

MARCOS VINICIUS SANCHES ABREU

**PROPOSTA METODOLÓGICA PARA REDEFINIÇÃO DE
LIMITES CENSITÁRIOS A PARTIR DA ESTATÍSTICA
ESPACIAL E INTEGRAÇÃO DOS DADOS DO CENSO AO
CADASTRO TERRITORIAL MULTIFINALITÁRIO**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Engenharia Civil, para obtenção do título de *Magister Scientiae*.

VIÇOSA
MINAS GERAIS - BRASIL
2011

**Ficha catalográfica preparada pela Seção de Catalogação e
Classificação da Biblioteca Central da UFV**

T

A162p
2011

Abreu, Marcos Vinicius Sanches, 1985-

Proposta metodológica para redefinição de limites censitários a partir da estatística espacial e integração dos dados do censo ao cadastro territorial multifinalitário / Marcos Vinicius Sanches Abreu. – Viçosa, MG, 2011. xiv, 148f. : il. (algumas col.) ; 29cm.

Orientador: Joel Gripp Junior.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 144-148

1. Cadastros. 2. Censo demográfico. 3. Análise espacial (Estatística). 4. Sistemas de informação geográfica - Processamento de dados. I. Universidade Federal de Viçosa. II. Título.


CDD 22. ed. 621.3678


MARCOS VINICIUS SANCHES ABREU

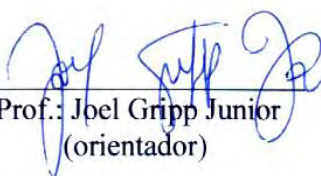
**PROPOSTA METODOLÓGICA PARA REDEFINIÇÃO DE
LIMITES CENSITÁRIOS A PARTIR DA ESTATÍSTICA
ESPACIAL E INTEGRAÇÃO DOS DADOS DO CENSO AO
CADASTRO TERRITORIAL MULTIFINALITÁRIO**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Engenharia Civil, para obtenção do título de Magister Scientiae.

Aprovada: 25 de novembro de 2011.


Prof^a.: Andrea Flávia Tenório Carneiro


Prof^a.: Aline Werneck B. de Carvalho


Prof.: Joel Gripp Junior
(orientador)

*Aos meus pais Marcos e Ana Rita
À minha irmã, Maryanna
E à minha esposa Eliane*

*Mais do que máquinas, precisamos de humanidade
Mais que inteligência, precisamos de afeição e doçura
(Charlin Chaplin)*

AGRADECIMENTOS

Agradeço primeiramente a Deus, pela inteligência e capacidade de realizar meus sonhos e nunca me abandonar.

Aos meus pais, Marcos e Ana Rita, pelo amor e apoio incondicionais, fazendo mais do que podem para me ajudar a alcançar meus objetivos. Minha eterna gratidão à minha mãe, pelas noites viradas nas correções ortográficas.

A toda a minha família, por me ajudarem em minha formação pessoal, espiritual e profissional e, em especial, a minha irmã, com quem dividi e ainda divido esperanças, sonhos e conquistas.

Ao meu orientador, professor Joel Gripp Júnior, por caminhar junto a mim neste trabalho, assumindo-o e acreditando na minha capacidade para conduzi-lo.

À minha comissão orientadora, Nilcilene das Graças Medeiros e Nerílson Terra Santos, por todas as dicas que sempre somaram ao proposto inicialmente neste trabalho.

Às professoras Andrea Flávia Carneiro Tenório e Aline Werneck Barbosa de Carvalho, por terem aceitado participar da banca avaliadora com tanta espontaneidade. Agradeço, principalmente, pela sensibilidade e destreza aplicadas na avaliação e na colocação das considerações.

Aos colegas da pós em Informações Espaciais da UFV, pelas trocas de experiências e pelo apoio dado ao meu trabalho, fundamentais para o desenvolvimento das idéias apresentadas.

Aos professores e funcionários do Departamento de Engenharia Civil – UFV, pelos serviços prestados e pelo zelo em seus trabalhos. Destaque especial é dado aos professores do setor de Engenharia de Agrimensura e Cartográfica, pela cumplicidade nestes dois anos de trabalho.

Ao futuro grande engenheiro Renato Viana, pela ajuda na montagem dos dados utilizados nesta pesquisa. Ao grande matemático-amigo-irmão Lucas Carvalho Silva, pelas orientações e pela ajuda técnica.

À família iPlanus e a meu amigo Anderson Meira, pela parceria e pelas conversas sempre produtivas, incentivando-me cada vez mais.

Aos meus irmãos da República Casa do Renato, pelo companheirismo e por me ajudar a tornar a vida mais leve durante todos esses anos.

A todos os meus amigos, que mais ou menos presentes, estiveram sempre ao meu lado.

Por fim, agradeço à minha metade, Eliane, minha esposa, por ser um Porto Seguro e colo de Deus aqui na Terra. Obrigado por me apoiar sempre.

SUMÁRIO

RESUMO	viii
ABSTRACT.....	x
1. INTRODUÇÃO	1
2. OBJETIVOS E JUSTIFICATIVAS.....	5
2.1. Objetivos específicos.....	5
2.2. Justificativas	6
3. REVISÃO DE LITERATURA	8
3.1. CADASTRO TERRITORIAL MULTIFINALITÁRIO	8
3.1.1. Cartografia Cadastral	11
3.1.2. Cadastros Temáticos.....	12
3.2. GEOPROCESSAMENTO APLICADO	14
3.2.1. Modelagem de Dados, Ontologia e Banco de Dados Geográficos	15
3.2.2. Geoestatística Aplicada.....	20
3.2.3. Segmentação de Superfícies	26
3.2.4. Avaliação Estatística e Propagação de Erros	29
3.2.5. Indicadores de Autocorrelação Espacial.....	33
3.3. INFORMAÇÕES SOCIOECONÔMICAS.....	38
3.3.1. Dados e Indicadores Socioeconômicos.....	38
3.3.2. Censos Demográficos.....	40
3.3.3. Censo Demográfico do IBGE	41
3.3.3.1. Vantagens na utilização de setores censitários na caracterização socioeconômica intraurbana.....	43
3.3.3.2. Desvantagens na utilização de setores censitários na caracterização socioeconômica intraurbana.....	45
3.4. ZONEAMENTO TERRITORIAL SOCIOECONÔMICO	48
3.4.1. Relações de Vizinhança na Configuração dos Territórios	48
3.4.2. Problema de Unidades de Áreas Modificáveis e Zoneamento Territorial.....	50
4. MATERIAIS E MÉTODOS	54
4.1. ÁREA DE ESTUDO.....	56
4.2. MATERIAIS UTILIZADOS.....	57
4.3. METODOLOGIA.....	58

4.3.1. Avaliação da Homogeneidade Socioeconômica dos Setores Censitários.....	58
4.3.2. Geração das zonas homogêneas	64
4.3.2.1. Geração das Superfícies Contínuas	69
4.3.2.2. Discretização da Superfície Contínua.....	71
4.3.2.3. Determinação das Fronteiras Socioeconômicas	72
4.3.2.4. Zoneamento Homogêneo	72
4.3.2.5. Determinação dos Setores Censitários a partir do Zoneamento Homogêneo	74
4.3.3. Avaliação da Homogeneidade Interna dos Setores Propostos.....	78
4.3.4. Proposta de integração das regiões socioeconômicas e Setores Censitários ao Cadastro Territorial Multifinalitário.....	79
5. RESULTADOS E DISCUSSÕES	82
5.1. Avaliação da Homogeneidade Socioeconômica dos Setores Censitários	82
5.2. Avaliação das Zonas Homogêneas	98
5.2.1. Geração das Superfícies Contínuas.....	99
5.2.2. Discretização da Superfície Contínua	104
5.2.3. Determinação das Fronteiras Socioeconômicas.....	107
5.2.4. Zoneamento Homogêneo	111
5.2.5. Determinação dos Setores Censitários a partir do Zoneamento Homogêneo	112
5.3. Avaliação da Homogeneidade Interna dos Setores Propostos	114
5.4. Proposta de integração das regiões socioeconômicas e Setores Censitários ao Cadastro Territorial Multifinalitário	125
5.4.1. Integração dos Setores Censitários ao CTM	125
5.4.1.1. Compatibilização dos Limites Censitários aos Parcelares.....	128
5.4.2. Considerações finais sobre a proposta de integração	140
6. CONCLUSÕES E RECOMENDAÇÕES.....	142
REFERÊNCIAS BIBLIOGRÁFICAS	144

RESUMO

ABREU, Marcos Vinicius Sanches, M.Sc., Universidade Federal de Viçosa, novembro de 2011. **Proposta metodológica para redefinição de limites censitários a partir da estatística espacial e integração dos dados do censo ao cadastro territorial multifinalitário.** Orientador: Joel Gripp Junior, Coorientadores: Nilcilene das Graças Medeiros e Nerilson Terra Santos.

O presente trabalho busca relacionar duas das principais fontes de dados espaciais municipais: o Cadastro Territorial Multifinalitário (CTM), o qual é responsável por informações físicoterritoriais, sendo um inventário territorial oficial de um município, armazenando dados espaciais de diversos temas de interesse; e o Censo Demográfico do IBGE, o qual disponibiliza dados demográficos e socioeconômicos agregados por setores censitários, sendo estes a menor unidade para fins de coleta de dados com limites físicos identificáveis em campo, abrangendo um conjunto de cerca de 250 domicílios cada. Destaca-se que ambos, muitas vezes, divergem numa base de dados municipais, ou não possuem compatibilidade necessária à realização de análises espaciais precisas. Por ser o CTM o principal gerenciador e integrador de informações municipais, essa dissertação propõe a compatibilização da estrutura geográfica do censo à estrutura proposta pelo CTM. Para isso, num primeiro momento, realizou-se um estudo sobre o grau de heterogeneidade socioeconômica interna aos setores censitários, cuja definição dos limites não leva em consideração critérios de homogeneidade da população inserida, acarretando em informações errôneas, podendo mascarar a realidade. A segunda parte envolveu a aplicação de ferramentas da Estatística Espacial associadas ao Geoprocessamento para a obtenção de áreas homogêneas internamente, as quais baseiam uma nova proposição de definição de limites censitários, sendo realizados testes de agregação de dados, buscando minimizar os Efeitos de Zoneamento dos Problemas de Unidades de Áreas Modificáveis (*MAUP*). Por último, fez-se a proposta conceitual para a integração sistemática dos limites censitários ao CTM, destacando-se a sugestão de um novo conceito dado ao setor censitário. Como resultados, verificou-se que indicadores socioeconômicos provenientes de dados censitários agregados por setores são sensíveis à variabilidade interna dos próprios indicadores. Com o segundo teste,

percebeu-se que tal sensibilidade pode ser minimizada a partir de agregações de dados homogêneos entre si, diminuindo os Efeitos de Zoneamento do MAUP e validando a proposta de detecção de regiões homogêneas. Quanto à integração Censo Demográfico – CTM, tem-se um conjunto de regras e conceitos sugeridos para que seja possível sua realização. Como contribuição, destaca-se a necessidade de sistematizar a utilização dos dados censitários, principalmente sob a nova ótica do CTM, tal como proposto pelas Diretrizes Nacionais para o Cadastro, publicadas pela portaria nº 511 de 2009 do Ministério das Cidades.

ABSTRACT

ABREU, Marcos Vinicius Sanches, M.Sc., Universidade Federal de Viçosa, November, 2011. **Methodological proposal for redefinition of boundaries of census tract with spatial statistic and integration of census data to Multipurpose Cadastre.** Adviser: Joel Gripp Junior, Co-Advisers: Nilcilene das Graças Medeiros and Nerilson Terra Santos.

This work seeks to relate the two main sources of municipal spatial data: Multipurpose Cadastre (MC), which is responsible for physical and territorial information, can be an official inventory of a municipality, storing spatial data from various sources of interest, and the demographic census, which provides demographic and socioeconomic data aggregated by census tracts, which are the smallest unit with identifiable physical boundaries in the field, covering a range of about 250 homes each. It is noteworthy that both often diverge in a local database or do not have compatibility necessary to perform accurate spatial analysis. As the MC is the leading manager and integrator of local information, this paper proposes the compatibility of the geographical structure of the census to the structure proposed by MC. To do so, at first, there was a study on the degree of socioeconomic heterogeneity internal to census tracts, whose definition of the limits does not take into account the criteria of homogeneity of the population included, resulting in loss of information and may mask the reality. The second part involved the application of tools of spatial statistics associated with the GIS to obtain internally homogeneous areas, which are the base of proposing a new definition of census boundaries, and testing for data aggregation in order to minimize the effects of Zoning of the Modifiable Areal Unit Problems (MAUP). Finally, it was made the conceptual proposal for the systematic integration of the census boundaries with the MC, especially being suggested a new concept to the census tract. As a result, it was found that socioeconomic indicators from census data for aggregated sectors are sensitive to the variability of the same. With the second test, it was realized that such sensitivity can be minimized from aggregations of homogeneous data with each other, reducing the effects zoning of the MAUP and validating the proposed detection of homogeneous regions. Regarding the integration of the Census to MC, we have a set of rules and

concepts suggested that its realization is possible. As a contribution, we highlight the need to systematize the use of census data, particularly under the new perspective of the MC, as proposed by the Guidelines for the National Cadastre, published by decree No. 511 of 2009 of the Ministry of Cities.

1. INTRODUÇÃO

Diante do intenso crescimento urbano das cidades brasileiras nos últimos anos, percebe-se cada vez mais a necessidade de processos qualificados de planejamento e gestão urbana. Seja para planejar ou gerenciar uma cidade com eficiência e eficácia, ou para cumprir obrigações legais impostas ao município, a administração pública necessita de informações confiáveis e disponibilizadas de maneira rápida e dinâmica (GARCIA, 2007).

Informações organizadas e bem gerenciadas se tornam a principal ferramenta para o apoio às decisões, pois delas surge o conhecimento. Não há ações sobre determinado fenômeno sem seu prévio conhecimento. Além disso, o equilíbrio urbano depende diretamente da interação entre processos socioeconômicos, opções de planejamento e de políticas públicas. Não conhecer a dinâmica desses processos se torna um obstáculo ao desenvolvimento urbano.

Tal conhecimento norteia o município na formulação e implementação de políticas públicas de desenvolvimento urbano e regional, estabelecendo diretrizes para a garantia do respeito aos direitos do cidadão relacionados à vida urbana. Aspectos que devem ser garantidos, como direito à moradia digna, à terra urbanizada, ao saneamento ambiental, ao trânsito seguro, além de meios de geração de renda e acesso a educação, saúde, informação, cultura, esporte, lazer, segurança pública, trabalho e participação (BRASIL, 2004) necessitam, portanto, de informações consistentes e atualizadas.

Nota-se, assim, a importância de uma qualificada gestão da informação municipal, para que se conheçam os problemas e as vocações da cidade, permitindo-se a minimização dos aspectos negativos e maximização dos aspectos positivos. Desse modo, destacam-se informações quantitativas e qualitativas referentes a infraestrutura, mobilidade urbana, aspectos de educação, de renda e de saúde, entre outros.

Ressalta-se que na maioria das vezes a informação municipal é referente a fenômenos distribuídos no espaço, podendo ser chamada também de Informação Territorial, Informação Espacial ou Informação Geográfica. Estima-se que 60% da informação utilizada no setor público sejam georreferenciáveis (IBGE, 2008c). Assim, o aspecto *localização* aparece como fundamental no processo de leitura da cidade.

Ramos *et al* (2007) discutem a necessidade de saber *onde* se encontram os problemas socioambientais para que as políticas públicas sejam bem orientadas. Os

autores ainda destacam a necessidade de se identificar e conhecer os diferentes *territórios intraurbanos*, principalmente aqueles relacionados às mesmas condições socioeconômicas populacionais.

Ramos (2002) explica que o significado de território se aproxima ao de comunidade que, citando Milton Santos, são “espaços seletivamente transformados pelas exigências funcionais da sociedade urbana”. Assim, um território não pode ser delimitado simplesmente por condicionantes físicos, logísticos ou administrativos, como bairros ou zonas, e sim por níveis de interação social da população inserida. Quanto maior essa interação, maior pode ser a homogeneidade interna em um território. Assim, encontrar regiões homogêneas socioeconomicamente é encontrar os territórios socioeconômicos.

Uma das formas de identificar a homogeneidade local é realizar a análise espacial de dados socioeconômicos. Definem-se análises espaciais como a mensuração de “propriedades e relacionamentos, levando em conta a localização espacial do fenômeno em estudo de forma explícita” (DRUK *et al*, 2004).

Ramos (2002) diz que, “como regra geral, os dados socioeconômicos são disponibilizados como características estatísticas de um determinado grupo de indivíduos habitantes de uma parcela ou zona geográfica”, que podem ser representados por bairros, distritos ou setores. Uma das fontes de dados dessa natureza é o censo realizado pelo IBGE, em que os mesmos podem ser disponibilizados por setores censitários.

Percebe-se, então, que a disponibilidade de informações socioeconômicas está, na maioria das vezes, condicionada a limites físicos e operacionais de levantamentos – logísticos, ou administrativos, como no caso dos setores censitários (RAMOS *et al*, 2007), não respeitando as relações socioespaciais populacionais que definem ou podem definir os territórios. Ramos *et al* (2007) afirmam que dados agregados por setores censitários não obedecem critérios de homogeneidade interna da população por eles apresentados.

Abreu (2009) cita que dados agregados por áreas estão sujeitos aos efeitos do *MAUP* (*Modifiable Areal Unit Problem* – Problemas de Unidades de Áreas Modificáveis), citando seus dois efeitos: escala e zoneamento. Com base no *MAUP*, Martin (1996) diz que restringir a compreensão do território com sua representação em

forma de zonas induz a erros interpretativos no que diz respeito à percepção da totalidade do fenômeno estudado.

Os efeitos do *MAUP* nunca poderão ser totalmente removidos. Assim, duas abordagens distintas são encontradas na literatura. Uma delas procura métodos e modelos matemáticos capazes de minimizar as instabilidades do problema. A outra, por sua vez, reconhece o problema e o associa à característica interna da zona, formulando critérios capazes de gerar subdivisões territoriais em função da homogeneidade e tamanho da população (RAMOS, 2002).

Outra fonte de dados populacionais, além do Censo Demográfico, é o cadastro socioeconômico populacional, como um dos temas do Cadastro Territorial Multifinalitário (CTM). Segundo as Diretrizes Para a Criação, Instituição e Atualização do Cadastro Territorial Multifinalitário (CTM) nos Municípios Brasileiros, publicadas em 8 de dezembro de 2009, o CTM pode ser definido como o inventário territorial oficial e sistemático do município. Ele é embasado no levantamento dos limites de cada parcela, sendo esta a menor unidade de cadastro, definida como uma parte contígua da superfície terrestre com regime jurídico único. Todas as outras unidades territoriais, como logradouros, praças, cursos d'água, glebas, etc. são definidas por uma ou mais parcelas.

Os artigos 21 e 22 das diretrizes supracitadas indicam que, para ser multifinalitário, o CTM deve ser modelado de forma a atender às necessidades de diferentes usuários, com base num sistema de referência único. Ao mesmo tempo, deve ser aberto à integração de diversos temas, como os cadastros econômico, físico, geoambiental, redes de infraestrutura, socioeconômico, etc., de acordo com as questões sociais, econômicas, administrativas ou outras que atendam aos municípios.

Assim sendo, as informações socioeconômicas também podem ser disponibilizadas em níveis parcelares. Alguns trabalhos sugerem alternativas para o cadastro socioeconômico. Loch e Erba (2007) recomendam a geoespacialização de pacientes e/ou alunos a partir dos endereços cadastrados nas instituições de saúde e ensino, respectivamente. Por sua vez, Gonçalves (2008) propõe modelos de diagramas de classes num Banco de Dados Geográficos direcionados a essas instituições, sendo cadastradas as unidades institucionais, no lugar de seus usuários.

Ainda nesse campo temático, no trabalho apresentado por Sumariva e Philips (2010), as informações socioeconômicas cadastradas se referiam também aos temas

enfrentados na comunidade pesquisada, como grau de satisfação de serviços públicos, precariedade de áreas de lazer, etc..

Percebe-se, portanto, que as informações socioeconômicas cadastradas dependem do motivo pelo qual se realiza o cadastro. Mesmo assim, questões básicas, como aspectos de renda, educação, demografia, entre outros, geralmente são levantadas, uma vez que são essenciais para o conhecimento da população e, conseqüentemente, das condições em que esta vive.

Certamente que dados no nível de parcelas podem retratar fielmente as condições socioeconômicas de um município, uma vez que estão na menor unidade espacial cadastral. No entanto, existem informações socioeconômicas que não podem ser tornadas públicas, devido à segurança da informação e à preservação do informante, como consta no artigo 18 das diretrizes citadas anteriormente.

Desse modo, é possível estruturar regiões homogêneas internamente, pelas quais dados socioeconômicos podem ser agregados e disponibilizados para os munícipes, garantindo as informações confidenciais constantes no CTM. Além disso, para fins de planejamento e gestão de políticas urbanas, dados agregados por zonas são suficientes para retratarem a população, desde que sejam homogêneos internamente.

Informações socioeconômicas podem ser um dos temas integrados ao CTM: Cadastro de Zonas Homogêneas Socioeconomicamente. O conceito de Zonas Homogêneas aqui apresentado é diferente daquele que se refere à Planta de Valores Genéricos (PVG) ou referente ao Planejamento de Transportes apresentado em Manzato e Silva (2007).

Desse modo, têm-se, ao final, zonas homogêneas que podem determinar a representação dos territórios socioeconômicos intraurbanos, objetivando auxiliar nos processos de planejamento urbano e gestão municipal, principalmente no tocante à elaboração e ao direcionamento de políticas públicas que busquem a diminuição ou a erradicação das desigualdades socioespaciais, conforme um dos princípios da Política Nacional de Desenvolvimento Urbano (BRASIL, 2004).

2. OBJETIVOS E JUSTIFICATIVAS

O presente trabalho propõe uma metodologia para identificar zonas homogêneas do ponto de vista socioeconômico, com o objetivo de redefinir os limites dos setores censitários e integrá-los ao Cadastro Territorial Multifinalitário, sendo inserido no tema Zoneamento Socioeconômico. São abordados aspectos relativos à homogeneidade interna de zonas geográficas, estudando propostas para minimização dos efeitos do *MAUP*.

A hipótese testada é a de que indicadores quantitativos contínuos espacialmente são capazes de representar condições socioeconômicas populacionais. A partir de sua representação computacional por modelo de dados matricial, podem ser identificadas regiões de mesmas características estatísticas, determinando a homogeneidade interna, cujas bordas são identificadas e, posteriormente, amarradas às parcelas cadastrais.

Busca-se, portanto, definir uma metodologia para a identificação de regiões homogêneas socioeconomicamente, gerando subsídios para um zoneamento socioeconômico populacional, o qual abalize a definição dos limites dos setores censitários, integrando-os ao Cadastro Territorial Multifinalitário

2.1. Objetivos específicos

Como objetivos específicos, destacam-se:

- Realizar estudos relacionados à geração de indicadores socioeconômicos através de dados censitários;
- Avaliar os efeitos de Zoneamento dos Problemas de Unidades de áreas Modificáveis nos dados censitários de Viçosa – MG;
- Propor a utilização de interpoladores geoestatísticos para indicadores socioeconômicos, buscando a geração de uma superfície contínua para os mesmos, permitindo verificar os padrões espaciais das condições socioeconômicas representadas;
- Analisar padrões espaciais através da detecção de agrupamentos nos pixels representados nos indicadores socioeconômicos presentes nas superfícies contínuas, determinando, então, fronteiras que delimitem um zoneamento socioeconômico homogêneo;

- Estabelecer novos limites censitários a partir do zoneamento obtido, buscando-se a minimização dos Efeitos de Zoneamento dos Problemas de Unidades de Áreas Modificáveis;
- Propor a integração dos Setores Censitários ao Cadastro Territorial Multifinalitário através do estabelecimento de regras de restrições topológicas e de vizinhança.

2.2. Justificativas

A justificativa deste trabalho é baseada no fato de que limites de setores ou zonas de planejamento são definidos por fatores administrativos e logísticos (RAMOS *et al*, 2007), indo de encontro ao conceito de dependência espacial, baseada na Primeira Lei da Geografia – Lei de Tobler, que diz que todas as coisas se parecem, mas as coisas mais próximas entre si se parecem mais que as mais distantes (ABREU, 2009), (RAMOS, 2002) e (WONG, 1995).

A definição de um zoneamento homogêneo permite a coleta e disponibilização de dados socioeconômicos, que servirão de base para a formulação de indicadores, norteando estudos e implementação de projetos não apenas por parte da gestão municipal, como também no âmbito acadêmico e privado.

Sem haver um estabelecimento de limites segundo critérios de homogeneidade socioeconômica, as informações dessa natureza, obtidas em cada região não homogênea, possuem grandes distorções. Por consequência, inferências errôneas acabam servindo de base informacional para elaboração de estratégias de desenvolvimento de políticas públicas e investimentos privados.

A necessidade de integração com o Cadastro Territorial Multifinalitário é explícita, uma vez que, por definição, este é um inventário municipal, contendo informações de todo o município de forma sistematizada e atualizada, sendo, portanto, a principal base de informações municipais.

Sendo o CTM o responsável por gerenciar e integrar qualquer base de informações municipais, sobretudo, no contexto deste trabalho, as informações

socioeconômicas, deve haver um mecanismo que associe tais informações ao próprio CTM.

De outro lado, os setores censitários são tidos, muitas vezes, como única base de dados de natureza socioeconômica geoespacializada num município. Este fato traz a necessidade de integrá-los à base de dados do CTM, tendo, eles, o papel de retratar de maneira adequada do ponto de vista espacial – amarrada às parcelas georreferenciadas do cadastro – a realidade socioeconômica municipal.

3. REVISÃO DE LITERATURA

3.1. CADASTRO TERRITORIAL MULTIFINALITÁRIO

Historicamente, o cadastro de terras esteve presente na vida da humanidade desde aproximadamente 4000 antes de Cristo. Loch e Erba (2007) traçam um pequeno histórico da sua evolução técnica e conceitual. Destaca-se, neste ínterim, o Cadastro Napoleônico, criado na França por Napoleão Bonaparte, no século XVIII, o qual influenciou e ainda influencia os principais sistemas cadastrais existentes na atualidade (BRANDÃO, 2003).

Os primeiros cadastros imobiliários foram concebidos com o propósito de arrecadação, sendo conhecidos como Cadastros Econômicos ou Fiscais. A base de dados para cálculos de impostos é proveniente de levantamentos e documentos cartográficos que compõem o Cadastro Geométrico ou Físico. Com o tempo, percebeu-se que o cadastro poderia ser utilizado junto com o Registro de Imóveis, constituindo o Cadastro Jurídico (ÁGUILA e ERBA, 2007).

Águila e Erba (2007) ainda afirmam que o cadastro é aquele que contribui com a distribuição equitativa das cargas tributárias, promove a segurança jurídica da propriedade imóvel e cria as bases para o planejamento urbano e regional.

Além disso, segundo os mesmos autores, a consolidação da Agenda 21, aprovada durante a Conferência das Nações Unidas sobre o Meio Ambiente e Desenvolvimento, associada à Segunda Conferência das Nações Unidas sobre Assentamentos Humanos – HABITAT II – ampliou a visão do cadastro. Os dois eventos “deixam clara a importância da informação territorial confiável como apoio aos processos de tomada de decisões orientadas a preservar o meio ambiente e a promover o desenvolvimento sustentável” (ÁGUILA e ERBA, 2007).

Tem-se, portanto, que o Cadastro Territorial Multifinalitário consiste em um “inventário territorial oficial e sistemático do município”, conforme definem as Diretrizes Para a Criação, Instituição e Atualização do Cadastro Territorial Multifinalitário (CTM) nos Municípios Brasileiros (Portaria Ministerial nº 511), as quais são denominadas nessa dissertação, a partir deste ponto, de Diretrizes para o Cadastro.

Garcia (2007) define a identificação e caracterização de unidades, indivíduos e elementos como a principal característica do CTM, indicando que, muito mais que dar rigor jurídico, fiscal e tributário às propriedades urbanas, deve haver a capacidade de integrar informações de diferentes áreas do governo local.

Essa definição vai ao encontro das Diretrizes para o Cadastro, supracitadas, que dizem que o CTM deve atender às questões sociais, ambientais e econômicas da administração pública, além da segurança jurídica da sociedade. Assim, qualquer sistema de informações municipais deve se basear no CTM, principalmente no que diz respeito à representação geométrica.

Para tal representação, a unidade territorial utilizada no CTM é a parcela. Ela consiste numa parte contígua e contínua da superfície terrestre com regime jurídico único. Philips (2010) afirma que “a identificação e o levantamento único do limite entre duas parcelas e a demarcação do limite pela sequência de vértices comuns” caracterizam a contiguidade das parcelas, conforme pode ser visto na Figura 3.1.

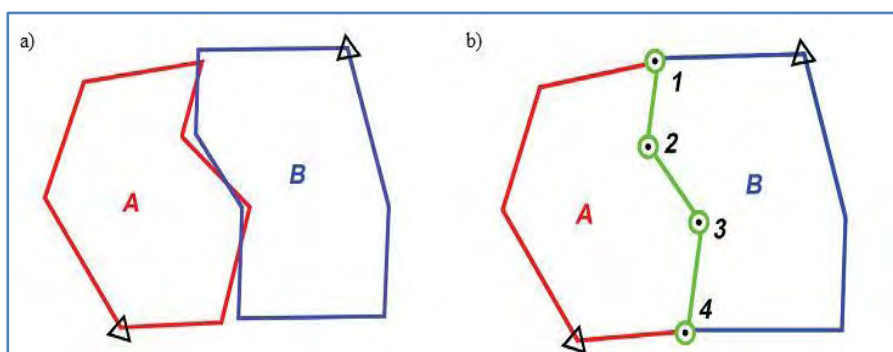


Figura 3.1 Exemplos de parcelas não contíguas em a) e contíguas em b).

Fonte: PHILIPS (2010)

Por sua vez, a continuidade de uma parcela é definida quando a sequência de linhas que formam os limites da mesma configura um polígono fechado (PHILIPS, 2010). Portanto uma parcela não pode ser formada por dois polígonos isolados. A Figura 3.2 exemplifica o conceito de continuidade.

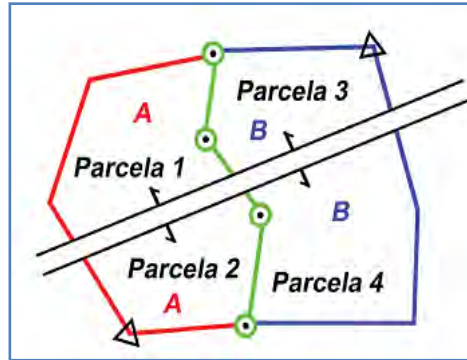


Figura 3.2. As parcelas 1, 2, 3 e 4 são contínuas e os imóveis A e B, descontínuos

Fonte: PHILIPS (2010)

Philips (2010) ainda explicita o conceito de regime jurídico único a partir de um exemplo hipotético, o qual se reproduz neste trabalho. A Figura 3.3 mostra uma propriedade cuja estrutura interna foi alterada a partir de alguns eventos, tais quais são apresentados. A partir das alterações, a propriedade passou a ser representada por 7 parcelas.

A parcela 1 representa a área remanescente da propriedade que ficou do outro lado da estrada que foi construída na região. A parcela 2 consiste na faixa de domínio da estrada, a qual foi desapropriada. A parcela 3 é o resultado de um contrato de usufruto, tendo um regime jurídico diferente da parcela 4, que é a residência do proprietário.

A parcela 5, por seu lado, constitui uma área de preservação ambiental, restringindo seu uso e ocupação. Já a parcela 6 é uma área de servidão para acesso às praias, a qual é desconsiderada para fins tributários por ser de uso público. Por último, a parcela 7 é de propriedade da União, sendo concedido seu uso ao dono do imóvel.

Percebe-se que todas as parcelas são contíguas e contínuas. A partir delas, o imóvel pode ser representado sob pontos de vistas diferentes. Para fins legais, existem dois imóveis, com duas matrículas diferentes no registro de imóveis, sendo o primeiro constituído pela parcela 1 e o segundo, pela soma das parcelas 3, 4, 5 e 6.

Para fins econômicos, existe um imóvel, diferente dos anteriores, o qual é formado pelas parcelas 1, 3, 4 e 7, já que as demais não geram tributos territoriais. Destaca-se que a parcela 7 é a única registrada pela Secretaria de Patrimônio da União.

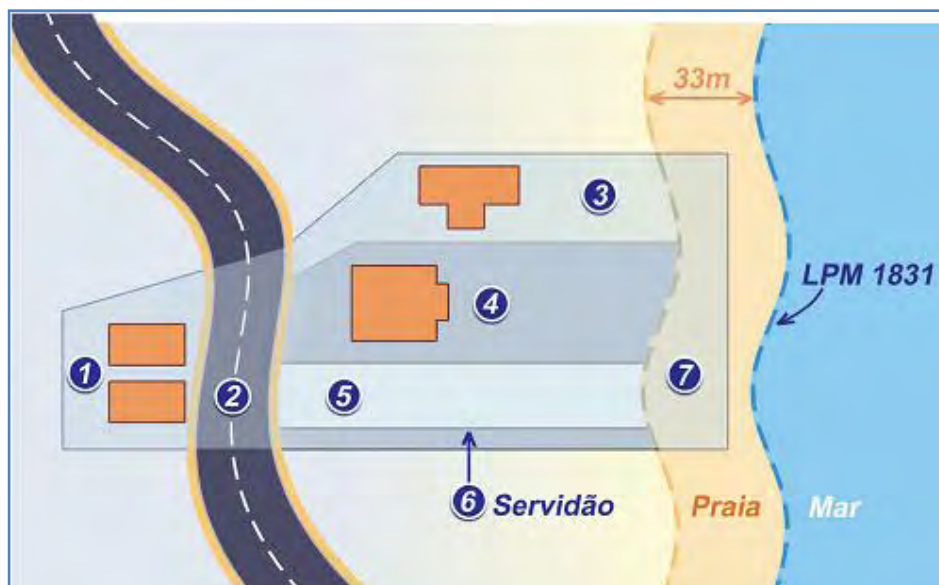


Figura 3.3. Diferentes parcelas que compõem um imóvel

Fonte: PHILIPS (2010)

Para que as parcelas sejam diferenciáveis entre si, cada uma deve ter um identificador único e estável, que serve de ligação entre ela mesma e o tema cadastrado no CTM. Todas as demais unidades territoriais, como logradouros, praças, cursos d'água, glebas, etc. são definidas a partir dos limites de uma ou mais parcelas. Quaisquer dados ou informações podem estar associados ao CTM, uma vez que o caráter multifinalitário do cadastro é assegurado nos artigos 21 e 22 das Diretrizes para o Cadastro (ABREU *et al*, 2010b).

3.1.1. Cartografia Cadastral

O artigo 8º das Diretrizes para o Cadastro define a Carta Cadastral como sendo a representação cartográfica do levantamento sistemático do território municipal. Todo o levantamento e as representações de limites territoriais devem estar contidos nela.

Define-se ainda que todo o levantamento cadastral, seja ele por topografia, geodésia, aerofotogrametria ou outros métodos com precisão compatível, deve estar amarrado ao Sistema Geodésico Brasileiro (SGB). Infere-se, portanto, que todo o Mapa Urbano Básico (MUB), definido pelos limites cadastrais, possa ser agregado a outras fontes de dados, dando o caráter multifinalitário do cadastro.

Um MUB é constituído de elementos como perímetro urbano, loteamentos, quadras, logradouros e outras divisões (GARCIA, 2007), os quais devem ser determinados pelos limites das parcelas. Obviamente que existem outros elementos urbanísticos que podem ser representados num MUB e que têm a forma linear (redes de água e esgoto) e pontual (postes, árvores, poços de visita, etc.), tal como explicita Loch e Erba (2007).

As Diretrizes para o Cadastro não definiram a projeção cartográfica mais adequada para as representações da Carta Cadastral, uma vez que tal feito é de responsabilidade da Comissão Nacional de Cartografia (CONCAR). Até que isso aconteça, é indicada a utilização do sistema Universal Transverso de Mercator (UTM), caso o município já não tenha outro em utilização.

É certo, porém, que toda Carta Cadastral deve obedecer aos padrões da Infraestrutura Nacional de Dados Espaciais (INDE) e às normas da Cartografia Nacional. Busca-se, portanto, adequar todas as informações territoriais a padrões espaciais que permitam não só a precisão necessária, como também garanta que o cruzamento com dados externos (análises espaciais) seja feito com critério garantido pela Cartografia Digital.

3.1.2. Cadastros Temáticos

Define-se como conteúdo mínimo do CTM a caracterização geométrica da parcela, seu uso, identificador único, localização e proprietário. Porém, o cadastro não é apenas a visão cartográfica que mostra a estrutura fundiária da cidade (LOCH e ERBA, 2007). Todos os temas a serem cadastrados, de acordo com as diversas necessidades do município, devem estar amarrados à Carta Cadastral, como define o artigo 20 das Diretrizes para o Cadastro.

Sendo assim, temas importantes, como rede viária, redes de água, esgoto e drenagem de águas pluviais, uso do solo, zonas homogêneas, equipamentos urbanos, unidades de saúde e de educação, segurança pública, hidrografia urbana, transporte urbano, etc. (GARCIA, 2007; CONÇALVES, 2008) podem ser considerados como agregados ao conteúdo mínimo do CTM, dando o caráter multifinalitário do mesmo.

Todas essas informações devem estar contidas no Sistema de Informações Territoriais (SIT), tal como definido nas mesmas diretrizes. Este sistema é constituído,

portanto, pelos dados do conteúdo mínimo, dados do registro de imóveis e dados temáticos. Ressalta-se, no entanto, que as Diretrizes para o Cadastro propõem apenas a estruturação de um cadastro básico, que possibilite a integração a cadastros específicos.

Para isso é importante que todos os temas sejam relacionados às parcelas. Um logradouro cadastrado, por exemplo, deve estar associado a todas as parcelas que representam o mesmo regime jurídico relacionado ao mesmo logradouro. Torna-se fácil, portanto, recuperar os serviços urbanos oferecidos pela gestão municipal em uma dada parcela, tais como pavimentação, iluminação pública, abastecimento de água, coleta de esgoto, drenagem de águas pluviais, etc., por meio de simples consultas ao banco de dados do SIT.

O mesmo se diz do cadastro de redes de abastecimento de água, como um segundo exemplo. A cada parcela estariam associadas, indiretamente, as dimensões das tubulações, a zona de pressão a qual está localizada a parcela, a qualidade da água, entre outros dados associados à própria rede. Seguindo o raciocínio, poder-se-ia aplicar a ideia a vários temas, no entanto, fica claro que qualquer um destes pode ser agregado ao SIT, inclusive o socioeconômico.

Com o intuito de exemplificar o tema socioeconomia no CTM, Loch e Erba (2007) sugerem a geoespacialização ou o endereçamento de pacientes e/ou alunos a partir dos endereços cadastrados nas instituições de saúde e ensino, respectivamente. Por outro lado, num processo de Modelagem de Banco de Dados, Gonçalves (2008) propõe modelos de diagramas de classes direcionados a essas mesmas instituições, sendo cadastradas as unidades institucionais, em vez de seus usuários. Em qualquer proposta, a entidade cadastrada está relacionada à parcela a que pertence (ABREU *et al*, 2010b).

Contudo, a realização e a atualização de um cadastro socioeconômico podem ser inviáveis financeiramente para uma prefeitura sem uma visão sistêmica de gestão territorial, uma vez que não há retorno financeiro direto, como no caso do cadastro imobiliário. Porém a integração interinstitucional proposta pelas Diretrizes do Cadastro traz a compreensão que a manutenção de todo o cadastro deve prover das receitas geradas pelo Imposto Predial e Territorial Urbano (IPTU) (CUNHA e ERBA, 2010).

É certo que o retorno do investimento feito no cadastro socioeconômico da população é indireto. No entanto, informações qualificadas e atualizadas sobre a população são a principal chave para a leitura eficaz da realidade municipal.

Processos de planejamento urbano requerem levantamento de dados socioeconômicos e físicoterritoriais, compatibilizados cartograficamente e temporalmente. Loch e Erba (2007) afirmam que o conhecimento da situação socioeconômica da população e sua distribuição geográfica é condição básica para que a justiça social seja construída.

3.2. GEOPROCESSAMENTO APLICADO

O geoprocessamento, segundo Câmara et al (2001), é a “disciplina do conhecimento que utiliza técnicas matemáticas e computacionais para o tratamento da informação geográfica”. Xavier-da-Silva (2007), por sua vez, considera o geoprocessamento um ramo da tecnologia da computação eletrônica de dados. Este autor ainda define o termo como sendo uma ferramenta computacional aplicada ao ramo científico da geoinformação.

O geoprocessamento é aplicado a todo o ramo da ciência em que a informação geográfica é essencial. Áreas do conhecimento, tais como Cartografia, Análise de Recursos Naturais, Transportes, Comunicações, Energia, Planejamento Urbano e Regional, Monitoramento Agrícola, Análise de Marketing, entre outras, foram influenciadas pelo tratamento da geoinformação (CÂMARA *et al*, 2001).

Pode-se considerar, portanto, o geoprocessamento como um conjunto de técnicas relacionadas a: *coleta* (Cartografia, Sensoriamento Remoto Aéreo e Orbital, Topografia, Geodésia, Questionários Alfanuméricos); *armazenamento* (Banco de Dados Geográficos); *tratamento e análise* (Modelagem de Dados, Estatística Clássica e Espacial, Aritmética Lógica, Análise de Redes, Análise Topológica, Lógica *Fuzzy*, etc.) e; *uso integrado* (Sistemas Baseados em Localização – LBS, Mapeamento Automatizado e Gerenciamento de Facilidades – AM/FM, Desenho Assistido por Computador – CAD, etc.).

A seguir, apresentam-se segmentos do geoprocessamento relevantes ao entendimento do trabalho aqui desenvolvido.

3.2.1. Modelagem de Dados, Ontologia e Banco de Dados Geográficos

Modelo de Dados é uma abstração da realidade, sendo uma forma de representá-la. A modelagem geográfica consiste num compromisso entre sintetizar o conhecimento sobre determinado fenômeno e simultaneamente providenciar informação com conteúdo tão completo quanto o possível (MATOS, 2008).

O fenômeno em questão pode estar inserido em diversas áreas do conhecimento, como geologia, agronomia, saúde, sociologia, entre outros (DRUK *et al.*, 2004). Assim, a variável espacial pode ser o teor de argila no solo, temperatura, casos de determinado tipo de doença, crimes cometidos, indicadores socioeconômicos, lotes urbanos, redes de serviços, parques, entre outros (ASSUNÇÃO, 2008).

Câmara *et al* (2001) citam o paradigma dos quatro universos na modelagem de dados geográficos. O primeiro é o Universo do Mundo Real ou Ontológico, o qual consiste no mundo tal como é, com suas entidades da realidade a serem modeladas, onde se definem os conceitos que representam cada entidade.

Ontologia é uma palavra de origem grega que tem por objetivo definir conceitos associados às coisas. Sua utilização em Sistemas de Informações Geográficas é recente (SILVA, 2006). A ontologia é responsável pelo estabelecimento de conceitos explícitos sobre determinada entidade geográfica de acordo com o domínio do conhecimento que se utiliza do modelo de dados que represente essa mesma entidade.

Um lote urbano, por exemplo, pode ter várias concepções a partir dos diferentes domínios em uma prefeitura (PINHO *et al*, 2007). Para a concessionária de abastecimento de água, por exemplo, a determinação das dimensões do lote pode não ser de nenhum interesse, uma vez que o que importa é sua localização, a qual pode ser representada por um ponto. No entanto, de acordo com o significado dado a essa entidade, um departamento de cadastro necessita da representação dos seus limites, bem como de suas propriedades geométricas.

Portanto, o estabelecimento das ontologias serve de base à concepção das representações geométricas (campos ou objetos) que se devem utilizar para a modelagem dos dados, bem como de sua organização e de seus relacionamentos com outros dados presentes num Banco de Dados Geográficos (BDG) (QUEIROZ e FERREIRA, 2006). Esse procedimento, por sua vez, consiste no Universo Matemático ou Formal, formando o segundo universo.

Os autores supracitados ainda afirmam que a Modelagem de Banco de Dados Geográficos se divide em duas estruturas complementares: *modelo conceitual* e *esquema conceitual*. O primeiro se refere à técnica utilizada para a modelagem bem como suas notações. Já o segundo se refere ao conjunto de diagramas utilizados para representar graficamente o resultado da modelagem conceitual, ou seja, é nesta estrutura que se representam a organização e os relacionamentos dos dados num BDG.

Existem diversas propostas metodológicas para a modelagem de dados geográficos (QUEIROZ e FERREIRA, 2006). Na literatura, um dos modelos mais utilizados é o OMT-G (*Object Modeling Technique for Geographic Applications*), baseado na linguagem UML (*Unified Modeling Language*). Assim, o modelo utiliza os diagramas de classes da UML, introduzindo primitivas geográficas (formas geométricas – ponto, linha, polígono e superfície).

Classes, relacionamentos e restrições de integridade são três conceitos que baseiam o modelo OMT-G. As classes representam as entidades geográficas (contínuas ou discretas) ou não geográficas (dados descritivos, não espaciais – também chamadas de classes convencionais) e são dispostas sob a forma de diagrama, o qual descreve sua estrutura e seus relacionamentos e é conhecido também por Diagrama de Classes.

As classes são representadas por uma caixa retangular, na qual são indicados o nome da classe, os atributos pertencentes à classe, as operações associadas e um símbolo que indica a geometria representada, conforme a Figura 3.4. A única diferença para as classes convencionais é que esta não possui a representação geométrica no canto superior esquerdo.

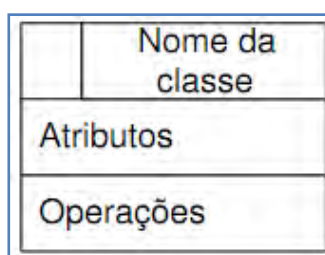


Figura 3.4 Notação gráfica para as classes georreferenciadas no modelo OMT-G

Fonte: Casanova *et al* (2005)

Os relacionamentos definem o tipo de interação entre as classes, podendo ser associação simples (baseada em atributos alfanuméricos), relacionamentos topológicos

em rede e relacionamentos espaciais. Todos são indicados por linhas que conectam as classes. Quando se tem associação simples, as linhas são contínuas e, quando há um dos outros dois relacionamentos, são tracejadas.

Existem vários tipos de relacionamentos espaciais, os quais variam com os diferentes modelos ou as diferentes aplicações. Uma das formas deste relacionamento é a *agregação*, a qual considera que uma das classes é formada pela agregação das instâncias de outra classe, ou seja, há uma relação “todo-parte” (CASANOVA *et al*, 2005). A Figura 3.5 exemplifica a representação dessa relação. O losango que aparece no final da linha tracejada indica a relação de agregação.

