



MARIA DO CARMO DIAS BUENO

**Grade estatística:  
uma abordagem para ampliar o  
potencial analítico de dados censitários**

Campinas  
2014





UNIVERSIDADE ESTADUAL DE CAMPINAS  
Instituto de Filosofia e Ciências Humanas

MARIA DO CARMO DIAS BUENO

**Grade estatística:  
uma abordagem para ampliar o  
potencial analítico de dados censitários**

ORIENTADOR: PROF. DR. ÁLVARO DE OLIVEIRA D'ANTONA

Tese de Doutorado apresentada ao Instituto  
de Filosofia e Ciências Humanas da  
Universidade Estadual de Campinas para a  
obtenção do título de Doutora em Demografia

Campinas  
2014

Ficha catalográfica  
Universidade Estadual de Campinas  
Biblioteca do Instituto de Filosofia e Ciências Humanas  
Cecília Maria Jorge Nicolau - CRB 8/338

B862g Bueno, Maria do Carmo Dias, 1962-  
Grade estatística : uma abordagem para ampliar o potencial analítico de dados censitários / Maria do Carmo Dias Bueno. – Campinas, SP : [s.n.], 2014.

Orientador: Álvaro de Oliveira D'Antona.  
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Filosofia e Ciências Humanas.

1. População. 2. Censo. 3. Distribuição espacial da população. 4. Análise espacial . 5. Integração de dados. I. D'Antona, Álvaro de Oliveira, 1967-. II. Universidade Estadual de Campinas. Instituto de Filosofia e Ciências Humanas. III. Título.

Informações para Biblioteca Digital

**Título em outro idioma:** Statistical grid : an approach to extend the analytical power of census data

**Palavras-chave em inglês:**

Population

Census

Spatial distribution of a population

Spatial analysis

Data integration

**Área de concentração:** Demografia

**Titulação:** Doutora em Demografia

**Banca examinadora:**

Álvaro de Oliveira D'Antona [Orientador]

Roberto Luiz do Carmo

Eduardo José Marandola Júnior

Flávia da Fonseca Feitosa

Christovam de Castro Barcellos Neto

**Data de defesa:** 21-11-2014

**Programa de Pós-Graduação:** Demografia



UNIVERSIDADE ESTADUAL DE CAMPINAS  
INSTITUTO DE FILOSOFIA E CIÊNCIAS HUMANAS

A Comissão Julgadora dos trabalhos de Defesa de Tese de Doutorado, em sessão pública realizada em 21 de novembro de 2014, considerou a candidata MARIA DO CARMO DIAS BUENO aprovada.

Este exemplar corresponde à redação final da Tese defendida e aprovada pela Comissão Julgadora.

Prof. Dr. Álvaro de Oliveira D'Antona

Álvaro de Oliveira D'Antona

Prof. Dr. Roberto Luiz do Carmo

Roberto Luiz do Carmo

Prof. Dr. Eduardo Jose Marandola Junior

Eduardo Jose Marandola Junior

Profª. Dra. Flávia da Fonseca Feitosa

Flávia da Fonseca Feitosa

Prof. Dr. Christovam de Castro Barcellos Neto

Christovam de Castro Barcellos Neto



## RESUMO

Dois problemas relacionados com a utilização de dados agregados são bastante conhecidos e citados na literatura. O primeiro deles diz respeito à não coincidência entre as unidades para as quais os dados são disponibilizados e aquelas para as quais os dados são necessários; o segundo está relacionado com a estabilidade temporal das unidades de agregação. Como solução para esses problemas este trabalho propõe a utilização de células regulares dispostas em um sistema de grade para a agregação e disseminação de dados censitários, denominada “grade estatística”. As células desta grade têm pequenas dimensões, podendo ser consideradas como “tijolos” que se juntam para formar qualquer recorte espacial desejado e não se alteram ao longo do tempo. A utilização de dados agregados em unidades de pequenas dimensões proporciona também um aumento do potencial analítico dos dados. Foram selecionadas duas unidades da federação – Pará e São Paulo - para a criação da metodologia de geração da grade estatística e para a execução de aplicações práticas. A abordagem selecionada é híbrida, mesclando agregação e desagregação. A primeira utiliza os microdados censitários associados aos seus atributos de localização para a agregação dos dados; a segunda utiliza métodos espaciais e/ou estatísticos juntamente com dados secundários para a realocação espacial dos dados. Os resultados obtidos demonstraram que a abordagem híbrida é viável e deve ser utilizada quando existe uma grande variação na qualidade dos dados e a área de estudo abrange grandes extensões. Os dados obtidos apresentaram uma boa qualidade e a sua utilização permitiu a execução de análises com um maior nível de desagregação espacial, além de facilitar a integração de dados agregados em unidades geográficas diferentes.

**Palavras-chave:** População, Censo, Distribuição espacial da população, Análise espacial, Integração de dados.



## ABSTRACT

Two problems associated with the use of aggregate data are well known and reported in the literature. The first concerns the mismatch between those units for which data is available and those for which data is needed; the second is related to the temporal stability of the aggregation units. As a solution to these problems, this study proposes the use of regular cells arranged in a grid-like pattern, called a “statistical grid”, to aggregate and disseminate census data. The cells of this grid have small dimensions; they may be considered as “bricks” that come together to fill any desired spatial area and remain unchanged over time. The use of aggregate data in units of small dimensions also provides an increased analytical power. Two states – Pará and São Paulo – were selected to create a methodology for generating a statistical grid and to implement practical applications. A hybrid approach was selected, merging aggregation and disaggregation approaches. The first uses the census microdata associated with its location attributes; the second uses spatial and/or statistical methods and ancillary data. The results demonstrated that the hybrid approach is viable and should be used in the event of a large variation in data quality and when the study area encompasses large expanses. The data obtained are of good fit and their use enabled the execution of analysis with a higher level of spatial disaggregation, in addition to facilitating the integration of data available in different geographical units.

**Keywords:** Population; Census; Spatial distribution of a population; Spatial analysis; Data integration.



## SUMÁRIO

<b>INTRODUÇÃO .....</b>	<b>1</b>
<b>CAPÍTULO 1 - BASES CONCEITUAIS E METODOLÓGICAS.....</b>	<b>9</b>
1.1 Informações Espaciais e Ciências Sociais.....	9
1.1.1 O contexto espacial.....	9
1.1.2 Informações Espaciais e Demografia.....	14
1.2 A geografia dos censos .....	18
1.2.1 Unidades geográficas.....	21
1.2.2 Representação espacial.....	25
1.2.3 Mudanças de suporte espacial.....	27
1.2.4 Soluções propostas.....	35
1.3 O sistema de grades regulares .....	37
1.3.1 Definição e história.....	38
1.3.2 Vantagens .....	39
1.3.3 Desafios .....	48
1.3.4 Abordagens.....	52
1.3.5 Usos.....	63
<b>CAPÍTULO 2 - DADOS E MÉTODOS .....</b>	<b>67</b>
2.1 Censo Demográfico 2010 .....	67
2.1.1 Inovações.....	68
2.1.2 Unidade geográfica de coleta de dados .....	71
2.1.3 Coleta das informações.....	73

2.2 Dados .....	76
2.2.1 Dados estatísticos .....	76
2.2.2 Dados vetoriais .....	79
2.3 Grade Estatística .....	79
2.3.1 Definição geométrica e espacial .....	79
2.3.2 Abordagem híbrida.....	86
2.3.3 Procedimentos .....	95
<b>CAPÍTULO 3: RESULTADOS E ANÁLISES.....</b>	<b>107</b>
3.1 Aplicação 1: Adaptação a recortes geográficos diversos.....	116
3.1.1 Introdução .....	116
3.1.2 Materiais e métodos.....	117
3.1.3 Resultados .....	119
3.1.4 Discussão.....	123
3.2 Aplicação 2: Integração de dados.....	125
3.2.1 Introdução .....	126
3.2.2 Materiais e métodos.....	127
3.2.3 Resultados .....	131
3.2.4 Discussão.....	139
3.3 Aplicação 3: Análise espacial.....	141
3.3.1 Introdução .....	142
3.3.2 Materiais e métodos.....	142
3.3.3 Resultados .....	145
3.3.4 Discussão.....	156

<b>CONSIDERAÇÕES FINAIS.....</b>	<b>159</b>
<b>REFERÊNCIAS.....</b>	<b>167</b>
<b>APÊNDICE A - Avaliação das dimensões das células .....</b>	<b>189</b>
<b>APÊNDICE B - Avaliação de métodos de desagregação .....</b>	<b>217</b>



## **AGRADECIMENTOS**

Agradeço a todos que contribuíram, de forma direta ou indireta, para a realização deste trabalho, em especial:

Ao meu orientador, Dr. Álvaro de Oliveira D'Antona, pela orientação e incentivo.

Ao Professor David Martin, pela generosidade no compartilhamento dos seus conhecimentos.

Ao IBGE, pela oportunidade proporcionada e pela cessão dos dados.

À CAPES, pelo apoio financeiro durante o Programa de Doutorado Sanduíche no Exterior.

Aos colegas do IBGE e da UNICAMP, pela torcida e pela disposição em ajudar.

À minha família, pelo incentivo para prosseguir pelo caminho que escolhi.



## LISTA DE FIGURAS

Figura 1 - Geografia do censo no Brasil, 2010.....	20
Figura 2 - Representação espacial de dados.....	26
Figura 3 - Aspectos do MAUP.....	30
Figura 4 - MAUP: Efeito de Escala.....	30
Figura 5 - Influência da escala na percepção de um fenômeno.....	31
Figura 6 - MAUP: Efeito de Zoneamento.....	32
Figura 7 - Influência do zoneamento na percepção de um fenômeno.....	33
Figura 8 - Estrutura hierárquica de células.....	44
Figura 9 - Técnicas para transferência de suporte espacial.....	53
Figura 10 - Tela do aplicativo do Censo 2010 para computador de mão.....	70
Figura 11 - Exemplo de codificação de Quadra/Face.....	71
Figura 12 - Árvore de decisão para escolha da abordagem e da técnica de desagregação.....	94
Figura 13 - Procedimentos para agregação de dados pontuais.....	96
Figura 14 - Procedimentos para agregação de dados lineares.....	98
Figura 15 - Procedimentos para incorporação direta de dados de setores censitários em células.....	100
Figura 16 - Procedimentos para desagregação de dados utilizando malha viária como dado auxiliar.....	101
Figura 17 - Procedimentos para desagregação de dados utilizando classificação de uso/cobertura das terras como dado auxiliar.....	104
Figura 18 - Procedimentos para desagregação de dados utilizando ponderação zonal.....	105
Figura 19 - Distribuição da população utilizando setor censitário e grade estatística, Pará, 2010.....	111
Figura 20 - Distribuição da população utilizando setor censitário e grade estatística, São Paulo, 2010.....	112
Figura 21 - Localização das unidades de conservação selecionadas, Pará, 2010.....	118
Figura 22 - Distribuição da população nas unidades de conservação selecionadas, Pará, 2010.....	121
Figura 23 - Adequação da grade estatística aos limites das UCs.....	125

Figura 24 - Mapa de suscetibilidade com divisão de setores censitários, São Sebastião, 2010. ....	130
Figura 25 - Compatibilização de escalas.....	141
Figura 26 - Delimitação das áreas de estudo para a análise do índice Global e Local de Moran, Limeira, 2010. ....	144
Figura 27 - Índice Local de Moran – Raça/Cor, Limeira, 2010. ....	149
Figura 28 - Índice Local de Moran – Alfabetização, Limeira, 2010.....	150
Figura 29 - Índice Local de Moran - Razão entre crianças e jovens + adultos, Limeira, 2010.....	151
Figura 30 - Índice Local de Moran - Razão de idosos e jovens + adultos, Limeira, 2010. ....	152
Figura 31 - Índice Local de Moran – Renda média mensal domiciliar <i>per capita</i> , Limeira, 2010.....	153
Figura 32 - Índice Local de Moran - Razão de sexos, Limeira, 2010. ....	154
Figura 33 - Índice Local de Moran – Densidade de população, Limeira, 2010....	155
Figura 34 - Grade estatística: células com dados suprimidos pelo critério de sigilo, Limeira, 2010.....	157

## LISTA DE GRÁFICOS

Gráfico 1 - Setores censitários e domicílios sem dados de localização por faixa de ausência de localização, PA e SP (litoral), 2010. ....	92
Gráfico 2 - Estrutura etária no interior e entorno de duas unidades de conservação, Pará, 2010. ....	122
Gráfico 3 - Renda domiciliar média <i>per capita</i> em salários mínimos no interior e entorno de duas unidades de conservação, Pará, 2010. ....	123
Gráfico 4 - Autocorrelação espacial ( <i>z-score</i> ) de acordo com a distância, Limeira, 2010. ....	146



## LISTA DE QUADROS

Quadro 1 - Família de grades estatísticas do Japão. ....	46
Quadro 2 - Confidencialidade: valor limite e método utilizado.....	49
Quadro 3 - Escala e dimensão de cartas, folhas topográficas e cadastrais. ....	83
Quadro 4 - Sumário das aplicações utilizadas para analisar o potencial analítico dos dados censitários agregados em grade estatística.....	115
Quadro 5 - Unidades de Conservação Ambiental selecionadas, Pará.....	119
Quadro 6 - Classe de suscetibilidade por compartimento fisiográfico. ....	129



## LISTA DE TABELAS

Tabela 1 - Resolução espacial das unidades geográficas, Pará e São Paulo, 2010. ....	22
Tabela 2 - Setores censitários povoados por situação, Brasil, 2010. ....	73
Tabela 3 - Quantidade de setores censitários povoados nas operações de pré-coleta e coleta, UF, 2010. ....	75
Tabela 4 - Área e população por situação, Pará e São Paulo, 2010. ....	84
Tabela 5 - Quantidade de células povoadas e não povoadas, por dimensão, Pará e São Paulo, 2010. ....	108
Tabela 6 - Quantidade de setores censitários por abordagem e técnica para geração de grade estatística, Pará e São Paulo, 2010. ....	108
Tabela 7 - Quantidade de células da grade estatística por abordagem, Pará e São Paulo, 2010. ....	110
Tabela 8 - Unidades e população atingidas e não atingidas pelo critério de sigilo estatístico, Pará e São Paulo, 2010. ....	114
Tabela 9 - População residente, total e por sexo, e domicílios no interior das unidades de conservação selecionadas, 2010. ....	119
Tabela 10 - População por sexo de acordo com os compartimentos fisiográficos, São Sebastião, 2010. ....	131
Tabela 11 - Distribuição da população por faixa etária de acordo com os compartimentos fisiográficos, São Sebastião, 2010. ....	132
Tabela 12 - Distribuição de renda domiciliar per capita em faixas de Salário Mínimo (SM) de acordo com os compartimentos fisiográficos, São Sebastião, 2010. ....	132
Tabela 13 - Tipo de ocupação do imóvel de acordo com os compartimentos fisiográficos, São Sebastião, 2010. ....	133
Tabela 14 - Espécie de unidade doméstica de acordo com os compartimentos fisiográficos, São Sebastião, 2010. ....	134
Tabela 15 - Distribuição da população por classe de suscetibilidade e compartimento fisiográfico, São Sebastião, 2010. ....	135

Tabela 16 - Faixa etária da população por classe de suscetibilidade e compartimento fisiográfico, São Sebastião, 2010. ....	136
Tabela 17 - Distribuição de renda domiciliar <i>per capita</i> em faixas de Salário Mínimo (SM) por classe de suscetibilidade e compartimento fisiográfico, São Sebastião, 2010. ....	137
Tabela 18 - Tipo de ocupação do imóvel de acordo com a classe de suscetibilidade e o compartimento fisiográfico, São Sebastião, 2010. ....	138
Tabela 19 - Espécie de unidade doméstica de acordo com a classe de suscetibilidade e o compartimento fisiográfico, São Sebastião, 2010. ....	139

## INTRODUÇÃO

A demografia se ocupa do estudo das populações humanas, sendo que uma parte desses estudos se baseia na agregação dos indivíduos em unidades geográficas, geralmente político-administrativas, enquanto outra parte tem por base as unidades sociais, como as famílias. A maioria dos dados utilizados pela demografia tem origem em pesquisas domiciliares, sejam elas censitárias ou amostrais, e como esses dados estão agregados em unidades de área, e estas áreas têm uma localização no espaço, pode-se dizer que o espaço sempre esteve presente na demografia (VOSS, 2007). Apesar desta presença inegável do espaço, na maioria dos estudos demográficos ele é tratado apenas como um repositório inócuo, não havendo nenhuma interação com os fenômenos estudados (SWEENEY, 2002).

A maioria dos dados sociodemográficos tem um elemento espacial – os crimes são cometidos num determinado lugar, o desemprego é calculado para uma cidade ou região, a localização de áreas com resíduos perigosos ocorre a certa distância de núcleos populacionais, a população aumenta ou diminui nas proximidades de grandes centros, as doenças e mortes estão distribuídas irregularmente no espaço. No entanto, os procedimentos convencionais de análise demográfica geralmente não utilizam outras camadas espaciais de informação que estão subjacentes aos fenômenos analisados, sejam elas camadas de estruturas físicas (estradas) ou naturais (rios, vegetação), ou camadas de localização de eventos (ocorrência de crimes) ou de serviços (estabelecimentos de saúde).

Nos últimos anos, as tecnologias espaciais ou geotecnologias têm ganhado um maior destaque em diversas áreas da ciência devido, principalmente, à melhoria da relação custo-capacidade-benefício dos sistemas computacionais e da popularização de ferramentas de visualização e manipulação de dados espaciais. Além disso, a crescente disponibilidade de dados espaciais, que permitem a espacialização de variáveis demográficas, sociais e econômicas, também tem contribuído para o aumento do interesse por geotecnologias (WACHTER, 2005).

Hoje em dia, a visualização de dados sob a forma de mapas se tornou uma tarefa quase imprescindível em estudos que envolvam dados passíveis de serem espacializados. Fica bastante evidente a utilização cada vez maior de Sistemas de Informação Geográfica (SIG) e de técnicas de análise espacial na área da demografia, como observado por Matthews (2003) a partir da quantificação de sessões e oficinas oferecidas em encontros como o *Population Association of America Annual Meeting*. No entanto, ainda existe uma carência de modelos e métodos nos quais o espaço seja tratado como fator interveniente nos processos demográficos. É necessário que o mesmo tratamento dado pelos demógrafos à dimensão temporal, trabalhando em termos de período e coorte, seja dado à dimensão espacial. É necessária também uma utilização maior das geotecnologias nos estudos demográficos, além do efetivo reconhecimento do papel do espaço nos processos sociais (NICHD, 2007).

Alguns fatores são citados na literatura como sendo responsáveis pela dificuldade de incorporação do espaço nas Ciências Sociais (PEÑA, 2006). O primeiro se refere à unidade espacial em que a maioria dos dados sociodemográficos está disponível, ou seja, áreas geográficas com delimitação político-administrativa (bairro, município, estado) ou operacional (setores censitários). Além dessas unidades nem sempre serem as mais indicadas para o estudo que se deseja realizar, ainda há dificuldades para a integração de dados de origens diversas, já que, em geral, os dados são usualmente disponibilizados em unidades geográficas que atendem às necessidades específicas da sua própria área de atuação.

Nas últimas décadas, no campo dos estudos ambientais, surgiram diversas iniciativas no sentido de estabelecer unidades de análise para descrever, monitorar e analisar o ambiente natural, como unidades de relevo e unidades de paisagem, sem mencionar as clássicas bacias hidrográficas; no entanto, com relação aos estudos da sociedade humana, continuam prevalecendo os sistemas tradicionais de unidades político-administrativas (BACKER, 2008). Isso se deve ao fato de que as estatísticas sociodemográficas e econômicas oficiais atendem aos

objetivos dos governos, quais sejam os de direcionar o planejamento e a execução de políticas públicas, além de monitorar a evolução das mesmas.

