a necessidade da existência de relacionamento entre as tabelas de polígono e de pontos, mas apenas que haja uma sobreposição espacial, como a mostrada na figura.

Apesar do exemplo da Figura 4.12, retornar como resultado da consulta, um desenho, este pode ser composto pelos dados descritivos da tabela (registros), da mesma forma que ocorre em consultas normais em banco de dados.

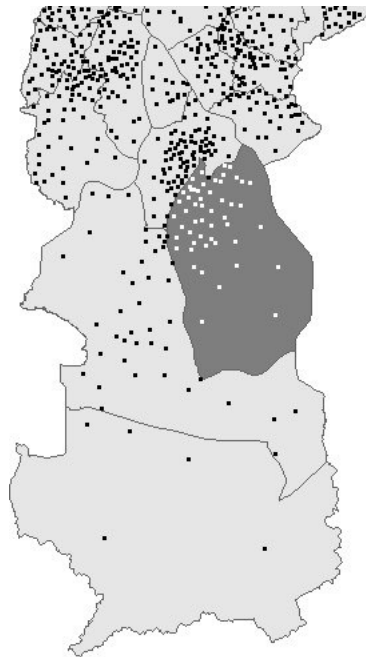


Figura 4.12: Pontos contidos em um polígono (CASANOVA et al., 2005)

Na Figura 4.13 é mostrado outro exemplo onde é gerado um polígono em torno de uma linha. Este polígono é denominado *buffer*, pois é uma representação temporária baseada em um critério de distância do objeto de origem (exemplo: 3km em torno da linha). Posteriormente são selecionados todos os polígonos que tocam este buffer (polígonos em negrito).

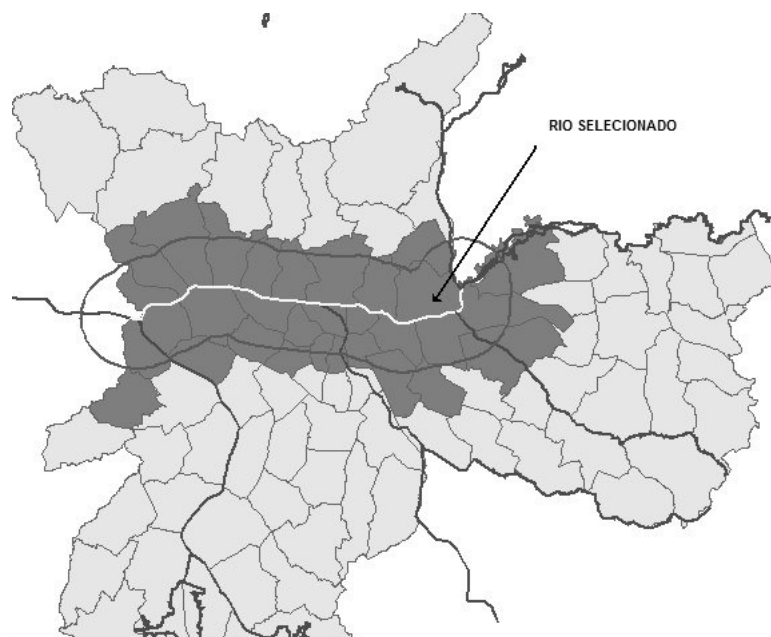


Figura 4.13: Criação de um buffer ao redor de um rio e a identificação dos polígonos que tocam este buffer (CASANOVA et al., 2005)

Para realizar as consultas espaciais foi utilizado neste trabalho o componente MapObjects da ESRI (ESRI, 2004), o qual possibilita a utilização de operadores espaciais.

5 *Sistema Proposto*

Neste capítulo será apresentado em detalhes o sistema GeoCom, fundamentando-se nos capítulos anteriores.

5.1 Descrição do Sistema

O sistema GeoCom foi desenvolvido com o objetivo de proporcionar um ambiente de análise comercial da rede de distribuição de energia. Para isso, utiliza-se do SIG e dos métodos de classificação anteriormente discutidos neste trabalho (capítulo 5).

O desenvolvimento foi realizado sobre uma base de dados real, a qual contém o levantamento de todos os componentes elétricos da rede de distribuição de média tensão.

Para a integração com esta base de dados espacial, foi utilizada a mesma linha de tecnologia da companhia, sendo o sistema de gerenciamento de dados Oracle interligado com o SDE, combinação que permite a manipulação dos dados espaciais utilizando a arquitetura em camadas como forma de integração.

A grande maioria das variáveis de configuração do sistema são armazenadas na base de dados, como escalas de visualização e simbologias, proporcionando um bom ambiente de manutenção.

5.1.1 Ferramentas de navegação

Como funcionalidades básicas de todo sistema de informação geográfica, as ferramentas de navegação também estão presentes no GeoCom (Figura 5.1).



Figura 5.1: Ferramentas de navegação do sistema GeoCom

Foram implementadas as seguintes ferramentas de navegação:

- Visão total: independente da escala atual, esta função retorna o mapa para a visualização total da base de dados.
- Ampliação de escala: esta função aumenta a escala em 50% para cada clique na tela.
- Diminuição de escala: esta função diminui a escala em 50% para cada clique na tela.
- Ampliação da área selecionada: nesta função o planejador seleciona uma área através do clique e arrasto do mouse. A área selecionada será ampliada de forma a ocupar todo o espaço disponível na tela de visualização do mapa.

5.1.2 Controle de camadas

O sistema apresenta um ambiente gráfico que retrata o mundo real de forma dinâmica e consistente, onde os dados são apresentados em camadas, sendo cada camada composta por dados do mesmo tipo (linha, ponto ou polígono). Para facilitar o manuseio, foi implementado um controlador de visualização destas camadas (Figura 5.2).

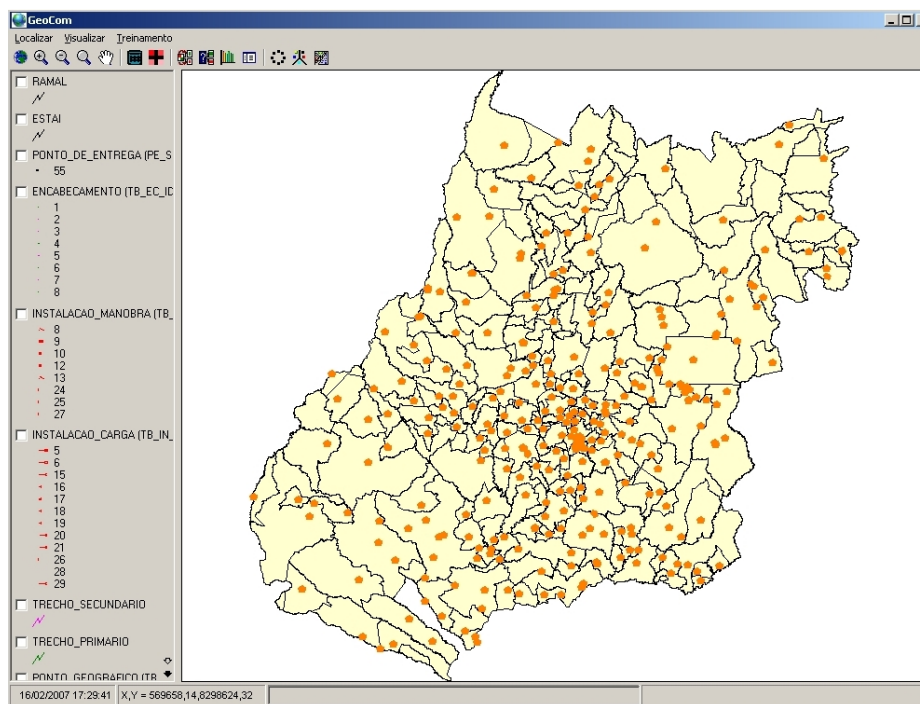


Figura 5.2: Tela principal do sistema GeoCom

5.1.3 Controle de escala

Pode-se observar, na Figura 5.2, que apenas as camadas município e subestação são mostradas, pois a escala é muito pequena e a representação dos demais elementos da rede de distribuição não traria benefício algum.

Isso ocorre porque foi implementado no sistema um controle automático de escala, onde para cada *zoom* é calculado a escala da tela e comparada com a escala de visualização das camadas presente no banco de dados, em tempo de execução. Caso a escala da tela atual seja menor que a escala de uma determinada camada, esta será exibida automaticamente.

Na Figura 5.3 são apresentadas as camadas de quadra e da rede primária (linhas verdes). Entretanto, as demais camadas não são mostradas nesta escala.

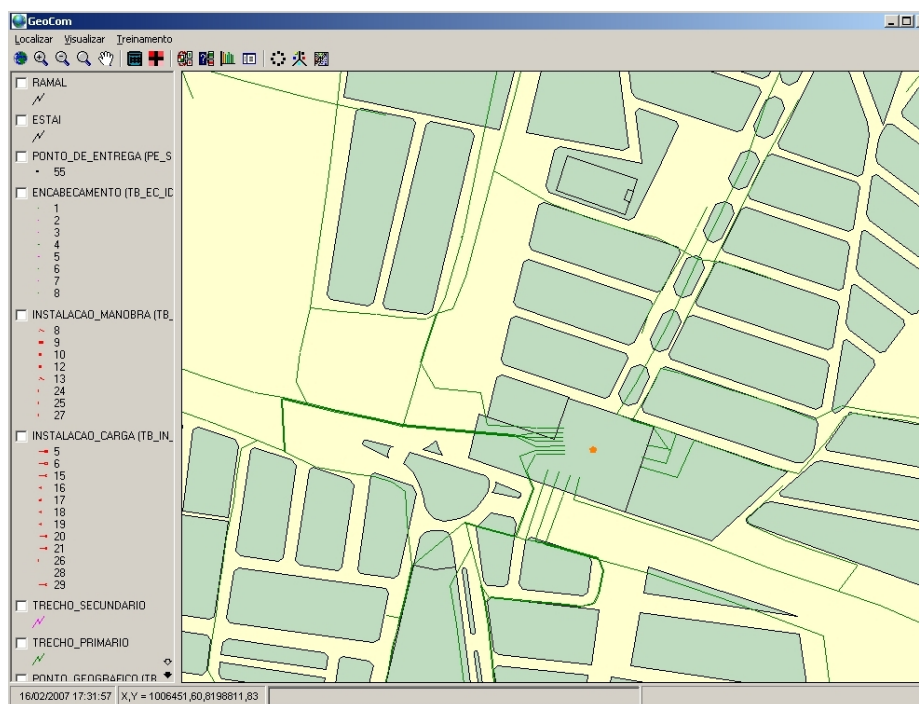


Figura 5.3: Exemplo de controle de escala

O controlador de visualização permite ainda que o planejador tenha liberdade de mostrar ou ocultar uma determinada camada, ignorando o controle automático. Dessa forma, uma camada pode ser vista em qualquer escala da tela. Isso, apesar de ser prático, pode sobrecarregar o banco de dados devido ao volume de informações que são carregadas na tela. A Figura 5.4 mostra a camada de trecho primária sendo vista fora de escala.

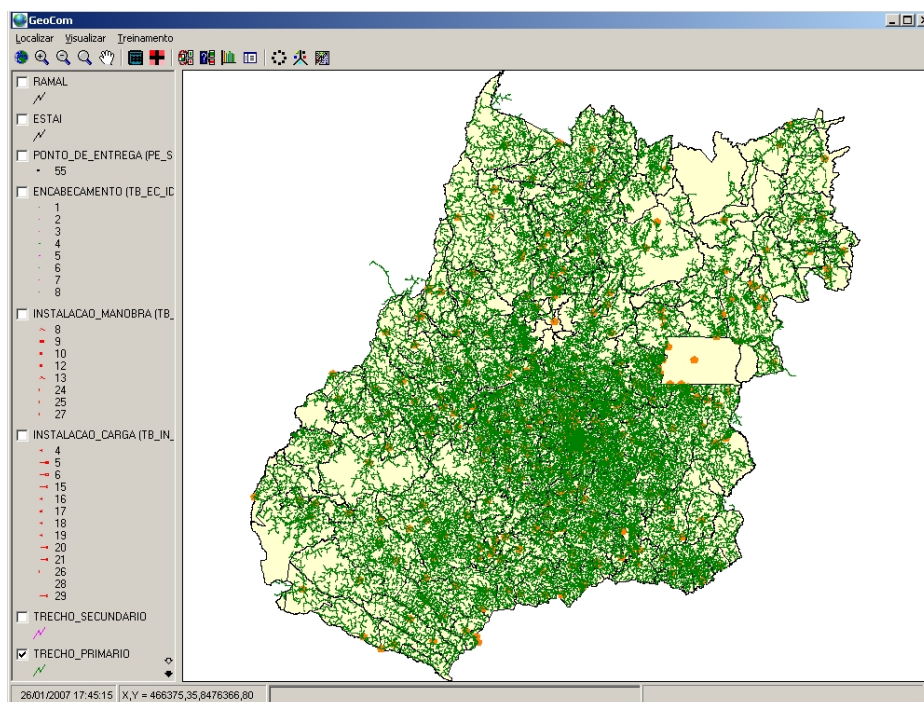


Figura 5.4: Exemplo de uma camada ativa fora de escala

5.1.4 Controle de símbolos

Um outro controle desenvolvido é o redimensionamento dos símbolos da rede de distribuição. Isto é necessário para manter a proporção entre as camadas e os símbolos apresentados na tela.

As camadas espaciais (municípios, quadras, trechos de rede, ...) são armazenadas no banco de dados e controladas pela camada de controle de acesso (seção 4.3.2), já os símbolos são desenhos ligados a elementos do tipo ponto. Assim, os símbolos necessitam de um tratamento diferenciado, sendo alterados em tempo de execução pela aplicação.

Esse controle é efetuado a cada ampliação realizada pelo planejador, sendo fundamental para uma boa representação da rede. A necessidade deste controle pode ser observada quando comparamos as Figuras 5.5 e 5.6.

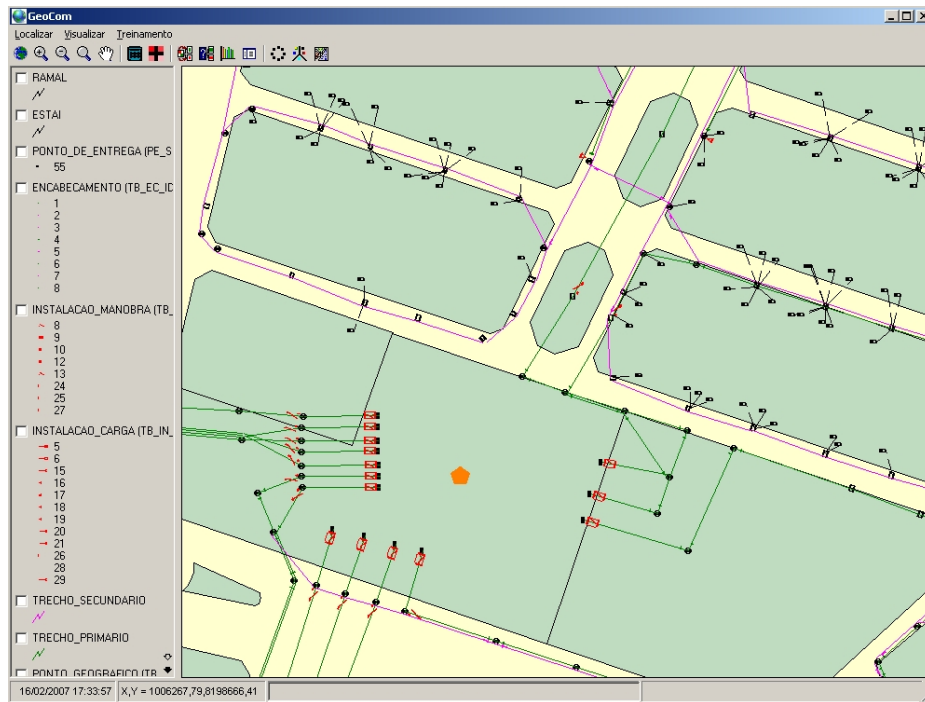


Figura 5.5: Exemplo de controle de símbolos

Observe que na Figura 5.5 os símbolos são apresentados em tamanho menor que na Figura 5.6, mantendo assim a proporção do desenho, já que as camadas que não utilizam simbologia (exemplo: quadras) também são redimensionadas automaticamente devido à mudança de escala da tela.

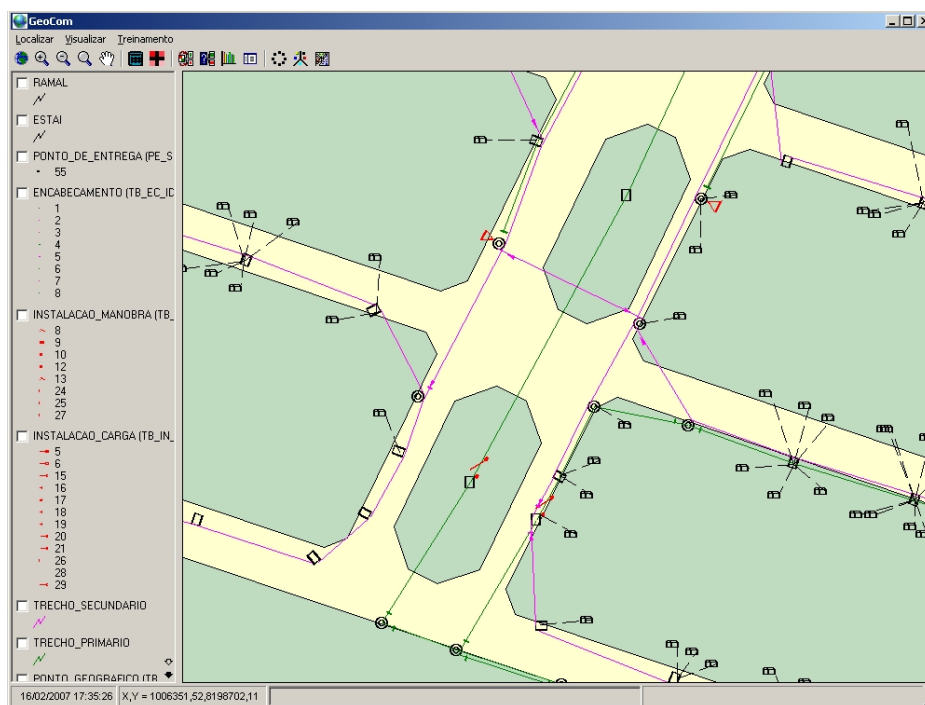


Figura 5.6: Exemplo de controle de símbolos

5.1.5 Histograma de Consumo

Para apresentar a distribuição de consumo de um determinado município foi disponibilizada a geração de um histograma de consumo (Figura 5.7). Este histograma aplica um degrade, na cor vermelha, sobre as quadrículas do município selecionado. Assim, quadrículas com coloração mais escura possuem maior consumo agregado que as quadrículas com coloração mais clara.

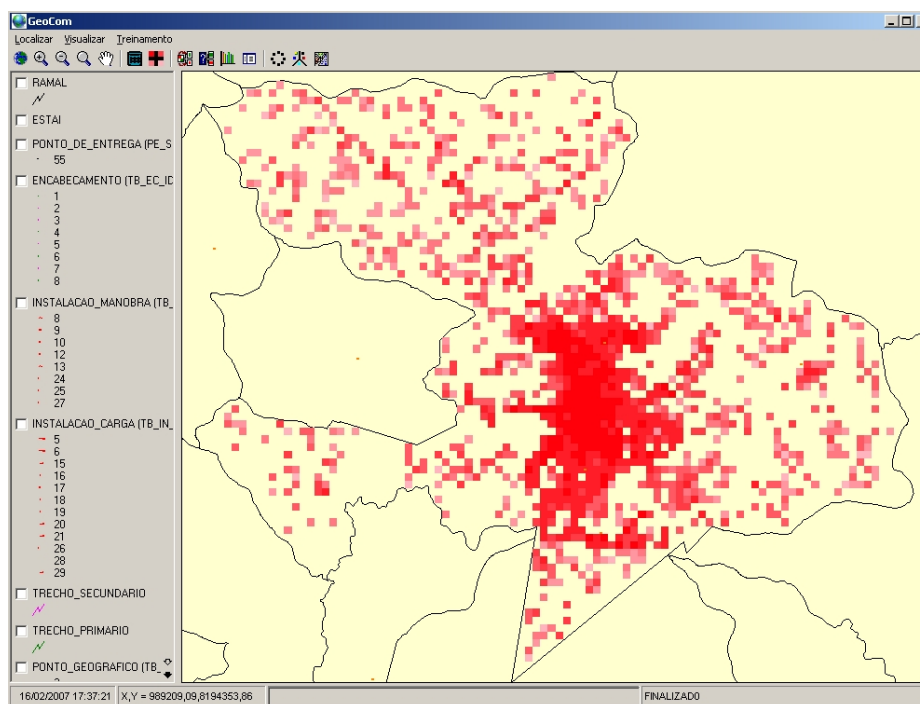


Figura 5.7: Histograma de consumo

Este histograma, juntamente com o produto da classificação das quadrículas, potencializa o processo de análise da rede de distribuição, já que o planejador tem uma informação a mais para a tomada de decisão da localização de construção de novas subestações e para a análise da classificação das quadrículas.

5.2 Metodologia Geral proposta para o Agrupamento

Para possibilitar o agrupamento e classificação das curvas de carga com o apoio do SIG foi proposta uma metodologia que possui quatro etapas, cada uma com suas respectivas atividades.

5.2.1 Coleta de informações - etapa 1

Nesta etapa está prevista a obtenção das informações necessárias para o agrupamento, ocorrendo o ajuste destas informações na etapa seguinte. São obtidos do setor comercial os dados de consumo de energia dos consumidores em kWh, referente aos 12 meses que compõem o cenário. A partir destes dados é avaliado o consumo de energia das entidades da rede elétrica georreferenciadas (pontos de entrega e estações transformadoras).

5.2.2 Pré-carga de dados - etapa 2

A base do agrupamento e classificação das áreas da rede de distribuição é a geração das curvas de carga. Estas curvas são geradas por quadrícula, onde cada quadrícula corresponde a uma grade uniforme de 500 x 500 m.

O processo de geração da curva de carga por quadrícula ocorre da seguinte forma:

1. Primeiramente é selecionada a área de abrangência do estudo (município).
2. Para cada quadrícula são somados os consumos médios (kWh) dos clientes presentes na área da mesma.
3. O consumo obtido na quadrícula é dividido por classe de clientes.
4. As curvas de carga são normalizadas.

Como os clientes não são entidades espaciais, estes são identificados em uma quadrícula através dos transformadores e dos pontos de entrega (pontos de medição) presentes na área da quadrícula.

A curva de carga da quadrícula obtida é referente ao total dos consumos dos últimos doze meses e a composição das curvas das quadrículas de um município gera o cenário de consumo do mesmo. Assim, por exemplo, para gerar o cenário de 2005 basta carregarmos o histórico de consumo daquele ano (etapa 1) e o sistema automaticamente identificará que os dados são referentes ao cenário de consumo de 2005.

O consumo total obtido em uma quadrícula é dividido em classes, gerando uma curva de classe por consumo. Devido a variabilidade de magnitude encontrada nas curvas de carga, o processo de normalização se faz necessário, sendo o intervalo $[-1,1]$ definido neste trabalho. Foram determinadas 10 classes possíveis de clientes:

1. Residencial de 0 a 50 kWh
2. Residencial de 50 a 200 kWh
3. Residencial de 200 a 400 kWh
4. Residencial acima de 400 kWh
5. Comercial de 0 a 200 kWh
6. Comercial acima de 200 kWh
7. Industrial abaixo de 1000 kWh
8. Industrial acima de 1000 kWh
9. Rural
10. Outros

Na geração das curvas por município o sistema calcula tanto pelos transformadores presentes na quadrícula quanto pelos pontos de entrega. Isso é feito para proporcionar ao planejador a opção de agrupar as curvas por transformador ou pontos de entrega, já que a quantidade de clientes na quadrícula pode ser diferente.

Note na Figura 5.8 que clientes alimentados por um transformador podem estar localizados em outra quadrícula, já que estes transformadores atendem seus clientes através das redes de baixa tensão (linhas na cor magenta), as quais normalmente possuem extensões que ultrapassam a área de uma quadrícula (500 x 500 m).

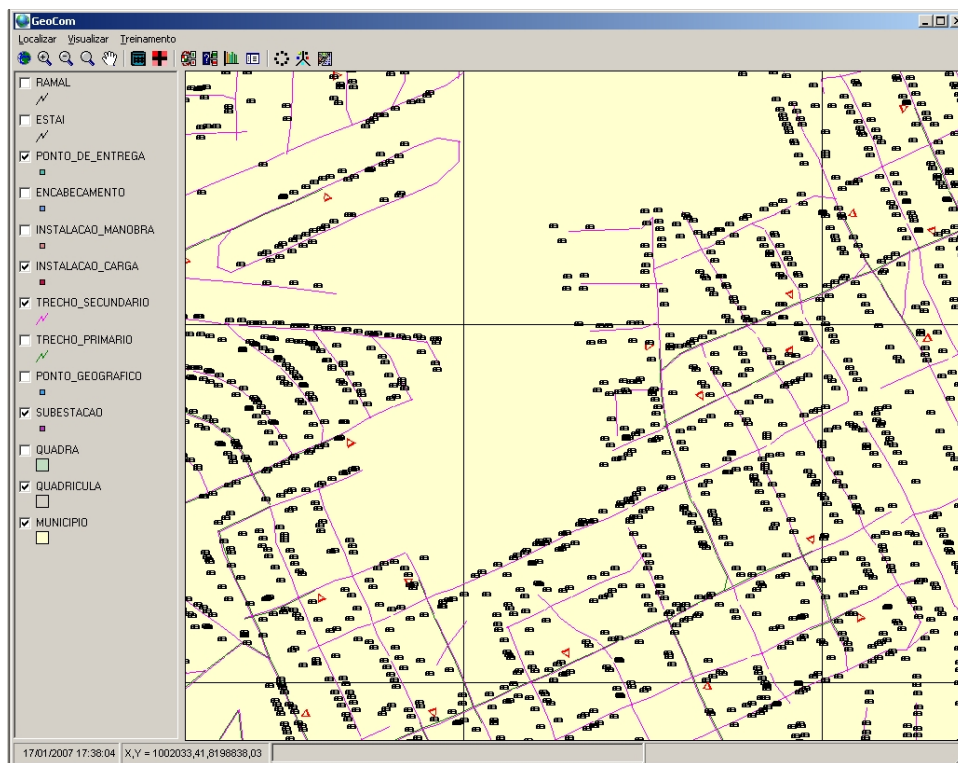


Figura 5.8: Exemplo de distribuição dos clientes nas quadriculas.

Vê-se na Figura 5.9, a qual mostra o resultado do cálculo das curvas de carga no município de Anápolis, que a quantidade de clientes é diferente quando calculados por transformador ou por ponto de entrega.

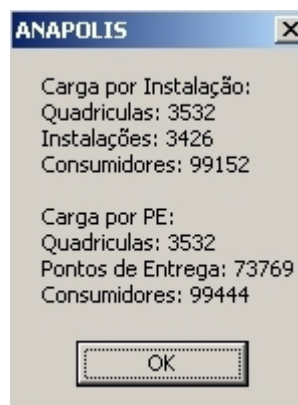


Figura 5.9: Diferença de clientes obtidos por transformador e por ponto de entrega

5.2.3 Agrupamento e classificação - etapa 3

Após a geração das curvas de carga aplica-se o processo de agrupamento das quadriculas gerando-se grupos que posteriormente serão rotulados. O processo de agrupamento pode ser

realizado através dos métodos: Kohonen em Anel, K-Médias e Fuzzy C-Médias.

Para iniciar o agrupamento deve-se selecionar o município, o ano base do cenário e preencher os dados solicitados pelo método escolhido. Segue uma descrição do preenchimento dos dados por cada método proposto:

- Kohonen em Anel (seção 3.2.3, p.40): Para este método os dados de entrada são: a taxa de aprendizagem, responsável pela atualização do nó vencedor; o raio de vizinhança, que define da quantidade de nós vizinhos que serão atualizados pelo nó vencedor; o número de classes de saída e o número de iterações para o treinamento.

A janela de configuração 'Kohonen' apresenta os seguintes campos e opções:

- Município:** ANAPOLIS
- Ano Base:** 2006
- Taxa de Aprendizado:** 0,3
- Raio de vizinhança:** 3
- Nº de Saídas:** 15
- Nº de Iterações:** 30000
- Selecione a carga:**
 - Por Instalação
 - Por Ponto de Entrega
 - Ambos
- Selecione os atributos para o agrupamento:**
 - Residencial de 0 a 50
 - Residencial de 50 a 200
 - Residencial de 200 a 400
 - Residencial acima de 400
 - Comercial de 0 a 200
 - Comercial acima de 200
 - Industrial de 0 a 1000
 - Industrial acima de 1000
 - Rural
 - Outros

Botões: Executar, Fechar

Figura 5.10: Configuração do treinamento pelo método Kohonen em Anel

- K-Médias: Os dados de entrada deste método são: o erro de agrupamento (normalmente igual a zero), que define o critério de parada, e o número de agrupamentos de saída do treinamento.

k-Médias

Município: ANAPOLIS

Ano Base: 2006

Nº de Saídas: 20

Erro: 0

Selecione a carga

Por Instalação

Por Ponto de Entrega

Ambos

Selecione os atributos para o agrupamento

Residencial de 0 a 50

Residencial de 50 a 200

Residencial de 200 a 400

Residencial acima de 400

Comercial de 0 a 200

Comercial acima de 200

Industrial de 0 a 1000

Industrial acima de 1000

Rural

Outros

Executar Fechar

Figura 5.11: configuração do treinamento pelo método K-Médias

- Fuzzy C-Médias (seção 3.3, p.45): Os dados de entrada deste método são os mesmos do K-Médias, porém há a adição do índice de fuzzificação, que define a distância permitida entre uma curva e o centro do grupo que esta será inserida. O critério de parada deste método é a variável ε , a qual mede a diferença entre o grau de pertinência de uma iteração atual e a anterior.

Fuzzy C-Means

Município: MARZAGAO

Ano Base: 2006

Índice de Fuzzificação: 1,25

Erro ε : 0,01

Nº de Saídas: 10

Selecione a carga

Por Instalação

Por Ponto de Entrega

Ambos

Selecione os atributos para o agrupamento

Residencial de 0 a 50

Residencial de 50 a 200

Residencial de 200 a 400

Residencial acima de 400

Comercial de 0 a 200

Comercial acima de 200

Industrial de 0 a 1000

Industrial acima de 1000

Rural

Outros

Executar Fechar

Figura 5.12: Configuração do treinamento pelo método Fuzzy C-Médias

Para este método a saída do agrupamento é uma tabela que apresenta o grau de pertinência das curvas aos *clusters* definidos no treinamento (Figura 5.13)

Qtd	Quadricula	Ai1	Ai2	Ai3	Ai4	Ai5	Ai6	Ai7	Ai8	Ai9	Ai10
67	96851415	99	0	0	0	0	0	0	1	0	0
68	96951415	5	54	0	0	0	0	10	0	30	0
69	97001415	97	0	0	0	0	0	2	0	0	0
70	97051415	57	1	0	0	0	0	42	0	0	0
71	97151415	5	14	0	1	0	0	7	0	71	1
72	97201415	0	0	0	0	0	0	0	0	0	100
73	97251415	0	0	0	0	100	0	0	0	0	0
74	97301415	0	0	0	0	0	0	0	99	0	0
75	97001440	3	12	0	0	0	0	85	0	0	0
76	97051440	1	94	0	0	0	0	4	0	1	0
77	97101440	2	1	0	0	0	0	96	0	0	0

Figura 5.13: Tabela de saída do agrupamento pelo método Fuzzy C-Médias

Neste formulário o planejador deve inserir o grau de pertinência mínimo (índice de defuzzificação), necessário para a curva pertencer a uma classe específica. Após a escolha do índice de defuzzificação o sistema mostra as curvas que ficariam fora da classificação (linhas marcadas em amarelo). Na Figura 5.13 nota-se que o planejador digitou um índice de 70% e o sistema acusou que 169 curvas ficarão fora da classificação. O planejador tem liberdade de refazer a avaliação do índice quantas vezes for necessário. Por fim, aplica-se a classificação (defuzzificação) com o índice escolhido.

Para todos os métodos, é dada a possibilidade de mudança dos dados de entrada pela desativação das classes de consumo. Isso acarreta uma mudança nos grupos gerados e aumenta a flexibilidade do sistema.

Os valores aplicados para o treinamento não são padrões, sendo necessário a sensibilidade do planejador para a obtenção de um bom conjunto de valores por município. Esta sensibilidade é obtida ao se analisarem as classes geradas por vários treinamentos diferentes.

5.2.4 Visualização e análise dos resultados - etapa 4

Nesta seção são apresentadas as formas de visualização e os índices de comparação dos métodos de agrupamento e classificação empregados.

5.2.4.1 Visualização das classes

Terminado o treinamento e agrupamento das curvas de carga, podem-se visualizar as classes geradas. Primeiramente, o planejador deve escolher o ano-base do treinamento e o tipo de obtenção dos clientes, se por transformador ou por ponto de entrega (ver Figura 5.14).

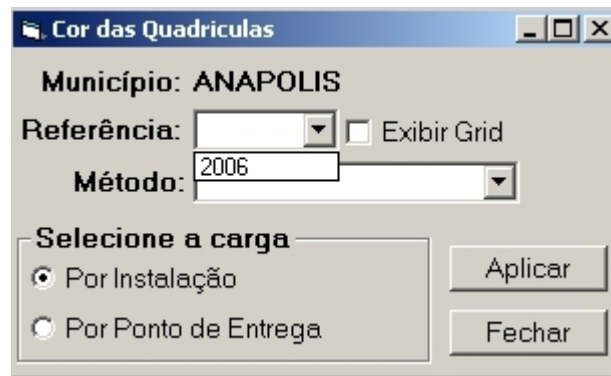


Figura 5.14: Seleção dos critérios para visualização das classes

Após a escolha dos parâmetros a visualização é montada e disponibilizada ao planejador (Figura 5.15). As quadrículas de mesma classe são coloridas da mesma cor, proporcionando uma visão clara das áreas com curvas de carga semelhante.