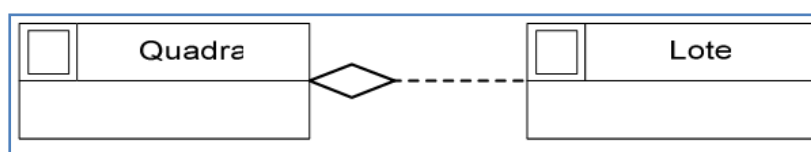


Figura 3.5 Agregação Espacial

Fonte: Casanova *et al* (2005)

A agregação espacial se divide em três tipos: subdivisão espacial (o objeto primitivo se subdivide, gerando novos objetos derivados); união espacial (o objeto derivado é gerado pela união dos objetos primitivos) e; continência (um objeto contém outros objetos) (BORGES, 1997).

Quanto às restrições de integridade, Borges (1997) afirma que algumas aplicações geográficas dependem de relacionamentos topológicos, que precisam ser explicitados num BDG. Assim, algumas regras de dependência espacial (importante diferenciar esse termo, baseado apenas nas características espaciais, do mesmo termo empregado em Estatística Espacial, visto adiante) devem ser seguidas.

A autora ainda enumera os tipos de regras a serem respeitadas para o caso de agregação espacial. Tais regras não são explicitadas nesta seção, sendo aquelas necessárias para o entendimento da proposta de integração dos setores censitários ao CTM apresentadas junto à mesma, em seção apropriada.

Outras questões envolvendo a modelagem de banco de dados poderiam ser citadas. No entanto, acredita-se que o conteúdo focado aqui seja o suficiente para o entendimento da aplicação desses conceitos nesta dissertação.

Seguindo com o processo de modelagem de dados, como apresentado por Câmara *et al* (2001), a partir da concepção de como os dados se estruturam e se relacionam, surge o terceiro universo, chamado de Universo da Representação, onde se distinguem se os dados serão representados por modelos vetoriais ou matriciais (QUEIROZ e FERREIRA, 2006).

As variáveis espaciais podem ser representadas ou abstraídas em quatro classes geográficas: pontos, linhas, polígonos ou superfícies. Os primeiros são utilizados para representar fenômenos que ocorrem em um lugar específico, como postes, árvores, poços artesianos, etc..

As linhas conectam pelo menos dois pontos, representando eventos isolados (falhas geológicas, direção dos ventos, rotas de animais, etc.), eventos conectados em estrutura de árvores (hidrografia, por exemplo) ou eventos conectados em forma de redes (sistema viário, abastecimento de água, coleta de esgotos, cabos telefônicos, etc.).

Já os polígonos são usados para representar áreas fechadas, conectando pelo menos três linhas. Eles podem ser isolados (fronteiras não compartilhadas), adjacentes (fronteiras compartilhadas) ou aninhados (polígonos internos a outros).

Por último, as superfícies são representações de fenômenos contínuos, cujos valores possíveis se dão sem interrupção ao longo do espaço. Podem representar relevo, temperatura, nutrientes do solo, indicadores socioeconômicos, etc..

Qualquer uma das classes geográficas pode ser representada por modelos de dados vetoriais e/ou matriciais. Modelos vetoriais reproduzem as feições concebidas na cartografia impressa, sendo definidos pelas coordenadas dos vértices de cada feição. A eles podem ser associados atributos da localização geográfica, armazenados em tabelas estruturadas em forma de banco de dados, recebendo o nome de *Modelo Vetorial Relacional*.

Este modelo é o mais utilizado para representar feições pontuais, lineares e poligonais. A superfície – ideia de continuidade espacial – é representada indiretamente por isolinhas (feições lineares) ou diretamente por Redes Irregulares de Triângulos (*TIN* – *Triangular Irregular Network*).

Já o modelo matricial consiste numa discretização do espaço em células, cujas posições são definidas por números de linhas e de colunas, em conjunto com as coordenadas da primeira célula e a dimensão espacial dessas células, também chamada de resolução espacial.

Cada célula possui um único código associado, o qual indica o valor da variável representada na localização referente a essa mesma célula. Esse modelo também é capaz de representar feições pontuais, lineares e poligonais, destacando-se, porém, sua eficiência em representações de superfícies.

Uma limitação dessa representação espacial é o alto nível de generalização, causando perda de características geográficas. Em contrapartida, é mais utilizada em análises espaciais, principalmente no concernente à álgebra de mapas.

A Figura 3.6 abaixo mostra uma comparação entre os dois modelos de dados para um mesmo ambiente real.

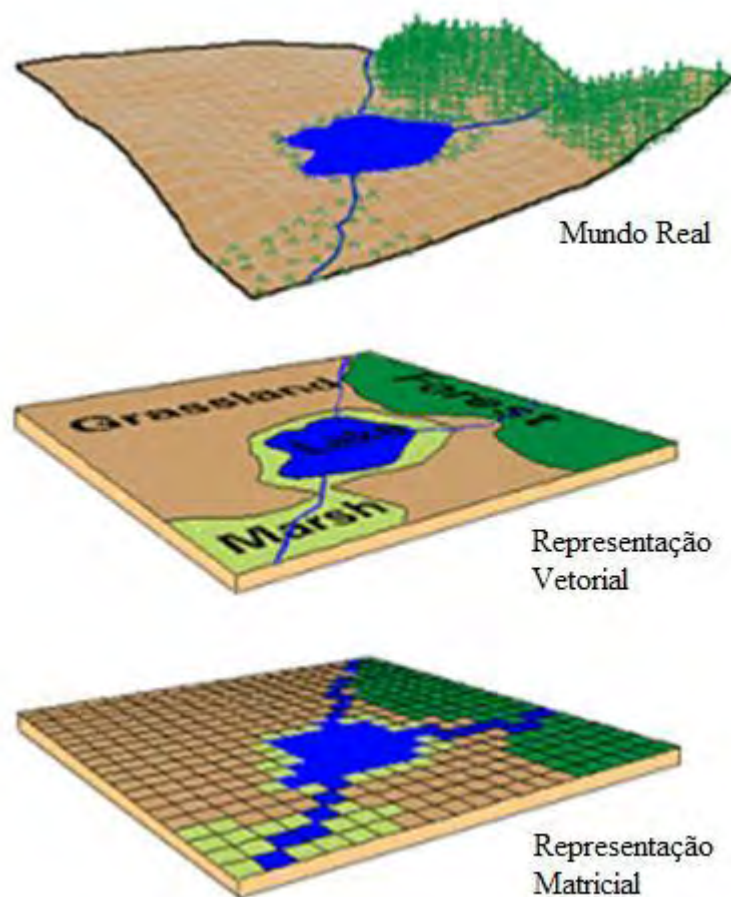


Figura 3.6. Representações matricial e vetorial para um mesmo ambiente real.

Fonte: Câmara e Monteiro (2003)

O último universo, por sua vez, é o da Implementação, no qual se definem as estruturas de dados para a implementação das geometrias selecionadas nos modelos de dados escolhidos. Aspectos relacionados a algoritmos, geometria computacional,

hardware, processamento e otimização, além da utilização das bibliotecas para a construção de aplicativos geográficos (CASANOVA *et al*, 2005) são contemplados neste universo.

Resumidamente, a concepção de um modelo deve contemplar os seguintes aspectos:

- Estabelecimento dos conceitos relacionados à definição do fenômeno estudado (definição das ontologias), ou seja, sobre o que deve ser modelado, como se comporta num ambiente georreferenciado, quais são seus relacionamentos e como se organizam no modelo de banco de dados proposto, de acordo com as próprias ontologias definidas;
- Identificação de quais entidades devem ser levantadas e como devem ser representadas;
- Análise da adequação do levantamento do dado quanto à adequação a escala (nível de generalização), resolução (menor dimensão a considerar), perspectiva (projeção cartográfica), tempo (variação temporal deve ser considerada) e erro associado (espacial e/ou de atributo);
- Concepção da estrutura de dados (vetores ou matrizes);
- Identificação da incidência de erros; e
- Definição de procedimentos de gestão e atualização das informações.

3.2.2. Geoestatística Aplicada

Uma das formas de se obter uma superfície com os valores de uma variável contínua no espaço é através da interpolação de amostras pontuais georreferenciadas. As amostras podem estar regularmente ou irregularmente espaçadas e a cada ponto deve estar associado um valor dessa variável.

Após a coleta ou obtenção dessas amostras, um estudo criterioso sobre o interpolador espacial a ser utilizado deve ser feito, existindo duas categorias dos mesmos: os determinísticos e os estocásticos. Os primeiros consideram apenas as configurações espaciais das amostras, não se preocupando com a variabilidade da mesma. Encontram-se aqui os interpoladores do inverso da distância, vizinhos mais próximos, triangulação, etc., como destacam Burrough e McDonell (1998).

Em contrapartida, na segunda classe, o processo de interpolação se baseia na variabilidade espacial das amostras pontuais, consideradas, neste caso, variáveis regionalizadas, cuja teoria foi formulada por Matheron em 1963 (SCHULER, 2007). Segundo Vicente (2004), a variável regionalizada possui como características a localização, a anisotropia (variações diferentes em diferentes direções, as quais dependem das características do fenômeno) e a continuidade (baseada na Primeira Lei da Geografia, melhor abordada adiante).

Em geoestatística, segundo Felgueiras (1999), “os valores dos atributos são representados como variáveis ou campos aleatórios.” Ele ainda diz que o método de interpolação conhecido como Krigagem permite inferir valores de um atributo espacial a partir de amostras pontuais para qualquer região. Burrough e McDonnell (1998) acrescentam que, em Krigagem, o atributo é conhecido como variável regionalizada.

A Krigagem considera a continuidade e a distância entre dois pontos, podendo ser um amostral e um não-amostral (SANTOS, 2008). Os pontos não-amostrais, no geral, apresentam-se igualmente espaçados, para serem representados computacionalmente por uma grade regular (estrutura matricial).

Ramos (2002) apresenta dois métodos de interpolação por Krigagem: a Krigagem Ordinária e a Krigagem Ordinária por Indicação. Felgueiras (1999) diz que a Krigagem Ordinária possui um algoritmo mais robusto que outros estimadores por Krigagem, por trabalhar com médias ou tendências estimadas localmente, em contrapartida aos outros, que preestabelecem uma média estacionária.

A ponderação deste tipo de interpolação é baseada na distância que separa duas amostras e numa medida de dependência espacial entre elas. Por sua vez, a análise desta dependência espacial entre as amostras é fundamentada no conceito da continuidade espacial, que é modelada a partir da ideia de autocorrelação espacial.

Como citado anteriormente, a continuidade espacial se baseia na Primeira Lei da Geografia de Waldo Tobler, a qual diz que as coisas tendem a se parecer, porém, as coisas mais próximas entre si se parecem mais do que as coisas mais distantes (JAKOB, 2002), (WONG, 1995) (ISAAKS e SRIVASTAVA, 1989). Vicente (2004) acrescenta que a avaliação da continuidade espacial se dá através dos conceitos de regressão linear para analisar a dependência espacial entre duas variáveis, ou entre uma variável em duas posições diferentes.

Em geostatística, a maneira mais apropriada para se obter informações sobre a dependência espacial, através da análise da autocorrelação espacial, é a utilização da função semivariograma (γ). Vicente (2004) e Isaaks e Srivastava (1989) explicam e detalham a origem e os princípios dessa função.

Como o semivariograma descreve a continuidade espacial, e esta se baseia na Primeira Lei de Tobler, a função semivariograma demonstra que as diferenças entre amostras separadas por um vetor h é menor quando h também é menor, ou seja, quanto menor o valor da função, mais provável é a continuidade espacial da variável em estudo.

O semivariograma é construído a partir de um conjunto de valores obtidos por uma equação que relaciona a semivariância entre as amostras e a distância que as separa, conforme ilustrado em Vicente (2004). A autora chama este gráfico de semivariograma experimental, acrescentando que deve ser ajustado um modelo matemático a ele de modo a representar a tendência da função semivariograma em relação a uma dada distância. O gráfico é plotado a partir da Equação 1.

$$\gamma(h) = \frac{\sum(v_i - v_j)^2}{2N(h)} \quad (1)$$

onde,

$N(h)$ é número de pares de pontos que estão a uma determinada distância h e;

v_i e v_j são amostras nas localizações i e j .

Como pode ser visto na Figura 3.7, a função semivariograma possui 4 parâmetros, a saber:

- Patamar – nível onde o semivariograma se estabiliza e considera-se que não haja mais dependência espacial entre as observações;
- Alcance – distância h a partir da qual não há dependência espacial;
- Efeito Pepita – apresenta a descontinuidade do semivariograma para distâncias menores que a menor distância de interesse entre as amostras;
- e
- Contribuição – o quanto de informação os pares de pontos estão fornecendo.

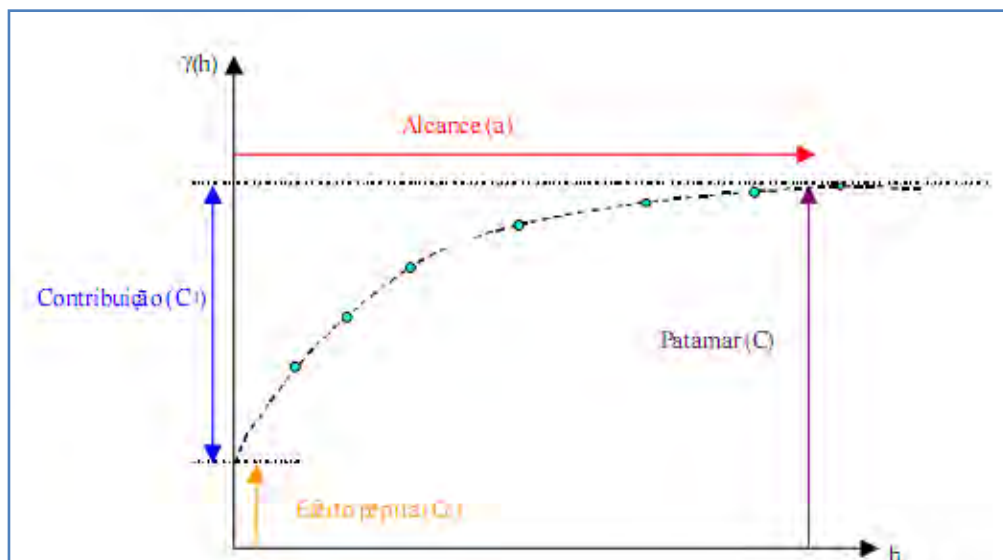


Figura 3.7. Exemplo de semivariograma experimental.

Fonte: Vicente (2004).

Este modelo é ajustado de maneira interativa com o pesquisador, com base nos diferentes semivariogramas teóricos existentes. Estes, por sua vez, devem representar, de maneira satisfatória, a tendência apresentada pelo semivariograma experimental, permitindo que as estimativas por Krigagem apresentem variâncias mínimas (SCHULER, 2007).

O modelo teórico é utilizado porque, na maioria das vezes, torna-se impossível obter os parâmetros necessários para a modelagem da dependência espacial das amostras a partir do semivariograma experimental (SANTOS, 2008).

A função do modelo teórico do semivariograma varia de acordo com o próprio modelo e se apresenta genericamente como na Equação 2.

$$\gamma(h) = f(C_0, C_1, a, h) \quad (2)$$

sendo,

C_0 o efeito pepita;

C_1 a contribuição;

a o alcance;

h o vetor distância entre dois pontos.

Camargo (1997) apresenta, de forma detalhada, os principais modelos teóricos (esférico, exponencial e gaussiano – apresentados a seguir), citando também a modelagem aninhada (ou complexa) e da anisotropia (distribuição mais intensa numa

direção e menos em outra). A Figura 3.8 foi retirada de Camargo (1997) e mostra o comportamento dos três modelos mais utilizados em geoestatística. Ressalta-se obviamente que outros modelos podem ocorrer, dependendo do fenômeno modelado.

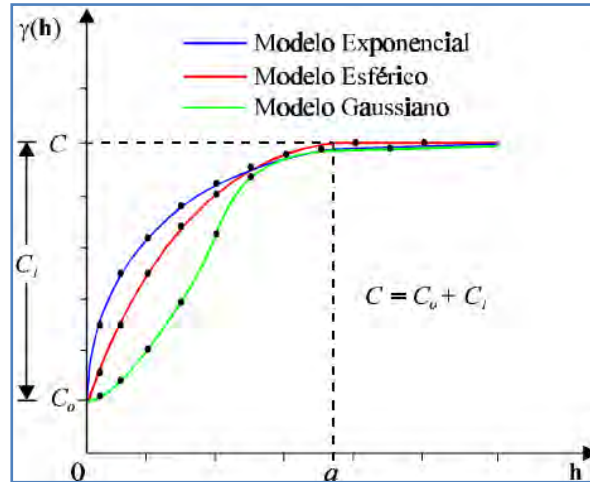


Figura 3.8. Representação gráfica dos principais modelos de semivariograma.

Fonte: Camargo (1997).

Os modelos matemáticos dos semivariogramas teóricos Esférico, Exponencial e Gaussiano são dados pelas Equações 3, 4 e 5, respectivamente, dadas abaixo:

$$\gamma(h) = \begin{cases} 0 & , |h| = 0 \\ C_0 + C_1 \left[1,5 (|h| / a) - 0,5 (|h| / a)^3 \right] & , 0 < |h| \leq a \\ C_0 + C_1 & , |h| > a \end{cases} \quad (3)$$

$$\gamma(h) = \begin{cases} 0 & , |h| = 0 \\ C_0 + C_1 \left[1 - \exp(-|h| / a) \right] & , |h| \neq 0 \end{cases} \quad (4)$$

$$\gamma(h) = \begin{cases} 0 & , |h| = 0 \\ C_0 + C_1 \left[1 - \exp(-|h| / a)^2 \right] & , |h| \neq 0 \end{cases} \quad (5)$$

Definido e ajustado o modelo teórico, obtêm-se os valores das semivariâncias entre os pontos amostrados e os pontos não amostrados, em função das distâncias entre os mesmos. Aplica-se a equação do modelo teórico escolhido em cada par amostral ou não amostral, sendo a distância entre eles a variável independente, obtendo-se, assim, a semivariância entre o par.

Em seguida, parte-se para a determinação dos pesos atribuídos às amostras, utilizando a modelagem do semivariograma, segundo o método de interpolação escolhido.

Camargo (1997) mostra que, para a obtenção de um estimador não tendencioso e sem a necessidade de conhecimento prévio da média do atributo em estudo (características da interpolação por Krigagem Ordinária), basta que o somatório dos pesos seja igual a 1. Para isso, utiliza-se o multiplicador de Lagrange. Assim, o estimador de Krigagem Ordinária é dado na Equação 6.

$$Z^*(x_0) = \sum \lambda_i \cdot Z(x_i), \quad (6)$$

sendo,

$Z^*(x_0)$ o valor do atributo a ser estimado na posição X_0 ;

$\sum \lambda_i$ o somatório dos pesos, com i variando de 1 a n , sendo n o número de valores amostrados. Em Krigagem Ordinária deve ser igual a 1;

$Z(x_i)$ o valor do atributo a em cada amostra posicionada em X_i .

Minimizando a variância do erro sob a condição $\sum \lambda_i = 1$, segundo cita Camargo (1997), os pesos são obtidos a partir de um sistema de equações de Krigagem Ordinária, que pode ser escrito sob notação matricial, como mostra a Equação 7.

$$K \cdot \lambda = k \quad \text{ou} \quad \lambda = K^{-1} \cdot k, \quad (7)$$

onde,

K é a matriz de semivariância entre as amostras, obtidas pela Equação 1;

k é a matriz de semivariância entre as amostras e o ponto a ser estimado, obtidos pela equação do modelo empírico escolhido;

λ o vetor dos pesos.

Cada semivariância é inserida na matriz correspondente (K ou k). A partir da equação 7, obtêm-se os pesos entre as amostras e cada ponto não amostral. O passo seguinte é obter o valor do atributo em um ponto não amostral, aplicando-se a Equação 6.

Ao final do processo de interpolação, tem-se um conjunto de pontos não amostrais, regularmente espaçados, representados por grades matriciais, contendo o valor estimado para a variável de estudo naquele local.

3.2.3. Segmentação de Superfícies

Em algumas aplicações de grades matriciais em geoprocessamento, tem-se o objetivo de encontrar áreas de mesmas características. Aqui se mostram algumas técnicas aplicadas a imagens orbitais de Sensoriamento Remoto, as quais se assemelham em sua estrutura e seu modelo de dados com as superfícies contínuas obtidos pela interpolação citada anteriormente.

Um método bastante utilizado é a segmentação de imagens, a qual consiste em um reconhecimento de feições. Ela busca agrupar pixels com características similares, identificando bordas desses agrupamentos. Medeiros (2003) cita que os algoritmos de segmentação se baseiam na similaridade e descontinuidade dos níveis de cinza numa imagem monocromática. Neste caso, pode-se assumir que cada valor da superfície contínua utilizada corresponde a um valor de cinza, para se aplicar as técnicas de segmentação.

Coutinho (1997) diz que a segmentação tem como produto final regiões possuidoras de uma mesma propriedade, sendo elas formadas por conjuntos de pixels contíguos, espalhados bidirecionalmente e apresentando uniformidade.

Santos (2002) afirma que o nível de subdivisão da imagem em regiões depende do problema a ser resolvido, ou seja, deve-se saber o motivo de se utilizar a segmentação e o critério de finalização do processo envolvido, o qual segue o princípio de isolamento dos objetos de interesse. Dois métodos são citados na literatura: detecção de bordas, o qual se baseia em descontinuidades; e crescimento de regiões, cujo critério é de similaridade.

Na detecção de bordas, a ideia consiste em agrupar pontos até que o objeto de interesse seja reconstruído. Medeiros (2003) compara uma imagem em níveis de cinza a uma superfície topográfica para explicar o crescimento de regiões. A autora diz que, se as bacias hidrográficas fossem imersas, a água necessária para preencher uma bacia poderia se encontrar com a água necessária para preencher a bacia vizinha. O dique construído que evita essa junção de águas forma a linha que segmenta as regiões. A essas linhas dá-se o nome de *Watersheds*.

A técnica de *Watersheds* está associada aos mínimos locais nas imagens, ou seja, aos menores valores de brilho locais. Uma vez que existe uma diversidade de alvos que se misturam, pode ocorrer um excesso de segmentação, o que se torna um problema na utilização desta técnica em Sensoriamento Remoto. Algumas técnicas vêm sendo propostas para minimizar este efeito indesejável, conforme ainda cita a autora. A Figura 3.9 ilustra a aplicação de tal técnica.

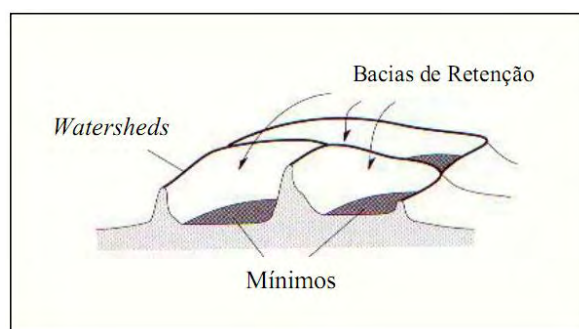


Figura 3.9. Detecção de bordas por watersheds.
Fonte: MEDEIROS (2003)

Já no crescimento de regiões, cada pixel é rotulado inicialmente como uma região. As regiões espacialmente adjacentes são agrupadas segundo algum critério de similaridade, inferior ao limite definido pelo usuário. Assim, as regiões são reagrupadas de acordo com um limiar de agregação (tamanho mínimo a ser aceito para uma região – também definido pelo usuário). Pode-se pensar em um critério de similaridade envolvendo o conceito de autocorrelação espacial, por exemplo.

O processo é repetido até que nenhum outro agrupamento possa ser feito. Este processo gera regiões com contorno fechado, eliminando regiões com número de pixels pequeno. Um dos problemas desta técnica é o fato de que alguns pixels da borda podem ser agregados a uma das regiões vizinhas.

Uma segunda abordagem é a classificação não supervisionada. A classificação de imagens pode ser entendida como a geração de uma nova imagem em que cada pixel possui um novo valor representando um tema (classe ou categoria). Essa associação a categorias se baseia nos níveis de cinza das bandas das imagens e a relação destes com os fenômenos apresentados na imagem. Mais uma vez, os níveis de cinza podem ser substituídos pelos valores assumidos pela superfície contínua interpolada.

Os métodos de classificação automática podem ser divididos em duas categorias, segundo o grau de participação do usuário: classificação supervisionada e não supervisionada. Na classificação supervisionada, o analista contribui com o conhecimento a respeito da área *a priori*, a partir de amostras de treinamento, para estabelecer as classes. Assim, as áreas onde as classes aparecem são identificadas e impostadas no ambiente computacional para que, a partir dos valores correspondentes a estas regiões, sejam calculados os parâmetros estatísticos para cada categoria “classificada” (média e variância, por exemplo).

Já na classificação não supervisionada, o analista, inicialmente, não participa da seleção das classes. Essa tarefa é realizada a partir de algoritmos computacionais que analisam o conjunto de dados e nele identificam as classes mais frequentes. Os algoritmos mais conhecidos baseiam-se nos princípios de agrupamento estatístico (*cluster*).

Os parâmetros mais utilizados para definir os *clusters* são a média, variância e covariância, pelos quais a proximidade espectral dos pixels no espaço multidimensional é definida. Os *clusters* constituem, portanto, classes espectrais. Trata-se de uma operação exploratória para verificar o que é estatisticamente separável ou não, quais as classes mais evidentes e sua relação com o que se deseja discriminar.

O *Isodata* é uma das técnicas de classificação não-supervisionada mais empregadas. Ele funciona através de iterações, nas quais a imagem é sucessivamente varrida e os *clusters* vão sendo modificados, com alterações dos parâmetros estatísticos, agregação de novos pixels, divisão ou fusão de clusters. Na primeira iteração, determina-se um número N de pixels (sementes) em torno dos quais começam a ser gerados os primeiros *clusters*.

As sementes funcionam como média em torno das quais vão se formar os agrupamentos iniciais. Cada pixel da imagem é então agrupado ao *cluster* cuja média

está mais próxima. Na iteração seguinte, as médias dos *clusters* são recalculadas e os pixels reagrupados.

Todo o processo é repetido tantas vezes quantas forem as iterações programadas, nas quais cada pixel pode ser reenquadrado em novos *clusters*, em função da proximidade estatística. O processo pode ser interrompido, porém, quando numa determinada iteração não ocorrer modificação alguma em relação à iteração anterior, indicando que a convergência foi atingida ou quando for atingido o número máximo de iterações.

Após cada iteração os seguintes critérios são aplicados:

- *Clusters* com desvio padrão muito grande são subdivididos para formar *clusters* menores;
- *Clusters* que estão estatisticamente muito próximos são fundidos; e
- *Clusters* com um número pequeno de pixels são eliminados.

Em aplicações em imagens de Sensoriamento Remoto, o processo é terminado com a classificação de cada *cluster* de acordo com a cobertura do solo que ele representa. No entanto, para a metodologia que é proposta neste trabalho, basta identificar as fronteiras de cada *cluster*, não importando, necessariamente, o padrão socioeconômico representado por ele.

3.2.4. Avaliação Estatística e Propagação de Erros

Para se tomar conhecimento de fenômenos estudados, a coleta de dados é essencial. Muitas vezes, a variável observada possui diferentes realizações, assumindo certo grau de variabilidade, o qual pode depender dos mecanismos de coleta ou até mesmo da natureza estocástica dos dados (PAULINO e SOARES, 2004).

Na área de informações espaciais, por exemplo, tem-se, corriqueiramente, a coleta de distâncias e ângulos entre pontos de referência e pontos temáticos. Para um maior controle de qualidade, busca-se obter várias medições dessas mesmas grandezas, assumindo-se um valor médio para ambas.

Neste contexto, a variabilidade, possivelmente, é causada pelos mecanismos de coleta. A calibração dos instrumentos é fator dominante e sua inexistência pode causar o que se chama de erros sistemáticos. Muitas vezes, erros grosseiros – imperícia do

operador do instrumento – são considerados como inerentes aos erros sistemáticos (PAULINO e SOARES, 2004).

A variabilidade causada pela natureza dos dados se refere a medidas quantitativas de uma população estudada ou de uma amostra dessa população. Trazendo para o contexto do urbanismo, exemplificando a importância dos estudos de variabilidade, a leitura da cidade é uma das peças-chave para um eficaz processo de planejamento de políticas públicas, como aponta Nahas (2009). Uma das possíveis leituras é a distribuição da renda municipal, sendo muitas vezes obtidas a partir de censos demográficos.

Para garantir a preservação da identidade dos informantes do censo, geralmente tem-se a coleta e divulgação de resultados censitários em pequenas zonas intraurbanas (tema melhor abordado na seção 3.4.2.). Assim, a informação de renda da população é agregada nessas zonas, podendo-se ter a renda total ou a renda média da população residente em cada zona, mas nunca de cada residência desta zona. Este último é um exemplo prático de quando a variabilidade de um conjunto de dados depende da natureza estocástica do próprio dado.

Ressalta-se, no entanto, que essa natureza estocástica pode ser forçada pelo método de coleta, como no caso dos censos, e não exatamente pelas condições naturais do fenômeno estudado.

O trabalho aqui apresentado disserta sobre a coleta de dados agregados por zonas e sua influência nos processos de leitura da cidade, considerando que o método de coleta exerce papel importante na medida de variabilidade do fenômeno.

Independente de qual seja o gerador da variabilidade, em estatística, utilizam-se alguns indicadores da mesma. Destaca-se, para este trabalho, a variância e o coeficiente de variação.

A variância consiste na relação matemática entre a soma dos quadrados dos desvios em torno da média pela quantidade populacional e é representada pela Equação 8. Na verdade, esse é o conceito de variância populacional. Muitas vezes, torna-se impossível obter os valores de toda a população, partindo-se, então, para a variância amostral, conforme a Equação 9.

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n} \quad (8)$$

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad (9)$$

onde,

σ^2 é a variância populacional;

μ é a média populacional;

s^2 é a variância amostral;

\bar{x} é a média amostral;

x_i é o valor *iésimo* valor medido e;

n é o tamanho da população.

Em muitos casos, a variável de interesse é dependente da variável medida, existindo, portanto, um modelo matemático relacionando as duas grandezas, tal como $Y = f(X)$. Faz-se necessário também obter a variabilidade dos valores gerados por esse modelo. De outra forma, tem-se a necessidade de se conhecer a variância do valor Y .

Para tal determinação, utiliza-se a *Lei da Propagação de Variâncias*, a qual está presente em diversas áreas do conhecimento. Em informações espaciais, por exemplo, as coordenadas de pontos de interesse, obtidas por meio de levantamento de ângulos e distâncias – observações da topografia clássica – podem trazer consigo suas medidas de variabilidade em torno de uma média posicional.

Uma vez que as observações supracitadas podem ser obtidas com sua variância, é possível propagá-las de modo a encontrar a variância das coordenadas topográficas. Na verdade, seguindo essa ideia, a qualquer processo de obtenção de valores, através de modelos matemáticos que relacionam variáveis aleatórias independentes, pode ser atribuído o modelo de propagação de variâncias.

Paulino e Soares (2004) demonstram a obtenção genérica de uma função de propagação de variâncias. A Equação 10 apresenta o modelo para uma única variável, cabendo à Equação 11 a indicação genérica da função para duas variáveis.

$$\sigma_Y^2 = [f'(X = \mu)]^2 \sigma_x^2 \quad (10)$$

$$\sigma_Y^2 = \left[\frac{\partial f}{\partial x_1} \Big|_{\mu} \right]^2 \sigma_{x_1}^2 + \left[\frac{\partial f}{\partial x_2} \Big|_{\mu} \right]^2 \sigma_{x_2}^2 + 2 \left[\frac{\partial f}{\partial x_1} \Big|_{\mu} \right] \left[\frac{\partial f}{\partial x_2} \Big|_{\mu} \right] \sigma_{x_1 x_2} \quad (11)$$

onde,

σ_Y^2 é a variância de Y ;

$f'(X = \mu)$ e $\frac{\partial f}{\partial x} |_{\mu}$ são as derivadas da função Y em $x = \mu$;

σ_x^2 é a variância de x e;

$\sigma_{x_1x_2}$ é a covariância entre x_1 e x_2 .

A aplicação para modelos com mais de duas variáveis segue a mesma ideia, tomando-se o cuidado de se verificar as covariâncias entre elas. É certo, porém, que nem todas as aplicações possuem variáveis correlacionadas, simplificando, portanto, o modelo matemático.

Uma característica da variância é que sua unidade de medida é o quadrado da unidade de medida do fenômeno estudado. Por isso, muitas vezes, utiliza-se o desvio-padrão, sendo este a raiz quadrada da variância, tendo, portanto, a mesma unidade do fenômeno estudado.

De outro lado, o coeficiente de variação (CV) também mede o quanto houve de dispersão em torno de uma média. Sua diferença, no entanto, é que ele é adimensional, possuindo um caráter genérico, o qual permite a comparação entre as dispersões de diferentes fenômenos, com diferentes unidades de medidas, tal como níveis de renda e anos médios de estudos – variáveis censitárias.

Seu valor é dado pela divisão entre o desvio-padrão e o valor médio, conforme apresenta a Equação 12.

$$CV = \frac{\sigma_x}{\mu} \quad (12)$$

Um cuidado que se deve ter com o coeficiente de variação, em sua interpretação, é quanto ao valor médio estudado. Por exemplo, caso haja uma distribuição de valores tais quais a média seja o valor 5 e o desvio-padrão seja 2, ter-se-ia um CV maior que se a média fosse 80 com o mesmo desvio-padrão. Portanto, a variabilidade em torno da média foi a mesma, sendo, porém, a interpretação dos resultados diferente.

Deste modo, é certo afirmar que a análise da variabilidade de uma medida estatística deve ser feita a partir de um conjunto de parâmetros que a meçam, trazendo maior grau de informações quanto ao fenômeno em estudo.

3.2.5. Indicadores de Autocorrelação Espacial

O conceito de dependência espacial pode ser entendido como a tendência a que o valor de uma ou mais variáveis associadas a uma determinada localização assemelhe-se mais ao valor de suas amostras vizinhas do que ao restante das localizações do conjunto amostral (RAMOS, 2002).

Ramos (2002) cita que “a característica que distingue a análise estatística dos dados espaciais é que seu foco principal está em inquirir padrões espaciais de lugares e valores, a associação espacial entre eles e a variação sistemática do fenômeno por localização.”

A Estatística Espacial, portanto, busca verificar se existe ou não dependência espacial num conjunto de dados, podendo o padrão espacial ser aleatório, disperso ou agrupado. A medição do grau de dependência espacial está amarrada à definição da proximidade ou vizinhança entre as feições medidas. Isto é, é necessário obter o grau de influência de um local em relação ao seu vizinho. Como o foco deste trabalho é em dados agregados por áreas, os locais citados nesta seção são representados por polígonos.

Esta informação sobre a vizinhança é armazenada na Matriz de Proximidade, a qual é quadrada (número de linhas e colunas igual ao número de polígonos), podendo ser obtida em função da distância ou da adjacência entre os polígonos. O primeiro caso abrange pesos inversamente proporcionais à distância entre os centróides das áreas, podendo ser definida também uma distância limite. Quanto à adjacência ou condição topológica, a matriz pode ser binária ou estocástica.

Dado o arranjo poligonal genérico ilustrado na Figura 3.10, a matriz de proximidade binária é dada pela inserção do número um na célula correspondente a dois polígonos adjacentes e do número zero, caso contrário.

Quanto à matriz estocástica, o princípio de construção é o mesmo, com a diferença de que os valores armazenados numa linha são divididos pelo somatório de

valores presentes na linha, o que dá uma ideia de graus de influência diferentes a partir do número de vizinhos presentes em um polígono (WONG e LEE, 2005).

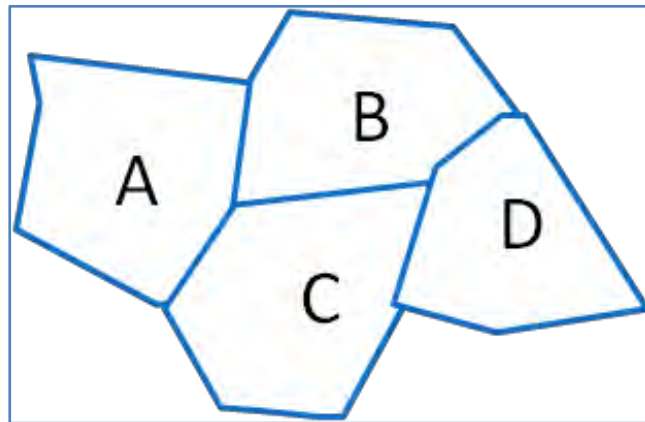


Figura 3.10 Arranjo poligonal genérico.

Tabela 4.1 Matriz de Proximidade Binária

	A	B	C	D
A	0	1	1	0
B	1	0	1	1
C	1	1	0	1
D	0	1	1	0

Tabela 4.2 Matriz de Proximidade Estocástica

	A	B	C	D
A	0	0,5	0,5	0
B	0,33	0	0,33	0,33
C	0,33	0,33	0	0,33
D	0	0,5	0,5	0

Existem vários indicadores de dependência espacial, podendo ser classificados como globais ou locais. DRUK *et al.* (2004) diz que

Os indicadores globais de autocorrelação espacial, como o Índice de Moran, fornecem um único valor como medida da associação espacial para todo o conjunto de dados, o que é útil na caracterização da região de estudo como um todo. Quando lidamos com grande número de áreas, é muito provável que ocorram diferentes regimes de associação espacial e que apareçam máximos locais de autocorrelação espacial, onde a dependência espacial é ainda mais pronunciada.

Dentre os indicadores globais, um dos mais conhecidos e utilizados é o Índice de Moran, o qual varia de -1 (quando há Forte Autocorrelação Negativa, indicando Padrão Disperso) a 1 (quando há Forte Autocorrelação Positiva, indicando Padrão Agrupado). Seu modelo matemático é dado pela Equação 13.

$$I = \frac{n \sum \sum w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{W \sum (x_i - \bar{x})^2} \quad (13)$$

onde,

n é o número de polígonos;

w_{ij} é o valor do peso entre os polígonos i e j ;

W é o somatório dos valores presentes na matriz de proximidade e;

x é o valor do polígono i ou j .

Outro medidor de autocorrelação espacial global é a *General G Statistics*. Sua principal característica é indicar se a dependência espacial é entre polígonos com baixos valores (quando G é grande) ou altos valores (quando G é pequeno). A estatística é representada pela Equação 14.

$$G = \frac{\sum \sum w_{ij} x_i x_j}{\sum \sum x_i x_j} \quad (14)$$

onde,

w_{ij} é o valor do peso entre os polígonos i e j ;

x é o valor do polígono i ou j .

Como indicador local (*LISA - Local Indicator for Spatial Autocorrelation*), o Índice Local de Moran aparece como referência. Para se identificar agrupamentos e quantificar a associação espacial em função de uma vizinhança preestabelecida, utiliza-se a Equação 15.

$$I_i = \frac{(x_i - \bar{x}) \sum_j w_{ij} (x_j - \bar{x})}{s^2} \quad (15)$$

onde,

w_{ij} é o ponderador na matriz de vizinhança W para as localizações i e j ;

x_i é o valor de x na posição i ;

\bar{x} é a média global de x e;

s^2 é a variância global de x .

A Equação 4.3 também pode ser escrita na forma da Equação 16.

$$I_i = \frac{Z_i Wz_i}{s^2} \quad (16)$$

onde,

$Z_i = (x_i - \bar{x})$ é o valor do desvio do atributo no polígono i em relação à média;

$Wz_i = \sum_j w_{ij} (x_j - \bar{x})$ é o valor médio dos desvios dos objetos vizinhos a i e;

s^2 é a variância global de x .

O Índice Local de Moran, que trabalha a partir da análise de covariância entre diferentes unidades de área, possui como interpretação mais direta aquela em que valores significativamente altos e positivos indicam a presença de *cluster* de regiões, independente se os atributos são altos ou baixos. Valores significativamente baixos e negativos indicam um padrão de transição entre um regime espacial e outro (NEVES, 2000).

Para determinar quais clusters representam valores altos ou baixos, utilizam-se os valores de Z_i e Wz_i . Se ambos são positivos ou negativos, então I_i será positivo. Mas, se forem plotados esses valores num gráfico de dispersão dividido em quadrantes, pode-

se observar quais pares indicam presença de clusters de valores altos e quais de valores baixos.

Pares presentes nos quadrantes Q1 indicam o padrão *alto-alto* (valores altos de Z_i e de WZ_i), Q2 indicam padrão *baixo-baixo*, Q3 indicam *alto-baixo* e Q4 indicam *baixo-alto*. Assim, admite-se que o padrão Q1 representa área com valor alto cercada por áreas com valores com mesma situação. De modo análogo, Q2 indica valores baixos com vizinhos baixos. O quadrante Q3, por sua vez, indica valores altos com vizinhos baixos, e Q4, valores baixos com vizinhos altos. Todas estas situações podem ser observadas na Figura 3.11.

Essa análise é feita a partir do BoxMap e/ou do MoranMap. Basicamente indicam a mesma coisa, com a diferença de que o segundo apenas existe se há significância estatística de 95%.

Além do Índice Local de Moran, existem também as estatísticas G_i e G_i^* , que medem a concentração total ou falta de concentração de pares X Y, sendo estes vizinhos, e o mapa de espalhamento de Moran, que não mede diretamente a correlação espacial, mas aponta padrões desta correlação.

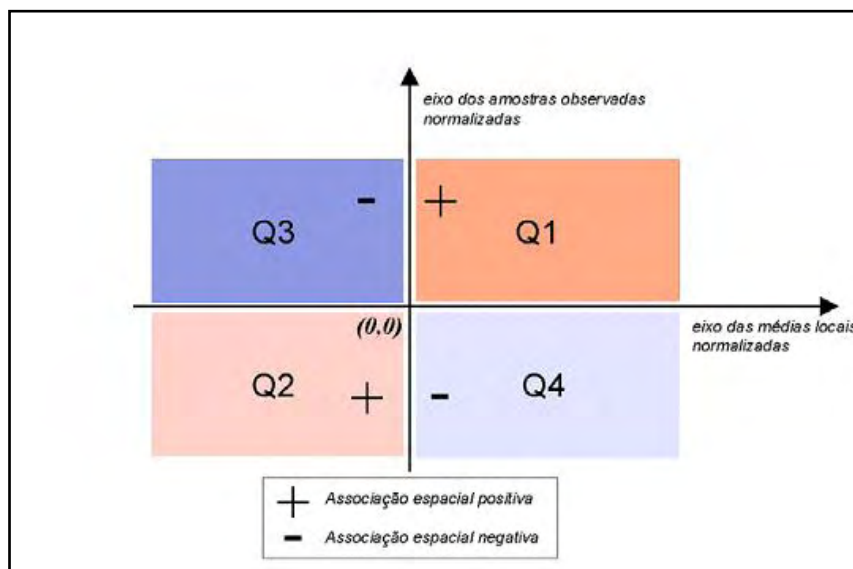


Figura 3.11. Esquema do Gráfico de Espalhamento de Moran.

Fonte: RAMOS (2002)

Ramos (2002), Genovez (2002) e Feitosa (2005) apresentam experimentos relacionados à identificação de *Clusters* e visualização do padrão de agrupamento em

suas áreas de trabalho. Abreu *et al* (2011) utilizaram a análise de *clusters* para identificar áreas segregadas socioespacialmente.

Seguindo a mesma aplicação de Abreu *et al* (2011), a partir da identificação de polígonos (pixels transformados) autocorrelacionados, tem-se caracterizadas as regiões homogêneas socioeconomicamente, do ponto de vista estatístico.

3.3. INFORMAÇÕES SOCIOECONÔMICAS

3.3.1. Dados e Indicadores Socioeconômicos

O conhecimento dos territórios intraurbanos é imprescindível para uma eficaz leitura da cidade e orientação das políticas públicas num município. Ramos (2002) diz que grande parte das características desses territórios é de natureza socioeconômica. Para se conhecer, portanto, um território, deve-se ter em mãos uma representação formal dessas características: os dados socioeconômicos.

Martin (1996) define dados socioeconômicos como dados relacionados a um membro de uma população ou à própria população. No contexto da análise espacial intraurbana, esses dados podem estar associados a endereços ou agregados em zonas geográficas. Ainda segundo o autor, há tempos que autoridades governamentais se interessam por mapear perfis socioeconômicos, como níveis educacionais, concentração de renda, condições de saúde, etc..

Além dos valores de cada variável, como renda familiar e anos de estudos dos cidadãos, alguns indicadores podem ser estabelecidos para ampliar a compreensão das condições populacionais. Um indicador que ainda é utilizado para comparar o poderio econômico dos países é o Produto Interno Bruto (PIB), o qual era erroneamente utilizado para inferir sobre a riqueza e, conseqüentemente, sobre o padrão de vida de uma população (NAHAS, 2009).

Nos anos de 1990, o Programa das Nações Unidas para o Desenvolvimento (PNUD) passou a utilizar o Índice de Desenvolvimento Humano (IDH) para quantificar a qualidade de vida. A metodologia de cálculo desse índice foi adaptada para várias escalas de análise, desde estados, regiões, municípios, distritos e até setores censitários

do IBGE (ABREU *et al*, 2011).

Contudo, esses indicadores não são os melhores medidores de condições socioeconômicas da população. Critérios de renda, educação e longevidade nem sempre são suficientes para uma leitura mais precisa de uma realidade, como aponta Nahas (2009), citando a necessidade de mensuração do acesso ao saneamento básico, habitação, transporte, etc., sintetizando o que se pode chamar de medidas da *qualidade de vida*.

Vitte (2009), citando Allardt (1993), diz que a qualidade de vida pode ser sintetizada em três verbos: *ter, amar e ser*. O primeiro deles representa as necessidades básicas, como recursos econômicos, condições de moradia, emprego, condições de trabalho, saúde e educação. O segundo verbo representa o crescimento pessoal, sendo avaliado através do grau de participação do indivíduo nas decisões e atividades que influenciam sua vida, nas atividades políticas, nas oportunidades de lazer e no desfrute da natureza. Por fim, o terceiro verbo revela os vínculos e contato com a comunidade local, com a família, com colegas de trabalho, participação em organizações e associações e suas amizades.

Percebe-se que nos dois últimos há um grau de subjetividade que torna difícil sua mensuração quantitativa. Além disso, estão ligados diretamente ao indivíduo, e não a um conjunto de pessoas. Por sua vez, o primeiro pode ser medido e avaliado a partir do quanto é fornecido ou oferecido como oportunidades à população residente em um local (NAHAS, 2009).