Essas unidades político-administrativas podem variar substancialmente em tamanho e forma de uma região para outra, assim como podem variar também no tempo. Os limites administrativos também funcionam como barreiras artificiais na representação dos fenômenos socioeconômicos e ambientais (GRASLAND; MADELIN, 2006). Por essas razões, a utilização dessas unidades coloca sérios obstáculos às tentativas de combinar diferentes conjuntos de dados, como por exemplo, dados ambientais e demográficos, tarefa bastante comum em estudos de População e Ambiente (de SHERBININ *et al.*, 2002).

A utilização de dados demográficos oriundos de pesquisas domiciliares disponibilizados por unidade de coleta é a situação ideal, não por ser a escala mais apropriada para observar e modelar fenômenos sociais, mas porque oferece a possibilidade de se trabalhar em todas as escalas e em qualquer tipo de compartimentação espacial (GRASLAND; MADELIN, 2006). Infelizmente, existem restrições à divulgação de dados desta forma devido a questões legais de manutenção do sigilo estatístico, principalmente se considerarmos as pesquisas realizadas pelos órgãos estatísticos oficiais. Na realidade, a maioria dos dados estatísticos existentes é disponibilizada em agrupamentos de áreas, e a hipótese de homogeneidade interna dessas áreas é considerada verdadeira, não importando a forma ou dimensão dessas unidades. E ainda há a questão da dificuldade de comparação dos dados ao longo do tempo, pois tanto as unidades operacionais - os setores censitários - quanto as unidades político-administrativas podem ter seus limites alterados a qualquer momento.

Existem outros dois fenômenos problemáticos relacionados com a utilização de dados agregados em unidades de área e que também colocam dificuldades na incorporação do espaço nas Ciências Sociais. Um deles é o “problema da unidade de área modificável” ou *Modified Area Unit Problem* – MAUP, no original em inglês (OPENSHAW, 1984). Este fenômeno está relacionado com dois efeitos: o efeito escala ou efeito de agregação, que diz respeito à existência de inferências diferentes obtidas quando o mesmo conjunto

de dados é agrupado em unidades de área cada vez maiores, e o efeito de agrupamento ou zoneamento, que trata da variabilidade dos resultados devido às diferentes formações das unidades de área, com variações nas suas formas, mas com a escala mantendo-se aproximadamente a mesma (GOTWAY; YOUNG, 2002).

O outro fenômeno problemático é conhecido como “falácia ecológica” (ROBINSON, 1950). Este fenômeno ocorre quando as análises baseadas em dados agregados levam a conclusões diferentes daquelas baseadas em dados individuais. A falácia ecológica também apresenta dois componentes: um relacionado com o agrupamento dos indivíduos e outro relacionado com a distribuição das variáveis nesses agrupamentos.

Diversos estudos realizados concluem que os problemas da falácia ecológica e da unidade de área modificável são inerentes aos dados agregados em áreas, não podendo ser removidos, nem tampouco ignorados (CARVALHO *et al.*, 2004; OPENSHAW, 1996). Apesar da existência de críticas à afirmação de que para minimizar o impacto do problema da unidade de área modificável devem-se utilizar dados no menor nível de agregação possível (OPENSHAW, 1996), esta opção é defendida por alguns estudiosos do assunto (CARVALHO *et al.*, 2004; GRASLAND; MADELIN, 2006; GOODCHILD, 1992). Seguindo esta linha de raciocínio, Schuurman *et al.* (2006) mostram que diversos pesquisadores sugerem a desagregação dos dados censitários, ou seja, a quebra dos setores censitários em unidades menores, com o objetivo de minimizar os problemas ocasionados tanto pela falácia ecológica quanto pelo problema da unidade de área modificável.

Os problemas citados acima fragilizam os estudos demográficos, levando os pesquisadores a não considerarem a dimensão social nesses estudos ou a utilizarem unidades com uma extensão inadequada, sendo que ambos os caminhos levam a resultados incoerentes e incertos.

Diante da problemática apresentada, surgem as seguintes questões:

- a) Como obter dados para a realização de análises sociodemográficas em áreas que não são coincidentes com unidades político-administrativas?

- b) Como efetuar análises temporais de dados demográficos, tendo em vista que as unidades utilizadas para coletar e disseminar dados se alteram ao longo do tempo?
- c) Qual o melhor procedimento para integrar espacialmente dados oriundos de diferentes fontes e com agregação em diferentes unidades espaciais?

Sugerimos que uma possível solução para essas questões está em agregar os dados em unidades geográficas que sejam independentes de limites político-administrativos ou de limites operacionais e que tenham dimensões suficientemente pequenas para permitir a sua conformação a diferentes recortes espaciais. Essas características nos levam a uma unidade geográfica composta por um conjunto células de forma regular, agrupadas em um sistema hierárquico e que daqui por diante denominaremos de “grade estatística”.

O objetivo geral desta pesquisa é contribuir para a determinação de uma metodologia para geração de dados censitários agregados em células regulares, o que constitui um produto inédito no Brasil. O produto obtido com esta metodologia pretende viabilizar a geração de análises que solucionem os questionamentos apresentados e aumente o potencial analítico dos dados, principalmente no que diz respeito a análises espaciais. Os objetivos específicos desta pesquisa são a investigação de técnicas de agregação e desagregação de dados, a investigação das melhores dimensões de células para conjugar a discretização de dados e a confidencialidade dos respondentes, a construção de uma metodologia para a geração de uma grade regular para a disseminação de dados censitários e a avaliação desta grade em relação às formas convencionais de disseminação, investigando suas vantagens, desvantagens e potenciais na realização de estudos de população.

O primeiro capítulo trata das bases e referências conceituais e metodológicas, trazendo uma visão sobre o papel do espaço nas pesquisas científicas, os motivos da sua inclusão ou não em algumas áreas específicas e o reconhecimento da importância do espaço como plano de análise, o que, aliado aos atuais avanços tecnológicos, faz com que o mesmo ressurgira e se torne cada

vez mais presente na vida cotidiana e nos estudos analíticos. A segunda parte do capítulo discute as unidades espaciais de análise, que são responsabilizadas pelas principais dificuldades para a incorporação do espaço nas análises demográficas. Nesta parte do capítulo, são mostradas as principais dificuldades envolvendo a utilização de unidades de área em estudos dos mais diferentes campos do conhecimento. Também são discutidos os problemas inerentes aos dados agregados, como a falácia ecológica e o problema da unidade de área modificável, bem como as propostas para minimizá-los. A terceira parte deste capítulo detalha a abordagem de grades regulares como unidade de agregação de dados, além de mostrar um histórico de sua utilização.

O segundo capítulo inicia com uma exposição sobre o censo no Brasil e segue descrevendo os dados utilizados e a metodologia proposta para a geração de uma grade regular para disseminação de dados censitários. Vários aspectos metodológicos são abordados, discutindo as opções feitas com base em estudos empíricos, em revisões bibliográficas e na qualidade dos dados existentes. No item relacionado com a metodologia, são discutidas as opções que foram adotadas com relação à definição geométrica da grade estatística, incluindo o sistema de projeção cartográfica e a dimensão das células. Esta última questão expõe de forma explícita o dilema entre as necessidades do usuário, que deseja ter dados detalhados e desagregados, e as necessidades do produtor de dados, que tem a obrigação de manter a confidencialidade. Em seguida, são mostrados os procedimentos operacionais para espacialização dos dados de acordo com a abordagem selecionada e, dentro de cada abordagem, da técnica escolhida. Todos os procedimentos são detalhados e ilustrados, mostrando o fluxo de operações realizadas para tratar os dados cartográficos e estatísticos até se obter a espacialização dos dados censitários em um sistema de unidades celulares.

O terceiro capítulo traz os resultados obtidos com o emprego da metodologia proposta e apresenta três aplicações práticas envolvendo a utilização de grade estatística em estudos de população. O objetivo dessas aplicações é avaliar o ganho de potencial analítico promovido pela utilização de dados agregados em unidades geográficas discretas e regulares. A primeira aplicação

demonstra a utilização da grade estatística em áreas de interesse cujos limites não coincidem com as unidades de divulgação tradicionais dos censos, como é o caso de algumas unidades de conservação ambiental. Na segunda aplicação, é mostrada a integração de dados ambientais físicos e dados demográficos, mostrando o desafio de utilizar dados de fontes diferentes, reunidos com o objetivo de fornecer elementos para estudos e análises da interação da dimensão humana e dos processos e sistemas ambientais. Já a terceira aplicação ilustra a utilização da grade estatística em estudos intraurbanos, empregando técnicas de análise espacial para explorar o comportamento de algumas variáveis demográficas no território.

Finalmente, é apresentada uma avaliação geral da metodologia proposta e são sugeridas pesquisas futuras, visando ao avanço no conhecimento e desenvolvimento de soluções para as questões aqui abordadas.



## **CAPÍTULO 1 - BASES CONCEITUAIS E METODOLÓGICAS**

*...we are becoming increasingly aware that we are, and  
always have been, intrinsically spatial as well as temporal beings...  
(Edward Soja)*

Este capítulo traz as bases e referências conceituais e metodológicas utilizadas para a elaboração desta tese. No início, é apresentada uma discussão sobre o papel do espaço nas pesquisas científicas, sua progressiva inclusão em algumas áreas específicas, particularmente na Demografia, o papel protagonizado pelos avanços tecnológicos no interesse crescente pelo espaço e as dificuldades e avanços para a sua efetiva utilização pelos pesquisadores. Depois, são discutidas as unidades espaciais de análise, mostrando seus objetivos e suas limitações. Também são discutidos os problemas inerentes aos dados agregados em unidades de área, como a falácia ecológica e o problema da unidade de área modificável, bem como as propostas para minimizá-los. Em seguida, apresentamos o sistema de grade regular utilizado como estrutura de armazenamento de dados, suas vantagens e desafios, além das abordagens utilizadas para o povoamento das células desta grade. Finalmente, mostramos alguns exemplos de grades implantadas para o mapeamento da distribuição e/ou caracterização da população nos níveis global, continental e nacional.

### **1.1 Informações Espaciais e Ciências Sociais**

#### **1.1.1 O contexto espacial**

O espaço faz parte da nossa vida cotidiana, pois as pessoas e o ambiente, tanto o natural quanto o construído, ocupam um lugar na Terra, ou seja, tudo está em algum lugar. Poderíamos dizer de uma maneira simplista que o espaço nada mais é do que a superfície do nosso planeta. Superfície esta que,

apesar das dimensões, a cada dia que passa se torna menos relevante se considerarmos o avanço das tecnologias de transporte e comunicação (MOUW, 2000). Mas não podemos concluir que o espaço é importante apenas porque tudo acontece em algum lugar, mas porque saber onde as coisas acontecem é crucial para o entendimento de como e porque elas acontecem (WARF; ARIAS, 2009).

Muitas teorias sociais relacionam o comportamento individual ou de um grupo familiar com o contexto da localização espacial deste indivíduo ou grupo familiar (RINDFUSS; STERN, 1998), ou seja, procuram explicar esse comportamento tendo em conta um contexto social e espacial (ENTWISLE, 2007). Seguindo este raciocínio, para termos uma visão completa de um fenômeno precisamos compreender a relação entre a ocorrência deste fenômeno em uma determinada região e as diversas dimensões da realidade que se apresentam naquele espaço, desde os aspectos físicos e ambientais, até os aspectos humanos e sociais.

Isto nos remete ao modelo de camadas, comum em textos introdutórios de SIG, que nos propõe a sobreposição de uma série de camadas transparentes, cada uma delas representando uma variável ou fator relevante ao estudo em questão, de modo a permitir a visualização dos locais mais afetados, que seriam aqueles que apresentam uma maior opacidade (MCHARG, 1969). Esta funcionalidade, incorporada aos programas de geoprocessamento, se tornou uma das principais características dos SIGs. Com isso, podemos concluir que uma das características do espaço é ser um suporte para a integração de informações, sejam elas quantitativas ou qualitativas (TAYLOR, 2010).

O espaço também pode proporcionar outra forma de integração: a integração da ciência com as políticas públicas (GOODCHILD; JANELLE, 2004). O estudo da sociedade a partir de diferentes pontos de vistas oriundos de diversas áreas disciplinares – economia, ciências sociais, educação e saúde pública, por exemplo – gera uma série de dados que podem ser integrados espacialmente para o desenvolvimento e implantação de políticas que buscam promover o desenvolvimento social e humano. E esta integração nos leva a distinguir a

natureza interdisciplinar do espaço. A utilização do espaço como plano de análise se desenvolveu inicialmente no campo da Geografia, mas nas últimas décadas, com a ressurgência do tema, existem manifestações em diversas outras áreas do conhecimento, como as Ciências Sociais e Humanas, a Estatística e a Biologia (WACHTER, 2005). Mesmo o espaço sendo um tópico comum em muitas disciplinas, há diferenças na maneira com que cada uma delas utiliza o espaço na condução dos seus estudos analíticos, ou, como enfatizado por Arias (2010), o espaço fala mais de uma linguagem.

Podemos destacar algumas disciplinas nas quais o espaço tem um papel significativo e que apresentam cada vez mais estudos que o envolvem. Primeiro temos a Epidemiologia e a Geografia da Saúde ou Geografia Médica (ROSENBERG, 1998), fazendo inferências a partir de padrões espaciais, como John Snow e seus estudos sobre o cólera já em meados do século XIX (SNOW, 1855). Já no século XX, temos, por exemplo, a modelagem da difusão de doenças no espaço-tempo, como proposto inicialmente pelo geógrafo sueco T. Hagerstrand em 1953 (HAGERSTRAND, 1967). Mais recentemente, temos diversos estudos tratando de questões relativas à transmissão de doenças contagiosas, tendo esses estudos enfoques globais (LINARD; TATEM, 2012; TATEM *et al.*, 2012) e locais (STODDARD *et al.*, 2013; CARVALHO; NASCIMENTO, 2012). Há também diversos estudos recentes nesta área realizados com a utilização de dados inovadores, como dados de telefonia móvel (TATEM *et al.*, 2009). Nesta área específica, outro tipo de estudo bastante comum está relacionado com o planejamento e acesso a serviços de saúde. Como exemplo, temos o estudo desenvolvido por Oliveira, Travassos e Carvalho (2004) e Oliveira *et al.* (2011), que investiga a influência da distância entre o usuário e o prestador de serviços nos padrões de internação, e o estudo de Muller, Cubas e Bastos (2010), que discute a utilização de sistemas de informação geográfica como ferramenta de gestão de dados territoriais para auxiliar o planejamento e gestão de serviços de saúde da família.

A Economia é outra área do conhecimento que utiliza o espaço em seus estudos, com grande utilização de técnicas de econometria e estatística espacial aplicadas a temas como crescimento econômico, desigualdade, mercado de trabalho, migração, sustentabilidade e outros, além do grande arcabouço metodológico desenvolvido por pesquisadores como Anselin, Florax e Rey (2004).

Finalmente, temos a Demografia, que está representada por estudos relacionados com planejamento urbano e regional, como segregação espacial, *urban sprawl*<sup>1</sup>, mobilidade e vulnerabilidade (VILLAÇA, 1997; FONSECA; DAVANZO; NEGREIROS, 2002; CUNHA *et al.*, 2006; MARANDOLA JR; HOGAN, 2006; OJIMA, 2007), e também pela sua área mais formal, com estudos sobre fecundidade (PANDIT; BAGCHI-SEN, 1993; WEEKS *et al.*, 2004), mortalidade (BALK *et al.*, 2004b; JAMES *et al.*, 2004) e migração (JOHNSON *et al.*, 2005). Não podemos nos esquecer da avaliação demográfica mais comum e sempre atual - o mapeamento da distribuição populacional – que apresenta um grande número de trabalhos desenvolvidos utilizando diferentes técnicas e abordagens.

Apesar de algumas iniciativas isoladas, como o simbólico estudo de John Snow em 1854, apenas no início do século XX podemos perceber um movimento mais claro em direção à incorporação do espaço nas investigações acadêmicas. Em 1920, alguns sociólogos e geógrafos da Universidade de Chicago, notadamente Robert Park e Ernest Burgess, começaram a explorar as interferências causadas pelo espaço urbano no comportamento social, buscando estabelecer relações de causalidade. Daí surgiu a Teoria das Zonas Concêntricas<sup>2</sup> (PARK; BURGESS; MACKENZIE, 1925). Essas pesquisas envolviam a confecção de mapas, cujo objetivo era revelar a distribuição espacial dos problemas sociais e permitir a comparação entre áreas diferentes. Burgess tinha um grande interesse

---

<sup>1</sup> O *urban sprawl* usualmente é associado a um desenvolvimento urbano pouco ou não planejado na periferia das cidades, possivelmente à custa de terrenos rurais (EWING *et al.*, 2002).

<sup>2</sup> O modelo de Burgess explica a organização das cidades através de zonas concêntricas distintas. A parte central da cidade, a mais interna dos círculos, é constituída pelo bairro comercial e de negócios. As demais zonas são formadas por áreas de agrupamento residencial de populações de diferentes classes sociais, além de zonas de transição e áreas reservadas para a expansão da cidade (PARK; BURGESS; MACKENZIE, 1925).

por mapas e os usava extensivamente; ele também estava sempre à procura de dados, tendo sido um dos pesquisadores da área social daquela época que mais utilizou dados censitários (BULMER, 1984). Para grande parte dos cientistas sociais esta espacialização ecológica foi longe demais e após a Segunda Guerra Mundial a Escola de Chicago perdeu a sua influência, mantendo-se viva apenas em algumas áreas da economia urbana e da geografia urbana. A maior crítica com relação a esse modelo diz respeito à não incorporação das complexas dimensões sociais e culturais envolvidas no processo de urbanização, o que o tornaria muito simplista (SOJA, 2009).

No período pós-Segunda Guerra Mundial, surge a Nova Geografia que desenvolve metodologias derivadas das ciências exatas, com um claro predomínio da abordagem espacial, com o surgimento de inúmeros métodos de análise, destacando-se entre eles a estatística espacial (CHRISTOFOLETTI, 1976).

Neste contexto espacial, o papel dos SIGs merece ser destacado devido a sua enorme associação com o tema. Apesar de ter surgido nos anos 60, apenas nos anos 80 é que os SIGs realmente se firmaram no mercado e na academia, principalmente devido à criação de centros de pesquisa voltados para o desenvolvimento, utilização e ensino do geoprocessamento, como o NCGIA – *National Centre for Geographical Information and Analysis* -, localizado na Califórnia, USA. Ainda nesta mesma década, os SIGs se beneficiaram da popularização e barateamento dos computadores pessoais e do desenvolvimento de tecnologias digitais, como a criação de bancos de dados relacionais e de ferramentas de análise espacial. Até então os SIGs estavam confinados a um grupo especializado de utilizadores que tinham conhecimentos para manipular os aplicativos e também tinham conhecimentos de cartografia. Apenas após a sua popularização, a partir dos anos 90, pessoas de outras áreas se aventuraram a descobrir as ferramentas e funções que ele proporciona e explorar as possibilidades de integração de dados e realização de análises espaciais (CÂMARA; DAVIS; MONTEIRO, 2001). No final dos anos 90, surge a Ciência da

Informação Geográfica ou *GIScience*, que é a ciência por detrás da tecnologia das informações espaciais, sendo essencialmente multidisciplinar, englobando a cartografia, geodésia, fotogrametria, estatística espacial e outras ciências relacionadas com informações e tecnologias geoespaciais (GOODCHILD, 1997).