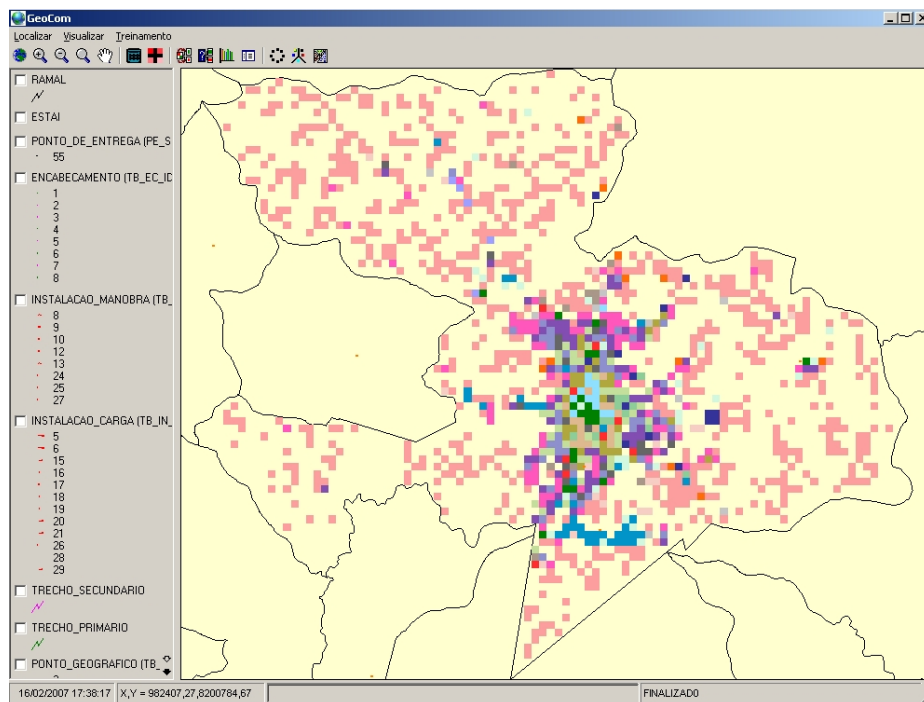


Figura 5.15: Visualização das classes geradas

Além da coloração das quadrículas, o planejador pode obter um gráfico de todas as curvas

de carga das quadrículas da classe selecionada. Basta selecionar uma quadrícula e o gráfico é apresentado (ver Figura 5.16). Observe que os padrões das curvas são próximos, confirmando assim a exatidão do processo de agrupamento. Além das curvas são mostrados: o número de curvas contidas na classe, o método utilizado e o número da classe.

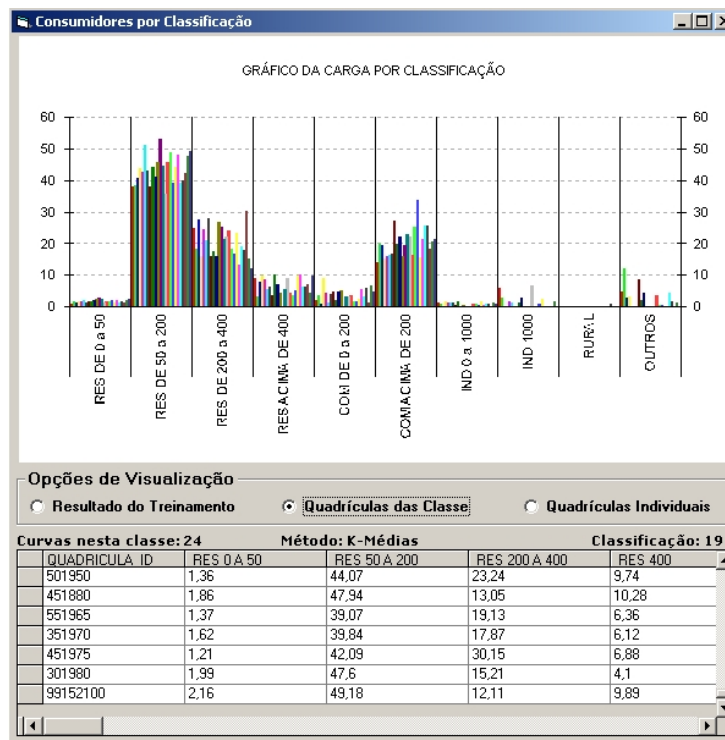


Figura 5.16: Visualização das curvas de carga das quadrículas de mesma classe

Há a opção de visualizar a curva de carga de uma quadrícula específica escolhendo a opção quadrículas individuais e selecionando a linha da planilha referente a quadrícula desejada (Figura 5.17).

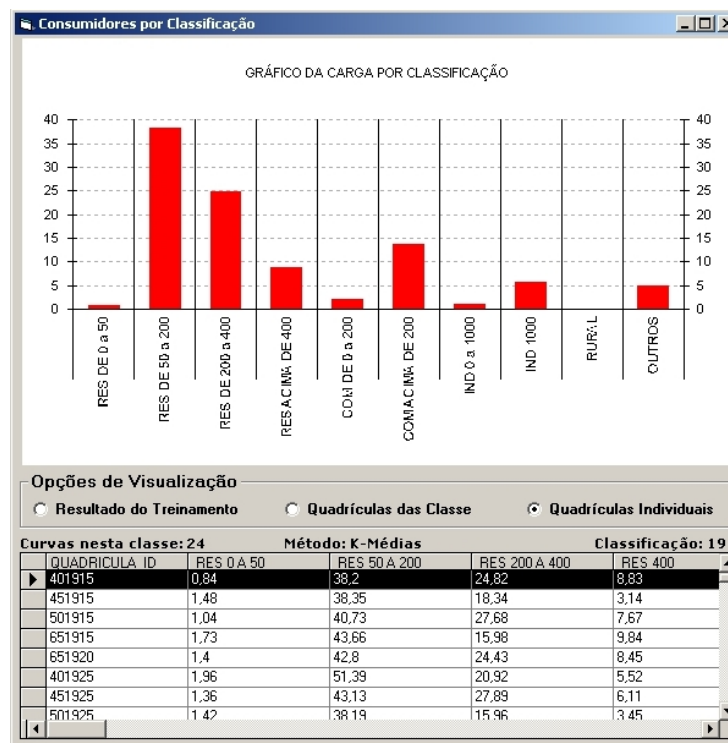


Figura 5.17: Visualização de uma determinada curva de carga

Caso o planejador queira visualizar a área da quadrícula da linha da planilha selecionada basta clicar duas vezes na linha e será feito um *zoom* na tela do mapa para a quadrícula específica (ver Figura 5.18).

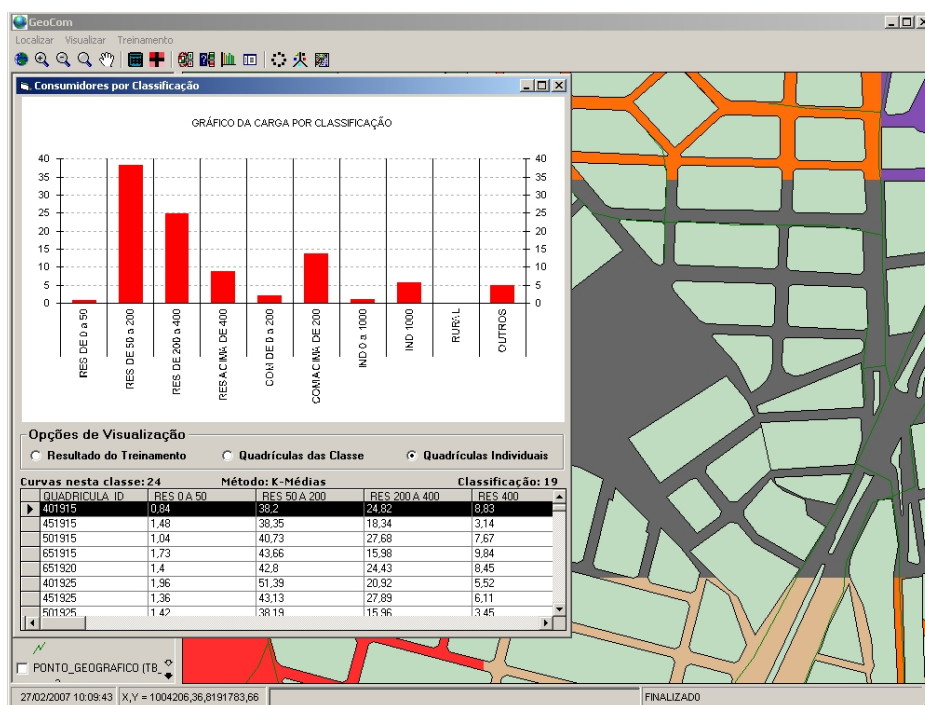


Figura 5.18: Visualização de uma quadrícula selecionada na planilha.

Para uma melhor análise da classificação realizada, há também a opção de visualizar o gráfico da classe. Para isso a opção Resultado do Treinamento deve ser selecionada. Nota-se na Figura 5.19 que o gráfico da classe corresponde ao gráfico apresentado na Figura 5.16, que representa as curvas de carga das quadrículas que participam desta classe.

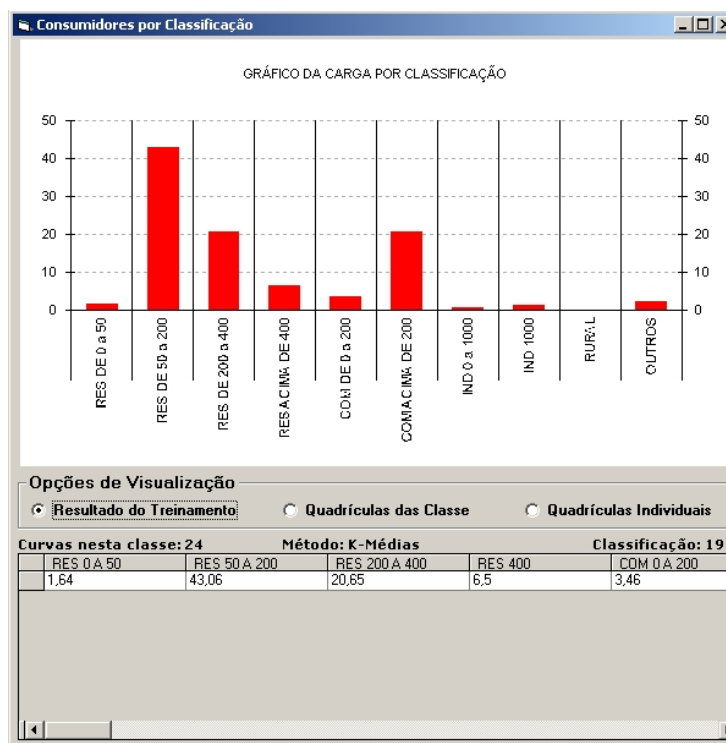


Figura 5.19: Visualização da curva do centróide de um grupo

Caso haja a necessidade de se obter um gráfico detalhado de uma determinada quadrícula, basta clicar no botão "gerar gráfico da quadrícula" e posteriormente clicar na mesma.

Caso a opção de treinamento tenha sido por instalação, serão apresentados o gráfico da quadrícula, as instalações presentes nesta quadrícula e os clientes presentes em cada instalação, conforme Figura 5.20.

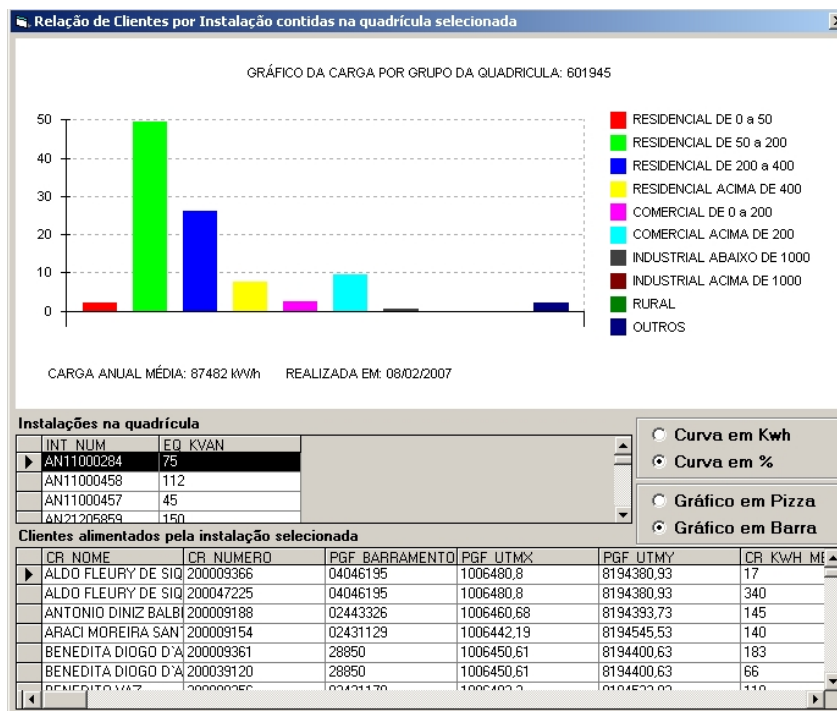


Figura 5.20: Visualização dos detalhes da quadriculas por Instalação

Entretanto, caso a opção de treinamento tenha sido por ponto de entrega, serão apresentados o gráfico da quadricula e apenas a relação de todos os clientes presentes nesta quadriculas, já que o posicionamento das instalações é desconsiderado nesta opção (Figura 5.21).

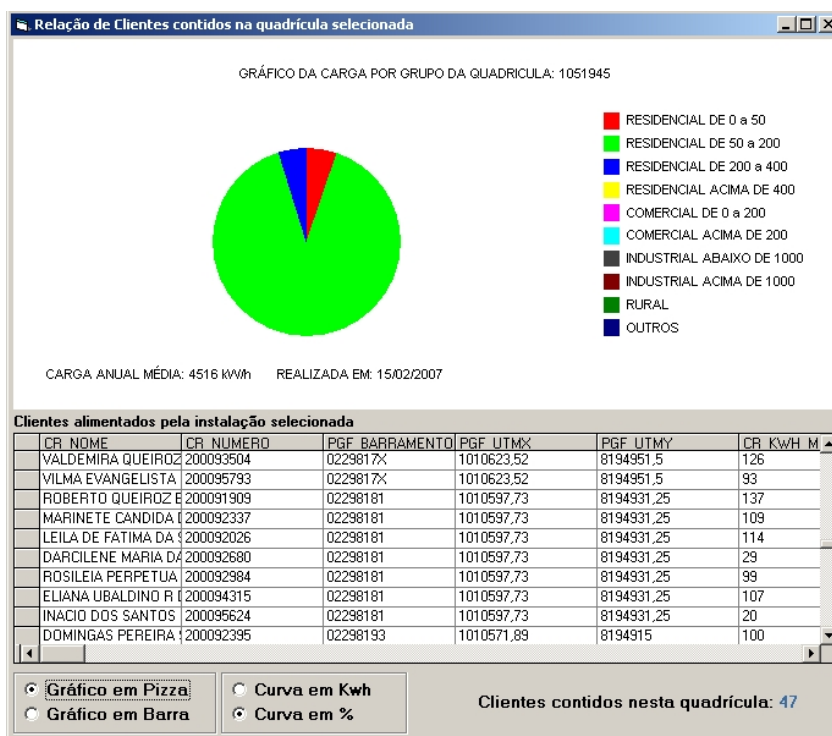


Figura 5.21: Visualização dos detalhes da quadriculas por Ponto de Entrega

A Figura 5.21 mostra um gráfico gerado com a opção Gráfico em Pizza.

5.2.4.2 Medidas de desempenho - índice intra-grupo (IAG)

Conforme apresentado no Capítulo 3, o índice Intra-Grupo (IAG) foi utilizado para fazer uma comparação dos resultados do agrupamento. Onde, através da função de custo baseada no erro quadrático médio, ele indica a similaridade dos padrões de entrada presentes em um mesmo grupo.

A Figura 5.22 mostra o resultado do treinamento pelo método K-Médias no município de Anápolis - GO.

Classes	consres0_50	consres50_200	consres200_400	consres400	conscom0_200	conscom200	consind0_1000	consind1000	cor
0	0	1,93027128	0	0	0	0	0	0	
1	0,30575741	8,48173285	4,03935098	1,89950195	0,74153304	11,85679625	6,05658215	0	
2	0,40186611	2,59778456	0	0	0	1,12036776	0	0	
3	1,87360895	6,30035308	0,67264828	2,86905347	9,09119977	1,3017321	0	0	
4	0	0	0	0	0	0	0	0	
5	0,21660984	1,36949062	4,34664276	0	0,20300151	2,16562739	0	0	
6	1,19566506	1,20927869	3,85511364	0	1,27923105	0	0	0	
7	1,65954279	12,30230589	0,65338315	11,76729298	0,2694119	1,48884544	0,01963034	2,28978236	
8	2,88690864	0	12,26490895	8,11420547	2,95533932	13,44066215	2,3856737	6,28628223	
9	0,04976767	1,11833282	1,25488903	1,66620295	0,51321356	0,64179224	0,25468808	0,44388486	

Figura 5.22: Medida de desempenho IAG - exemplo do método Fuzzy C-Médias

A medida da distância das curvas ao centróide de seu agrupamento é apresentada por faixa de consumo e totalizada na última coluna. Por fim, todas as distâncias são somadas e o erro quadrático total é apresentado no formulário.

Portanto, para um mesmo município, o método que obtiver menor erro quadrático total terá realizado uma melhor classificação das curvas de carga.

5.2.4.3 Medidas de desempenho - índice inter-grupo (IEG)

Este índice determina a relação entre os grupos, ou seja, o quão distantes eles estão de uma curva referencial. Esta curva referencial é obtida pela média das curvas do treinamento.

A Figura 5.23 mostra o resultado do treinamento pelo método K-Médias no município de Anápolis - GO.

Cálculo do Índice Inter-Grupo (IEG)

Município: ANAPOLIS

Referência: 2005 Método: K-Médias

Seleção a carga:
 Por Instalação
 Por Ponto de Entrega

Erro quadrático por classe Erro quadrático total: 811

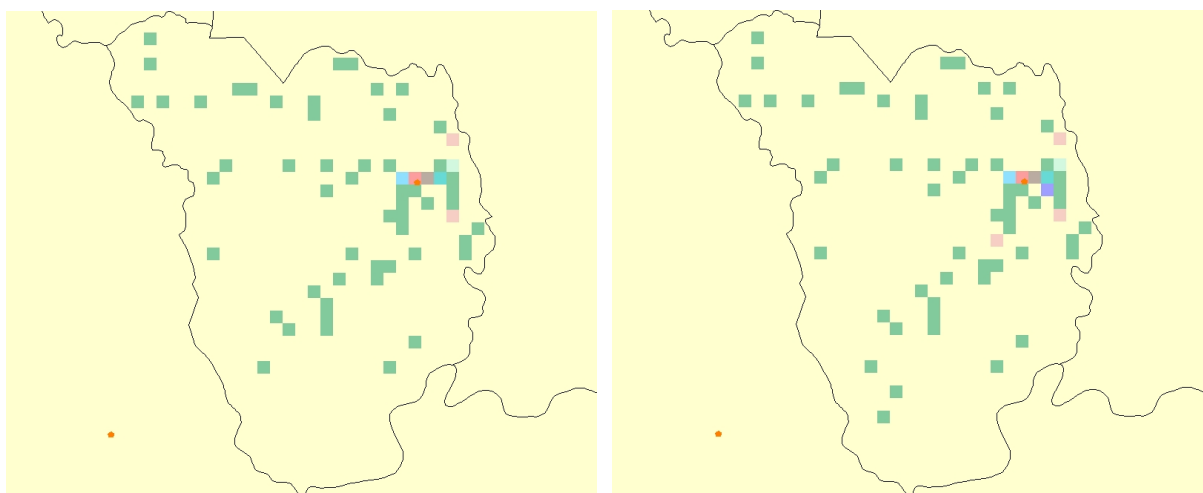
Classes	consres0_50	consres50_200	consres200_400	consres400	conscom0_200	conscom200	consind0_1000	consind1000	cor
2	0,004	0,263	0,184	0,081	0,025	1,667	0,017	0,077	
3	0,039	0,847	0,368	0,16	0,044	1,855	0,034	0,154	
4	2,003	1,435	0,552	0,241	0,000000001E-02	2,044	0,051	0,231	
5	2,024	1,937	0,64	2,16	0,000000001E-02	2,125	0,000000001E-02	0,308	
6	2,088	3,273	0,731	2,241	0,131	2,314	0,000000001E-02	0,385	
7	2,119	3,537	2,531	2,611	0,148	2,456	0,102	0,405	
8	2,168	4,949	2,896	2,649	0,181	2,533	0,109	0,44	
9	2,185	5,25	3,834	3,983	0,31	4,264	0,124	0,446	
10	2,224	6,302	4,678	4,104	0,547	6,041	0,233	0,778	
11	2,256	6,844	4,836	4,147	0,572	6,19	0,25	0,855	

Figura 5.23: Medida de desempenho IEG - exemplo do método Fuzzy C-Médias

Este índice não permite afirmar que quanto menor o valor, melhor é o resultado, mas apenas que, se ele tiver um valor alto, os grupos estão mais separados ou, se ele tiver um valor baixo, há uma semelhança maior entre os padrões de grupos diferentes.

5.2.4.4 Análise do crescimento horizontal

Após a etapa de classificação o planejador tem a opção de realizar uma análise do crescimento horizontal da rede. Para essa análise é necessário que haja a carga e classificação de pelo menos dois cenários, Figuras 5.24(a) e 5.24(b).



(a) Cenário 2005

(b) Cenário 2006

Figura 5.24: Município de Marzagão - cenários 2005/2006

O processo de análise do crescimento horizontal é simples, basta selecionar o cenário inicial, o cenário final e qual o método de classificação para visualização (Figura 5.25).

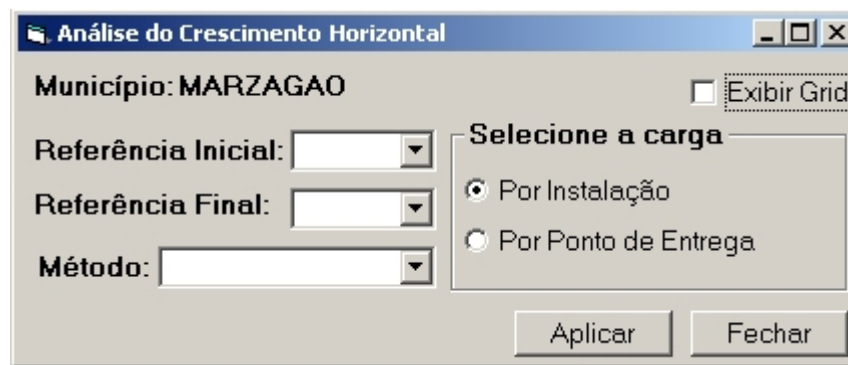


Figura 5.25: Tela de análise do crescimento horizontal

Neste processo, os consumos das quadrículas dos cenários seleccionados são comparados. Posteriormente são identificadas as quadrículas que possuem um consumo diferente de zero no cenário final e um consumo igual a zero no cenário inicial.

O resultado desta análise sobre o município de Marzagão é mostrado na Figura 5.26.

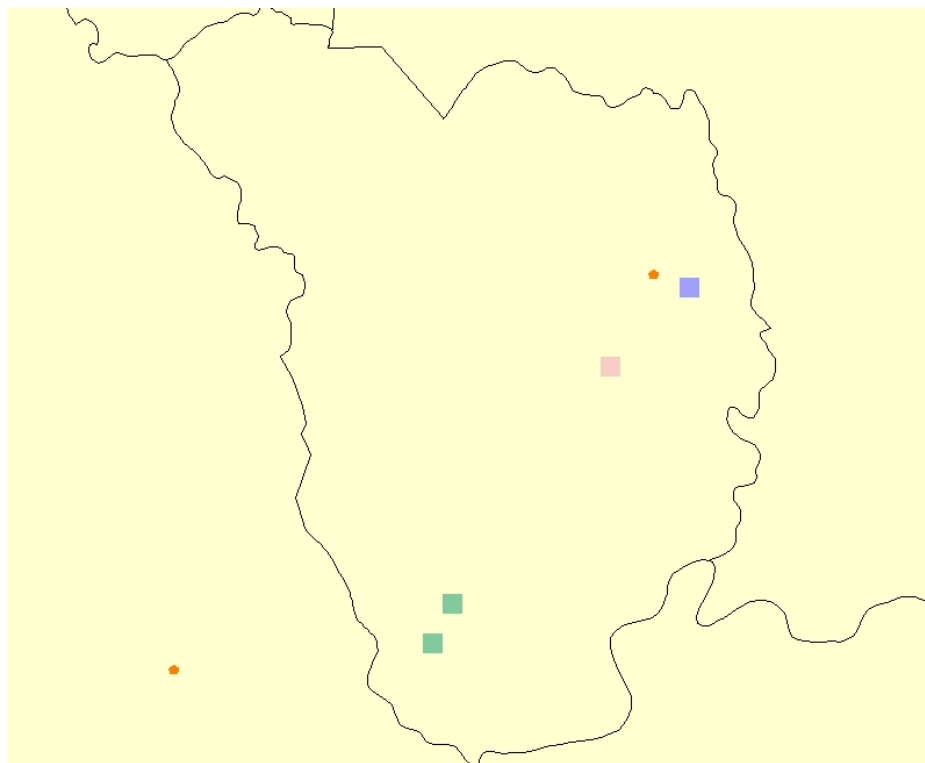


Figura 5.26: Crescimento Horizontal - Marzagão 2005/2006

O método de classificação é solicitado para que, além da visualização das quadrículas que identificam o crescimento horizontal da rede, o planejador possa saber também qual o perfil deste crescimento. Assim, as quadrículas são apresentadas com sua classificação do cenário final escolhido para a análise.

5.2.4.5 Análise de crescimento vertical

Outra possibilidade de análise é a do crescimento vertical. Neste processo, o planejador seleciona o cenário inicial, o cenário final, o método de classificação e a taxa mínima de crescimento, a qual determina o crescimento de consumo mínimo que a quadrícula deve possuir para ser identificada (Figura 5.27).

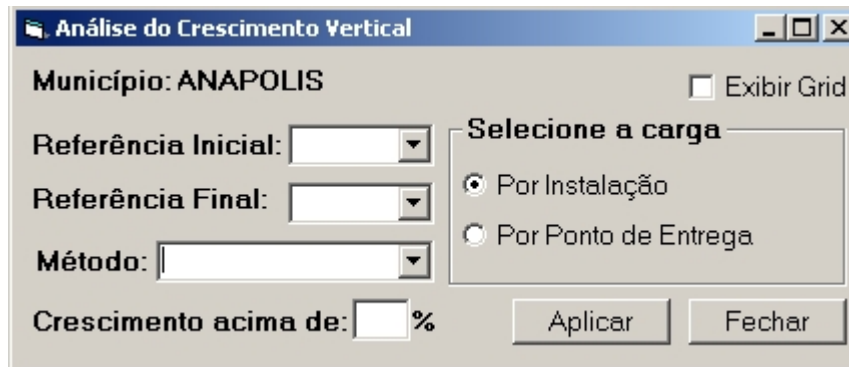


Figura 5.27: Tela de análise do crescimento vertical

Desta forma, todas as quadrículas que obtiverem um crescimento igual ou superior a taxa de crescimento informada serão apresentadas (Figura 5.28).

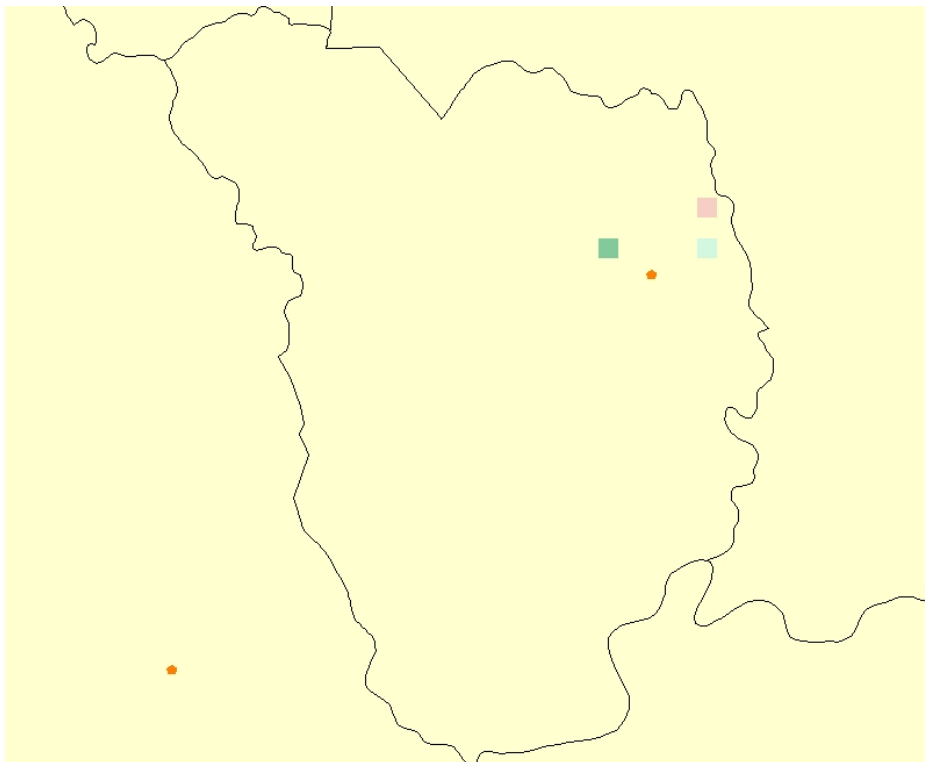


Figura 5.28: Crescimento Vertical - Marzagão 2005/2006

São desconsideradas as quadriculas que no cenário inicial continham consumo igual a zero, pois estas são avaliadas pelo processo do crescimento horizontal.

6 *Estudo de Caso*

6.1 Considerações Gerais

Neste capítulo são aplicadas as técnicas de agrupamento e classificação descritas no Capítulo 3 com o objetivo de comparar os desempenhos dos algoritmos implementados.

Conforme apresentado no capítulo anterior, os algoritmos de agrupamento e classificação foram implementados em Visual Basic 6.0, utilizando-se do Sistema de Informação Geográfica (SIG).

6.2 Metodologia de Análise dos Dados

Os dados utilizados neste trabalho foram obtidos da Companhia Energética de Goiás (CELG) e são referentes aos municípios de Aparecida de Goiânia, Anápolis e Rio Verde. Estes dados são georreferenciados e retratam o posicionamento real da rede de distribuição da empresa.

Para a aplicação das técnicas de agrupamento e classificação, foram criadas curvas de consumo por classe de clientes (curva de carga) posicionadas em uma mesma quadrícula (quadrado de 500 x 500 metros), as quais, devido a diferença de amplitude de consumo, foram normalizadas de -1 a 1 para o treinamento.

Entretanto, nos estudos de caso apresentados a seguir, as curvas dos centróides são normalizadas de 0 a 100% para facilitar sua interpretação. Para a obtenção e análise destas curvas, foi desenvolvida uma metodologia, a qual é descrita no capítulo 5, seção 5.2.

O processo de visualização das classificações dos municípios é realizado pela coloração das quadrículas, onde quadrículas de mesma classe possuem a mesma cor. Essas cores são atribuídas às quadrículas seguindo o padrão da Figura 6.1.

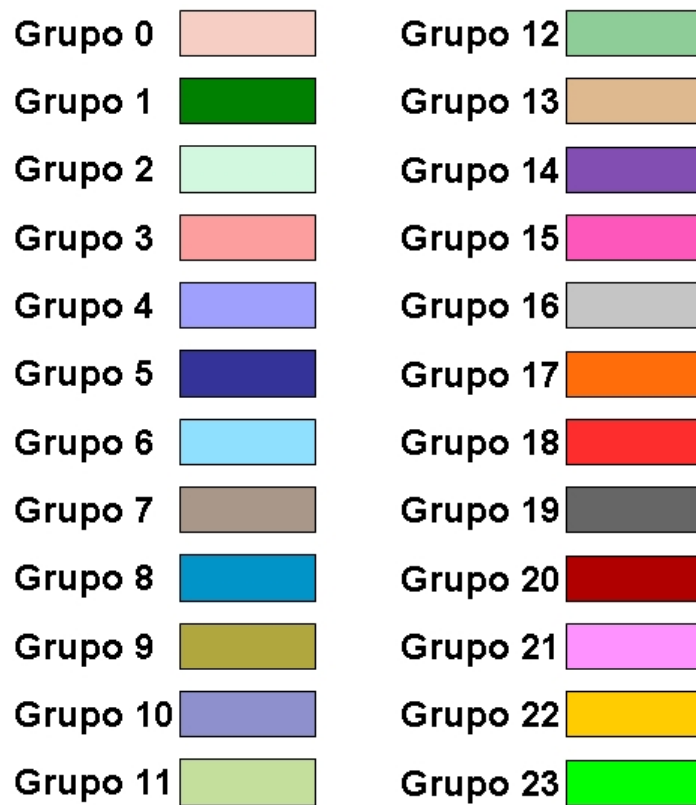


Figura 6.1: Legenda de cores da classificação

Assim, as cores das quadrículas mais os gráficos das curvas dos centróides compõem as ferramentas básicas de visualização e análise das classificações.

6.3 Considerações sobre as Técnicas Empregadas

O processo de extração do conhecimento utilizando as técnicas de agrupamento e classificação é uma tarefa difícil. Para auxiliar os planejadores nesta extração e na decisão sobre como utilizar este conhecimento, foram desenvolvidas ferramentas de visualização e análise dos resultados obtidos.