Assume-se, portanto, que indicadores do primeiro nível de satisfação são suficientes para se ter uma leitura eficiente da realidade socioeconômica urbana. Nahas (2009) apresenta dois indicadores utilizados pela Prefeitura de Belo Horizonte, criados pela parceria entre a Secretaria Municipal de Planejamento da mesma prefeitura e uma equipe multidisciplinar da PUC/MG: Índice de Qualidade de Vida Urbana (IQVU) e o Índice de Vulnerabilidade Social (IVS).

Os dois índices são compostos basicamente por dados sobre habitação, saneamento básico, saúde, educação, transporte, segurança, abastecimento, assistência social, cultura, esportes, outros serviços de infraestrutura, parâmetros ambientais, outros serviços urbanos e variáveis econômicas (NAHAS, 2009).

Diversos indicadores, menos robustos, também foram propostos ou adaptados

para outros municípios. Abreu *et al* (2011) sugeriram o cálculo do IDH por setores censitários, além de realizar a análise de tendências espaciais através da geoestatística. Em outro trabalho, Abreu *et al* (2010a) identificaram padrões de segregação socioespacial a partir da análise de autocorrelação espacial em dados agregados por setores censitários.

Feitosa (2005) procurou desenvolver e avaliar técnicas de mensuração da segregação residencial, buscando captar a dimensão espacial e sua distribuição no Município de São José dos Campos - SP. Ramos (2002) propôs algumas metodologias de análise espacial que englobam boa parte de problemas e necessidades quando se trata de dados socioeconômicos, como minimizar os Problemas das Unidades de Áreas Modificáveis, identificar padrões de agrupamentos de determinado fenômeno e gerar uma superfície de tendências da distribuição de algumas variáveis.

3.3.2. Censos Demográficos

Segundo Oliveira (2008), as operações censitárias remontam aos tempos antigos. A autora faz uma breve revisão da evolução histórica dos censos populacionais no mundo. Em seu estudo, é dado um destaque especial aos censos australiano, brasileiro, canadense, estadunidense e português.

Na Austrália, os censos são realizados a cada 5 anos, sob a responsabilidade do *Australian Bureau of Statistics* (ABS). A unidade de coleta e disponibilização de dados censitários é a seção de recenseamento, a qual cobre todo o país, evitando sobreposição ou omissão. No entanto o ABS está propondo a utilização da malha de quarteirões como menor unidade de divulgação dos dados. O país está passando por nova estruturação da sua geografia estatística, buscando-se definir uma nova unidade de recenseamento, baseada por área estatística (OLIVEIRA, 2008).

Já no Canadá, como aponta Oliveira (2008), a responsabilidade pelo censo é do *Statistics Canada* (STATCAN). O censo também é realizado a cada 5 anos. Como unidade básica de recenseamento, o país utiliza os *Census Tracts*, os quais agregam uma população entre 2500 e 8000 pessoas e encontram-se localizados nas áreas metropolitanas e grandes aglomerados. O STATCAN possui um processo diferenciado de coleta e divulgação dos dados, utilizando a internet, o qual se estabeleceu a partir de 2001 (OLIVEIRA, 2008).

Os Estados Unidos da América tem um intervalo de 10 anos entre um censo e outro, sendo realizado pelo *United States Census Bureau* (USCB). Todos os dados censitários são coletados a partir de uma plataforma computacional móvel para posteriormente serem associados às unidades de área escolhidas para a divulgação, cujos limites são definidos a partir da participação de moradores locais, desde que estes residam em áreas espaciais (OLIVEIRA, 2008)

Em Portugal, por sua vez, O Instituto Nacional de Estatísticas (INE) é o órgão que realiza os censos. Seu intervalo também é de 10 anos e sua configuração Cartográfica se assemelha a do Brasil nos termos das definições dos limites de recenseamento (OLIVEIRA, 2008).

Martin (2000) apresenta a proposta realizada no censo-teste do Reino Unido em 1996. Neste processo, foram criados dois zoneamentos censitários diferentes, um dos quais foi criado para fins de coleta de dados, sendo baseado nos endereços de cada domicílio. O segundo foi definido para a divulgação dos resultados do censo, considerando a homogeneidade social e o tamanho da população inserida.

Destaca-se que, no que diz respeito à cartografia dos censos, os países analisados utilizam-se da cartografia básica nacional em grandes escalas para a definição de suas áreas de coleta de dados, sem, no entanto, haver preocupações com unidades do Cadastro Territorial, o que faz com que este trabalho apresente uma abordagem inovadora na área dos censos.

A seguir, apresenta-se uma breve descrição da estrutura do censo demográfico no Brasil.

3.3.3. Censo Demográfico do IBGE

Uma das fontes de dados socioeconômicos nacionais é o Censo Demográfico, que é realizado regularmente a cada dez anos, consistindo em um levantamento de todo o território brasileiro. Para a garantia de que não haja sobreposição ou falha na cobertura de áreas recenseadas, o IBGE trabalha com os setores censitários, que consistem em uma “unidade territorial criada para fins de controle cadastral da coleta” de dados (IBGE, 2003) e (IBGE, 2011).

Os setores censitários possuem limites de fácil identificação no campo, permitindo a recuperação de suas fronteiras tanto no momento da coleta de dados

quanto em escritório (representação digital dos mesmos). Além disso, suas dimensões são compatíveis com cerca de quatro a cinco quadras urbanas, fato este que permite analisar a distribuição espacial intraurbana das variáveis disponibilizadas no censo.

O Censo Demográfico de 2000, disponibilizado no nível de setores censitários, possui mais de 3200 variáveis (IBGE, 2003), as quais informam características sobre o sexo, a idade, a renda e a escolaridade das pessoas, além de condições gerais dos domicílios, como padrão de construção, número de banheiros, abastecimento de água, coleta de esgoto e de resíduos sólidos, entre outras informações. Ressalta-se que, em julho de 2011, foram divulgados os resultados da sinopse do Censo Demográfico 2010 (IBGE, 2011), contendo apenas informações sobre os domicílios e idade e sexo dos moradores.

No intervalo entre o censo de 2000 e o de 2010, o IBGE realizou a contagem populacional em alguns municípios brasileiros (IBGE, 2008a). Esta coleta de dados possui apenas informações sobre as características dos domicílios e das pessoas residentes. É um conjunto menor de variáveis, porém, de grande valor para a leitura do município, permitindo-se fazer algumas análises espaço-temporais da realidade populacional.

Todavia, os setores censitários têm seus limites modificados de um levantamento para o outro. Isso acarreta em dados agregados de maneiras diferentes. Uma simples diferença na agregação espacial de dados pode modificar completamente o resultado desta agregação, como é visto na seção 3.4.2. deste trabalho.

Muitas vezes, as modificações dos limites censitários se devem ao aumento no número de domicílios de uma época para outra. Cada setor abrange aproximadamente de 250 a 400 domicílios. Logo, um setor com quantidade superior ao limite máximo é desmembrado. Mesmo assim, alguns limites podem mudar independente da quantidade de domicílios presentes, como mostra a Figura 3.12, extraída de Abreu *et al* (2010b).

De qualquer forma, a utilização de dados censitários agregados por setores censitários é primordial para uma qualificada gestão pública. O uso destes dados não impõe grandes investimentos para um município, uma vez que o IBGE disponibiliza os limites censitários em arquivo digital no formato *shapefile*, que, apesar de ser arquivo proprietário, não possui patente de utilização, além das planilhas contendo todas as variáveis do censo e da contagem populacional.

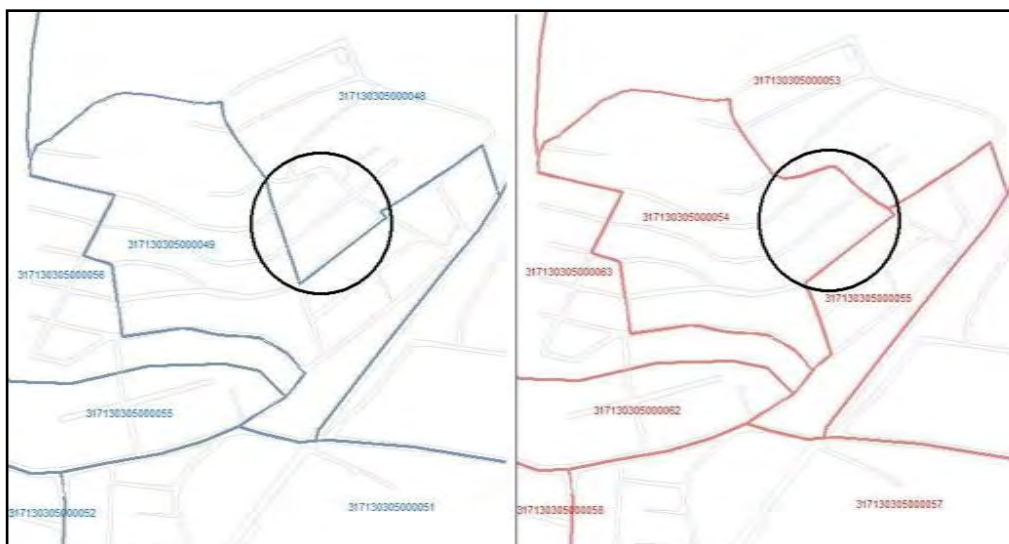


Figura 3.12. Setores Censitários 2000 em azul e setores censitários 2007 em vermelho, em Viçosa – MG.

Fonte: ABREU *et al* (2010b).

A seguir, são apresentadas as vantagens e desvantagens da utilização dos setores censitários na caracterização socioeconômica intraurbana.

3.3.3.1. Vantagens na utilização de setores censitários na caracterização socioeconômica intraurbana

A primeira vantagem é o fato de o censo ser um levantamento sistemático, abrangendo todo o território, assegurando a plena cobertura do país (IBGE, 2003). Assim, todo o território de um município é levantado. A disponibilização dos dados censitários se dá de forma espacializada, permitindo estudos em alguns locais numa cidade, ou então compreendendo a distribuição espacial de variáveis e indicadores por todo o município, como bem apresentado na literatura, podendo-se citar, entre outros, Ramos (2002), Feitosa (2005) e Abreu (2009), conforme visto na seção 4.3.1.

Por serem espacializadas, as informações censitárias podem ser cruzadas com outras, provenientes de diferentes fontes de dados espaciais. Tem-se, como exemplo, a quantificação de moradores presentes na área de influência de um PSF, ou então o número de idosos considerados usuários potenciais de determinada linha de ônibus.

Além disso, como segunda vantagem, tem-se que os setores censitários disponibilizam extenso conjunto de variáveis, que indicam aspectos populacionais e

domiciliares, como idade, sexo, alfabetização, renda do chefe de família, número de sanitários por domicílio, presença de água encanada, coleta de esgoto e de lixo e etc..

Esse conjunto de variáveis é muito maior que aqueles encontrados nas secretarias e empresas públicas. Como mostrado anteriormente, todas essas informações podem ser utilizadas em processos decisórios, mesmo que não sejam detalhadas em nível de parcelas, o que não impede uma boa leitura minudenciada da cidade, já que diagnósticos socioeconômicos não necessitam de pormenores em grandes escalas.

Como outra vantagem enumerada neste trabalho, pode-se pensar na atualização, também sistemática, desses dados. Apesar de o censo ser realizado a cada 10 anos, a avançada tecnologia utilizada pelo IBGE na realização da Contagem Populacional de 2006 e no Censo Agropecuário de 2007, além dos avanços já alcançados para a realização do Censo 2010 (IBGE 2010), permitem pensar que não seria difícil atualizar os censos em períodos menores de tempo, como 5 anos, de acordo com o proposto em IBGE (2008c).

O uso de PDA's e receptores GPS permite a coleta de dados de maneira mais rápida e precisa, com crítica imediata no momento em que os dados são coletados, preenchimento de quesitos obrigatórios, controle no preenchimento por saltos automáticos, transmissões diretas dos dados coletados para o sistema central, dispensa de transporte de grandes volumes de papel e manuseio e georreferenciamento de todas as unidades visitadas nas áreas rurais (estabelecimentos agropecuários, de saúde e educação), como se verifica em IBGE (2008b).

É importante salientar que as coordenadas das unidades recenseadas auxiliam na estruturação do Cadastro Nacional de Endereços para Fins Estatísticos – CNEFE, como aponta IBGE (2008c).

Ainda em IBGE (2008b), o então presidente da instituição, Doutor Eduardo Pereira Nunes, apresenta uma proposta para um modelo operacional para um censo contínuo. O objetivo “é produzir informações sobre características da população e dos domicílios sempre atualizadas com grande detalhamento geográfico e temático” (IBGE, 2008b).

Portanto, além de maior frequência de atualizações, a tecnologia da informação possibilita que dados, coletados de forma digital, sejam processados de maneira mais rápida e segura, permitindo acréscimo de novas informações, que podem ser mais detalhadas, como nível de alfabetização de todos os moradores, rendimento nominal,

total de domicílios, entre outras.

Certamente que outras vantagens poderiam ser abordadas, pois são inúmeras as aplicações de dados censitários em análises espaciais intraurbanas. Além disso, o avanço das tecnologias e das ciências de coletas e processamentos de dados mostra que a atualização em maior frequência é possível, principalmente se apoiada em um sistema de gestão eficiente e seguro.

3.3.3.2 Desvantagens na utilização de setores censitários na caracterização socioeconômica intraurbana

Ramos (2002) explica que o significado de território se aproxima ao de comunidade que, citando Milton Santos, são “espaços seletivamente transformados pelas exigências funcionais da sociedade urbana”. Assim, um território não pode ser delimitado simplesmente por condicionantes físicos ou logísticos, mas sim por níveis de interação social da população inserida no espaço urbano.

Ramos (2002) aponta o geoprocessamento e a análise geográfica quantitativa como principais ferramentas para a criação de modelos capazes de identificar e representar os diferentes territórios intraurbanos. A proposta apresentada na seção 4.3.2 desta dissertação é um exemplo disso.

Portanto, um território intraurbano não é necessariamente definido por limites físicos ou logísticos. Todavia, um setor censitário é definido exatamente por esses dois aspectos, correndo o risco de não possuir homogeneidade interna do ponto de vista do comportamento socioeconômico de uma localidade. Aqui se encontra a primeira desvantagem em sua utilização para o entendimento socioeconômico intraurbano.

Por outro lado, não se pode esperar que o IBGE identifique as zonas homogêneas intraurbanas de todos os municípios e delimite cada setor censitário a partir de outros critérios que não sejam o físico e o logístico. Para um país de dimensões continentais, criar e gerenciar os limites de todos os setores censitários a partir de critérios de homogeneidade poderiam se tornar inviável tanto tecnicamente quanto financeiramente, consistindo numa segunda desvantagem.

Com o acelerado crescimento populacional urbano, periodicamente, novos setores censitários têm de ser criados ou outros desmembrados, de modo a compatibilizar o tamanho da população de cada setor, obedecendo ao limite de 400

domicílios (IBGE 2003). Contudo, nem sempre os limites de um setor num censo são os mesmos em outro censo. A Figura 3.12, apresentada anteriormente, indica os limites de setores censitários do Censo de 2000 e da Contagem Populacional de 2007, respectivamente, na região do Bairro Santo Antônio, em Viçosa, Minas Gerais. Nota-se a mudança brusca ocorrida na representação da área nos dois levantamentos.

Além disso, o IBGE busca conformar os limites censitários com os limites administrativos ou outras feições de interesse. Desse modo, encontram-se setores cujos limites são linhas de cumeada, topos de morro, cursos d'água e, mais comumente encontrados, os eixos centrais dos logradouros. A busca por limites físicos muitas vezes inviabiliza a representação de territórios intraurbanos e impede que a relação de vizinhança seja considerada na definição dos limites censitários.

Muitas vezes, os setores censitários vão de encontro ao conceito de dependência espacial, baseado na Primeira Lei da Geografia. Como os eixos centrais de logradouros constituem divisa setorial, as casas de um lado da rua pertencem a um setor, enquanto casas do outro lado pertencem a outro setor. Isto implica dizer que um logradouro não possui homogeneidade socioeconômica ao longo de seus trechos, contrariando o conceito de valor único por trecho de logradouro presente no conceito de planta genérica de valores (GARCIA, 2007). Dá-se um destaque à exceção da afirmação anterior quando ocorre a heterogeneidade a partir do momento em que alguns logradouros constituem divisas territoriais, tais como grandes avenidas, ou divisas entre áreas altamente segregadas.

Para regiões centrais da cidade, esse fato pode não apresentar grandes diferenças no que diz respeito à homogeneidade, contudo, nas regiões periféricas, notam-se setores censitários que englobam áreas que não apresentam proximidade urbana, impedindo relações de vizinhança, sendo, portanto, setores com grande heterogeneidade interna. Isto pode ser verificado quantitativamente observando-se as variâncias de algumas informações presentes nas planilhas disponibilizadas nos censos. A figura 3.13 apresenta uma região no Município de Timóteo, Minas Gerais, em que um setor censitário de 2000 engloba três possíveis territórios ou parte de territórios diferentes, divididos por cursos d'água.

A figura 3.14, por sua vez, apresenta dois setores censitários que possuem como limite em comum um eixo central de logradouro, dividindo o Bairro Inácio Martins em dois. O setor à direita agrega as residências da direita às residências do Bairro Julia

Molla. Por outro lado, o setor à esquerda agrega as residências aos Bairros Boa Vista e Vau Açu. Verifica-se, portanto, que a vizinhança existente no eixo de logradouro que divide os dois setores foi ignorada, aumentando a heterogeneidade nos dois setores, por incluir residências que não possuem proximidade urbana entre si.

Todavia, entende-se que o IBGE precisa estabelecer limites de fácil identificação no campo, tanto para facilitar os recenseadores nos levantamentos, quanto para facilitar a representação e recuperação dos setores censitários, sendo mais objetivos nas delimitações.

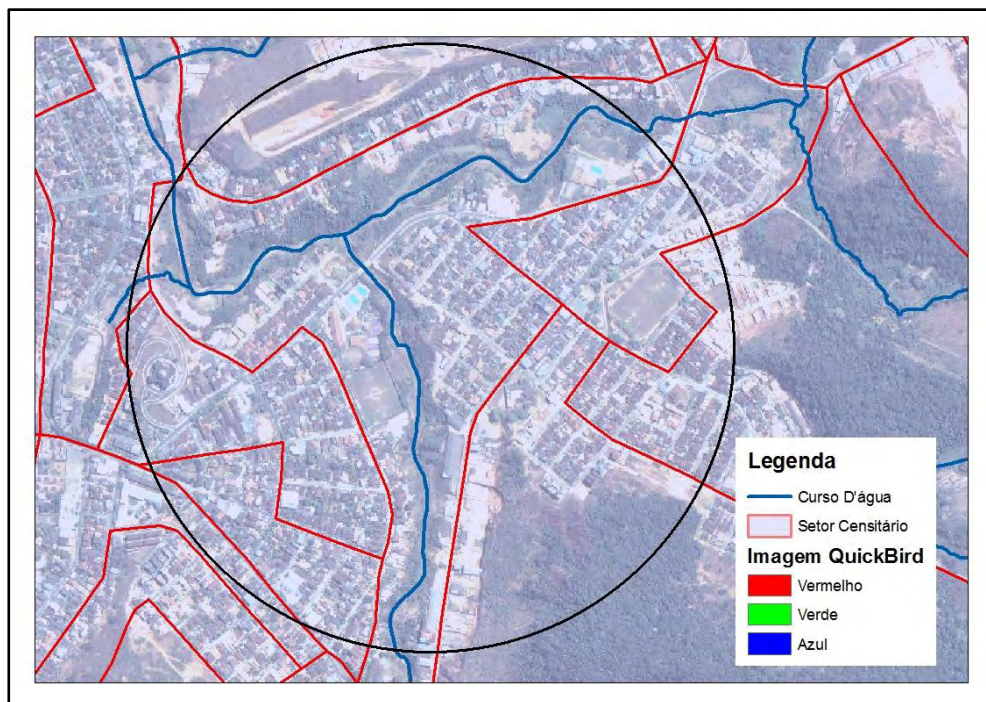


Figura 3.13 Setor censitário 2000 de Timóteo – MG

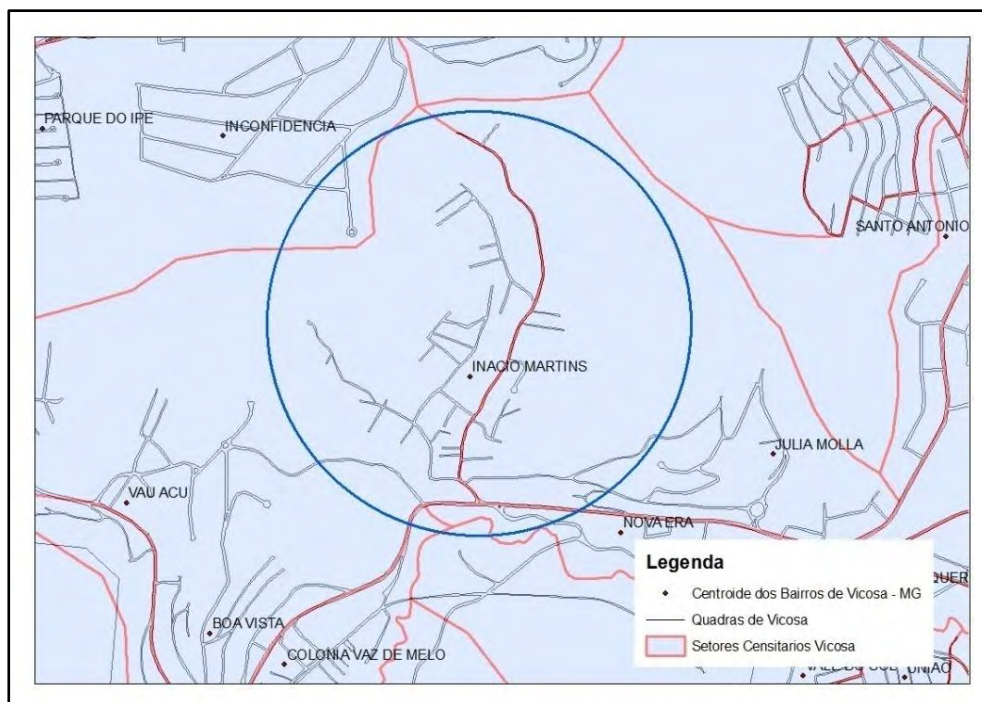


Figura 3.14 Setores censitários no Bairro Inácio Martins, Viçosa – MG

3.4. ZONEAMENTO TERRITORIAL SOCIOECONÔMICO

3.4.1. Relações de Vizinhança na Configuração dos Territórios

Para Bourdieu (1998) citado por Portugal (2009), características econômicas, culturais, políticas e educacionais da sociedade, denominados por ele de capital social e de capital econômico, permitem que as pessoas se mantenham distantes física, social e simbolicamente das pessoas e coisas indesejáveis e, por outro lado, próxima às desejáveis. Essa proximidade no espaço físico permite as realizações das relações sociais entre as pessoas e, a partir desta lógica, as cidades se estruturam física e socialmente (PORTUGAL, 2009).

Portugal (2009) discursa sobre como a Lei de Terras, a partir de 1850, estabeleceu o regime de propriedade fundiária e influencia até hoje a venda de lotes urbanos como grande oportunidade de negócios. O acesso à terra se condiciona à posse econômica, construindo um espaço hierarquizado. A autora ainda explana sobre a formação deste espaço, dizendo que, a partir da abolição da escravatura e da vinda de imigrantes

estrangeiros para o Brasil, associado à “proteção” latifundiária, surgiram novos loteamentos em centros urbanos de muitas cidades.

A partir do início do século XX, o Estado se viu, então, obrigado a reger esse tipo de ocupação, estabelecendo regulamentações do uso do solo e cobrando impostos territoriais. Essa intervenção, aliada à localização dos lotes, passa a constituir um dos fatores de caracterização da paisagem urbana. Assim, além da posse do capital, o Estado também caracteriza o espaço a partir de interesses do capital especulativo (FERREIRA, 2005, citado por PORTUGAL, 2009). Nesse sentido, a opção do Estado por manter uma estrutura adequada e bem equipada em algumas regiões da cidade está atrelada à tendência de obter benefícios para si, através de pagamento de taxas e atração de investimentos para o município (SINGER 1979, apud PORTUGAL, 2009).

As relações sociais e econômicas estabelecidas pela sociedade ocorrem no espaço a partir da proximidade dos lotes, o que dá um aspecto de nivelamento e igualdade (PORTUGAL, 2009). As pessoas ocupam os espaços físicos que a renda lhes permite, cabendo, portanto, à população de baixa renda o que não interessa às camadas da sociedade com maior poder aquisitivo.

Santos (2009), citando Correa (2005), afirma que o surgimento de favelas, ocupações irregulares de encostas e invasões de terrenos públicos ou privados é a produção do próprio espaço pela camada de baixa renda, sendo, então, fruto dos meios de produção capitalistas.

Cupertino *et al* (2009) mostram que o processo de produção do espaço urbano por seus diferentes atores, gerando espaços diferentes entre si, os quais bloqueiam as relações sociais entre suas categorias sociais, é fruto do que se chama de autosegregação e de segregação imposta. O primeiro é quando a camada mais alta busca morar em locais distantes daquilo que é indesejado. O segundo é quando não sobram opções de onde e como morar para a classe baixa. Ou seja, terras sem atrativos e infraestrutura não convidam grupos das classes sociais mais elevadas (CORREA, 2000 apud PIRES e CUNHA, 2009).

Portanto a presença de infraestrutura é responsável, mesmo que indiretamente, pela formação de territórios socioeconômicos intraurbanos. Por conseguinte, as relações sociais construídas dentro de um território estão pautadas em como o poder público olha para esse território.

Os equipamentos e serviços urbanos oferecidos por uma prefeitura a uma localidade podem configurá-la como um território. Ressalta-se que a presença de infraestrutura influencia diretamente na qualidade de vida da população, conforme citado anteriormente, na seção 3.3.1.. Identificar regiões com índices de qualidade de vida semelhantes é também identificar territórios intraurbanos.

A questão passa a ser a definição da escala de análise desses índices. Muitas vezes, a agregação de dados desconsidera as relações sociais existentes num determinado local, como um logradouro, por exemplo, que, na definição dos setores censitários, é utilizado como um limite.

3.4.2. Problema de Unidades de Áreas Modificáveis e Zoneamento Territorial

Segundo Wong (1995), a agregação de dados, como na divulgação dos resultados do Censo Demográfico, envolve várias escalas espaciais e diferentes agrupamentos de porções superficiais. Como resultado, tem-se que as mesmas variáveis estatísticas podem ter resultados diferentes, dependendo de como forem agregadas geograficamente. Este fato é consequência do *Modifiable Areal Unit Problem – MAUP* (Problemas de Unidades de Áreas Modificáveis).

Wong (1995) ainda cita que este termo foi usado pela primeira vez por Openshaw e Taylor, em 1979, embora Gehlke e Biehl, em 1934, já tivessem levantado o problema. Robinson, em 1950, colocou esta questão metodológica no contexto da Falácia Ecológica, como aponta Wong (1995).

Segundo Dias *et al* (2002), a Falácia Ecológica, termo utilizado em ciências sociais e epidemiológicas, é o fenômeno que envolve a conclusão imprópria sobre as relações no nível individual a partir de dados e informações agregados por área. É o que ocorre quando se faz inferência sobre a renda de uma família a partir do valor da renda média do setor censitário a que pertence essa família.

Ramos (2002) cita que o *MAUP* é composto de dois tipos de efeitos. O Efeito de Escala ocorre quando a variação da escala de análise implica num diferente resultado estatístico para um mesmo conjunto de dados. Como exemplo, Abreu (2009) mostra este efeito comparando o Índice de Desenvolvimento Humano (IDH) no nível municipal e no nível de setores censitários, para o Município de Viçosa, MG.

O trabalho supracitado mostra que o município tem um índice de 0,929 na dimensão educação do desenvolvimento humano, levando-se a inferir, erroneamente, que toda a população viçosense tem um alto grau de escolaridade. No entanto, ao se analisar a distribuição local dos níveis de escolaridade, percebe-se que as regiões periféricas da cidade apresentam valores muito baixos nesta mesma dimensão, destacando-se o Bairro Nova Viçosa, com um valor de 0,620, como mostra a Figura 3.15.

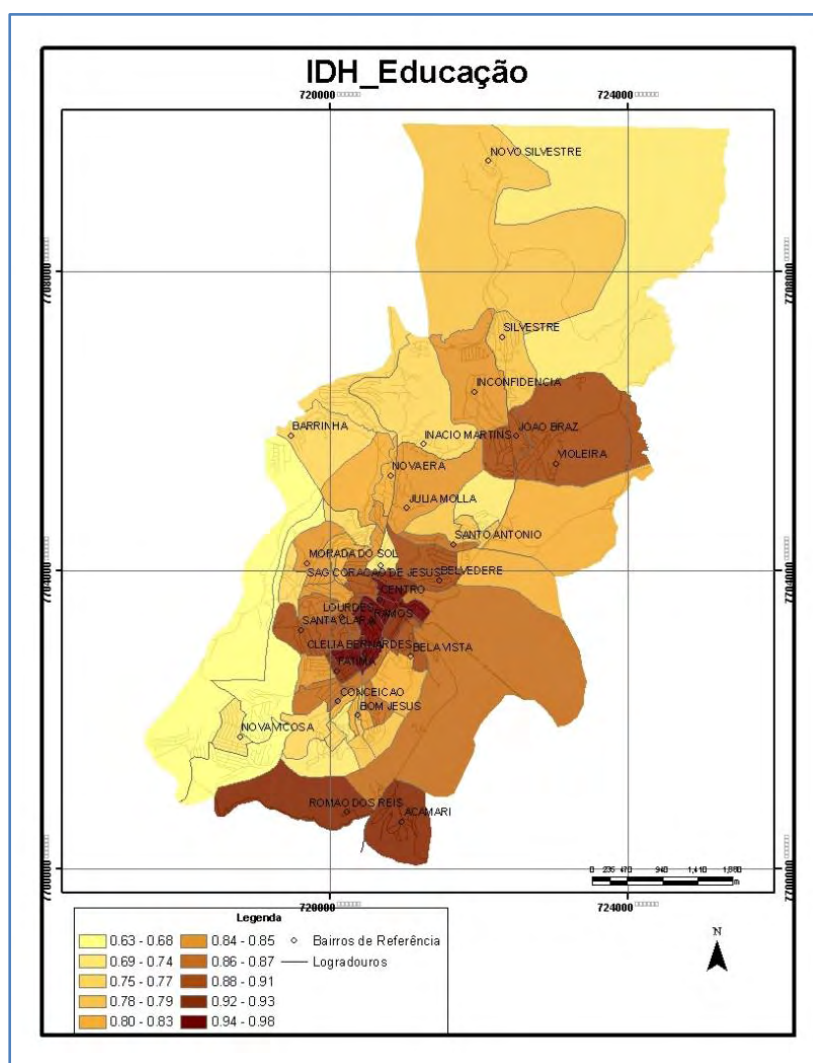


Figura 3.15 – IDH-Educação Viçosa 2000

Fonte: Abreu *et al* (2011)

O segundo efeito é o de Zoneamento, que consiste na variação de um resultado estatístico a partir de diferentes agrupamentos em uma dada escala. Logo, destaca-se que os dados agregados por áreas estão condicionados à forma como foram coletados

e/ou agregados, ou seja, os limites dos polígonos definem seus valores. Caso um limite de um polígono fosse diferente, certamente o valor representado por esse polígono seria diferente.

Assim sendo, caso os dados do censo também fossem agregados por bairros, os valores das variáveis representados por esses bairros seriam diferentes dos apresentados por setores censitários. Nota-se que a escala é a mesma e o que mudou foi o limite de agregação dos dados.

Seguindo essa linha de pensamento, não faria sentido realizar uma análise temporal nos dados contidos naqueles setores censitários apresentados na Figura 3.12, pois uma vez que seus limites são diferentes, os valores do setor censitário de 2007 não refletem apenas a variação que poderia ter ocorrido no tempo, mas também a inserção de novas residências no mesmo.

Os efeitos do *MAUP* nunca poderão ser totalmente removidos. Nesse sentido, duas abordagens distintas são encontradas na literatura. A primeira procura métodos e modelos matemáticos capazes de minimizar as instabilidades do *MAUP*. Nesse âmbito, Ramos (2002) cita autores como Openshaw, Wrigley, Green, Flowerdew, Bailey e Gattrel.

Ramos (2002) apresenta evidências do *MAUP* dentro de análises de dados agregados por área para a cidade de São Paulo - SP. Além disso, ele apresenta um método estatístico muito utilizado para o tratamento das instabilidades impostas pelo *MAUP* a dados socioeconômicos do tipo taxas populacionais e proporções, quando estes estão agregados em zonas projetadas sem se respeitar os critérios de homogeneidade e tamanho da população.

Dias *et al* (2002) aplica a mesma metodologia para tratar a taxa de mortalidade infantil nos bairros do Rio de Janeiro - RJ. Trata-se da *Estimação Empírica de Bayes*, proposta por Bailey e Gattrel, em 1995, “que consiste na estimação de novas taxas ou proporções a partir de um conhecimento a priori sobre um conjunto de variáveis observado” (RAMOS, 2002).

A outra abordagem dada ao *MAUP* reconhece o problema e o associa à característica interna à zona, formulando critérios capazes de gerar subdivisões territoriais em função da homogeneidade e tamanho da população.

Ramos (2002) diz que

um sistema de zonas é uma forma de descrição muito simples. As bordas da zona devem conter em si significado substantivo, como, por exemplo, o lugar onde elas representem fronteiras ou transições segundo algum padrão específico. As áreas delimitadas devem, também, conter significado geográfico.

Essa foi a proposta apresentada em Martin (2000), na qual o autor mostra como encontrar o melhor zoneamento para as áreas de saídas dos dados do censo do Reino Unido. Essas áreas de saída se comportam como os setores censitários no censo realizado pelo IBGE no Brasil. A proposta buscava gerar zonas com igual distribuição populacional, tamanhos aproximadamente iguais e homogeneidade socioespacial.

Quanto à construção de uma zona que represente um território, deve-se tomar cuidado com as questões abordadas anteriormente. Openshaw (1996) diz que as fronteiras das zonas devem abranger apenas os microdados homogêneos entre si, e somente desse modo um território estaria bem representado.

Ramos (2002) discorre sobre o que pode definir a existência ou não de territórios. Para cada comportamento ou cada condição populacional analisada, podem-se ter diferentes configurações territoriais. Quando se busca a homogeneidade de uma variável socioeconômica, por exemplo, deve-se maximizar a correlação interna às áreas, tendo em contrapartida a minimização da correlação entre as áreas (RAMOS, 2002) e (MARTIN, 2000).

O que se nota é que, para se ter um zoneamento socioeconômico, deve-se atentar a dois fatores determinantes. O primeiro é o indicador socioeconômico espacializado mais adequado do ponto de vista de sua eficácia e/ou do ponto de vista das variáveis que se tem à disposição para sua construção. A segunda questão a ser abordada é a forma como este indicador é espacializado (a partir de pontos – microdados – ou superfícies) e como é feita a agregação espacial.

4. MATERIAIS E MÉTODOS

Uma zona deve conter um significado geográfico, conforme afirma Openshaw (1996) ao dizer que “suas bordas devem ser definidas de modo que produzam dados que se conformam com o que pode representar um modelo, minimizando o erro desse modelo. O sistema de zoneamento age, portanto, como um detector de padrões espaciais”.

Martin (2000) propõe a criação das áreas de saída do censo do Reino Unido, baseado nos critérios de forma, homogeneidade e tamanho da população. Estes critérios foram propostos por Openshaw em 1977, conforme cita Ramos (2002), acrescentando ainda densidade e compacidade. Estes fatores garantem uma agregação controlada de dados.

Ramos (2002) cita que diferentes características socioeconômicas se dão a partir da interação existente entre o espaço geográfico e a identidade coletiva da população que ocupa esse espaço. Logo, a partir do estabelecimento da escala de análise espacial (bairros, quadras, lotes, etc.), é possível desenvolver medidas quantitativas de condições socioeconômicas da população residente em cada território e representá-las espacialmente.

Alguns fatores socioeconômicos populacionais podem ser tratados como sendo contínuos no espaço. Essa ideia é defendida por Martin (1996) e trabalhada em Ramos (2002), sob a justificativa de que indicadores socioeconômicos possuem essa continuidade espacial, tanto matematicamente quanto geograficamente.

A proposta que apresenta neste trabalho é a identificação de regiões homogêneas a partir de uma superfície contínua que represente a distribuição espacial de indicadores socioeconômicos. Deste modo, buscou-se verificar, conforme apresenta Martin (2000), que regiões homogêneas procuram maximizar a correlação interna e minimizar a correlação externa às áreas.

Assim, se a superfície é gerada por interpoladores geoestatísticos, os quais são baseados na continuidade espacial, pode-se dizer que a 1ª Lei da Geografia está influenciando os valores dos pixels nos modelos de dados matriciais, ou seja, pixels vizinhos tendem a ser mais parecidos entre si do que pixels afastados.

Quando isso não acontece, tem-se uma variação espacial abrupta do padrão representado nas células. Portanto, uma discretização dessa superfície pode mostrar fronteiras ou limites entre os diferentes padrões de indicadores socioeconômicos. Essa discretização pode ser realizada a partir da segmentação ou da classificação não supervisionada da superfície.

A metodologia de pesquisa é de natureza exploratória e aplicada, por meios documentais. São explorados métodos geostatísticos para a interpolação de indicadores socioeconômicos espacializados e geração da superfície de indicadores, além dos métodos de agregação espacial para a geração de zonas homogêneas.

Num primeiro momento, é testada a homogeneidade espacial da atual configuração dos setores censitários da área urbana de Viçosa, MG. A proposta se baseia no estabelecimento de alguns indicadores socioeconômicos a partir de variáveis de renda e educação, às quais possuem variâncias associadas. O entendimento de homogeneidade se dá a partir da análise dessas variâncias e de suas propagações nos indicadores definidos.

A segunda parte consiste em realizar testes de agregação de dados. A partir da sua representação por meio de superfícies contínuas, segmenta-se a mesma, buscando-se encontrar áreas com certo grau de semelhança, definindo-se, então, as fronteiras de homogeneidade.

Dadas essas fronteiras, definem-se as regiões homogêneas, as quais determinam a criação de novos limites censitários. A agregação dos dados originais se dá por meio dessas regiões, testando-se, também, o grau de variabilidade interna dos dados agregados.

Por último, com o encontro das bordas homogêneas, ou limites censitários homogêneos, propõe-se a conexão dos setores censitários ao Cadastro Territorial Multifinalitário, buscando-se também a amarração dos limites das duas representações (zonas homogêneas e setores censitários homogêneos) aos limites parcelares propostos pelas Diretrizes para o Cadastro. Neste ponto, discute-se sobre a integração dos dados do Censo do IBGE ao CTM, apontando vantagens, desvantagens e proposições que minimizem as desvantagens.

Adianta-se, aqui, que a proposição feita é de cunho conceitual, não havendo aplicação prática que modifique a atual estrutura de setores censitários para a área estudada. Tal feito seria possível apenas trabalhando-se com dados originais do censo,

advindos dos levantamentos do censo – o que não foi possível neste trabalho – além, claro, da validação do IBGE para a proposta de integração de dados censitários ao CTM.

Portanto, pretende-se iniciar, com a apresentação da metodologia indicada e dos resultados, uma discussão sobre uma conexão sistemática dos setores censitários com a base cadastral alvitrada pelas Diretrizes do Cadastro.

4.1. ÁREA DE ESTUDO

A aplicação desta pesquisa é de natureza genérica, podendo ser realizada em qualquer município brasileiro. A primeira parte do trabalho foi aplicada ao Município de Viçosa, que está localizado ao norte da Zona da Mata de Minas Gerais, a 229 km de Belo Horizonte, situando-se entre as latitudes de 20° 41' 20" S e 20° 49' 35" S e, entre as longitudes de 42° 49' 36" W e 42° 54' 27" W, a uma altitude média de 650 metros, abrangendo uma área de 300,2 km². A Figura 4.1 apresenta o município com destaque para sua área urbana.

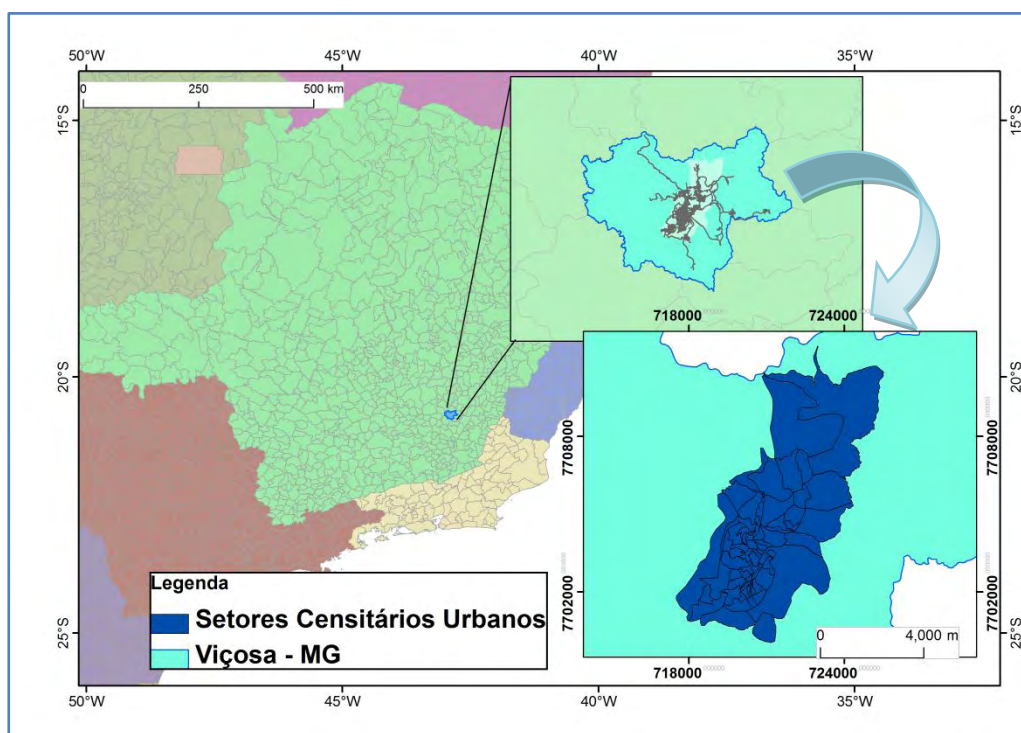


Figura 4.1 - Área de estudo

Ressalta-se que a segunda parte do trabalho utilizou dados genéricos, advindos de Isaaks e Srivastava (1989), pois não foi possível obter os microdados do censo do IBGE para o município escolhido devido, principalmente, ao fato de eles não terem sido divulgados até o fechamento desta dissertação.

Quanto à terceira parte da pesquisa, a proposta de integração foi ilustrada em algumas regiões do distrito sede de Viçosa-MG. Destaca-se, no entanto, que os lotes urbanos apresentados foram também simulados para este trabalho por não haver a representação espacial dos mesmos no Departamento de Cadastro da Secretaria Municipal de Fazenda.

Portanto, para a recuperação de seus limites, lançou-se mão da vetorização manual, a partir da interpretação visual das feições representadas numa imagem IKONOS, datada de 2007, ortorretificada por SANTOS (2008), cuja extensão compreende a Bacia do Ribeirão São Bartolomeu (Viçosa – MG).

É muito importante destacar que, em momento nenhum, este trabalho defende a utilização de imagens orbitais, mesmo que de alta resolução espacial, na obtenção de limites de parcelas urbanas para fins cadastrais. A aplicação dada aqui é apenas ilustrativa, já que tal operação foi escolhida devido ao baixo custo associado ao levantamento dos lotes dos trechos selecionados na simulação, uma vez que a imagem pertence ao Departamento de Engenharia Civil da UFV.

4.2. MATERIAIS UTILIZADOS

Foram utilizados neste trabalho os seguintes materiais:

- Dados
 - Dados do Censo 2000, agregados por setores censitários, para o Município de Viçosa – MG, disponibilizados pelo Projeto Viçosa Digital, através de planilhas no formato *xls;
 - Malha dos setores censitários de Viçosa – MG, compatibilizados com o Mapeamento Urbano Básico (MUB) da cidade, também disponibilizada pelo Projeto Viçosa Digital, no formato *shapefile*;
 - Arquivos digitais (*shapefile*) das quadras e eixos de logradouros do MUB de Viçosa – MG, igualmente pertencentes ao Projeto Viçosa Digital;

- Imagem Ikonos, datada de 2007, com resolução espacial de 1m, ortorretificada por Santos (2008), disponibilizada através do Plano de Segurança da Água/UFV - São Bartolomeu, Viçosa – MG;
 - Tabela de dados presente em Isaaks e Srivastava (1989), contendo 470 pontos com respectivas coordenadas e valores de duas variáveis genéricas V e U.
- Softwares
 - Suíte ArcGIS 9.3.1, utilizados para edições de dados, interpolação por Krigagem, geração dos Índices Globais de Autocorrelação Espacial e geração de *layouts*;
 - Idrisi Andes, para a detecção de padrões espaciais dos pixels das superfícies geradas;
 - TerraView 4.0.0, para a geração dos Índices Locais de Moran, uma vez que esse software também gera este indicador quando não há significância estatística, ao contrário do ArcGIS;
 - Pacotes Microsoft Office 2007.

4.3. METODOLOGIA

4.3.1. Avaliação da Homogeneidade Socioeconômica dos Setores Censitários

O IBGE, no Censo 2000, dividiu suas variáveis em duas classes: Variáveis Alfabéticas e Variáveis Numéricas. As primeiras continham os nomes das subdivisões geográficas das pesquisas censitárias, indicando sua existência ou não-existência, dando, quando possível, sua identificação (IBGE, 2003).

As Variáveis Numéricas, por sua vez, foram divididas em “branco” (quando havia alguma omissão do valor para a proteção ao informante do dado), “0” (para as ocorrências nulas) e “valor” (para as ocorrências não-nulas ou não-omitidas) (IBGE, 2003).

Por sua vez, a classe “valor” também se dividiu em “número de casos” (indicando a quantidade de pessoas, domicílios, etc.), “total” (soma de valores de

rendimento, número de anos de estudos, etc.), “média” e “variância” (IBGE, 2003). Assim, pode-se dizer que existem dois tipos de variáveis de valor: aqueles que indicam somas e os que indicam médias. A Figura 4.2 ilustra a classificação das variáveis do censo.