Mais recentemente, principalmente nas duas últimas décadas, a utilização de dados espaciais ultrapassou os muros das universidades e atingiu a vida cotidiana. Podemos distinguir como causas para esta emergência os avanços na área da computação, com o aumento da capacidade dos computadores, o aumento da quantidade e qualidade de programas que trabalham com dados espaciais, o desenvolvimento de múltiplas técnicas de análise, e também o aumento extraordinário da quantidade de dados espaciais coletados e disponibilizados.

Podemos perceber que os assuntos geoespaciais tomaram proporções inimagináveis na atualidade, o que levou Taylor (2010) a dizer que estamos testemunhando o surgimento da Era da Localização. Apesar do interesse pelo espaço nos estudos analíticos ter se iniciado há mais de um século, esta nova era parece coincidir com a proliferação do uso de aparelhos com sistemas de localização (TAYLOR, 2010), como automóveis e telefones celulares, bem como com a disponibilização de serviços de mapas na internet (RUMSEY, 2009), como o *Google Earth* e o *Google Maps*, que surgiram em 2005.

### 1.1.2 Informações Espaciais e Demografia

De uma maneira geral, os cientistas sociais sempre deram mais importância aos motivos que levam aos acontecimentos sociais do que ao local onde eles ocorrem (WEEKS, 2004). Diante de uma série de variáveis com uma significância aparentemente maior, a localização geográfica era simplesmente vista como algo menor (GUTMANN; STERN, 2007), ou óbvia demais para merecer atenção (DAHMS, 2009) ou até mesmo incidental (ZELINSKY, 1966).

Entrando, num campo mais específico das Ciências Sociais, a Demografia, alguns pesquisadores afirmam que esta ciência tem uma longa tradição espacial (VOSS, 2007; WEEKS, 2004). Isso seria justificado pelo fato de que a Demografia se interessa pelo homem não como indivíduo, mas como um grupo, seja um grupo familiar ou um grupo mais amplo ou população (KOSINSKI, 2000). E partindo desta definição, poderia se dizer que como esses grupos estão inseridos em uma região geográfica, a demografia seria uma ciência essencialmente espacial. No entanto, essa justificativa nos permite concluir apenas que havia uma consciência do espaço por parte dos demógrafos (WEEKS, 2004). Esta consciência pode ser vista nos estudos em que basicamente eram feitas comparações de alguma característica demográfica entre regiões geográficas diferentes, sendo que essas regiões podiam ser divisões administrativas (países, estados, etc.) ou não administrativas (áreas urbanas e rurais, por exemplo). Esta consciência espacial também está presente nos estudos de migração, já que este fenômeno envolve um lugar de origem e um de destino; está presente nas explicações para o declínio da fecundidade a partir da difusão de ideias que se dá não apenas de uma pessoa para outra, mas também de um local para outro; está presente nas diferenças das taxas de mortalidade que estão relacionadas com a presença ou com a ausência de determinados fatores de risco em locais específicos (CASTRO, 2007).

A análise comumente realizada em demografia era, e ainda é, a comparação de características e tendências em áreas distintas. Estas análises comparativas nos estudos demográficos continuam até hoje, agora com a utilização de uma vasta gama de dados digitais e espaciais, bem como com a utilização de ferramentas específicas para a manipulação desses dados, como os SIGs. Desta maneira, a visualização de mapas mostrando variáveis demográficas se tornou uma grande ferramenta descritiva, capaz de revelar padrões impossíveis de serem percebidos em tabelas. No entanto, esta ampla utilização de mapas e a percepção de padrões espaciais não significa que o espaço foi incorporado analiticamente nos estudos de população. Para que isso seja possível há a

necessidade de criação de modelos onde o espaço tenha uma relação causal explicitada em relação aos fenômenos demográficos, que é uma das maiores dificuldades para a incorporação definitiva e efetiva do espaço nas análises (NICHHD, 2002).

Um fato que merece ser destacado é que a demografia sempre explorou a questão temporal nas suas análises, colocando o tempo em termos de idade, período e coorte, mas, nunca desenvolveu nada semelhante com relação ao espaço (MOUW, 2000). Isto se deve principalmente ao grande número de elementos relacionados entre si que estão presentes no espaço (BARCELLOS; MACHADO, 1998) e que adicionam um grande grau de dificuldade nesta abordagem.

A partir do final do século passado, surge a necessidade de integrar as questões relacionadas à população, meio ambiente e desenvolvimento sustentável, que estão intimamente entrelaçadas e colocam problemas desafiadores referentes a teorias e métodos (MARTINE, 2007; HOGAN, 2007). Há a necessidade de novas teorias para lidar com a complexidade do assunto, e também de novos métodos para lidar com as extensões espacial e temporal (LUTZ; PRSKAWETZ; SANDERSON, 2002). Basicamente nesta época, começa, então, a utilização de metodologias não utilizadas tradicionalmente nas Ciências Sociais, como, por exemplo, o Sensoriamento Remoto, para ajudar na compreensão das dimensões humanas no contexto das mudanças ambientais globais, área em que a componente espacial é extremamente forte (STERN; YOUNG; DRUCKMAN, 1992; LIVERMAN *et al.*, 1998).

Assim, além da crescente e justificada necessidade, o desenvolvimento de novas metodologias de coleta de dados, a farta disponibilização de dados espaciais acurados e detalhados e a criação de ferramentas para manipular esses dados, também impulsionaram a incorporação do espaço nos estudos sociais e demográficos (CASTRO, 2007; MATTHEWS, 2003; VOSS, 2007; WACHTER, 2005). Segundo Logan, Zhang e Xu (2010) este fenômeno foi mais evidente em áreas que trabalhavam utilizando dados agregados em algum tipo de unidade

geográfica, que é o caso de alguns pesquisadores no campo da Demografia que já tinham alguma afinidade com SIGs. O conhecimento das relações entre as pessoas e seus ambientes físicos e sociais possibilitou novas visões e entendimentos dessas complexas interações, levando a grandes avanços científicos que, provavelmente, vão continuar acontecendo ainda por muito tempo. A incorporação de dados de Sensoriamento Remoto e de técnicas de econometria espacial são estratégias que estão sendo amplamente utilizadas para estabelecer a compreensão do relacionamento entre as atividades humanas e as mudanças ambientais (MATTHEWS, 2003), como pode ser visto em Liverman *et al.* (1998).

Apesar de todos os avanços em direção à incorporação de informações e análises espaciais nos estudos populacionais, não se pode afirmar que isso seja definitivamente uma realidade. Na década passada, Matthews (2003) já notava que a presença de temas espaciais em conferências como a *Population Association of America Annual Meeting* era cada vez maior. Mas, ao contrário do que era de se esperar, esta crescente popularidade não levou uma grande quantidade de pesquisadores a incorporar as questões espaciais aos seus estudos. Uma das possíveis explicações para esta ocorrência é o fato de que apenas poucos pesquisadores conseguem entender e manipular dados espaciais, e, conseqüentemente, usá-los de maneira integrada nas suas pesquisas demográficas (MATTHEWS, 2003). O interesse pelas questões espaciais no âmbito das Ciências Sociais foi acompanhado pelo reconhecimento de se ter que criar e aplicar uma agenda de treinamento específica, tanto para estimular o raciocínio espacial, quanto para ensinar a utilizar ferramentas de manipulação de dados e de execução de análises espaciais e geoestatísticas (MOUW, 2000; NICHHD, 2002; MATTHEWS, 2003; GOODCHILD; JANELLE, 2004; RUMSEY, 2009; MATTHEWS; JANELLE; GOODCHILD, 2011).

Podemos concluir com o que foi visto neste item que o desenvolvimento da área de Tecnologia da Informação é essencial para o aumento do interesse por parte dos pesquisadores pelos assuntos relacionados com informação e localização espacial. Porém, algumas dificuldades ainda permanecem e são

apontadas como motivos para a não incorporação de informações espacializadas nos estudos sociais. Uma dessas dificuldades é decorrente da inadequação entre a unidade geográfica em que os dados demográficos são apresentados e a unidade geográfica de interesse dos pesquisadores (BALK; YETMAN, 2004; GRASLAND; MADELIN, 2006; GOODCHILD; ANSELIN; DEICHMANN, 1993; GOODCHILD; LAM, 1980; HOGAN, 1992, 2001; KAMPEL, 2004; LINHARES, 2004; OPENSHAW, 1996, MARTINE; SCHENSUL, 2013). Outra dificuldade também relatada com frequência na bibliografia é a limitação à realização de estudos utilizando séries temporais de dados censitários, tendo em vista que os limites dos setores censitários se alteram de um censo para outro (HOGAN; D'ANTONA; CARMO, 2008; MARTINE; SCHENSUL, 2013). Essas questões relacionadas com a unidade espacial de análise serão o foco do próximo item.

## **1.2 A geografia dos censos**

Os censos de População e Habitação podem ser definidos como sendo o processo de coletar, compilar, avaliar, analisar e publicar os dados estatísticos, econômicos e sociais, relativos a todas as habitações e seus residentes, em um país ou região delimitada de um país (NAÇÕES UNIDAS, 2008).

Para a realização desta operação faz-se necessária a delimitação de unidades de trabalho, ou seja, unidades territoriais definidas para a coleta de dados. Essas unidades territoriais são conhecidas como áreas de enumeração ou setores censitários. Essas áreas serão atribuídas a um recenseador, que terá a tarefa de realizar a coleta de dados em todas as edificações ali incluídas, em um determinado período de tempo (IBGE, 2003). Tomando como base o modelo de censo adotado pela maioria dos países, entre eles, o Brasil, pode-se constatar que no interior de cada uma dessas áreas de enumeração estão localizados os domicílios e dentro deles os indivíduos. Desta maneira, no que se refere ao Censo de População, a unidade básica é o indivíduo e, em um nível superior, a família, e no que se refere ao Censo de Habitação, é o domicílio. No entanto, em termos de

localização espacial, o domicílio é a unidade básica, uma vez que é a menor unidade de referência geográfica para os indivíduos residentes num mesmo domicílio (XIE, 2006). A agregação destas unidades básicas gera o primeiro nível de unidades geográficas da operação censitária, que nada mais é do que a unidade operacional de coleta, denominada setor censitário ou área de enumeração.

As unidades operacionais de coleta, mesmo sendo criadas *a posteriori*, podem ser vistas como a origem de um sistema hierárquico de unidades superficiais que são utilizadas para divulgar os dados censitários. Realizando operações de agregação sequenciais, cada nível dará origem a um nível superior, gerando a estrutura hierárquica geográfica completa do censo. Na prática, os setores censitários são criados de maneira que um conjunto dessas unidades operacionais preencha completamente a superfície do menor nível administrativo municipal permitindo, assim, a criação de uma malha hierárquica aninhada de unidades.

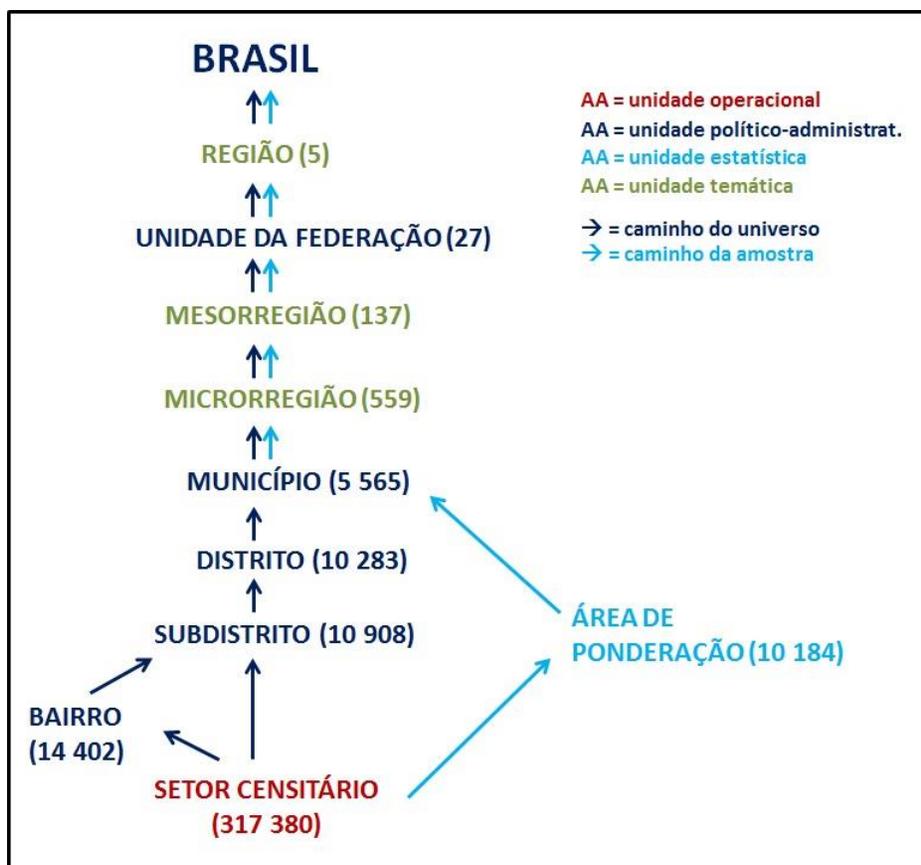
A Figura 1 ilustra a geografia básica do censo no Brasil utilizada em 2010. Um dos caminhos, o caminho do questionário curto, se inicia com os setores censitários e vai em direção às unidades político-administrativas e temáticas até atingir o nível mais alto que é o país. Outro caminho, o caminho do questionário longo, tem início no mesmo ponto e daí segue para uma unidade estatística resultante da agregação de setores censitários - as áreas de ponderação -, depois vai em direção a uma unidade político-administrativa de nível hierárquico mais alto - o município - e, então, segue coincidente com o caminho anterior. As quantidades de cada unidade são apresentadas ao lado do seu nome. Podemos ver que a ordem de grandeza do número de setores censitários (300 mil) é bastante superior às demais unidades, indicando que eles representam pequenas porções do território. Deve-se ressaltar que a quantidade de bairros existentes (14.402) é extremamente pequena em relação à quantidade de municípios (5.565). A explicação para esta diferença reside no fato de que o IBGE incorpora

apenas as divisões intramunicipais oficiais e estas são realmente pouco frequentes no Brasil.

Cabe ressaltar que nem todas as unidades geográficas do censo necessitam ter um papel administrativo (NAÇÕES UNIDAS, 2008), como as Unidades da Federação ou os Municípios na Figura 1. Algumas dessas unidades podem exercer apenas um papel estatístico, não sendo utilizadas para a administração do território, mas apenas para divulgação de dados, como no caso dos Setores Censitários e das Áreas de Ponderação.

Desta maneira, podemos constatar que existem diversas unidades geográficas utilizadas na pesquisa censitária, e que elas podem causar algumas dificuldades na execução de análises, como veremos no próximo item.

Figura 1 - Geografia do censo no Brasil, 2010.



Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

### 1.2.1 Unidades geográficas

As unidades geográficas, de uma maneira geral, variam em forma e em dimensão e, para avaliar o seu grau de particionamento Tobler *et al.* (1997) desenvolveram uma medida que nos dá uma ideia da precisão espacial envolvida na divisão de uma região em unidades. Esta medida, denominada resolução espacial, pode ser utilizada como parâmetro de comparação entre diferentes unidades de uma mesma região ou entre diferentes regiões divididas em unidades semelhantes. Uma maneira simples de entender o significado de resolução espacial é pensar em termos de granularidade, assim, quanto melhor a resolução espacial, maior a quantidade de “grãos” de pequeno tamanho dentro da região de interesse. Podemos compreender esta medida também como sendo a escala espacial do particionamento, uma vez que a escala espacial é definida pela quantidade e pelo tamanho das unidades utilizadas para dividir uma região de interesse (GOTWAY; YOUNG, 2004).

O cálculo da resolução espacial é feito pela extração da raiz quadrada do quociente entre a área de uma região e o número de unidades que compõem esta região. Isso significa que uma região dividida em muitas unidades geográficas pequenas tem uma resolução espacial melhor do que uma região de extensão equivalente, mas dividida em poucas unidades geográficas de grande extensão. A Tabela 1 mostra os valores da resolução espacial para o estado do Pará, considerando-se diferentes unidades geográficas. Cabe ressaltar que esta resolução é um valor médio, pois as unidades geográficas têm formas irregulares e dimensões variadas.

Podemos observar que, à medida que a quantidade de unidades aumenta, a resolução espacial diminui, já que a área do estado permanece a mesma. Uma resolução espacial maior não significa que o número que expressa a medida seja grande, pelo contrário, o número é pequeno; o termo maior está relacionado com a granularidade ou a quantidade de unidades, e não com o valor numérico da medida.

Tabela 1 - Resolução espacial das unidades geográficas, Pará e São Paulo, 2010.

Unidade geográfica	Área (km <sup>2</sup> )		Número de unidades		Resolução espacial (km)	
	PA	SP	PA	SP	PA	SP
Estado	1 247 955	248 223	1	1	1 117,12	498,22
Município	1 247 955	248 223	143	645	93,42	19,62
Setor Censitário	1 247 955	248 223	8 933	68 296	11,82	1,91

Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Seguindo o raciocínio apresentado acima, para melhorar ou aumentar a resolução espacial dos dados demográficos oriundos de pesquisas censitárias seria necessário alterar a unidade geográfica utilizada para a sua disseminação, tornando-a menor e, conseqüentemente, mais próxima da unidade básica de coleta dos dados (o domicílio). Esta melhoria na resolução tem diferentes objetivos, um deles é obter dados estatísticos para pequenas áreas, possibilitando análises mais detalhadas e acuradas. Outro objetivo igualmente importante é facilitar a transferência dos dados para um recorte espacial diferente daquele para o qual os dados estão disponibilizados. As diferenças entre a unidade de origem, ou a unidade em que os dados são disponibilizados, e a unidade de destino, ou a unidade na qual se deseja ter os dados, tem sido alvo de antigas e incessantes discussões por parte dos usuários de dados e pesquisadores (GOODCHILD; LAM, 1980; GOODCHILD *et al.*, 1993; OPENSHAW, 1996; HOGAN, 1992, 2001; MARTIN, 2003; KAMPEL, 2004; LINHARES, 2004; GRASLAND; MADELIN, 2006; OJIMA; MARTINE, 2012; MARTINE; SCHENSUL, 2013).

Na fase de disseminação, os dados coletados no nível individual e/ou domiciliar e já agregados em setores censitários – as unidades de origem -, precisam ser agregados novamente em outras unidades para atender aos objetivos dos governos com relação à sua utilização, quais sejam os de direcionar o planejamento e a implantação de políticas públicas, além do monitoramento das

mesmas. Essas unidades geralmente são definidas por critérios político-administrativos e podem ser consideradas as unidades de destino, no caso da administração pública. Historicamente, as agências de estatística sempre divulgaram dados obedecendo às unidades político-administrativas (BACKER, 2008; GUZMAN; SCHENSUL; ZHANG, 2013), procurando atender as demandas municipais, nacionais e internacionais por dados que possam direcionar e monitorar as políticas públicas (BUENO; DAGNINO, 2011).