Conforme apresentado no Capítulo 3, foram usadas três técnicas para o agrupamento e classificação das curvas de carga, onde duas usam a abordagem convencional (Kohonen em Anel e K-Médias) e uma usa a abordagem nebulosa (Fuzzy C-Médias).

Para a obtenção de resultados satisfatórios, é importante a determinação de alguns parâmetros tais como, tipo de normalização, conjunto de entrada, número de saída e outras. Entretanto, a obtenção dos parâmetros adequados é um processo empírico, pois depende do nível de experiência do planejador e do número de testes realizados.

O desempenho destas técnicas pode ser avaliado por meio de três fatores (ZANON; LIMA, 2003):

- **Convergência:** é a capacidade de um modelo aprender um determinado conjunto de dados definido no procedimento de treinamento.
- **Generalização:** é a habilidade de um modelo em reconhecer padrões que não estão no conjunto de treinamento.
- **Escalabilidade:** refere-se ao procedimento de ajustar os parâmetros de um modelo, para que o mesmo consiga convergir e generalizar satisfatoriamente.

Para a obtenção dos parâmetros adequados e conseqüentemente dos melhores resultados, foram realizadas inúmeras simulações em cada estudo de caso apresentados neste capítulo.

6.3.1 Mapa Auto-Organizável (SOM)

Neste trabalho, foi utilizada a variação da rede SOM, denominada Kohonen em Anel, a qual possui uma camada de saída unidimensional e uma vizinhança em forma de anel, onde o primeiro nó do anel é vizinho do último.

A seguir são apresentados os parâmetros utilizados.

6.3.1.1 Conjunto de entrada

Os conjuntos de entrada são divididos por município e são compostos pelas curvas de consumo médio dos clientes por classe (curvas de carga), criando-se cenários anuais para cada município. Essas curvas são obtidas pelas quadrículas presentes no interior dos limites do município, sendo excluídas do treinamento as quadrículas que possuem consumo igual a zero.

Como a base de dados é espacial e essas curvas de carga são obtidas pela distribuição espacial da rede elétrica, a criação de um cenário possibilita guardar uma "fotografia" da rede de distribuição dos anos anteriores. Essa "fotografia" da rede não mais poderá ser obtida, já que a rede elétrica é dinâmica e está em constante expansão.

Portanto, a todo começo de um novo ano deve ser criado o cenário de cada município, garantindo assim uma boa base de dados para treinamento e análises futuras da rede de distribuição.

6.3.1.2 Normalização

Na etapa 2 da metodologia proposta neste trabalho (Capítulo 5, seção 5.2.2), a qual ocorre antes de iniciar o agrupamento e classificação, os dados são normalizados por curva, de duas formas: em uma faixa entre -1 e 1 (equação 6.1) e em uma faixa entre 0 e 100 (equação 6.2). A segunda normalização é para visualização dos resultados do treinamento via relatórios e gráficos.

$$ValorNormal = 2 * \left(\frac{x_{ij} - x_{min}}{x_{max} - x_{min}} \right) - 1 \quad (6.1)$$

onde, x_{min} - Valor mínimo encontrado em x . x_{max} - Valor máximo encontrado em x .

$$ValorNormalPerc = \left(\frac{x_{ij} - x_{min}}{x_{max} - x_{min}} \right) * 100 \quad (6.2)$$

onde, x_{min} - Valor mínimo encontrado em x . x_{max} - Valor máximo encontrado em x .

6.3.1.3 Número de saídas

O número de saídas para cada município foi determinado através da equação 6.3.

$$N_s = \sqrt{N_q} \quad (6.3)$$

Onde N_q é número de quadrículas com curvas distintas após a normalização.

6.3.1.4 Inicialização dos pesos

O treinamento é fortemente dependente da inicialização dos pesos. Assim, se a inicialização for feita com valores que estão distantes dos padrões de entrada, estes nós não serão atualizados. Para evitar esse problema, foram selecionados valores aleatórios dos dados de entrada, conforme proposto por Silva (SILVA, 2003). Os dados de entrada que compõem o escopo de seleção são as curvas distintas resultantes da normalização. O processo de inicialização dos pesos ocorre apenas na fase de agrupamento.

6.3.1.5 Número de iterações

O limite máximo de iterações consideradas foi de 6.000, onde nas 1000 primeiras iterações ocorre a fase de ordenação e nas restantes o ajuste fino. Para o processo de classificação dos dados é necessária apenas uma iteração, sendo o restante utilizada para o processo de agrupamento.

6.3.1.6 Taxa de aprendizagem e raio de vizinhança

Foram realizados vários testes com estes parâmetros e verificou-se que os melhores resultados encontrados estavam de acordo com os apresentados em (KOHONEN, 1997).

Assim a taxa de aprendizado foi inicializada com o valor 0,9 e o raio de vizinhança foi calculado da seguinte forma:

$$R_c = \left(\frac{N_c}{2} \right) - 1 \quad (6.4)$$

Onde N_c é o número de saídas selecionadas para o treinamento.

6.3.2 K-Médias

Este método possui cinco parâmetros de inicialização para seu treinamento:

1. Conjunto de entrada;
2. Normalização;
3. Número de centróides;
4. Inicialização dos centróides;
5. Critério de parada.

Os parâmetros conjunto de entrada, normalização, número de centróides e inicialização dos centróides são inicializados como no método Kohonen em anel, descrito anteriormente.

No K-Médias, não há a necessidade de definir o número de iterações, mas sim um critério de parada, o qual determina a distância máxima entre os centróides da iteração atual e da anterior.

6.3.3 Fuzzy C-Médias

O Fuzzy C-Médias, que é um método fuzzy, possui um parâmetro denominado índice de fuzzificação (descrito no Capítulo 3, seção 3.4), o qual controla o grau de fuzzificação do treinamento. Os demais parâmetros de inicialização deste método são os mesmos descritos no K-Médias.

As simulações realizadas utilizaram um índice de fuzzificação variando em 1.25, 1.5 e 2, buscando o valor que produzisse o melhor resultado.

6.4 Estudo de Caso 1 - Anápolis

Localizada a 54 quilômetros de Goiânia, Anápolis é considerada a capital industrial de Goiás com a criação do Distrito Agroindustrial de Anápolis - DAIA. É o terceiro maior município do estado em população e o segundo no ranking de competitividade e desenvolvimento, e compõe a região mais desenvolvida do Centro-Oeste.

A Tabela 6.1 mostra o consumo e o número de clientes por classe no município de Anápolis.

Tabela 6.1: Consumo e número de consumidores por classe - Anápolis 2006 (CELG, 2007)

Classe de consumidores	Consumo (GWh)	Consumo (%)	Número de consumidores	Consumidores (%)
Residencial	157,17	32,89%	102.926	88,14%
Industrial	170,07	35,59%	748	0,64%
Comercial	65,62	13,73%	10.392	8,90%
Rural	12,85	2,69%	2.263	1,94%
Poder Público	21,2	4,44%	385	0,33%
Iluminação	29,14	6,10%	-	-
Serviço Público	21,71	4,54%	47	0,04%
Próprio	0,13	0,03%	7	0,01%
Total	477,89	100%	116.774	100%

A classe industrial detém o maior consumo, seguida da residencial, na qual se encontram 88,14% dos consumidores do município.

A seguir, são apresentadas as aplicações dos métodos de agrupamento e classificação sobre o município de Anápolis, onde pode ser constatado, como produto final, o perfil de distribuição dos clientes.

Para todos os métodos são apresentadas as curvas dos centróides dos treinamentos. Estas curvas mostram a porcentagem de consumo por classe, sendo que o somatório das amplitudes

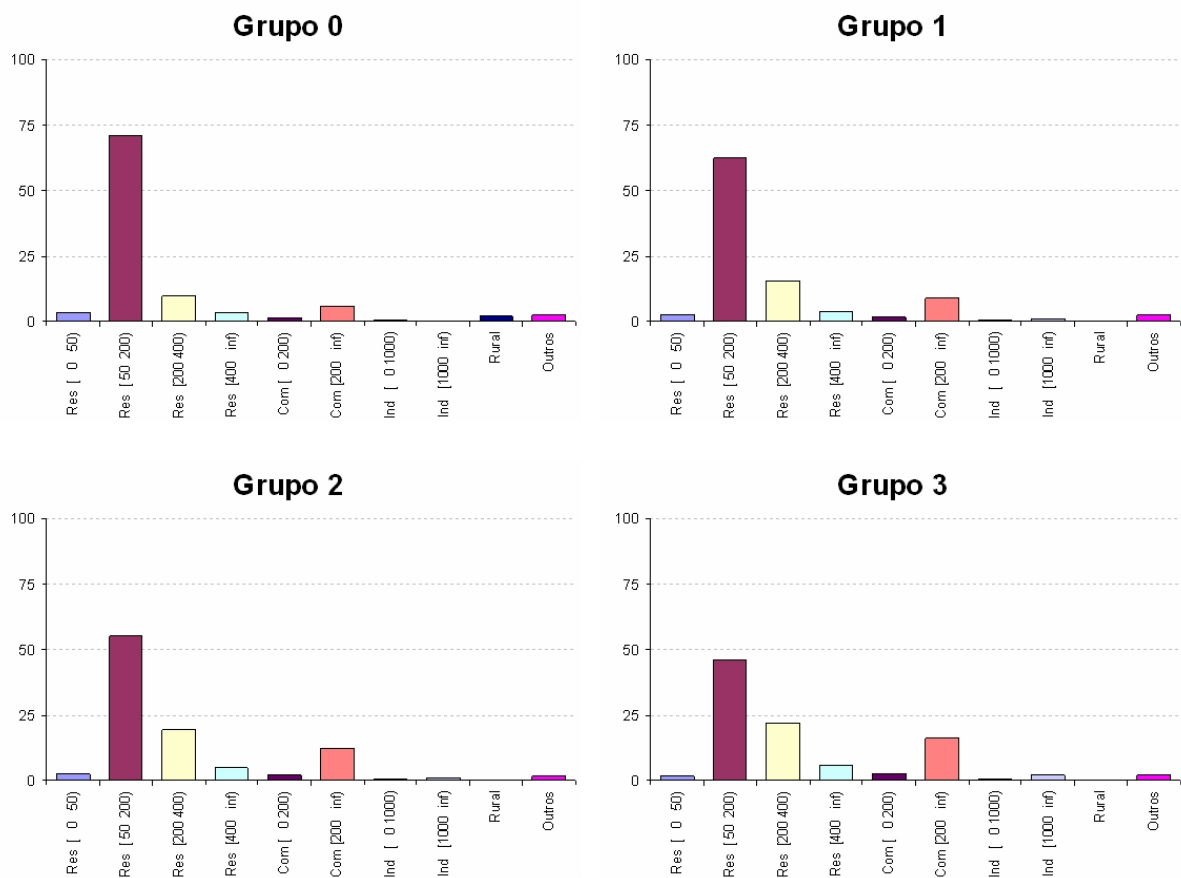
de cada faixa é igual a 100%.

6.4.1 Kohonen em anel

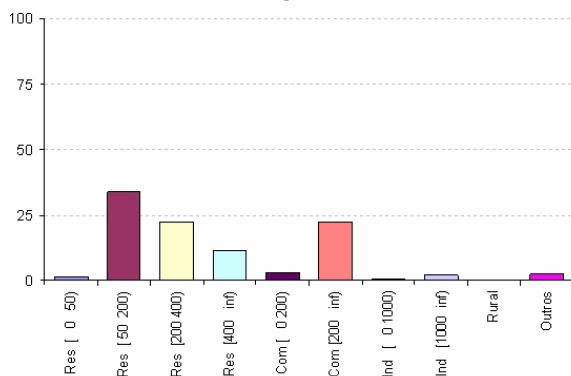
Este método, diferente dos demais, possui um vetor de atualização dos pesos, o qual possui um valor inicial alto que é diminuído durante o treinamento. Isso provoca uma relação de vizinhança entre os nós do anel, onde nós vizinhos possuem perfis semelhantes.

Essa vizinhança pode ser vista no gráfico 6.2, onde nós vizinhos possuem curvas com perfil próximo variando apenas a porcentagem das classes que compõem estes perfis. Como a topologia empregada é do tipo em anel, nota-se que o perfil do grupo 19 é próximo ao perfil do grupo 0, caracterizando a topologia empregada.

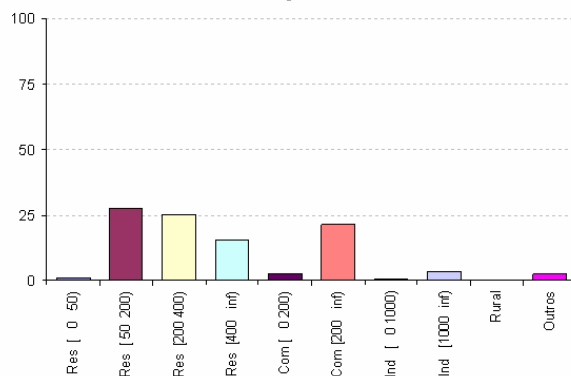
Apesar deste método produzir centróides menos distribuídos, ele permite identificar, como produto final da classificação, quadrículas que possuem o perfil semelhante, mas com uma composição percentual ligeiramente diferente.



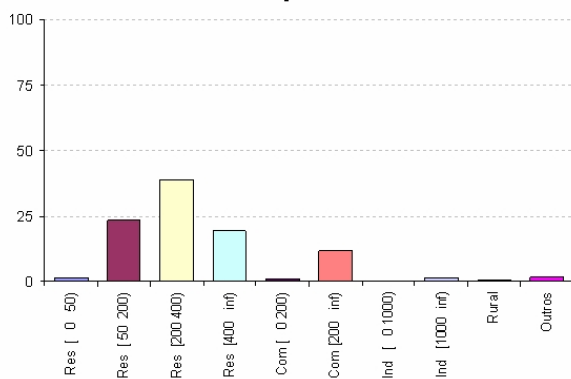
Grupo 4



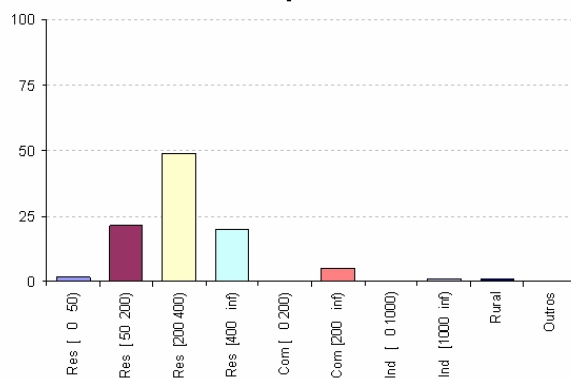
Grupo 5



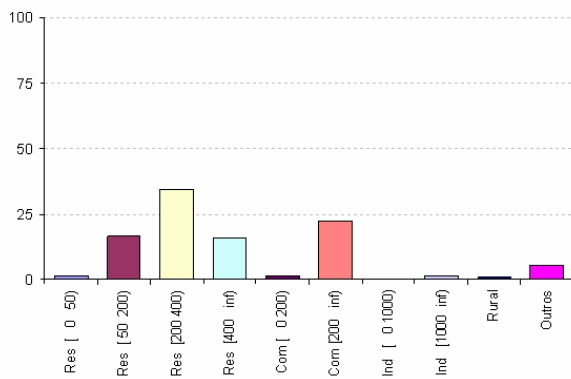
Grupo 6



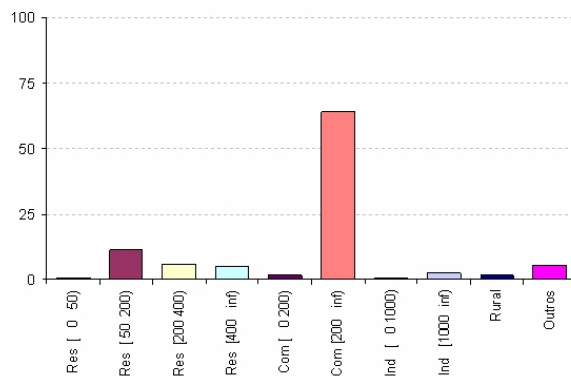
Grupo 7



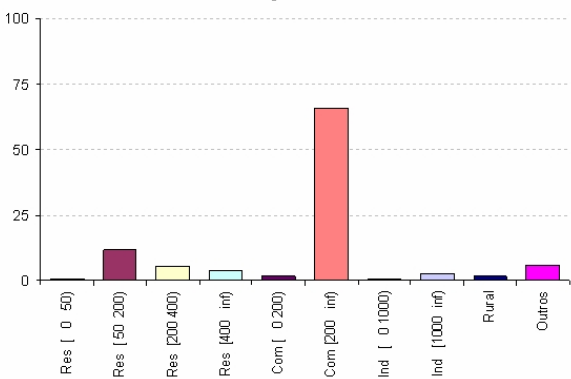
Grupo 8



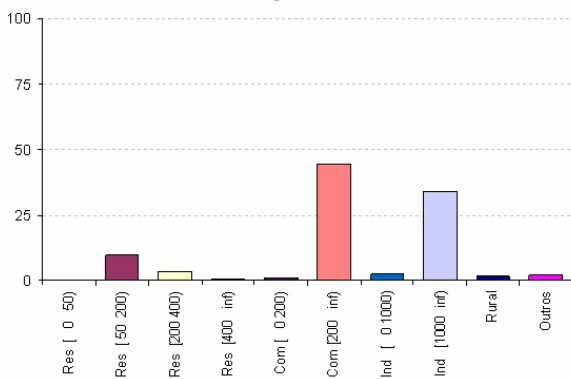
Grupo 9



Grupo 10



Grupo 11



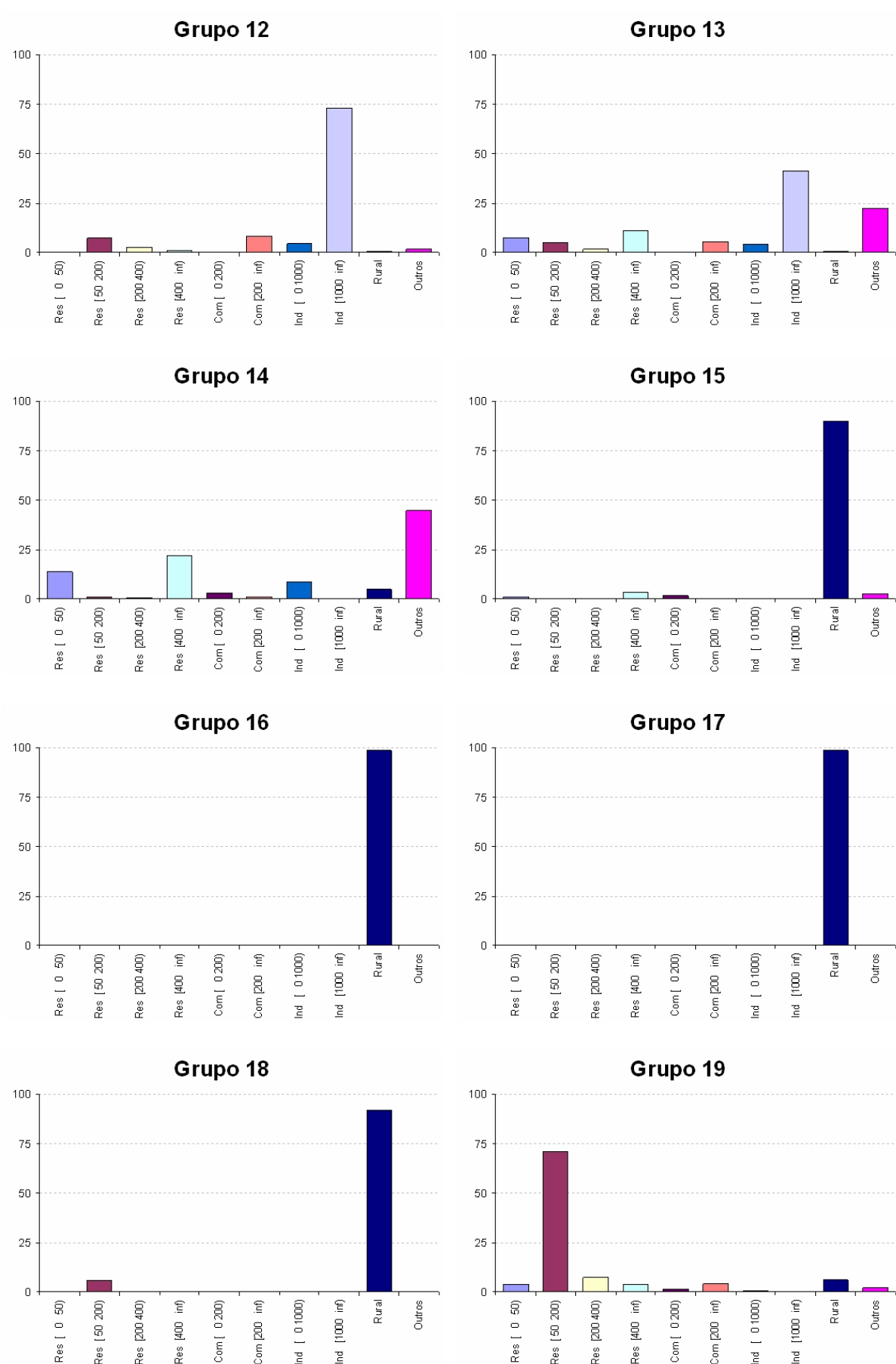


Figura 6.2: Gráficos dos centróides do treinamento pelo método Kohonen em Anel - Anápolis

Conforme já mencionado, o processo de classificação não atualiza os centróides provenientes do agrupamento. Assim, devido ao processo de atualização feito pelo vetor de pesos, poderão existir grupos que no processo de classificação ficarão sem elementos. Isso ocorre porque o centróide gerado pela ação do vetor de pesos pode se distanciar das amostras do treinamento.

Isso pode ser confirmado pela Tabela 6.2, que mostra a relação final do processo de classificação, onde os grupos 11 e 13 ficaram sem quadrículas. A última linha da tabela apresenta as quadrículas do município que possuem consumo igual zero para o período do cenário e portanto não entram no agrupamento e classificação.

Na Figura 6.3 é mostrado o resultado da classificação pelo método Kohonen em Anel. Através da legenda de cores (Figura 6.1), vê-se a disposição das quadrículas do grupo 12. Estes grupos compõem as quadrículas com clientes industriais que possuem um consumo médio acima de 1000 kWh/ano. Observa-se que estas quadrículas concentram-se na região sul do perímetro urbano da cidade de Anápolis.

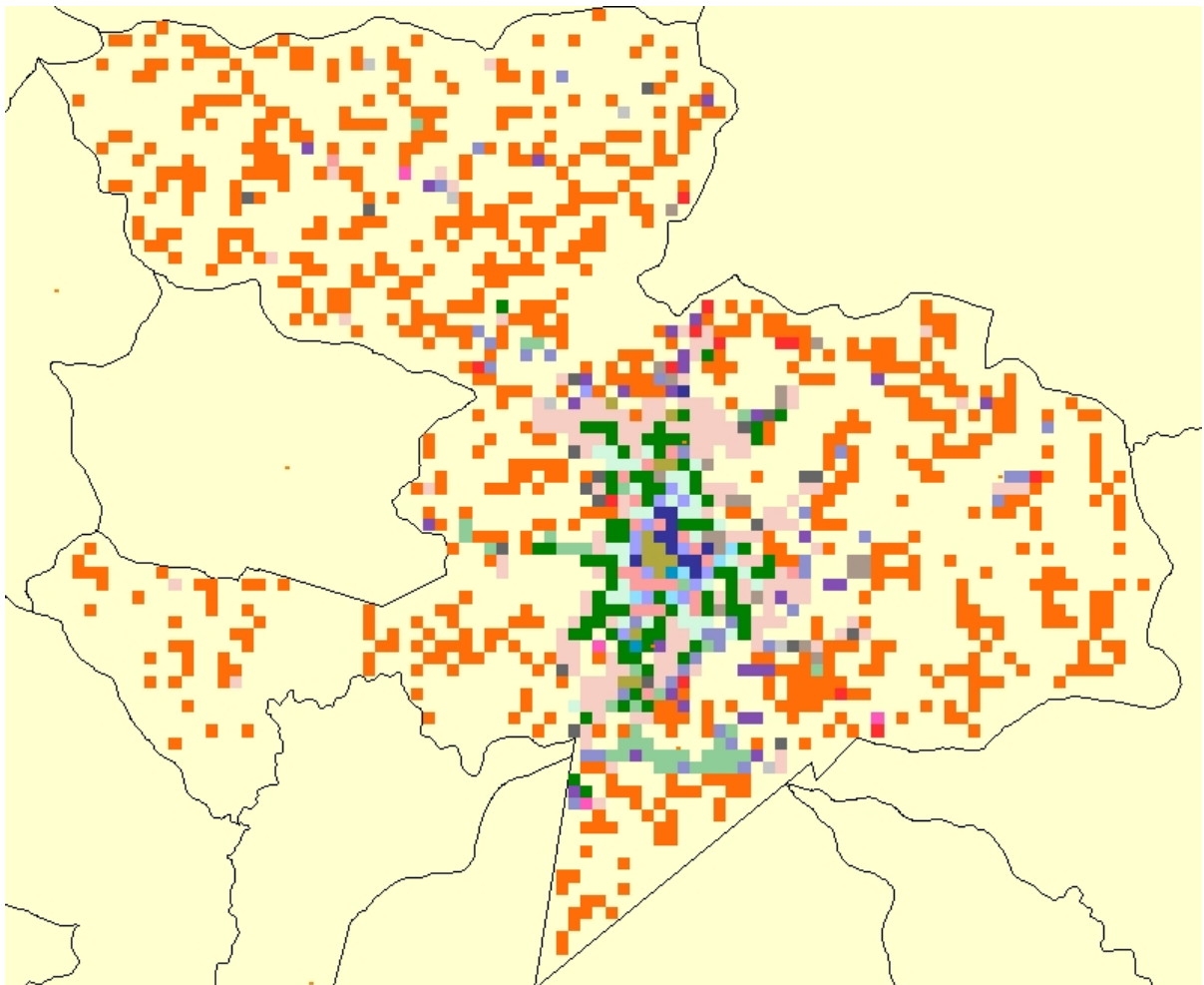


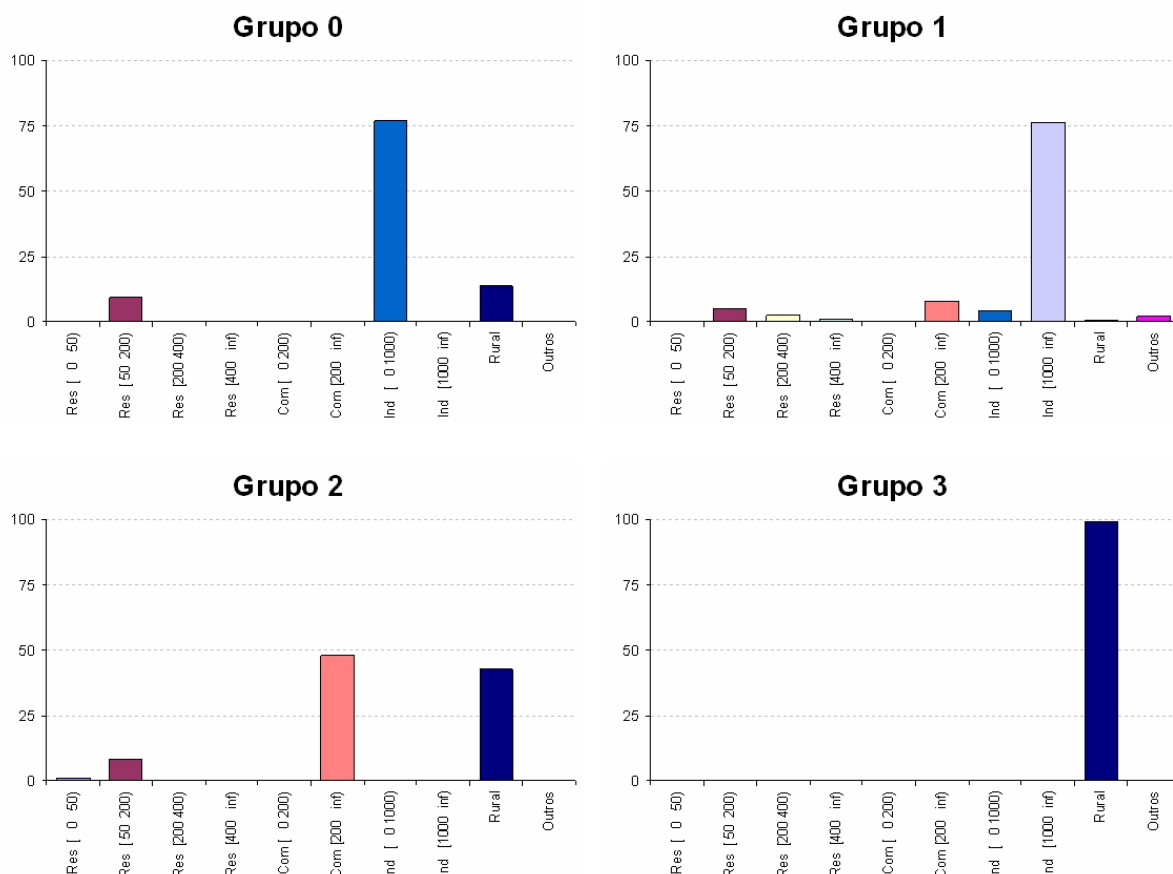
Figura 6.3: Tela do resultado da classificação pelo método Kohonen em Anel

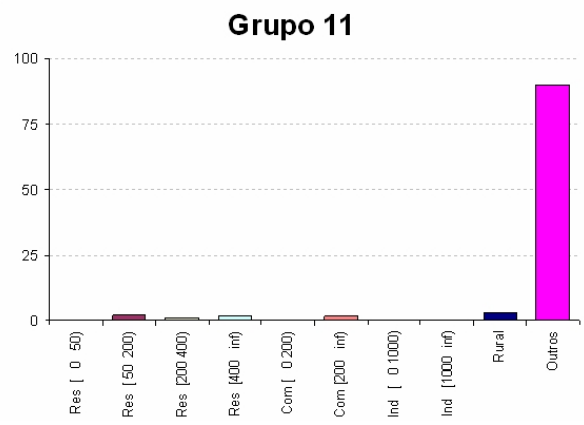
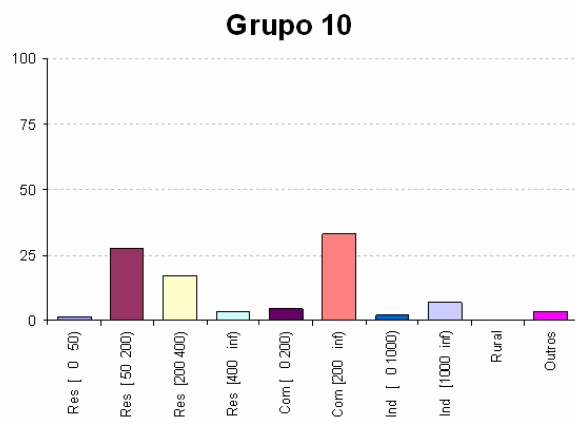
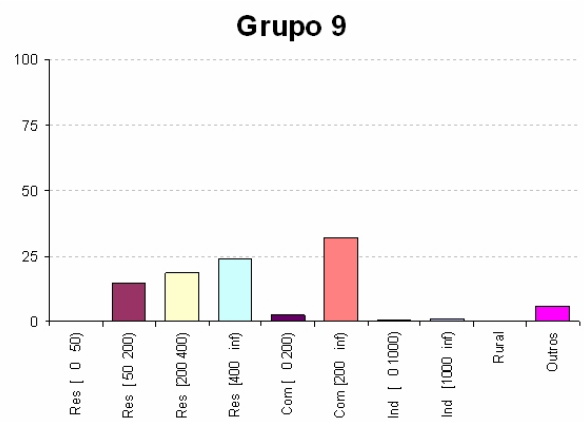
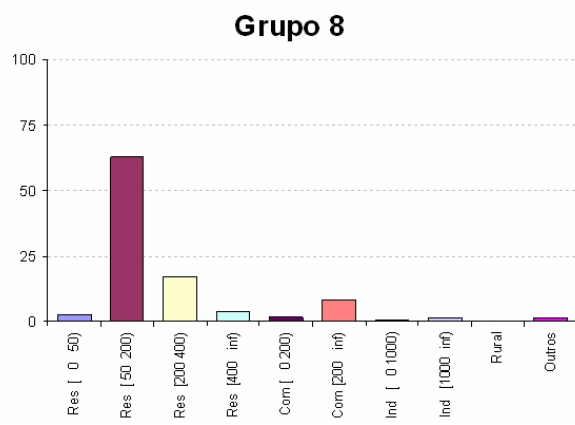
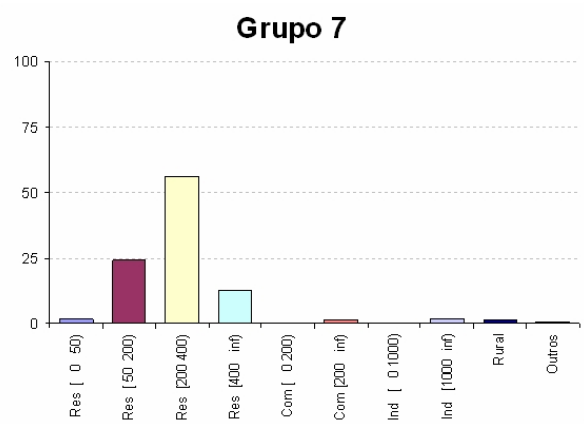
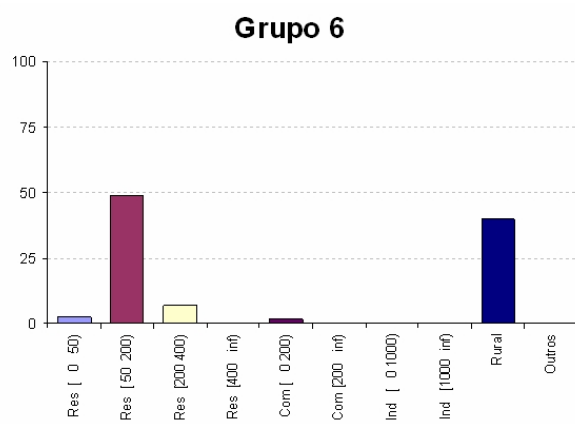
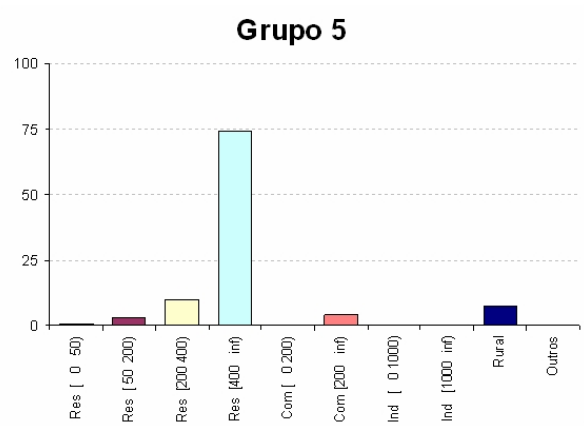
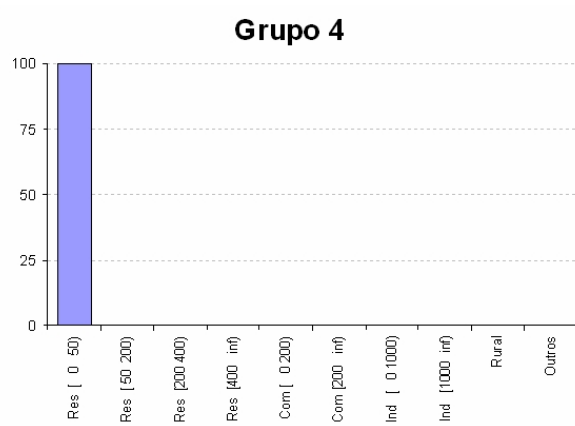
6.4.2 K-Médias

Na técnica K-Médias, como mostra a Figura 6.4, não há uma relação de vizinhança entre os centróides dos grupos gerados como ocorre no Kohonen Anel, ocasionando grupos com perfis mais distribuídos.