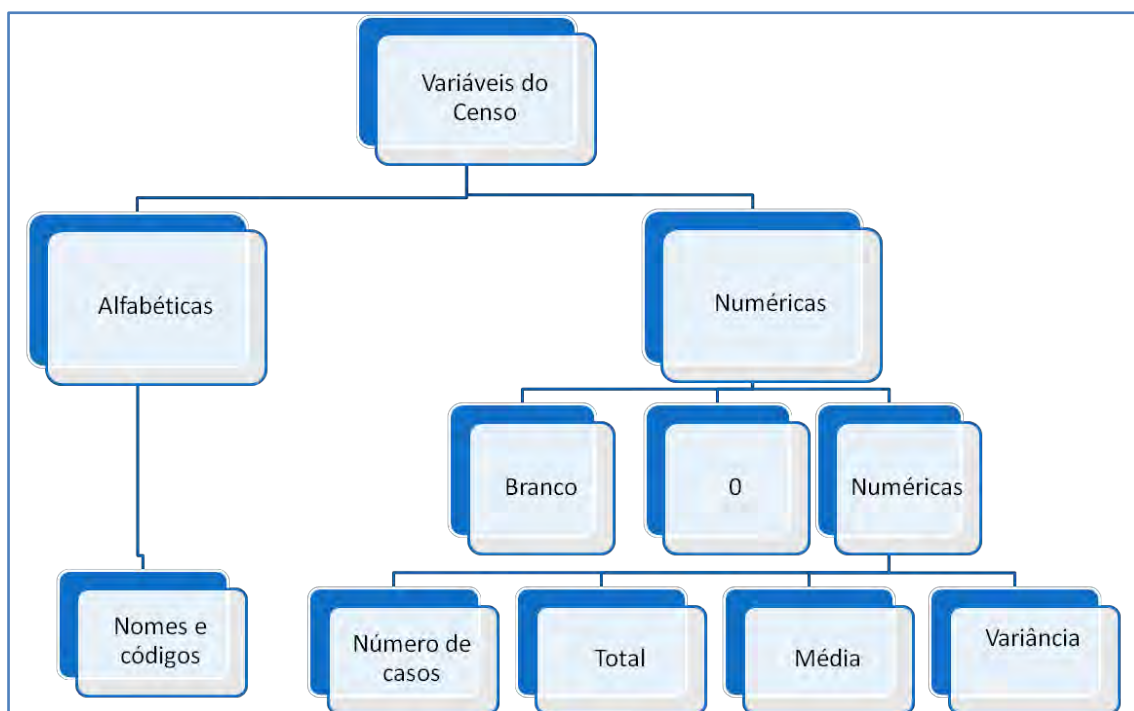


Figura 4.2 Tipologias das Variáveis do Censo 2000

Pensando-se em dados agregados por área, as variáveis do tipo soma não são afetadas pelo *MAUP*. Portanto, não é feita nenhuma inferência em nível de indivíduo a partir de dados agregados, não ocorrendo, então, a Falácia Ecológica, ou seja, se determinado arranjo espacial dos limites dos setores censitários indicam que existem 300 pessoas num setor A e, em outro possível arranjo, o setor A tiver 350 pessoas, essa diferença não implica em leitura errônea da área da cidade correspondente ao setor A.

Por outro lado, as variáveis do tipo média são completamente afetadas pelos efeitos do *MAUP*. Imagina-se que, no mesmo contexto anterior, o valor da renda média por responsáveis por domicílios do setor A passasse de R\$ 1000,00 a R\$ 1500,00 apenas com a inserção de alguns indivíduos e/ou a exclusão de outros.

Os valores apresentados neste exemplo hipotético podem parecer exagerados, porém, na maioria das cidades brasileiras existem regiões de verdadeiro contraste social, como apresentam Portugal (2009) e Cupertino *et al* (2009).

Uma forma de avaliar se um setor censitário engloba áreas contrastantes socioeconomicamente é verificar o grau de variabilidade das variáveis médias nesse setor. Assim, quanto menor é seu valor, menor é a heterogeneidade dos dados coletados. Melhor dizendo, os efeitos do *MAUP*, ou o grau de Falácia Ecológica, são menores quando se têm variáveis do tipo médias com baixos níveis de dispersão.

Esta seção deste trabalho busca avaliar o grau de dispersão apresentado nas variáveis médias relativas ao rendimento financeiro e ao nível de escolaridade dos responsáveis por domicílios nos setores censitários do Censo 2000, realizado pelo IBGE, para a área urbana do Município de Viçosa – MG.

Num primeiro momento, executou-se a leitura e análise das variâncias. Os setores com altos valores neste quesito indicam a presença de diferença socioeconômica, isto é, diferentes graus de renda e de escolaridade.

Em análises de dados censitários, um setor pode ser considerado homogêneo do ponto de vista da renda, porém, heterogêneo quando se analisa a escolaridade. De modo a verificar o grau de variabilidade, ou melhor, de heterogeneidade, utilizou-se o coeficiente de variação, pois, além de dimensões diferentes, os dois níveis (educação e renda) são medidos em unidades diferentes (anos e reais, respectivamente).

A comparação por coeficiente de variação permite, então, avaliar o quão mais heterogêneo é um setor numa condição do que na outra. Permite-se verificar qual das duas variáveis induz a uma maior análise errônea das condições de vida dos residentes num setor censitário.

Além da leitura das variáveis do censo, em muitos casos, a gestão pública utiliza indicadores socioeconômicos, como apontado na seção 3.3.1.. Abreu *et al* (2011) trabalharam com uma proposta metodológica para o cálculo do Índice de Desenvolvimento Humano (IDH) por setores censitários, utilizando somente os dados do Censo 2000, e a variabilidade desse indicador é avaliada nesta seção.

O IDH mede condições de renda, longevidade e educação e, para isso, os autores utilizaram as variáveis “valor” dos tipos *número de casos* (como número de alfabetizados e pessoas por faixa etária) e *médias* (renda média e anos médios de estudos dos responsáveis por domicílios).

Todas as variáveis foram condensadas a fim de formar uma síntese da qualidade de vida da população viçosense no ano de 2000. Uma forma de avaliar a influência da

variabilidade dessas variáveis do tipo média é aplicar a propagação das variâncias ao IDH.

As equações utilizadas por Abreu *et al* (2011) são reproduzidas abaixo.

$$IDH = \frac{IDH_R + IDH_L + IDH_E}{3} \quad (17)$$

sendo,

$$IDH_R = \frac{\log_{10} RPC - \log_{10} RMI}{\log_{10} RMA - \log_{10} RMI} \quad (18)$$

$$IDH_L = \frac{(EV - 25)}{(85 - 25)} \quad (19)$$

$$IDH_E = \frac{Tx_{AC15} + 2Tx_{7-14} + MC}{4} \quad (20)$$

onde,

IDH_R , IDH_L e IDH_E são os subindicadores de renda, longevidade e educação, respectivamente;

RPC é a renda *per capita*, obtida pela divisão da renda total dos responsáveis por domicílios pela população total residente no setor;

RMI e RMA são os valores mínimos (R\$ 3,90) e máximos (R\$ 1560,17), respectivamente, de Paridade de Poder de Compra, calculados em reais para o ano 2000 pelo Programa das Nações Unidas para o Desenvolvimento (ABREU *et al*, 2011);

EV é a expectativa de vida da população;

Tx_{AC15} e Tx_{7-14} são as taxas de alfabetização da população acima de 15 anos e entre 7 e 14 anos, respectivamente e;

MC é a média normalizada dos anos de estudos dos responsáveis por domicílios.

Para realizar a propagação das variâncias das variáveis RPC e MC , necessita-se derivar as equações acima em relação às mesmas, conforme segue abaixo.

$$\frac{\partial IDH_R}{\partial RPC} = \frac{1}{\ln 10 RPC (\log_{10} RMA - \log_{10} RMI)} \quad (21)$$

$$\frac{\partial IDH_E}{\partial MC} = \frac{1}{4} \quad (22)$$

$$\frac{\partial IDH}{\partial RPC} = \frac{1}{3 \ln 10 \, RPC (\log_{10} RMA - \log_{10} RMI)} \quad (23)$$

$$\frac{\partial IDH}{\partial MC} = \frac{1}{12} \quad (24)$$

Considerando que as unidades das variáveis são diferentes, não faz sentido prático propagar as variâncias dos subindicadores de renda e de educação, cujas unidades são R\$² e Anos², respectivamente, para o IDH. Ressalta-se que as equações utilizadas para o cálculo do IDH foram concebidas de modo que o resultado fosse adimensional, variando de 0 a 1. Por outro lado, não se pode garantir que as variâncias sejam adimensionais, como no indicador. Neste sentido, a variabilidade do IDH não é avaliada considerando as duas variáveis do tipo média.

As variâncias dos subindicadores são avaliadas conforme as equações abaixo:

$$\sigma_{IDH_R}^2 = \left[\frac{\partial IDH_R}{\partial RPC} \right]^2 \sigma_{RPC}^2 \quad (25)$$

$$\sigma_{IDH_E}^2 = \left[\frac{\partial IDH_E}{\partial MC} \right]^2 \sigma_{MC}^2 \quad (26)$$

Uma modificação foi realizada em relação ao trabalho de Abreu *et al* (2011). Não é possível obter a variância da renda *per capita*, uma vez que os dados apresentados nas planilhas do censo não permitem a recuperação desta informação. Assim, neste trabalho optou-se por avaliar as condições de renda nos setores censitários a partir do rendimento nominal mensal por pessoa responsável por domicílio – denominado, aqui, por Renda por Responsável por Domicílio (RPR) –, o qual consiste numa média que inclui os rendimentos iguais a zero. Esta informação já é presente nas planilhas do censo, contendo, inclusive, sua variância.

A escolha desta forma de avaliação se dá não apenas pelo fato de já haver a informação de variabilidade, mas também se justifica por qualquer outra fonte de renda familiar que não seja obtida pelo responsável pelo domicílio ser excluída do levantamento censitário. O IBGE considera como renda aquela obtida por trabalho próprio ou por outras fontes de renda, as quais não incluem rendas de outros moradores

das residências (IBGE, 2003). Avalia-se, portanto, apenas a renda do responsável por domicílio.

Com essa modificação realizada, alguns setores censitários passaram a conter um indicador de renda maior que 1. Isso ocorreu graças ao fato de que alguns rendimentos médios ultrapassaram o valor máximo da Paridade do Poder de Compra (PPC) calculado pelo PNUD para o ano 2000 (R\$ 1560,17), conforme apresentou Abreu *et al* (2011). Optou-se, neste trabalho, por converter os valores obtidos nos cálculos para o intervalo entre 0 e 1. Assim, os valores foram reescalados linearmente, conforme mostra a Equação 23.

$$IDH_R = \frac{\text{Valor Encontrado}}{\text{Valor Máximo}} \quad (27)$$

A diferença entre o indicador de renda encontrado em Abreu *et al* (2011) e o encontrado neste trabalho se dá nos valores absolutos, porém, com uma pequena variação relativa aos setores em sua vizinhança. Assim, os setores com os valores mais altos continuaram com os valores mais altos, ocorrendo-se o mesmo com os mais baixos.

As maiores mudanças relativas a vizinhanças ocorreram devido ao fato de se considerar a renda média apenas por responsáveis por domicílios, e não mais por pessoas residentes nos domicílios, conforme o trabalho de Abreu *et al* (2011). De qualquer forma, essa diferença não traz mudanças significativas na configuração da distribuição de renda na área urbana da cidade, segundo é apresentado nos resultados deste trabalho.

Todos os dados utilizados neste procedimento estão tabelados, conforme a organização disponibilizada pelo IBGE. As tabelas contêm um campo referente ao código do setor censitário, o qual é utilizado como chave estrangeira para a junção com os arquivos vetoriais poligonais que representam os próprios setores.

Deste modo, após a realização dos cálculos dos indicadores e da propagação de suas variâncias, todas as tabelas foram espacializadas, permitindo a edição cartográfica por mapas coropléticos. As variáveis utilizadas para os cálculos dos indicadores e suas variâncias foram Var03 e Var04 (renda média e sua variância, respectivamente) e Var10 e Var11 (média dos anos de estudos e sua variância, respectivamente).

4.3.2. Geração das zonas homogêneas

Os procedimentos realizados para a geração de zonas homogêneas são ilustrados no modelo sistêmico apresentado na Figura 4.3. Eles se baseiam nos conceitos apresentados na revisão de literatura deste trabalho. Cada passo é explicado em seção exclusiva, sendo abordados os aspectos essenciais para sua execução. Tendo em vista a importância dada à integração dos setores censitários homogêneos ao CTM, esta é apresentada em seção própria.

Observa-se que é apenas uma modelagem conceitual. Assim, todas as características de cada etapa não são colocadas como exatas, permitindo-se – até sugerindo-se – adaptações aos procedimentos que melhor se adéquem à realidade local ou às técnicas apresentadas, já que foi dada uma abordagem simplificada das mesmas.

No entanto, se seguida a linha de raciocínio, ao menos o fluxo das tarefas deve ser respeitado, pois é a proposta metodológica sugerida pelo autor.

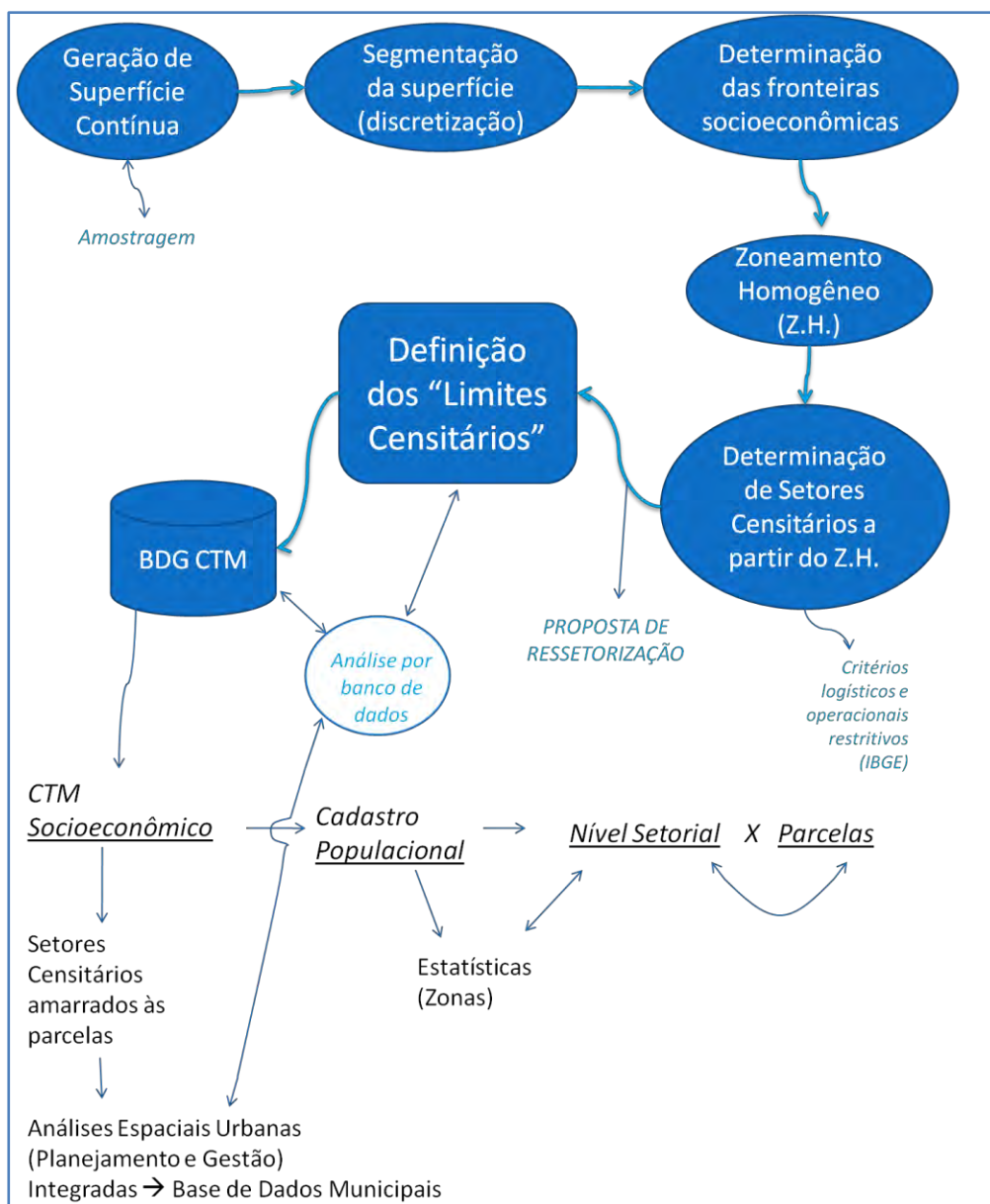


Figura 4.3. Fluxograma sistêmico da Modelagem Socioeconômica e Proposta de Ressetorização Censitária

Escolha das Variáveis

É certo que, antes dos passos apresentados, deve-se atentar para as variáveis socioeconômicas utilizadas. O contexto apresentado na seção anterior engloba aspectos de duas variáveis socioeconômicas encontradas no censo do IBGE, classificadas neste trabalho como variáveis do tipo média, quais sejam: renda e educação. Conforme

apresentado na seção 3.3.1, um indicador socioeconômico deve trabalhar com o maior número possível de variáveis, permitindo-o retratar fielmente uma localidade.

Porém, no contexto de modelagem conceitual trazido neste trabalho acadêmico, numa tentativa de simplificar o processo de apresentação do modelo proposto, é suficiente a utilização de apenas duas variáveis.

O ideal seria dispor de um arcabouço com as mesmas variáveis trabalhadas na seção 4.3.1. Neste caso, seriam necessários dados em nível individual de todas as residências – ou de uma amostra significativamente grande e bem espacializada – contendo os valores de renda e escolaridade de cada unidade.

Tal necessidade esbarra na política de privacidade dos informantes de dados estatísticos. Em IBGE (2003), apresentam-se as possíveis formas de identificação de um informante, que são evitadas no processo de divulgação dos resultados dos censos. Tal procedimento é chamado de desidentificação dos microdados censitários.

A primeira forma de identificação é direta, quando um arquivo é liberado abertamente. A segunda é a identificação por atributo, a qual ocorre quando uma informação sensível sobre o informante é liberada. A última, por sua vez, é a identificação por inferência, a qual pode ser alcançada a partir da localização da informação. A agregação geográfica é, inclusive, uma forma de proteger os informantes, uma vez que há a inserção de um mínimo de domicílios (IBGE, 2003).

Além disso, os microdados do Censo de 2000 não foram coletados com identificação geográfica individual, mas sim codificados a partir da identificação do setor censitário aos quais pertencem. Foi apenas a partir da Contagem Populacional de 2007 (IBGE, 2008b) que o GPS foi integrado ao levantamento censitário, permitindo a coleta das coordenadas georreferenciadas de cada localidade recenseada.

A limitação apresentada, associada à proteção dos dados estatísticos, torna impossível a comparação de uma nova agregação dos dados originais com os dados disponibilizados de forma agregada pelos setores censitários do Censo de 2000.

Uma alternativa buscada para a realização dos testes da metodologia proposta seria a obtenção dos microdados, ou pelo menos o acesso aos mesmos, do Censo de 2010, replicando o teste apresentado na seção anterior nos dados agregados por setores censitários. Todavia, duas situações inviabilizaram este ensaio.

A primeira é a forma com que o IBGE possibilita o acesso ao tipo de informação. O Centro de Disseminação de Dados e Informações (CDDI), ancorado à

política de privacidade dos informantes de dados estatísticos do próprio IBGE, possui um conjunto de regras de acesso por parte de pesquisadores externos ao órgão.

O acesso aos microdados não desidentificados só é permitido com: submissão, ao CDDI, do projeto pelo pesquisador; julgamento pelo Comitê de Avaliação de Acesso a dados não desidentificados; em caso de autorização, o acordo com o usuário, que envolve os termos de compromisso, condições de acesso e de pagamento; a preparação do produto na sala de acesso a dados restritos; e a liberação do produto, após a avaliação do atendimento à preservação do sigilo das informações individualizadas.

Já a segunda limitação é quanto ao fato de que esse tipo de atendimento, com os dados do universo do Censo 2010, somente seria realizado após a liberação dos resultados do universo pelo IBGE, que estava marcado para setembro de 2011. Observa-se que, findado o mês de setembro, tais dados ainda não foram disponibilizados ao público. A espera pela liberação vai de encontro com a data limite para a finalização desta pesquisa.

Devido ao exposto, corroborado com o fato de que se foca, aqui neste trabalho, numa modelagem conceitual, é dada uma aplicação da proposta a variáveis genéricas, sendo estas, porém, reais, o que confere veracidade aos dados envolvidos na simulação, evitando qualquer possibilidade tendenciosa na síntese dos mesmos.

A agregação de dados a partir de áreas homogêneas pode ser aplicada a qualquer situação, pois o conceito de homogeneidade intra-áreas é o mesmo, independente da aplicação dada. É óbvio, entretanto, que, em aplicações socioeconômicas urbanas, alguns cuidados devem ser observados, como a descontinuidade espacial, causada por vias de circulação, zoneamento urbano ou acidentes geográficos (RAMOS, 2002).

Na busca por dados confiáveis para o teste metodológico, objetivando também a minimização de custos, refletiu-se por utilizar os dados apresentados no livro *An Introduction to Applied Geostatistics*, de Isaaks e Srivastava (1989), denominados genericamente de variáveis U e V, as quais se baseiam em dados de elevação da área do Walker Lake, situado em Nevada, no oeste dos Estados Unidos da América. Foi dado um aspecto geral ao conjunto amostral, não sendo tocadas questões quanto à escala e às coordenadas da configuração espacial, focando-se apenas na aplicação da metodologia sugerida.

A dupla de conjunto de valores apresenta certa semelhança com a realidade das duas tipologias de variáveis estudadas na seção anterior – renda e escolaridade. O

esclarecimento de tal afirmação é dado sob dois aspectos, os quais são indissolúveis: distribuição estatística e configuração espacial das amostras.

Quanto ao primeiro ponto, existe certa variabilidade espacial nos valores de U e V, que se assemelham a uma distribuição de valores de renda e escolaridade. Os dados não se apresentam tendenciosos, com concentrações de valores semelhantes em apenas uma parte da área.

No plano da configuração espacial dos dados, nota-se na Figura 4.4, a distribuição irregularmente espaçada, no geral, e com pequenas concentrações de pontos em alguns locais, apresentando certa dispersão espacial em outras localidades. Tal desenho se assemelha à representação de áreas residenciais urbanas (maior concentração) e rurais (maior dispersão), respectivamente.

Destaca-se, ainda, que alguns pontos possuem valor nulo para a variável U. Isaaks e Srivastava (1989) omitiram esses dados propositalmente com fins didáticos em seu livro. Para o prosseguimento desta pesquisa, assumiu-se que não havia pontos nessas localidades, mesmo que haja valores para a variável V.

Adotou-se, nesta simulação, que este conjunto de pontos representa as residências de uma cidade fictícia e que os valores das variáveis genéricas V e U correspondam aos valores de duas variáveis socioeconômicas levantadas por um Censo Demográfico nesta cidade. Os pontos com valores de U omitidos por Isaaks e Srivastava (1989) e excluídos nesta simulação localizam-se no que poderia ser a zona rural da imaginária área de estudos – a qual é excluída dos estudos apresentados, uma vez que se dá um foco na modelagem conceitual nas áreas urbanizadas das cidades.

Os autores dos dados dividiram a área de estudos (considerando os valores de U) em 7 agrupamentos, indicados pelos pontos escuros. Observa-se que um destes grupos é bem maior, tendo grande extensão no sentido norte-sul, englobando pequena área a oeste. Nessa simulação, considerou-se esta área como sendo o distrito sede da imaginária cidade, agregando uma área urbana um pouco afastada. As demais zonas foram consideradas núcleos urbanos dos demais distritos. Lembrando-se, obviamente, que a quantidade de “residências” aqui simuladas é bem menor que a quantidade normalmente existente na maioria das cidades brasileiras.

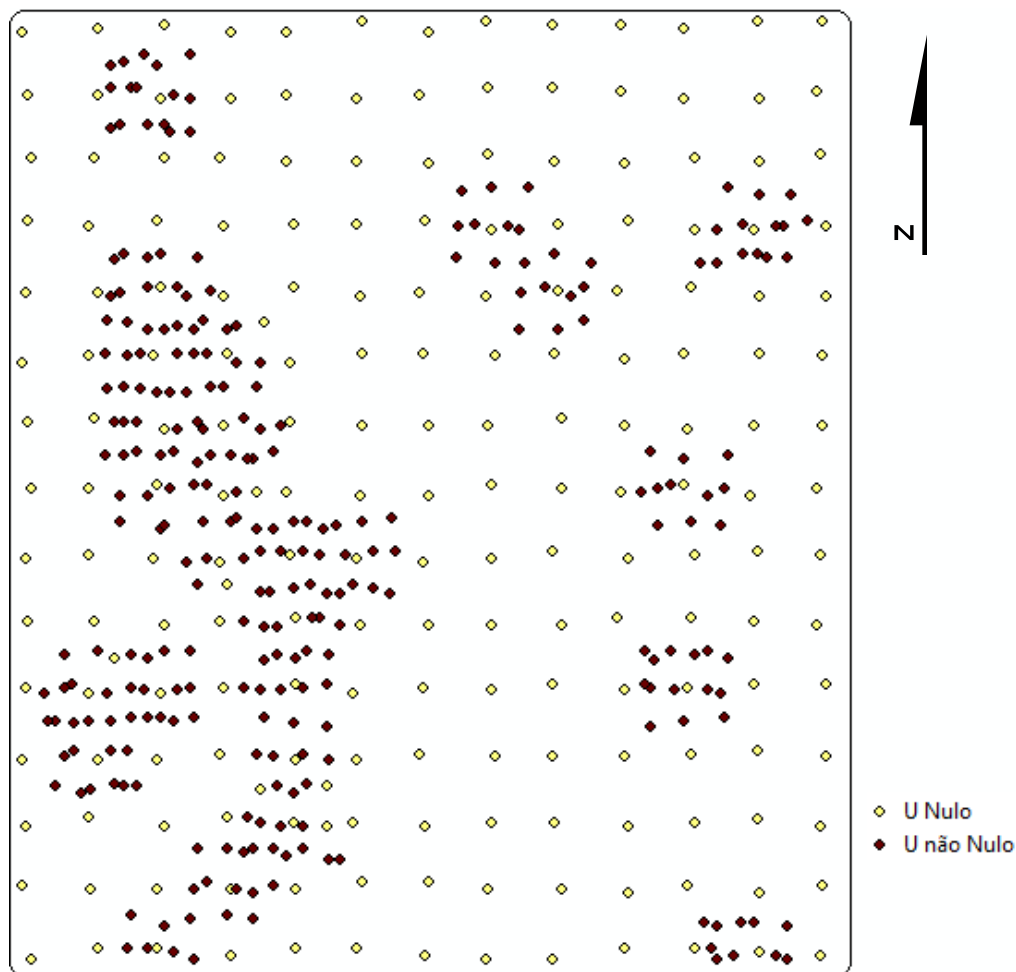


Figura 4.4. Distribuição espacial dos dados utilizados na simulação

Fonte: ISAAKS e SRIVASTAVA (1989).

4.3.2.1. Geração das Superfícies Contínuas

Focando-se na modelagem conceitual, após escolhidos as variáveis ou os indicadores socioeconômicos que melhor representem as condições socioeconômicas da população local, passa-se para a modelagem das superfícies contínuas por geoestatística, uma para cada variável. Ressalta-se que a modelagem da dependência espacial das amostras é independente para U e para V.

Na verdade, a aplicação da krigagem neste trabalho não tem intenção de estimar valores das variáveis em pontos não amostrados, como na concepção do interpolador. Propõe-se sua aplicação apenas para indicar as variações locais dos valores das variáveis, como forma de modelar a continuidade espacial. Daí a importância deste

interpolador, uma vez que a obtenção dos pesos se baseia no conceito de dependência espacial.

Outra questão é quanto à utilização do conceito associado à palavra amostragem. O que se sugere aqui não necessariamente é a obtenção de amostras das variáveis, pois o IBGE coleta dados em todos os domicílios das cidades brasileiras. Pensando assim, corroborado pelo exposto no parágrafo anterior, não faz sentido dizer que os domicílios são amostras coletadas para a estimação de variáveis em locais onde há vazios urbanos. Todavia, a separação de alguns locais para uma validação do modelo é apropriada.

A modelagem da continuidade ou dependência espacial, na conceituação deste trabalho, serve apenas para identificar padrões espaciais das variáveis estudadas. Ainda, o conjunto de pontos utilizados tem uma distribuição irregularmente espaçada, tal como apresentada na Figura 4.4. Lembra-se que a Krigagem não é influenciada por tendências apresentadas no conjunto de pontos. Assim, uma concentração de dados causada por aglomerados residenciais não prejudicaria a análise, fato que traz nova justificativa para sua utilização neste processo.

Para a simulação proporcionada por este trabalho, utilizaram-se as variáveis U e V, conforme citado anteriormente. Com a exclusão dos locais onde a variável U era nula, o total de pontos caiu de 470 para 275. Nota-se que a amostragem é irregularmente espaçada, apresentando conformidade semelhante a um desenho urbano.

Dando início à modelagem, num primeiro momento, fez-se análise exploratória das duas variáveis. Buscou-se conhecer seus parâmetros, tais como médias, variâncias, coeficientes de assimetria e de curtose, valores mínimos, máximos e quartis. Essa tarefa é importante para se conhecer o conjunto de dados, o que auxilia o processo de análise da semivariância.

Em seguida, partiu-se para a análise da continuidade espacial das duas variáveis a partir dos semivariogramas, buscando-se também obter seus parâmetros. É sempre necessário verificar se existe anisotropia, analisando, na superfície do semivariograma, se há diferentes padrões a partir de diferentes direções. Investigaram-se, portanto, as direções e tolerância angulares e as larguras de bandas para as duas variáveis.

Além disso, outro ponto importante é a validação cruzada entre os valores medidos e estimados pela interpolação. Chama-se esse passo de predição dos erros, os quais devem ter média igual a zero. Na prática isso nem sempre ocorre, devendo o valor médio desse erro ser o mínimo possível.

Foram geradas, portanto, duas superfícies que modelam a continuidade e distribuição espacial dos valores de U e V, respectivamente. É ressaltado que esse procedimento foi realizado no pacote de *softwares ArcGIS 9.3.1*, advertindo-se, no entanto, que qualquer outro que realize modelagem geoestatística poderia ser utilizado.

4.3.2.2. Discretização da Superfície Contínua

Neste momento da modelagem conceitual, parte-se para a obtenção das áreas de valores semelhantes para as superfícies geradas. Em processamento digital de imagens, costumam-se empregar como ferramentas de detecção de padrões espaciais a segmentação ou a classificação automática. A partir dos valores de cada pixel na imagem, os algoritmos reconhecem aqueles que possuem semelhanças e adjacências, formando agrupamentos ou *clusters*.

Trazendo a ideia para a realidade estudada aqui, detectar padrões espaciais consiste em encontrar as fronteiras de áreas com pixels semelhantes. Melhor dizendo, seria encontrar regiões que apresentam valores das variáveis ou dos indicadores socioeconômicos semelhantes.

Ao obter essas fronteiras, encontram-se os locais onde o padrão de continuidade espacial muda, trazendo a sensação de haver a oposição à Primeira Lei da Geografia. Assim, constata-se que nesses lugares há uma mudança nas condições socioeconômicas, acarretando em heterogeneidades locais. Por isso, tais fronteiras estabelecem os limites de homogeneidade interna às áreas.

No entanto, neste momento, busca-se apenas encontrar tais regiões, não importando quais valores ou padrões socioeconômicos (altos ou baixos) estão sendo detectados, importando apenas encontrar onde há diferenças. A utilização do algoritmo *Isodata* é ideal nesse caso, devido ao fato de que, a partir de critérios estatísticos, as áreas de mesmos valores e parâmetros estatísticos são encontradas.

Assim, foram gerados, neste trabalho, três procedimentos de detecção de *clusters*, sendo um para a variável U, outro para V e o último para uma composição das duas superfícies. Tal escolha se deu pelo fato de que algumas áreas podem ter padrões de homogeneidade de valores de pixels semelhantes. Então, resolveu-se testar as possíveis diferenças entre a junção das fronteiras (de U e de V) e as fronteiras da composição VU.

Como os valores de U e V possuem grande amplitude, as duas superfícies foram reescaladas para 255 níveis de cinza, permitindo a aplicação do algoritmo escolhido. Por ser um método iterativo, foram realizados vários testes, com mudanças nas iterações, nos graus de similaridade, nos tamanhos mínimos das áreas e no número de *clusters* desejados.

Como tal detecção não está baseada no número de classes temáticas presentes numa classificação de imagens, o qual auxilia na escolha da quantidade de *clusters* no processo, o número de agrupamentos escolhidos foi igual ao máximo possível detectado pelo algoritmo implementado no *software IDRISI Andes*, o qual foi escolhido neste passo.

4.3.2.3. Determinação das Fronteiras Socioeconômicas

Como o objetivo da detecção de padrões era o de apenas obter as fronteiras entre as áreas homogêneas, os polígonos que representam as classes ou agrupamentos são convertidos para linhas. Deste modo, há um conjunto de linhas para cada variável estudada na detecção de homogeneidade.

Para o caso apresentado, foram encontrados três conjuntos de feições lineares, como exposto na seção anterior. O que se pretendeu mostrar é que as fronteiras obtidas pela composição VU são ligeiramente diferentes daquelas obtidas pela junção ou sobreposição das fronteiras de U e de V. Buscou-se testar, portanto, se a diferença verificada exerce alguma influência nos resultados da agregação.

4.3.2.4. Zoneamento Homogêneo

Por conseguinte, as linhas foram recortadas segundo o limite considerado para cada distrito do município fictício. Além disso, foram ajustadas todas as bordas formadas pelas feições lineares e subtraídas todas as linhas geradas a partir de polígonos indesejáveis causados pela presença de pixels isolados na superfície interpolada. A Figura 4.5 ilustra tais polígonos.

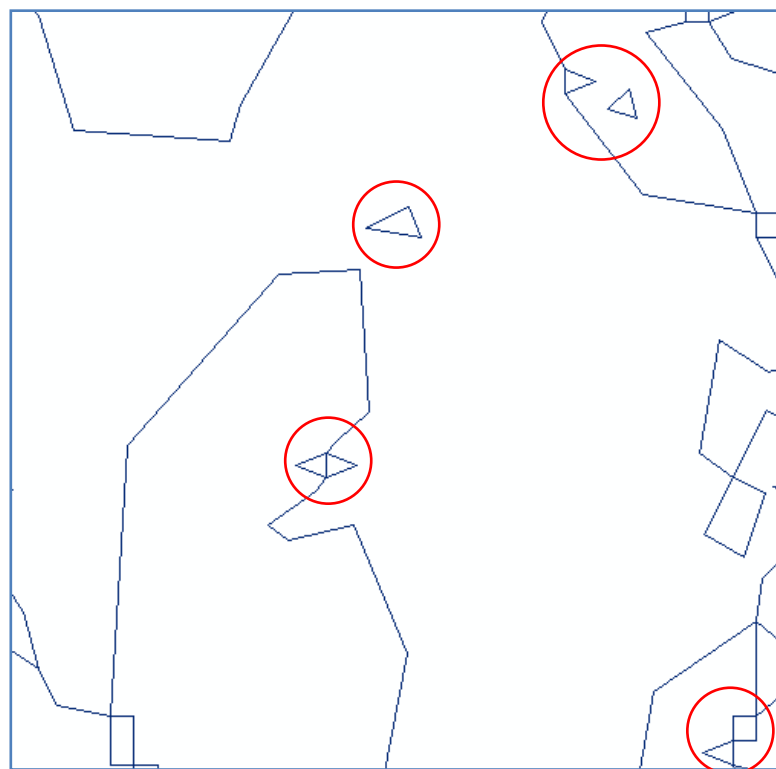


Figura 4.5 Exemplo de fronteiras indesejáveis

Este tipo de procedimento foi obtido com simples edições de modelos de dados vetoriais lineares, que podem ser realizadas em qualquer SIG. É certo, contudo, que tais edições devem considerar as propriedades topológicas das fronteiras encontradas.

Para a metodologia realizada neste experimento, utilizou-se, novamente, o *Software ArcGIS 9.3.1*. Para isso, porém, necessitou-se converter os *clusters* encontrados de polígonos matriciais para linhas vetoriais no *IDRISI*, havendo a posterior conversão de formatos de dados para se trabalhar no software de edição vetorial.

As linhas indicando fronteiras de homogeneidade, anteriormente editadas, sugerem os limites das áreas homogêneas. Algumas dessas áreas não possuíam pontos coletados. Isso se deve ao fato de as fronteiras terem sido definidas a partir de duas superfícies, o que não significa dizer que havia pontos em toda a superfície. Tais zonas foram definidas como *Setores Incompletos*, sendo, posteriormente, excluídos da montagem da base de dados.

Todas as áreas encontradas foram consideradas *Zonas Homogêneas*. Na proposta conceitual, os tamanhos dessas zonas podem variar consideravelmente em diferentes regiões do município em estudo. A variação dessas dimensões pode ser bastante ampla,

envolvendo mais residências que o permitido pelo IBGE na criação dos setores censitários. Por outro lado, a quantidade mínima requerida pelo órgão pode também não ser atingida.

Assim, os analistas devem usar alguns critérios para o desmembramento ou para a fusão de algumas zonas homogêneas na criação dos limites censitários, apresentados na próxima seção.

4.3.2.5. Determinação dos Setores Censitários a partir do Zoneamento Homogêneo

No modelo, para a criação de setores censitários a partir das condições apresentadas anteriormente, propõe-se a análise dos valores das variáveis representados em cada residência localizada nas áreas modificáveis. Caso haja desmembramento, o critério é criar uma linha fronteira que separe os conjuntos de pontos/residências mais heterogêneos entre si, agrupando os mais homogêneos. É importante não separar as residências localizadas no mesmo logradouro, exceto quando este for muito extenso, fato discutido na Seção 4.3.4.

Por outro lado, caso haja fusão de zonas, deve-se escolher aquelas que agregam pontos/residências mais homogêneas entre si. É certo, porém, que isso nem sempre vai ocorrer, uma vez que não existe total homogeneidade num logradouro. Mesmo assim, é válida a proposta do parágrafo anterior de se deixar num mesmo setor os dados do mesmo logradouro.

A justificativa para esse critério é fundamentada em algumas definições abordadas na seção 3.4.2, replicadas neste espaço. A primeira diz que as bordas das zonas geográficas devem ser definidas de modo que produzam dados que se conformam com o que pode representar um modelo, minimizando o erro desse modelo. O sistema de zoneamento age, portanto, como um detector de padrões espaciais (OPENSHAW, 1996).

Complementando, quando se busca a homogeneidade de uma variável socioeconômica, por exemplo, deve-se maximizar a correlação interna às áreas, tendo em contrapartida a minimização da correlação entre as áreas (RAMOS, 2002) e (MARTIN, 2000).

A Figura 4.6 ilustra as definições acima. São dados quatro pontos, cada um com o valor indicado. No caso de os dois pontos da esquerda serem agregados, tem-se um

polígono com valor médio igual a 5 e desvio igual a 4. O outro polígono também possuiria média igual a 5, contendo desvio igual a 3.

Noutra situação, considerando a proximidade dos valores representados, o polígono de cima tem média igual a 1,5 e desvio 0,5, seguido pelo polígono abaixo, com média igual a 8,5 e desvio também igual a 0,5.

Nota-se que, no primeiro caso, há maior homogeneidade entre as áreas e maior heterogeneidade interna às áreas se comparado à segunda situação, a qual possui resultados opostos (baixa correlação entre as áreas e alta correlação interna às áreas).

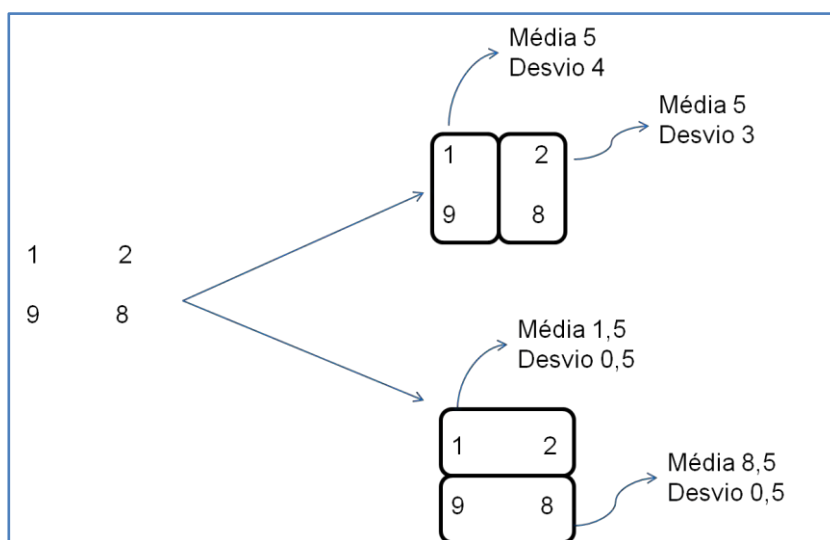


Figura 4.6 Exemplo de agregação homogênea

Na contramão desses conceitos, para o estabelecimento dos limites censitários, o IBGE estipula apenas os limites mínimos e máximos de residências em cada setor, além da facilidade de acesso dos recenseadores aos domicílios pesquisados, que deve ser garantida. No modelo aqui proposto, as zonas homogêneas com menos que o limite inferior em quantidade de pontos devem ser conectadas a zonas vizinhas de mesmas condições, ignorando as fronteiras socioeconômicas criadas na segmentação.

Neste trabalho, considerou-se um mínimo de 3 pontos em um setor, evitando a existência de setores com poucos pontos. Levando-se em conta o traçado das fronteiras socioeconômicas, a criação dos setores censitários foi realizada apenas com base nos critérios modelados anteriormente, ignorando os critérios logísticos restritivos do IBGE, uma vez que os dados utilizados não possuíam logradouros indicando possíveis rotas de acesso aos locais coletados.

Para facilitar a identificação de pontos de valores próximos, utilizou-se a reclassificação da legenda e das cores de seus valores em 10 quantis. Deste modo, se dois pontos de valores (e, conseqüentemente, cores) próximos, presentes em setores que contenham menos de três pontos, estão separados por um limite, esse limite é excluído, acoplando um setor ao outro. A Figura 4.7 ilustra o procedimento acima descrito.

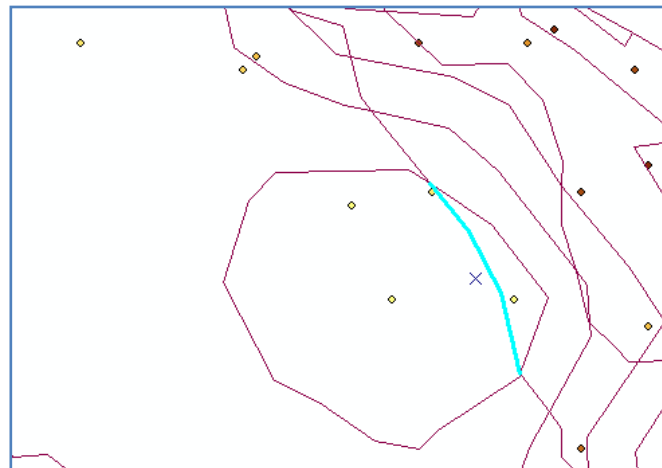


Figura 4.7 Exemplo de linha a ser excluída para a agregação de zonas homogêneas

Em alguns casos, foi necessário verificar a distribuição nas duas variáveis, dado que um setor com apenas um ponto possuía alto valor em uma delas. Analisando-se o valor na outra variável, foi possível perceber qual linha de divisa deveria ser subtraída. A Figura 4.8 exemplifica tal fato.

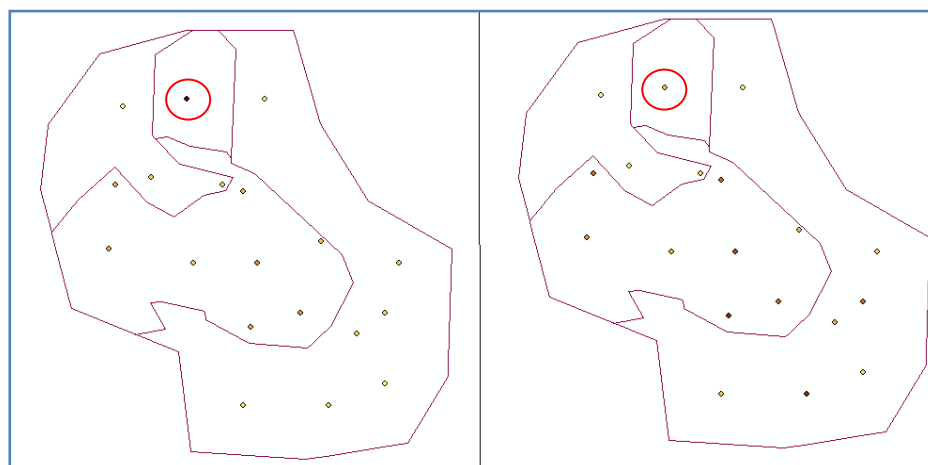


Figura 4.8 Exemplo de análise simultânea das duas variáveis

Com tais operações, foram definidos os *Limites Censitários*. Numa modelagem real, destaca-se que tais limites deveriam considerar também as quantidades mínimas e máximas permitidas de residências inseridas e facilidade de acesso por parte do recenseador.

Em verdade, a única diferença entre o que o IBGE propõe e o trazido nesta dissertação é a *incorporação da homogeneidade socioeconômica*. Por fim, definidos os novos limites censitários, os dados coletados são agregados e suas estatísticas são calculadas, sem alteração nenhuma no procedimento realizado pelo IBGE.

Adicionalmente, como forma de obter referências para avaliar a metodologia proposta, foram criados sete diferentes arranjos poligonais. Os limites dos três primeiros foram estabelecidos de forma aleatória do ponto de vista homogêneo. Consideraram-se apenas o número mínimo de três pontos para cada setor. Isso foi uma tentativa de replicar o procedimento do IBGE na definição dos limites, pensando-se apenas nos fatores logísticos e operacionais. Estes arranjos foram chamados de A1, A2 e A3, sendo a letra A uma abreviação para *Aleatório*.

Em seguida, foram criados dois novos arranjos, os quais foram baseados nos limites homogêneos definidos. Entretanto, consideram-nos sem rigor, pois os processos de desmembramento/fusão não levaram em consideração os critérios de homogeneidade de valores individuais propostos.

Além disso, um dos arranjos foi baseado nas fronteiras advindas da segmentação das variáveis V e U sobrepostas, chamado de HS1 (Homogêneo Sobreposto 1), enquanto o outro se fundamentou nos limites criados pela composição VU, sendo denominado de HC1 (Homogêneo Composto 1).

Por fim, criaram-se os dois últimos conjuntos de zonas utilizando maior rigor, buscando-se a homogeneidade dos valores representados nos pontos a serem agregados. Analogamente, como explanado acima, os dois modelos de zoneamento foram batizados de HS2 e HC2, respectivamente.