No entanto, se o objetivo da análise dos dados censitários está relacionado com questões ambientais, por exemplo, a unidade geográfica deve permitir a compreensão das relações demográficas e ambientais, devendo, portanto, ser um recorte mais próximo de alguma unidade ambiental, como as bacias hidrográficas (HOGAN, 1991). Já no caso de a análise estar relacionada com a vulnerabilidade a algum evento físico, como deslizamento de encosta ou inundação, outra unidade geográfica deverá ser utilizada. Infelizmente, temos que reconhecer que raramente os dados demográficos são reportados em unidades relacionadas com o meio físico ou ambiental (DEICHMANN, 1996). Podemos citar outro exemplo, como estudos que devem ser elaborados para alguma área geográfica específica, como a área de cobertura de estabelecimentos de saúde ou de educação, cuja finalidade poderia ser a avaliação dos Objetivos de Desenvolvimento do Milênio<sup>3</sup>, que apesar de serem globais e nacionais, necessitam de dados agregados em unidades cuja resolução seja compatível com a realidade local. Indiscutivelmente, nenhuma dessas unidades de destino citadas acima coincide com alguma divisão político-administrativa. E não deveriam mesmo coincidir, já que essas unidades têm sua delimitação feita por critérios ambientais, físicos, gerenciais ou de acessibilidade, e não por critérios relacionados com a administração do território. Grande parte das agências de estatística não incorporou essa demanda de disponibilização de dados em diferentes recortes

---

<sup>3</sup> Em setembro de 2000, as nações integrantes da Organização das Nações Unidas - ONU firmaram um compromisso para combater a extrema pobreza e outros males da sociedade. Esta promessa acabou se concretizando nos 8 Objetivos de Desenvolvimento do Milênio que deverão ser alcançados até 2015.

espaciais não político-administrativos na sua rotina de trabalho. Mesmo as demandas oriundas da área ambiental não foram atendidas, apesar de apresentarem importância e demanda cada vez maiores.

Analisando esta questão do ponto de vista das agências de estatística, podemos ver que existem justificativas para este comportamento. Cada setor de serviços ou de planejamento tem suas próprias necessidades e costuma utilizar unidades geográficas específicas, delimitadas para satisfazer o seu próprio gerenciamento e planejamento. No caso da área acadêmica a questão é ainda mais complicada, pois cada pesquisador gostaria de ter os dados agregados para uma determinada região. Segundo Judson (2007), o grande problema relacionado com dados no século XXI é que, do ponto de vista dos usuários, as agências de estatística deveriam disponibilizar os dados com maior rapidez, agregados em níveis geográficos cada vez menores e devidamente integrados ou com possibilidades de integração com outros tipos de dados. Na verdade, seria uma tarefa bastante árdua gerar produtos estatísticos de disseminação que atendessem a todas essas demandas, considerando-se os recursos humanos, financeiros e o tempo de execução. E sem esquecer os requisitos técnicos que precisam ser atendidos.

Um problema que decorre da utilização de unidades operacionais como unidades de disseminação dos dados é a dificuldade de realizar comparações temporais. Como mencionado anteriormente, as unidades operacionais tem o objetivo de facilitar o planejamento e o gerenciamento da operação censitária, não sendo criadas objetivando a disseminação de dados ao público. Assim, de um censo para outro, essas unidades costumam ser alteradas, e geralmente esta alteração se faz pela divisão de um setor em dois ou mais setores. Isso se dá principalmente em áreas de expansão urbana, onde a densidade populacional aumenta dia após dia e onde a transformação de áreas rurais em áreas urbanas ocorre frequentemente.

Apesar disso, a publicação de dados agregados em unidades operacionais ocorre em diversos países, inclusive no Brasil, enquanto em alguns

países existem unidades derivadas das unidades operacionais para a divulgação, como USA, Canadá e Austrália (NAÇÕES UNIDAS, 2007b), e em outros, existem unidades criadas especificamente para esta finalidade, como Inglaterra e País de Gales (ONS, 2012). Em todos os casos, mas, mais frequentemente no primeiro, a consequência direta para o usuário é a dificuldade de comparação temporal dos dados, já que as unidades se alteram com o passar do tempo (EXETER *et al.*, 2005; NORMAN; REES; BOYLE, 2003) e há a necessidade de se realizar uma compatibilização entre as unidades. Isso, no entanto, não impede que a análise seja efetuada, desde que as informações de comparabilidade, que permitem que esta compatibilização seja realizada, sejam divulgadas juntamente com os dados. Na prática, mesmo havendo a divulgação das informações que permitam a comparabilidade temporal das unidades geográficas, a realização desta compatibilização é uma tarefa complexa, árdua e demorada.

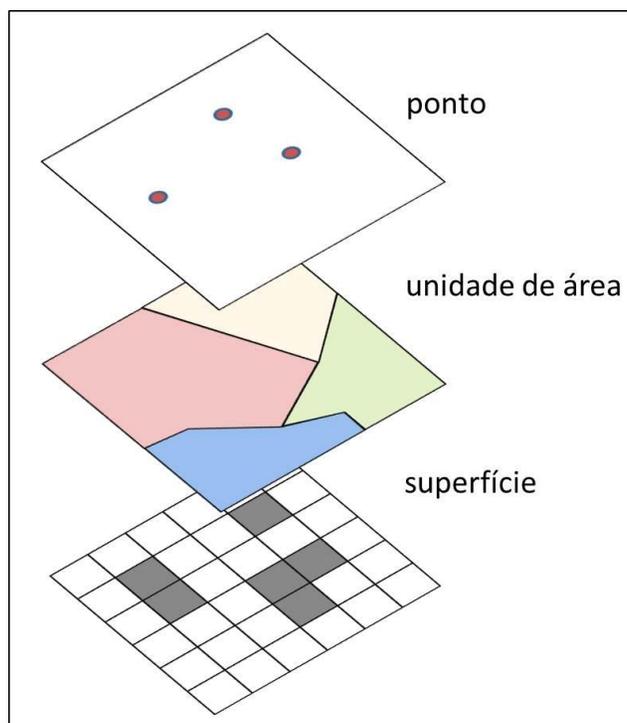
As unidades político-administrativas também são passíveis de alteração ao longo do tempo, havendo tanto a divisão quanto a união de unidades. Essas alterações também ocasionam dificuldade em análises temporais e, de certa maneira, até maiores do que no caso dos setores censitários, já que nem sempre existem informações que permitem a compatibilização das unidades antigas com as novas.

### 1.2.2 Representação espacial

Como vimos anteriormente, as unidades básicas dos censos demográficos são as pessoas e os domicílios. A representação espacial dessas unidades, realizada com a utilização de ferramentas de geoprocessamento, é feita através de objetos discretos vetoriais representados por pontos. No caso dos domicílios, a representação é direta, com a utilização das coordenadas geográficas que melhor representam espacialmente o objeto; no caso dos indivíduos, a representação é indireta, sendo utilizada a representação do

domicílio como uma aproximação. A Figura 2 ilustra as três possibilidades de representação espacial de dados: ponto, unidade de área e superfície.

Figura 2 - Representação espacial de dados.



Fonte: Elaborado pelo autor.

A utilização desses dados individualizados e representados como pontos é a situação ideal para a realização de qualquer pesquisa ou análise envolvendo localização geográfica e dados censitários. Primeiramente, por ser a escala mais apropriada para observar e modelar fenômenos sociais, mas também porque oferece a possibilidade de trabalhar em múltiplas escalas, utilizando agregações sucessivas para alcançar escalas diferentes e em qualquer tipo de compartimentação espacial, com a agregação dos pontos e suas características associadas em qualquer recorte territorial desejado (GRASLAND; MADELIN, 2006). Devido a questões relacionadas com a proteção da identidade dos informantes ou confidencialidade, como é comumente conhecida no meio

estatístico, isso não é possível, e os dados básicos devem ser agrupados em alguma unidade de área antes que sejam liberados para o público. Independente de qual seja a unidade de área utilizada, a forma de representação espacial mais comum é a forma vetorial, com a utilização de polígonos para representar as unidades nas quais os dados básicos são agregados.

Os dados básicos também podem ser transformados em taxas ou indicadores a partir de manipulações dos dados discretos (pontos) ou dos dados de área (polígonos) e, então, ser representados como superfícies. Esta forma de representação é tecnicamente mais apropriada a esses dados, por representar um modelo de distribuição contínua no espaço (KAMPEL, 2004). No entanto, ela não é comumente utilizada em produtos de disseminação de dados estatísticos. O motivo para isso é que, nestes casos, o tipo de dado que está sendo disseminado é uma contagem das variáveis investigadas, que é um dado discreto. Após a agregação desses dados discretos em alguma unidade geográfica, a forma mais usualmente utilizada para representá-los é através de unidades de área (CARVALHO *et al.*, 2004).

### 1.2.3 Mudanças de suporte espacial

Todas as variáveis existentes na base de dados censitários são passíveis de mapeamento através da utilização das mais variadas técnicas de representação cartográfica, embora os mapas de densidade populacional sejam, com certeza, os mais comumente elaborados e discutidos. Isso se deve não apenas à importância da distribuição da população sobre o terreno em diferentes áreas, como, por exemplo, economia e planejamento urbano, mas, também porque a densidade populacional é uma medida intrinsecamente espacial, e, como tal, a melhor forma de representação é através de mapas. A técnica mais tradicionalmente utilizada para a elaboração de mapas de densidade é a coroplética, que é a divisão dos dados em faixas de valores e sua respectiva associação com cores. No entanto, esta forma de representação apresenta alguns

problemas (LANGFORD; UNWIN, 1994). Um desses problemas está relacionado com o tamanho das unidades, o que pode levar a uma interpretação errônea do mapa. Outro problema existente está relacionado com a utilização de dados coletados individualmente e agregados em unidades zonais. Como consequência, para cada alteração na forma ou limite dessas unidades teremos um mapa final com aparência totalmente distinta, uma vez que com essas alterações as unidades podem se enquadrar em outra faixa de valores.

Esses problemas e vários outros surgem basicamente devido à mudança de suporte espacial dos dados. Toda informação geográfica está associada a um conjunto de pontos de localização que são modelados utilizando-se objetos. Esses objetos e as métricas utilizadas para medir a interação entre eles são conhecidos como suporte espacial da informação (OLEA, 1991). Como exemplo de suporte espacial de pontos, podemos citar a localização (latitude e longitude) dos domicílios em uma pesquisa domiciliar; para suporte espacial de área, temos como exemplo os setores censitários e as divisões municipais. A mudança de suporte de uma variável cria uma nova variável, que apesar de estar relacionada com a original, tem propriedades estatísticas e espaciais diferentes (GOTWAY; YOUNG, 2002).

Todos os problemas relacionados com mudança de escala, agregação, desagregação e inferência entre unidades de suporte diferente fazem parte de um grupo de problemas que é denominado em geoestatística de problemas de mudança de suporte ou *change of support problem* - COSP (GOTWAY; YOUNG, 2002, 2004, 2007; PLUMEJEAUD *et al.*, 2010). No contexto dos censos, estamos basicamente restritos a dois tipos de mudança de suporte: de ponto (dados domiciliares e individuais) para área (dados agregados em unidades geográficas) e de área (setores censitários, por exemplo) para área (outro recorte espacial qualquer).

A revisão bibliográfica apontou dois problemas principais que estão relacionados com a agregação de dados individualizados em áreas: o problema da unidade de área modificável e a falácia ecológica. Esses problemas são

considerados como inerentes aos dados agregados em áreas e, apesar de serem de difícil detecção e mensuração e, conseqüentemente, de difícil correção ou eliminação, eles não devem ser ignorados (OPENSHAW, 1984; GOTWAY; YOUNG, 2002; CARVALHO *et al.*, 2004; GRASLAND; MADELIN, 2006, FLOWERDEW, 2011). Ambos podem ser vistos como casos específicos de problemas de mudança de suporte espacial (YOUNG; GOTWAY, 2007) e veremos, em seguida, com mais detalhes, suas causas e efeitos.

### 1.2.3.1 Problema da unidade de área modificável

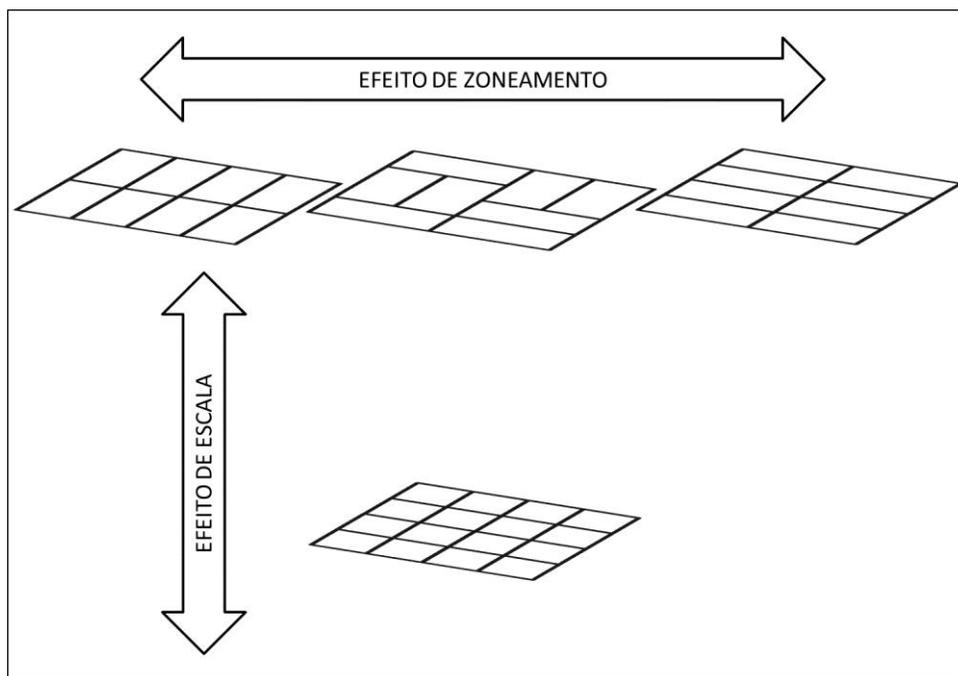
Este problema foi primeiramente detectado por Gehlke e Biehl (1934) e foi amplamente estudado por Openshaw (1984) e por Openshaw e Taylor (1979, 1981), sendo que este último criou o nome pelo qual o problema é conhecido - *Modifiable Area Unit Problem* ou MAUP.

O MAUP envolve dois efeitos (WONG, 1996), como pode ser visto na Figura 3. O efeito de escala, como o próprio nome sugere, é devido à mudança de escala, o que na prática significa que a quantidade de unidades é alterada. Na ilustração, a área de estudo na parte inferior apresenta dezesseis unidades, enquanto na parte superior temos apenas oito unidades, o que configura uma mudança de escala. O efeito de zoneamento ocorre numa mesma escala, e a quantidade de unidades não se altera; a alteração acontece no arranjo das unidades que são agrupadas gerando configurações diferentes.

O efeito de escala está relacionado com diferenças nas medidas estatísticas obtidas para uma ou mais variáveis quando as mesmas são agrupadas em unidades com diferentes. Este efeito está ilustrado na Figura 4. Podemos observar que o valor da média se mantém constante, enquanto a variância diminui à medida que a agregação aumenta. Desta maneira, temos um efeito de suavização, havendo perda de heterogeneidade e uma conseqüente distorção nos dados. Este efeito está presente, por exemplo, quando realizamos análises demográficas em diferentes níveis de unidades político-administrativas,

como municípios e estados. Os resultados obtidos para os municípios podem ser bastante diferentes daqueles obtidos para estados, não significando que há uma real variação nos dados, já que esta diferença pode ser causada pelo efeito de escala.

Figura 3 - Aspectos do MAUP.



Fonte: Adaptado de WONG, 1996.

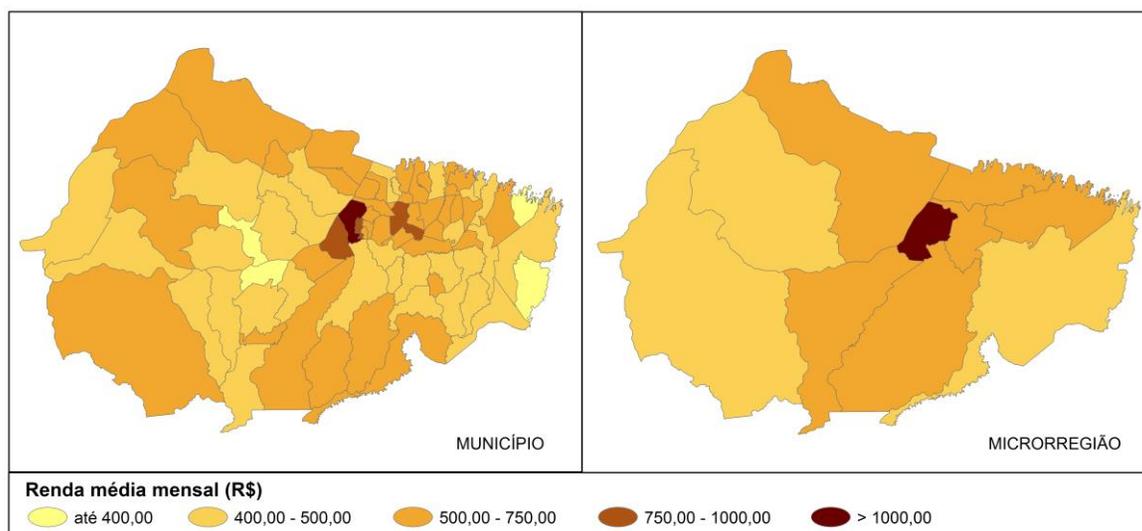
Figura 4 - MAUP: Efeito de Escala.

2	4	6	1	3	3.5	3.75	3.75
3	6	3	5	4.5	4		
1	5	4	2	3	3		
5	4	5	4	4.5	4.5	3.75	3.75
$\bar{x} = 3,75$ $\sigma^2 = 2,60$				$\bar{x} = 3,75$ $\sigma^2 = 0,50$		$\bar{x} = 3,75$ $\sigma^2 = 0,00$	

Fonte: Adaptado de JELINSKI; WU, 1996.

A Figura 5 mostra dois mapas feitos para a mesma variável – renda média mensal por habitante. No caso do mapa por municípios, podemos ver uma variação da variável ao longo do território, havendo concentrações de valores na faixa de 500,00 a 750,00 nas partes norte e sul e na faixa de 400,00 a 500,00 na parte central. Quando observamos o mapa feito por microrregiões, não encontramos nenhuma microrregião na faixa de valores até 400,00 nem na faixa 750,00 – 1.000,00. O que aconteceu foi uma homogeneização dos valores da variável quando houve a mudança de escala representada no segundo mapa. Uma maior quantidade de valores foi considerada em cada microrregião, já que temos uma menor quantidade de unidades, mas com dimensões maiores. Isso nos leva a um resultado diferente daquele obtido para os municípios e, em termos estatísticos, isso significa que houve uma diminuição da variância causada pela diminuição na quantidade de unidades.

Figura 5 - Influência da escala na percepção de um fenômeno.