Esse método gerou centróides com uma composição não encontradas nas curvas geradas pelo método Kohonen Anel. São exemplos as curvas dos grupos 4 e 5. Isso possibilita identificar algumas áreas de forma mais precisa.

Como o método K-Médias realiza a atualização dos centróides dos grupos pelo média das curvas presentes nos grupos, ela é mais dependente da inicialização que o método Kohonen Anel. Essa média generaliza um pouco mais o centróide de alguns grupos. Entretanto, caso a inicialização seja satisfatória, melhores grupos são formados, diminuindo assim a generalização da classificação. Foi o que ocorreu com os grupos 4 e 5, por exemplo.





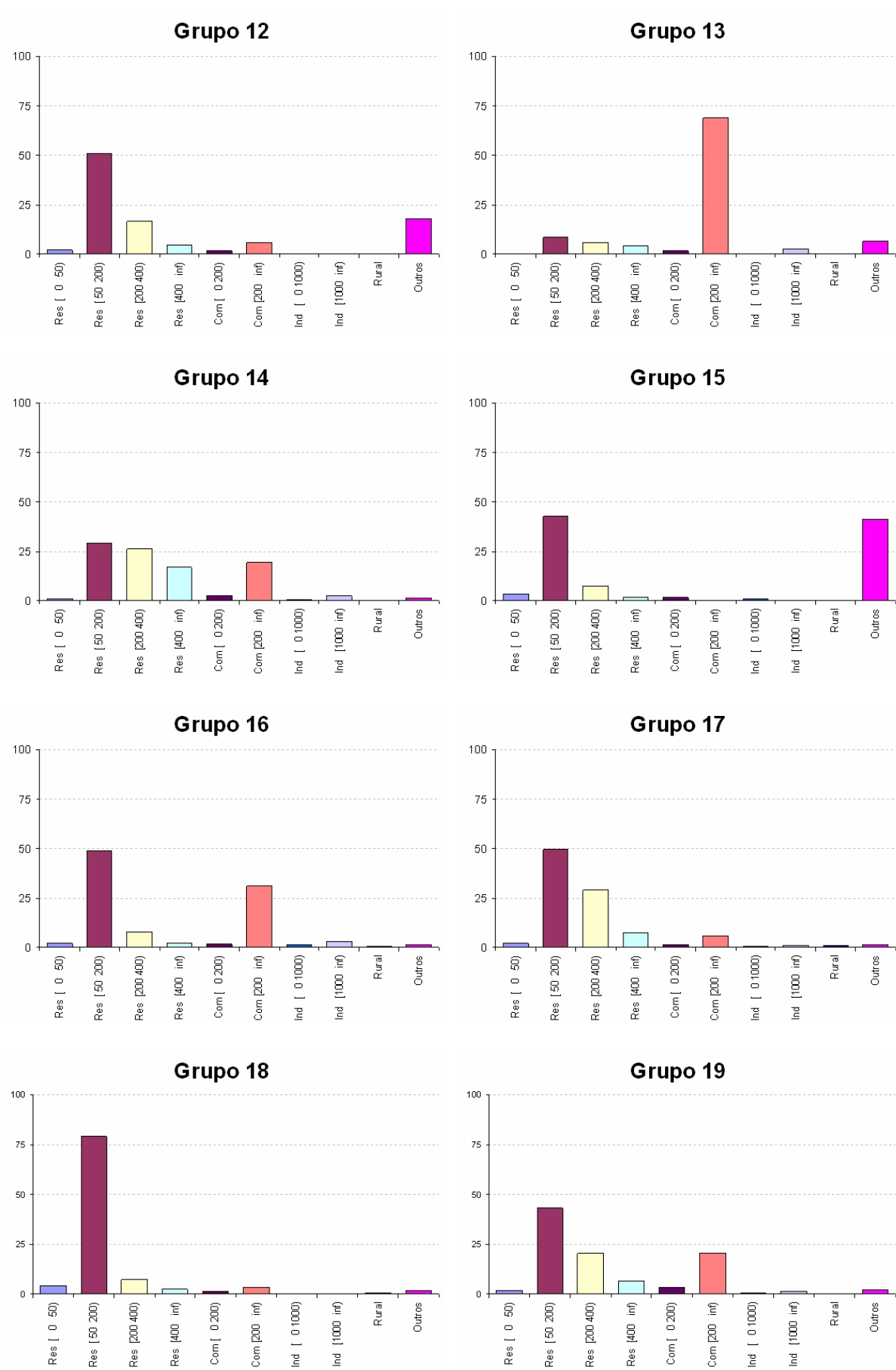


Figura 6.4: Gráficos dos centróides do treinamento pelo método K-Médias - Anápolis

Um outro exemplo do resultado da diferença de atualização dos centróides é que este método gerou apenas um grupo expressivamente rural (grupo 3) enquanto o método Kohonen gerou quatro (grupos 15, 16, 17 e 18) devido a ação do vetor de atualização.

As quadrículas pertencentes aos grupos 0 e 1 compõem as áreas com presença de indústrias no município, totalizando 45 quadrículas (Tabela 6.2), sendo que destas 5 são referentes a indústrias com consumo anual de 0 a 1000 kWh.

Uma boa parte das quadrículas estão presentes no grupo 18, o qual tem perfil predominantemente residencial, confirmando assim os dados fornecidos na Tabela 6.1 que mostram que 32,89% do consumo e 88,14% dos consumidores são residenciais. Há uma predominância no grupo de consumidores com média de consumo de 50 a 200 kWh.

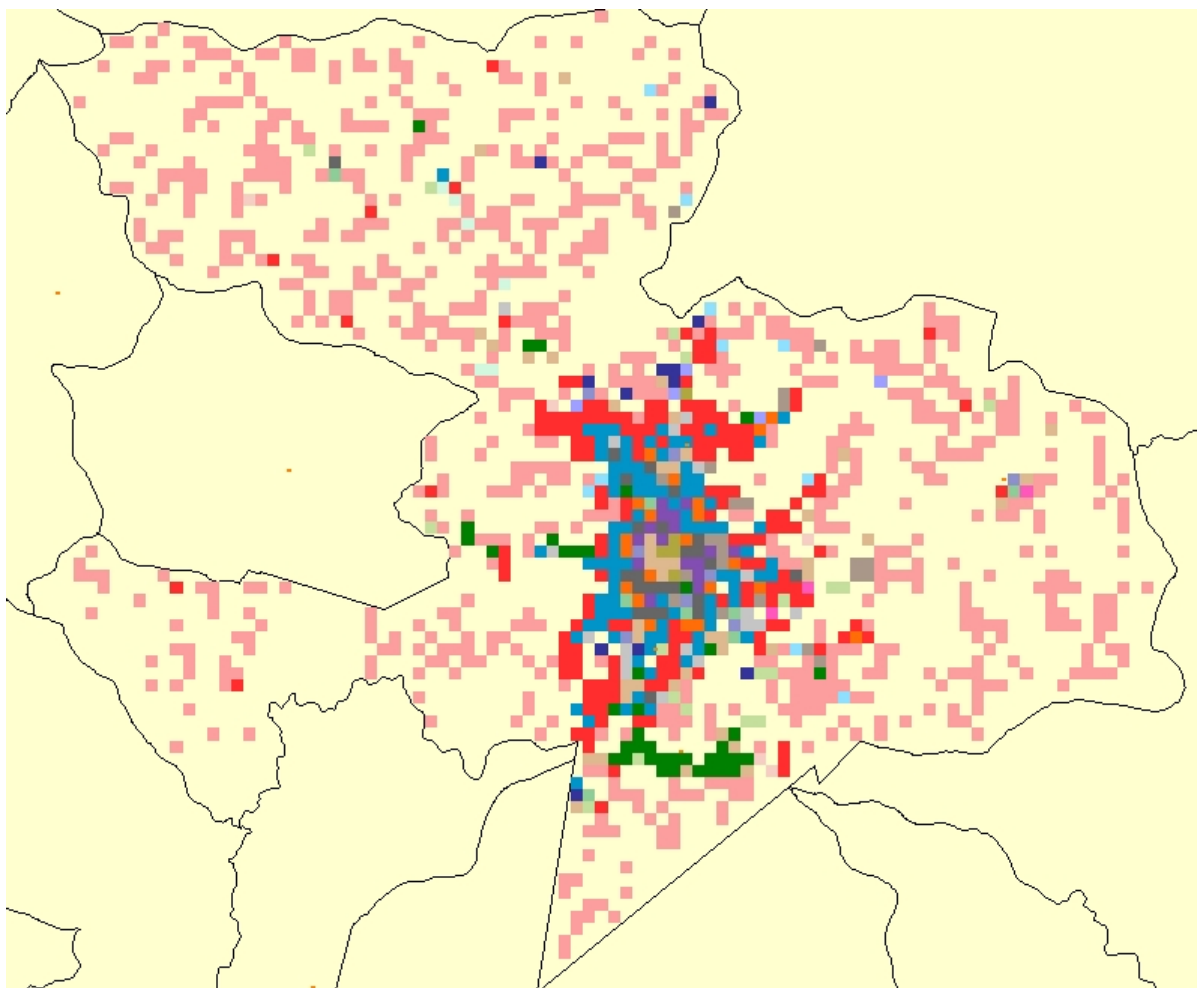
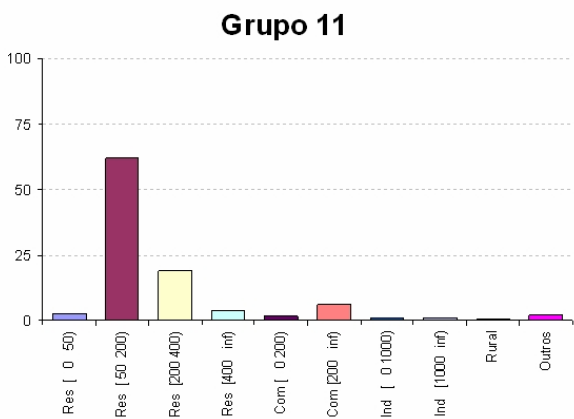
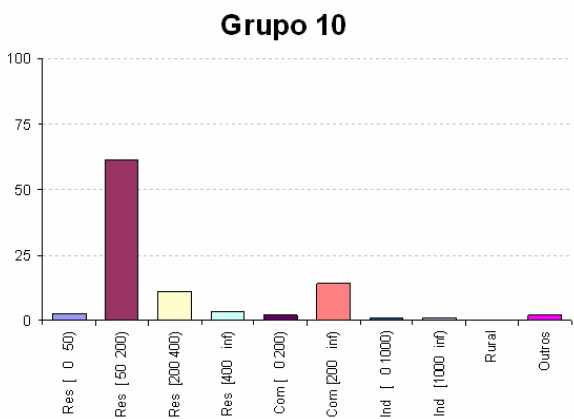
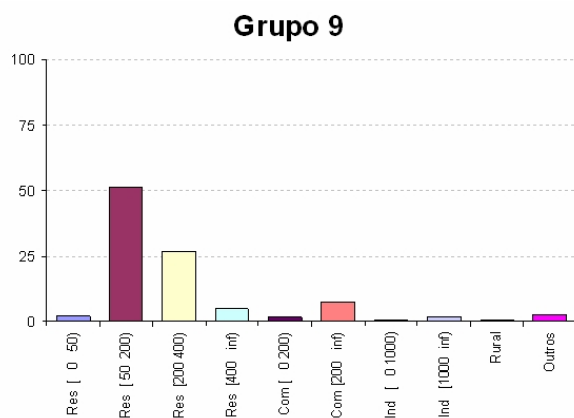
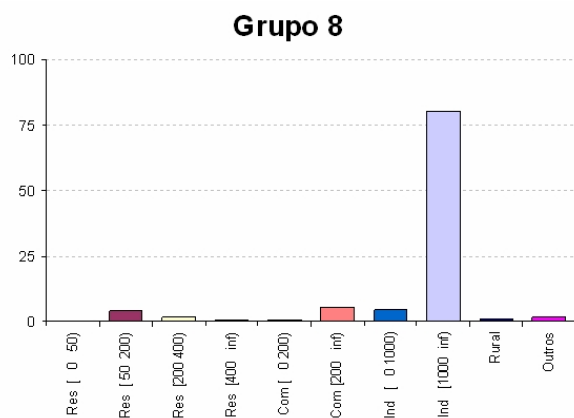
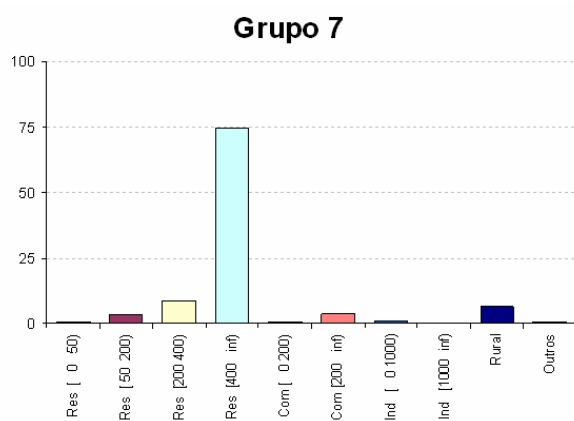
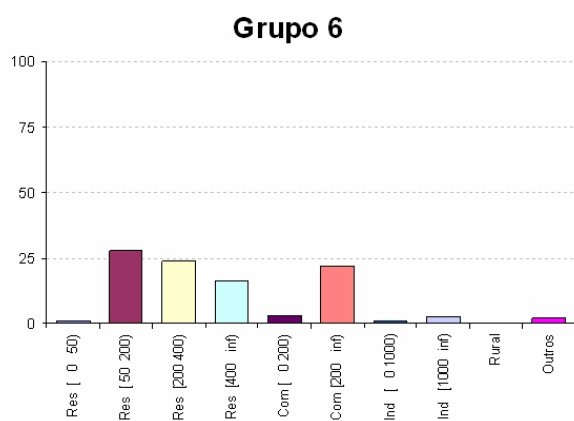
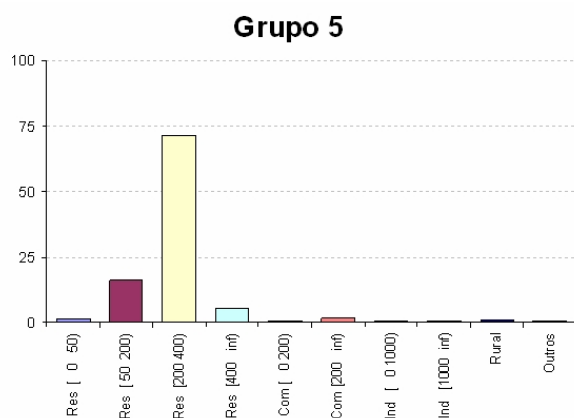
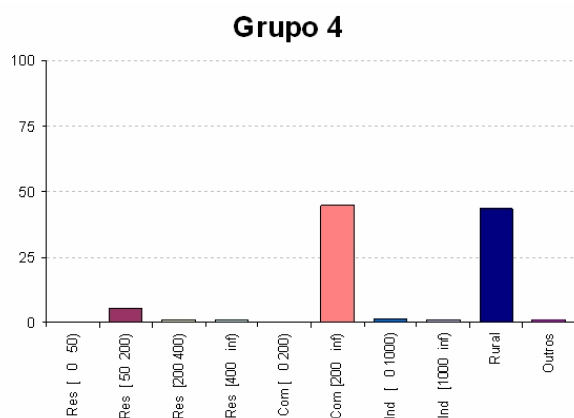


Figura 6.5: Tela do resultado da classificação pelo método K-Médias



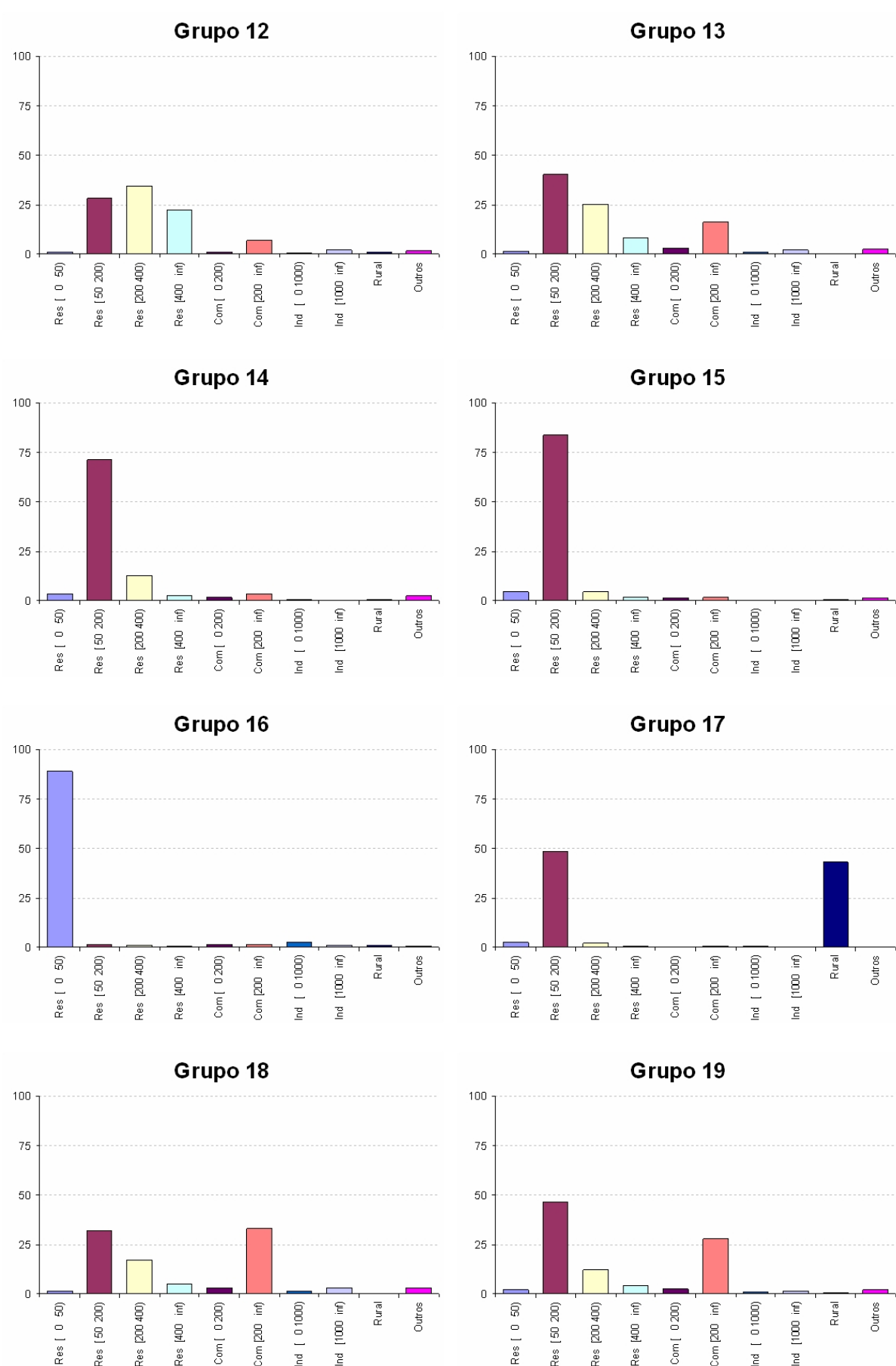


Figura 6.6: Gráficos dos centróides do treinamento pelo método Fuzzy C-Médias - Anápolis

Neste método houve uma melhor distribuição das quadrículas entre os grupos (Tabela 6.2). Isso se deve a uma discrepância maior entre os centróides gerados em comparação ao método K-Médias. Observe que, igualmente ao K-Médias, não houve neste método a ocorrência de grupos sem quadrículas como no Kohonen em Anel.

Essa melhor distribuição dos centróides pode ser explicado pela forma de atualização dos mesmos. O centróide V_i calculado pela equação 3.9, visto como o centro do *cluster* A_i , é a média ponderada dos dados em A_i .

Essa forma de cálculo do centróide diminui a influência da inicialização e melhora o processo de classificação por levar em consideração a pertinência de seus elementos aos demais centróides.

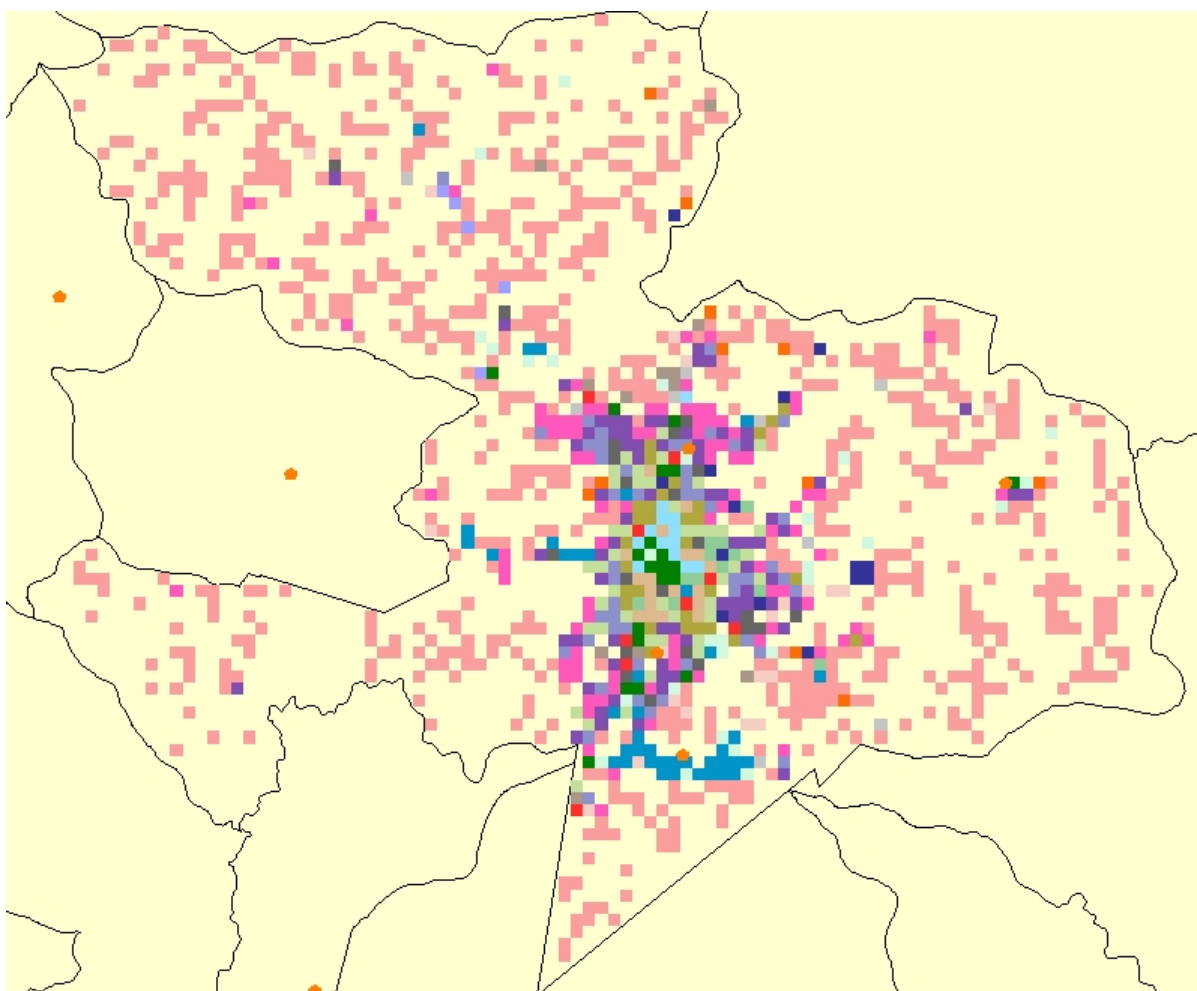


Figura 6.7: Tela do resultado da classificação pelo método Fuzzy C-Médias

Na Figura 6.7, proveniente da classificação pelo método Fuzzy C-Médias, nota-se claramente uma maior distribuição das cores no perímetro urbano que nos métodos anteriores. Já a área rural, composta pelo grupo 3, apresentou praticamente a mesma composição de quadrículas

dos métodos anteriores.

6.4.4 Distribuição das classificações

Nesta seção é apresentada a Tabela 6.2 que contém o resultado das classificações dos métodos citados anteriormente.

Tabela 6.2: Relação de quadrículas por grupo - Anápolis 2006

Grupos	Número de Quadrículas		
	Kohonen	K-Médias	Fuzzy
0	163	5	20
1	83	40	20
2	38	6	28
3	23	779	778
4	14	5	5
5	12	12	14
6	5	10	13
7	21	22	12
8	2	89	40
9	14	5	30
10	37	8	48
11	-	18	43
12	41	8	12
13	-	44	18
14	31	16	72
15	4	3	93
16	6	18	9
17	770	21	11
18	13	164	9
19	20	24	22
-	2234	2234	2234

O processo de treinamento começa com a inicialização dos centróides. Essa inicialização é realizada por meio da escolha aleatória de curvas do conjunto de treinamento. Assim, por exemplo, um mesmo grupo recebe perfis de curvas iniciais diferentes, ocasionando que, para

treinamentos distintos, um conjunto de quadrículas pode receber um número de grupo diferente. Entretanto, seu perfil de classificação continuará o mesmo, mudando apenas o número do grupo.

Portanto, a Tabela 6.2 deve ser analisada em conjunto com os gráficos dos grupos de cada método apresentado. Por exemplo, utilizando os gráficos, nota-se que a maior parte da área rural de Anápolis é identificada pelos grupos 17 (Kohonen Anel), 3 (K-Médias) e 3 (Fuzzy C-Médias).

6.4.5 Medidas do agrupamento

Nesta seção, serão apresentados os resultados dos índices de medição do agrupamento e classificação.

Primeiramente, na Tabela 6.3 é mostrada a relação do tempo de processamento das classificações efetuadas no município de Anápolis.

Conforme apresentado, o método K-Médias obteve o menor tempo computacional na execução da classificação. Esse tempo foi muito inferior ao tempo gasto pelo método Kohonen em Anel, o qual executou 6000 iterações, sendo destas 1000 para o ajuste fino da rede. Já o tempo medido no método Fuzzy C-Médias é apenas o relativo a duração do processo de agrupamento, visto que para a classificação é necessário uma análise do índice de defuzzificação a ser aplicado.

Os estudos de caso foram realizados em um servidor IBM X-Series 226, com dois processadores (Intel Xeon) de 3.2 GHz, 4 GB de memória RAM, 5 discos de 70 GB com 10.000 rpm configurados em RAID 5 e com sistema operacional Windows 2000 Service Pack 4.

Tabela 6.3: Tempo de processamento dos métodos - Anápolis

Método	Tempo
Kohonen Anel	1 h 56 min 04 seg
K-Médias	1 min 38 seg
Fuzzy C-Médias	4 min 43 seg

A Tabela 6.4 mostra o resultado de medição dos Índices Intra-Grupo (IAG), o qual indica o quanto os padrões de entrada que estão em um mesmo grupo são similares.

O método que obteve o melhor resultado foi o método Fuzzy C-Médias seguido de perto pelo método K-Médias. Isso está relacionado à boa distribuição das quadrículas pelos grupos. O método Kohonen em Anel obteve o pior resultado, já que sua distribuição não seguiu os demais métodos, tanto que dois dos seus grupos ficaram sem elementos (Tabela 6.2).

Tabela 6.4: Resultados dos Índices Intra-Grupo (IAG) - Anápolis

Método	IAG
Kohonen Anel	565
K-Médias	465
Fuzzy C-Médias	426

Por fim, a Tabela 6.5 apresenta os resultados do Índice Inter-Grupo (IEG), o qual mede o quão distantes os grupos estão de uma curva referencial.

Novamente o método Fuzzy C-Médias obteve um melhor resultado que os demais métodos. Entretanto, este índice não permite afirmar que quanto menor o valor, melhor é o resultado, mas apenas que se ele tiver um valor alto os grupos estão mais separados ou se ele tiver um valor baixo significa que há uma semelhança maior entre os padrões de grupos diferentes.

Tabela 6.5: Resultados dos Índices Inter-Grupo (IEG) - Anápolis

Método	IEG
Kohonen Anel	819
K-Médias	795
Fuzzy	790

6.5 Estudo de Caso 2 - Aparecida de Goiânia

Aparecida de Goiânia é um município brasileiro do estado de Goiás que faz divisa com o município de Goiânia e sua população estimada em 2005 era de 435.323 habitantes (IBGE/CIDADES, 2007).

Aparecida de Goiânia é o segundo maior colégio eleitoral do estado e sua Área é de $288km^2$ representando 0,0848% do Estado. Após sua emancipação em 1963 Aparecida de Goiânia passou a ser o alvo de inúmeros assentamentos promovidos principalmente pelo governo do estado, o que a impulsionou na classificação de um dos maiores índices de crescimento populacional do Brasil.

Em seus aspectos econômicos, a pecuária, com a criação de gado bovino de corte e leite é uma das atividades na sua pequena extensão rural.

No município onde predomina a indústria extrativa de areia, pedra e barro para construções, a agricultura não é expressiva dentro de uma pequena área territorial rural, visto que 70% do

seu território encontra-se hoje ocupado por grande proliferação imobiliária, cujos lotes e áreas diversas estão ocupadas por moradias e setores industriais.

Segue abaixo a Tabela 6.6 que mostra o consumo e o número de clientes por classe no município de Aparecida de Goiânia.

Tabela 6.6: Consumo e número de consumidores por classe - Aparecida 2006 (CELG, 2007)

Classe de consumidores	Consumo (GWh)	Consumo (%)	Número de consumidores	Consumidores (%)
Residencial	185,21	58,30%	127.034	93,20%
Industrial	57,33	18,05%	928	0,68%
Comercial	44,05	13,87%	7.842	5,75%
Rural	0,82	0,26%	149	0,11%
Poder Público	8,49	2,67%	307	0,23%
Iluminação	17,90	5,63%	-	-
Serviço Público	3,86	1,22%	46	0,03%
Próprio	0,02	0,01%	1	0%
Total	317,68	100%	136.308	100%

As classes residencial e Comercial são responsáveis por 72,17% do consumo de energia do município e essa ocupação urbana será constatada na aplicação dos métodos de agrupamento e classificação a seguir.