Todos os dados originais foram agregados aos setores criados, calculando-se os valores totais, médios, mínimos e máximos, além das variâncias de U e V e o número de pontos agregados em cada polígono. Essas operações foram realizadas pela função de junção espacial, presente no *Software ArcGIS 9.3.1*, ressaltando-se que tal função também se encontra presente em qualquer outro SIG.

4.3.3. Avaliação da Homogeneidade Interna dos Setores Propostos

Com os sete arranjos de setores criados, foram estabelecidos alguns critérios avaliadores para o padrão de homogeneidade intra-área (maximização da correlação interna) e de heterogeneidade interáreas (minimização da correlação entre as áreas).

Foram utilizados dois critérios de avaliação do primeiro padrão citado anteriormente: análise visual dos valores agregados de V e U, acompanhados de seus respectivos desvios-padrão, os quais deveriam ser baixos, e; avaliação da autocorrelação dos valores de desvio-padrão, na tentativa de encontrar agrupamentos poligonais com baixos desvios.

Para a primeira avaliação intra-área, as classes de valores, em cada uma das quatro variáveis de análise (V, U, DP_V e DP_U), foram padronizadas, permitindo assim a comparação entre os arranjos. A escolha se baseou na divisão em 5 quantis de todos os arranjos. Em seguida, verificou-se qual arranjo possuía o menor valor de classe, sendo a escala de valores deste alterada com a inclusão do maior valor de classes presente nos demais arranjos. Deste modo, todos os valores possíveis de ocorrência nos polígonos foram contemplados na classificação da legenda.

Em acréscimo, a análise de autocorrelação dos desvios-padrão também foi dividida em duas partes. Na primeira, utilizou-se o Índice Local de Moran, aplicado aos valores de desvio-padrão das duas variáveis estudadas. Nesta análise, buscou-se encontrar quais polígonos possuíam relação de dependência espacial com seus vizinhos, sendo mais adequado o arranjo que possuísse maior quantidade de agrupamentos locais de baixo desvio-padrão.

No entanto, a utilização desta avaliação pode não trazer grandes informações caso os níveis de agrupamento não sejam significativos estatisticamente. Para avaliar estes casos, utilizou-se a *General G Statistics* para se conhecer se havia ou não autocorrelação de padrões baixos. Em outras palavras, caso localmente não houvesse significância de agrupamentos de baixos valores, globalmente poderia ser notada alguma tendência para isso.

Já a minimização da correlação entre as áreas foi avaliada a partir das análises de autocorrelação global de Moran, aplicados aos valores de V e de U agregados. Assim, caso não houvesse algum padrão de dependência espacial num dado arranjo, então a correlação interáreas, para as duas variáveis, estaria minimizada.

Nesses procedimentos, foram utilizados os *softwares ArcGIS 9.3.1*, para os cálculos dos indicadores globais e *TerraView 4.0.0* para o indicador local. Este último foi escolhido pelo fato de que o primeiro apenas apresenta os resultados do Índice Local de Moran quando há significância estatística, o que não ocorre no segundo.

4.3.4. Proposta de integração das regiões socioeconômicas e Setores Censitários ao Cadastro Territorial Multifinalitário

Como visto na seção 3.1.2, o cadastro socioeconômico surge como importante fonte de informação para a tomada de decisões. Porém, nem todas as informações interessantes à caracterização socioeconômica da população estão presentes nas secretarias e empresas públicas e privadas.

Pode-se usar como exemplo o nível de escolaridade de chefes de família, uma vez que os órgãos responsáveis pela informação sobre educação a detêm apenas para os alunos matriculados, excluindo do seu cadastro as pessoas que não estudam mais, seja por opção ou por condição. A distribuição de renda também é outro aspecto que dificilmente se encontra em bancos de dados de prefeituras, principalmente as de cidades de pequeno porte.

Outro exemplo a ser dado é a quantidade e a localização de casas do meio urbano que possuem água proveniente de poços, além daquela disponibilizada pela entidade responsável pelo abastecimento de água tratada. Dificilmente, uma concessionária de abastecimento de água tem este tipo de informação em seu cadastro de clientes, uma vez que o que lhe interessa é saber o quanto é consumido por seus clientes e o quanto será cobrado em cada registro cadastrado em seu banco de dados, além do que, obviamente, este tipo de fonte de abastecimento nem sempre é legalizado.

Essas informações estão presentes nos dados das pesquisas censitárias, sendo os mais completos e confiáveis no que diz respeito à dimensão socioeconômica. Ramos (2002) discorre sobre a natureza de dados socioeconômicos e a relaciona com os conceitos de territórios intraurbanos apresentados na literatura do Urbanismo. O autor ainda diz que, “como regra geral, os dados socioeconômicos são disponibilizados como características estatísticas de um determinado grupo de indivíduos habitantes de uma parcela ou zona do território”, que podem ser representados por bairros, distritos ou setores.

O autor supracitado ainda mostra que a escala de análise de fenômenos intraurbanos está ligada ao nível de desagregação de dados espaciais utilizados na análise. Assim, quanto mais desagregados forem os dados, melhor será a observação de padrões em níveis mais locais.

Conclui-se, portanto, que os setores censitários tornam-se fundamentais para a análise de fenômenos intraurbanos, contribuindo para os processos decisórios por parte dos planejadores urbanos e gestores municipais, uma vez que constituem a menor unidade de análise espacial de dados censitários.

Neste sentido, pondera-se sobre a integração dos setores censitários ao CTM, de maneira sistemática, uma vez que ambos proveem dados e informações municipais necessários para o desenvolvimento municipal, permitindo a caracterização socioterritorial, apresentando, no entanto, desarticulações espaciais entre si em algumas localidades, como é apresentado posteriormente.

Destaca-se que a discussão aqui apresentada se iniciou no trabalho intitulado *Proposta de Integração dos Setores Censitários ao CTM auxiliando o Cadastro Socioeconômico: Vantagens e Desvantagens*, de Abreu *et al* (2010b), publicado nos anais do IX Congresso de Cadastro Técnico Multifinalitário e Gestão Territorial, ocorrido em Florianópolis, Santa Catarina, em outubro de 2010.

A proposição foi realizada a partir de análises documentais presentes na literatura nas áreas de CTM e Análises Espaciais Urbanas, discussões com outros profissionais e estudantes da área de Cadastro Territorial, visitas à sede do IBGE em Viçosa – MG, além de concepções e experiências práticas do autor deste trabalho.

Portanto, a partir das vantagens e desvantagens da utilização dos setores censitários para a caracterização socioeconômica urbana, apresentadas nas seções 3.2.2.1 e 3.2.2.2, respectivamente, as quais abalizam essa integração, focou-se em proposições que eliminem ou minimizem as desvantagens, baseando-se principalmente nos resultados apresentados nos testes realizados nesta dissertação.

Como resultado, é proposta uma nova ontologia que defina os setores censitários sob a ótica do CTM. Em seguida, sugere-se como se dá a integração em um modelo de banco de dados. Ressalta-se que em nenhum momento é realizado um estudo mais aprofundado sobre a modelagem de BDG, muito menos tal modelo se resume à ideia que aqui se apresenta, a qual busca uma abordagem simples para o entendimento do conceito alvitrado.

Por fim, são dados exemplos simulados da integração dos setores censitários ao CTM, sendo realizada uma pequena simulação em algumas localidades do Município de Viçosa – MG.

5. RESULTADOS E DISCUSSÕES

A partir da metodologia proposta, apresentam-se aqui os resultados encontrados, buscando-se associar a uma análise crítica das soluções e simulações encontradas.

5.1. Avaliação da Homogeneidade Socioeconômica dos Setores Censitários

Tendo-se realizada a espacialização das tabelas que contêm os cálculos dos indicadores socioeconômicos e suas propagações de variâncias, a partir da análise visual de mapas temáticos, pode-se concluir sobre o grau de heterogeneidade dos setores censitários de Viçosa-MG, no ano de 2000. Ressalta-se, no entanto, que são apresentadas apenas as figuras representando os setores censitários, sem as preocupações cartográficas.

Observa-se, inicialmente, que a configuração espacial dos dados censitários apresenta valores agregados por áreas, os quais sofrem os efeitos do *MAUP* (Problema de Unidades de Áreas Modificáveis). Possivelmente, se os setores censitários tivessem outra configuração espacial, então uma ou outra região apresentaria valor bem distinto do apresentado aqui.

A ideia apresentada é a de que o Efeito de Zoneamento exerce maior influência que o Efeito de Escala, uma vez que todos os setores contemplam tamanhos populacionais semelhantes, não trazendo grandes variações em escala. Melhor dizendo, um novo zoneamento censitário deveria conter setores que possuísem o mesmo tamanho, não agregando maior população que no arranjo original. O ideal é que esses setores tivessem homogeneidade interna, ou seja, contemplassem variáveis do tipo médias com as menores variâncias possíveis.

Nota-se, na Figura 5.1, maiores valores de renda por responsáveis por domicílios nos setores censitários da região central da cidade, compreendendo, principalmente, os Bairros de Ramos, Clélia Bernardes, Centro e a parte norte do Bairro de Fátima. Recebe destaque o setor dos Bairros Acamari e Romão dos Reis, devido ao fato de o primeiro

consistir num condomínio fechado, cujos moradores são, em sua maioria, professores da Universidade Federal de Viçosa, e o segundo um bairro de classe baixa.

Ainda no que diz respeito à distribuição da renda, destacam-se como setores com valores realçados positivamente os correspondentes aos Bairros de Lourdes e Santa Clara, além dos Bairros João Braz, Recanto da Serra (o qual também é um condomínio fechado de classe alta) e Violeira.

Com valores extremamente baixos, destacam-se os setores das regiões de Vau-Açu, Nova Viçosa, Estrelas, norte do Bairro Santo Antônio (conhecido como Morro da Coruja) e Bairro Sagrado Coração de Jesus (popularmente chamado de Rebenta Rabicho).

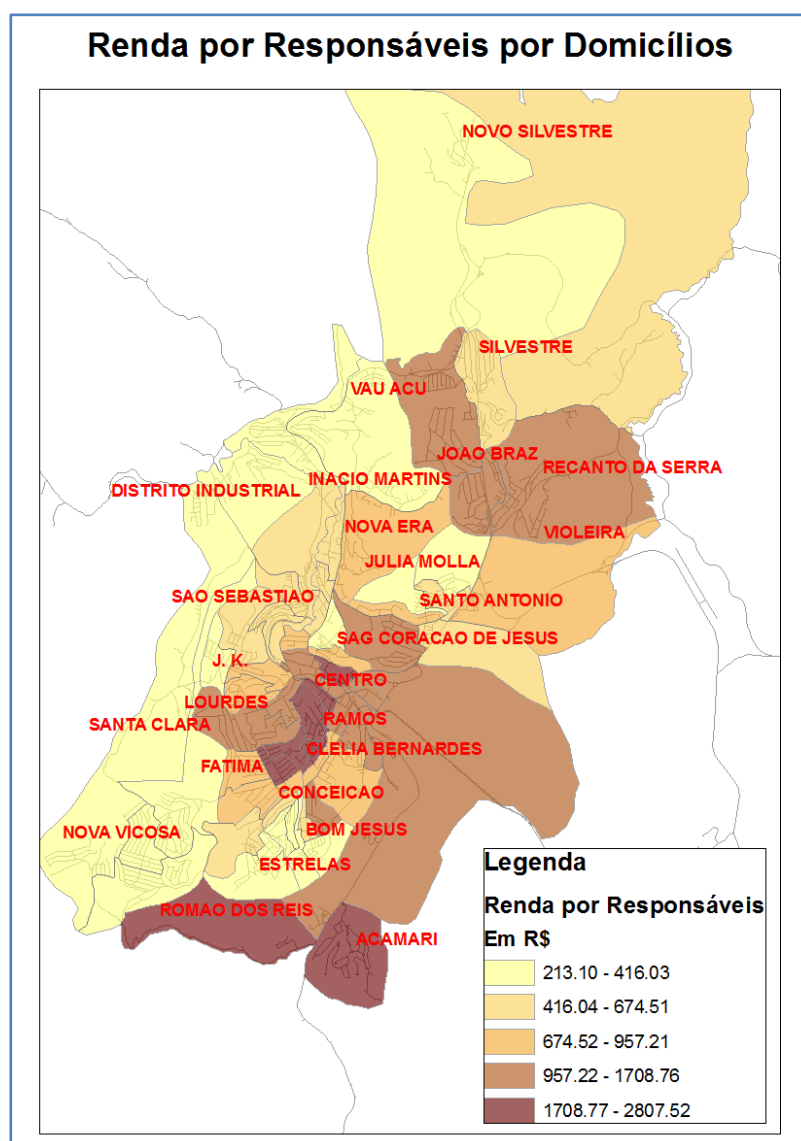


Figura 5.1 Distribuição da Renda por Responsáveis por Domicílios

Analisando a distribuição espacial dos desvios-padrão da renda, na Figura 5.2 destacam-se, com os maiores valores, as mesmas áreas citadas anteriormente para os maiores valores de renda, com exceção do Bairro Clélia Bernardes. Neste grupo de grande desvio-padrão se insere o setor censitário que contém os Bairros Recanto da Serra e Violeira. Isso permite inferir sobre a variabilidade dos dados presentes nestes setores, trazendo a ideia de grande heterogeneidade interna dos mesmos.

Nota-se uma grande quantidade de setores censitários com valores consideravelmente altos, apresentando certo grau de heterogeneidade, tais sejam aqueles dos Bairros João Braz, Lourdes, Santa Clara, Fátima, Conceição e parte do Santo Antônio.

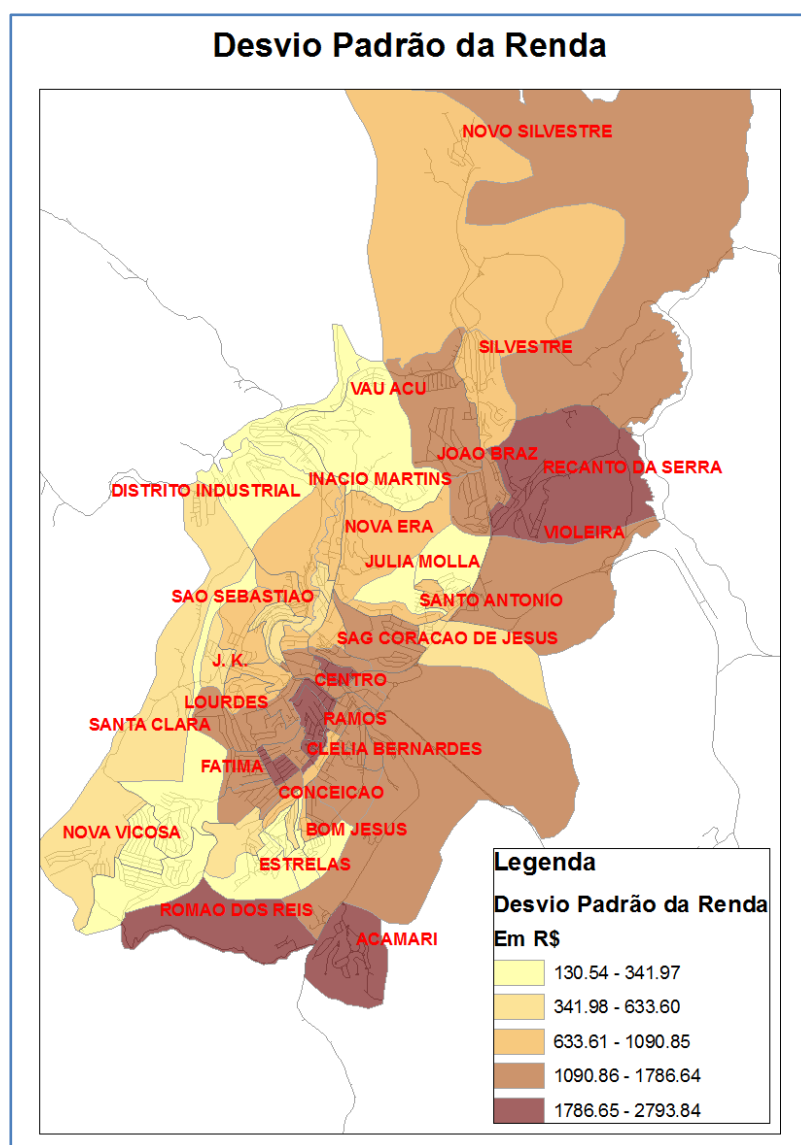


Figura 5.2 Distribuição do Desvio-Padrão da Renda por Responsáveis por Domicílios

Enfatiza-se, por outro lado, os setores com menor grau de variabilidade. Percebe-se que nenhum setor com baixo padrão de diversidade de renda possui alto valor de renda. Portanto, os setores com baixos valores de renda tendem a ter menor variância, sendo mais homogêneos, destacando-se Nova Viçosa, Estrelas e norte do Santo Antônio (Morro da Coruja).

Como os setores de grande concentração de renda também possuíram grandes dispersões, infere-se que a distribuição é desigual internamente ao setor, sendo bem influenciada pelos valores de renda da classe mais alta. Este fato parece bem óbvio quando se lembra que valores médios sofrem interferências dos extremos num conjunto de dados.

Deve-se ter cuidado, portanto, ao afirmar que a Região Sul de Viçosa, composta pelos Bairros Acamari, Monte Verde, Vila Alves, Romão dos Reis e Rua Nova, é de classe alta. A alta variabilidade associada ao alto valor de Renda por Responsável por Domicílio, o qual é influenciado por valores extremos – e como não há renda negativa, os valores extremos aqui discutidos só podem ser os altos – permite entender que existem alguns domicílios com baixos valores de renda. A mesma conclusão se pode tirar de todos os setores com altos valores de renda e de variância.

Por outro lado, analisando-se a Figura 5.3, verifica-se que o padrão espacial dos anos médios de estudos permanece praticamente constante em relação à renda. A diferença entre os dois padrões espaciais (renda e escolaridade) se dá na espacialização da variabilidade da média de anos de estudos.

Os setores centrais, correspondendo aos Bairros de Ramos, Clélia Bernardes e Centro possuem os menores valores de desvio-padrão. Os demais setores com altos valores de anos médios de estudos possuem altos valores de variância, como observado na Figura 5.4.

Toma-se, por exemplo, os setores censitários do Acamari e Romão dos Reis e do Recanto da Serra e Violeira. No primeiro setor, tem-se uma média de 12,03 anos com um desvio-padrão de 6,01 anos. Portanto, consiste num setor que tem uma variação média entre 6,02 e 18,04 anos de estudos.

O segundo setor possui 8,8 anos com desvio de 5,3, variando de 2,5 a 14,1 anos de estudos. Nota-se que ambos consideram apenas um desvio-padrão. Ressalta-se,

ainda, que o máximo acatado pelo IBGE no levantamento é 15 anos de estudos (IBGE, 2003 e ABREU *et al*, 2011).

Alguns setores com baixos valores médios de anos de estudos apresentaram grande variância, tais como os Bairros de Lourdes, Santa Clara, Conceição e Bom Jesus. Deste modo, pondera-se sobre a heterogeneidade interna dos setores censitários quanto ao padrão educacional, através da qual se tem um mascaramento de condições de escolaridade dos residentes nos mesmos.

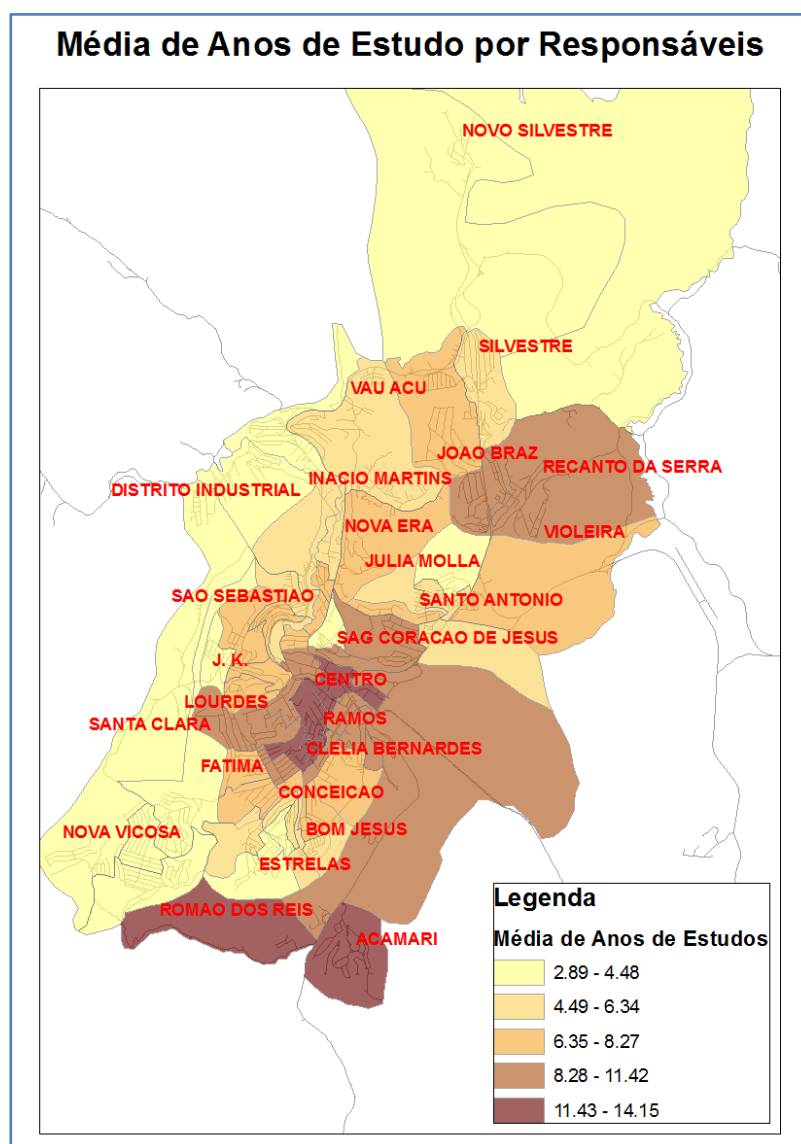


Figura 5.3 Distribuição da Média dos Anos de Estudos por Responsáveis por Domicílios

Tal afirmação é verificada, *in loco*, quando se tem dois setores censitários contendo, juntamente com condomínios fechados ocupados por professores universitários, comunidades carentes, ainda com padrões rurais, como o Romão dos Reis e a Violeira. Esclarece-se, entretanto, que não é pretensão deste trabalho fazer referências sobre padrões de ruralidade e urbanidade no modo de vida das pessoas e implicações deste padrão em suas condições educacionais.

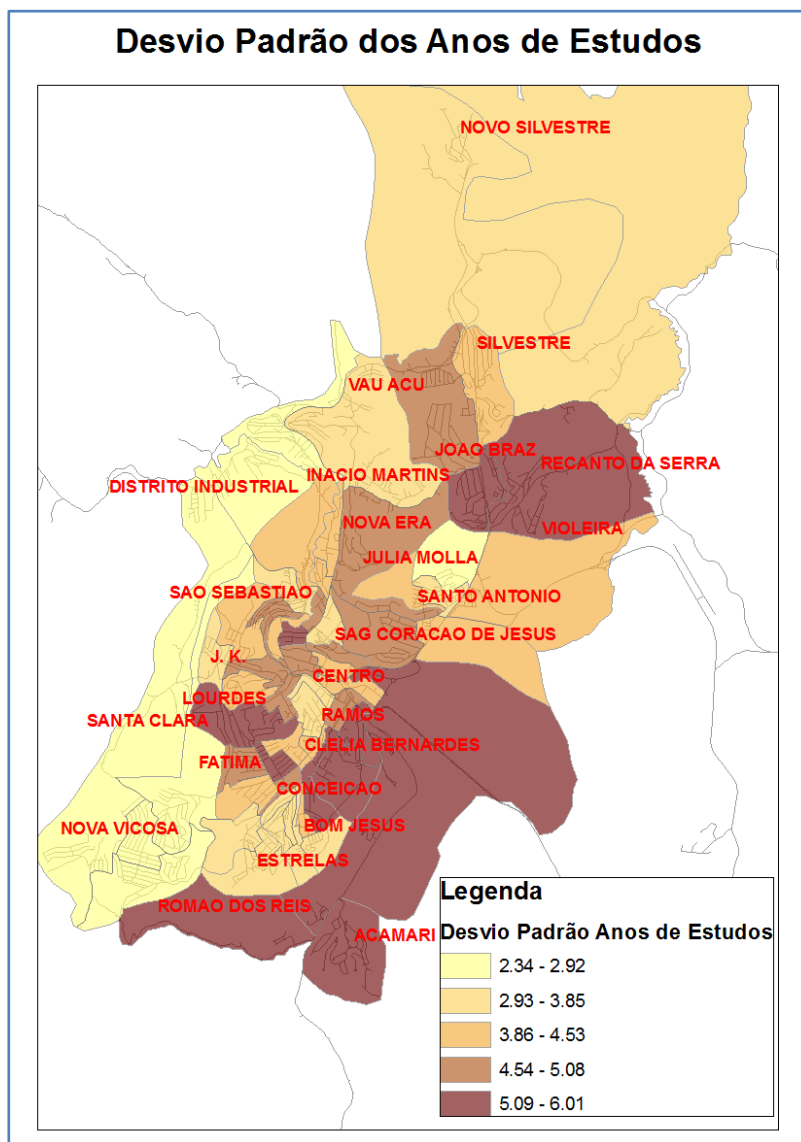


Figura 5.4 Distribuição do Desvio-Padrão dos Anos Médios de Estudos dos Responsáveis por Domicílio

Como abordado anteriormente, a utilização de indicadores socioeconômicos é de grande valia na gestão pública, apoiando diagnósticos e servindo de base para os

processos de planejamento de políticas públicas. Abreu *et al* (2011) tiveram essa intenção ao trabalhar numa proposta de cálculo e espacialização do IDH por setores censitários para a área urbana do Município de Viçosa - MG, motivados principalmente pelo fato de que as políticas sociais municipais não observavam a distribuição intraurbana de condicionantes socioeconômicos.

Contudo, os autores não analisaram as variabilidades dos indicadores, as quais são trabalhadas aqui. Lembra-se que houve uma pequena mudança na metodologia de cálculo do indicador de renda, apresentado na seção 4.3.1.

A Figura 5.5 mostra a espacialização da renda apresentada pela metodologia empregada por Abreu *et al* (2011), enquanto a Figura 5.6 representa a distribuição de renda adotada neste trabalho, que aborda apenas a distribuição de renda por responsáveis por domicílios, e não mais a distribuição por todos os residentes, como naquela metodologia.

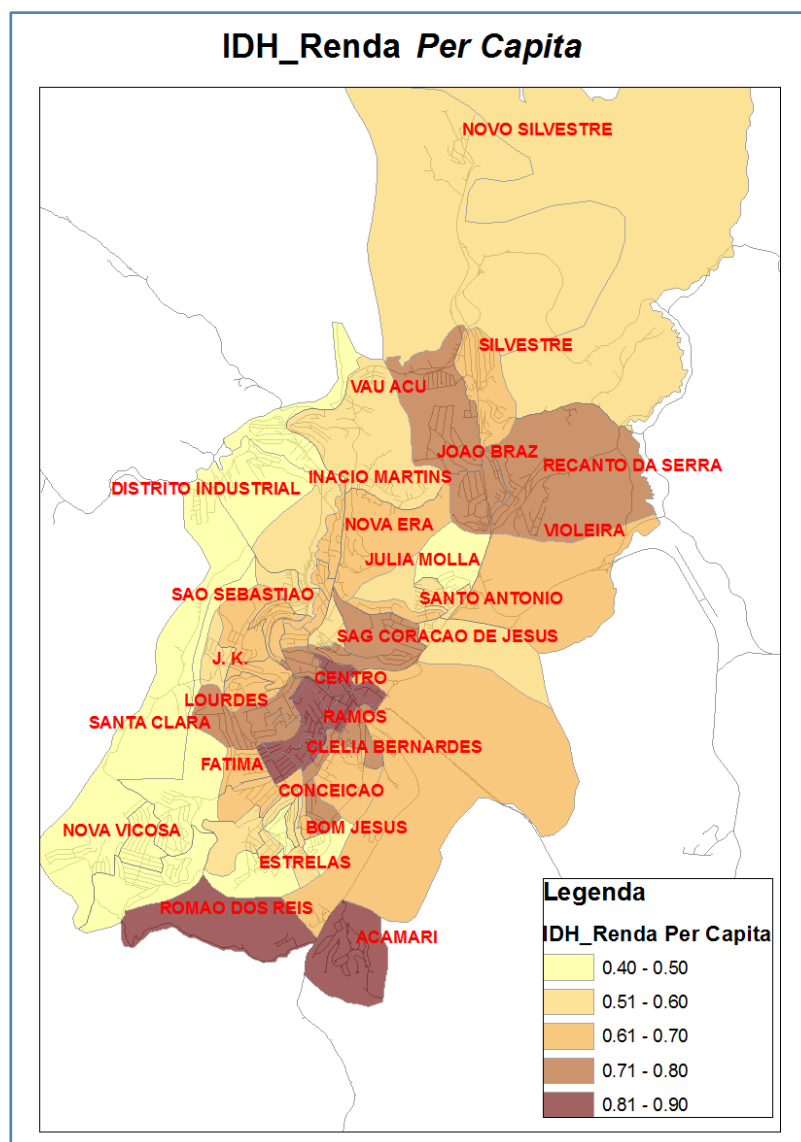


Figura 5.5 IDH Renda *per capita*

Comparando-se as duas figuras, nota-se que a principal mudança se dá nos valores absolutos do indicador de renda. Enquanto a primeira apresenta um intervalo entre 0,39 e 0,90, a segunda apresenta valores entre 0,60 e 1. No entanto, como o intuito desta comparação é apresentar as diferenças locais, percebe-se que, num contexto geral, a configuração espacial relativa aos maiores e menores valores sofre poucas mudanças.

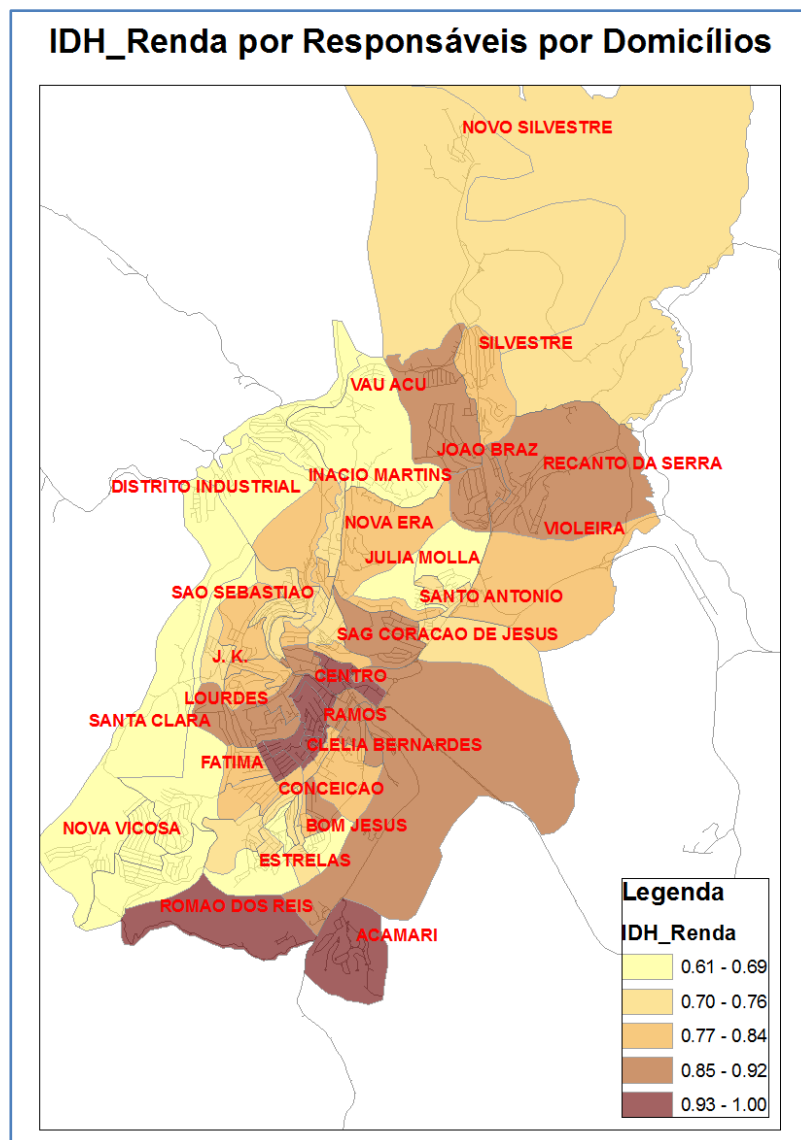


Figura 5.6 IDH Renda por Responsáveis por Domicílios

A amplitude dos valores passa de aproximadamente 0,50 para 0,40. Ressalta-se que os setores censitários foram divididos em 5 classes de intervalos iguais, permitindo a percepção de quais setores mudaram de classe. Percebe-se, assim, uma pequena diferença na área central, em que dois setores censitários pertencentes às regiões dos Bairros de Ramos e Clélia Bernardes, foram, ambos, rebaixados para uma classe anterior. Entretanto, destaca-se que seus valores absolutos aumentaram.

Uma pequena alteração também é perceptível nos setores censitários das regiões do Bairro Vau-Açu, no extremo noroeste, do Bairro São Sebastião, na área centro-oeste, e do Bairro Estrelas, na área centro-sul da cidade. Também houve uma redução da

classe, passando, contudo, para um valor absoluto do indicador maior que na outra metodologia.

Destaca-se que, como não houve mudanças significativas na configuração espacial dos valores altos e baixos do IDH Renda, apesar das mudanças nos valores absolutos, o novo indicador pode ser utilizado na continuação deste trabalho. Assim, a partir desta análise qualitativa, a variável Renda por Responsável por Domicílio pode ser usada para a posterior verificação da heterogeneidade dos setores censitários, cuja representação é dada na Figura 5.7.

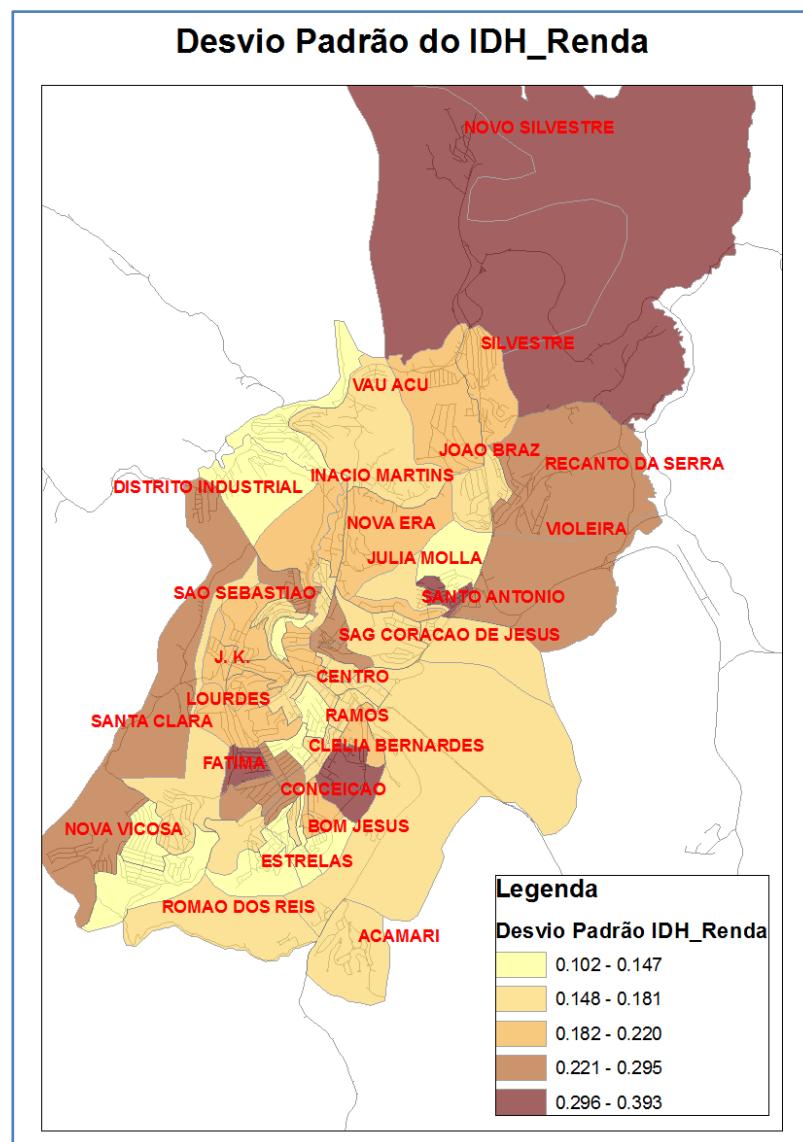


Figura 5.7 Distribuição do Desvio-Padrão do IDH Renda

A principal consideração a ser feita é notar a diferença entre o padrão espacial dos desvios da distribuição da renda e do IDH Renda. Os setores com maior variabilidade no valor da renda apresentaram menor dispersão no indicador analisado. Isso pode ser explicado pela Equação 21, em cujo denominador aparece o valor da Renda por Responsável por Domicílio (RPR), o qual é alto nestes setores, diminuindo o valor das variâncias dos mesmos setores.

É interessante observar que alguns setores com menores desvios na distribuição da renda continuam apresentando baixos desvios na distribuição do indicador, como os casos da parte leste de Nova Viçosa, dos Bairros Estrelas, J.K., São Sebastião, Vau-Açu e a parte norte do Bairro Santo Antônio. Destaca-se, ainda, que os setores que representam esses bairros são de baixa renda.

Já do ponto de vista dos altos valores de desvio-padrão do indicador, apontam-se como realce as áreas oeste do Bairro de Fátima, parte leste ao Bairro da Conceição e uma área central do Bairro Santo Antônio. Destacam-se também os setores na segunda maior classe apresentada na Figura 5.7, principalmente aquele que compreende a faixa que liga o Nova Viçosa e o Distrito Industrial.

Assim, percebe-se que o grau de variabilidade é influenciado pelo valor da Renda por Responsável por Domicílio, além, claro, pelo valor da variância da própria RPR. Tal fato possui leve semelhança com a ideia do coeficiente de variação, deixando-se claro, obviamente, que a propagação de variâncias não tem nada a ver com tal coeficiente, sendo apenas uma coincidência matemática.

Por sua vez, a Figura 5.8 mostra o indicador de educação, IDH Escolaridade, espacializado. Destaca-se que, numa análise geral, a configuração de valores altos e baixos permanece semelhante ao indicador de renda. O baixo valor do setor censitário da região do Bairro Sagrado Coração de Jesus é enfatizado por ser vizinho de setores com valores consideravelmente altos.

Os setores correspondentes ao Bairro Nova Viçosa se ressaltam com valores extremamente baixos, contradizendo um ponto forte da Cidade de Viçosa – MG, que é a qualidade da educação de algumas instituições de ensino presentes em seu território (ABREU, 2009). O contraste espacial apresentado tanto na Figura 5.3 quanto na Figura 5.8 mostra que o acesso ao ensino de qualidade não é para toda a população viçosense.

Por sua vez, a Figura 5.9 expõe os desvios-padrão do indicador educacional do desenvolvimento humano. Ela apresenta um padrão espacial igual ao da Figura 5.4, fato

explicado pela Equação 22, que mostra que a função derivada do IDH Educação é uma constante. Logo, pela Equação 26, tem-se que para encontrar as variâncias do indicador educacional basta dividir as variâncias dos anos médios de estudos dos Responsáveis por Domicílios por 4. Devido a esse fato, as análises dos valores apresentados tornam-se dispensáveis, já que seriam redundantes nesta seção.

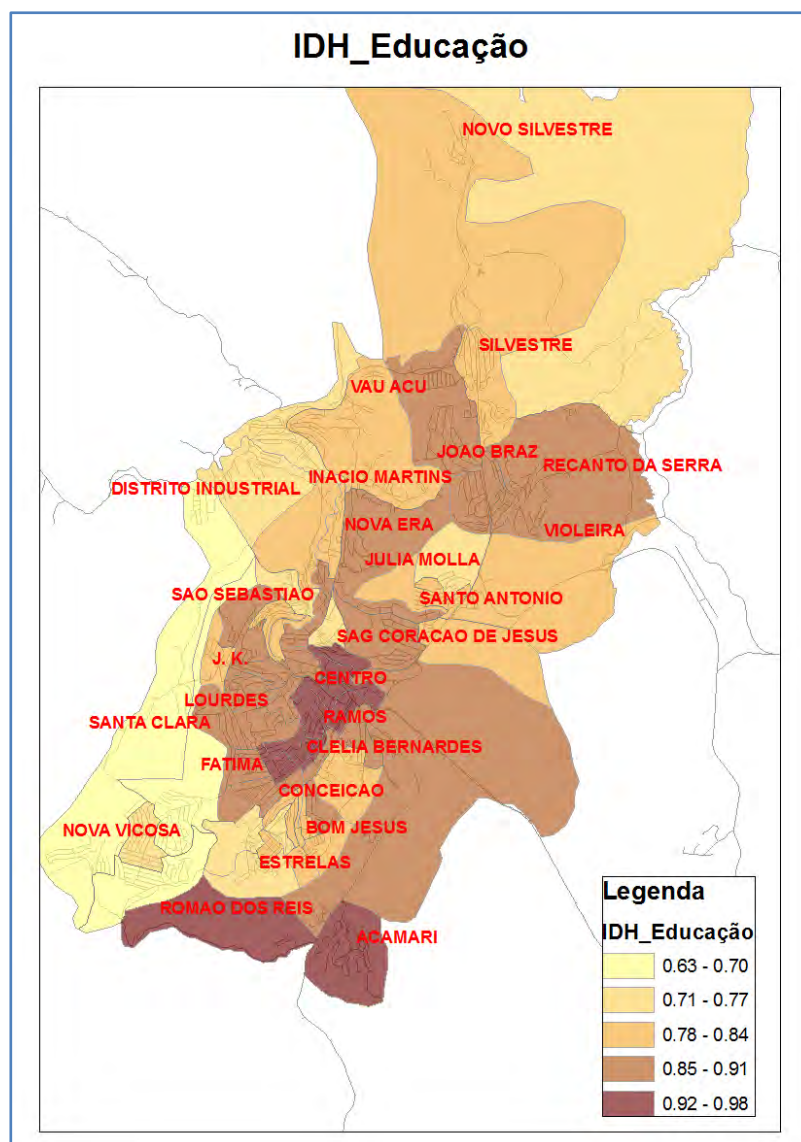


Figura 5.8 IDH Educação

Nota-se, ainda na Figura 5.9, que os valores de desvio-padrão do IDH Educação de alguns setores ultrapassam a amplitude do próprio indicador, que é de 0 a 1. Tal ocorrência indica uma grande variabilidade, podendo ser explicada pelo fato de que havia maiores valores de variância da média dos anos de estudos.

A intenção desta parte do trabalho era obter uma variância para o IDH, mostrando quais setores censitários possuíam maior heterogeneidade do indicador socioeconômico escolhido. Porém, como abordado na seção 4.3.1, as unidades das variâncias dos dois indicadores utilizados, renda e educação, não são compatíveis, sendo inviável a sua propagação para o Índice de Desenvolvimento Humano.

Assim, optou-se por verificar o grau de influência das duas variáveis no IDH a partir do coeficiente de variação. Num primeiro momento, avaliou-se o indicador de renda, o qual é apresentado na Figura 5.10, mais adiante.

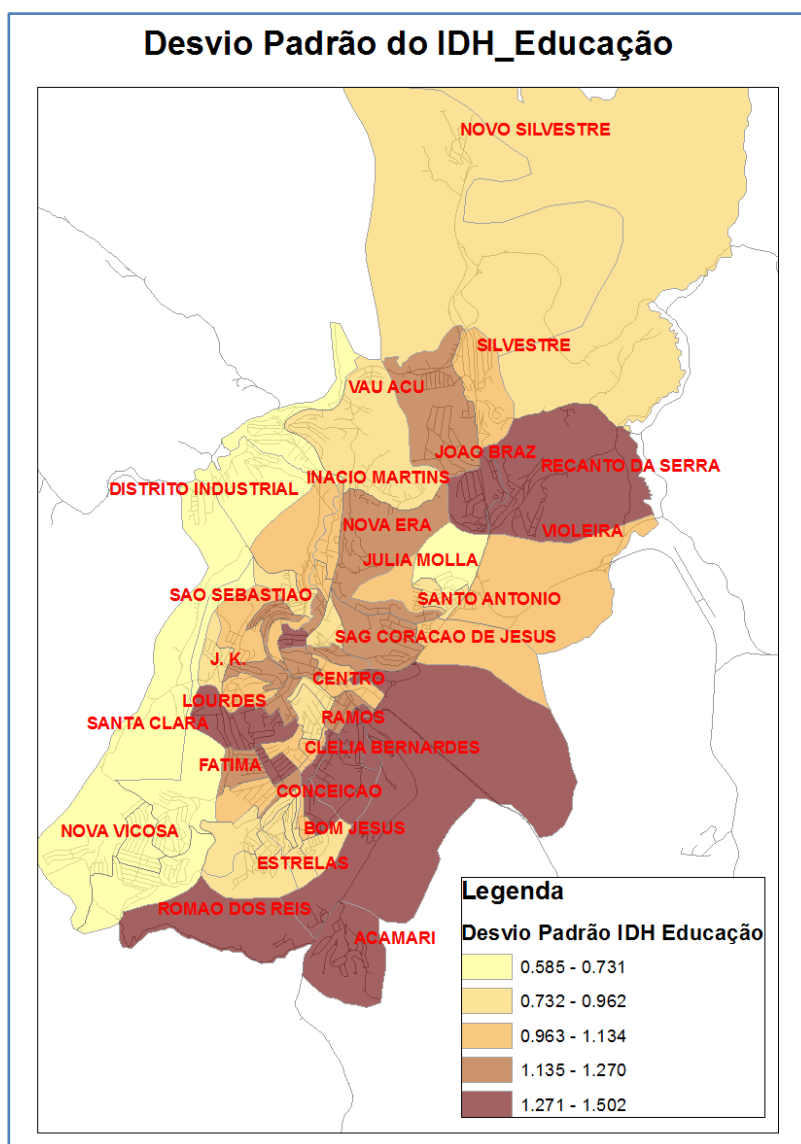


Figura 5.9 Distribuição do Desvio-Padrão do IDH Educação

Como pode ser visto na Figura 5.10, nenhum setor apresenta grande dispersão em torno do valor médio do IDH Renda. Todos eles apresentaram um coeficiente de variação menor que 1, ou também interpretável como menor que 100%. Verifica-se, portanto, que, apesar de haver grandes variâncias nos valores relativos às Rendas por Responsáveis por Domicílios, o que faz o desvio-padrão ser até maior que o valor médio observado, elas não exercem alto grau de variabilidade no indicador de renda.

Já a Figura 5.11 mostra o coeficiente de variação do indicador de educação, IDH Educação. Diferentemente do apresentado acima, a grande maioria dos setores censitários possui maior dispersão em torno do valor médio.