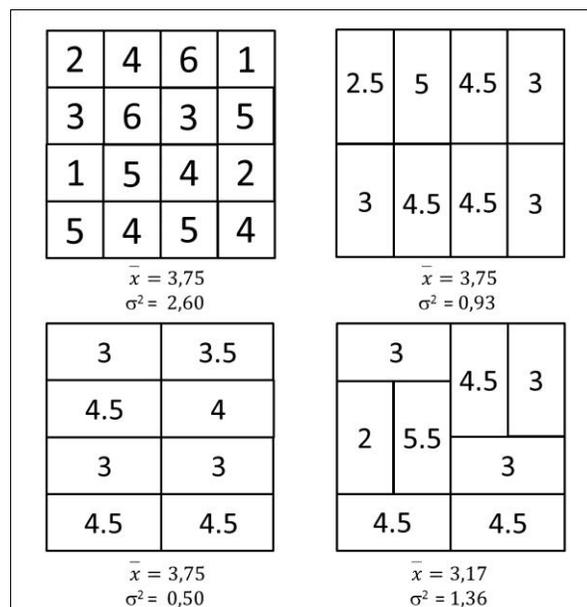


Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

O segundo efeito do MAUP, conhecido como efeito de zoneamento ou de agrupamento, está relacionado com a variabilidade dos resultados estatísticos

obtidos em função das diversas maneiras que as unidades podem ser agrupadas. A Figura 6 exemplifica este efeito. Podemos observar que os agrupamentos foram feitos tanto na direção horizontal, quanto na vertical. A média se manteve igual, enquanto a variância sofreu modificações bem expressivas. Todos os agrupamentos têm a mesma quantidade de unidades – oito –, mas as medidas estatísticas variam substancialmente de um para outro. Quando as variáveis estudadas têm uma autocorrelação espacial<sup>4</sup> alta, o agrupamento provavelmente irá aumentar esta correlação; se as variáveis têm uma autocorrelação espacial baixa, o agrupamento irá diminuir esta correlação. É importante ressaltar que o efeito de zoneamento não é muito pronunciado quando as unidades são agregadas de maneira aleatória no espaço. Isso acontece porque quando unidades contíguas são agregadas, a autocorrelação espacial entre elas se altera e torna o problema mais aparente (GOTWAY; YOUNG, 2002).

Figura 6 - MAUP: Efeito de Zoneamento.

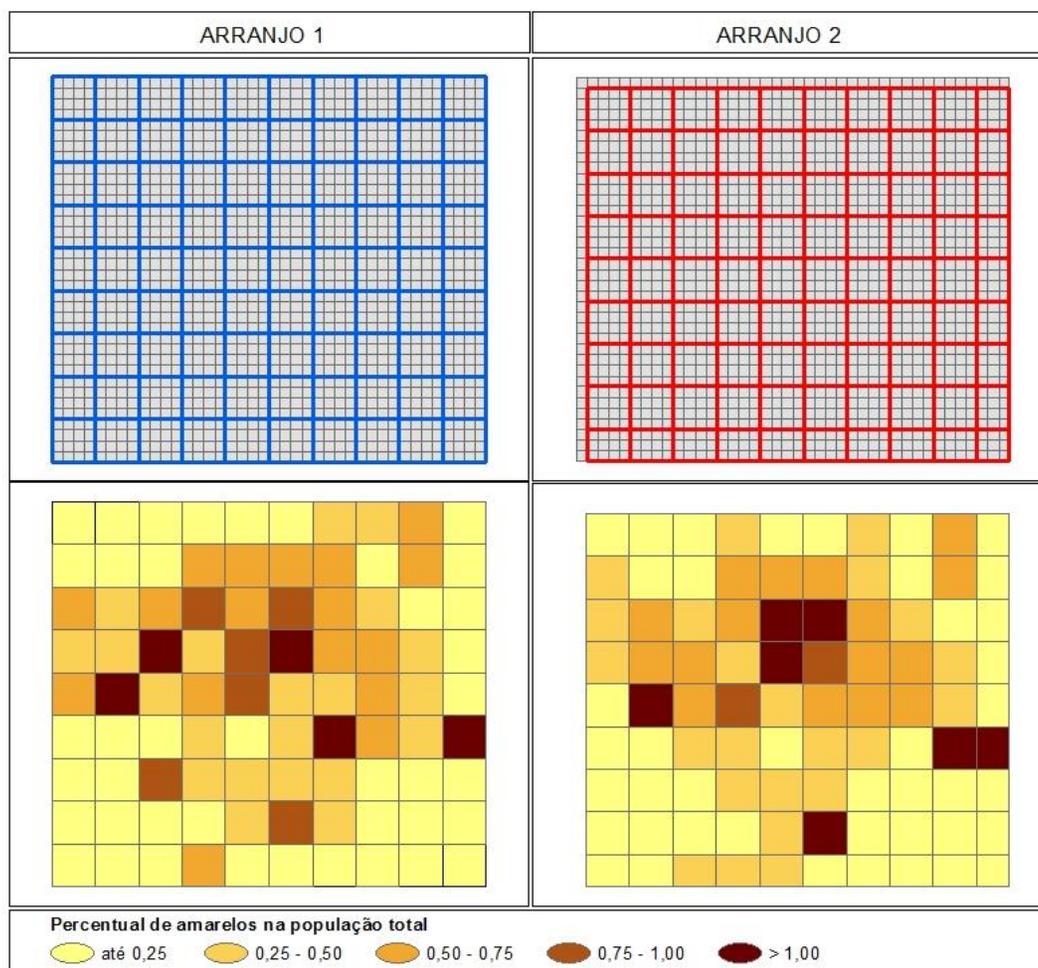


Fonte: Adaptado de JELINSKI; WU, 1996.

<sup>4</sup> Autocorrelação espacial é uma medida do grau em que um conjunto de feições espaciais tende a se apresentar agrupado (autocorrelação espacial positiva) ou disperso (autocorrelação espacial negativa) no espaço.

A Figura 7 mostra um exemplo da influência do zoneamento em mapas temáticos. Os dois mapas mostram a distribuição percentual de população de cor amarela no total da população. No mapa da esquerda, as células foram agrupadas de acordo com o Arranjo 1. No mapa da direita, utilizou-se um agrupamento ligeiramente diferente - o Arranjo 2 -, com um pequeno deslocamento de uma célula na horizontal e na vertical em relação ao Arranjo 1. Os dois mapas apresentam resultados bem diferentes causados pela forma diferente de agrupamento das unidades espaciais iniciais.

Figura 7 - Influência do zoneamento na percepção de um fenômeno.



Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

É importante ressaltar que os dois aspectos do MAUP ocorrem simultaneamente, apesar do efeito de zoneamento ocorrer numa mesma escala, e o efeito de escala pressupor uma alteração de escala (OPENSHAW; TAYLOR, 1979, 1981; OPENSHAW, 1984; WONG, 1996), conforme mostra a Figura 3.

De acordo com o que foi apresentado, podemos dizer que é recomendável que as conclusões de análises que envolvam dados agregados em unidades areais sejam cuidadosamente analisadas para verificar se os resultados são consistentes em outras escalas e outros arranjos espaciais, de modo a descartar a possibilidade de que possam ser resultantes do problema da unidade de área modificável (FOTHERINGHAM, 1998).

#### 1.2.3.2 Falácia Ecológica

Quando temos unidades espaciais nas quais são agregados dados individuais, podemos observar que quanto maior a extensão dessas unidades, menor a variação do fenômeno estudado. Isso acontece porque a média dos valores observados tem uma variação menor do que os próprios valores. Isso ocorre de maneira indiretamente proporcional ao tamanho da unidade de agregação, ou seja, quanto maior a unidade de agregação e, conseqüentemente, maior a quantidade de valores observados, menor a variação da média e, logo, menor a variação do fenômeno em questão. Isso pode levar a um erro inferencial conhecido como “falácia ecológica” (ROBINSON, 1950), que significa que uma análise baseada em dados agregados pode levar a conclusões diferentes daquelas que seriam obtidas com a análise dos dados individuais. Este problema ocorre porque as análises que utilizam dados agregados em áreas não fazem distinção entre as associações espaciais criadas pelo agrupamento dos dados e as associações reais dos dados individuais existentes antes do agrupamento (OPENSHAW, 1984). A suavização decorrente da agregação dos dados é a causa tanto da falácia ecológica quanto do efeito de agregação do MAUP.

Para termos uma ideia mais clara do que é a falácia ecológica, vamos lançar mão de um exemplo. Numa análise utilizando setores censitários, observa-se que existe uma correlação positiva entre as variáveis “número de pessoas com baixo nível de instrução” e “número de domicílio sem esgotamento sanitário”, ou seja, a proporção de pessoas com um baixo nível de instrução aumenta com o aumento da quantidade de domicílios que não apresentam esgotamento sanitário. Esta correlação foi estabelecida no nível de setor censitário, mas não pode, no entanto, ser estendida com um alto grau de certeza para o nível individual, de modo que se possa afirmar que na maioria dos domicílios sem esgotamento sanitário existem moradores com baixo nível de instrução.

#### 1.2.4 Soluções propostas

É fato que os problemas decorrentes da mudança de suporte espacial apresentados acima influenciam os resultados de diversos tipos de análises estatísticas. Reconhecer que isso é um problema nos leva à busca de soluções, ou, pelo menos, a ter consciência de estar trabalhando com números imprecisos. Este é o caminho seguido por alguns pesquisadores e entre as soluções propostas podemos citar a utilização de dados não agregados, a definição de um agrupamento ótimo ou a modelagem com variáveis agrupadas (GRASLAND; MADELIN, 2006). Esta listagem não pretende ser exaustiva, havendo diversas outras soluções citadas na literatura.

A proposta de utilização de dados individuais é com certeza uma solução, mas está longe de poder ser adotada porque, na prática, a disseminação de dados estatísticos oficiais individualizados não é possível devido às questões de confidencialidade.

Entre as propostas de correção ou minimização dos efeitos do MAUP, temos a definição automatizada de agrupamentos, que é feita com o estabelecimento de regras estatísticas e computacionais para gerar soluções ótimas para uma análise específica. Esta solução foi proposta inicialmente por

Openshaw (1977) e foi posteriormente desenvolvida por Openshaw e Rao (1995). Exemplos da utilização desta solução podem ser vistos em Martin (1998), Openshaw e Alvanides (1999), Martin, Nolan e Tranmer (2001) e Cockings e Martin (2005). Esta solução pressupõe um grande conhecimento dos fenômenos envolvidos e o resultado final é válido apenas para o caso particularmente analisado ou outro que possa ser considerado similar. De certa forma, esse é um fator limitante para uma aplicação generalizada, mas não impede a sua utilização como um meio de minimizar os efeitos do MAUP.

Outra proposta para tratar os efeitos do MAUP e também da falácia ecológica é a criação de uma estrutura de modelagem hierárquica com a utilização de grupos de variáveis que são medidas no nível individual e, então, são utilizadas para ajustar as variâncias e covariâncias no nível agrupado (STEEL e HOLT, 1996; HOLT *et al.*, 1996). O principal problema desta proposta está na disponibilidade das variáveis individuais, tendo em vista as questões relacionadas com a restrição de disseminação de dados estatísticos oficiais individualizados.

A partir de metade da década de 2000, começaram a surgir algumas propostas para solucionar os problemas relacionados com mudança de suporte de dados utilizando técnicas de geoestatística (GOTWAY; YOUNG, 2007). A técnica proposta por Gotway e Young (2007) tem por objetivo tratar os problemas derivados de agregação e desagregação espacial de dados (efeito de escala do MAUP), bem como realizar a predição de valores para unidades espaciais sobrepostas. Essa técnica é relativamente recente e existem poucos casos de aplicação publicados para que seja possível avaliar os seus resultados.

Outro caminho também adotado em relação ao MAUP vai em direção contrária ao proposto nos parágrafos anteriores: não encarar o MAUP como sendo um problema, mas como sendo um fato que colabora para aumentar o conhecimento sobre o relacionamento entre as variáveis e as unidades que compõem a análise. Jelinski e Wu (1996) argumentam que o MAUP não deve ser visto como um problema, já que oferece informações importantes para o entendimento dos sistemas reais. Este ponto de vista também é defendido no

relatório final de um estudo técnico realizado pelo *European Observation Network for Territorial Development and Cohesion* – ESPON<sup>5</sup> com o objetivo de compreender e propor soluções para o MAUP (GRASLAND; MADELIN, 2006). Este relatório estimula uma maior exploração dos dados, criando, por exemplo, representações multiescalares do mesmo fenômeno, de modo a expor a sua complementaridade, como anteriormente proposto por Sui (1999) e Mennis (2002). Sui (1999) ressalta que estudos realizados em uma única escala e utilizando uma unidade zonal específica não podem levar a resultados confiáveis, já que não é possível saber exatamente o grau com que os resultados foram afetados pela agregação de dados.

Ainda há outro caminho possível: ignorar o problema. Claramente esta é a opção mais simples e fácil, e a consequência é a existência de incertezas nos resultados das análises. Infelizmente, este caminho é bastante utilizado, já que a mensuração dos efeitos do MAUP é uma tarefa complicada (DARK; BRAM, 2007).

### **1.3 O sistema de grades regulares**

A proposta de criação de um sistema de grades regulares para disseminação de dados estatísticos surge a partir da necessidade de integração de dados de origens diversas e agregados em unidades geográficas incompatíveis, além da necessidade de se ter dados agregados em unidades menores e sem variação ao longo do tempo. Todas essas necessidades se tornam mais evidentes a partir da utilização de ferramentas de geoprocessamento nas análises socioeconômicas e demográficas, devido principalmente à facilidade de se criar unidades artificiais de análise nesses ambientes, como *buffers* baseados em pontos, linhas ou áreas (RUSANEM *et al.*, 2001). Este sistema não é recente, apesar de ainda ser relativamente pouco utilizado e explorado, mas a sua produção tem se tornado cada vez mais comum, tendo em vista as

---

<sup>5</sup> ESPON é um programa que congrega diversas iniciativas relacionadas com o planejamento e integração territorial dos países membros da União Europeia.

tecnologias espaciais atualmente disponíveis e empregadas na realização dos censos, como utilização de GPS, geoprocessamento e sensoriamento remoto (NAÇÕES UNIDAS, 2000; 2009).

Nos itens seguintes, abordaremos as vantagens e desvantagens deste sistema, as técnicas para a sua geração e alguns exemplos de aplicação em nível internacional e nacional.

### 1.3.1 Definição e história

Um sistema de células regulares dispostas em forma de grade pode ser utilizado como uma base para a geração de um suporte geográfico estável para a disseminação de dados. Geográfica e socialmente falando, essas unidades são totalmente arbitrárias, não tendo um significado que possa ser transportado para o mundo real, pois não consideram a distribuição de nenhum processo ou fenômeno subjacente (GRASLAND; MADELIN, 2006). No entanto, as células servem perfeitamente ao propósito de receptáculo, permanecendo estáveis ao longo do tempo, apresentando uma forma regular e simples, com dimensões suficientemente pequenas para atuar como tijolos na construção de qualquer recorte geográfico desejado (NORMAN; REES; BOYLE, 2003; GUZMÁN; SCHENSUL; ZHANG, 2013), além de atenderem as demandas de disseminação de dados para pequenas áreas.

De acordo com Tobler *et al.* (1997), a ideia de utilizar um sistema de grade para disponibilizar dados surgiu primeiramente em 1851 com a publicação de um mapa de observações de baleias com a divisão dos oceanos em quadriláteros de 5°. Este sistema foi utilizado simplesmente por ser o mais conveniente para o local, uma vez que não existem divisões políticas ou administrativas nos oceanos.

Em 1856, um mapa de densidade populacional utilizando grades foi utilizado durante uma pesquisa relacionada com a construção de ferrovias na Suécia (SEHLIN, 2011).

Segundo informações do *Japan Statistics Bureau* (s.d.), o sistema de grade surgiu em 1929 como uma proposta metodológica de análise feita por um geógrafo finlandês chamado Graneau. Em 1969, o Japão utilizou pela primeira vez este sistema para divulgar estatísticas derivadas de diversas pesquisas para a área metropolitana de Tóquio e, desde então, tanto o governo quanto empresas privadas, realizam a coleta e a manutenção de dados regionais neste sistema, agora abrangendo todo o país.

Na década de 70, outros países, principalmente do norte europeu, também adotaram esse sistema, como a Finlândia, que disponibiliza dados censitários em grade desde 1970, e a Inglaterra, que gerou uma grade com os dados do censo de 1971, mas além de não disponibilizar os dados publicamente, esta produção não foi continuada nos censos seguintes.

Os dados em grade também são utilizados para visualização de dados, como é o caso de alguns mapas da publicação de 1980, *People in Britain: a Census Atlas*, e do Atlas Populacional da China, publicado em 1987.

### 1.3.2 Vantagens

O sistema de grade apresenta uma série de vantagens quando comparado com o sistema tradicional de disseminação de dados censitários através de unidades geográficas que representam divisões político-administrativas ou operacionais, a saber:

- a. Estabilidade espaço-temporal;
- b. Adaptação a recortes espaciais;
- c. Hierarquia e flexibilidade;
- d. Versatilidade;

- e. Interpretação cartográfica;
- f. Identificação simples;
- g. Utilização em modelagens e
- h. Minimização dos efeitos do MAUP.

### 1.3.2.1 Estabilidade espaço-temporal

Uma característica do sistema de grades é a sua independência de recortes político-administrativos comumente utilizados para a agregação de dados censitários, o que traz vantagens relacionadas com a estabilidade temporal. Esta estabilidade dos limites das unidades ao longo do tempo é vista como uma das principais vantagens deste sistema (MARTIN, 2000; TAMMILEHTO-LUODE *et al.*, 2000; RUSANEM *et al.*, 2001; TAMMILEHTO-LUODE, 2011), coisa que não acontece com unidades geográficas diretamente ligadas com a administração do território (municípios, distritos e outras), nem com unidades operacionais ou funcionais (setores censitários). No primeiro caso, isso acontece porque essas unidades estão sujeitas a alterações nos seus limites físicos, que pode ser devido à criação de novas unidades, que surgem à medida que as regiões vão se tornando mais desenvolvidas e, conseqüentemente, almejam uma maior autonomia no gerenciamento de seu território, ou mesmo pela redefinição dos limites em decorrência de questionamentos legais ou imprecisões conhecidas e que necessitam de correção. No segundo caso, a causa das alterações é inteiramente operacional, uma vez que os limites dos setores censitários é definido de acordo com o quantitativo de população existente em uma determinada área. Assim, à medida que essas áreas vão crescendo em número de moradores, é necessário que haja uma revisão nos limites para que os parâmetros que regulam a coleta de dados nessas unidades sejam mantidos. O caso mais comum é a divisão de um setor em dois ou mais setores, mas, ocasionalmente, pode acontecer a junção de setores.

Esta característica de estabilidade do sistema de grades pode ser utilizada de duas maneiras diferentes na realização de análises temporais. A primeira requer que a agência nacional de estatística adote este sistema e o mantenha nos próximos censos. Hoje, se quisermos realizar uma análise temporal utilizando a malha de setores censitários, primeiro teremos que ajustar esta malha utilizando os dados de comparabilidade divulgados juntamente com os resultados do censo. Esta tarefa não é simples nem rápida, necessitando do emprego de uma grande quantidade de horas de muitos técnicos habilitados. Com a utilização das grades associadas aos dados censitários, esta tarefa pode ser realizada pronta e rapidamente, sem a necessidade de um processamento de dados demorado ou pesado. A segunda maneira de explorar a estabilidade do sistema de grades é utilizá-lo para transformar os dados passados e criar uma base de dados histórica comparável em termos de unidade espacial. Em ambos os casos, a consequência direta da estabilidade espaço-temporal das unidades espaciais é uma maior rapidez na identificação de mudanças ocorridas em qualquer variável analisada ao longo do tempo.

#### 1.3.2.2 Adaptação a recortes espaciais

Outra característica das grades que oferece vantagem em relação às unidades geográficas tradicionais é a sua forma simples, regular e de pequenas dimensões, o que permite que cada célula funcione como um “tijolo” que pode ser agrupado de diversas formas para se adaptar a diferentes recortes geográficos (TAMMILEHTO-LUODE, 2011). A necessidade de ter dados socioeconômicos e demográficos em unidades espaciais que não sejam as tradicionais unidades político-administrativas é cada vez maior, tendo em vista a crescente necessidade de realização de avaliações e pesquisas envolvendo áreas vulneráveis com relação a algum aspecto físico, ambiental e/ou social. Essa demanda vem sendo feita desde o final dos anos 80 e começo dos anos 90, quando começa a surgir uma maior consciência do relacionamento entre população e meio ambiente, e

continua até hoje, uma vez que a mesma não foi devidamente atendida em todos os níveis de precisão necessários. A área ambiental utiliza com frequência unidades espaciais derivadas de características físicas do meio ambiente, como bacias hidrográficas ou unidades de relevo. Isso causa alguns problemas quando se deseja integrar a dimensão humana aos estudos ambientais, pois geralmente os indicadores sociais são disponibilizados em unidades político-administrativas. Referências a essas questões podem ser vistas, por exemplo, em Clarke e Rhind (1992), Hogan (1992), Deichmann (1996), Ojima e Martine (2012) e Martine e Schensul (2013).