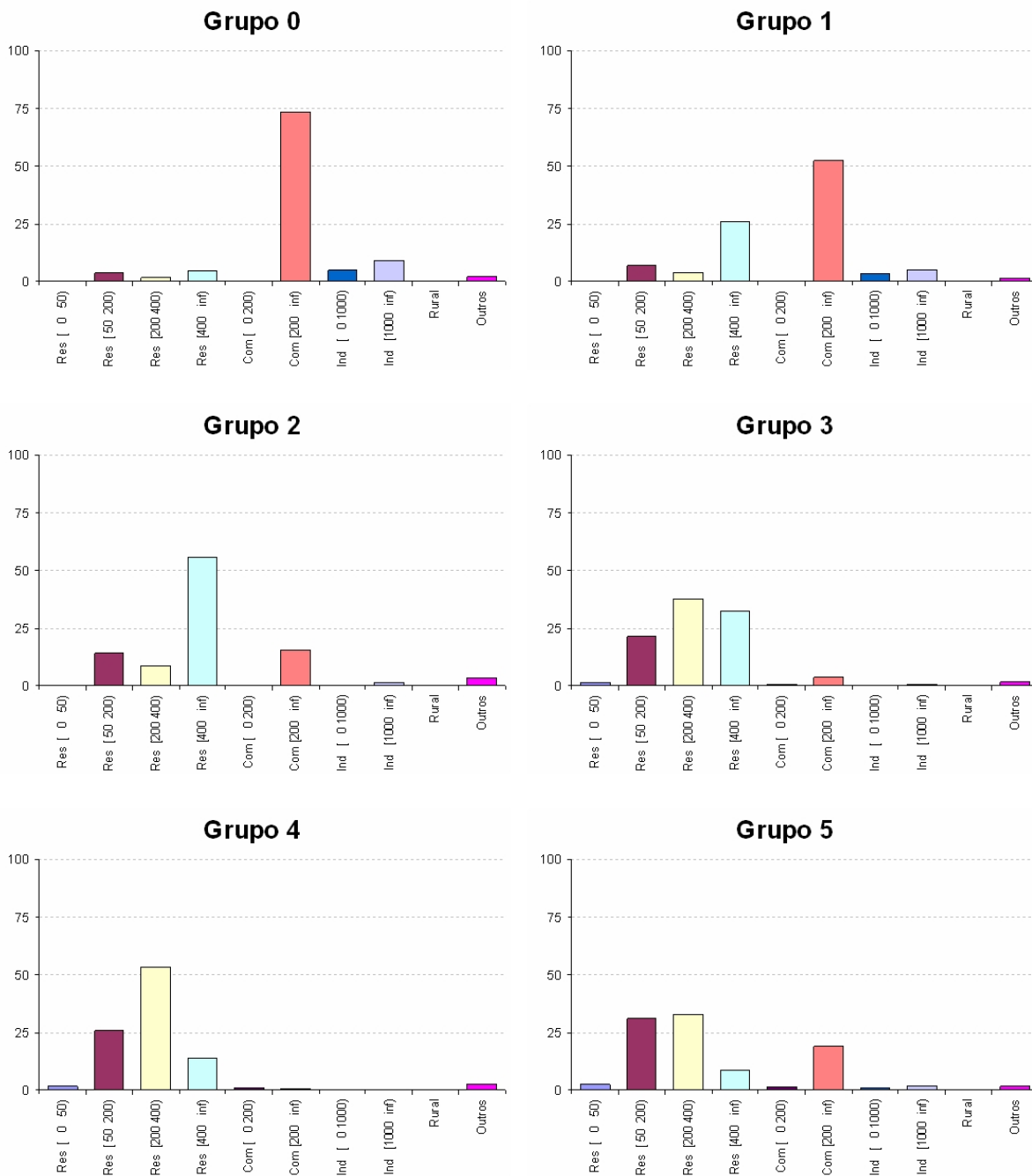
6.5.1 Kohonen em anel

Este método apresentou uma maior distribuição dos centróides para o município de Aparecida de Goiânia do que para Anápolis, ocasionando uma melhor distribuição das quadrículas (Tabela 6.7). Isso ocorreu porque 80% das curvas de carga deste município são distintas, o que privilegia este método, pois ele produz uma faixa de centróides que distribui melhor as curvas com perfis próximos.

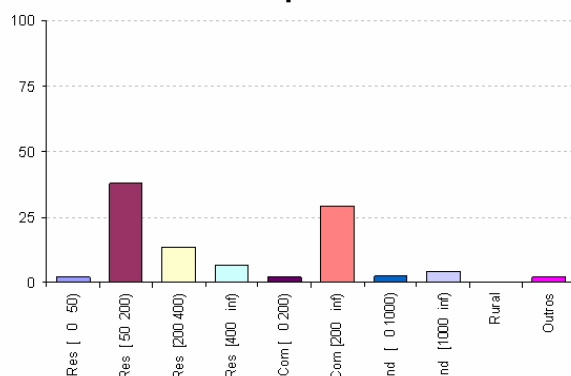
Foi fornecido como parâmetro de entrada a formação de vinte e três grupos devido ao número de curvas distintas na normalização (equação 6.8).

Os grupos formados que ficaram sem quadrículas possuem curvas semelhantes nos demais centróides, o que prejudicou menos o processo de classificação. Houve também poucos grupos com uma quantidade inferior a cinco quadrículas, aumentando assim a escalabilidade do processo.

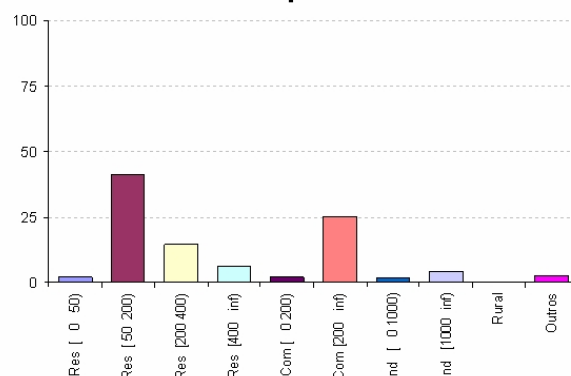
Como este método gera centróides que podem se distanciar do conjunto de treinamento, conseqüentemente ele apresenta composições de perfis que não são encontrados no conjunto de treinamento. Assim uma composição de consumo rural e industrial (0 a 1000 kWh), grupo 20, não foi identificado no processo de classificação das quadrículas.



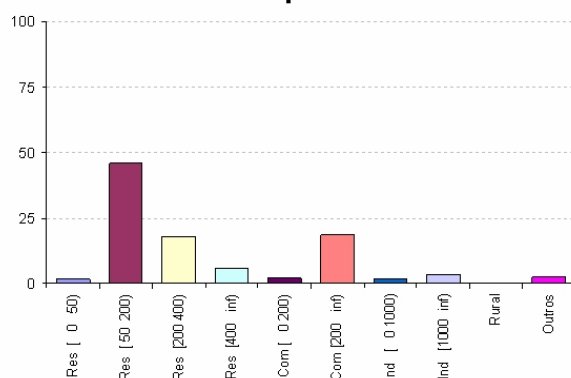
Grupo 6



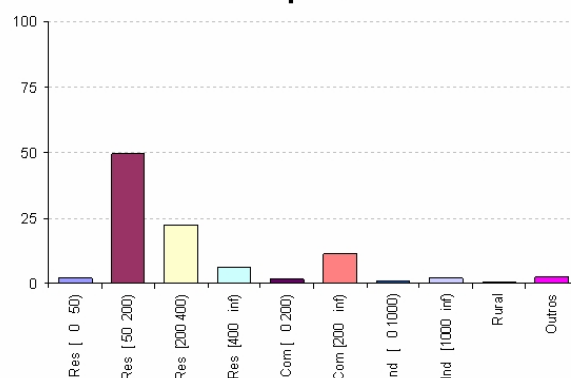
Grupo 7



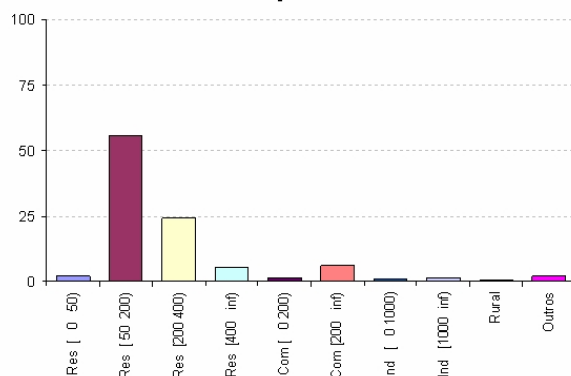
Grupo 8



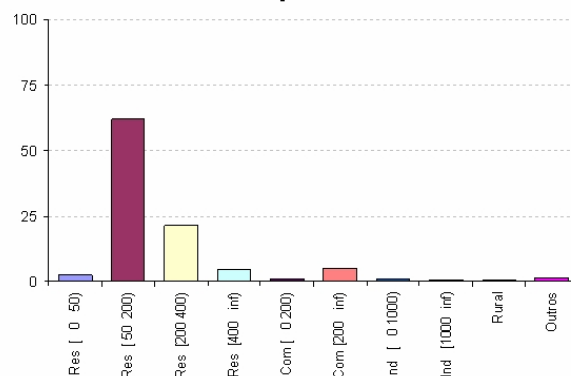
Grupo 9



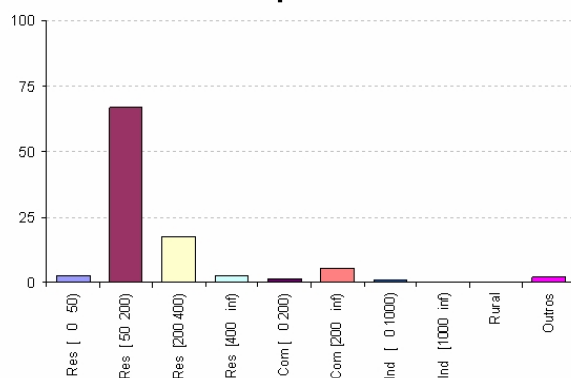
Grupo 10



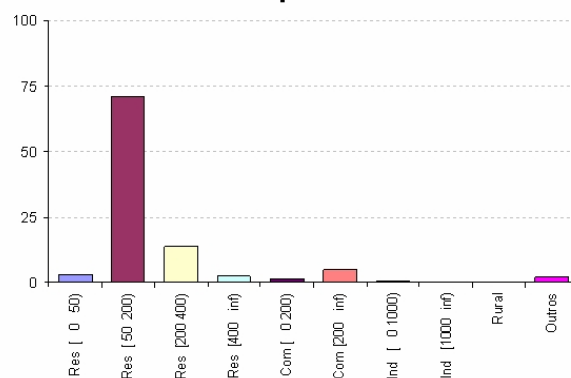
Grupo 11



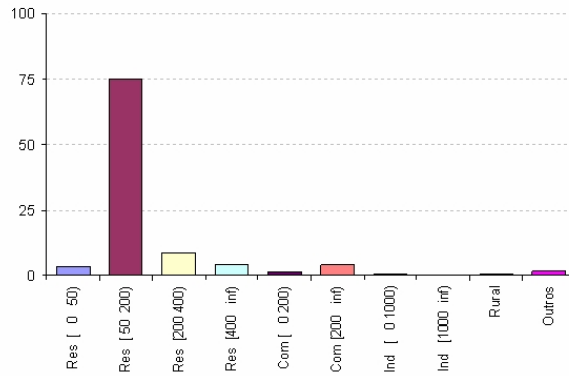
Grupo 12



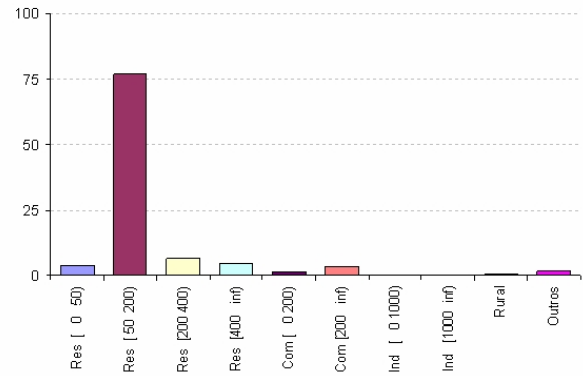
Grupo 13



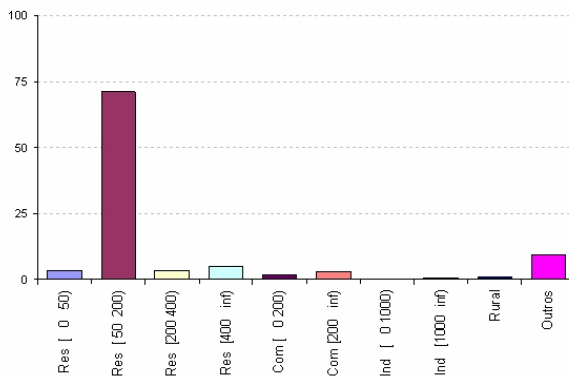
Grupo 14



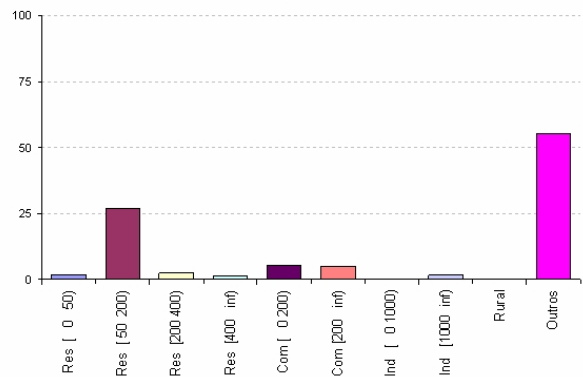
Grupo 15



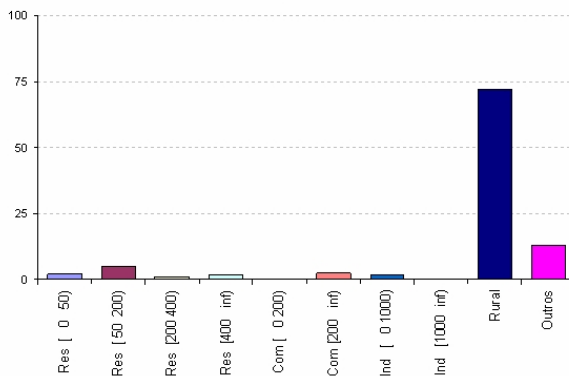
Grupo 16



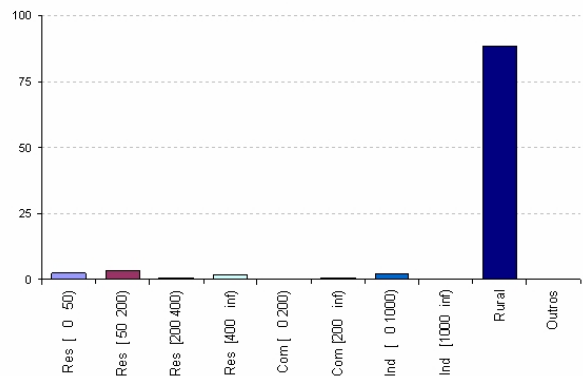
Grupo 17



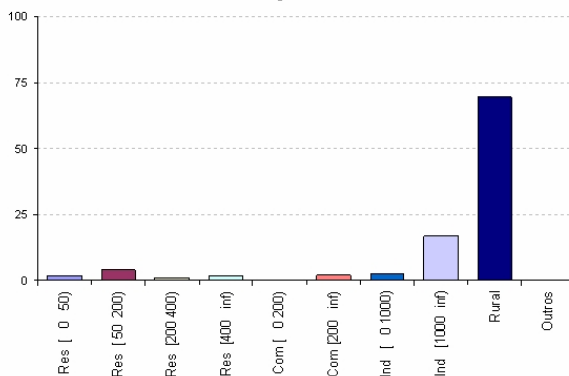
Grupo 18



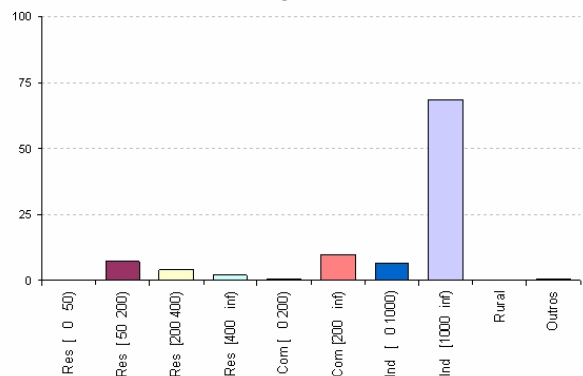
Grupo 19



Grupo 20



Grupo 21



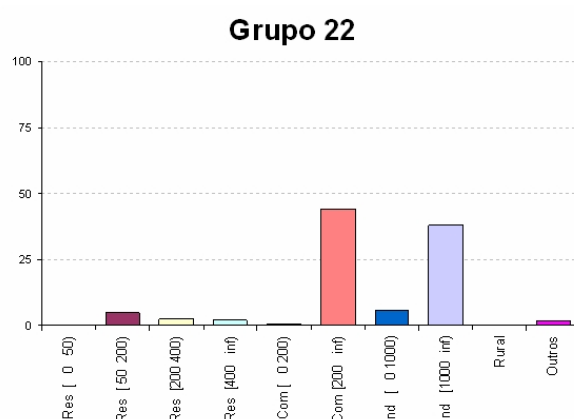


Figura 6.8: Gráficos dos centróides do treinamento pelo método Kohonen em Anel - Aparecida de Goiás

Na Figura 6.9 é apresentado o resultado da classificação de Aparecida de Goiânia pelo método Kohonen Anel. Observe a grande variação de classes no perímetro urbano discutida anteriormente.

As áreas rurais foram classificadas praticamente com a curva do grupo 19 que mostra uma classe ainda livre da presença urbana. Entretanto estas áreas estão inter-mescladas com áreas de perfil residencial, mostrando uma área rural pouco homogênea, com pequenas aglomerações nas extremidades sudoeste e sudeste do município.

O grupo 21 apresenta as quadriculas com um perfil predominante industrial (acima de 1000 kWh), as quais estão concentradas na região leste do município.

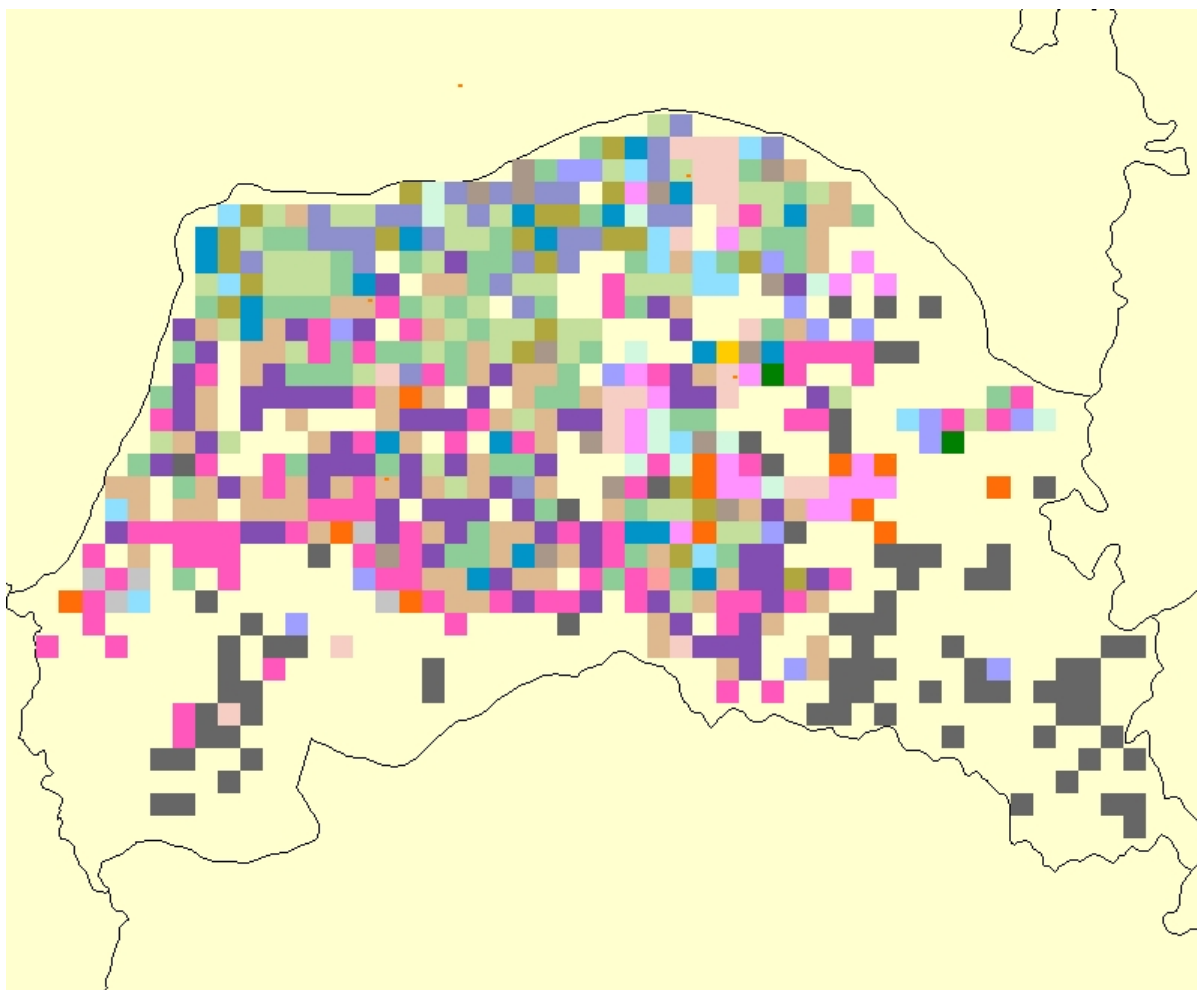


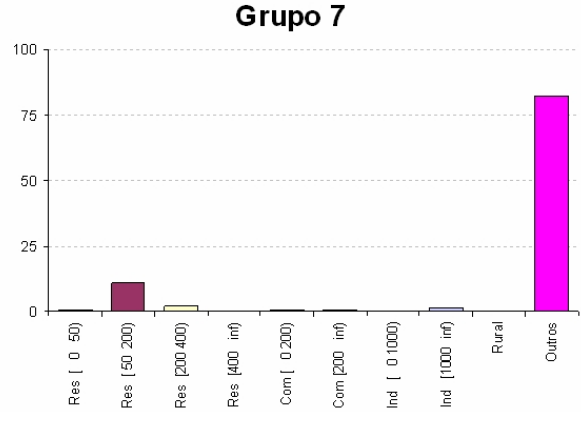
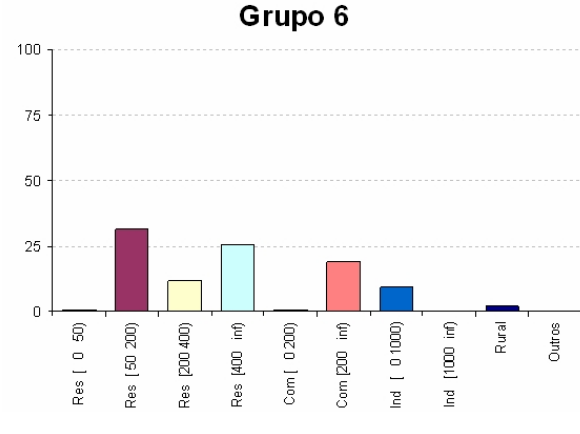
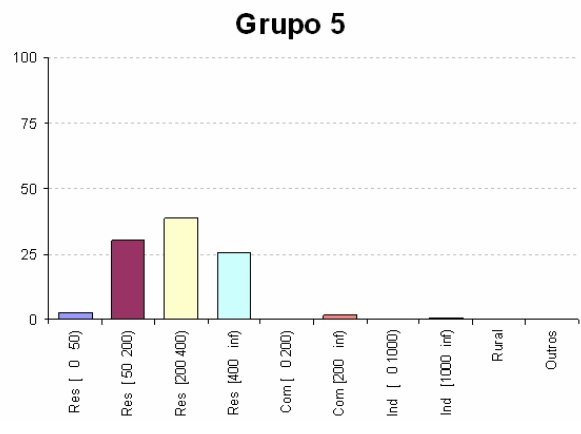
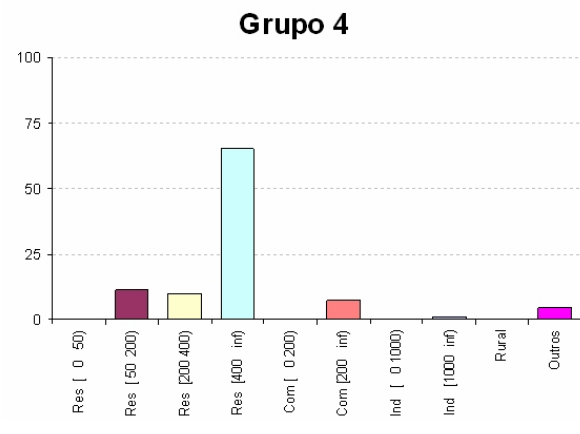
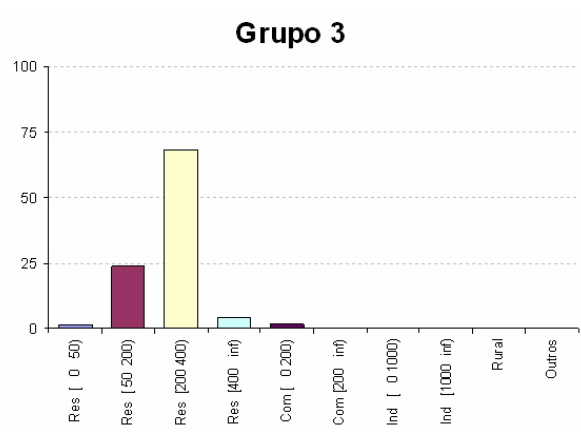
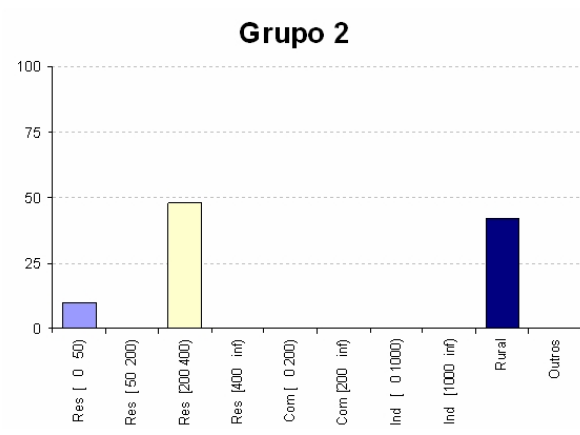
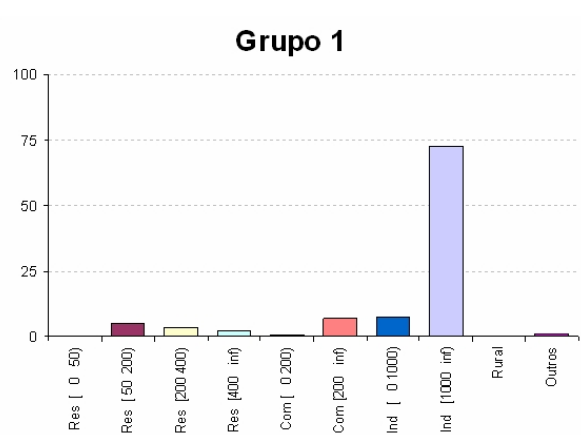
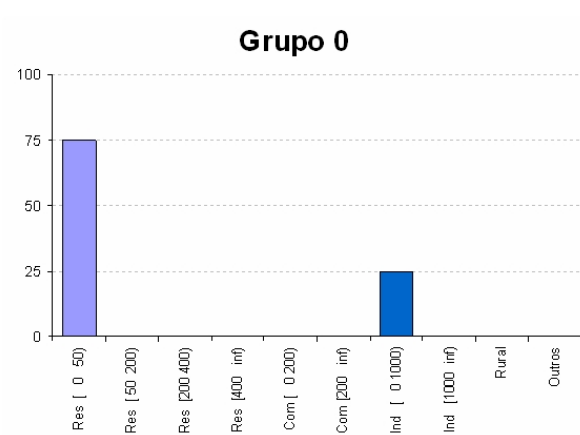
Figura 6.9: Tela do resultado da classificação pelo método Kohonen em Anel

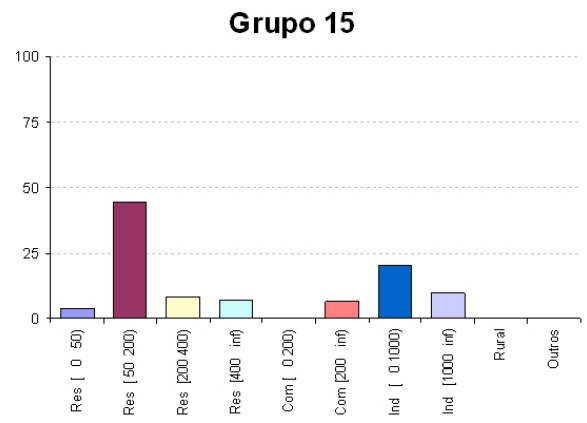
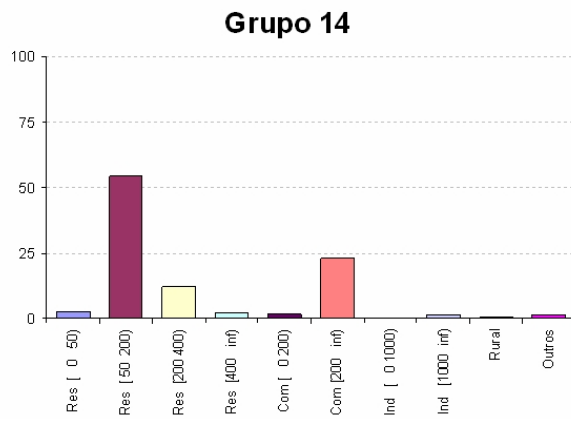
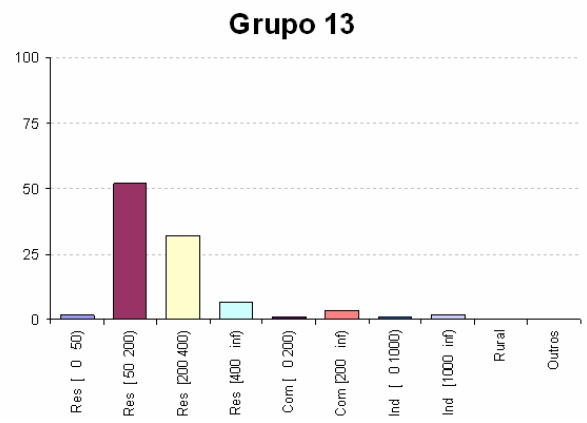
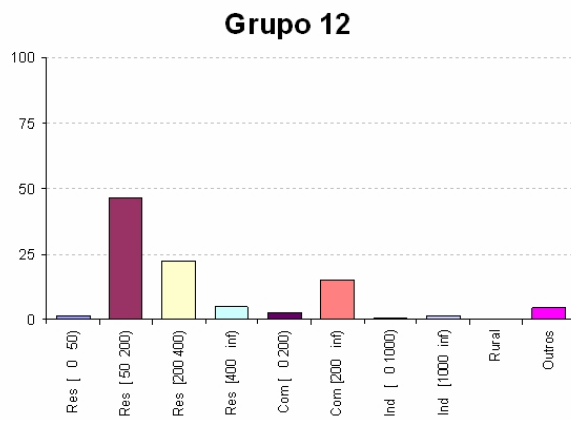
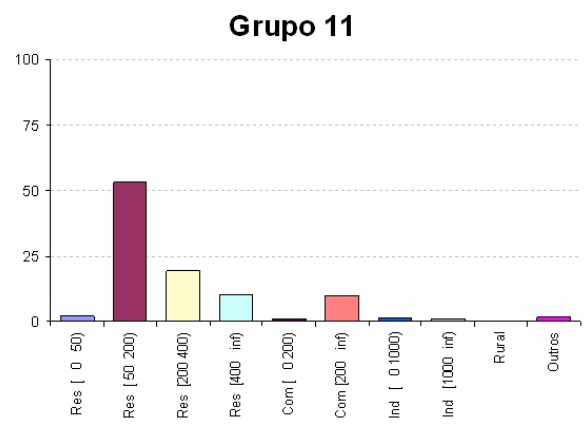
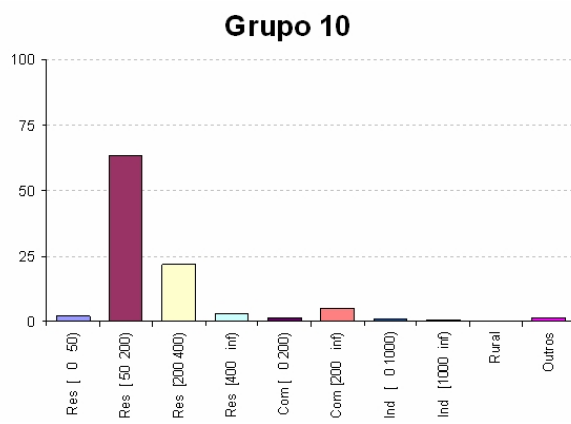
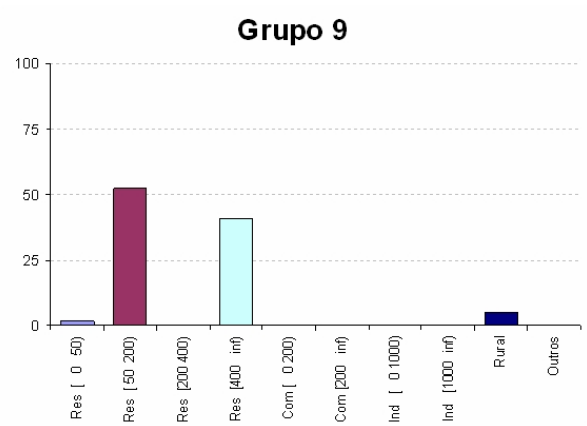
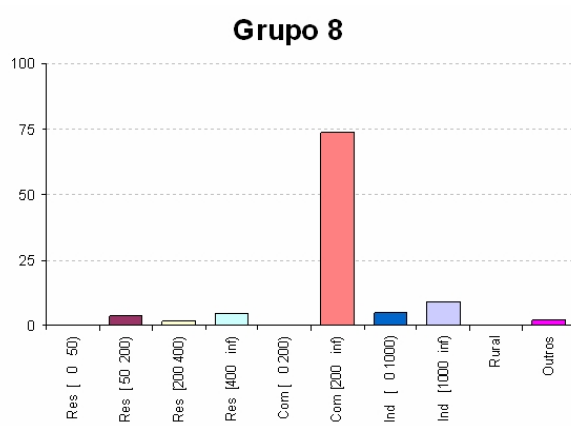
6.5.2 K-Médias

O método K-Médias novamente apresentou centróides bem distribuídos e alguns não encontrados no método Kohonen em Anel. São exemplos os grupos 0, 2 e 7.