Destacam-se, contudo, aqueles setores com maior homogeneidade, quais sejam os setores de Nova Viçosa, os setores a oeste da Região do Vau-Açu, além do norte do Bairro Santo Antônio (Morro da Coruja) e do Bairro de Ramos.

Enquanto o último consiste num bairro mais desenvolvido, os outros são ocupados por famílias de baixa renda. A homogeneidade é notada, portanto, independentemente do padrão educacional apresentado nos setores, podendo ser altos ou baixos.

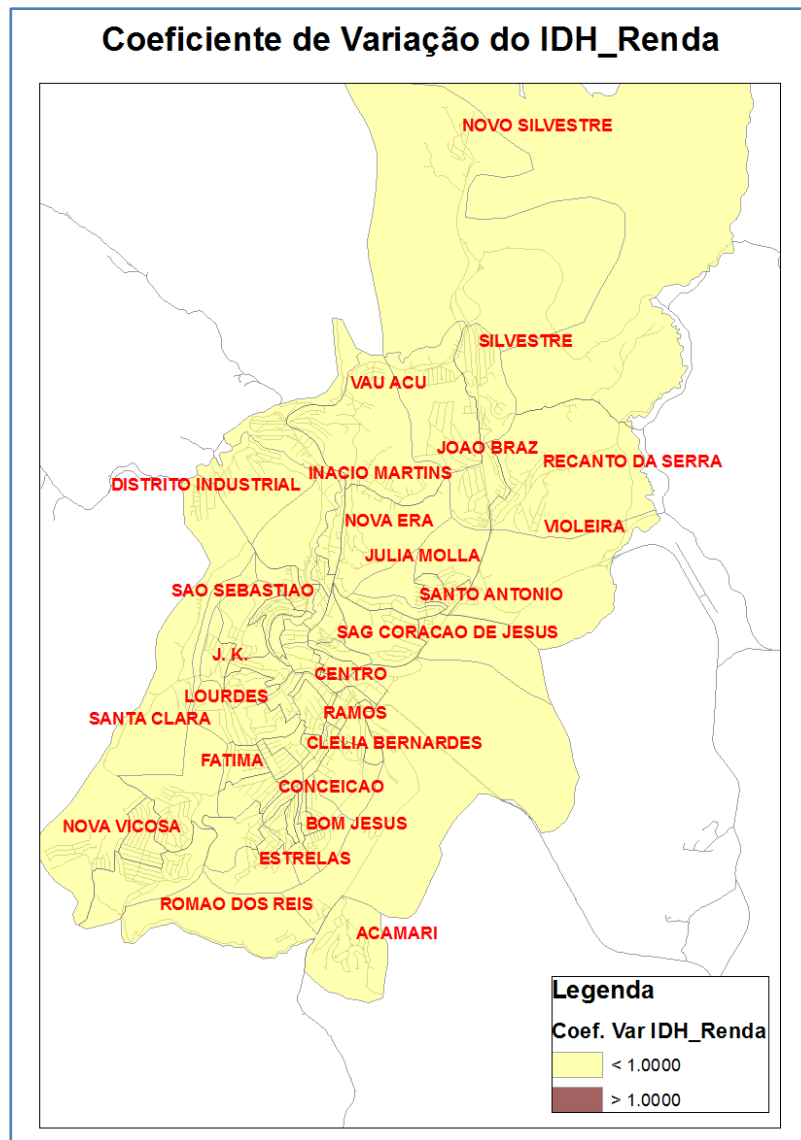


Figura 5.10 Coeficiente de Variação do IDH Renda

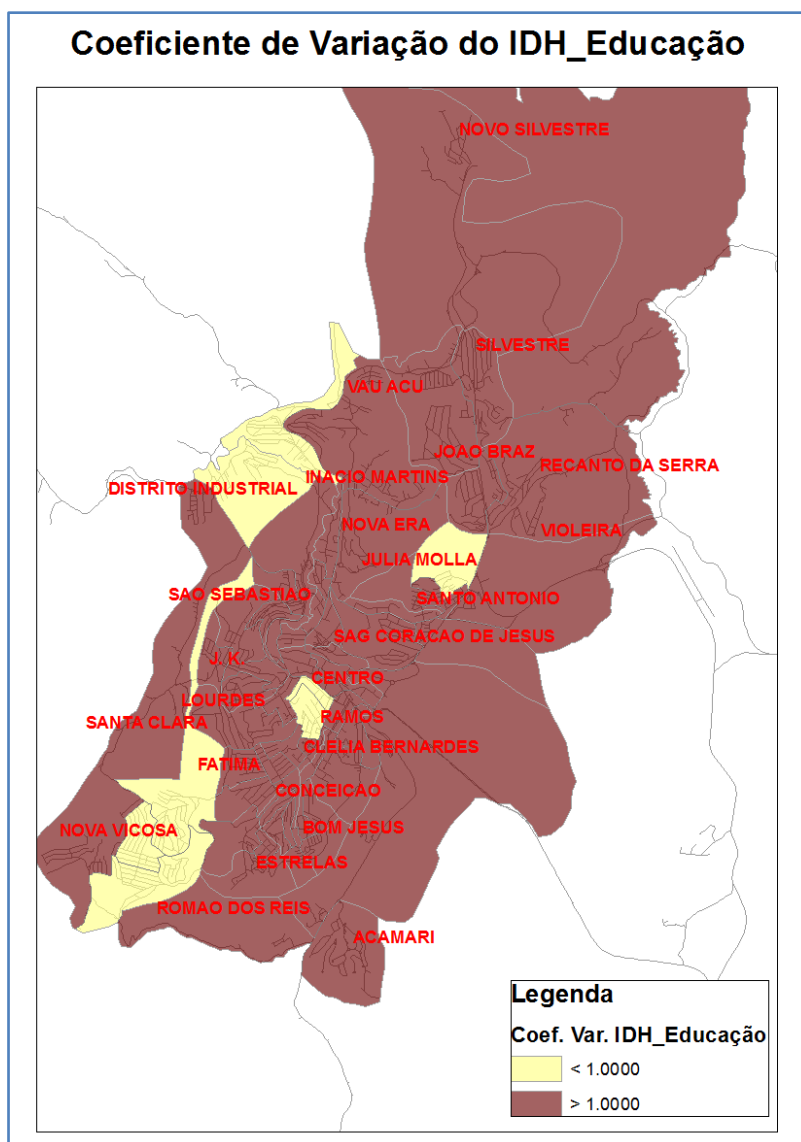


Figura 5.11 Coeficiente de Variação do IDH Educação

Com os resultados apresentados, podem-se tirar algumas conclusões acerca do nível de homogeneidade dos setores censitários. Alguns setores são naturalmente compostos por residentes que apresentam as mesmas características socioeconômicas. Por outro lado, percebe-se uma grande quantidade de setores com grande variabilidade interna.

Isso certamente mascara alguns diagnósticos socioeconômicos, feitos em processos de leitura da cidade. Como saber, por exemplo, da existência de comunidades mais carentes quando se tem num mesmo setor censitário comunidades com maior padrão de vida, como são os casos dos setores censitários do Acamari e Romão dos Reis e do Recanto da Serra e Violeira?

É sabido que alguns municípios brasileiros buscam coincidir os limites censitários com os limites dos bairros, permitindo cruzar informações sociodemográficas com informações administrativas, provindas de ações e políticas públicas destinadas aos bairros.

Tal coincidência pode ser interessante para o poder público municipal, porém, como visto nesta seção, caso os limites censitários sejam estabelecidos sem critérios de homogeneidade, os mesmos podem mascarar a realidade, não retratando as condições socioeconômicas verdadeiras.

Inferir que algumas comunidades possuem determinado padrão de renda ou de educação, através de leituras de dados agregados por setores censitários, pode ser errôneo. Na verdade, em leituras deste tipo ocorre a Falácia Ecológica (FE). Entretanto, quando se tem um bom retrato da realidade socioeconômica, com setores mais homogêneos, possivelmente, os efeitos da FE podem trazer menores erros em processos de diagnósticos municipais.

5.2. Avaliação das Zonas Homogêneas

O modelo proposto e testado nas variáveis genéricas U e V tem seus resultados apresentados adiante. Ressalta-se, novamente, que os procedimentos indicados não são necessariamente únicos, devendo-se preservar apenas a ordem de tarefas. Isso acontece porque a maioria dos procedimentos é iterativa, dependendo muitas vezes da experiência e da sensibilidade do analista.

Ademais, em cada conjunto de dados, os pontos/residências apresentam valores diferentes, com arranjos espaciais diferentes. Assim, a dependência espacial modelada na interpolação varia significativamente, tendo como conseqüência a possibilidade de se encontrar diferentes números de agrupamentos (*clusters*) no processo de detecção de padrões espaciais dos pixels da superfície contínua.

Além disso, o desenho urbano, o relevo e os limites geográficos de uma determinada cidade, na qual se aplica esta metodologia, pode determinar também a criação dos limites censitários, trazendo simplificações ou modificações expressivas nas fronteiras socioeconômicas determinadas.

De qualquer forma, apresentam-se os resultados comentados e discutidos de cada procedimento presente na metodologia proposta neste trabalho. Verifica-se, ao

final, que a consideração da homogeneidade espacial dos dados traz resultados que se destacam quando comparados a padrões aleatórios, podendo-se inferir sobre o ponto de vista socioeconômico.

5.2.1. Geração das Superfícies Contínuas

Após a exclusão dos 195 pontos, foi realizada a análise exploratória dos dados, verificando a distribuição espacial e estatística de seus valores. Notam-se, na Figura 5.12, a distribuição espacial dos pontos com a definição dos limites de área de estudos.

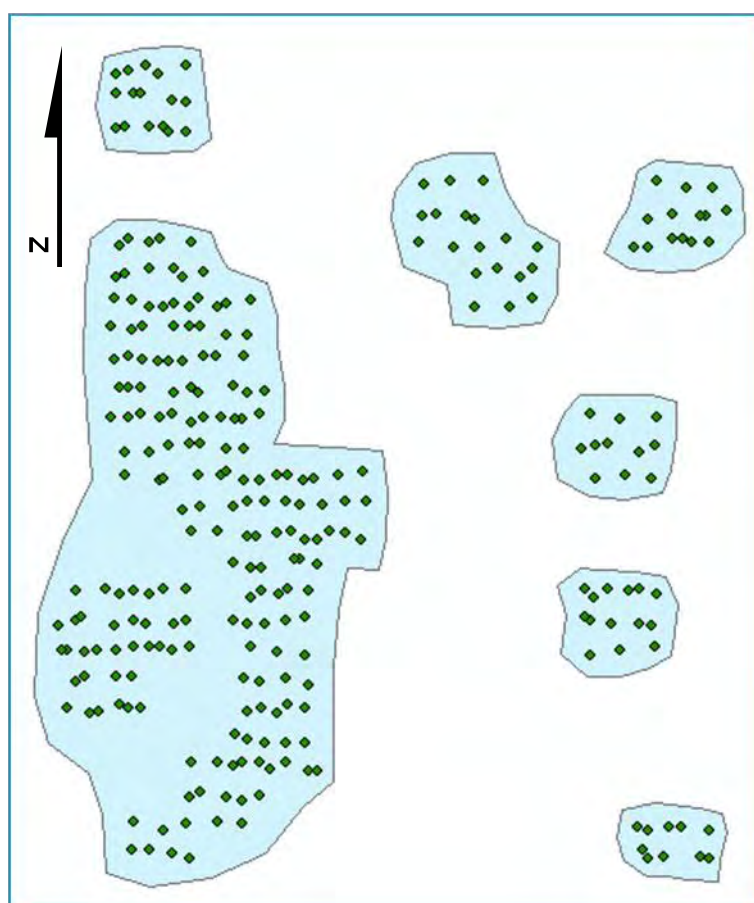


Figura 5.12 Distribuição espacial dos pontos estudados

Quanto à distribuição estatística, têm-se na Tabela 5.1 os valores médios, desvios-padrão, mínimo, máximo e coeficientes de assimetria e de curtose de cada variável. Registra-se que, como U e V são variáveis genéricas, as unidades de medida utilizadas por Isaaks e Srivastava (1989) foram p.p.m., não sendo importantes para este trabalho.

Tabela 5.1 – Parâmetros Estatísticos das Variáveis V e U

	V	U
Média	548,92	613,17
Desvio – Padrão	283,88	768,76
Mínimo	0	0
Máximo	1528,92	5190,10
Coef. Assimetria	0,27	2,24
Curtose	3,33	9,78

Nota-se que a distribuição de V se aproxima da distribuição normal, enquanto a de U é bastante assimétrica, com grande concentração de valores pequenos. Em contrapartida, essa variável possui alguns valores extremamente grandes, levando algumas agregações centrais a possuírem grandes variâncias, como apresentado mais adiante.

A Figura 5.13 mostra uma análise exploratória espacial, retratando tendências locais a altos ou baixos valores para ambas as variáveis. É interessante notar que o lado leste da área de estudos, o qual foi considerado na simulação como sendo a área urbana do distrito sede do município fictício, apresenta os maiores valores para U e V. Tal fato também apoia a utilização destes dados para a simulação, uma vez que essa característica é notada na maioria dos municípios brasileiros, ou seja, o distrito sede concentra as maiores condições socioeconômicas da população.

Nota-se também que a variável U possui 6 realizações de altíssimos valores no sentido noroeste da área de estudos. São exatamente nos setores criados presentes nesse local que aparecem as maiores variâncias. Após essa análise, resolveu-se não retirar esses pontos, classificando-os como possíveis *outliers*, uma vez que o IBGE não faria a exclusão de pontos extremos em sua base, pois foram levantados e representam uma realidade.

Mais além, a krigagem suaviza a superfície gerada, não aparecendo na superfície uma pequena área extrema, o que aconteceria numa interpolação determinística, por exemplo.

Se comparadas com as variáveis renda e educação, trabalhadas anteriormente, percebe-se certa semelhança com a distribuição dos valores, excetuando-se, obviamente, a presença dos extremos valores de U.

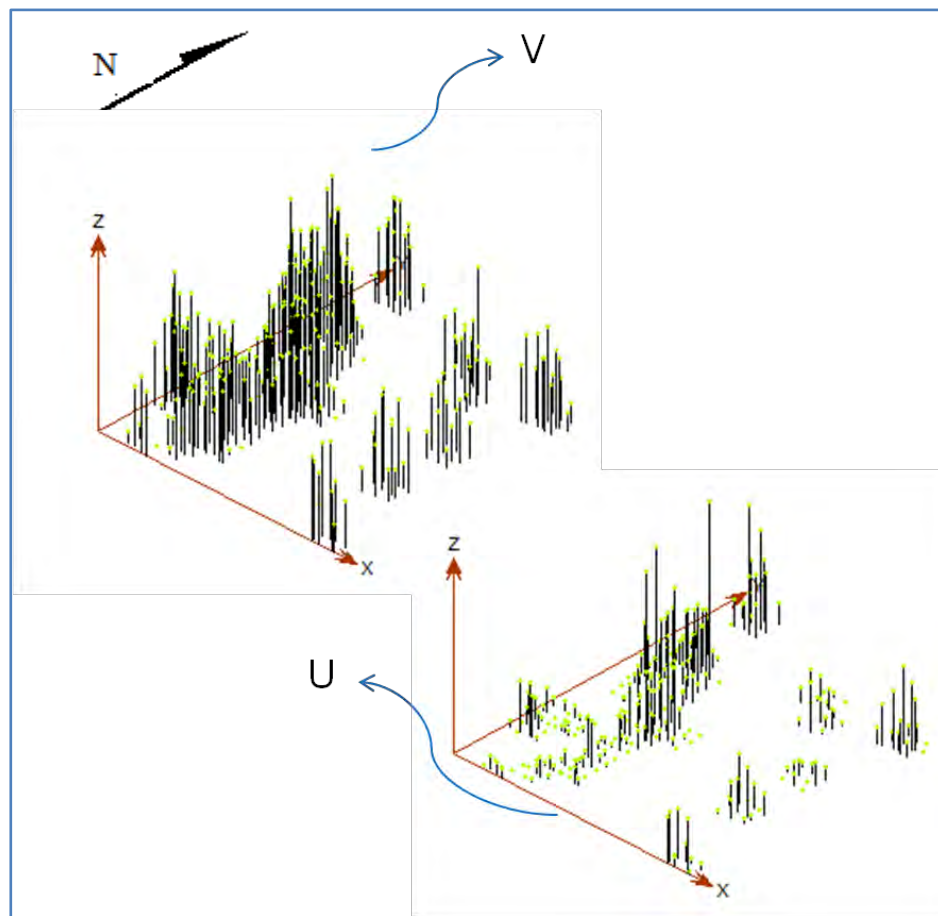


Figura 5.13 Distribuição espacial dos pontos estudados

Em seguida, o processo de modelagem da dependência espacial dos pontos realizado pelos semivariogramas constatou anisotropia para as duas variáveis, o que pode ser percebido analisando a Figura 5.13 novamente. Nota-se que a continuidade espacial é mais acentuada no sentido norte-sul, seja pelos valores apresentados nas variáveis ou mesmo pela descontinuidade espacial existente no sentido leste-oeste. Os parâmetros dos semivariogramas são dados na Tabela 5.2.

Tabela 5.2 – Parâmetros dos Semivariogramas das Variáveis V e U

	V	U
Alcance	30,5	140
Patamar	88500	785000
Efeito Pepita	14000	502500
Contribuição	74500	282500
Modelo Teórico	Esférico	Gaussiano
Dir. Anisotropia	343°	2°

A Figura 5.14 mostra os semivariogramas de V e U, respectivamente.

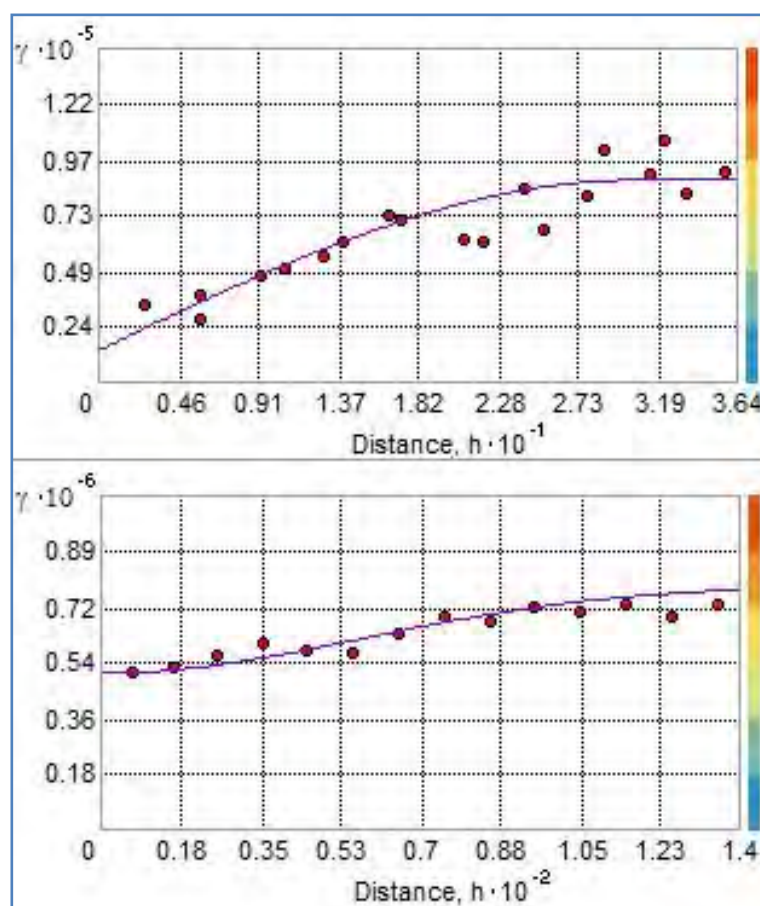


Figura 5.14 Ajuste dos semivariogramas teóricos sobre os empíricos

Dada a modelagem dos semivariogramas, realizaram-se as interpolações, gerando-se as duas superfícies contínuas. É certo que as áreas externas aos polígonos limitantes também obtiveram valores, no entanto, serão desconsideradas da análise, pois

o objetivo não era estimar valores nesses locais. As Figuras 5.15 e 5.16 mostram as variáveis V e U, respectivamente.

Aqui não se preocupa com os valores apresentados nas classes em ambas as superfícies. O que interessa, nesta proposição, é apenas mostrar como se dá a variação espacial das variáveis, o que auxilia na visualização de áreas com valores aproximados, ou melhor, mais homogêneos.

É certo que, visualmente, é fácil identificar alguns limites, mas lança-se mão das ferramentas do Processamento Digital de Imagens, que, com critérios estatísticos, é mais apropriada para determinar tais fronteiras de homogeneidade.

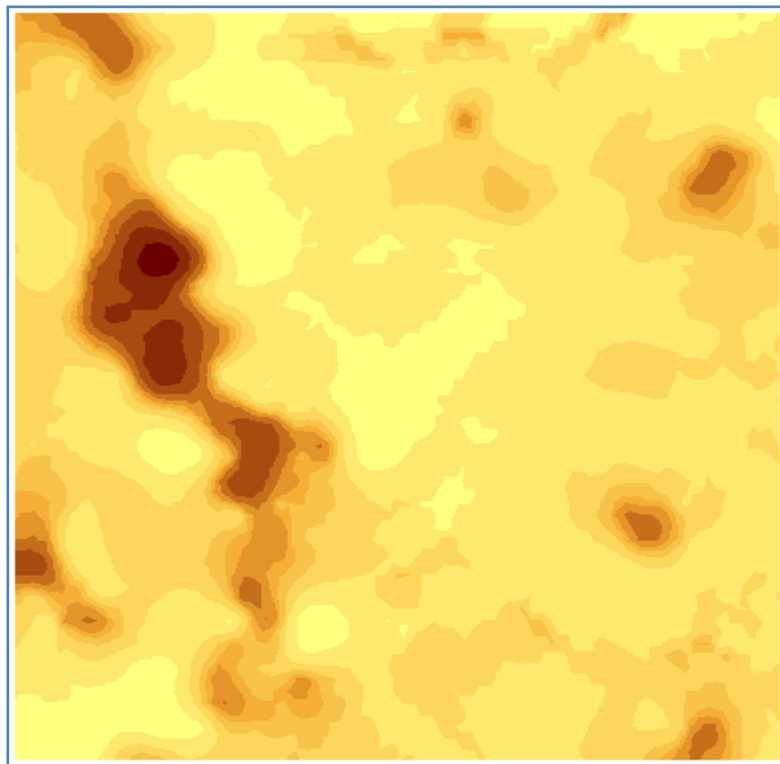


Figura 5.15 Superfície contínua da variável V

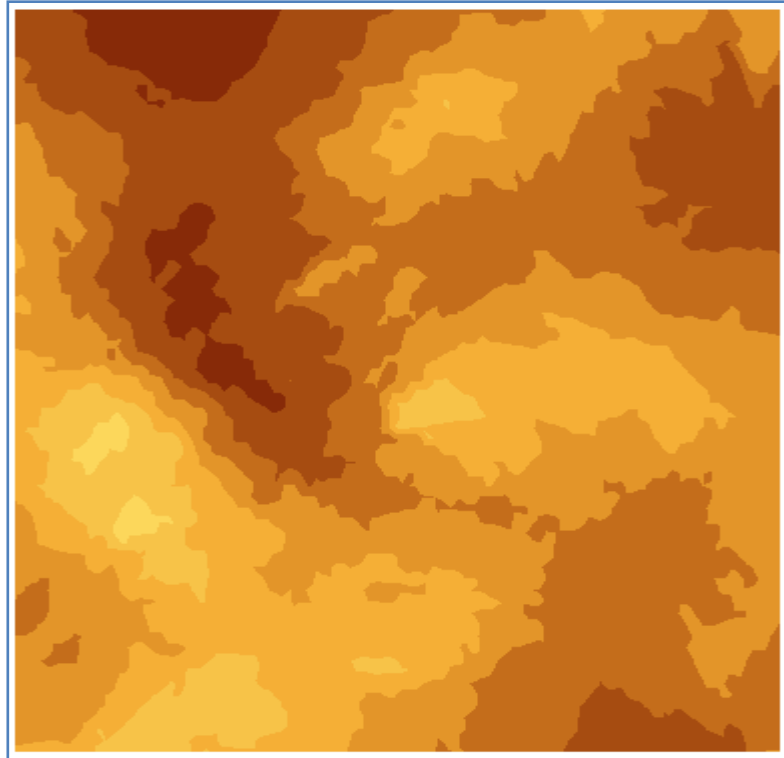


Figura 5.16 Superfície contínua da variável U

5.2.2. Discretização da Superfície Contínua

Com a utilização do algoritmo *Isodata*, presente no *software IDRISI ANDES* (comando *Isoclust*), foi possível realizar os testes de detecção de padrões dos valores de pixels. Após vários testes realizados em cada variável e na composição das duas, verificou-se que mudanças em alguns parâmetros não alteravam os resultados. Em verdade, quando havia alteração, era muito abrupta, não sendo validada por não corresponder ao que se esperava após a análise visual das superfícies.

As Figuras 5.17, 5.18 e 5.19 mostram os *clusters* detectados para as variáveis V, U e para a composição VU, respectivamente. Os parâmetros ideais para os procedimentos para a variável V foram 6 *clusters* possíveis, com mínimo de 50 pixels para cada em 3 iterações máximas. A variável U teve 6 *clusters* possíveis, com 50 pixels cada em 5 iterações. Por último, a composição trabalhou com 3 iterações para detectar 12 *clusters* com 30 pixels cada.

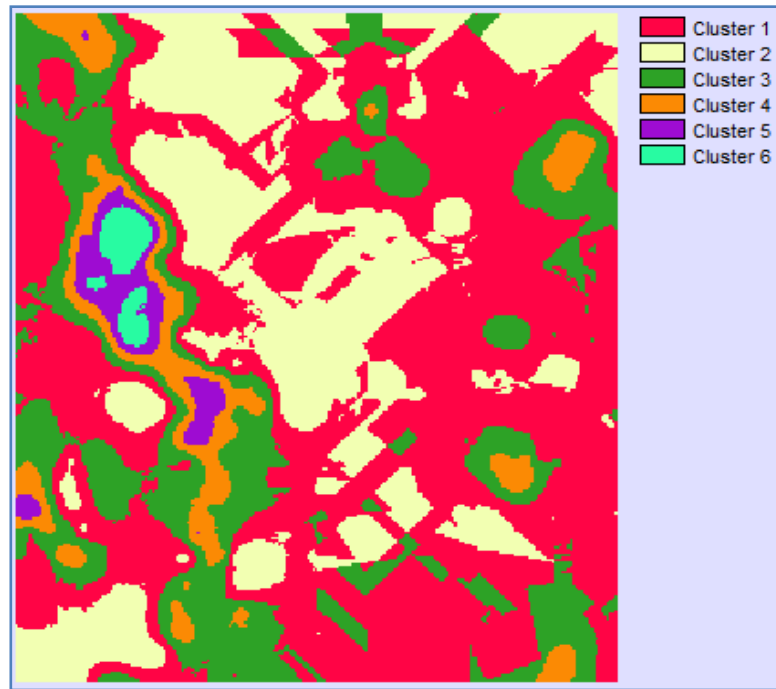


Figura 5.17 *Clusters* da variável V

Nota-se que os polígonos amarelos (*Cluster 2*) na área central da imagem não possuem um padrão bem definido como aqueles onde havia pontos. Isso mostra que a interpolação para inferência de valores em locais onde não há coleta, para esta metodologia, pode não ser interessante. Por exemplo, o que poderia garantir que os pequenos polígonos vermelhos (*Cluster 1*) realmente possuem valores diferentes que os polígonos amarelos em sua volta, uma vez que não houve dado ali para comprovar isso?

A mesma análise feita anteriormente é válida para as duas figuras seguintes, principalmente na Figura 5.19, cujo meio é representado por *clusters* da composição de duas variáveis estimadas nesse local. De qualquer forma, o que interessa neste procedimento é apenas encontrar fronteiras desses padrões, o que torna dispensável a identificação de classes (riqueza ou pobreza, alta ou baixa escolaridade – genericamente neste modelo V e U).

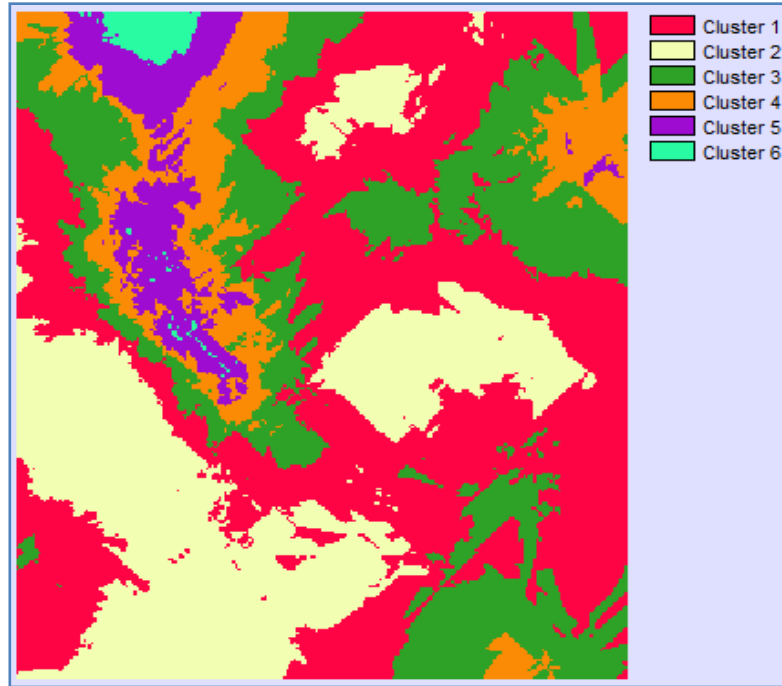


Figura 5.18 *Clusters* da variável U

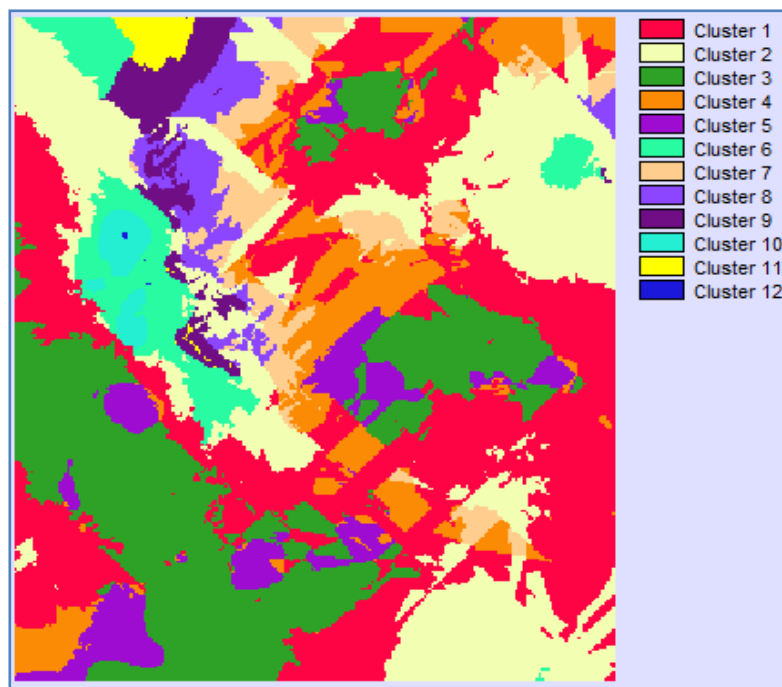


Figura 5.19 *Clusters* da composição VU

5.2.3. Determinação das Fronteiras Socioeconômicas

Com os polígonos gerados, converteram-se suas bordas em feições lineares. As Figuras 5.20, 5.21, 5.22 e 5.23 mostram a sobreposição das fronteiras encontradas às suas superfícies geradoras.

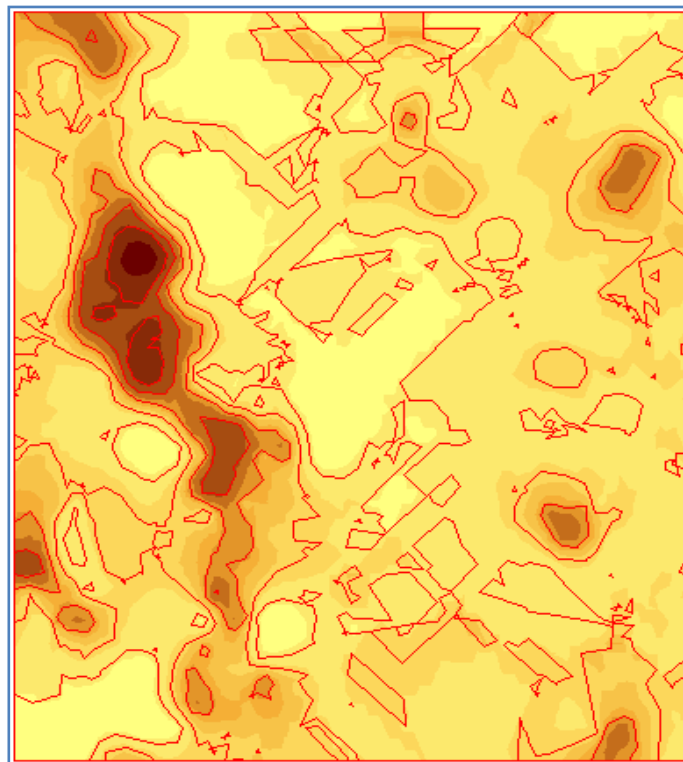


Figura 5.20 Sobreposição das fronteiras da variável V

Nota-se que em alguns locais, nas quatro figuras, as fronteiras não coincidem com as classes apresentadas. Isso porque, para a classificação das legendas das duas superfícies, não se usou o mesmo critério estatístico que o algoritmo *Isodata*. Ressalta-se aqui apenas a consistência desse algoritmo para a proposta.

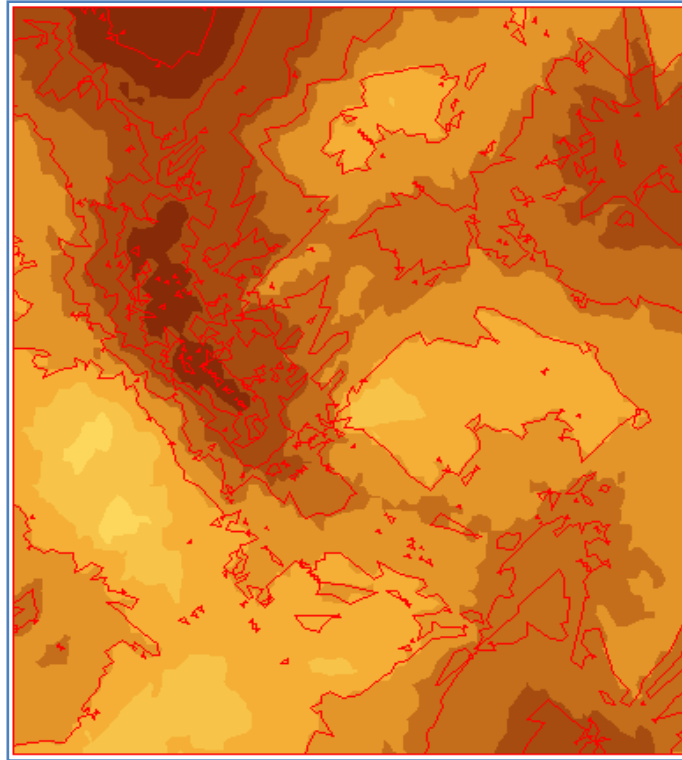


Figura 5.21 Sobreposição das fronteiras da variável U

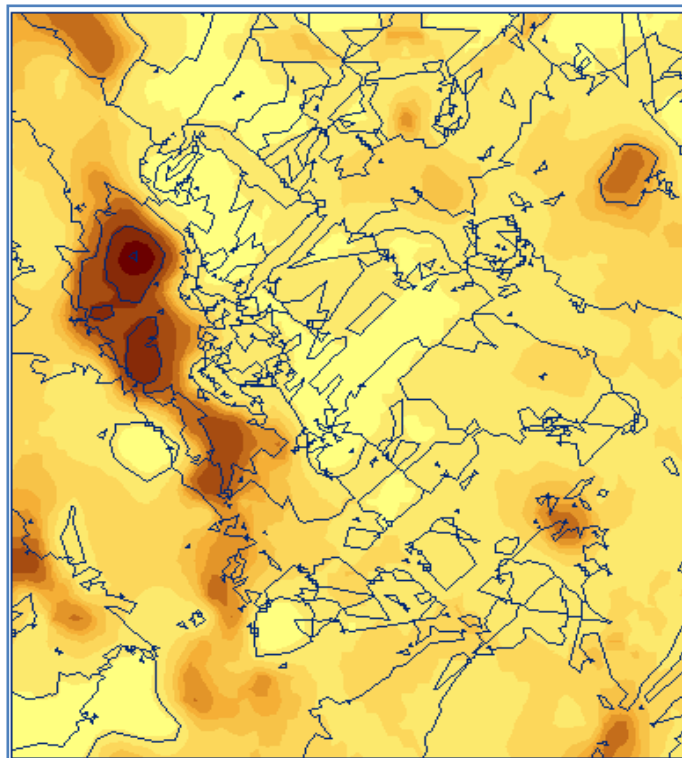


Figura 5.22 Sobreposição das fronteiras da composição VU à superfície U

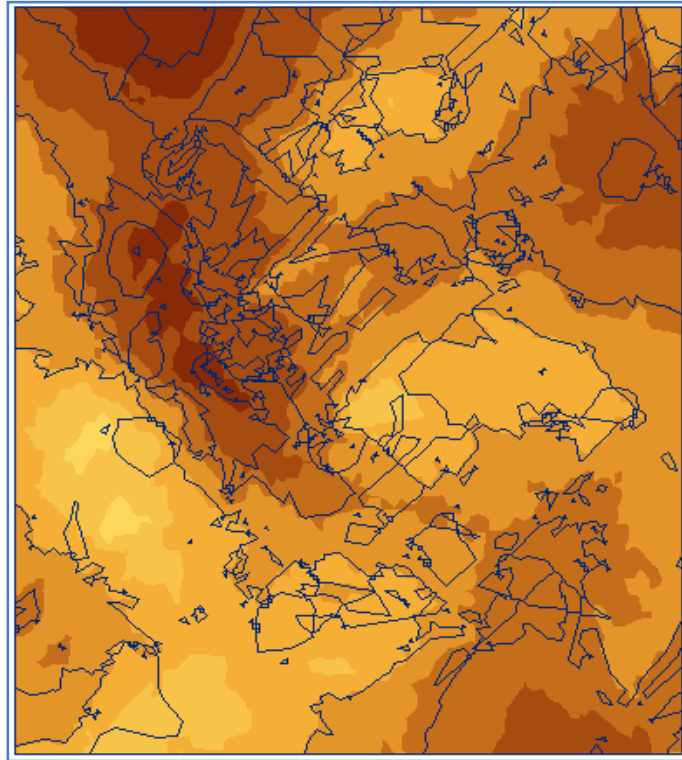


Figura 5.23 Sobreposição das fronteiras da composição VU à superfície V

Como pode ser visto na Figura 5.24, os limites da composição VU não correspondem exatamente com os limites da sobreposição das fronteiras de V às de U, representadas na Figura 5.25.

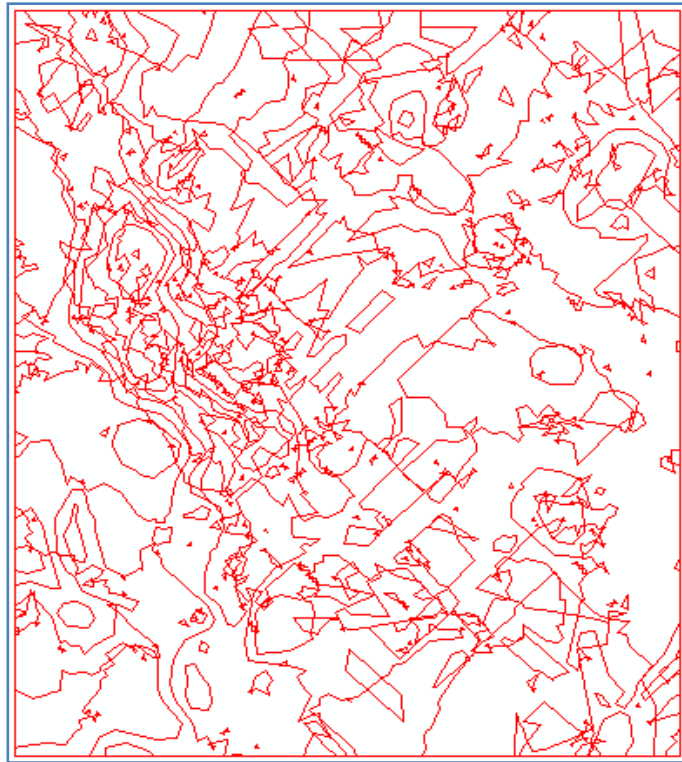


Figura 5.24 Sobreposição das fronteiras de V e U

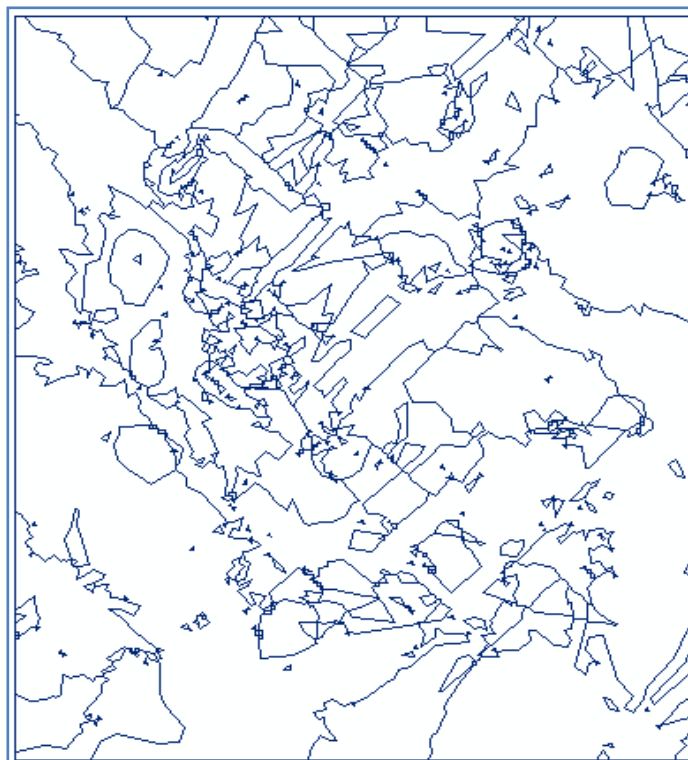


Figura 5.25 Fronteiras da composição VU

5.2.4. Zoneamento Homogêneo

Após a eliminação de pequenos polígonos indesejados, edição topológica e ajustes das bordas com os perímetros dos distritos do município fictício, foram estabelecidos os limites do zoneamento homogêneo. Destacam-se a presença de dois zoneamentos, sendo um com a sobreposição de V a U e outro derivado da composição VU.

Nas Figuras 5.26 e 5.27, têm-se os dois zoneamentos homogêneos estabelecidos.



Figura 5.26 Zoneamento Homogêneo para a sobreposição de V a U

Como a Figura 5.26 apresenta mais fronteiras que a Figura 5.27, nota-se que o algoritmo aplicado individualmente ocasiona um resultado diferente do apresentado na segmentação da composição. Isto é, a análise de detecção de padrões por bandas pode ter trazido um resultado mais consistente por verificar a presença de *cluster* nas duas superfícies ao mesmo tempo.

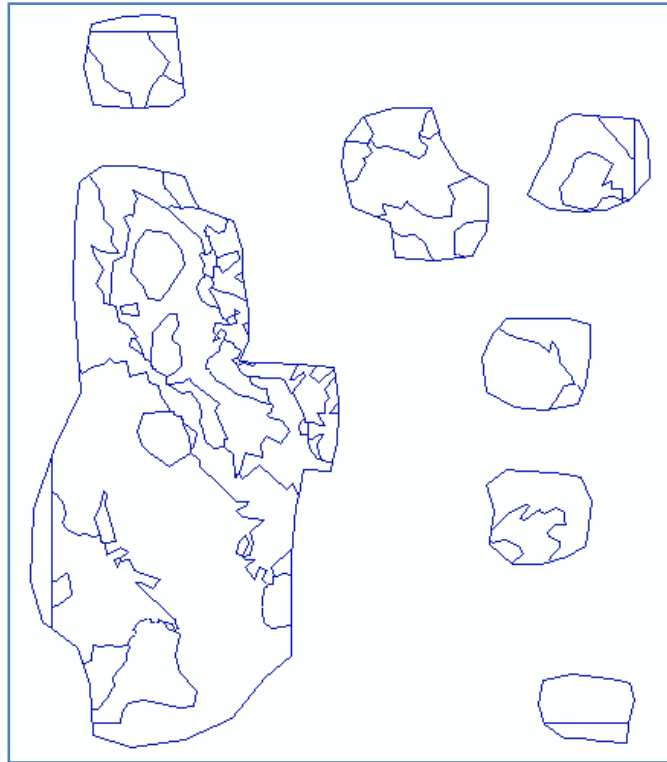


Figura 5.27 Zoneamento Homogêneo para a composição VU

É importante lembrar que todos os passos envolvidos até o momento, relativos ao zoneamento homogêneo, visam auxiliar o estabelecimento dos setores censitários. Busca-se usar os limites apresentados nas figuras anteriores, os quais devem ser respeitados, à medida do possível, para a confecção dos limites censitários.

É certo, contudo, que o próprio zoneamento socioeconômico homogêneo pode ser utilizado pela gestão pública, permitindo não só inferir sobre as condições populacionais, mas também para dar apoio aos processos de planejamento de políticas públicas.

De qualquer forma, os setores censitários homogêneos gerados neste trabalho foram baseados nestes limites homogêneos.

5.2.5. Determinação dos Setores Censitários a partir do Zoneamento Homogêneo

Como abordado na Seção 4.3.2.5, foram criados 7 arranjos de setores censitários, sendo três deles de forma aleatória (A1, A2 e A3), 2 homogêneos sem rigor (HS1 e HC1) e 2 homogêneos com rigor (HS2 e HC2). A Figura 5.28 apresenta tais arranjos.