Independente das relações entre população e ambiente, existe a necessidade de integração de dados disponibilizados em unidades não coincidentes em quase todas as áreas do conhecimento que utilizam dados estatísticos, sejam eles derivados de pesquisas domiciliares ou de registros administrativos. A área da saúde sofre bastante com estas questões de integração de dados, uma vez que geralmente suas análises envolvem unidades espaciais que representam regiões de abrangência dos serviços de saúde ou de prevalência ou ocorrência de determinadas doenças e que não apresentam nenhuma relação com unidades político-administrativas. A área da saúde sempre investiu muito em modelos espaço-temporais, pois desde muito tempo se sabe da estreita relação entre a disseminação de doenças e a distribuição e o movimento das populações no tempo e no espaço (HAGERSTRAND, 1967). Talvez por isso, esta área do conhecimento esteja entre aquelas que buscam avançar tecnicamente na busca de soluções metodológicas para enfrentar os problemas aqui discutidos, como pode ser visto em Linard *et al.* (2010) e em Linard e Tatem (2012).

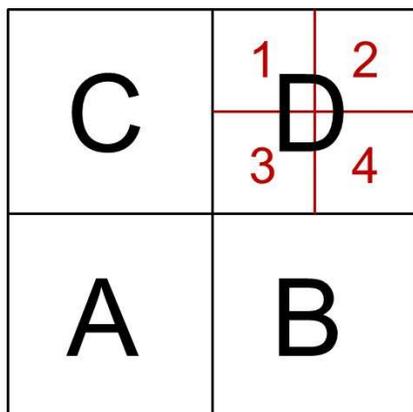
### 1.3.2.3 Hierarquia e flexibilidade

Os sistemas de grade com base em células regulares oferecem uma estrutura hierárquica, de forma que a união de um conjunto de células adjacentes leve à formação de uma célula maior, pertencente a um nível hierárquico superior.

Esta estrutura hierárquica dá flexibilidade ao sistema, permitindo que, caso se queira, diversos níveis hierárquicos possam conviver dentro do mesmo sistema, ou seja, células de dimensões diferentes possam fazer parte do mesmo sistema. No caso de uma grade estatística, é usual que se tenham células com dimensões diferentes cobrindo áreas urbanas e áreas rurais, com o objetivo de melhor representar a distribuição da população. Esta característica hierárquica do sistema de grade permite também que células menores – das áreas urbanas - possam ser agregadas até atingirem as dimensões das células maiores – das áreas rurais -, e que, então, este novo arranjo celular seja utilizado nas análises. E mesmo as células maiores podem ser agregadas em outras ainda maiores, se assim for necessário. Isso demonstra o quanto o sistema pode ser flexível, podendo se adaptar a diferentes escalas, além de todas as outras características vantajosas do sistema, como comparabilidade temporal e espacial.

A Figura 8 mostra uma estrutura hierárquica de células. Nesta figura, podemos observar que o agrupamento de células adjacentes de um mesmo nível forma uma célula de maior dimensão, pertencente ao nível imediatamente superior. Por exemplo, as células menores no canto superior direito 1, 2, 3 e 4 somadas formam uma célula maior, a célula D. Desta maneira, pode ser construído um sistema de grades que conjugue células de dimensões diferentes, de modo que as células menores se encaixem perfeitamente nas maiores. Grades com células de tamanhos diferentes podem ser utilizadas para diferenciar áreas rurais e urbanas, permitindo que regiões mais densas sejam mais detalhadas do que regiões menos povoadas. Isto também oferece ao usuário a opção de ter toda a região em estudo representada por células de mesmo tamanho, bastando que para isso, seja feita uma agregação das células menores, inclusive dos seus atributos, até que atinjam a dimensão das células maiores.

Figura 8 - Estrutura hierárquica de células.



Fonte: Elaborado pelo autor.

#### 1.3.2.4 Versatilidade

Um sistema de grade pode ser construído utilizando tanto uma estrutura de dados vetorial<sup>6</sup> quanto matricial<sup>7</sup> e também pode ser facilmente convertido de uma estrutura para outra. Quando a abordagem metodológica utilizada para a geração da grade é a agregação de dados pontuais, a estrutura usualmente empregada é a vetorial; quando a abordagem metodológica utilizada é a desagregação de dados disponibilizados em unidades areais, a estrutura comumente empregada no produto final é a matricial ou *raster*. A vantagem de se ter a grade em uma estrutura vetorial é que todas as variáveis existentes podem ser alocadas na tabela de atributos e relacionadas com a geometria, facilitando as análises conjuntas dessas variáveis (STRAND; BLOCH, 2009). No caso da estrutura matricial, é necessária a geração de um arquivo para cada variável ou da associação aos dados espaciais de apenas uma variável a cada vez.

<sup>6</sup> As estruturas vetoriais representam as entidades geográficas através de três formas básicas: pontos, linhas e áreas (ou polígonos), definidas por suas coordenadas cartesianas (CÂMARA, 2005).

<sup>7</sup> As estruturas matriciais utilizam uma grade regular sobre a qual se representam os elementos geográficos (CÂMARA, 2005).

### 1.3.2.5 Interpretação cartográfica

Em termos cartográficos, mas sem esquecer que a escala é um fator importante, a utilização de células regulares e com dimensões iguais ou comparáveis em vez da utilização de unidades irregulares e com tamanhos diversos oferece ao leitor uma forma mais legível de interpretação do fenômeno que está sendo representado (GRASLAND; MADELIN, 2006). Isto é uma grande vantagem, principalmente quando o leitor não está familiarizado com as *nuances* que a interpretação de um mapa pode oferecer.

### 1.3.2.6 Identificação simples

A identificação única das células de uma grade pode ser feita de diferentes maneiras, tendo sempre como objetivo a simplicidade e, preferencialmente, permitindo que o posicionamento espacial esteja explícito na própria nomenclatura. No caso de o sistema ser baseado em Sistemas Cartográficos, existem regras e normas estabelecidas, e elas podem ser replicadas ou adaptadas à grade estatística. Outra forma de nomenclatura é a utilização direta das coordenadas de um dos cantos da célula e indicar o tamanho da mesma, que é semelhante ao sistema que é utilizado para definir o posicionamento espacial de uma imagem orbital em programas que trabalham com Sensoriamento Remoto (*world file*), ou, de maneira semelhante, mas utilizando as coordenadas do centro da célula. O *Infrastructure for Spatial Information in the European Community* - INSPIRE sugere a seguinte composição para identificação da célula: a dimensão da célula seguida pelas coordenadas vertical e horizontal do canto inferior esquerdo. Por exemplo, 1kmN2599E4695, que identifica uma célula de 1 km, com coordenadas do canto inferior esquerdo  $Y = 2.599.000$  m e  $X = 4.695.000$  m (INSPIRE, 2010). No caso do Japão, a nomenclatura se baseia no código da primeira partição do Sistema Cartográfico

acrescido de algarismos representando as divisões posteriores, como pode ser visto no Quadro 1.

Os dois sistemas apresentados acima atingem o objetivo de identificar unicamente as células e de explicitar a localização geográfica das mesmas, sendo que ambos podem ser adotados no caso de grades estatísticas.

Quadro 1 - Família de grades estatísticas do Japão.

<b>Partição</b>	<b>Método de divisão</b>	<b>Dimensão</b>	<b>Tamanho</b>	<b>Codificação</b>
Partição Primária	Divide o país em partes iguais medindo 0,67° de latitude por 1° de longitude	40' x 1°	~ 80 km	5339
Partição Secundária	Divide a Partição Primária em 64 (8 por 8)	5' x 7' 30"	~ 10 km	5339-23
Grade Básica	Divide a Partição Secundária em 100 (10 por 10)	30" x 45"	~ 1 km	5339-23-43
Grade Básica – Metade	Divide a Grade Básica em 4 (2 por 2)	15" x 22.5"	~ 500 m	5339-23-43-1
Grade Básica – Quartil	Divide a Grade Básica em 16 (4 por 4)	7.5" x 11.25"	~ 250 m	5339-23-43-1-4

Fonte: Adaptado de JAPAN STATISTICS BUREAU [s.d.].

### 1.3.2.7 Utilização em modelagens

Muitos modelos de simulação que utilizam dados socioeconômicos ou demográficos no seu processamento, o fazem sob a forma de grade e, usualmente, no formato matricial. Isso se deve ao fato de que a estrutura de grades tem uma geometria simples e computacionalmente eficiente, e sob a forma de matriz, facilita os numerosos cálculos iterativos que o modelo necessita executar, poupando tempo de máquina. Este fato já era visto como uma vantagem da utilização do sistema de grades no final dos anos 90 (TOBLER *et al.*, 1997), quando ainda não eram amplamente utilizadas as técnicas de simulação espacial hoje utilizadas. Temos, por exemplo, os modelos *Cellular Automata* que têm sido bastante utilizados na simulação de fenômenos urbanos, como

crescimento regional, *urban sprawl* e evolução do uso das terras. Esses modelos, como o próprio nome indica, utilizam células distribuídas num espaço gradeado como um de seus componentes.

#### 1.3.2.8 Minimização dos efeitos do MAUP

Openshaw e Rao (1995) contestam que a agregação de dados no menor nível geográfico possível seja uma solução para minimizar os impactos causados pelo problema da área modificável. No entanto, esse procedimento é defendido por outros estudiosos, como GOODCHILD (1992), SOUZA; TORRES (2003); SCHUURMAN *et al.* (2006) e CARVALHO *et al.* (2004), apenas para citar alguns. Por outro lado, uma das soluções propostas para o tratamento do MAUP pelo próprio Openshaw (1977) fala sobre o desenvolvimento de uma metodologia que permita criar regiões adequadas para o estudo de determinado fenômeno, as unidades ideais ou ótimas. Entendendo o sistema de grade como um conjunto de unidades básicas de pequenas dimensões, ele pode ser utilizado para realizar a agregação que conduz a geração de uma unidade maior e mais adequada ao fenômeno estudado (GRASLAND; MADELIN, 2006), conforme visto nos itens B e C. Assim, a utilização de grades pode ser considerada uma forma de minimizar os efeitos do MAUP.

De acordo com Wrigley (1995), os dois aspectos do MAUP ficam mais críticos quando se diminui o número de unidades, ou seja, à medida que a quantidade de unidades diminui e, conseqüentemente, a dimensão das mesmas aumenta, a severidade do MAUP tende a crescer. Isso significa que ao compararmos um sistema de grade, que tem uma grande quantidade de pequenas células, com um sistema de polígonos administrativos com poucas unidades de grande extensão, podemos deduzir que o MAUP se apresenta menos crítico no primeiro caso do que no segundo.

### 1.3.3 Desafios

O sistema de grade apresenta diversas vantagens, como acabamos de mencionar, mas, também apresenta alguns desafios que necessitam ser enfrentados, tendo em vista a sua crescente produção e utilização. Os principais desafios estão relacionados com a confidencialidade, a compatibilidade internacional e a análise dos dados.

#### 1.3.3.1 Confidencialidade

Quando tratamos de dados estatísticos agregados em pequenas áreas geográficas, como podem ser consideradas as células de um sistema de grade, a confidencialidade se apresenta como um dos maiores desafios. O dilema entre liberar os dados sem restrição correndo o risco de quebra do sigilo estatístico e suprimir dados que podem alterar os resultados de um estudo é difícil de ser resolvido. Praticamente todas as técnicas utilizadas para minimizar o risco de quebra de confidencialidade acarretam em algum grau de redução da quantidade e da qualidade das informações disponibilizadas (GUTMANN *et al.*, 2008). E, apesar de alguns estudos sobre o assunto, não há ainda uma fórmula para mensurar o potencial de perda de poder analítico dos dados em decorrência da aplicação dessas técnicas (KARR *et al.*, 2006).

O Quadro 2 mostra as regras de confidencialidade utilizadas por alguns países europeus na disseminação em grade de dados censitários da ronda 2010. O valor limite é o valor abaixo do qual não serão disseminados os dados para aquela célula, independente do tamanho da mesma. A primeira divisão da coluna Valor Limite apresenta os limites para a disseminação de variáveis globais, como total de pessoas e total de domicílios; a segunda divisão apresenta os limites para disseminação das demais variáveis. Cabe ressaltar que alguns países colocam o

gênero e estrutura etária como dados globais, como por exemplo, Eslováquia, enquanto outros colocam essas variáveis juntos com aquelas que sofrem restrição, como, por exemplo, Áustria. Podemos observar que, embora não exista uma recomendação sobre a técnica que deve ser utilizada para a manutenção do sigilo estatístico, a supressão de dados é a mais utilizada. Esta supressão pode ser feita de duas maneiras. A primeira se dá pela substituição do valor real por um número, geralmente o valor limite utilizado, e, neste caso, não se consegue identificar quais células tiveram seus valores suprimidos. Na segunda maneira, a substituição é feita por um conjunto especial de caracteres, e neste caso é possível identificar as células que tiveram os dados suprimidos. Não existe uma norma ou padrão para estabelecer os valores limite para realizar esta supressão, embora a utilização dos valores três e dez indivíduos esteja presente em mais de um país.

Quadro 2 - Confidencialidade: valor limite e método utilizado.

País	Valor Limite (hab.)		Método
	Variáveis globais	Demais variáveis	
Áustria	-	3	Ruído e supressão de dados
Dinamarca	-	20 *	Supressão de dados
Estônia	3	3	Supressão de dados
Finlândia	-	10	Supressão de dados
Noruega	-	10	Supressão de dados
Suécia	-	3	Supressão de dados
Irlanda	3	não divulga	Ruído
Portugal	-	não divulga	-

Fonte: Adaptado de KAMINGER, 2007 e EFGS, 2012.

\* Neste caso, a unidade é domicílio.

Outra forma de evitar a quebra de sigilo é a utilização de indicadores em vez dos valores brutos das variáveis, como proposto por MARTINE e SCHENSUL (2013), ou a utilização de classes em vez de valores, como feito pela Suíça (KAMINGER, 2007).

### 1.3.3.2 Compatibilidade internacional

Atualmente, parece haver um movimento na direção de utilizar sistemas de grades regulares para a representação de dados censitários devido, principalmente, ao seu grande potencial para análises, além de haver evidências de uma convergência para a utilização desta representação em modelos de população globais (MARTIN, 2000). No entanto, ainda não existem normas que promovam a compatibilidade espacial ou estatística deste produto. Desta forma, constatamos através da revisão bibliográfica, que são utilizadas diferentes dimensões para as células, diferentes sistemas de projeção cartográfica, além das diferentes variáveis demográficas contempladas e que nem sempre atendem à todas as necessidades dos usuários (TATEM *et al.*, 2012).

Com relação ao sistema de projeção cartográfico, existem algumas argumentações sobre a importância da manutenção da forma quadrada das células e também da igualdade da área de todas as células da grade, e, portanto, a defesa da utilização de sistemas de projeção planos, bem como a não alteração da projeção cartográfica após a geração da grade (TAMMILEHTO-LUODE, 2011). Por outro lado, também existem considerações contrárias, na qual a utilização de sistemas cartográficos não planos não é considerada um problema e, conseqüentemente, a existência de células com áreas diferentes é aceitável (STRAND; BLOCH, 2009).

Com relação às variáveis contempladas, as diferenças são grandes, havendo países que disseminam apenas a população total enquanto outros disseminam diversas variáveis sociodemográficas, além de variáveis relacionadas com as edificações e estabelecimentos empresariais e agropecuários. Um desafio

a ser vencido é a definição de um conjunto mínimo de variáveis que possa ser disseminado por todos os países e que atenda a diferentes usuários e múltiplos objetivos.

Todas essas questões estão relacionadas com a falta de padrões e normas internacionais que estabeleçam boas práticas a serem empregadas na geração de grades estatísticas. Este desafio terá que ser enfrentado em um curto espaço de tempo, tendo em vista que um crescente número de países está adotando este sistema para a divulgação de suas estatísticas. Também é importante que o produto venha acompanhado de seus metadados, de modo que o usuário possa ter conhecimento das características técnicas da grade, como projeção cartográfica, por exemplo, e das incertezas envolvidas na metodologia adotada para a integração com os dados estatísticos.

#### 1.3.3.3 Análise dos dados

A análise de dados agregados em pequenas unidades espaciais traz, sem sombra de dúvidas, uma maior acurácia na distribuição e identificação de eventos, uma vez que a resolução espacial é alta. No entanto, existe uma grande variação no valor das variáveis e, dependendo da escala utilizada, os dados podem se apresentar confusos e nebulosos, prejudicando a identificação de tendências ou padrões espaciais e, conseqüentemente, a obtenção de conclusões. Em unidades maiores, as medidas estatísticas são suavizadas, e a instabilidade diminui, mas isso também pode prejudicar as conclusões. Esta contradição está presente em todos os estudos chamados ecológicos, onde o foco é a comparação de grupos, em vez de indivíduos. Isto se deve ao aumento da homogeneidade interna em unidades pequenas, decorrente da diminuição da população contida na unidade, que, apesar de ser um fator positivo, traz consigo a instabilidade típica dos pequenos números, ocasionando grandes flutuações nas variáveis (CARVALHO; SANTOS, 2005). Teoricamente, existe um tamanho ideal para as unidades de análise que combina as vantagens tanto das unidades

pequenas quanto das grandes, sem, no entanto, incorporar as desvantagens. Como discutido anteriormente no item 1.3.2.8 o sistema de grade pode ser utilizado para construir essas unidades ótimas. Mas não podemos nos esquecer de que esta delimitação depende diretamente do fenômeno que está sendo estudado, não havendo uma unidade ótima aplicável a todos os casos. Na prática, isso nem sempre é possível de ser realizado, e então, lançamos mão de unidades de análise genéricas, que nem sempre têm relação com o objeto de estudo. A adoção destas unidades genéricas nos leva à necessidade de um maior cuidado na elaboração das conclusões.

Esta discussão está sendo realizada sob a luz de análises tradicionais efetuadas com unidades que apresentam uma extensão que não pode ser classificada como pequena ou micro. Até mesmo os setores censitários urbanos, localizados em regiões com densidade populacional extremamente alta, e, portanto, com área extremamente pequena, apresentam distorções nas suas taxas quando comparados com setores maiores. Não se sabe até que ponto as metodologias de criação de taxas e indicadores se aplicam a microáreas, e, no caso de serem aplicáveis, não se conhece os parâmetros que devem ser utilizados para nortear as análises (SOUZA; TORRES, 2003), pois não há o costume de se utilizar essas unidades em estudos e análises (RUSANEM *et al.*, 1993).