O grupo 7 identificou as áreas com presença predominante do poder público. No método Kohonen em Anel este perfil é identificado pela curva do grupo 17, possuindo praticamente a mesma quantidade de quadrículas identificadas (Tabela 6.7).

Isso ocorre em cada método porque estas curvas são diferentes dos demais centróides. Assim as quadrículas identificadas nestes grupos apresentaram uma distância euclidiana maior para os demais centróides. Assim curvas ligeiramente diferentes, em métodos distintos, podem apresentar uma mesma composição de quadrículas na classificação dependendo da composição dos demais centróides.





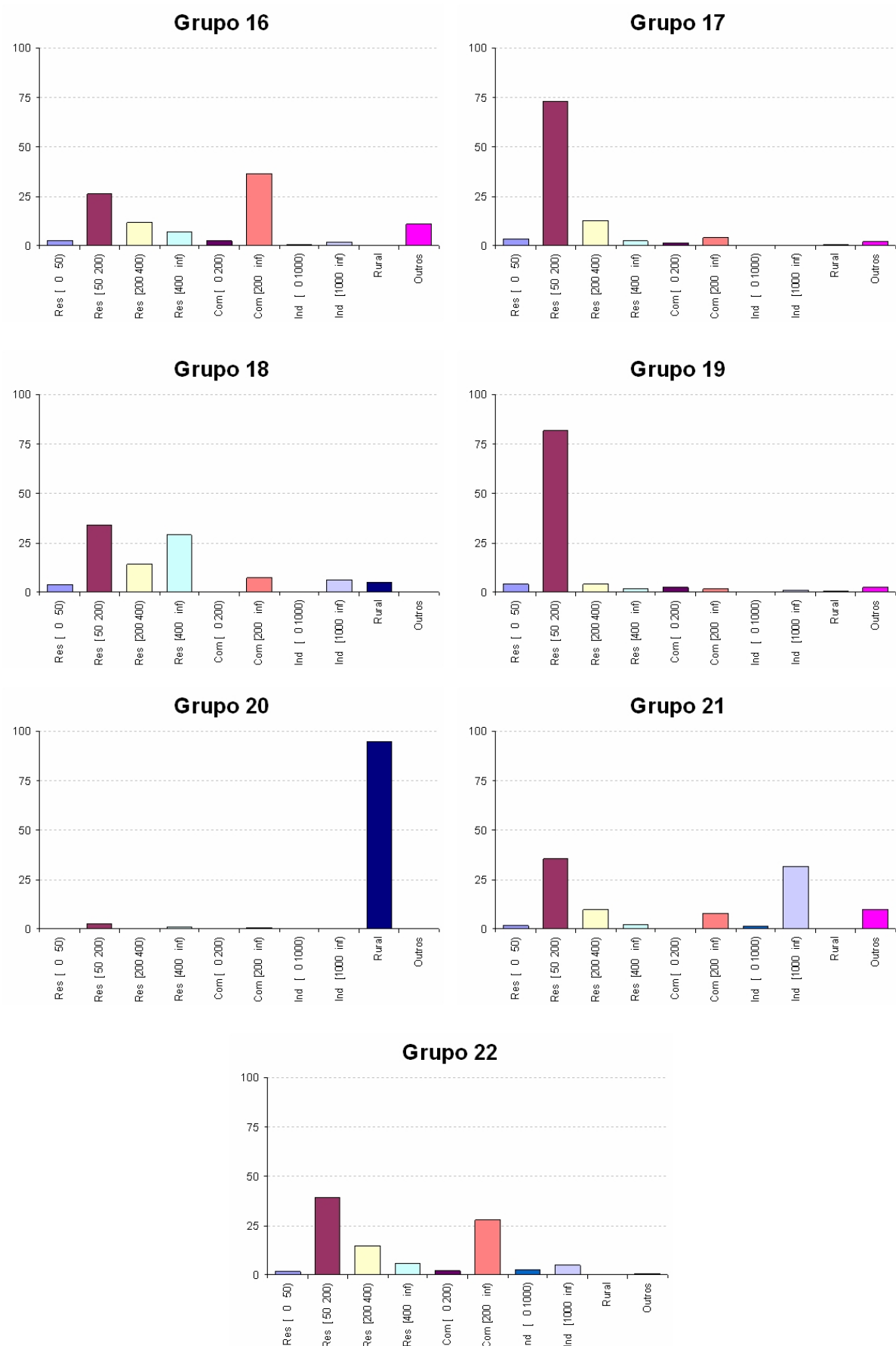


Figura 6.10: Gráficos dos centróides do treinamento pelo método K-Médias - Aparecida de Goiânia

Todos os grupos obtiveram quadrículas no processo de classificação, mas muitos grupos ficaram com um número inferior a seis quadrículas. Portanto pela Figura 6.11 notamos que este método produziu uma menor distribuição gráfica das quadrículas. Isso entretanto não implica em uma classificação de menor qualidade, pois o método Kohonen em Anel gerou centróides muito próximos, os quais foram generalizados no método K-Médias.

O grupo 20 identifica as áreas predominantemente rurais, as quais são as mesmas identificadas pelo método Kohonen em Anel (grupo 19), o que confirma ainda mais o processo de classificação destas áreas.

Os grupos 10 e 17 totalizam 250 quadrículas, ou seja, 40, 13% do conjunto de treinamento. Tanto o grupo 10 quanto o 17 tem o perfil predominante de clientes residências de 50 a 200 kWh. Assim enquanto o método Kohonen em Anel divide estas mesmas 250 quadrículas em mais grupos o método K-Médias faz uma melhor generalização pela falta de ação do vetor de atualização dos centróides.

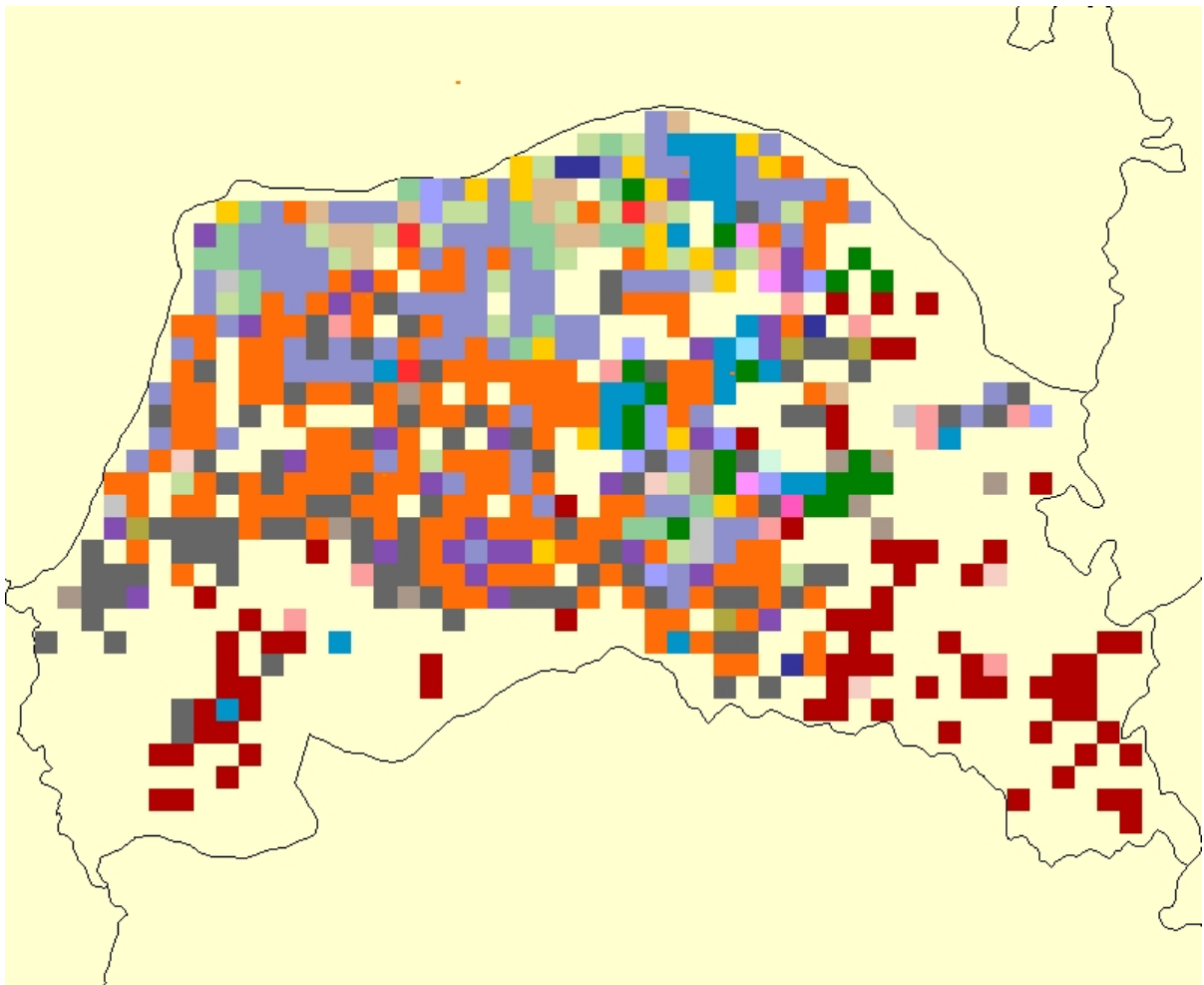


Figura 6.11: Tela do resultado da classificação pelo método K-Médias

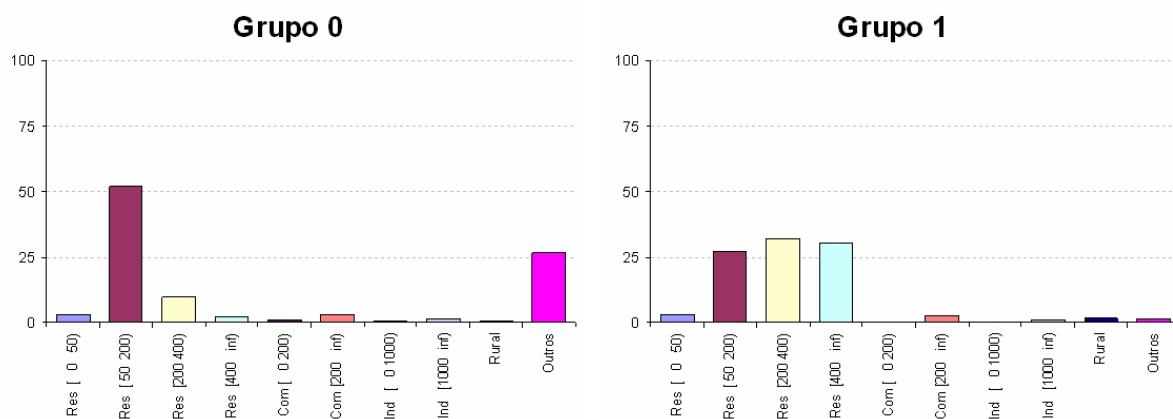
As áreas sem classificação no interior do perímetro urbano se deve ao tipo de carga utilizado no treinamento, o qual foi por instalação. Este tipo de carga baseia-se na localização espacial dos transformadores e não dos clientes. Assim sendo estas áreas sem classificação provavelmente possuem clientes, mas seus transformadores estão localizados nas áreas com classificação.

6.5.3 Fuzzy C-Médias

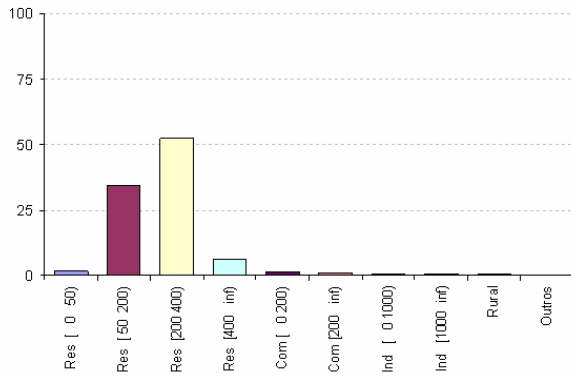
Este método, como no Kohonen em Anel, produziu uma maior distribuição no processo de classificação (Figura 6.13). Entretanto obteve uma melhor distribuição das quadrículas nos grupos formados, veja a Tabela 6.7. Observe que o Fuzzy C-Médias não produziu grupos com grande número de quadrículas como nos métodos anteriores.

A distribuição dos centróides é a responsável pela melhor distribuição das quadrículas, Figura 6.12. Os centróides formados neste método são próximos aos do K-Médias, mas essa pequena diferença implicou no considerável aumento da distribuição das quadrículas.

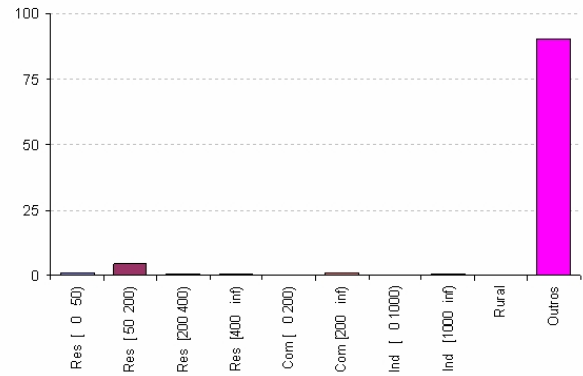
Apesar da proximidade dos centróides com os do K-Médias este método produziu ainda centróides exclusivos como é o caso do grupo 7, o qual recebeu nove quadrículas no processo de classificação (Tabela 6.7).



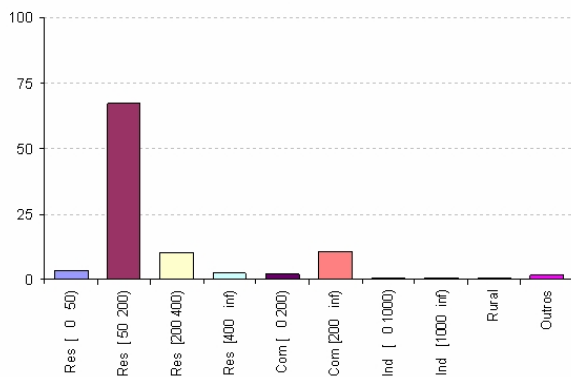
Grupo 2



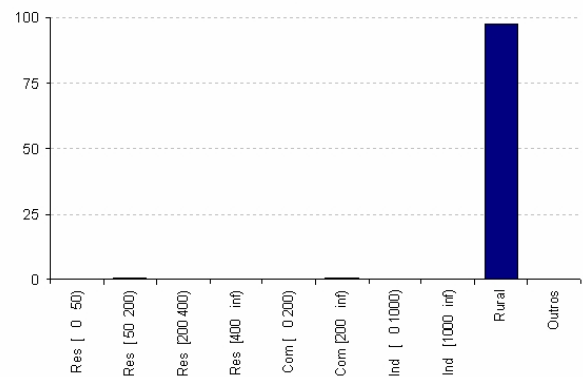
Grupo 3



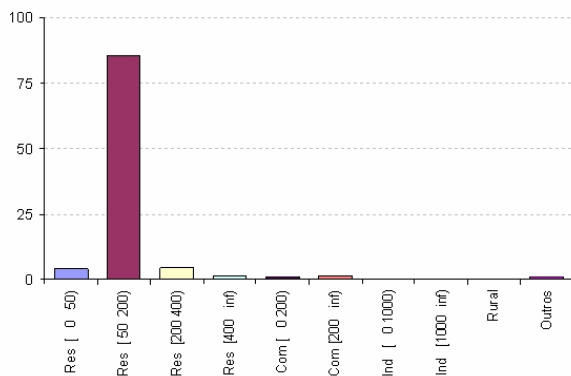
Grupo 4



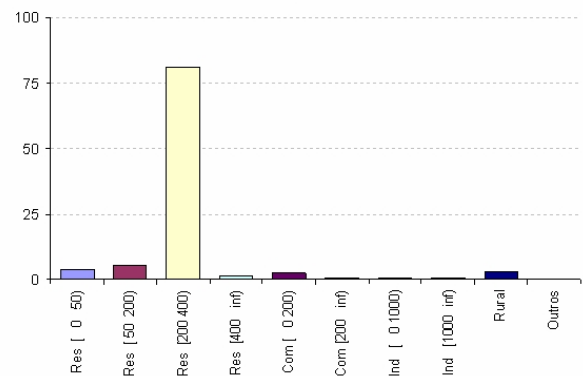
Grupo 5



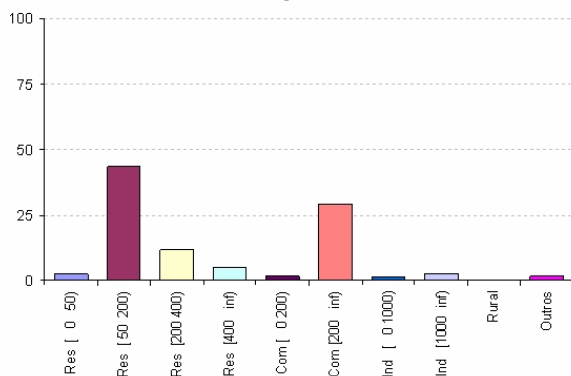
Grupo 6



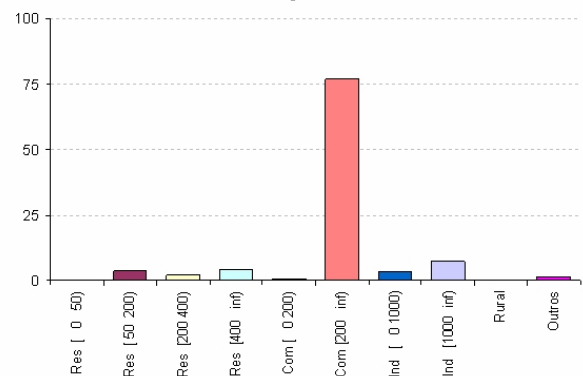
Grupo 7

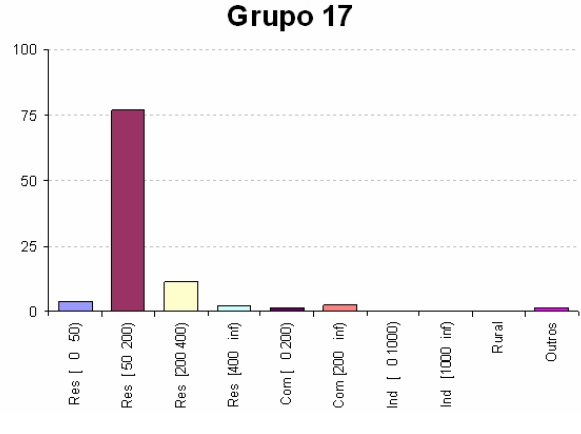
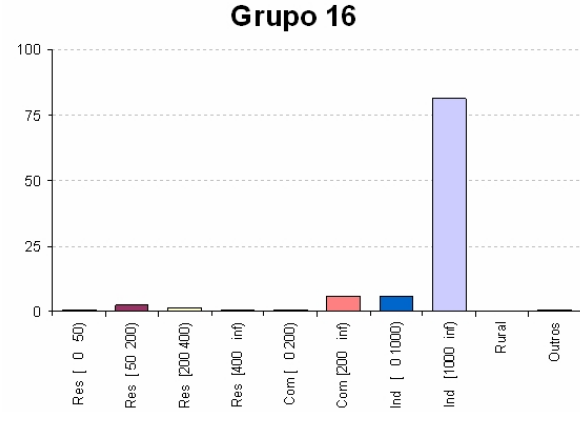
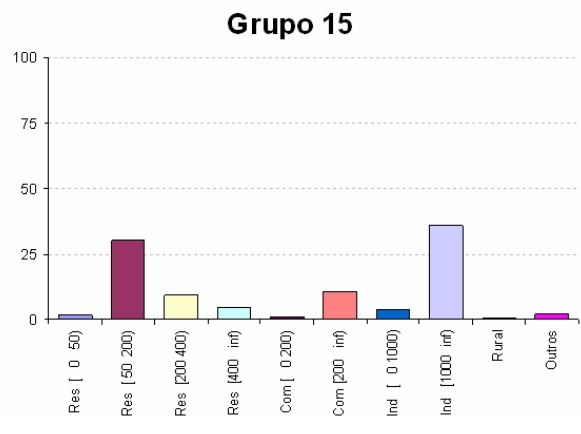
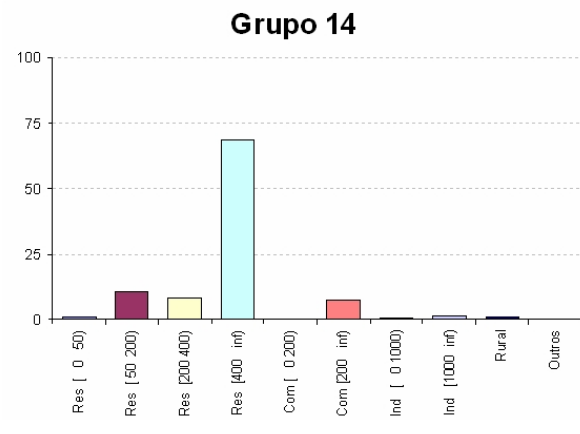
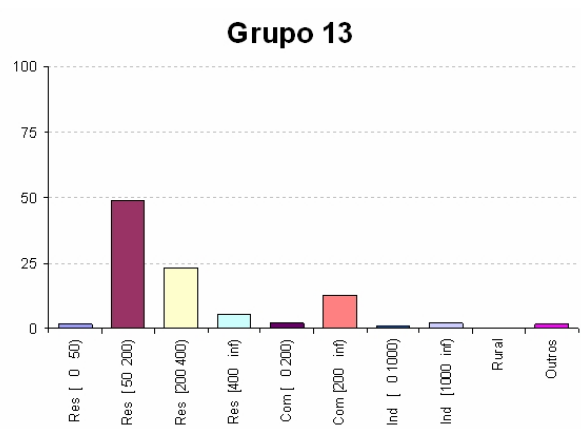
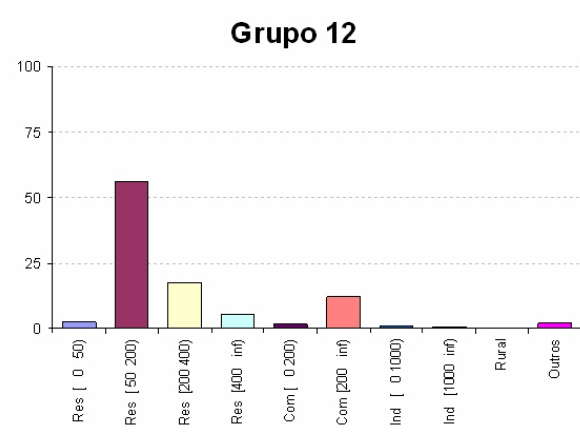
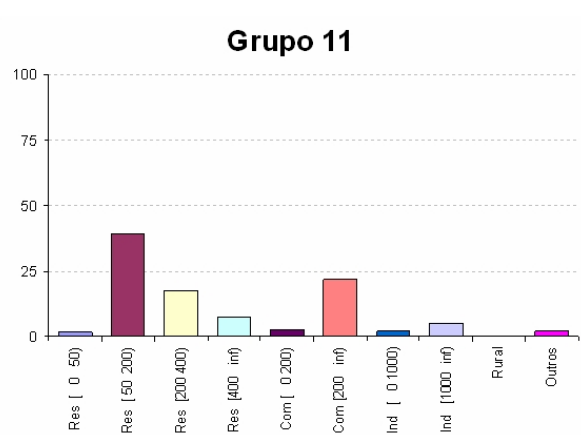
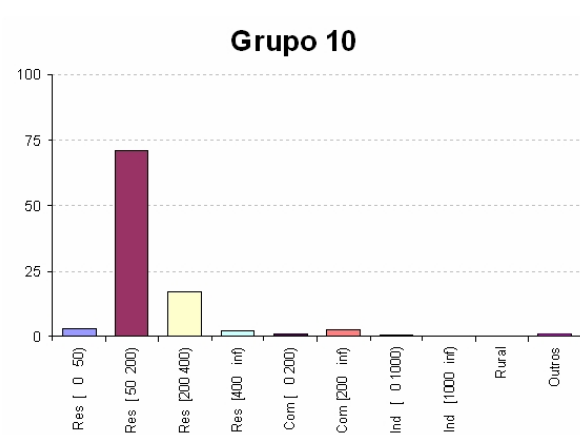


Grupo 8



Grupo 9





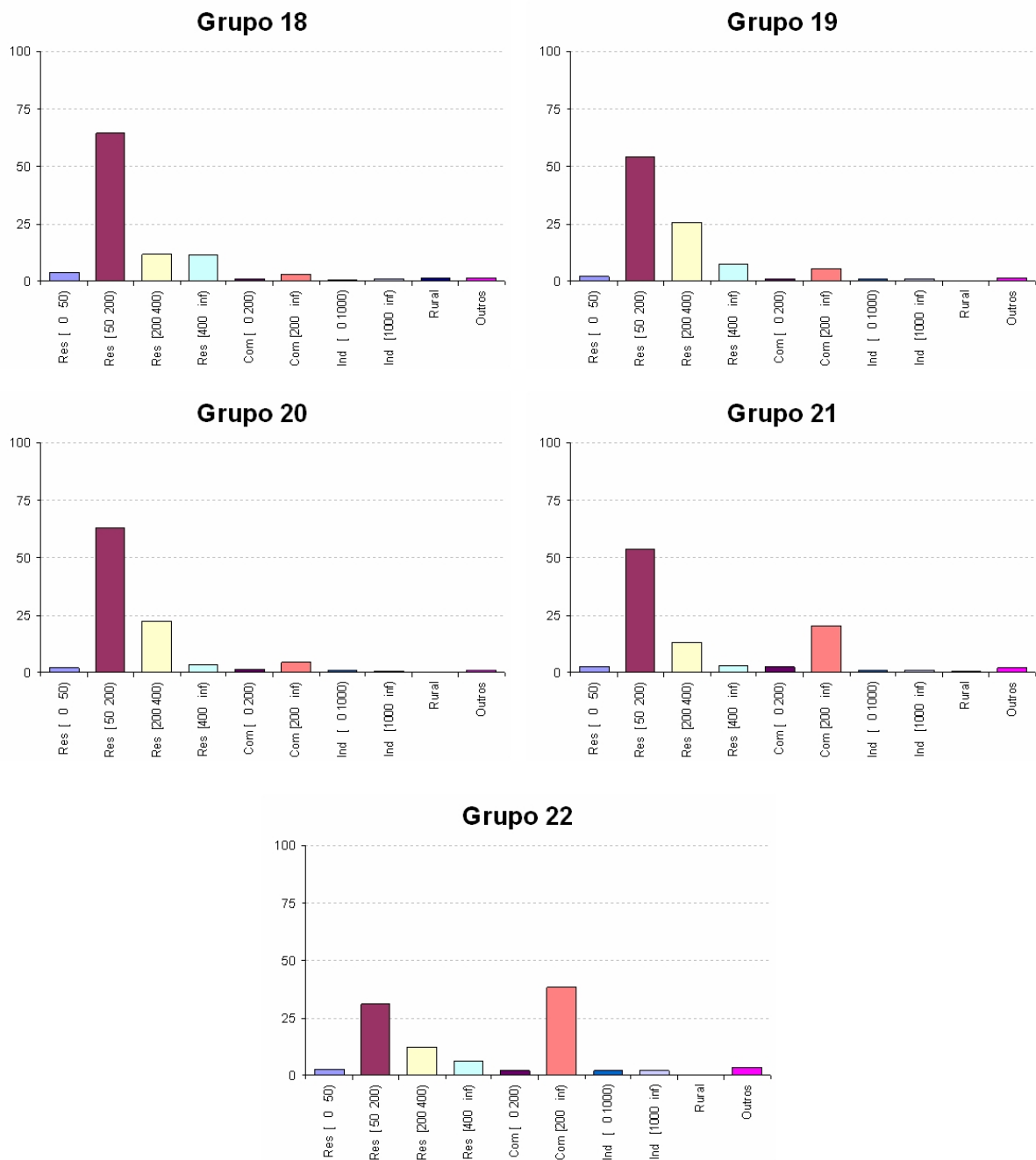


Figura 6.12: Gráficos dos centróides do treinamento do método Fuzzy C-Médias - Aparecida de Goiânia

Essa melhor distribuição das quadrículas é observado apenas nas classes com perfil residencial e comercial, pois a disposição das demais classes está em conformidade com os métodos anteriores. A composição de quadrículas do grupo 5 do Fuzzy C-Médias está idêntica a do grupo 20 do K-Médias (Tabela 6.7). No método Kohonen em Anel este perfil é identificado pelo grupo 19 que possui apenas cinco quadrículas a mais que os métodos anteriores.

Este município, como dito anteriormente, produziu uma grande variedade de curvas distin-

tas após a normalização. Isso foi constatado na etapa de defuzzificação e conseqüente classificação do conjunto de treinamento pela necessidade da diminuição do índice de defuzzificação, em relação ao treinamento de Anápolis, devido a alta distribuição do grau de pertinência das curvas aos grupos formados.

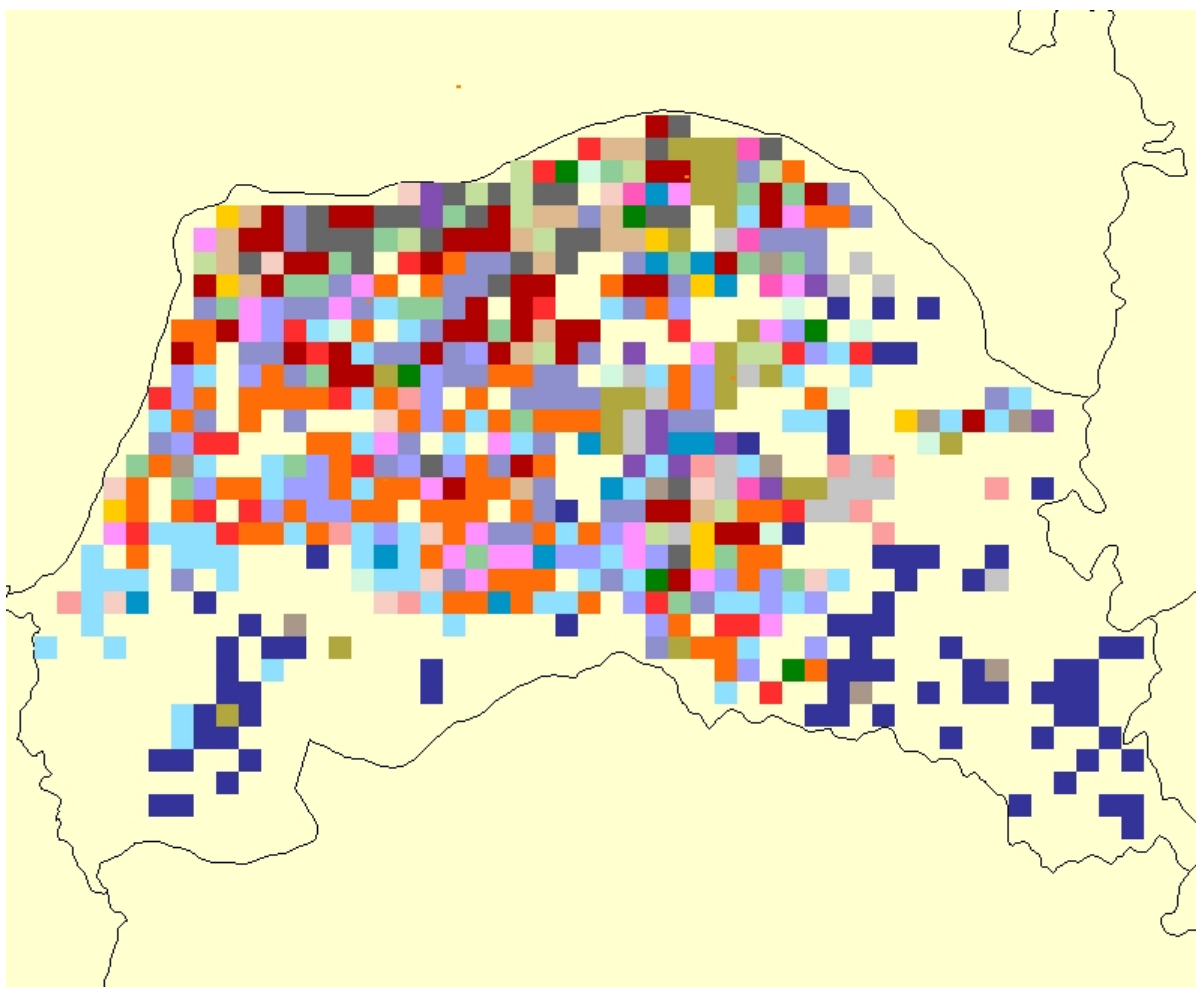


Figura 6.13: Tela do resultado da classificação pelo método Fuzzy C-Médias

6.5.4 Distribuição das classificações

Nesta seção é apresentada a Tabela 6.2 que contém o resultado das classificações dos métodos citados anteriormente no município de Aparecida de Goiânia.

Como já foi explicado não há relação entre o número de quadrículas dos métodos sobre um determinado grupo. Essa tabela deve ser analisada em conjunto dos gráficos dos centróides. Por exemplo, o grupo 5 do método Fuzzy C-Médias pode ser comparado com o grupo 20 do K-Médias e com o grupo 19 do Kohonen em Anel.

Tabela 6.7: Relação de quadrículas por grupo - Aparecida de Goiânia

Grupos	Número de Quadrículas		
	Kohonen	K-Médias	Fuzzy
0	22	4	11
1	2	19	6
2	12	1	9
3	1	12	10
4	16	12	37
5	-	4	75
6	14	1	68
7	14	11	9
8	21	25	12
9	22	4	25
10	28	93	60
11	54	24	14
12	63	20	24
13	82	13	17
14	74	28	11
15	80	1	5
16	5	6	19
17	12	157	81
18	-	3	23
19	80	90	27
20	-	75	49
21	20	3	23
22	1	17	8
-	396	396	396

Para o município de Aparecida foram identificadas 1019 quadrículas, sendo que 623 com consumo acima de zero kWh. Estas foram normalizadas e distribuídas na Tabela 6.7 como resultado do processo de agrupamento e classificação.