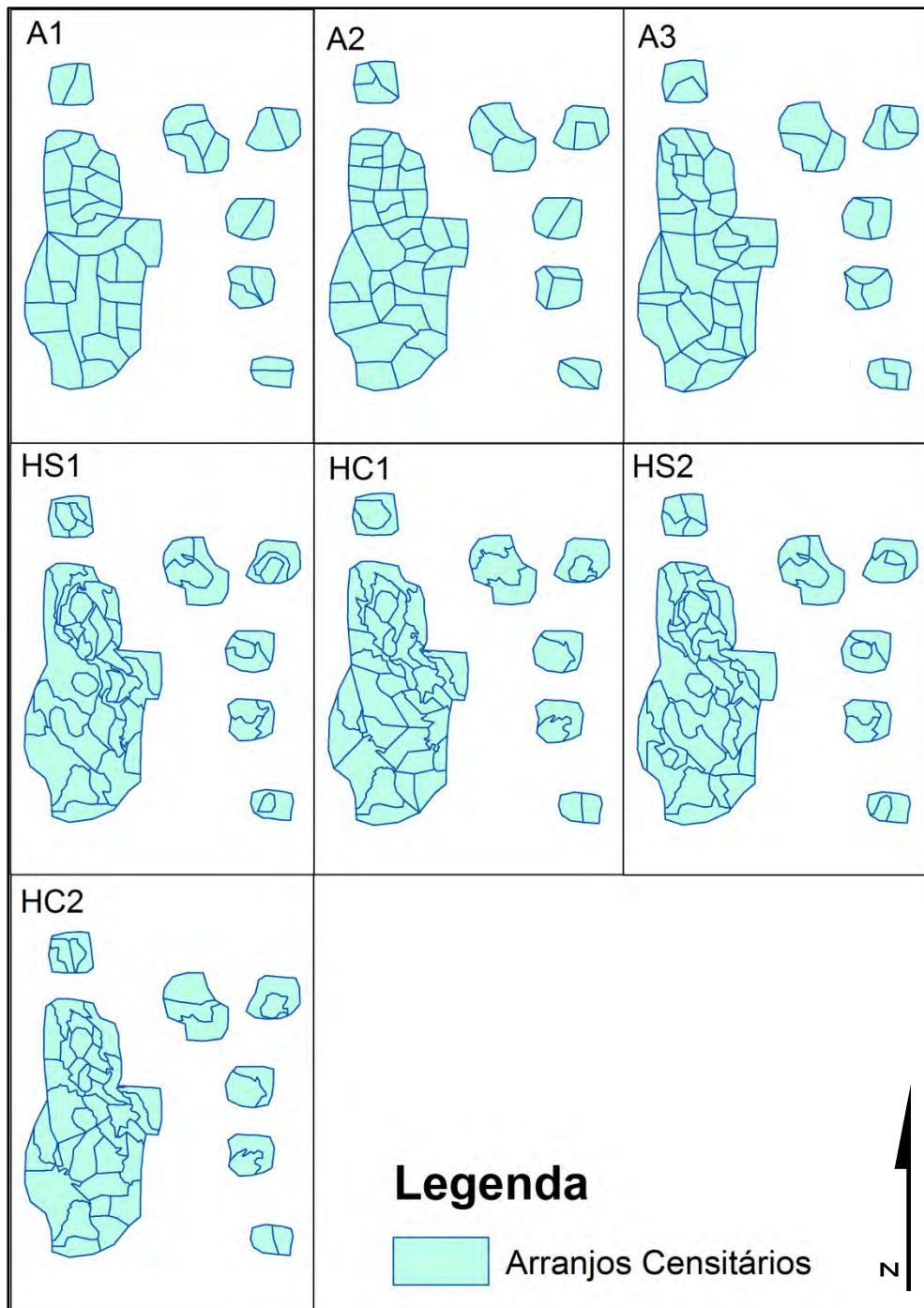


Figura 5.28 Arranjos poligonais dos Setores Censitários

Visualmente, os arranjos aleatórios mantêm um formato mais regular que os homogêneos. Isso pode ser explicado pelo fato de a superfície contínua ser suavizada e os *clusters* de pixels se apresentarem também suavizados. É certo, porém, que, numa situação real, o fato de os setores serem traçados, obedecendo ao desenho regular das quadras, faz com que suas formas também possam ser mais regulares.

Outra situação que permite a suavização das bordas dos setores apresentados neste trabalho é a pouca diferença no tamanho da área dos setores em comparação com as zonas homogêneas geradas, ou seja, a escala de análise é pequena em relação a uma cidade real.

Ainda assim, a abordagem mais simplificada dada ao teste aqui realizado permite maior facilidade de manipulação dos dados, auxiliando as comparações efetuadas entre os diferentes arranjos, bem como a confecção dos próprios limites censitários.

É importante salientar que tal simplificação não torna a análise reducionista, mesmo que possíveis alterações nas fronteiras homogêneas, que venham a ser causadas pela edição final dos setores censitários, agreguem ou desagreguem pontos/residências que alterem significativamente o valor médio e sua variância em uma variável analisada. Também devido a isso, o método rigoroso proposto na confecção dos setores censitários deve ser empregado, o qual permite verificar a presença de tais dados extremos, ou seja, aqueles que diferem do conjunto agregado num setor.

5.3. Avaliação da Homogeneidade Interna dos Setores Propostos

Após a geração dos setores censitários do município fictício, passa-se ao procedimento de agregação dos dados originais. Registra-se que a agregação de dados censitários é uma prática realizada não só pelo IBGE, mas por todos os órgãos de estatísticas censitárias internacionais e seu principal motivo de existência é proporcionar a confidencialidade e a não identificação dos informantes dos dados.

Deste modo, como é a menor unidade geográfica em que tais dados possam ser disponibilizados, é importante que os valores representados por eles sejam os mais fiéis às condições socioeconômicas da população inserida em seus limites. Recordar-se, então, dos efeitos dos Problemas de Unidades de Áreas Modificáveis em dados censitários e sua relação íntima com a Falácia Ecológica.

Quando se faz inferências sobre determinada pessoa, residência, logradouro ou qualquer outra referência geográfica que esteja agregada com outras de mesma categoria, deve-se lembrar que o valor deduzido da variável quantitativa está influenciado pelos efeitos do *MAUP*. Isso é uma relação intrínseca ao fenômeno geográfico agregado por áreas.

Conseqüentemente, qualquer leitura da cidade, presente em processos de diagnósticos populacionais, está propícia a erros. Em contrapartida, Openshaw (1996) cita a necessidade de se minimizar os erros num modelo de zoneamento que represente determinadas variáveis. A proposta é trazer, nesse contexto, a ideia de que as variáveis do tipo média devem ser trabalhadas no sistema de zoneamento censitário.

Minimizar os erros num modelo de dados agregados, os quais sofrem interferências do *MAUP*, seria minimizar a Falácia Ecológica. Se uma indução a uma pessoa ou família será feita a partir da característica de várias pessoas ou famílias residentes num mesmo setor, então que este setor represente pessoas ou famílias de características semelhantes, reduzindo a variabilidade da população.

É certo que tais erros nunca poderão ser eliminados e a inferência sempre terá consigo os efeitos de escala e zoneamento do *MAUP*. No entanto, caso a premissa acima ocorra, as leituras socioeconômicas numa cidade estarão sujeitas a um erro menor.

Como proposta de analisar se houve ou não homogeneidade interna nos arranjos chamados de homogêneos, buscou-se representar as duas variáveis, V e U, e seus respectivos desvios-padrão, pensando na lógica de que, se mais homogêneos são os setores, menor é sua variabilidade e menor é seu desvio-padrão, portanto, a Falácia Ecológica tem menor grau.

As Figuras 5.29 e 5.30 apresentam, nessa ordem, os valores médios das variáveis V e U agregadas por cada setor. Como cada arranjo está utilizando a mesma classificação, pode-se verificar a presença de mais polígonos de cor amarela naqueles arranjos homogêneos. Portanto, houve um “desmascaramento” de áreas com baixos valores, sendo melhor representadas.

Ainda assim, percebe-se que essas diferenças foram poucas e que a tendência de haver valores mais altos continua na parte noroeste do município fictício ou região norte do distrito sede. Isso já era esperado ao se fazer a análise exploratória dos dados. Uma maior concentração de pontos com valores altos se dava exatamente nessa região.

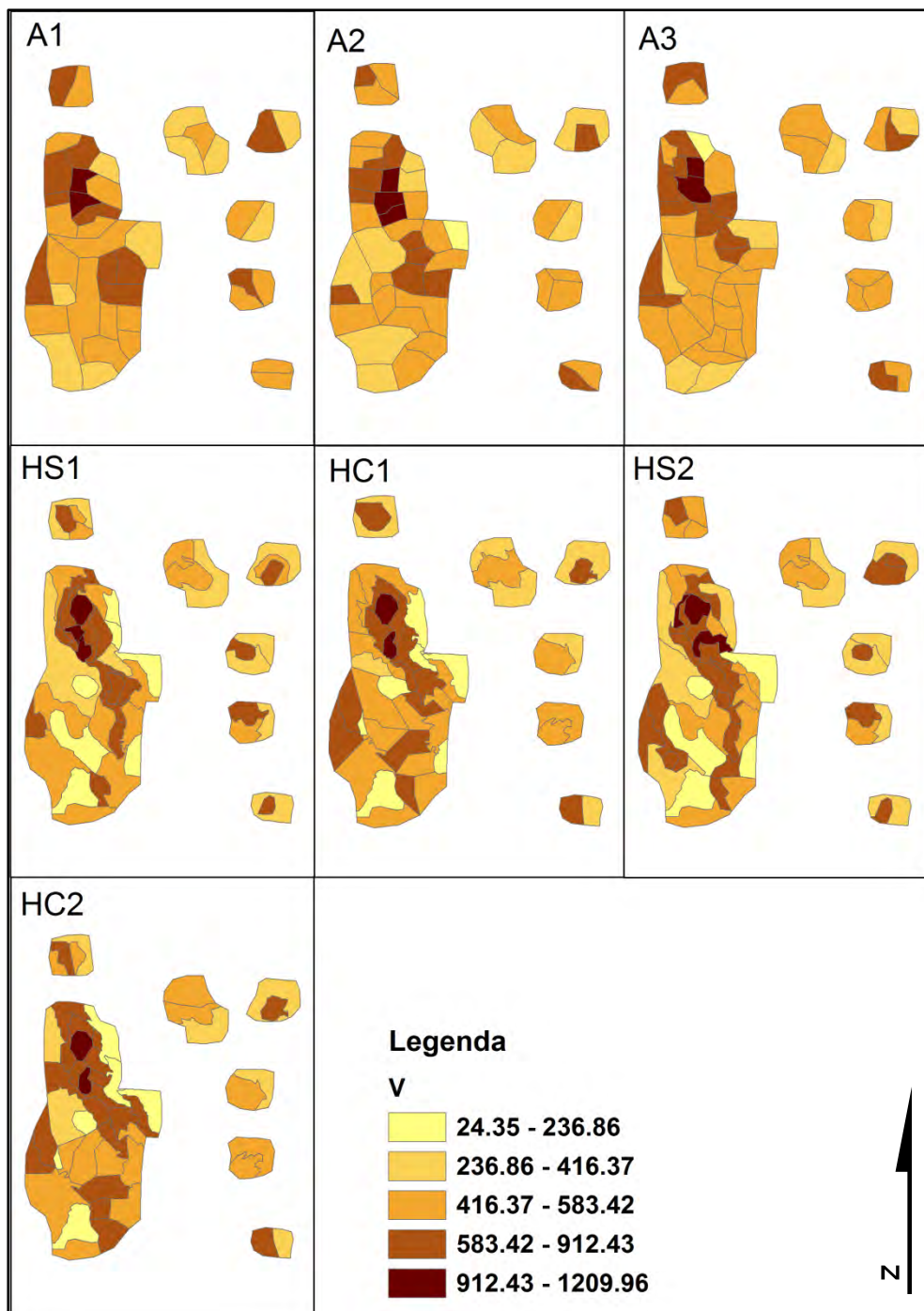


Figura 5.29 Média de V por Setores Censitários

Análise semelhante é feita na Figura 5.30 para a variável U. Houve também uma significativa diminuição de polígonos na segunda classe mais alta de valores.

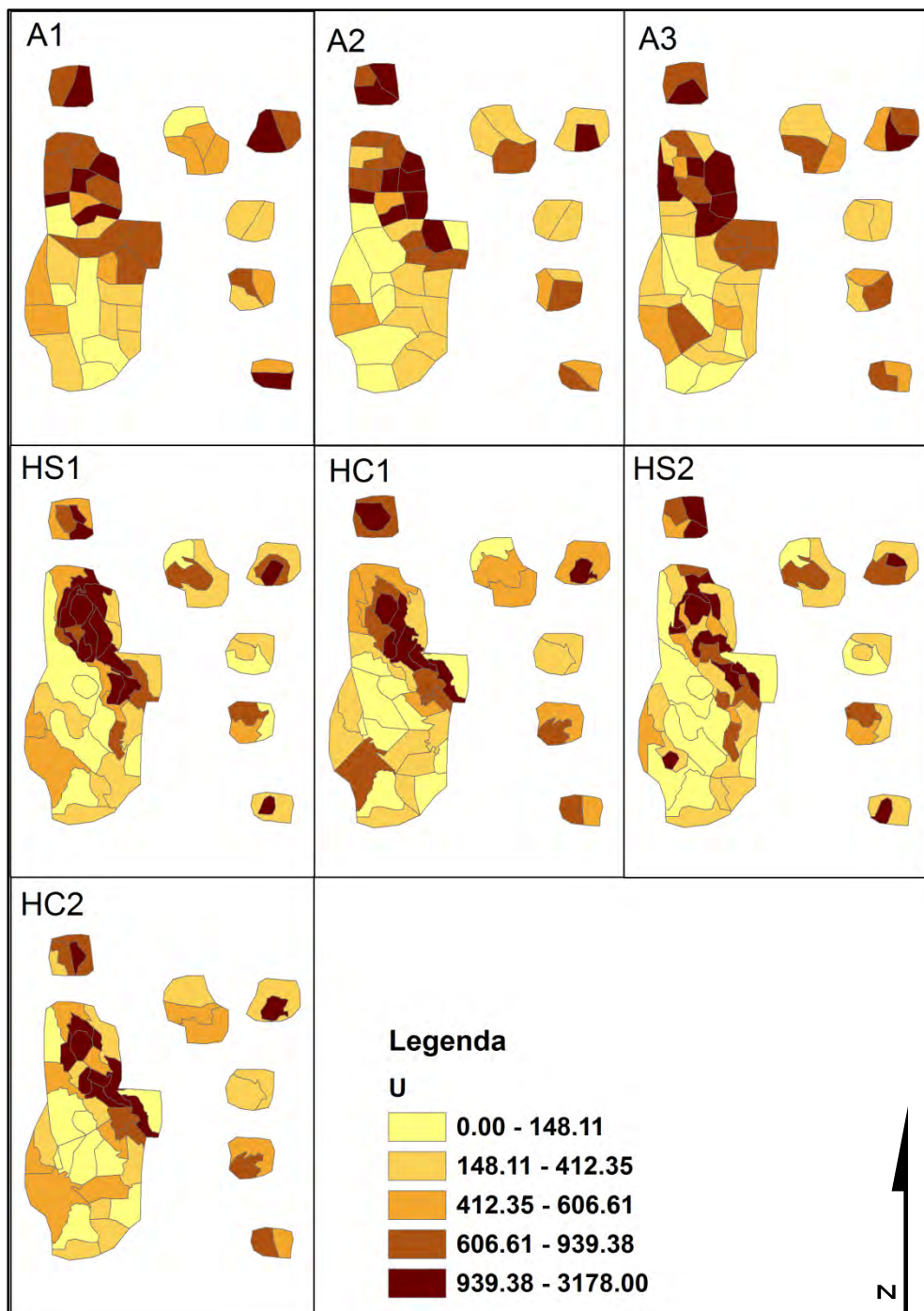


Figura 5.30 Média de U por Setores Censitários

Apesar da melhoria apresentada, nota-se a legitimidade do modelo de agregação proposto a partir da análise visual da variabilidade interna aos polígonos. E é aqui que é validada a hipótese de ser possível reduzir os erros associados à agregação de dados por zonas geográficas. As Figuras 5.31 e 5.32 indicam os desvios-padrão das variáveis V e U, respectivamente.

Nota-se, na Figura 5.31, que os erros, ou melhor, os graus de variabilidade dos dados agregados, foram minimizados à medida que se buscou maior rigor na setorização homogênea, com destaque maior no arranjo HS2. Assim, acredita-se que a detecção de padrões espaciais nos valores indicados nas superfícies contínuas é mais robusta quando é feita individualmente em cada superfície.

Já na Figura 5.32, a variabilidade não mudou tão bruscamente como na análise de V. Isso pode ser explicado pelo fato de que U não possui uma distribuição normal, havendo maior concentração de menores valores, apesar de haver discrepância considerável em seus dados extremos.

Para ambas as figuras, percebe-se que foi impossível excluir setores com as maiores variâncias. Tal fato é baseado na presença de um ou dois pontos em uma zona homogênea, devendo ser agregado a esses pelo menos um ponto de valor não tão semelhante.

É certo que, numa situação real, isso deve acontecer, pois o método rigoroso de agregação é limitado quando há menos pontos do que o limite inferior. Dessa maneira, sempre haverá a possibilidade de se encontrar polígonos com grande variabilidade.

No entanto, este trabalho agregou no mínimo três pontos em cada polígono, enquanto o IBGE agrega no mínimo 250 residências. Seguindo o raciocínio, a variância não é tão sensível a valores extremos como a média. Por conseguinte, a presença de poucas residências discrepantes num setor não aumentaria radicalmente o valor de variabilidade interna ao setor.

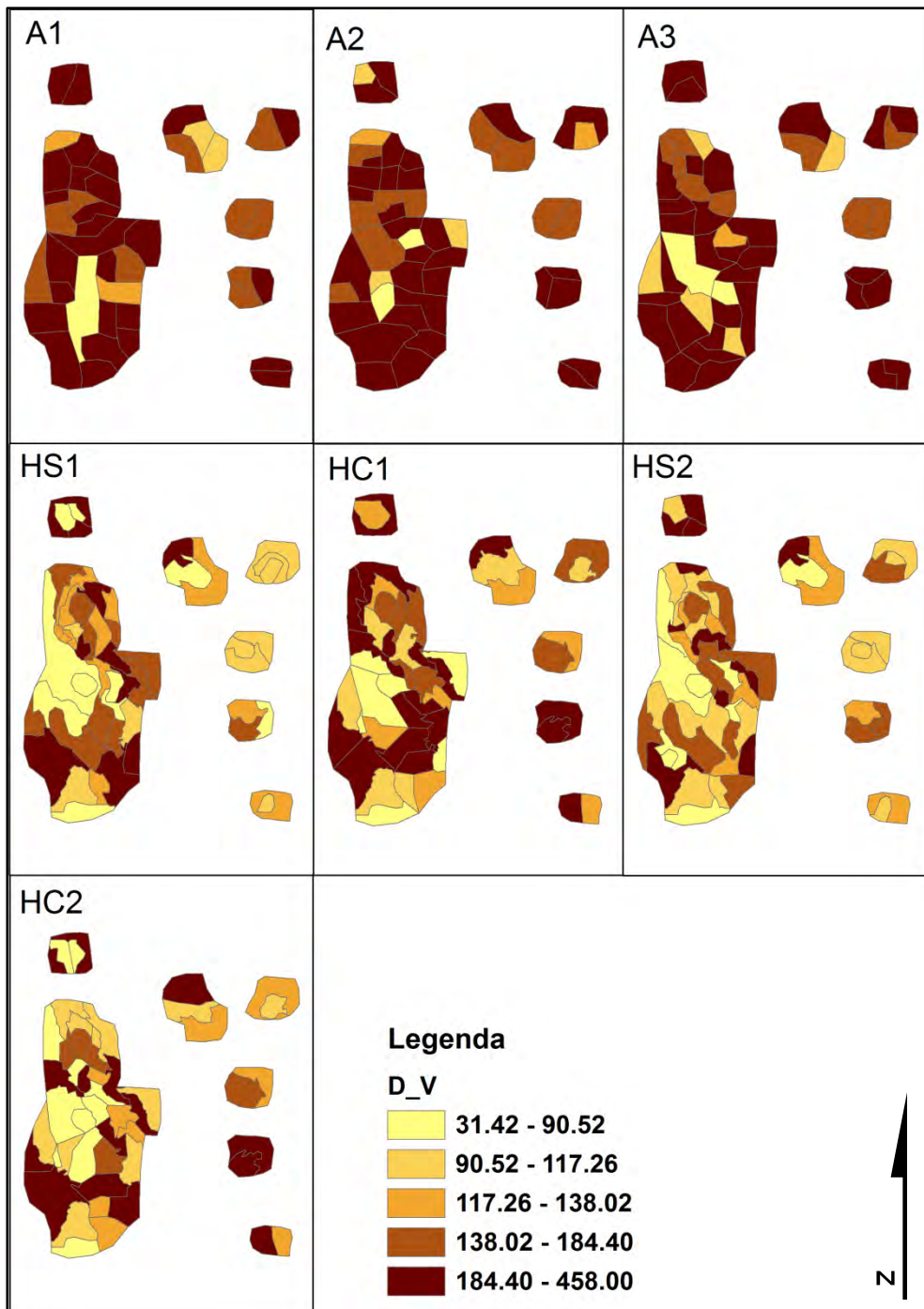


Figura 5.31 Desvio-Padrão de V por Setores Censitários

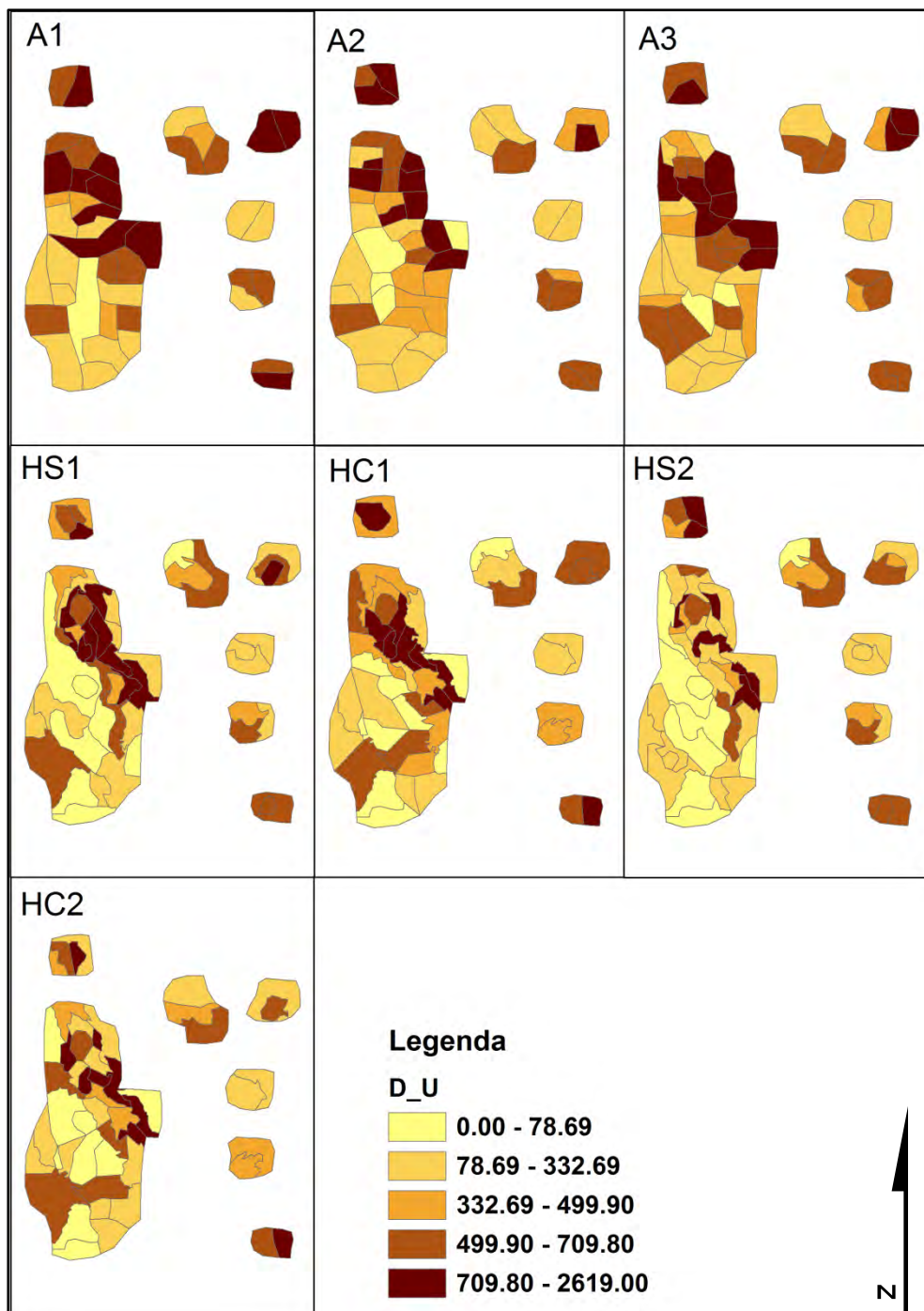


Figura 5.32 Desvio-Padrão de U por Setores Censitários

A análise visual auxilia na percepção do grau de homogeneidade interna aos setores censitários. Percebe-se que há uma tendência de valores mais baixos de variância (ou desvio-padrão, como representado nas figuras). Como forma complementar de avaliação, buscou-se, com a análise de autocorrelação espacial, uma medida quantitativa de presença de baixas variâncias.

Com a *General G Statistics*, foi verificado se havia existência de agrupamentos de setores com baixas variâncias. No entanto, na Tabela 5.3, percebe-se que a maioria dos arranjos não possuía nenhum tipo de agrupamento. Naqueles em que foi percebida alguma dependência espacial, esta era de polígonos com alta variabilidade.

Tabela 5.3 – Avaliação de agrupamentos de setores com baixas variâncias

<i>Índice General G Statistics</i>		
	D_U	D_V
A1	Aleatório	Aleatório
A2	Aleatório	Aleatório
A3	Alto	Aleatório
HS1	Alto	Aleatório
HC1	Alto	Aleatório
HS2	Aleatório	Aleatório
HC2	Aleatório	Aleatório

Buscando-se uma perspectiva diferente da global, o Índice Local de Moran foi aplicado às mesmas variáveis. A expectativa era a de se encontrar maior quantidade de polígonos com baixas variâncias autocorrelacionadas positivamente com vizinhos de baixos valores de variâncias também. Nas Figuras 5.33 e 5.34, estes polígonos estão representados na cor verde escura, para V e U, respectivamente.

A tonalidade mais clara representa polígonos com baixa variabilidade e vizinhos de polígonos com altos desvios em torno da média. Por sua vez, as cores vermelhas indicam polígonos de altas variâncias com vizinhança parecida. Finalizando, a cor laranja indica alta variabilidade com vizinhos de baixas variabilidades. Assim, o ideal é encontrar maior concentração de cores verdes e menor de cor vermelha.

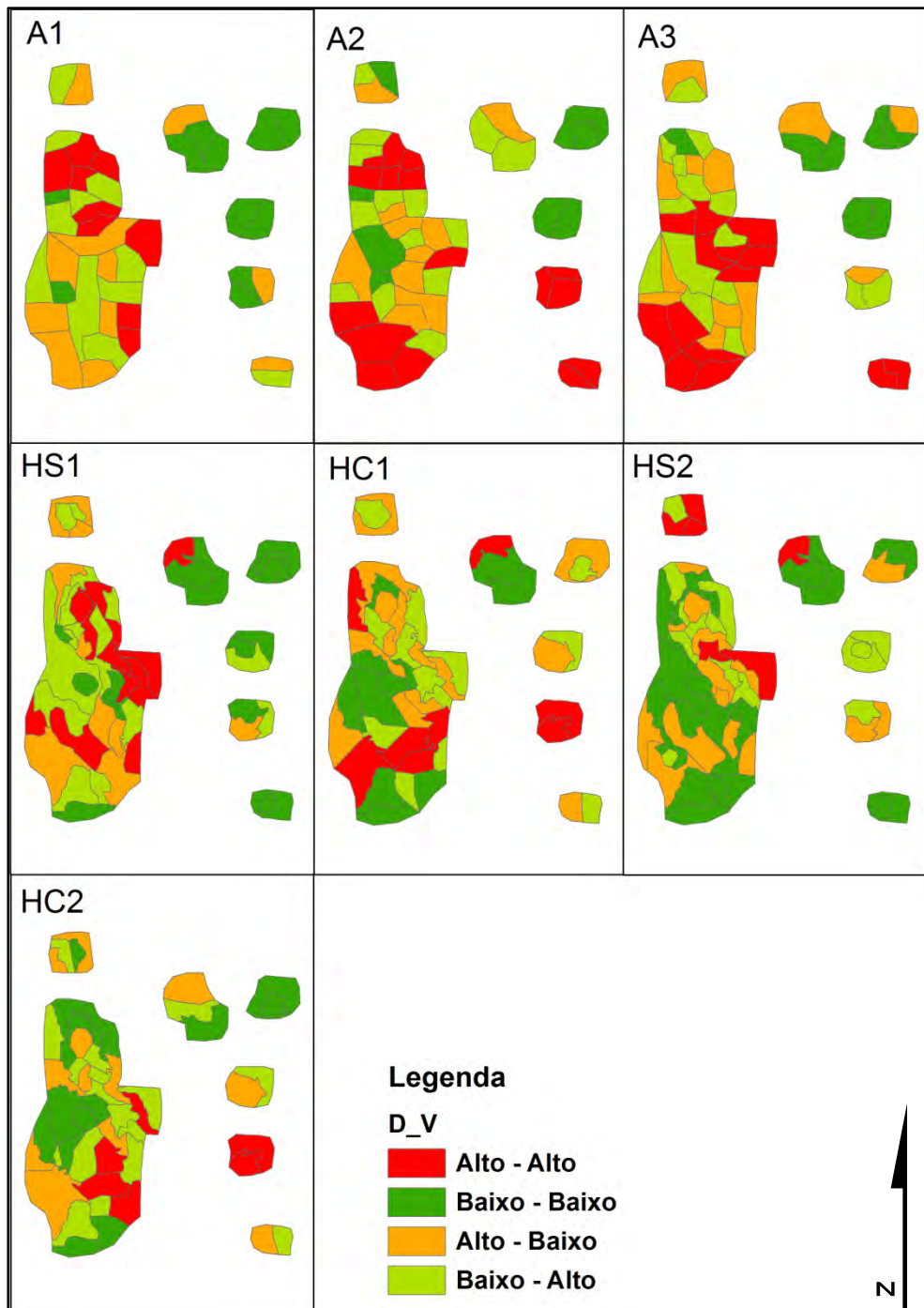


Figura 5.33 Autocorrelação Local do Desvio-Padrão de V

Nota-se que nos arranjos aleatórios há menor concentração de verde escuro, enquanto os arranjos homogêneos apresentam menos padrões altos. Tal análise parece contraditória com os resultados da *General G Statistics*, no entanto, a maioria dos valores de autocorrelação apresentados pelo indicador de Moran não possuíam significância estatística. Assim, apesar de estatisticamente as análises de autocorrelação

local serem falhas, percebe-se uma leve tendência de haver maior concentração de baixos valores de variância.

A mesma análise crítica é feita para a Figura 5.34, apresentando os desvios de U. Porém, é ressaltado que a quantidade de polígonos verdes escuros não variou consideravelmente, como o aumento de verdes claros. Já a redução de polígonos vermelhos foi mais intensa para os arranjos homogêneos rigorosos.

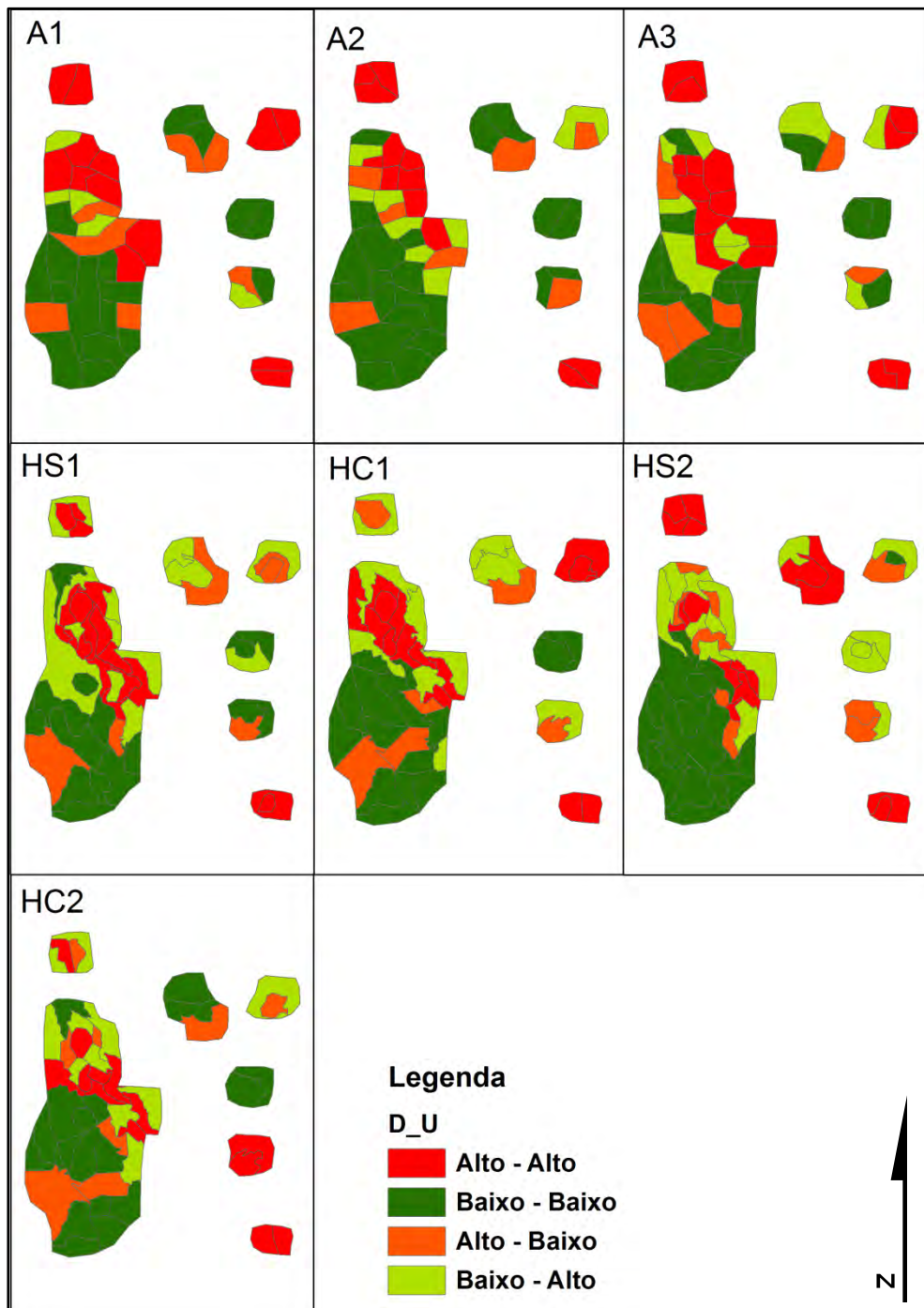


Figura 5.34 Autocorrelação Local do Desvio-Padrão de U

Para essa análise, é importante destacar que a autocorrelação local é baseada no valor médio de cada arranjo poligonal, não considerando a média de todos os arranjos ao mesmo tempo. Assim, a detecção de polígonos com baixas variâncias se dá para cada arranjo, não sendo possível haver comparação quantitativa entre os arranjos, mas sim apenas se há maior presença de arranjos locais baixos.

Quanto à minimização da correlação interáreas, a utilização do Índice Global de Moran mostrou que os arranjos possuíam padrões agrupados para as variáveis V e U. A presença de arranjos aleatórios é praticamente insignificante. Assim, nota-se que as setorizações não foram capazes de minimizar a correlação entre os setores. A Tabela 5.4 mostra seus resultados.

Tabela 5.4 – Avaliação de agrupamentos globais de setores

Índice Global de Moran		
	U	V
A1	Agrupado	Aleatório
A2	Agrupado	Agrupado
A3	Agrupado	Aleatório
HS1	Agrupado	Agrupado
HC1	Agrupado	Aleatório
HS2	Aleatório	Agrupado
HC2	Agrupado	Aleatório

5.4. Proposta de integração das regiões socioeconômicas e Setores Censitários ao Cadastro Territorial Multifinalitário

Dando continuidade ao estudo, apresenta-se aqui a proposta de integração dos setores censitários ao CTM, compatibilizando os limites dos mesmos aos limites das parcelas instituídas nas Diretrizes para o Cadastro. Este tema está numa seção diferente daquela em que se aborda a geração de setores homogêneos, porque as duas proposições podem ser consideradas independentes, apesar de ser sugerido como ideal trabalhar com setores homogêneos amarrados ao CTM.

Portanto, mesmo que não haja a geração das zonas homogêneas, seja devido a limitações de um município em sua criação ou por algum outro fator restritivo, pondera-se sobre a componente cadastral dos setores censitários. Portanto, a partir das vantagens e desvantagens de sua utilização para a caracterização populacional, desconsiderando a presença do *MAUP* em dados agregados, o que já foi avaliado e discutido anteriormente, apresenta-se a proposta de integração.

5.4.1. Integração dos Setores Censitários ao CTM

Fica clara a necessidade de se delimitar os setores censitários a partir de aspectos físicos identificáveis no campo. Em outros casos, a delimitação por eixo de logradouro garante a fixação de um limite censitário numa base cartográfica, respeitando-se o desenho urbano.

A proposta de integração dos setores ao cadastro sugere a amarração conceitual do limite censitário às parcelas cadastrais. Nesse caso, o limite passaria a ser definido pelo cadastro, e não mais por aspectos físicos. As atualizações dos limites censitários seriam automáticas numa modelagem de banco de dados, por exemplo.

Assim, o setor censitário englobaria as residências cadastradas nos logradouros pertencentes a esse setor, evitando preocupações com limites físicos. Definem-se, portanto, os *Limites Cadastrais*, que seriam atualizados assim que o cadastro imobiliário fosse atualizado. O gerenciamento do número de residências em cada setor também seria facilitado pelo próprio CTM.

Outra vantagem dessa integração é que os setores censitários estariam amarrados à rede de referência cadastral, garantindo o seu georreferenciamento, permitindo o

cruzamento preciso com outros planos de informações ou camadas. Tal fato contrasta com a estruturação atual, quando um limite censitário é uma linha de cumeada ou outra feição, da qual não haja representação ou esta esteja fora dos padrões cartográficos adotados na planta cadastral municipal. Isso pode acarretar em distorções nas análises espaciais, já que haverá certa subjetividade na representação deste setor.

Por outro lado, na situação proposta aqui, se todo o território deve ser cadastrado em parcelas, conforme sugerido nas Diretrizes para o Cadastro, e se o setor censitário for delimitado por parcelas, então possíveis distorções em análises espaciais causadas por cruzamento de dados em SIG seriam minimizadas ou até mesmo inexistentes. Certamente que não entra nessa análise a questão cadastral rural, uma vez que a definição dos limites dos setores censitários rurais obedecem apenas aos limites físicos e administrativos e a definição das parcelas rurais cadastrais nem sempre segue os limites administrativos.

Quanto à geração dos limites censitários e sua gestão, Costa *et al* (2005) mostram que o artigo 182 da Constituição Federal de 1988 e o princípio da preponderância do interesse comum tornam o município o principal ente federativo responsável pelo gerenciamento do seu território, garantindo o bem-estar de sua população e o cumprimento da função social propriedade urbana. Baseado nisso, o município se torna o agente mais apto a identificar suas zonas homogêneas, criar setores censitários que representem características socioeconômicas semelhantes e gerenciar os limites desses setores.

Este trabalho, assim sendo, defende uma parceria formal entre o IBGE e as prefeituras para a gestão dos setores censitários urbanos, uma vez que cada município tem responsabilidades primárias no gerenciamento do seu território.

O trabalho do IBGE poderia ser facilitado no que diz respeito à aquisição das bases informacionais municipais para a delimitação dos setores censitários. O órgão possui sedes regionais, as quais se responsabilizam pela aquisição e estruturação de bases cartográficas de um conjunto de municípios vizinhos às sedes.

Portanto, com a proposta da gestão dos limites dos setores censitários por município, partindo-se das Diretrizes para o Cadastro, o IBGE, a partir de seus escritórios regionais, participaria do processo de validação das bases cartográficas censitárias, minimizando sua participação na criação dessas bases.

Não se defende aqui a realização dos censos pelos municípios. O IBGE ainda

deveria ter o controle de todo o censo, garantindo, assim a sistematização do levantamento censitário e a confidencialidade das informações censitárias. Também, como os equipamentos utilizados pelo órgão censitário permitem a integração de bases de dados geoespaciais, todos os setores censitários provenientes da base municipal seriam associados aos bancos de dados espaciais utilizados para a coleta censitária.

Quanto ao censo rural, é necessário pensar numa parceria Prefeitura-IBGE-INCRA. O Cadastro Nacional de Imóveis Rurais ainda não está implementado, no entanto, seguindo-se o seu modelo conceitual, através do qual todos os estabelecimentos rurais estariam representados cartograficamente e amarrados ao SGB, a definição dos limites censitários rurais poderia, do mesmo modo, estar ligada ao cadastro.

É certo também que as estruturas cadastrais municipais (urbano e rural) ainda não estão unificadas. Aliado a isso, as diferenças constantes nos diversos temas impedem a unificação dos dois sistemas numa mesma estrutura de banco de dados. Porém, cartograficamente é possível realizar a integração cadastro-censo, já que as unidades cadastrais censitárias poderiam ser os limites dos imóveis rurais.

Em seguida, destaca-se que essa integração não visa substituir os cadastros socioeconômicos em nível de parcelas que já existam num município. Cadastros educacionais, de saúde, como citados anteriormente, ou de outras áreas de interesse são mais que necessários na gestão e planejamento territoriais. Os dados censitários apenas auxiliariam esses processos na prefeitura e órgãos de planejamento, na falta de informações cadastrais parcelares.

Além disso, dependendo da informação cadastrada por órgãos da prefeitura, não se pode torná-la pública devido à segurança da informação e preservação do informante. Caso os setores censitários sejam os representantes de territórios intraurbanos homogêneos do ponto de vista socioeconômico e estejam sistematicamente amarrados ao CTM, as informações acima descritas podem ser agregadas em nível desses setores, de modo que possam ser publicadas, garantindo a confidencialidade dos indivíduos que as disponibilizaram. São citados como exemplo informações como perfis de consumidores ou alguns casos de doenças que causem constrangimento ao paciente.

A seguir, são dados alguns detalhes quanto à compatibilização dos limites censitários aos limites parcelares, sendo mostrados exemplos ilustrativos.

5.4.1.1. Compatibilização dos Limites Censitários aos Parcelares

Inicialmente, pondera-se sobre as diferenças que existem entre as parcelas cadastrais, os setores censitários e os estabelecimentos rurais. Pensando-se em unidades de coleta de dados, as parcelas servem de referência para a obtenção de dados cadastrais. Os setores censitários, tratados nesta proposta como delimitados pelas parcelas, são unidades de coleta de dados censitários. A principal relação existente é o relacionamento geográfico (topológico) entre eles. Obviamente que, num cruzamento de dados num banco de dados geográficos, novas informações podem ser inferidas ao se associar as duas fontes de dados.

O objetivo do censo é proporcionar a leitura das condições de um conjunto de domicílios, os quais são associados às parcelas cadastrais. A mesma afirmação é feita aos estabelecimentos rurais quanto unidades de coleta cadastrais em comparação com os setores censitários.

Independentemente dessa reflexão, uma das peculiaridades, e um grande problema, do sistema de setorização censitária atual é a subjetividade como se determinam e representam os limites dos setores. Como citado anteriormente, os setores censitários “têm limites físicos identificáveis em campo” (IBGE, 2003). Na maioria dos casos, na área urbana, esses limites são os eixos dos logradouros, os quais são perfeitamente identificáveis.

No entanto, algumas feições definidoras de fronteiras entre setores são de difícil localização, trazendo certo grau de complexidade na assimilação dos limites em campo. Algumas vezes são citadas nascentes, curvas de nível ou até residências de proprietários de terra como referências locais para a demarcação dos perímetros censitários.

Os trechos abaixo foram retirados de alguns limites censitários de certos municípios de Minas Gerais, para a base de dados do Censo 2010 (IBGE, 2010), observando-se que aqui se replicam fidedignamente as palavras e termos utilizados em tais descrições, grifando-se palavras que indiquem maior subjetividade:

- “PONTO CULMINANTE DE UM MORRO COM COTA SUPERIOR A 980 M, NO LIMITE DA "RESERVA BIOLÓGICA SANTA CANDIDA”. Ponto inicial do setor censitário número 313670205090012 da cidade de Juiz de Fora.

- “SEGUE PELA RUA FRANKLIN DE PAULA MARQUES, RUA CIDADE DO SOL, RUA NEGRAO DE LIMA, RUA JOSE TEIXEIRA DA SILVA, RUA GERALDO GERHEIN, RUA SILVA MELLO ATE SEU ENCONTRO COM A RUA CIDADE DO SOL, DESTE ENCONTRO EM RETA ATE ATINGIR UMA LINHA DE VALETA, SEGUE ACOMPANHANDO ESTA LINHA DE VALETA ATE ATINGIR SEU FINAL NO LADO DIREITO DE UMA ESTRADA CONHECIDA COMO ESTRADA DO MOTEL, DESTE PONTO SEGUE EM LINHA RETA ATE ENCONTRAR UM PONTO SOBRE O ESPIGAO COM COTA APROXIMADAMENTE 800M DE ALTITUDE (...)”. Trecho de um limite do setor número 313670205090015, também de Juiz de Fora.
- “DO PONTO INICIAL, SEGUE PELO "LEITO DO RIO PARAIBUNA" ATE UM PONTO FRONTEIRO A ULTIMA CASA DA "RUA DR SAULO VILELA" (CASA ESTA EXCLUSIVE). DAI, SEGUE EM LINHA RETA, ATRAVESSANDO O "LEITO DA RFFSA", A "RUA DR SAULO VILELA" ATE ATINGIR, NO BARRANCO, UMA CERCA DE ARAME. SEGUE POR ESTA CERCA DE ARAME ATE ATINGIR UMA VALETA”. Trecho do limite do setor número 313670205180041, ainda da Cidade de Juiz de Fora.
- “(...) POR ESTA ATE A RUA MANOEL DE SOUZA BRASIL, POR ESTA, SENTIDO PARQUE DA LAPA GRANDE ATE ATINGIR A COTA DE ALTITUDE 700, DAI, SAINDO DA ESTRADA, RETORNA A DIREITA, MARGENANDO OS MORROS, CONTINUANDO PELA LINHA DA COTA DE ALTITUDE 700 ATE ATINGIR O MARCO 4 CRAVADO A MARGEM ESQUERDA DA BR-135 (...)”. Trecho do limite do setor censitário número 314330205000101, no Município de Montes Claros.
- “DO PONTO INICIAL CONTORNANDO A ANTIGA RESIDENCIA DE JOSE BASILIO (INCLUSIVE) EM DIVISA COM O SÍTIO SANTO ANTONIO (EXCLUSIVE) ALCANCA A ESTRADA PARA CASCALHO PELA QUAL SEGUE EM DIRECAO AO CENTRO ATE

UM VALO EM SUA MARGEM ESQUERDA (...)”. Trecho do limite do setor censitário número 317130325000001, no Município de Viçosa.