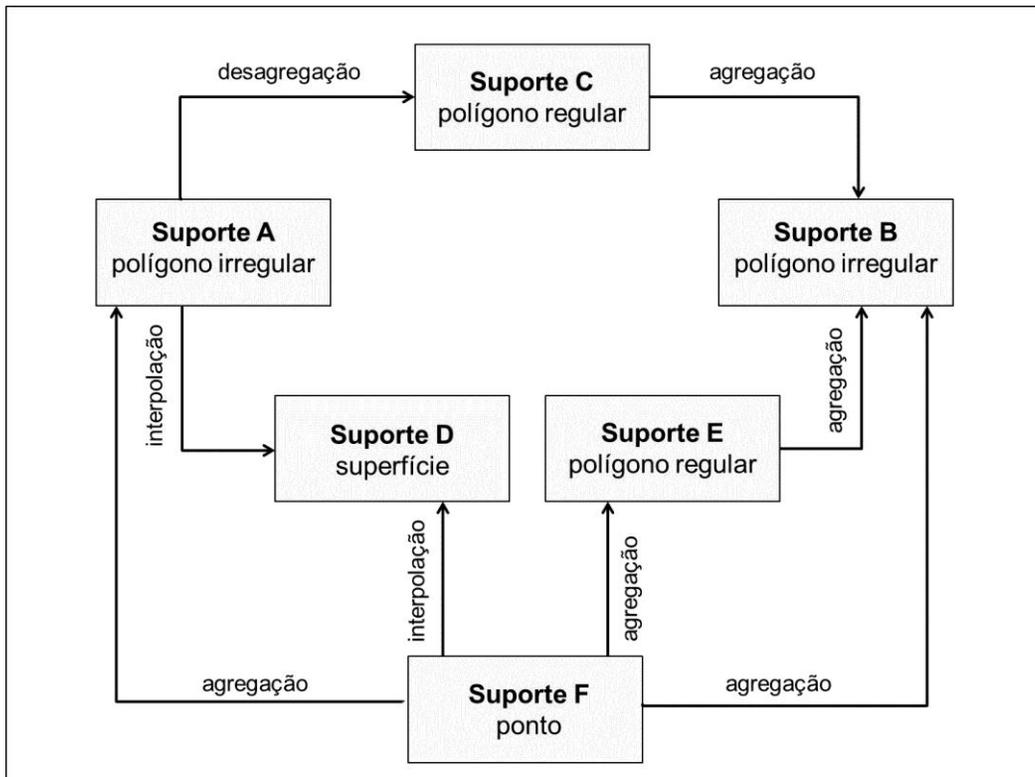
#### 1.3.4 Abordagens

A geração de um sistema de grades com objetivos estatísticos consiste em uma operação de transferência de suporte espacial de dados, onde os dados básicos existentes em um suporte (por exemplo, pontos ou polígonos) são transferidos para outro suporte (por exemplo, polígonos ou superfícies). As técnicas para esta transferência de suporte podem ser classificadas de acordo com três diferentes abordagens (PLUMEJEAUD *et al.*, 2010): agregação, desagregação e interpolação, conforme ilustra a Figura 9.

### 1.3.4.1 Agregação

A abordagem de agregação consiste em somar os dados básicos que se encontram dentro da unidade desejada seguindo uma regra espacial de inclusão (PLUMEJEAUD *et al.*, 2010). No caso de pesquisas censitárias, estamos somando os domicílios, representados espacialmente por pontos que se encontram dentro de uma área determinada que, por sua vez, é representada espacialmente por um polígono. Esta seria a solução ideal para os problemas decorrentes das constantes alterações nos limites das unidades espaciais utilizadas para agregar dados estatísticos, ou seja, ter os microdados das pesquisas domiciliares georreferenciados e agregá-los em uma unidade espacial qualquer (NORMAN *et al.*, 2003).

Figura 9 - Técnicas para transferência de suporte espacial.



Fonte: Adaptado de PLUMEJEAUD *et al.*, 2010.

A localização espacial dos domicílios pode ser feita diretamente, utilizando as coordenadas capturadas em campo, ou indiretamente, utilizando algoritmos para a transformação de endereços em pares de coordenadas. Este é o caso de alguns países nórdicos, como Finlândia e Suécia, onde diversos cadastros administrativos podem ser relacionados entre si através do endereço, que, por sua vez, tem suas coordenadas registradas em uma base de dados (NAÇÕES UNIDAS, 2007a). A operação de agregação desses pontos em uma unidade poligonal qualquer pode ser feita com a utilização de ferramentas simples, disponíveis na maioria dos programas que trabalham com informações espaciais.

A principal característica desta solução é a simplicidade, mas, em contrapartida, ela pressupõe a existência de dados georreferenciados ou, pelo menos, passíveis de serem georreferenciados. Com a utilização crescente de tecnologias nos censos, a adoção desta solução para a divulgação de dados censitários torna-se viável para um número cada vez maior de países. Segundo dados das Nações Unidas (2013), a rodada dos Censos 2010 mostrou um uso sem precedentes de tecnologias em todas as etapas da operação censitária. Estão incluídas nestas tecnologias a utilização de ferramentas de cartografia (GPS, SIG, imagens de satélite, fotografias aéreas e digitalização de dados) na fase de planejamento e o emprego de computadores de mão, internet, SIG e telefonia móvel na fase de enumeração.

Outra característica importante desta abordagem está relacionada com quem pode executá-la. A geração da grade tem sua origem nos dados básicos coletados para os domicílios e para os residentes e, como regra geral, apenas os produtores da pesquisa censitária podem acessar e manipular esses dados. Assim, no caso dos censos, apenas os institutos nacionais de estatística, que são os executores da pesquisa, têm acesso aos microdados identificados, ou seja, sem intervenções com a finalidade de retirar a identificação dos respondentes. Esta característica torna esta abordagem limitada do ponto de vista de desenvolvimento tecnológico e científico, uma vez que grande parte dos pesquisadores não pode testar e desenvolver técnicas para aperfeiçoar as

metodologias empregadas, pois não pode acessar os dados ou tem restrições para realizar este acesso.

Um desafio ainda permanece nesta abordagem quando se trata da geocodificação de edificações: não existe um padrão estabelecido para a coleta ou para os atributos que devem ser registrados (ZIMMERMAN; JIE, 2010). O resultado desta falta de padronização é a existência de diferentes pontos de referência para as coordenadas geográficas dos edifícios. Alguns países, como Áustria, Finlândia, Suíça e Estônia (área urbana), utilizam um ponto central do polígono da edificação; a Suécia e a Estônia (área rural) utilizam um ponto central da propriedade; Dinamarca e Noruega utilizam um ponto localizado na entrada principal da edificação (TAMMISTO, 2007). Cabo Verde utiliza uma abordagem mista, em que os edifícios que foram georreferenciados em escritório, diretamente sobre fotografias aéreas, têm um ponto localizado no centro da edificação, e aqueles que foram georreferenciados em campo têm um ponto localizado nas proximidades da entrada principal (BUENO *et al.*, 2009). Apesar de esta diferença não causar grandes incertezas, considerando-se as finalidades estatísticas da informação, uma base continental ou global construída a partir desses dados necessita de uma documentação estruturada sob a forma de metadados, de modo a informar o usuário acerca da acurácia dos dados, evitando que a utilização dos mesmos seja comprometida.

#### 1.3.4.2 Desagregação

Esta abordagem engloba técnicas que transferem os dados existentes em um suporte espacial para outro suporte espacial, através da criação de um denominador espacial comum entre as unidades de origem e de destino (PLUMEJEAUD *et al.*, 2010). Algumas das técnicas pertencentes a este tipo de abordagem são bem antigas e surgiram principalmente para resolver o também antigo problema de incompatibilidade entre áreas para as quais as estatísticas são publicadas e áreas para as quais se desejam ter essas estatísticas.

A técnica mais simples e mais amplamente utilizada é a ponderação zonal simples (RASE, 2001; REIBEL; AGRAWAL, 2007), que utiliza a densidade da variável na unidade de origem para a realocação na unidade de destino. A vantagem desta técnica é a sua simplicidade e a existência de ferramentas prontas em diversos programas de geoprocessamento; a desvantagem é a hipótese de distribuição uniforme da variável nas unidades de origem, o que é pouco provável de ser encontrado no mundo real, pois não são consideradas as variações locais da variável no interior das unidades (WU *et al.*, 2005; PLUMEJEAUD *et al.*, 2010; FRANÇA, 2012).

Com o objetivo de aumentar a acurácia dos resultados, surgiu a ponderação zonal modificada, que utiliza dados auxiliares como uma aproximação para estimar a distribuição da variável na unidade de destino. O método dasimétrico é o mais conhecido nesta categoria e tem sua origem nos trabalhos desenvolvidos por Wright (1936) com o objetivo de encontrar uma representação mais realista da distribuição da população do que aquela oferecida pelos mapas coropléticos. No caso de distribuição espacial da população, diferentes fontes de informação auxiliar podem ser utilizadas, considerando-se que a população não tem uma distribuição aleatória no espaço e que prefere ocupar áreas com características apropriadas (DEICHMANN, 1996). A classificação do uso das terras derivada de interpretações de imagens orbitais é a informação auxiliar mais comumente utilizada e a existência de numerosos estudos com testes robustos confirmando a sua acurácia tornam este dado auxiliar o padrão para o mapeamento dasimétrico (REIBEL; AGRAWAL, 2007), embora diversas outras informações também sejam utilizadas, como veremos mais adiante.

O método dasimétrico mais simples, dentre os que utilizam dados de sensoriamento remoto, é o binário (LANGFORD *et al.*, 1991), que utiliza apenas duas classes de uso das terras: habitada e inabitada. O total de população na unidade de origem (geralmente o setor censitário) é, então, redistribuído apenas na classe habitada, através da densidade populacional dasimetricamente derivada. A vantagem desta modalidade é a sua simplicidade, tendo em vista que

a classificação de uso das terras envolve apenas duas classes, muito embora esta divisão seja subjetiva e dependa de conhecimentos da região mapeada (EICHER; BREWER, 2001); a desvantagem é não considerar a existência de variações na densidade populacional nas áreas povoadas (MAANTAY; MAROKO; GERRMANN, 2007; FRANÇA, 2012). Alguns exemplos de estudos feitos utilizando esta técnica podem ser vistos em Langford e Unwin (1994) e em Holt *et al.* (2004).

Como colocado por Langford (2006), esta visão binária é simplista demais, e várias densidades populacionais convivem em uma mesma área, pois os fatores socioeconômicos e as diferenças físicas do ambiente contribuem para o surgimento de valores diferenciados deste indicador. E para considerar essas diferenças surge o modelo dasimétrico com diversas classes de uso das terras e com diferentes valores de densidade associados a elas. Nesta modalidade do método dasimétrico, o desafio é encontrar o valor mais adequado para a densidade de cada classe, e este desafio foi tratado de diferentes maneiras.

Eicher e Brewer (2001) propuseram o uso de apenas três classes de uso das terras, sendo que a população da unidade de origem deveria ser redistribuída na unidade de destino de acordo com percentuais arbitrados com base em amostras de campo ou alguma informação auxiliar. A desvantagem deste método é que ele não considera as eventuais diferenças entre regiões contendo a mesma classe de uso da terra e, portanto, associa um mesmo valor para a redistribuição da população para a classe como um todo (EICHER; BREWER, 2001; LANGFORD, 2006). Por exemplo, se o valor arbitrado para a redistribuição da população na classe área urbana foi de 80%, este valor será empregado igualmente em uma área urbana muito densa e em uma área urbana pouco densa, pois a classe é considerada como sendo homogênea. Com apenas três classes, é de se esperar que dentro de cada classe existam grandes variações. Como o percentual utilizado é um valor médio que representa a classe como um todo, isso significa que em uma região de estudo de pequena extensão é mais provável que o erro seja maior, pois os polígonos que constituem cada classe

serão mais heterogêneos, por isso, certamente os resultados serão melhores para estudos regionais do que para estudos locais (DONNAY; UNWIN, 2001).

Em vez de arbitrar um valor para as densidades em cada classe de uso das terras, podem ser utilizados modelos estatísticos para estabelecer a correlação entre a distribuição da população e a informação auxiliar, como proposto por Langford, Maguire e Unwin (1991) e por Yuan, Smith e Limp (1997). A vantagem desta abordagem é que a utilização de técnicas estatísticas, como a regressão, por exemplo, permite avaliar a acurácia dos resultados de maneira objetiva (WU *et al*, 2005).

Outro aperfeiçoamento no método dasimétrico foi feito por Mennis (2003), que utiliza amostragens empíricas para a determinação das densidades. Nesta modalidade, devem-se buscar zonas cobertas totalmente por apenas uma classe de uso das terras e obter a densidade populacional, que, então, será utilizada nas demais zonas onde esta classe é existente. Isso significa que as densidades para cada classe de uso são utilizadas como pesos relativos na redistribuição da população, o que assegura que a população total da unidade de origem permaneça a mesma (ZANDBERGEN; IGNIZIO, 2010).

As diferentes abordagens do método dasimétrico mostradas até agora utilizam classificações de uso das terras derivadas de imagens de sensores remotos como informação auxiliar, no entanto, outros tipos de dados podem ser utilizados. Informações também derivadas de imagens de Sensoriamento Remoto, como luzes noturnas e superfícies impermeáveis do solo, também são utilizadas como apoio ao método dasimétrico. As imagens de luzes noturnas são indicativas da presença humana e são valiosas para projetos de mapeamento da distribuição da população em escala global (SUTTON, 1997; TURNER; OPENSHAW, 2001). No Brasil, Miranda (1999) e Kappel (2004) utilizam este produto em estudos relacionados com a representação espacial da população na Amazônia. As superfícies impermeáveis têm uma grande correlação com a densidade populacional e têm a vantagem de não necessitarem de uma calibração complexa para a sua classificação a partir de imagens de sensores remotos

(ZANDBERGEN, 2011). A utilização desta informação tem como base a hipótese de que a densidade de edificações ou de população é proporcional ao grau de impermeabilidade. No entanto, como isso não é sempre verdadeiro, será necessário efetuar algumas edições na camada de superfícies impermeáveis para excluir algumas áreas que não são residenciais, como áreas comerciais e industriais, além da rede viária e ferroviária (STEINNOCHER *et al.*, 2010). Exemplos de utilização de superfícies impermeáveis como informação auxiliar para estudos de população podem ser vistos em Morton e Yuan (2009) e Steinnocher *et al.* (2010).

Outros tipos de informação de origens as mais variadas possíveis também podem ser utilizados para auxiliar a redistribuição da população através de métodos dasimétricos, como, por exemplo, redes de transporte, endereços e edificações, entre outros. A utilização de mapas de redes de transporte como informação auxiliar tem por base a observação de que a maioria das casas está situada ao longo das vias. Xie (1995) propôs três diferentes metodologias para utilizar esta informação em métodos dasimétricos. A primeira delas, a mais simples de todas, assume que a população está distribuída homoganeamente ao longo das vias; a segunda assume que existem diferenças na densidade populacional de acordo com a categoria da via e utiliza pesos para expressar essas diferenças; a terceira metodologia utiliza um cadastro de endereços espacial que distribui as casas ao longo dos dois lados da via e aplica uma taxa média de pessoas por casa. A restrição a este método é a falta de uma padronização para a atribuição de densidades populacionais para as diferentes classes de vias (XIE, 2006). Alguns exemplos de utilização da rede viária como informação auxiliar podem ser vistos em Mrozinski e Cromley (1999), Reibel e Bufalino (2005) e Brinegar e Popick (2010).

Outra informação que pode auxiliar a distribuição de população através de métodos dasimétricos é o endereço. A partir de uma listagem de endereços e com a utilização de algoritmos de geocodificação existentes em programas que tratam informações espaciais, pode-se obter um conjunto de pontos

representando as edificações. E, então, a densidade deste conjunto de pontos pode ser considerada uma estimativa da densidade populacional. Apesar de esta metodologia ser teoricamente promissora, na prática, ela enfrenta uma série de problemas (ZIMMERMAN, JIE, 2010). O primeiro deles está relacionado com a abrangência e a disponibilidade da informação. Em diversos países, a disponibilidade de cadastros de endereços com qualidade suficiente para permitir uma boa geocodificação não é universal, geralmente estando restrita a grandes centros urbanos. Este é o caso do Brasil e, de uma maneira geral, dos países em desenvolvimento. Em outros países, apesar da existência de um cadastro abrangente, a sua utilização não é permitida ao público, como nos Estados Unidos, onde o *Census Bureau* criou e mantém o *Master Address File* - MAF, que não é publicamente acessível. Além disso, os resultados obtidos com a execução de algoritmos de geocodificação não são totais, havendo geralmente entre 20 e 40% de endereços não geocodificados automaticamente (ZIMMERMAN; JIE, 2010). Quando o cadastro de endereços já traz em seus registros as coordenadas, não havendo necessidade de geocodificação, o problema está na falta de um padrão estabelecido para a coleta dessas coordenadas, como discutido na sessão anterior. Uma limitação ao uso de pontos georreferenciados de endereços para a distribuição de população através de métodos dasimétricos é a atribuição de uma densidade populacional uniforme por ponto ou edificação, pois se sabe que existem diferentes tipos de edificação residencial com diferentes densidades (MAANTAY; MAROKO; HERRMANN, 2007). Uma maneira de tratar esta questão é a incorporação de pesos para cada tipo de edificação (unifamiliar ou multifamiliar) ou de acordo com o tamanho da edificação, que pode ser medido tanto pelo número de quartos como pela área construída (ZANDBERGEN, 2011). Os trabalhos de Owens *et al.* (2010), Tapp (2010) e Zandbergen (2011) são exemplos de utilização de endereços como dado auxiliar em métodos dasimétricos.

O mapeamento cadastral, com a delimitação de lotes e edificações, também pode ser utilizado para auxiliar a construção de mapas de distribuição da

população em áreas urbanas, como proposto por Maantay, Maroko e Herrmann (2007). Uma restrição ao uso desta informação é a mesma apontada na utilização de endereços como informação auxiliar, qual seja, a dificuldade de estabelecer uma densidade média a ser utilizada, tendo em vista a diversidade de configurações que uma edificação pode apresentar (ZIMMERMAN, 2010). Outra restrição se relaciona com o fato de que esta informação nem sempre existe ou está disponível aos usuários, o que pressupõe a sua utilização apenas em pequenas áreas (ZIMMERMAN, 2010).

A utilização de edificações residenciais como informação auxiliar também é bastante comum, devido a sua correlação com a distribuição de população e também por ser utilizada na coleta de dados censitários (XIE, 2006; SRIDHARAN; QIU, 2013). Esta informação pode ser extraída de imagens de satélite de alta resolução, ortofotocartas ou imagens *LiDAR*<sup>8</sup>, embora seja necessária a associação com mapas cadastrais para a identificação da tipologia dos edifícios (residencial, comercial ou industrial) (XIE, 2006). A vantagem das imagens *LiDAR* em relação às outras imagens está na sua natureza tridimensional, possibilitando estimar a altura e o volume dos objetos e permitindo, conseqüentemente, um maior nível de detalhe no mapeamento das edificações (SRIDHARAN; QIU, 2013).

Finalmente, existem propostas que utilizam diversos dados simultaneamente. Podemos citar como exemplo o projeto *LandScan* (BHADURI *et al.*, 2007), que utiliza múltiplas fontes de dados auxiliares em uma adaptação do método dasimétrico. Entre os dados auxiliares utilizados estão declividade, malha viária, imagens de luzes noturnas e classificação de uso e cobertura das terras.

Uma avaliação geral da acurácia do método dasimétrico, principalmente com a utilização de uso das terras como informação auxiliar, foi feita por Zandbergen e Ignizio (2010), que apresentaram três fatores que influenciam os erros envolvidos na metodologia. O primeiro fator é o tamanho das unidades de

---

<sup>8</sup> **LiDAR** (*Light Detection and Ranging*) é uma tecnologia ótica de detecção remota que mede propriedades da luz refletida de modo a obter a distância de objetos e outras informações derivadas.

origem e de destino, uma vez que os erros são menores quando o método é aplicado em unidades de origem de pequena extensão e unidades de destino de grande extensão. O segundo fator é o tipo de organização espacial existente nas unidades de origem e destino: quanto mais essas unidades tiverem uma organização semelhante, menor será o erro. E, finalmente, a qualidade dos dados auxiliares, estando aqui incluída a resolução espacial, que deve ser maior do que a dos dados populacionais (TATEM *et al.*, 2007) e a classificação do uso das terras, que deve ser o mais detalhado possível.