6.5.5 Medidas do agrupamento

Serão apresentados, nesta seção, os resultados das medidas do processo de agrupamento e classificação do conjunto de treinamento do município de Aparecida de Goiânia.

A Tabela 6.8 apresenta apenas o tempo de processamento da etapa de agrupamento das quadrículas, já que para classificação é utilizada apenas um iteração para todos os métodos.

Novamente o método K-Médias obteve o menor resultado, mas com uma diferença expressiva para este município. Tanto para o K-Médias quanto para o Fuzzy C-Médias foi atribuído um erro igual a zero, entretanto o método Fuzzy demorou bem mais para convergir. Já o método Kohonen em Anel demorou proporcionalmente o mesmo tempo, pois foi utilizado o mesmo número de iterações que no estudo de caso anterior.

A expressiva diferença de tempo de processamento entre o estudo de caso 1 e 2 pode ser explicado pela diferença de quadrículas distintas após a normalização. Apesar de Anápolis possuir um maior número de quadrículas ela possui uma menor quantidade de quadrículas distintas que Aparecida de Goiânia. Isso interfere no método Kohonen em anel, que usa o número de iterações, e nos métodos K-Médias e Fuzzy C-Médias, que utilizam o critério de erro como condição de parada.

Tabela 6.8: Tempo de processamento dos métodos - Aparecida de Goiânia

Método	Tempo
Kohonen Anel	52 min 01 seg
K-Médias	32 seg
Fuzzy C-Médias	10 min 47 seg

O resultado de medição do Índice Intra-Grupo (IAG) é mostrado na Tabela 6.9 onde é observada uma maior aproximação dos valores que no estudo de caso anterior. Isso se deve ao perfil predominantemente residencial e comercial de quase toda a extensão do município de Aparecida de Goiânia, o que provoca uma menor generalização dos grupos quando comparado ao município de Anápolis.

O método Fuzzy C-Médias obteve novamente o menor valor seguido de perto pelo método K-Médias. Assim pelo parâmetro IAG o método Fuzzy C-Médias realizou a melhor classificação do conjunto de treinamento.

Tabela 6.9: Resultados dos Índices Intra-Grupo (IAG) - Aparecida de Goiânia

Método	IAG
Kohonen Anel	414
K-Médias	354
Fuzzy C-Médias	352

A Tabela 6.10 apresenta o resultado da medição do Índice Inter-Grupo. Neste treinamento o método K-Médias obteve o maior resultado, com até 22% acima do menor valor. Como já foi dito esse índice não permite afirmar que quanto menor o valor, melhor é o resultado.

Tabela 6.10: Resultados dos Índices Inter-Grupo (IEG) - Aparecida de Goiânia

Método	IEG
Kohonen Anel	717
K-Médias	880
Fuzzy C-Médias	723

Essa discrepância do método K-Médias pode ser explicado pela formação de grupos com poucas quadrículas (Tabela 6.7). Isso caracteriza a formação de grupos que se distanciam dos demais elementos do conjunto de treinamento. Os demais métodos obtiveram um resultado próximo.

6.6 Estudo de Caso 3 - Rio Verde

Ocupando uma área de 8.388 km^2 , o município de Rio Verde está localizado na microrregião Sudoeste do Estado de Goiás, a 220 km de Goiânia, capital do Estado.

No recente crescimento do agronegócio brasileiro, a cidade de Rio Verde tem se destacado, por contar com uma importante cooperativa agrícola (COMIGO) e com importantes unidades industriais que agregam valor à sua produção agrícola. Rio Verde é hoje o maior produtor de grãos do Estado produzindo arroz, soja, milho, algodão, sorgo, feijão e girassol. Conta ainda com importante plantel bovino, avícola e suíno.

Para estimular as micro e pequenas empresas, foi implantado em 18 de março 2004, o 5º distrito industrial de Rio Verde, o DIMPE (Distrito Industrial Municipal de Pequenas Empresas).

Seguem abaixo as Tabelas 6.11 e 6.12 que mostram o consumo e o número de clientes por classe no município de Rio Verde nos anos de 2005 e 2006, respectivamente.

Tabela 6.11: Consumo e número de consumidores por classe - Rio Verde 2005 (CELG, 2007)

Classe de consumidores	Consumo (GWh)	Consumo (%)	Número de consumidores	Consumidores (%)
Residencial	75,35	19,50%	41.981	85,64%
Industrial	209,29	54,16%	293	0,60%
Comercial	40,37	10,45%	3.278	6,69%
Rural	36,14	9,35%	3.134	6,39%
Poder Público	5,86	1,52%	286	0,58%
Iluminação	13,19	3,41%	-	-
Serviço Público	6,18	1,60%	46	0,09%
Próprio	0,08	0,02%	3	0,01%
Total	386,46	100%	49.021	100%

Através destas tabelas pode-se constatar um crescimento de 12,3% no consumo rural e de 15,4% no consumo industrial em apenas um ano. No cenário geral houve um aumento de 10,8% no consumo total seguido de um aumento de 5,2% no número de clientes.

Tabela 6.12: Consumo e número de consumidores por classe - Rio Verde 2006 (CELG, 2007)

Classe de consumidores	Consumo (GWh)	Consumo (%)	Número de consumidores	Consumidores (%)
Residencial	76,86	17,94%	44.007	85,64%
Industrial	241,61	56,41%	275	0,53%
Comercial	40,00	9,34%	3.651	7,08%
Rural	40,60	9,48%	3.282	6,36%
Poder Público	6,83	1,59%	316	0,61%
Iluminação	13,71	3,20%	-	-
Serviço Público	8,65	2,02%	55	0,11%
Próprio	0,09	0,02%	3	0,01%
Total	428,35	100%	51.589	100%

A classe rural de Rio verde possui um grande consumo comparado com os demais municípios de Goiás, com consumo 215,95% maior que a classe rural de Anápolis analisada anteriormente.

6.6.1 Considerações sobre o estudo de caso

Este estudo de caso, diferentemente dos demais, não tem a intenção de comparar os métodos de agrupamento e classificação.

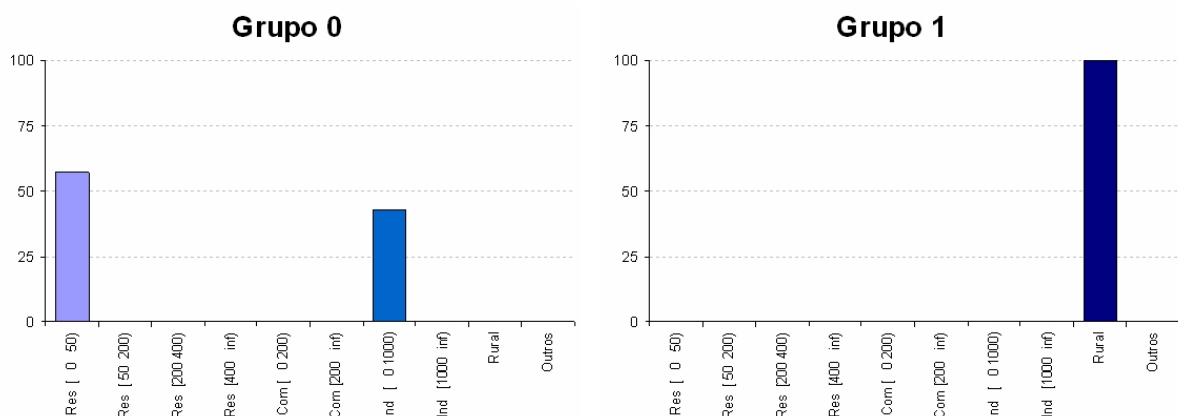
O objetivo deste estudo de caso é analisar as mudanças de perfil das quadrículas de um ano para o outro. Para isso, é realizado o agrupamento e classificação do cenário de 2005 e posteriormente os centróides gerados serão utilizados no processo de classificação no cenário do ano seguinte. Assim, não haverá alterações nos centróides para o treinamento de 2006, sendo feita apenas a classificação das quadrículas.

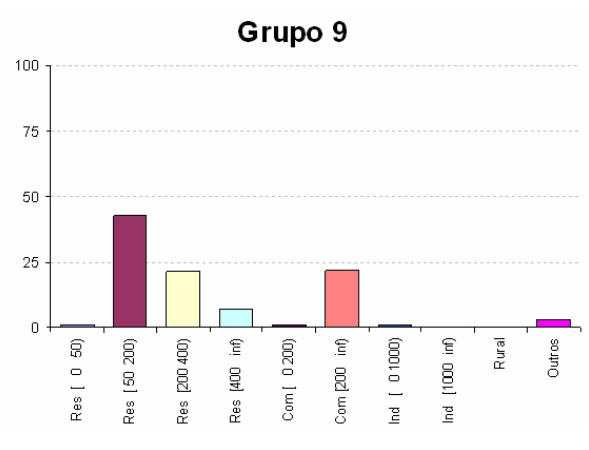
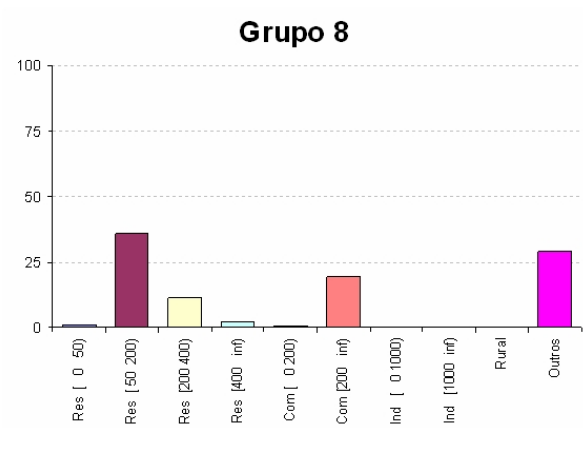
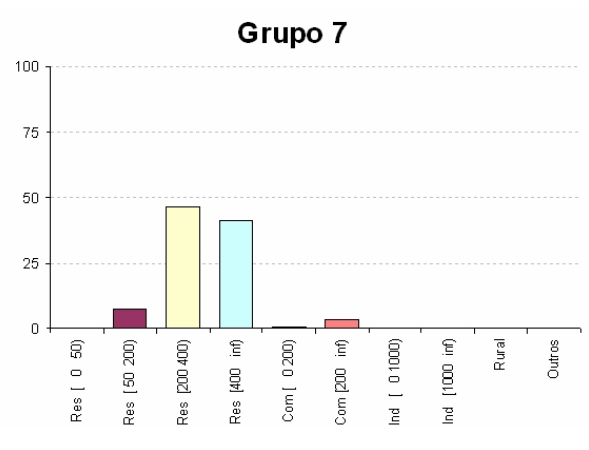
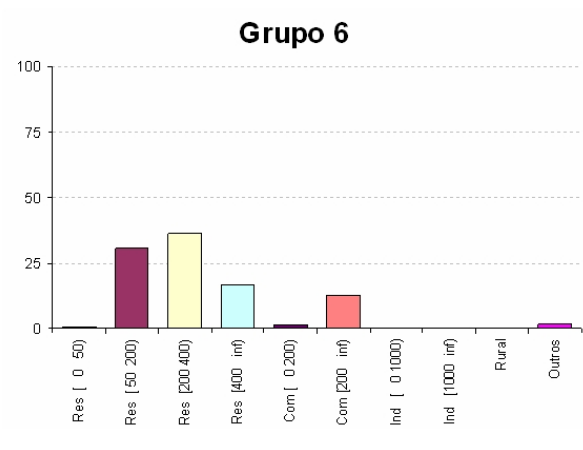
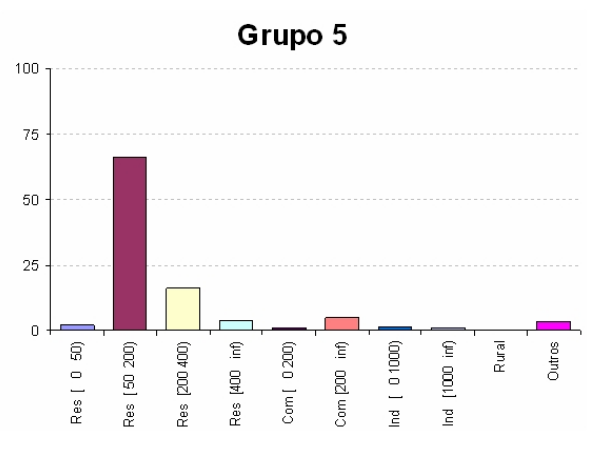
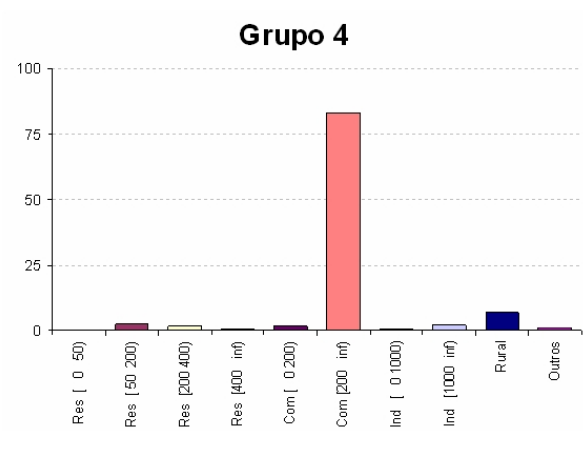
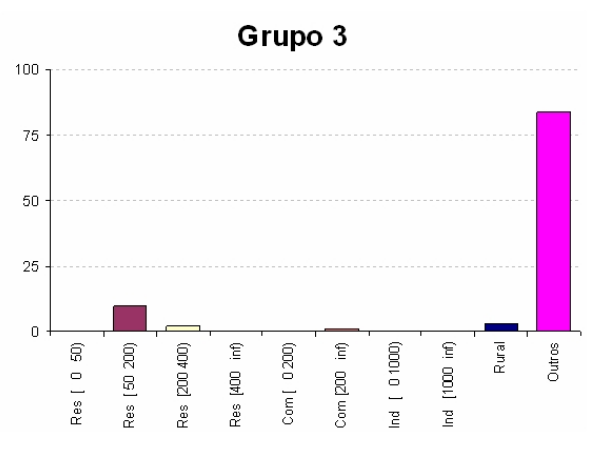
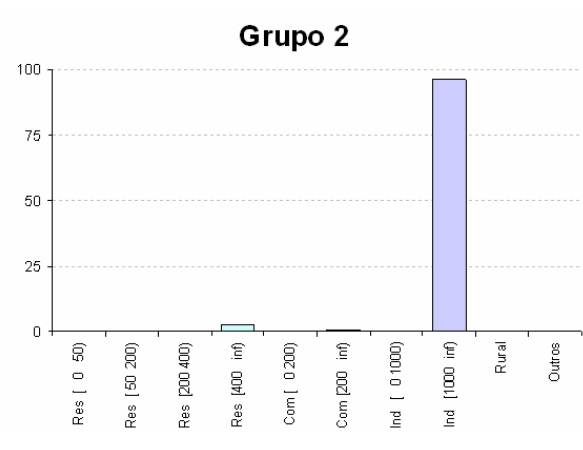
Outro objetivo deste estudo de caso é a análise do crescimento horizontal e vertical do cenário de 2005 para 2006. Entretanto, este estudo é meramente ilustrativo, pois necessitou-se montar o cenário de 2005, já que não houve a possibilidade de se conseguir a rede elétrica distribuída geograficamente para o ano de 2005.

Foi escolhido o método K-Médias para se realizar o agrupamento e classificação, onde o número de saídas determinado foi de 14 grupos.

6.6.2 Método K-Médias

O método K-Médias produziu novamente uma boa distribuição dos centróides. Conforme mostrado na Figura 6.14 foram gerados grupos com curvas bem definidas, onde há a predominância de apenas uma faixa de consumo.





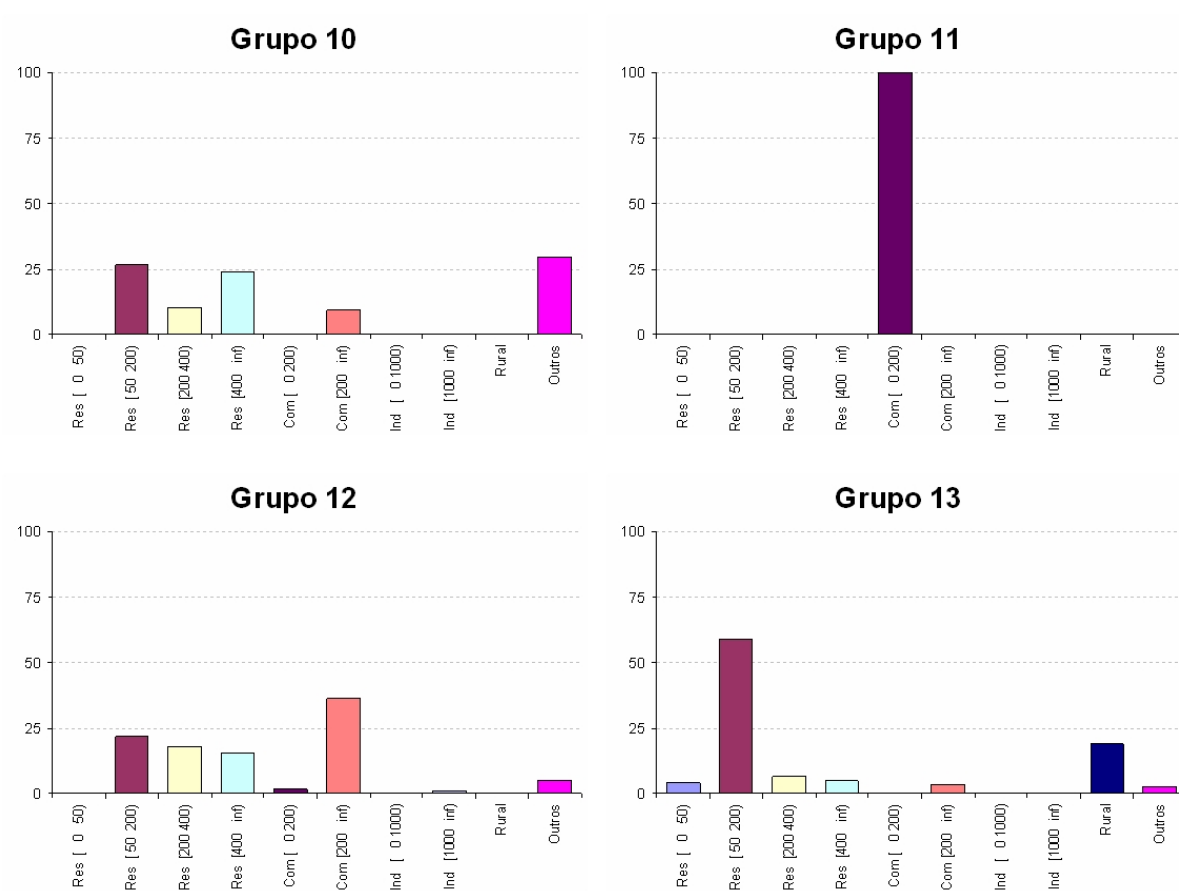


Figura 6.14: Gráficos dos centróides do treinamento pelo método K-Médias - Rio Verde

O município de Rio Verde possui uma extensa área rural, com identificação de 2059 quadriculas (Tabela 6.13). Devido a essa extensão rural o resultado gráfico da classificação não pode ser visualizado de forma satisfatória para todo o município, veja a Figura 6.15.

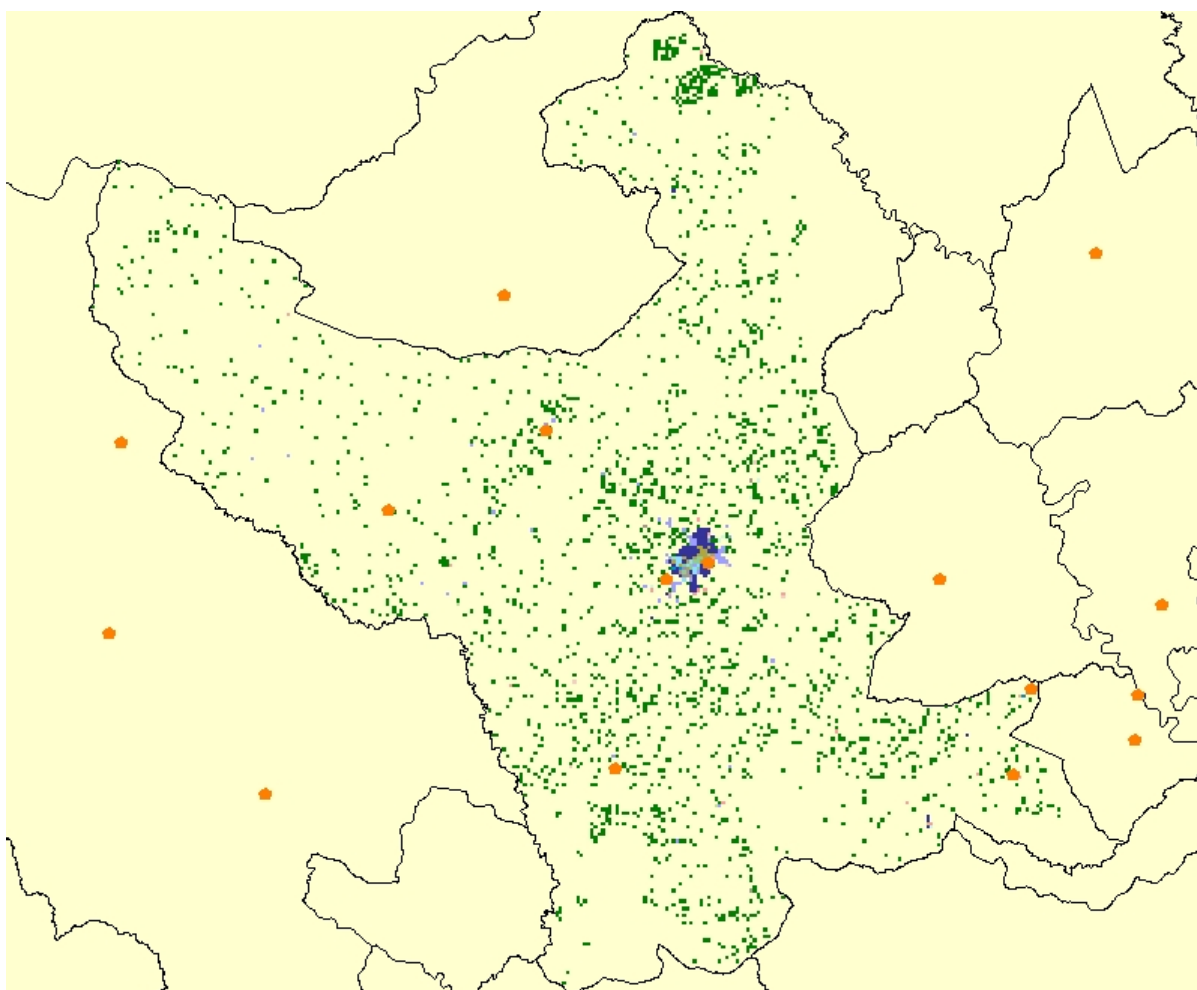


Figura 6.15: Tela do resultado da classificação pelo método K-Médias em Rio Verde - 2005

Para melhorar o processo de análise, foi feita uma ampliação à área urbana de Rio Verde, a qual será usada no decorrer deste estudo de caso (Figura 6.16).

Pode ser visto na Figura 6.16 uma grande presença de quadriculas do grupo 4 que possui perfil comercial acima de 200 kWh. Estas Quadriculas estão dispostas ao redor do perímetro urbano e estão presentes também na região rural.

As quadriculas do grupo 2, perfil industrial acima de 1000 kWh estão distribuídas em torno do perímetro urbano. A mesma distribuição ocorre com as quadriculas do grupo 3, perfil de presença do poder público.

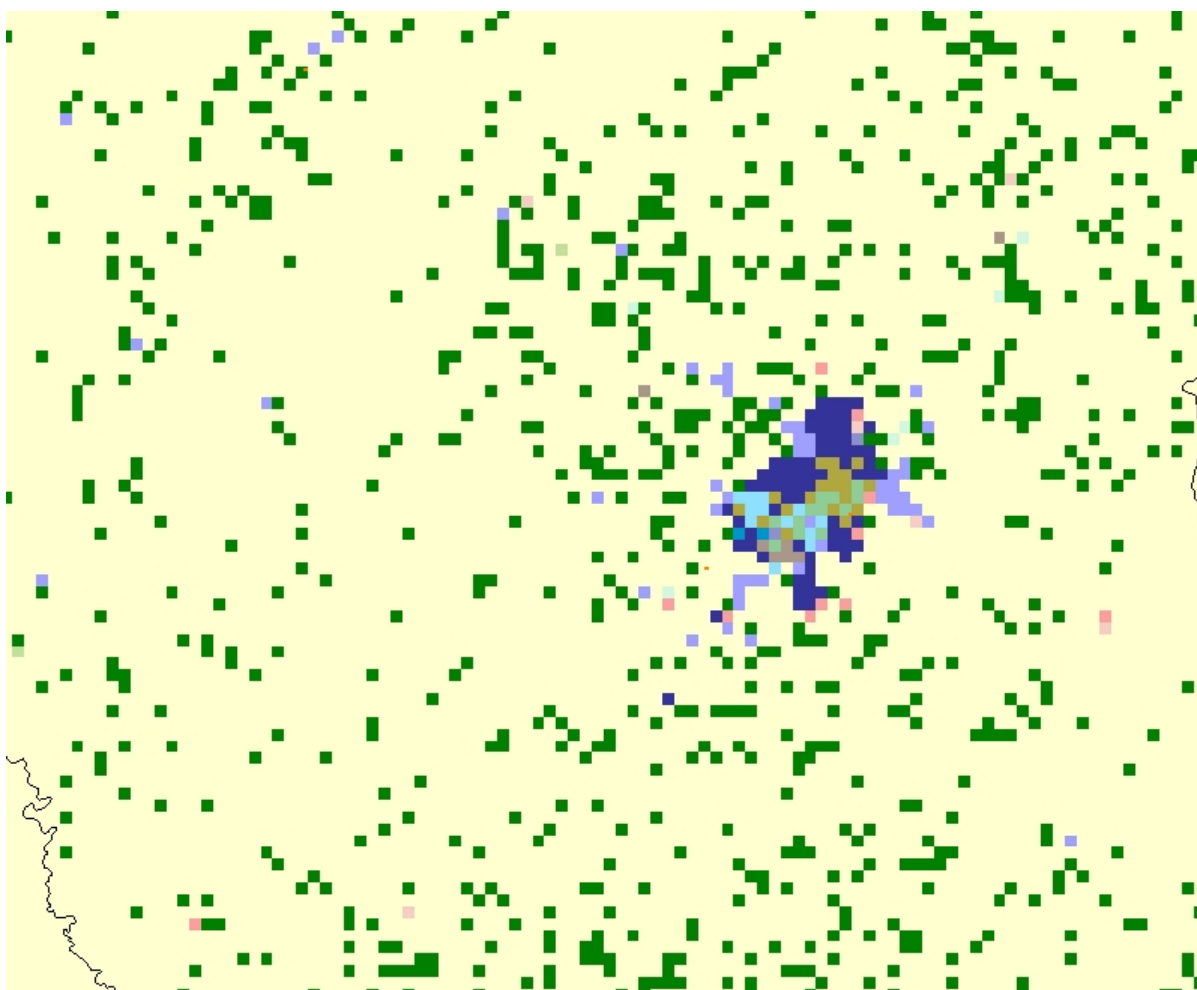


Figura 6.16: Zoom do resultado da classificação pelo método K-Médias em Rio Verde - 2005

Após a realização do processo de agrupamento e classificação do cenário de 2005 foi realizada a classificação do cenário de 2006. Esta classificação utilizou-se dos centróides produzidos pelo agrupamento do cenário de 2005 através de apenas uma iteração.

A Figura 6.17 mostra o resultado da classificação do cenário de 2006. Como não houve um novo agrupamento e conseqüente escolha aleatória de curvas de inicialização, as cores identificam os mesmos grupos nos dois cenários. Pode ser notada na Figura 6.17 uma pequena variação na composição dos grupos quando comparada com o cenário de 2005.

Essa pequena variação na composição dos grupos entre os cenários pode ser confirmada pela Tabela 6.13. A tabela mostra um grande crescimento do consumo rural com o aumento de 15% no número de quadriculas com perfil rural. Os demais grupos sofreram pequenas modificações, confirmado a apresentação das Figuras 6.16 e 6.17.

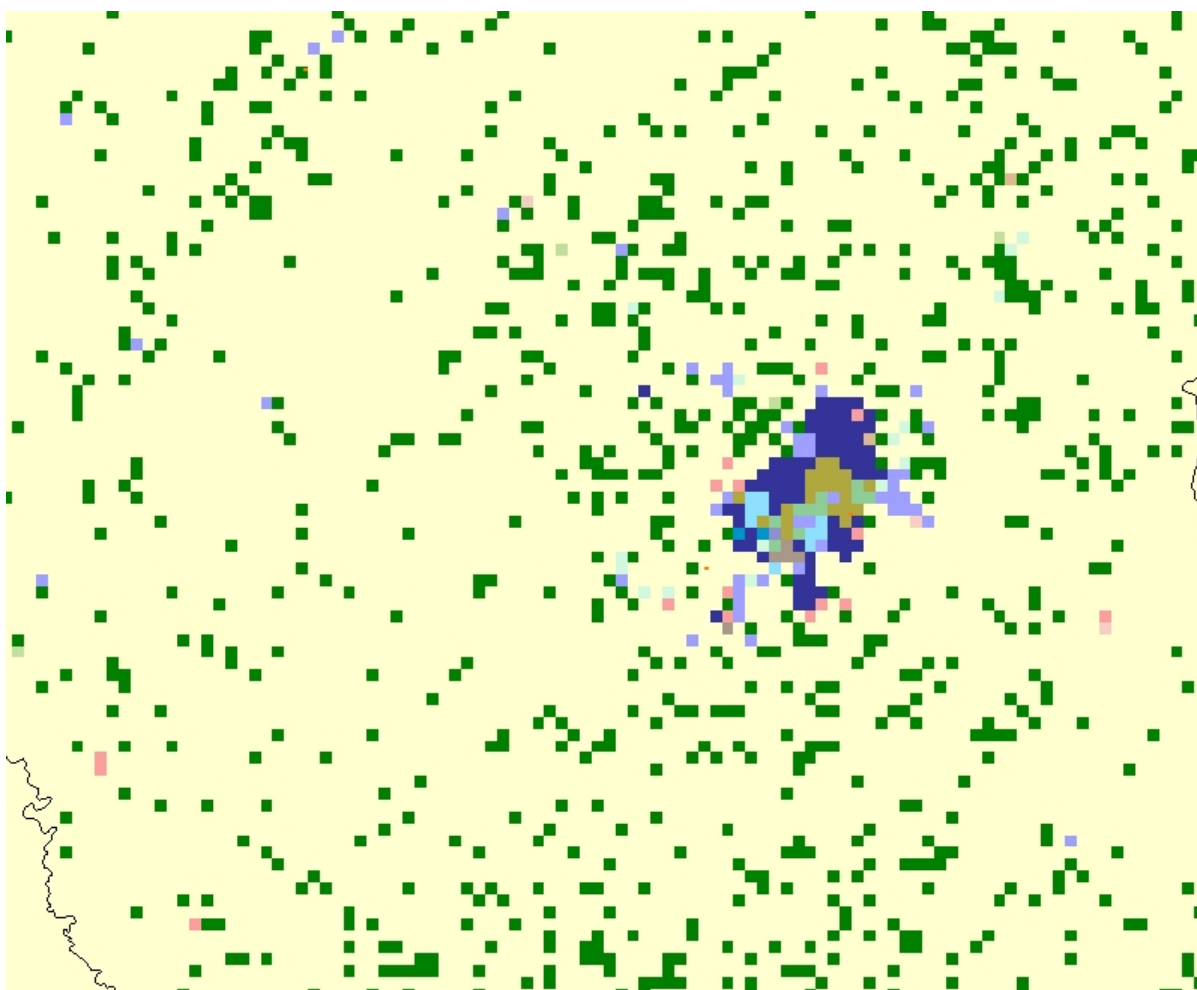


Figura 6.17: Zoom do resultado da classificação pelo método K-Médias em Rio Verde - 2006

6.6.3 Distribuição das classificações

Diferente dos demais estudos, o resultado das classificações apresentado nesta seção tem total relação quando analisado um grupo específico, conforme explicado anteriormente.

Foram identificadas 2295 quadrículas em 2005 e 2623 em 2006, um crescimento de 14,3%. Embora pareça grande o número de quadrículas com consumo acima de zero, o município possui ainda 30.318 quadrículas sem consumo, totalizando 91,3% das quadrículas do município.