Além da subjetividade dos pontos de referência, percebe-se que o método utilizado para o memorial descritivo dos limites censitários é completamente obsoleto, não sendo prático, principalmente quando se compara com o avanço das tecnologias empregadas em localização de feições. O padrão de memorial descritivo que poderia ser utilizado deveria se basear nas coordenadas dos vértices, sendo mais atual e moderno.

Assim, como se propõe a amarração dos limites censitários aos parcelares, a descrição dos mesmos pode ser dada a partir da descrição colocada às parcelas. Como todas essas têm seus vértices georreferenciados ao Sistema Geodésico Brasileiro, automaticamente, os limites censitários também estarão.

A amarração dos setores às parcelas implica numa mudança de paradigmas e também numa pequena mudança estrutural. Os limites deixariam de ser os eixos de logradouros, passando a ser os limites das parcelas.

Com essa proposta, uma pequena alteração na definição dos setores censitários é sugerida, baseada no novo significado dado aos mesmos. Na busca de se definir uma ontologia para os novos setores censitários, percebeu-se que o conceito dado pelo IBGE se modificaria com a nova proposta.

A primeira definição dada por IBGE (2003) é “O setor censitário é a menor unidade territorial, com limites físicos identificáveis em campo, com dimensão adequada à operação de pesquisas e cujo conjunto esgota a totalidade do Território Nacional, o que permite assegurar a plena cobertura do País”, com grifo do autor. Com a proposta, o grifo seria trocado por “limitados pelas parcelas cadastrais”.

O respeito aos limites das divisões político-administrativas do quadro urbano e rural legal e de outras estruturas territoriais de interesse continuaria, pois é a partir das parcelas que se definem tais limites, conforme é colocado no artigo 3º das Diretrizes para o Cadastro.

Considerando uma segunda definição, presente tanto em IBGE (2003), quanto em IBGE (2011), um setor censitário é “a unidade territorial criada para fins de controle cadastral da coleta”. Tal significado garante a total cobertura do país, pois os limites censitários foram concebidos inicialmente para fins de coleta de dados.

Uma nova aceção é trazida aqui considerando não apenas os resultados da avaliação dos efeitos de zoneamento apresentados anteriormente, mas, principalmente, as aplicações realizadas com os dados agregados por setores censitários. Amplo é o conjunto de obras, acadêmicas ou não, que se utiliza de tais dados.

Portanto, como fonte mais desagregada geograficamente, o setor censitário também pode ser conceituado não só como a unidade territorial criada para fins de controle cadastral de coleta, mas também de disponibilização de dados censitários.

Com base nas definições propostas, tem-se então que *o setor censitário é a menor unidade territorial, limitado pelas parcelas cadastrais, com dimensão adequada à operação de pesquisas e cujo conjunto esgota a totalidade do Território Nacional, o que permite assegurar a plena cobertura do País, sendo, portanto, a unidade territorial criada para fins de controle cadastral da coleta e disponibilização de dados censitários.*

Ampliando a definição, o setor censitário é uma agregação espacial das parcelas cadastrais, bem como uma região homogênea socioeconomicamente é uma agregação espacial do tipo *união espacial* de setores censitários. A Figura 5.35 apresenta uma proposta de um diagrama de classes, no modelo OMT-G, para a modelagem conceitual das três entidades num Banco de Dados Geográficos.

Ao exposto no parágrafo anterior, acrescenta-se que os dados censitários são associados aos setores censitários, sendo uma relação de um setor para um registro da tabela de dados censitários (relação 1 para 1). Tal associação já ocorre na conformação atual do censo, não sendo alterada, portanto. Adicionalmente, às parcelas também são associados os dados pertinentes, também na relação 1 para 1, lembrando-se que não necessariamente uma parcela representa um lote urbano, como definem as Diretrizes para o Cadastro.

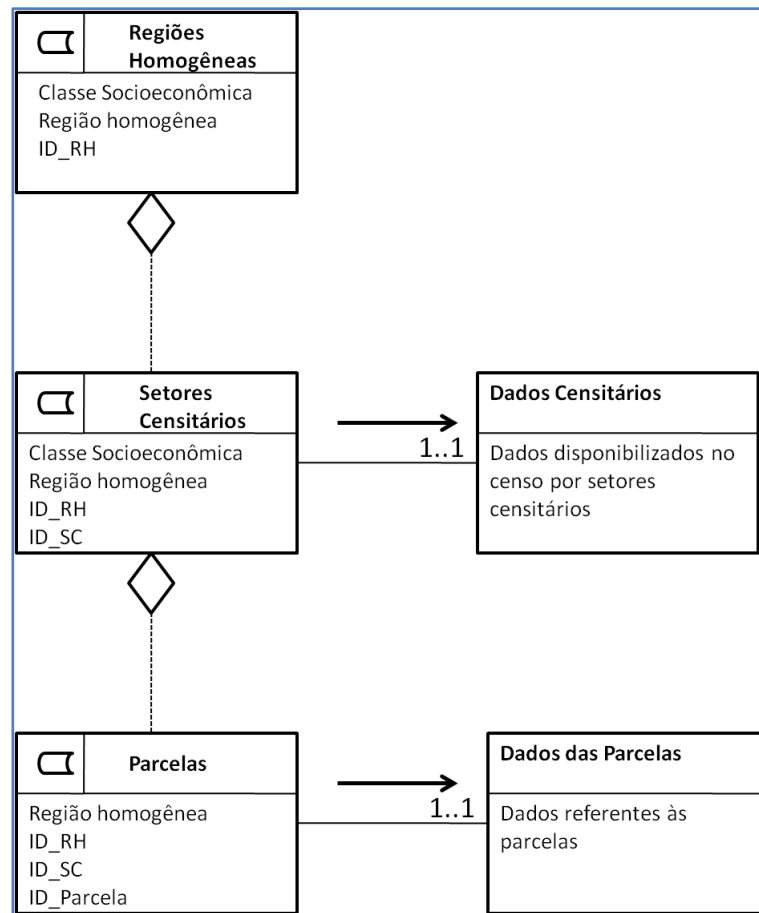


Figura 5.35 Modelagem Conceitual dos Setores Censitários

Como regras de restrição topológica para a definição dos novos setores censitários, enumeram-se aqui aquelas mostradas em Borges (1997), já presentes nas literaturas envolvendo o modelo OMT-G, quais sejam:

- A origem de um objeto derivado depende da união de pelo menos dois objetos disjuntos pertencentes à classe primitiva.
- O limite geográfico do objeto derivado deve coincidir com o limite geográfico externo formado pela união da geometria dos objetos pertencentes à classe primitiva, não podendo extrapolá-lo.
- A alteração do limite geográfico do objeto derivado só poderá ser feita através da alteração dos limites dos objetos primitivos.
- A exclusão de um dos objetos primitivos implica na alteração do limite do objeto derivado.
- A exclusão de todos os objetos primitivos que originaram o objeto derivado implicará na exclusão do objeto derivado.

Destaca-se que o objeto derivado ora pode ser a região homogênea, ora pode ser o setor censitário, enquanto o objeto primitivo (classe primitiva) ora pode ser o setor censitário, ora pode ser a parcela, respectivamente.

A partir dessas restrições topológicas, é fácil definir mecanismos automáticos de geração de memoriais descritivos para os setores censitários, quando e se necessário. Obviamente que, como os setores já existentes num banco de dados do CTM, a recuperação de seus limites por parte de usuários comuns seria simplificada, bastando apenas a consulta ao órgão responsável pelo CTM municipal.

As atualizações dos limites dos setores censitários já seriam feitas a partir das atualizações cadastrais, obtendo-se com certa facilidade o número de residências presentes em cada setor para a verificação da necessidade de haver desmembramento do mesmo.

Além da restrição topológica, propõem-se também as regras de restrição de vizinhança, as quais se produziram a partir do estudo da importância das relações de vizinhança nas configurações dos territórios, tanto no interrelacionamento sócio-pessoal, quanto em termos de distribuição espacial de infraestrutura urbana.

Tais restrições se aplicam também à configuração geométrica dos setores censitários, porém, abordando na forma como as parcelas se agregam. Atualmente, os setores são delimitados por eixos de logradouros, como pôde ser visto anteriormente. Isso implica em dizer que possíveis condições socioeconômicas semelhantes, encontradas num mesmo logradouro, são representadas em setores diferentes.

Com o objetivo de se ter setores os mais homogêneos socioeconomicamente possível, separar vizinhos frontais num logradouro significa ignorar uma possível homogeneidade. O estudo apresentado por Portugal (2009), aliado ao de Cupertino *et al* (2009) permite inferir que as principais relações sociais se dão entre pessoas de mesma classe social, o que muitas vezes está vinculado ao local de moradia dessas pessoas, dividindo, inclusive, os mesmos serviços urbanos prestados pela prefeitura.

Assim, propõe-se que os Limites Cadastrais, outrora definidos, englobem as parcelas num mesmo logradouro, não o dividindo ao meio, buscando evitar a separação de um mesmo território socioeconômico: o logradouro. Uma exceção é quando esse logradouro representa limite entre classe sociais. Obviamente que as zonas homogêneas não se conformam como a geometria regular das parcelas cadastrais. Todavia, como

forma de agregar essas zonas ao CTM, é proposta a compatibilização de seus limites às parcelas.

As regras de restrição de vizinhança se baseiam em três preceitos:

- Utilizar o conceito de dependência espacial na definição dos limites das zonas homogêneas, nas quais se baseiam os limites censitários.
- Considerar a homogeneidade interna a partir dos aspectos de homogeneidade socioeconômica. Não se deve desconsiderar a vizinhança de frente nos logradouros, desde que estes não representem uma heterogeneidade.
- Utilizar os limites das parcelas cadastrais, os quais servem de base para a definição dos limites administrativos (municipais ou distritais), garantindo, assim, que os setores censitários obedeam tais limites.

Portanto, as definições dos limites censitários e, por consequência, das zonas homogêneas devem adotar as seguintes considerações:

- Os limites das zonas homogêneas apenas servem de base para a identificação dessas zonas. Como são provenientes da segmentação de uma superfície contínua, advinda de uma interpolação, não se pode afirmar que são linhas verdadeiras, principalmente pela consideração de certo grau de subjetividade tanto na geração do indicador socioeconômico, quanto na obtenção de seus valores na própria interpolação;
- Sempre que possível, deve-se assumir que as feições que limitem fisicamente o convívio diário entre residentes num local sejam também fronteiras socioeconômicas, mesmo que a detecção de padrões homogêneos não tenha gerado ali um limite. Como exemplo, pode-se ter um curso d'água, uma gleba que constitua um vazio urbano ou mesmo um divisor de águas, caso este não possua ocupação antrópica (indo ao encontro com utilização das bacias hidrográficas como unidades de planejamento físico-territorial e socioeconômico);
- Deve-se evitar a utilização de um logradouro como divisa censitária, excetuando-se quando estes naturalmente repartem territórios, sendo importante não utilizar seu eixo central, por ir de encontro à utilização das parcelas para a caracterização de feições cadastrais;

- É permitida a divisão transversal de um logradouro, quando em obediência ao limite máximo de residências num setor censitário, no entanto buscando-se sempre criar essa divisa em locais que minimizem a separação de um território socioeconômico.

Tal proposta não se restringe a apenas essas regras, podendo a elas serem agregadas outras baseadas em situações que por ventura não foram concebidas neste trabalho. A seguir, são feitas algumas simulações com o intuito de ilustrar as condições acima apresentadas.

A Figura 5.36 apresenta as quadras, os lotes, o curso d'água e as fronteiras homogêneas passando pela região, ainda sem edição, uma vez que deve ser modificada em função da geometria das parcelas. É importante frisar que essas fronteiras homogêneas são simuladas.

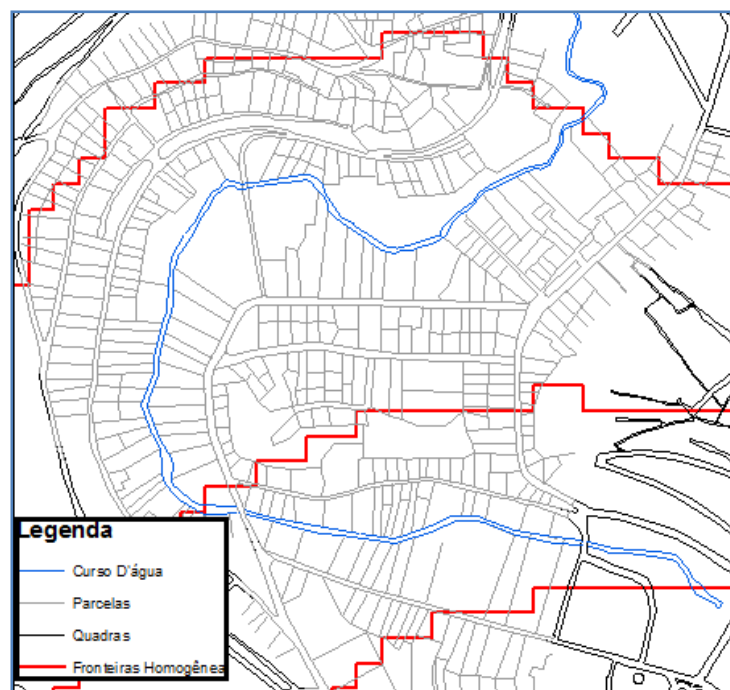


Figura 5.36 Situação cadastral sem setores censitários

Com base na fronteira socioeconômica apresentada e nos limites das parcelas, buscam-se conformar os limites censitários com esses, evitando-se a geração de uma divisa no meio de um logradouro, conforme mostra a Figura 5.37.

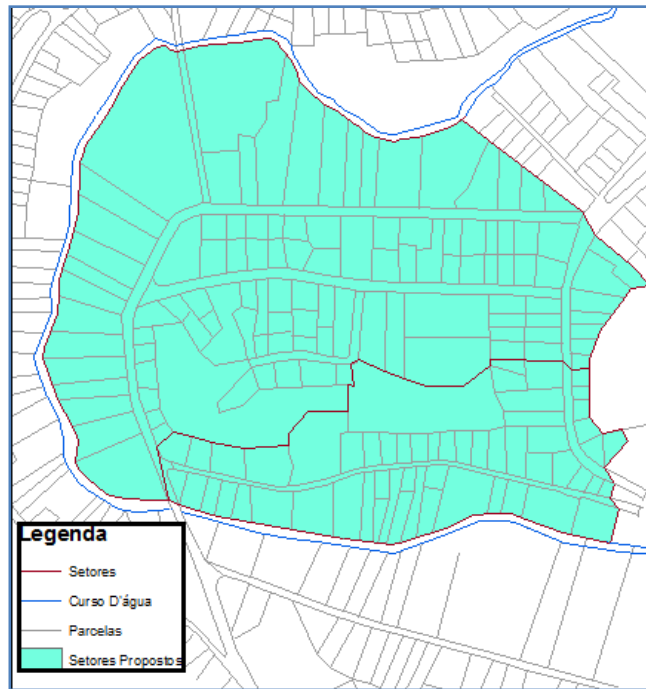


Figura 5.37 Setores censitários homogêneos

Como visto na modelagem de dados, a fronteira homogênea é obtida a partir dos limites censitários, os quais se basearam nos limites das parcelas, como pode ser visto na Figura 5.38.

Essa fronteira foi concebida originalmente a partir da transformação dos polígonos segmentados para linhas, por isso, sua conformação com “dentes de serra”. Já foi visto que as fronteiras não coincidem com os limites das parcelas, e por isso é buscada uma conformação com tais limites.

Contudo, em alguns casos, a fronteira homogênea pode ficar em um local que algumas parcelas ficam de fora da área homogênea, sendo, porém, pertencente a um logradouro presente nessa mesma área. Isso ocorre devido ao fato de a interpolação e a posterior segmentação indicarem apenas aproximadamente essas fronteiras. A Figura 5.39 ilustra tal ocorrido. A posição dessa fronteira é ajustada, conformando-se com as parcelas, conforme a Figura 5.40.

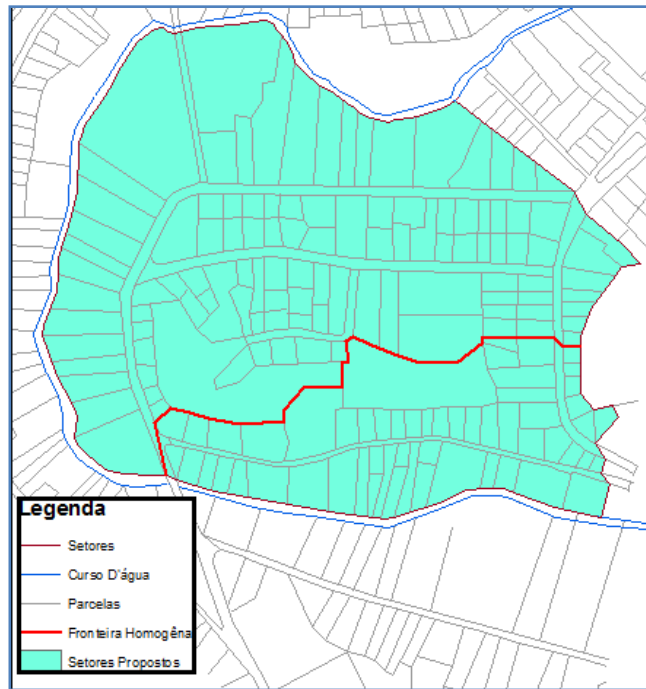


Figura 5.38 Fronteira Homogênea

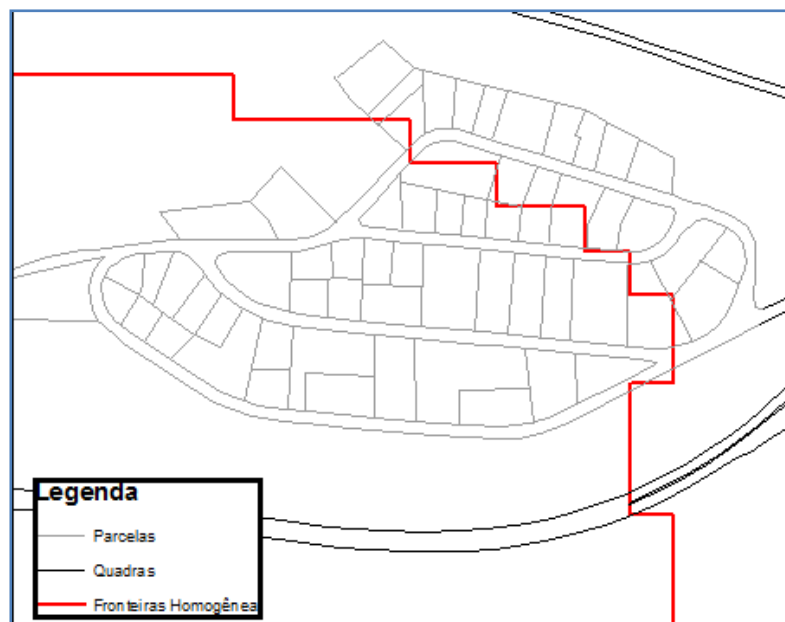


Figura 5.39 Posição da Fronteira Homogênea

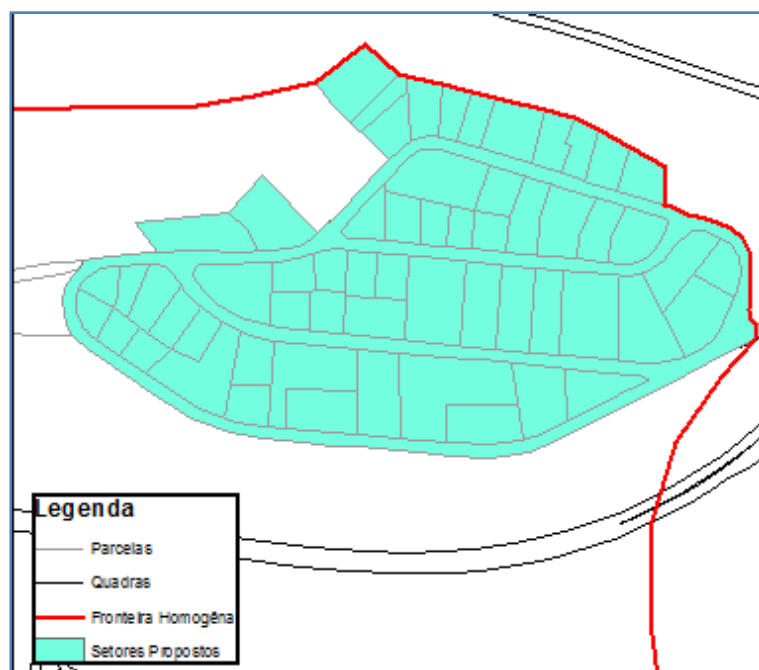


Figura 5.40 Setor censitário proposto e Fronteira Homogênea

Em outros casos, as fronteiras homogêneas podem ser ajustadas a logradouros especiais, tais como ruas estritamente comerciais ou grandes avenidas, onde não faz sentido supor que há relacionamento entre os moradores locais, não havendo, então, homogeneidade socioeconômica.

É o caso apresentado na Figura 5.41, em que uma avenida aparece como uma potencial fronteira homogênea. Como o lado direito da região é uma área isolada no desenho urbano, ela pode ser considerada como uma área homogênea. Nota-se, na Figura 5.42, como se deu o ajuste dos limites censitários e da fronteira homogênea.

Como dito anteriormente, a simulação se baseou na vetorização de algumas parcelas urbanas, não sendo possível proceder para toda área da cidade. No entanto, numa breve análise foi possível verificar essas três situações como as mais corriqueiras numa situação real. O ajuste mais abrupto das fronteiras homogêneas se dá quando se percebe que uma parte de um potencial território homogêneo não foi contemplada por elas. Observa-se que tal ajuste deve considerar a inserção de uma área numa zona homogênea em que ele se enquadre.

Destaca-se também que os vazios urbanos não foram modelados por se tratarem de glebas ainda não loteadas. Mesmo eles sendo representados por parcelas, propõe-se que o analista ou o responsável pela configuração dos setores os encaixem em setores censitários de acordo com algum zoneamento urbano preestabelecido na prefeitura. Tal

adequação é importante devido à condição de todos os setores cobrirem toda a área territorial nacional. O mesmo se diz quanto aos setores rurais, com a ressalva de que, enquanto o Cadastro Nacional de Imóveis Rurais ainda estiver incompleto, a integração cadastro-censo rural não é modelada seguindo os moldes da proposta aqui apresentada.

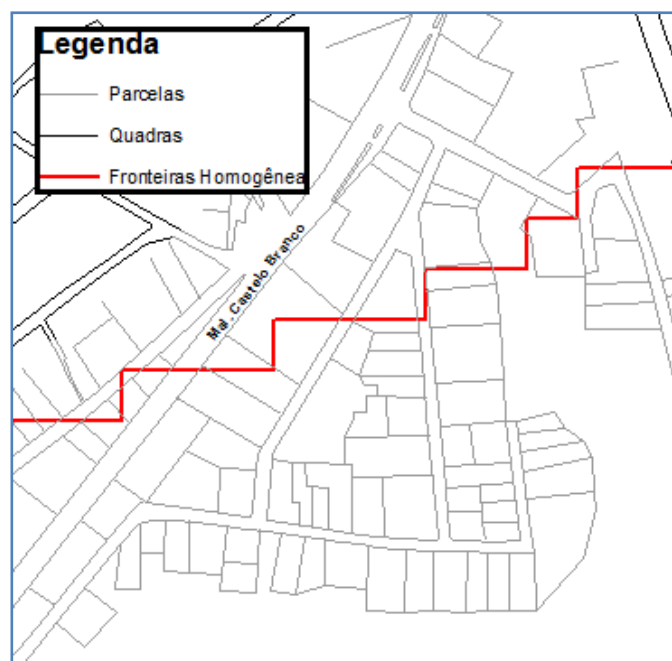


Figura 5.41 Posição da Fronteira Homogênea

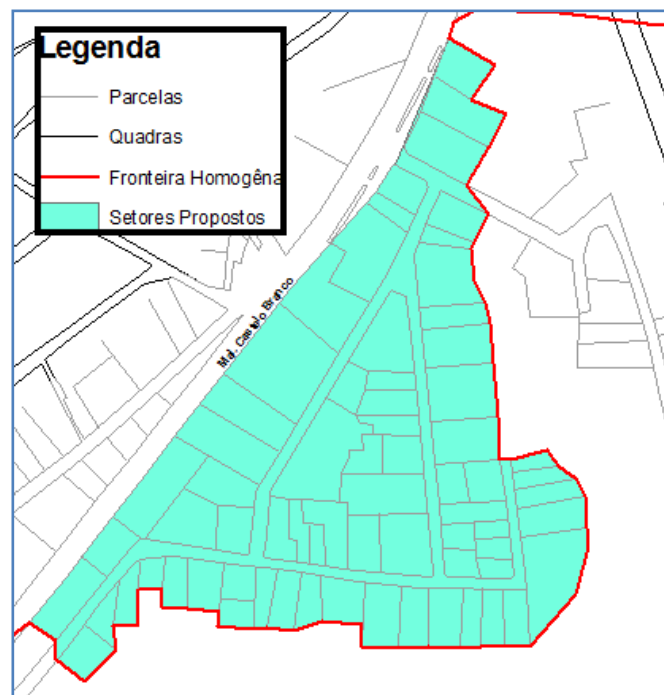


Figura 5.42 Setor censitário proposto e Fronteira Homogênea

5.4.2. Considerações finais sobre a proposta de integração

Verifica-se que a integração de informações territoriais em uma única base, consistindo em uma das funções do CTM, favorece a proposta deste trabalho, uma vez que se deve buscar unificar todas as informações sobre o território municipal, sejam elas provenientes de dados primários ou secundários. Portanto, associar os dados censitários ao território municipal consiste numa poderosa ferramenta para o entendimento da dinâmica espacial do município.

O desejo do IBGE em atualizar informações censitárias num menor intervalo de tempo comporta a adoção de análises espaço-temporais mais eficientes, permitindo que pequenas modificações nas condições populacionais sejam captadas e retratadas. Ainda em relação ao IBGE, porém em outra perspectiva, o trabalho facilitaria o órgão na realização dos censos e na disponibilização das informações censitárias.

A adesão dos dados censitários a um cadastro socioeconômico permite análises espaciais integradas à base de dados municipais. A amarração dos setores às parcelas cadastrais induz a consultas espaciais através de um banco de dados geográficos, que, se bem modelado, abre o leque de possibilidades de diagnósticos e prognósticos abaladores de tomadas de decisão.

A partir desse novo cadastro socioeconômico, pode-se pensar também em cadastro populacional, permitindo a obtenção de estatísticas importantes para o conhecimento dos habitantes de um município. Algumas dessas informações necessitam de dados no nível de parcelas (pacientes ou estudantes, por exemplo). Porém, estatísticas de características gerais da população podem ser obtidas por zonas geográficas.

Como visto anteriormente, quanto mais homogêneas essas zonas, menos distorcido pode ser o retrato populacional. Daí a importância de se ter a definição dos limites censitários a partir de um zoneamento socioeconômico homogêneo. Contudo, a proposta de integração não necessariamente passa pela obtenção das zonas homogêneas.

Caso não haja interesse por parte de um município na caracterização homogênea da população, o mesmo poderia adotar a amarração das parcelas ao CTM. É óbvio, entretanto, que tal adoção só é possível caso passe pela aprovação do IBGE.

Portanto, podem haver 4 situações distintas, as quais não são coexistentes e dependem exclusivamente das decisões quanto à aceitação da proposta dos órgãos

competentes. As duas primeiras são a situação atual, sem setores homogêneos, ou numa representação populacional homogênea, ambas sem formalização quanto ao CTM. Adicionalmente, as outras duas se diferem das anteriores pela integração ao CTM.

Por fim, este trabalho, como indicado nesta seção 5.4, defende a adoção de setores censitários homogêneos amarrados ao Cadastro Territorial Multifinalitário, conforme apontam as Diretrizes para o Cadastro, em seu artigo 5º, buscando-se a integração das informações populacionais, bem retratadas socioeconomicamente, ao *Sistema de Informações Territoriais (SIT)*.

6. CONCLUSÕES E RECOMENDAÇÕES

Com os resultados apresentados, conclui-se que o objetivo inicial de verificar a possibilidade de se encontrar áreas homogêneas a partir das técnicas de Geoprocessamento, aliadas à Estatística Espacial, foi alcançado.

A avaliação do grau de heterogeneidade interna dos dados agregados por setor censitário permite inferir sobre o quão mascarados podem estar os valores dos indicadores socioeconômicos, uma vez que apenas critérios logísticos e operacionais são considerados no estabelecimento dos limites censitários.

É importante destacar, portanto, que, em análises espaciais de indicadores agregados por área, deve estar sempre associado seu valor de incerteza, o que permite, ao menos, reconhecer o problema de instabilidade gerado pelo *MAUP*.

Por sua vez, os experimentos de agregação de dados realizados mostraram que arranjos poligonais aleatórios tendem a esconder baixos valores. Com isso, pode-se concluir que a utilização de setores censitários mais homogêneos, possivelmente, facilita encontrar valores extremos de condições socioeconômicas, em que tais setores apresentem menor variabilidade interna.

A homogeneidade interna provê maior confiabilidade nos resultados apresentados pelo censo. Num contexto geral, os dados municipais não se alteram e a tendência espacial de maiores e menores valores de variáveis do tipo média permanece quase que inalterada. Porém, numa análise quantitativa, percebeu-se que os setores homogêneos são mais representativos das populações locais.

As instabilidades representadas pelas variâncias internas são naturais nos dados agregados por área. O que se pretendeu com a metodologia para se encontrar fronteiras de homogeneidade socioeconômica foi apenas a possibilidade de se evitar a agregação de populações completamente diferentes socioeconomicamente.

Por fim, no quesito integração dos dados censitários ao CTM, a integração sistematizada das duas principais fontes de dados municipais é importante e, de certa forma, está contemplada nas Diretrizes para o Cadastro. O modelo apresentado, apesar de ter sido bem simples, busca mostrar a proposta de uma maneira fundamentada nos conceitos de dependência espacial associados ao Cadastro Territorial Multifinalitário.

Por não ser pretensão deste trabalho propor um modelo exaustivo, completamente aprofundado sobre a exploração ao máximo da utilização das informações censitárias agregadas por setores, sugere-se que trabalhos futuros iniciem uma discussão sobre a temática de integração através de focos multidisciplinares.

Sugerem-se, também, estudos mais aprofundados nos interpoladores geoestatísticos. É recomendado o estudo da krigagem por indicação, a qual permite a obtenção de incertezas locais baseadas nas Funções de Densidade de Probabilidades (FDPs). As informações acerca dessas incertezas tornam-se importantes, pois se deve lembrar que a krigagem gera uma superfície suavizada, a qual pode esconder ou mascarar pequenas variações locais. Portanto, a superfície das variâncias dos indicadores também pode ser discretizada, permitindo a obtenção de pequenas áreas escondidas pela superfície da variável estudada. Assim, as novas fronteiras seriam sobrepostas àquelas obtidas originalmente.

A discretização também é um aspecto importante a ser abordado, sugerindo-se estudo mais aprofundado sobre as segmentações de imagens. Recomenda-se um estudo sobre a detecção de agrupamentos de pixels a partir dos conceitos da *Lógica Fuzzy*, possibilitando melhores detecções de variações de padrões espaciais.

Por fim, espera-se que, a partir deste trabalho, surja uma discussão sobre tal proposta, avaliando sua aplicabilidade, num todo ou em partes.

REFERÊNCIAS BIBLIOGRÁFICAS

ABREU, M. V. S., **Estatística Espacial Aplicada à Análise da Distribuição do Índice de Desenvolvimento Humano Censossetorial de Viçosa – MG**. Monografia de conclusão do curso de Engenharia de Agrimensura. Universidade Federal de Viçosa, Viçosa, MG, 2009.

ABREU, M. V. S., OLIVEIRA, J. C., MEIRA, A. D., **Estatística Espacial Aplicada à Análise de Padrões de Segregação Socioespacial em Viçosa – MG**. Anais do III Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação Recife - PE, 27-30 de Julho de 2010a.

ABREU, M. V. S., GRIPP JR, J., SANTOS, A. P., **Proposta de Integração dos Setores Censitários ao CTM auxiliando o Cadastro Socioeconômico: Vantagens e Desvantagens**. Anais do IX Congresso de Cadastro Técnico Multifinalitário e Gestão Territorial. UFSC, Santa Catarina, 2010b.

ABREU, M. V. S., OLIVEIRA, J. C., ANDRADE, V. D. A., MEIRA, A. D., **Proposta Metodológica Para o Cálculo e Análise Espacial do IDH Intraurbano de Viçosa - MG**. *Rev. bras. estud. popul.*, São Paulo, v. 28, n. 1, jun. 2011.

ÁGUILA, M., ERBA, D. A., **El rol del Catastro em el Registro del Territorio**. In ERBA, D. A., *Catastro multifinalitário: aplicado a La definición de políticas de suelo urbano*. Cambridge, MA: Lincoln Institute of Land Policy, 2007.

ASSUNÇÃO, R. M. **Curso de Estatística Espacial**. Retirado de <http://www.est.ufmg.br/~assuncao/>. Acessado em outubro de 2008.

BORGES, K., A., V. **Modelagem de Dados Geográficos: Uma Extensão do Modelo OMT para Aplicações Geográficas**. Dissertação (Mestrado em Administração Pública – Tecnologias da Informação). Fundação João Pinheiro, Belo Horizonte, 1997.

BRANDÃO, A. C., **O Princípio da Vizinhança Geodésica no Levantamento Cadastral de Parcelas Territoriais**. Tese (Doutorado em Engenharia de Produção) – Curso de Pós-Graduação, UFSC, Florianópolis, 2003.

BRASIL. Lei nº 10.257, de 10 de junho de 2001. Estatuto da Cidade e Legislação Correlata. – 2. ed., atual. – Brasília: Senado Federal, Subsecretaria de Edições Técnicas, 2002.

_____. Política Nacional de Desenvolvimento Urbano. Ministério das Cidades, 2004.

BURROUGH, P. A. & MCDONNELL, R. A. **Principles of Geographical Information Systems**. Oxford; Oxford University Press, 1998. 333p.

CÂMARA, G., DAVIS, C., MONTEIRO, A. M., D'ALGE, J. C. **Introdução à Ciência da Geoinformação**. São José dos Campos, INPE, 2001 (*on-line*, 2ª. edição, revisada e ampliada).

CÂMARA, G., MONTEIRO, A. M. V., Notas de aula da disciplina SER-300- Estruturas de Dados Espaciais. INPE, 2003.

CAMARGO, E. C. G., **Desenvolvimento, Implementação e Teste de Procedimentos Geoestatísticos (Krigagem) no Sistema de Processamento de Informações Georreferenciadas (Spring)**. Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais. São José dos Campos, 1997.

CASANOVA, M., CÂMARA, G., DAVIS, C., VINHAS, L., RIBEIRO, G., (EDITORES). **Banco de Dados Geográficos**. São José dos Campos, MundoGeo, 2005.

COSTA, A. M., LOCH, C., SCHENINI, P. C., **Contribuições do Cadastro Técnico Multifinalitário como Ferramenta Integrada em Sistemas de Gestão Ambiental**. XII SIMPEP, Bauru, SP, 2005.

COUTINHO, A. C., **Segmentação e Classificação de Imagens Landsat TM para o Mapeamento dos Usos da Terra na Região de Campinas, SP**. Dissertação de mestrado em Ciências pelo Instituto de Biociências da USP, SP. 1997.

CUNHA, E. M. P, ERBA, D. A., **Manual de Apoio – CTM: Diretrizes para a Criação, Instituição e Atualização do Cadastro Territorial Multifinalitário nos Municípios Brasileiros**. Brasília: Ministério das Cidades, 2010.

CUPERTINO, E. A., PINTO, N. M .A, MARCOLINO, R. C., FERREIRA, A. A., RAIMUNDO, C. R., **A Emergência dos Condomínios Fechados nos Grandes Centros Urbanos: Um Recorte Bibliográfico**. III Simpósio de Ensino – SEN, Simpósios da UFV, UFV, Viçosa – MG, 21 a 24 de outubro de 2009.

DIAS, T. L.; OLIVEIRA, M. P. G.; CÂMARA, G.; CARVALHO, M. S. **Problemas de escala e a relação área-indivíduo em análise espacial de dados censitários**. Informática Pública 2002.

DRUCK, S.; CARVALHO, M.S.; CÂMARA, G.; MONTEIRO, A.V.M. (editores) **Análise Espacial de Dados Geográficos**. Brasília, EMBRAPA, 2004

FEITOSA, F. F. **Índices Espaciais para Mensurar a Segregação Residencial: O Caso de São José dos Campos (SP)**. Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais. São José dos Campos, 2005

FELGUEIRAS, C. A. **Modelagem Ambiental com Tratamento de Incertezas em Sistemas de Informações Geográficas: Paradigma Geoestatístico por Indicação**. Tese (Doutorado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais. São José dos Campos, 1999.

GARCIA, R. C., **O Que É Preciso Saber Sobre Cadastro Técnico Multifinalitário**. Caixa Econômica Federal, Rio de Janeiro, RJ, 2007.

GENOVEZ, P. **Território e Desigualdades: Análise Espacial Intraurbana no Estudo da Dinâmica de Exclusão/Inclusão Social no Espaço Urbano em São José dos Campos – SP**. Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2002.

GONÇALVES, R. P., **Modelagem Conceitual de Banco de Dados Geográficos para Cadastro Técnico Multifinalitário em Municípios de Pequeno e Médio Porte**. Dissertação (Mestrado em Engenharia Civil, ênfase em Informações Espaciais). Universidade Federal de Viçosa, Viçosa, MG, 2008.

IBGE, Fundação Instituto Brasileiro de Geografia e Estatística. **Censo Demográfico 2000 – Agregados por Setores Censitários dos Resultados do Universo**. 2ª Edição. Rio de Janeiro, 2003.

_____. **Contagem da População 2007 – Agregados por Setores Censitários**. Rio de Janeiro, 2008a.

_____. **FPM de 2008 e População de 2007**. Brasília, DF, 2008b.

_____. **Divisão Territorial do Brasil e Limites Territoriais**. Seminário Nacional com as Instituições Responsáveis por Limites Político-Administrativos. Brasília – DF, Julho de 2008c.

_____. **Base de Informações do Censo Demográfico 2010: resultados da Sinopse por setor censitário**. Rio de Janeiro, 2011.

ISAAKS, E. H. & SRIVASTAVA, R. M. **An Introduction to Applied Geostatistics**. New York, Oxford University Press, 1989. 561p.

JAKOB, A. A. E. **A Krigagem Como Método de Análises de Dados Demográficos**. In: Anais XIII Encontro da Associação Brasileira de Estudos Populacionais, Ouro Preto, 2002.

LOCH, C., ERBA, D. A., **Cadastro Técnico Multifinalitário: Rural e Urbano**. Cambridge, MA: Lincoln Institute of Land Policy, 2007.

MANZATO, G. G., SILVA, A. N. R., **Uma Estrutura Conceitual para a Definição de Regiões Urbanas Homogêneas**. Anais do XXI CONGRESSO DE PESQUISA E ENSINO EM TRANSPORTES, Rio de Janeiro, 2007.

MARTIN, D. **Geographic Information Systems: Socioeconomic Applications**. London: Routledge, 1996.

_____. **Towards an Integrated National Socioeconomic GIS: The Geography of the 2001 Census in England and Wales**. 3rd AGILE Conference on Geographic Information Science – Helsinki/Espoo, Finland, 2000.

MATOS, J. **Fundamentos de Informação Geográfica**. 5ª Edição. Editora LIDEL, Portugal, 2008.

MEDEIROS, N. G., **Segmentação Morfológica aplicada à Cartografia**. Dissertação de mestrado em Ciências Cartográficas. UNESP, SP. 2003.

MEIRELLES, M. P. S., **Análise Integrada do Ambiente Através de Geoprocessamento – Uma Proposta Metodológica Para Elaboração de Zoneamentos**. Tese de Doutorado, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1997.

NAHAS, M. I. P., **Indicadores Intra-urbanos como Instrumentos de Gestão da Qualidade de Vida Urbana em Grandes Cidades: uma Discussão Teórico-Metodológica.** In VITTE, C. C. S. e KEINERT, T. M. M., Qualidade de Vida, Planejamento e Gestão Urbana – Discussões Teórico-Metodológicas. Bertrand Brasil, Rio de Janeiro, 2009.

NEVES, M.; RAMOS, F.; CAMARGO, E.; CÂMARA, G.; MONTEIRO, A. **Análise Exploratória Espacial de Dados Sócio-Econômicos de São Paulo.** In: GIS Brasil2000, Salvador, 2000.

OLIVEIRA, A. M. C. **Os Sistemas de Informação Geográfica na elaboração de Operações Censitárias.** Trabalho de Projeto de Mestrado como requisito parcial para obtenção do grau de Mestre em Ciência e Sistemas de Informação Geográfica. Universidade de Nova Lisboa, Novembro, 2008.

OPENSHAW, S. **Developing GIS-Relevant Zone-Based Spatial Analysis Methods,** In: P. Longley, Batty, M. ed. Spatial Analysis: Modelling in a GIS Environment, New York: John Wiley & Sons, Inc., 1996, p. 390p.

PAULINO, C. D.; SOARES, P. **Probabilidades e Estatística: Notas sobre Erros e Propagação de Erros.** Setembro de 2004. Disponível em www.leb.esalq.usp.br/aulas/lce5702/PE_erros.pdf

PHILIPS, J., **Das Disposições Gerais,** In: Cunha, E. M. P., Erba, D. A., Manual de Apoio – CTM: Diretrizes para a Criação e Atualização do Cadastro Territorial Multifinalitário nos Municípios Brasileiros. Brasília: Ministério das Cidades, 2010.

PINHO, C. M. D., GOLTZ, E., CÂMARA, G. **Ontologias de Objetos Geográficos: O Lote Urbano.** In ALMEIDA, C. M., CÂMARA, G., MONTEIRO, A. M. V., Geoinformação em Urbanismo: Cidade Real x Cidade Virtual. Oficina de Textos, São Paulo, 2007.

PIRES, L. P.; CUNHA, V. G., **Segregação Urbana, Espaço Construído e Vivido: O exemplo do Novo Silvestre, Viçosa-MG- Brasil.** Encontro de Geógrafos da América Latina – EGAL, 3 a 7 de abril de 2009, Montevideu, Uruguai, ISBN 978 - 9974 - 8194 - 0 – 5. 2009.

PORTUGAL, J. G., **A Sociabilidade em Condomínios Fechados: o caso do Condomínio Residencial Recanto da Serra em Viçosa-MG.** Dissertação de mestrado em Economia Doméstica, Universidade Federal de Viçosa, UFV, 2009.

QUEIROZ, G. R., FERREIRA, K. R., **Tutorial Sobre Banco de Dados Geográficos.** INPE, GEOBRASIL, 2006.

RAMOS, F. R. **Análise Espacial de Estruturas Intraurbanas: O Caso de São Paulo.** Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais. São José dos Campos, 2002

RAMOS, F. R., CÂMARA, G., MONTEIRO, A. M. V., **Territórios Digitais Urbanos.** In ALMEIDA, C. M., CÂMARA, G., MONTEIRO, A. M. V., Geoinformação em Urbanismo: Cidade Real x Cidade Virtual. Oficina de Textos, São Paulo, 2007.

SANTOS, A. P., **Avaliação do Padrão de Exatidão Cartográfica em Imagem Ikonos e CBERS-2b da Bacia do Ribeirão São Bartolomeu, em Viçosa – MG.** Monografia de conclusão do curso de Engenharia de Agrimensura. Universidade Federal de Viçosa, Viçosa, MG, 2008.

SANTOS, C. D., **A Formação e Produção do Espaço Urbano: Discussões Preliminares Acerca da Importância das Cidades Médias para o Crescimento da Rede Urbana Brasileira.** Revista Brasileira de Gestão e Desenvolvimento Regional, v.5, n.1, p177-190, jan-abr/2009, Taubaté, SP, 2009.

SANTOS, E. A., **Utilização de Equações Diferenciais Parciais no tratamento de imagens orbitais.** Dissertação de mestrado em Ciências Cartográficas. UNESP, SP. 2002.

SANTOS, N. T. **Estatística Espacial Aplicada.** Apostila da Pós-Graduação em Estatística. UFV. 2008.

SCHULER, A. R. **Mapeamento Geoestatístico de Ensaio de Prospecção Geotécnica no Município do Rio Grande.** Monografia (Trabalho de Conclusão de Curso no Curso de Engenharia Civil da FURG). Fundação Universidade Federal do Rio Grande. Rio Grande, 2007.

SILVA, M. S., **Sistemas de Informações Geográficas: Elementos para o Desenvolvimento de Bibliotecas Digitais Geográficas Distribuídas.** Dissertação (Mestrado em Ciência da Informação). UNESP, Marília – SP, 2006.

SUMARIVA, M. N. L., PHILIPS, J., **Regularização e Reordenamento Fundiário da Localidade de Ponta Rasa, Bairro Roça Grande, em Imbituba/SC.** Anais do XXIV Congresso Brasileiro de Cartografia. Aracaju, SE, 2010.

VICENTE, J. **Estudos Comparativos de Métodos Geoestatísticos Aplicados em Agricultura de Precisão.** Dissertação (Mestrado em Engenharia Cartográfica) - Universidade Paulista “Júlio de Mesquita Filho” (UNESP), Presidente Prudente, 2004

VITE, C. C. S., **A Qualidade de Vida Urbana e sua Dimensão Subjetiva: uma Contribuição ao Debate sobre políticas Públicas e sobre a Cidade.** In VITTE, C. C. S. e KEINERT, T. M. M., Qualidade de Vida, Planejamento e Gestão Urbana – Discussões Teórico-Metodológicas. Bertrand Brasil, Rio de Janeiro, 2009.

WONG, D. **Aggregation Effects in Geo-Referencial Data.** In ARLINGHAUS, S. L. Pratical Handbook of Spatial Statistics. CRC Press (1995).

WONG, D.W.S., LEE, J. **Statistical Analysis of Geographic Information with ArcView Gis and ArcGIS.** Editora John Wiley & Sons, Inc., Hoboken, New Jersey. 2005

XAVIER-DA-SILVA, J., **Geoprocessamento em Estudos Ambientais: Uma Perspectiva Sistêmica.** In: MEIRELLES, M. S. P., CÂMARA, G., ALMEIDA, C. M., Geomática: Modelos e Aplicações Ambientais, Brasília, DF, EAMBRAPA Informação Tecnológica, 2007.