Tabela 6.13: Relação de quadrículas por grupo - Rio Verde

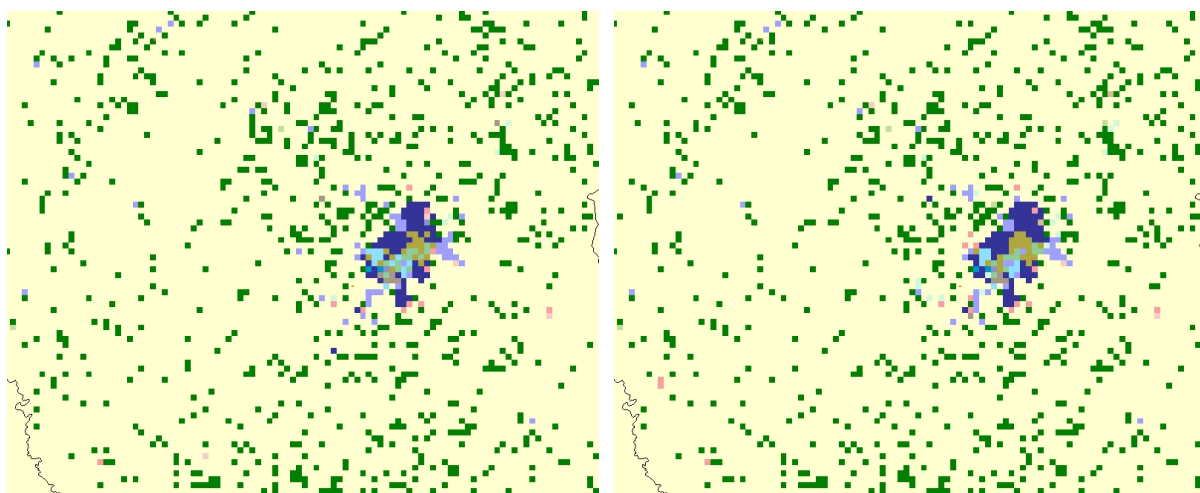
Grupos	Número de Quadrículas	
	2005	2006
0	7	3
1	2059	2372
2	7	16
3	20	26
4	62	61
5	79	82
6	12	10
7	8	5
8	3	3
9	19	24
10	2	1
11	3	6
12	12	10
13	2	4
-	30318	29990

Houve uma pequena variação do número de quadrículas entre os cenários, com um aumento expressivo apenas no grupo 1, que possui perfil rural.

6.6.4 Crescimento horizontal 2005/2006

Nesta seção é calculado o crescimento horizontal do município, onde os cenários inicial e final são comparados e identificadas as quadrículas que possuem um consumo diferente de zero no cenário final e que continham no cenário inicial um consumo igual a zero.

As Figuras 6.18(a) e 6.18(b) apresentam os cenários que serão base para a ilustração do crescimento horizontal. Através desta análise é possível identificar não apenas o sentido de crescimento da rede de distribuição, mas como este crescimento ocorre.



(a) 2005

(b) 2006

Figura 6.18: Município de Rio Verde - 2005/2006

A Figura 6.19 apresenta o resultado da análise do crescimento horizontal. Observe que apesar do cenário de 2006 apresentar um crescimento quando comparado com o cenário de 2005, ele não possui algumas quadrículas que estavam presentes no cenário anterior.

A grande maioria das quadrículas que compõem o crescimento possuem o perfil rural, o que está de acordo com as Tabelas 6.11 e 6.12 que mostram um crescimento deste setor.

Houve também a identificação de quadrículas com perfil residencial e comercial no perímetro urbano. Entretanto, o real crescimento do perímetro urbano será melhor analisado com o crescimento vertical, pois normalmente o crescimento nestas áreas utiliza-se de transformadores já existentes ou instalados em uma região próxima aos mesmos.

As quadrículas que representam o crescimento horizontal são apresentadas com a classificação atribuída no cenário final. Com isso não é constatado apenas o sentido do crescimento, mas também a forma desse crescimento.

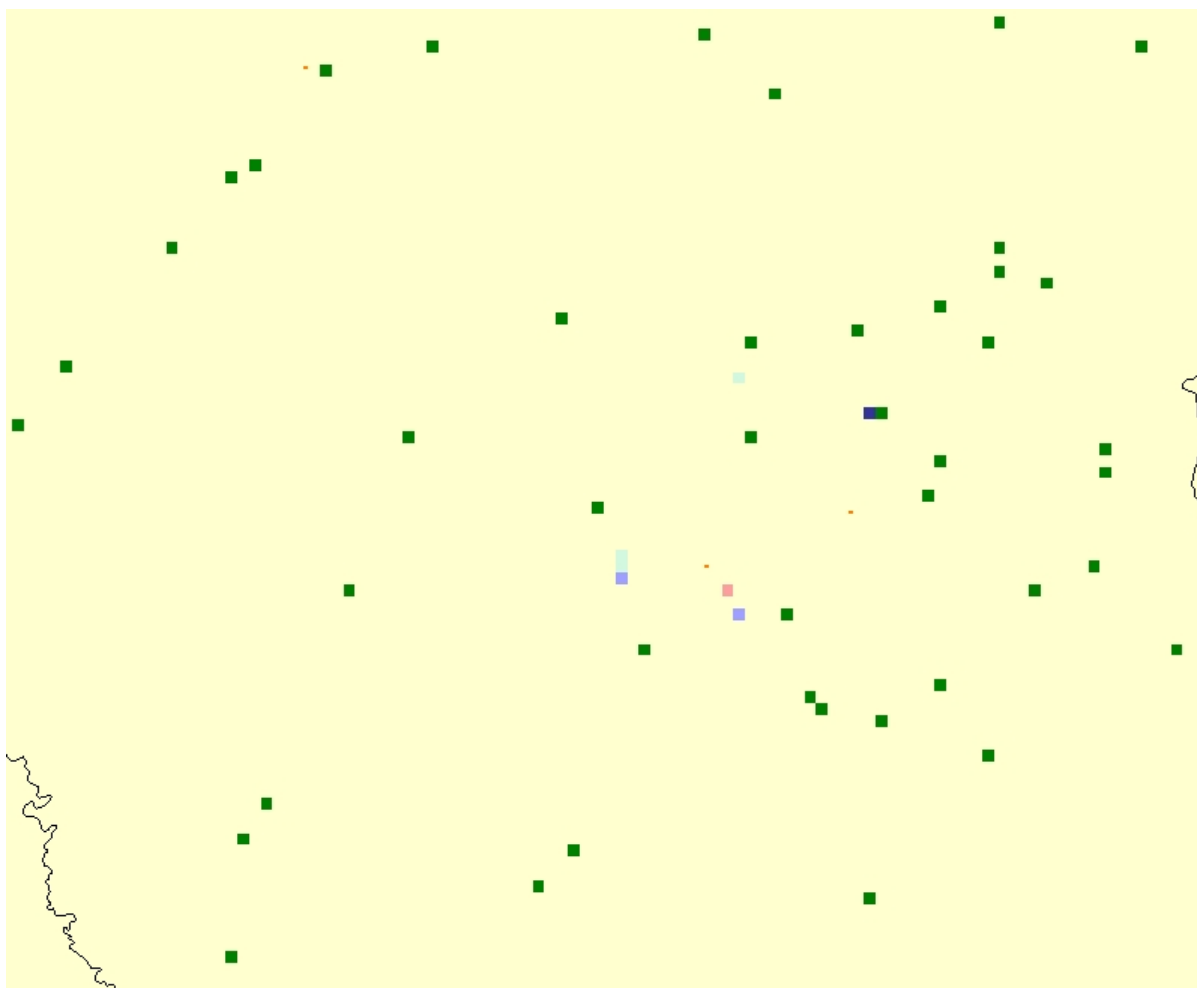


Figura 6.19: Crescimento Horizontal de Rio Verde - 2005/2006

6.6.5 Crescimento vertical 2005/2006

Neste processo, além dos cenários inicial e final, o planejador deve informar a taxa mínima de crescimento, a qual determina o crescimento de consumo mínimo que a quadrícula deve possuir para ser identificada.

São desconsideradas neste processo as quadrículas identificadas no crescimento horizontal, sendo analisadas apenas aquelas que já continham consumo no cenário inicial. Através desta análise, são identificadas as regiões com aumento de consumo, as quais podem ou não mudar seu perfil de consumo dependendo da forma de ocupação realizada neste crescimento.

A Figura 6.20 apresenta o resultado do processo de análise do crescimento vertical utilizando uma taxa mínima de crescimento de 10%.

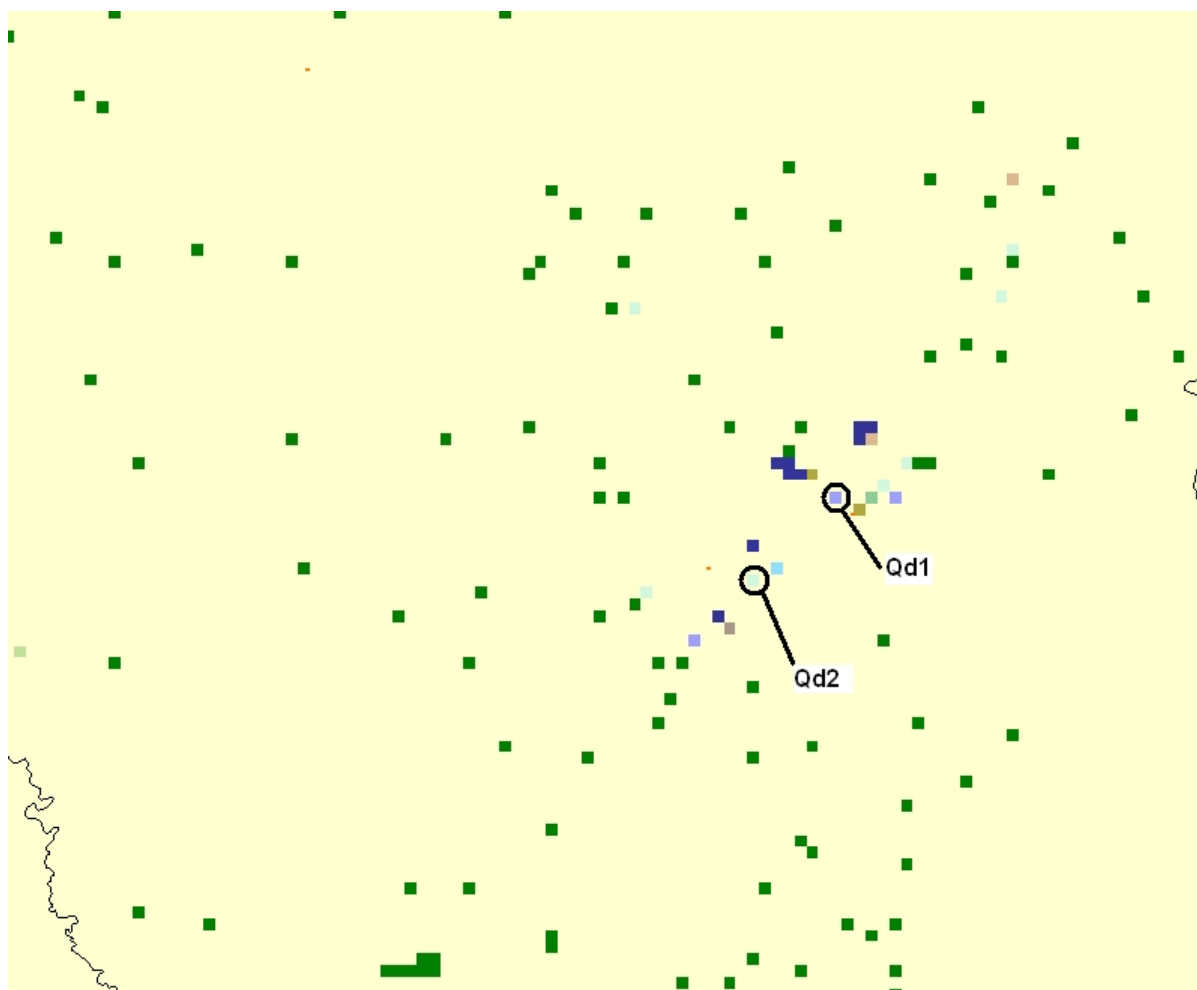


Figura 6.20: Crescimento Vertical de Rio Verde - 2005/2006

No resultado desta análise foram identificadas quadriculas de várias classes, com uma presença considerável do perímetro urbano. Algumas destas quadriculas mudaram de classe do cenário inicial para o final. Para constatar essa mudança serão analisadas duas quadriculas, as quais estão destacadas na Figura 6.20.

A quadricula denominada Qd1 possuía no cenário de 2005 um perfil residencial e comercial e era classificada como pertencente ao grupo 12. A Figura 6.21 mostra a consulta desta quadricula no cenário de 2005.

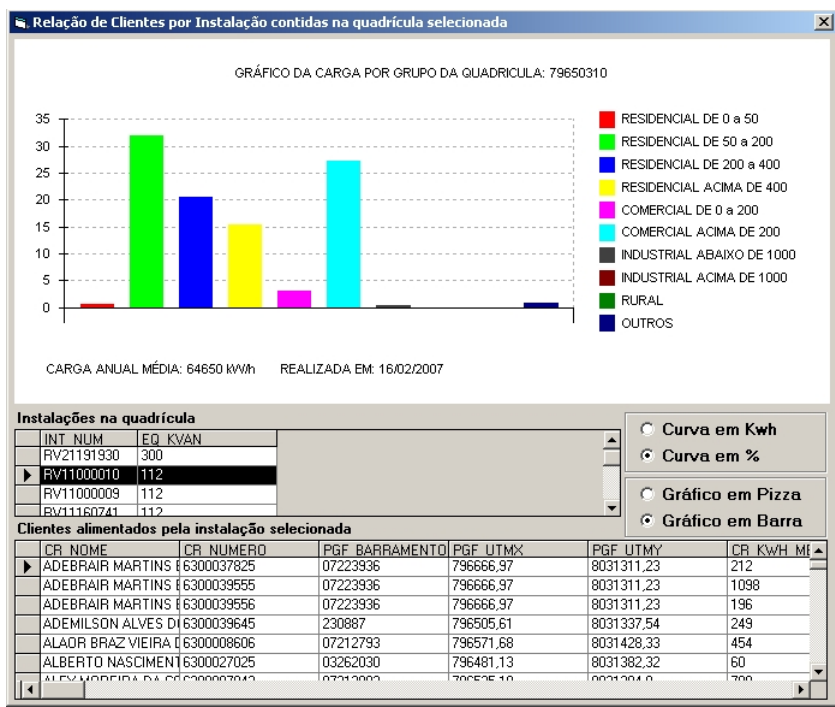


Figura 6.21: Consulta da quadricula Qd1 - Cenário 2005

Já no cenário de 2006 houve a inclusão de clientes comerciais com um alto consumo médio, o que mudou o perfil de consumo da quadricula. Ela passou a ter um consumo predominantemente comercial e com isso, passou a fazer parte do grupo 4.

A Figura 6.22 mostra a consulta desta quadricula no cenário de 2006 e identifica uma cooperativa que foi instalada na região em 2006, a qual é a principal responsável pela mudança de perfil de consumo desta área.

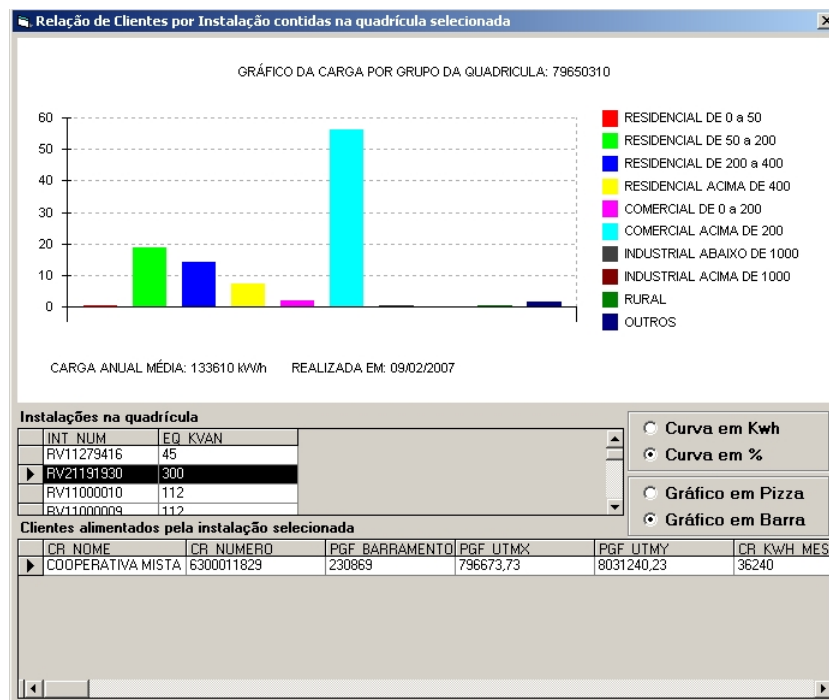


Figura 6.22: Consulta da quadricula Qd1 - Cenário 2006

A quadricula Qd2, entretanto, possuía em 2005 um perfil predominantemente comercial e era classificada no grupo 4. A Figura 6.23 mostra a consulta desta quadricula no cenário de 2005.

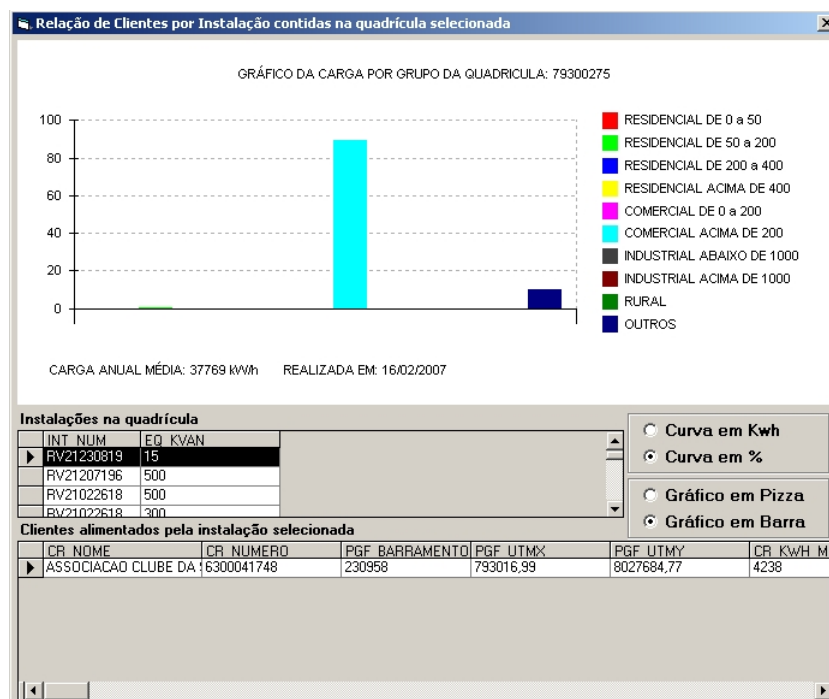


Figura 6.23: Consulta da quadricula Qd2 - Cenário 2005

Em 2006 foi instalada uma indústria na região da quadrícula mudando seu perfil predominante para industrial e conseqüentemente mudando sua classificação para o grupo 2. A seguir é apresentada a consulta desta quadrícula para o cenário de 2006 (Figura 6.24).

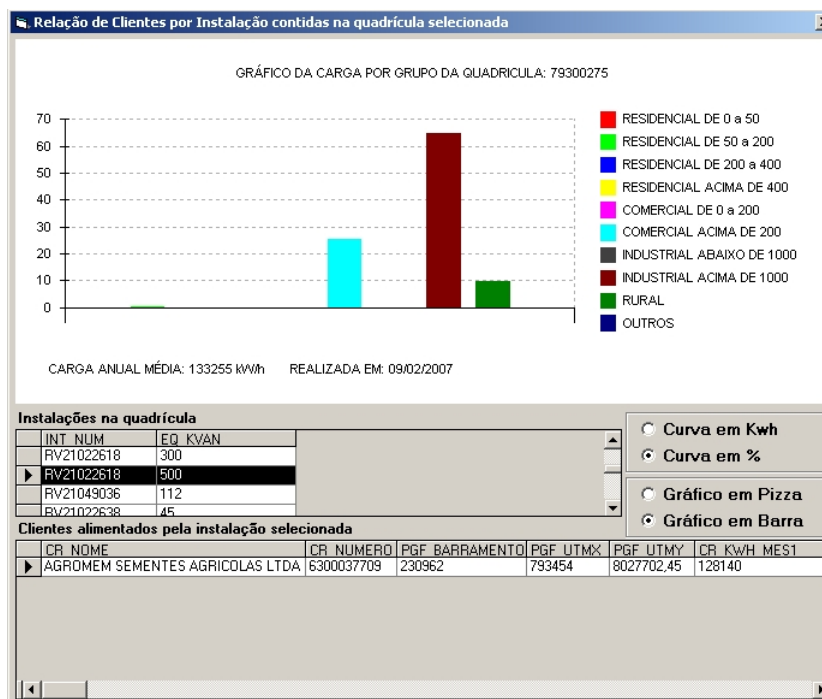


Figura 6.24: Consulta da quadrícula Qd2 - Cenário 2006

Assim, a análise do crescimento vertical identifica não apenas regiões com crescimento de consumo, mas também como esse crescimento afeta o perfil destas regiões.

7 *Conclusão*

O objetivo deste trabalho foi criar um programa, denominado GeoCom, que utilizando das funções do Sistema de Informação Geográfica (SIG) e de técnicas de classificação Kohonen emanel, K-Médias e Fuzzy C-Médias criasse um ambiente favorável à análise comercial de empresas de distribuição de energia elétrica que possuem um levantamento georeferenciado de sua rede de distribuição. Também foi proposta a comparação dos métodos de agrupamento e classificação empregados.

Os trabalhos encontrados na literatura especializada não dão muita importância ao processo de análise da distribuição do consumo dos clientes na área de concessão das companhias de distribuição de energia elétrica, mas sim ao estudo de previsão de carga. Os trabalhos que utilizaram do SIG para análise do setor elétrico não propuseram, em sua maioria, o desenvolvimento de uma ferramenta independente e sim utilizaram de softwares comerciais como ArcView e AutoCad, por exemplo.

Para realizar a análise do desempenho das técnicas empregadas e mostrar o funcionamento do sistema desenvolvido, foram utilizados os índices de medida intra-grupo (IAG) e inter-grupo (IEG) sobre os dados de consumo dos municípios de Anápolis (2006), Aparecida de Goiânia (2006) e Rio Verde (2005/2006).

Para a obtenção dos dados de consumo dos municípios, foi desenvolvida uma metodologia de análise espacial de mercado. Nesta metodologia, foram definidas dez faixas de consumo e o consumo de cada clientes foi agrupado em sua faixa específica. Estes clientes foram agrupados respeitando sua localização espacial através de quadrículas de 500x500 m.

O método Fuzzy C-Médias obteve o melhor resultado em todos os estudos de caso, formando grupos com um número de quadrículas mais distribuídos que os demais métodos. Além disso, este método, por usar um grau de pertinência para indicar o quanto um centróide representa uma determinada entrada, permite extrair conhecimento além do agrupamento produzido. Os melhores resultados neste método foram obtidos com um índice de fuzzificação igual a 1,25.

O método K-Médias obteve valores próximos ao Fuzzy C-Médias com um tempo de proces-

samento muito inferior e com o mesmo critério de parada. Isso se deve ao seu simples processo de atualização dos centróides, o qual utiliza da média aritmética.

O método Kohonen em Anel apresentou os piores resultados tanto nos índices quanto no tempo de processamento. Entretanto, este método, devido à ação de seu vetor de atualização de pesos, gerou grupos com centróides próximos, possibilitando identificar clientes com consumo ligeiramente diferentes. Através dos gráficos dos centróides gerados no agrupamento, foi confirmada a topologia em anel, já que o perfil da curva do último grupo é semelhante ao do primeiro.

Em todos os métodos foi utilizada a distância Euclidiana para medir a distância entre os padrões do conjunto de treinamento e os centróides dos grupos. Assim o que diferencia cada método é a forma como os centróides são atualizados, produzindo com isso informações diferentes como produto da classificação.

No município de Rio Verde foi feita uma análise do crescimento horizontal e vertical. Pôde ser visto, através do GeoCom, como e em que sentido ocorreu o crescimento de consumo na região. Foi analisada também a mudança de classe de determinadas quadrículas de um cenário para outro, ressaltando a atuação dos clientes na área estudada.

Apesar da utilização dos índices IAG e IEG, verificou-se que estes não são suficientes para se realizar uma medida de qualidade da classificação, necessitando assim que se realize também uma análise gráfica através do GeoCom.

Em termos de trabalho futuro pode ser feito o desenvolvimento do sistema GeoCom, possibilitando o agrupamento e classificação utilizando também a demanda como parâmetro da classificação.

Proporcionar também a utilização de outros índices de avaliação do desempenho do agrupamento e classificação para melhorar a comparação entre as técnicas, bem como possibilitar a escolha de outras medidas de distância objetivando produzir resultados mais satisfatórios.

Como visto em vários trabalhos na literatura especializada, o desenvolvimento de uma ferramenta de previsão de carga, agregando-se informações como planejamento urbano, uso do solo e distância das estradas às quadrículas seria de grande validade para o sistema GeoCom, produzindo uma ferramenta de análise completa da rede.

Referências

- ALVES, A. C. *Algoritmos Genéticos Aplicados ao Planejamento da Distribuição de Energia Elétrica em Curitiba e Região Metropolitana*. 89 p. Dissertação (Mestrado) — Universidade Federal do Paraná, Curitiba, Brasil, 2002.
- ANEEL. *Procedimentos de Distribuição - Planejamento da Expansão da Distribuição*. Agência Nacional de Energia Elétrica, 2006. Disponível em: <<http://www.aneel.gov.br/82.htm>>. Acesso em: 31/08/2006.
- BANDRÃO, T. P. V. *Um Estudo de Modelos Matemáticos para Expansão de Redes de Distribuição de Energia Elétrica*. 144 p. Dissertação (Mestrado) — Universidade Federal do Paraná, Curitiba, Brasil, 2004.
- BATISTA, L. B. *Aprendizagem e Recuperação de Imagens utilizando Mapas Auto-Organizáveis e Representação Log-Polar*. 113 p. Dissertação (Mestrado) — Universidade Federal de Campina Grande, Campina Grande, Brasil, 2004.
- BEZDEK, J. C. *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press, 1981.
- BITTENCOURT, J. R.; OSÓRIO, F. Fuzzyf - fuzzy logic framework: Uma solução software livre para o desenvolvimento, ensino e pesquisa de aplicações de inteligência artificial multiplataforma. *III Workshop sobre Software Livre, Porto Alegre*, 2002.
- BOULAXIS, N. G.; PAPADOPOULOS, M. P. Optimal feeder routing in distribution system planning using dynamic programming technique and gis facilities. *IEEE Transactions on Power Delivery*, 2002.
- CAMASTRA, F.; COLLA, A. M. Short-term load forecasting based on correlation dimension estimation and neural nets. *Congres ICANN '97 : artificial neural networks, Lausanne*, 1997.
- CARVALHO, A. C. P. L. F.; BRAGA, A. P.; LUDERMIR, T. B. *Fundamentos de Redes Neurais Artificiais*. Rio de Janeiro, RJ: RIO, 1998.
- CASANOVA, M. A. et al. *Banco de dados Geográficos*. Curitiba, PR: MundoGEO, 2005.
- CAVALCANTI, P. D. O. B. *GEOTUR: Ferramenta de Geoprocessamento Turístico*. Centro Universitário de João Pessoa, 2005. Disponível em: <<http://www.unipe.br/graduacao/computacao/projetos/tcc20052/GEOTUR%20-%20Ferramenta%20de%20Geoprocessamento%20Tur%20EDstico.pdf>>. Acesso em: 01/12/2006.
- CELG. *Relação Municipal de consumidores e consumo por classe - 2006*. [S.l.]: Departamento de Comercialização, 2007.

- CHOW, M.-Y.; ZHU, J.; TRAM, H. Application of fuzzy multi-objective decision marking in spatial load forecasting. *IEEE Transactions on Power Systems*, 1998.
- DORADO, E. D.; MIGUEZ, E.; CIDRÁS, J. Design of large rural low-voltage networks using dynamic programming optimization. *IEEE Transaction on Power System*, 2001.
- EMBRAPA. *EMBRAPA - Satélites de Monitoramento*. Empresa Brasileira de Pesquisa Agropecuária, 2006. Disponível em: <<http://www.sat.cnpm.embrapa.br/satelite/ikonos.html>>. Acesso em: 10/09/2006.
- ESRI. *Diagram of MapObjects 2.3 ActiveX Control and ActiveX Automation objects*. Environmental Systems Research Institute, 2004. Disponível em: <<http://downloads.esri.com/support/documentation/mo/MO23Diag.pdf>>. Acesso em: 12/11/2006.
- ESRI. *ArcSDE - Technology Overview*. Environmental Systems Research Institute, 2006. Disponível em: <<http://www.esri.com/software/arcgis/arcsgis/about/overview.html>>. Acesso em: 12/11/2006.
- FIDALGO, R. N. *Uma Infra-estrutura para Integração de Modelos, Esquemas e Serviços Multidimensionais e Geográficos*. Tese (Doutorado) — UFPE, Pernambuco, Brasil, 2005.
- FREDERICK, P. *Visualização Eficiente de Objetos Gráficos*. 89 p. Dissertação (Mestrado) — Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, Brasil, 1999.
- GUEMBAROVSKI, R. H. *Utilização de um Sistema de Informações Geográficas para a adequação do carregamento Elétrico de Transformadores*. 114 p. Dissertação (Mestrado) — Universidade Federal de Santa Catarina - Programa de Pós-Graduação em Engenharia de Produção, Florianópolis, Brasil, 1999.
- HAYKIN, S. *Redes neurais: princípios e práticas*. Porto Alegre, RS: Bookman, 2001.
- HWANG, K. J.; KIM, G. W. Spacial load forecasting model for eletrical distribution planning. *Korus 2004*, 2004.
- IBGE/CIDADES. *IBGE - Cidades*. Instituto Brasileiro de Geografia e Estatística, 2007. Disponível em: <<http://www.ibge.gov.br/cidadesat/default.php>>. Acesso em: 10/01/2007.
- IBGE/PAÍSES. *IBGE - Países*. Instituto Brasileiro de Geografia e Estatística, 2007. Disponível em: <<http://www.ibge.gov.br/paisesat/>>. Acesso em: 16/04/2007.
- INPE. *Anatomia de Sistemas de Informação Geográfica*. Instituto Nacional de Pesquisas Espaciais, 1996. Disponível em: <<http://www.dpi.inpe.br/geopro/livros/anatomia.pdf>>. Acesso em: 01/12/2006.
- KOHONEN, T. *Self-Organizing Maps*. Berlim: Springer-Verlag, 1997.
- LIBANIO, M.; LIMA, W. S. *Sugestões para Implementação do Módulo de Planejamento*. [S.l.], 2004.
- LOTUFO, A. D. P.; MINUSSI, C. R. Electric power systems load forecasting: A survey. *IEEE Power Tech '99 Conference, Budapest, Hungary*, 1999.

- MIRANDA, V.; MONTEIRO, C. Fuzzy inference applied to spatial load forecasting. *IEEE Power Tech '99 Conference, Budapest, Hungary*, 1999.
- QUEIROZ, R. R.; CÂMARA, R. *Arquitetura de Sistemas de Informação Geográfica*. Instituto Nacional de Pesquisas Espaciais, 2006. Disponível em: <<http://www.dpi.inpe.br/gilberto/livro/introd/cap3-arquitetura.pdf>>. Acesso em: 05/09/2006.
- RIEDEL, P. S. *Sistemas de Informação Geográfica*. UNESP - Curso de Pós-Graduação em Geociências, 2006. Disponível em: <<http://www.rc.unesp.br/igce/geologia/GAA01048/aulas.html>>. Acesso em: 10/11/2006.
- ROSS, T. J. *Fuzzy Logic with Engineering Applications*. Estados Unidos: McGraw-Hill, 1995.
- SILVA, A. B. *Sistemas de Informações Geo-Referenciadas - Conceitos e fundamentos*. Campinas, Brasil: Editora Unicamp, 2003.
- SILVA, J. C. M. "Ranking" de Dados Multidimensionais Usando Mapas Auto-Organizáveis e Algoritmos Genéticos. Dissertação (Mestrado) — Universidade Federal de Goiás, Goiânia, Brasil, 2000.
- SPERANDIO, M. *Classificação de Conjuntos Consumidores de Energia Elétrica via Mapas Auto-Organizáveis e Estatística Multivariada*. 105 p. Dissertação (Mestrado) — Universidade Federal de Santa Catarina - Programa de Pós-Graduação em Engenharia Elétrica, Florianópolis, Brasil, 2004.
- SPRING. *Spring - Tutorial de Geoprocessamento*. Instituto Nacional de Pesquisas Espaciais, 2006. Disponível em: <<http://www.dpi.inpe.br/spring/portugues/tutorial/introducaogeo.html>>. Acesso em: 09/09/2006.
- TSUTOMU, L. H. *Técnicas de Apresentação de Dados em Geoprocessamento*. 114 p. Dissertação (Mestrado) — Ministério da Ciência e Tecnologia - Instituto Nacional de Pesquisas Espaciais, Florianópolis, Brasil, 1997.
- VASCONCELOS, N. A. P. *Mapas Auto-Organizativos e suas Aplicações*. 97 p. Dissertação (Mestrado) — Universidade Federal do Rio de Janeiro, COPPE, Rio de Janeiro, Brasil, 2000.
- YONAMINE, F. et al. *Aprendizado não Supervisionado em Domínios Fuzzy - Algoritmo Fuzzy C-Means*. 2002.
- YU, Y. et al. Models and methods for urban distribution planning. *Power System Technology*, 2000, 2000.
- ZADEH, L. A. *Fuzzy logic-Incorporating Real-World Vagueness*. Center for Spatially Integrated Social Science, 2006. Disponível em: <<http://www.csiss.org/classics/content/68>>. Acesso em: 03/11/2006.
- ZAMORA, F. H. P. *Previsão Espacial de carga em sistemas de distribuição Utilizando Técnicas de Inteligência Artificial e Plataforma GIS*. Tese (Doutorado) — USP, São Paulo, Brasil, 2003.
- ZANON, D. M.; LIMA, W. S. *Mineração de Dados Aplicada à Análise de Curvas de Carga*. 138 p. Dissertação (Mestrado) — Universidade Federal de Goiás - Escola de Engenharia Elétrica e de Computação, Goiânia, Brasil, 2003.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)