#### 1.3.4.3 Interpolação

Podemos classificar os métodos de interpolação entre aqueles que são baseados em pontos e aqueles que são baseados em áreas. Na primeira classe, podemos citar os métodos de interpolação por polinômios, os que utilizam distância, os geoestatísticos (*kriging*) e os modelos de tendência utilizando diferentes funções (WU *et al.*, 2005). Todos esses métodos assumem que os centroides das áreas de enumeração são representativos da distribuição da população naquela área e, portanto, podem ser utilizados para a redistribuição da população. Um dos métodos desta classe que é mais difundido é o método proposto por Bracken e Martin (1989), no qual uma janela móvel ou *kernel* é posicionada sobre cada ponto, e a população alocada na nova unidade, que é composta por células regulares, é determinada por uma função de ponderação baseada no decaimento da distância entre o centroide e a célula. Uma abordagem alternativa foi utilizada por Honeycutt e Wojcik (1990) nos Estados Unidos, com a utilização de uma função Gaussiana circular para distribuir a população em torno do centroide. Os problemas relacionados com estes métodos dizem respeito à escolha do centroide para representar a unidade areal, já que caso esta área seja irregular e não simétrica, o mesmo pode estar localizado em uma posição que afeta os resultados, como por exemplo, fora do polígono (LAM, 1983; LIU, 2003). Outros problemas com estes métodos apontados por Martin (1996) estão

relacionados com a não preservação do valor total da população em cada zona de origem e com a não consideração de regiões desabitadas.

Na segunda classe de métodos de interpolação o método mais conhecido é o picnofilático proposto por Tobler (1979), que utiliza a população em unidades de área como base para a criação de uma superfície suavizada em grade. Este método restringe a movimentação de população entre as unidades areais, garantindo que a população permaneça constante. O método original utiliza como configuração espacial uma grade de pontos, enquanto um aprimoramento realizado por Rase (2001) utiliza uma rede triangular irregular. O método picnofilático é mais apropriado para os casos em que não se tem nenhuma informação auxiliar disponível e quando as unidades espaciais de origem são relativamente homogêneas (DEICHMANN, 1996). Também podem ser incluídos nesta classe os métodos propostos por Kyriakidis, nos quais são utilizados métodos geoestatísticos na realização de interpolações do tipo área-ponto (KYRIAKIDIS; YOO, 2003; KYRIAKIDIS, 2004; YOO; KYRIAKIDIS, 2006) e área-área (KYRIAKIDIS; SCHENEIDER; GOODCHILD, 2005; KYRIAKIDIS, 2011).

### 1.3.5 Usos

A emergência de estudos espacialmente explícitos se apoia no aumento da quantidade e da qualidade dos dados espaciais e em ferramentas específicas para o tratamento e coleta desses dados, como sistemas de informações geográficas e sensoriamento remoto, bem como nos avanços das técnicas de modelagem e estatística espacial. Diversos métodos para transformar dados populacionais originalmente relacionados com unidades administrativas irregulares para uma grade regular foram desenvolvidos a partir do início dos anos 1990, o que levou ao surgimento de diferentes bases de dados globais e continentais utilizando grades regulares como unidade espacial.

Podemos citar como exemplo de bases globais o Gridded Population of the World – GPW – e o Global Rural Urban Mapping Project – GRUMP – (CIESIN;

COLUMBIA UNIVERSITY; CIAT, s.d.), distribuídos pelo Socioeconomic Data and Applications Center – SEDAC –, localizado no Center for International Earth Science Information Network - CIESIN – da Columbia University, USA. O GPW utilizou inicialmente dados censitários em nível subnacional e um algoritmo de suavização (TOBLER *et al.*, 1995), mas métodos mais sofisticados foram utilizados nas demais versões (BALK; YETMAN, 2004). Atualmente, o GPW está na versão 3, e sua resolução espacial é de 2,5 minutos ou 5 km. O GRUMP, cujo objetivo é oferecer uma série temporal de dados de população para integração com outras bases de dados, encontra-se na versão 1 e tem resolução espacial de 30 segundos ou 1 km. Diferentemente do GPW, o GRUMP diferencia áreas urbanas e rurais devido à utilização de imagens de luzes noturnas na sua geração (BALK *et al.*, 2004a).

O projeto *LandScan* (BADHURI *et al.*, 2007), desenvolvido pelo *Oak Ridge National Laboratory*, também se configura como uma base de dados global. Apresenta resolução espacial de 30 segundos e utiliza dados de sensores remotos e os associa a diversos dados auxiliares para melhor distribuir espacialmente a população (DOBSON *et al.*, 2000).

Outro esforço de mapeamento global e continental utilizando grades foi realizado pelo Programa Ambiental das Nações Unidas – UNEP – (NAÇÕES UNIDAS, s.d.) utilizando dados de população nacionais redistribuídos para uma grade de 5 km com base na localização de áreas povoadas e malhas viárias. A base de dados oferece uma série histórica de densidade populacional com início em 1960 e indo até o ano 2000 para a África e América Latina; para a Ásia o dado está disponível apenas para 1995; a base global oferece a distribuição de população referente ao ano de 1990.

Os projetos continentais *AfriPop* e *AsiaPop* (LINARD *et al.*, 2012; AFRIPOP PROJECT, 2009; ASIAPOPPROJECT, 2011) foram criados com o objetivo de produzir mapas populacionais detalhados, com resolução de 100 m, abrangendo regiões menos desenvolvidas para disponibilização gratuita e tiveram o apoio de diversos pesquisadores, universidades e instituições. O mapeamento

para as Américas está em fase desenvolvimento e deverá ser disponibilizado brevemente (AMERIPOP PROJECT, 2012). Esses projetos fazem uso de dados de imagens LANDSAT, classificações de uso das terras e dados censitários nacionais (LINARD *et al.*, 2012; TATEM *et al.*, 2007; LINARD *et al.*, 2010).

Finalmente, temos o projeto GEOSTAT, que tem por objetivo a criação e disseminação de uma grade de população abrangendo os países da União Europeia (EFGS, 2012).

Com relação à utilização de grades em nível nacional, podemos citar como exemplo as agências nacionais de estatística de Suécia, Finlândia, Noruega e Japão. A Suécia oferece aos seus usuários um grupo de produtos estatísticos para geografias de pequenas áreas e entre elas estão as grades regulares com dimensões de 1 km em áreas rurais e 250 m em áreas urbanas. As centenas de variáveis oferecidas são sobre temas como população, renda, emprego e educação (SEHLIN, 2011). Na Finlândia, a base de dados em grade cobre todo o território finlandês, com dimensões de 1 km e 250 m, e contém estatísticas relacionadas com a estrutura da população, educação, ocupação principal, renda, família, edificação e local de trabalho (STATISTICS FINLAND, 2012). Já na Noruega a grade estatística tem dimensões de 250 m, 1 km e 5 km, e as variáveis disponíveis são relacionadas com população, edificações e estabelecimentos agropecuários (STRAND e BLOCH, 2009). Todas essas bases em grade têm como fonte um sistema inter-relacionado de registros administrativos. O Japão utiliza grades com forma retangular, baseadas em subdivisões da folha cartográfica regional na escala 1:200.000. As dimensões das células são de aproximadamente 250 m, 500 m, 1 km, 10 km e 80 km e as variáveis são derivadas dos censos de população e de empresas. A geração dos dados estatísticos utiliza uma abordagem de desagregação, partindo dos distritos de enumeração e utilizando o método de ponderação por área (NAOKI e MASAYUKI, 2012).

Podemos concluir a partir do que foi discutido neste capítulo que a adoção de um sistema de grades oferece uma série de vantagens em relação à

utilização de unidades operacionais ou político-administrativas para a disseminação de dados censitários, tornando-se uma opção viável de ser adotada. No próximo capítulo, apresentaremos a metodologia para a geração desta grade estatística contendo os dados do Censo Demográfico 2010 do Brasil.

## CAPÍTULO 2 - DADOS E MÉTODOS

*Population is the point of reference from which all the other elements are observed and from which they all, singly and collectively, derive significance and meaning. It is population which furnishes the focus.*  
(Glenn T. Trewartha)

Neste capítulo é apresentada a metodologia proposta para vencer os desafios enunciados no capítulo anterior e, conseqüentemente, alcançar o objetivo de aumentar o potencial analítico dos dados censitários utilizados em análises espaciais. Primeiramente, é feita uma breve apresentação do Censo Demográfico 2010 no Brasil. Em seguida, são descritos os dados utilizados neste estudo, tanto os estatísticos quanto os vetoriais. Após a apresentação dos dados, é feita a apresentação da metodologia proposta para a criação de uma grade estatística para disseminar dados censitários, expondo a definição das características geométricas e espaciais da grade, a abordagem geral utilizada para povoamento da mesma e os procedimentos operacionais adotados para a localização espacial dos dados estatísticos.

### 2.1 Censo Demográfico 2010

O Censo de População e Habitação<sup>9</sup> é uma operação estatística complexa, devido principalmente à sua abrangência, uma vez que investiga todos os domicílios do país. Em se tratando de Brasil, com uma área territorial de mais de 8,5 milhões de km<sup>2</sup>, onde se encontram cerca de 60 milhões de domicílios, esta operação toma proporções gigantescas, e sua complexidade aumenta proporcionalmente.

---

<sup>9</sup> A Organização das Nações Unidas denomina a operação sistemática de aquisição de informações acerca das características essenciais dos membros de uma população como Censo de População e Habitação (NAÇÕES UNIDAS, 2008). No Brasil esta operação é conhecida como Censo Demográfico (IBGE, 2010a).

No Brasil, quem tem a atribuição de realizar esta operação é o Instituto Brasileiro de Geografia e Estatística – IBGE –, órgão do governo federal ligado ao Ministério do Planejamento (IBGE, 2010a).

### 2.1.1 Inovações

Para enfrentar a complexidade da operação censitária, o IBGE utiliza tecnologias atuais em seu trabalho, aprimorando cada vez mais os métodos e técnicas de coleta, apuração e disseminação dos resultados.

Alguns avanços adotados pelo IBGE no Censo Demográfico 2010 merecem ser destacados devido a sua relação com o trabalho desenvolvido nesta tese. Estas inovações se aplicam à coleta de dados com utilização de dispositivos eletrônicos, ao mapeamento censitário e ao cadastro de endereços. A seguir, cada uma dessas inovações será apresentada com maiores detalhes.

#### 2.1.1.1 Coleta de dados eletrônica

Os equipamentos de coleta eletrônicos foram utilizados pela primeira vez na operação conjunta de Contagem Populacional e Censo Agropecuário realizada em 2007 e tiveram seu uso aprimorado no Censo Demográfico 2010 (IBGE, 2010a). O emprego desses equipamentos com dispositivos GPS acoplados auxiliou a compreensão do território por parte do recenseador, pois permitiu que a sua posição geográfica fosse visualizada em mapas. Os aparelhos também permitiram a coleta de informações de localização dos domicílios visitados pelo censo, bem como de outros pontos de interesse, como estabelecimentos de saúde e educação (IBGE, 2010a). Nas áreas rurais, com algumas exceções, não existe um traçado regular de vias que permita a codificação de quadra/face, então, nessas áreas foram capturados os pontos GPS de cada edificação, os quais foram incluídos no Cadastro de Endereços.

Os mapas, cadastros e questionário foram os elementos principais utilizados pela aplicação eletrônica desenvolvida especificamente para a coleta de dados do censo (IBGE, 2013). Todos esses elementos se relacionam por meio de uma série de códigos que permitiam a associação entre cada um dos elementos. Assim, o mapeamento censitário, o cadastro de endereços e o questionário podem ser relacionados, permitindo que os dados relativos a um determinado domicílio ou morador sejam localizados espacialmente, seja através de pontos GPS, seja através de um endereço que leva a um segmento específico de via no mapeamento censitário.

#### 2.1.1.2 Mapeamento censitário

O mapeamento censitário engloba uma série de mapas, cadastros e bancos de dados que têm o objetivo de oferecer um suporte cartográfico para as atividades censitárias. Para a atualização deste acervo, o IBGE estabelece parcerias com outros órgãos produtores de mapeamento e realiza, ele próprio, diversos trabalhos de campo e de escritório (IBGE, 2011a).

Para atender as demandas da operação de 2010, foram desenvolvidas diversas melhorias no mapeamento censitário (IBGE, 2010a), entre as quais destacamos:

- a. Integração das vertentes urbana e rural numa base de mapeamento digital única;
- b. Melhoria da geometria da malha de setores censitários e
- c. Associação do arruamento urbano com o Cadastro Nacional de Endereços para Fins Estatísticos – CNEFE –, através da codificação das quadras/faces de logradouros.

### 2.1.1.3 Cadastro de Endereços

Apesar do Cadastro de Endereços existir em meio digital desde 2005, apenas em 2010 ele foi associado ao mapeamento censitário (IBGE, 2010a). Isto permitiu que a aplicação eletrônica desenvolvida para a coleta de dados pudesse colocar os mapas e a lista de endereços juntos numa mesma tela, facilitando a compreensão do território por parte do recenseador, como pode ser visto na Figura 10.

Figura 10 - Tela do aplicativo do Censo 2010 para computador de mão.

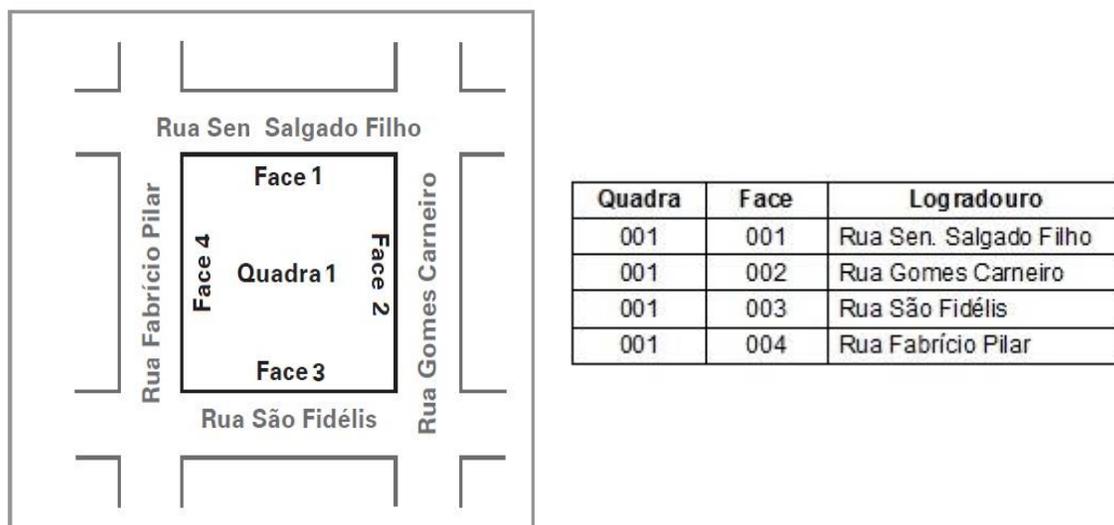


Fonte: Adaptado de IBGE, 2010b.

De acordo com as instruções para utilização da aplicação do Censo 2010 (IBGE, 2010b), dois caminhos poderiam ser utilizados para fazer a seleção da quadra/face de trabalho: escolhendo uma quadra/face específica na listagem ou selecionando graficamente no mapa. Esses caminhos são intercambiáveis, ou

seja, escolhendo um registro na listagem, a linha correspondente no mapa também é selecionada e tem sua cor alterada; selecionando uma quadra/face no mapa, uma linha da listagem é marcada. Este relacionamento entre as duas bases de dados somente foi possível de ser realizado nas áreas urbanas, onde a associação do cadastro de endereços com as linhas que representam as faces de quarteirão no mapeamento foi feita através de códigos. O código de cada quadra/face acoplado ao código do setor censitário cria uma identificação única, que permite a associação do cadastro de endereços ao mapeamento censitário e, conseqüentemente, a localização espacial dos endereços ali existentes. A codificação de uma quadra/face é ilustrada na Figura 11.

Figura 11 - Exemplo de codificação de Quadra/Face.



Fonte: Adaptado de IBGE, 2010b.

### 2.1.2 Unidade geográfica de coleta de dados

O mapeamento censitário tem a finalidade de planejamento, criação e delimitação dos setores censitários, que são as unidades operacionais da pesquisa censitária ou as unidades geográficas de coleta de dados (IBGE, 2013).

Cada setor censitário é classificado de acordo com a sua situação no território e também de acordo com a tipologia da ocupação humana daquele local. Assim, em uma primeira divisão, os setores são classificados como sendo urbanos ou rurais. No entanto, a adoção de uma classificação mais detalhada se faz necessária para melhor representar as situações encontradas *in situ*.

A Tabela 2 mostra a classificação dos setores censitários utilizada no Brasil até o momento, com as suas respectivas quantidades de setores povoados<sup>10</sup> em 2010. Esta classificação é utilizada desde a década de 1970 e está baseada primeiramente nas definições presentes na legislação municipal e, depois, nos casos em que não existe legislação, na observação do local (IBGE, 2003). Essas áreas de enumeração são revistas periodicamente para verificar se houve alguma alteração e, portanto, se a classe utilizada para descrever a ocupação do território deve ser alterada ou não. Na situação urbana, consideram-se as áreas urbanizadas e não urbanizadas internas ao perímetro urbano de cidades ou vilas e as áreas urbanas isoladas; a situação rural abrange toda a área situada fora desses limites, incluindo as extensões das áreas urbanas e os aglomerados rurais isolados, como povoados, núcleos e outros similares (IBGE, 2003). Associada a estas classes de situação há ainda uma classificação segundo a tipologia das construções, em que são consideradas algumas características que tornam necessária a adoção de um tratamento diferenciado na coleta. Os setores que englobam aglomerados subnormais, quartéis, alojamentos, embarcações, aldeias indígenas, presídios, asilos, orfanatos e outros são classificados como setores especiais (IBGE, 2003).

O critério utilizado para a delimitação dos setores censitários é a quantidade de domicílios existentes, os quais o recenseador é capaz de visitar no período destinado para a execução da coleta de dados, que é de aproximadamente um mês. Nas áreas urbanas, a quantidade de domicílios dentro de um setor censitário varia de 250 a 350, enquanto nas áreas rurais esse valor é

---

<sup>10</sup> Em 2010, o número total de setores censitários existentes no país era 317.380, sendo que 7.260 correspondem a áreas não habitadas, cuja delimitação foi realizada exclusivamente para atendimento a requisitos operacionais.

menor, variando entre 150 e 250 domicílios. Nas áreas rurais, a extensão do setor censitário é acrescentada a este critério, sendo que um setor idealmente não deve ter mais do que 500 km<sup>2</sup> (IBGE, 2003). Na prática, estas regras nem sempre são seguidas e é possível encontrar setores censitários com uma quantidade de domicílios superior ao estabelecido como ideal.

Tabela 2 - Setores censitários povoados por situação, Brasil, 2010.

		<b>Situação</b>	<b>Quantidade</b>	
Urbana	Cidade ou vila	Área urbanizada	227 250	
		Área não urbanizada	6 191	
	Área urbana isolada	3 273		
			<b>Total Urbano</b>	<b>236 714</b>
Rural	Aglomerado Rural	De extensão urbana	1 514	
		Isolado	Povoado	9 200
		Núcleo	236	
		Outros	1 281	
	Zona rural		61 175	
			<b>Total Rural</b>	<b>73 406</b>
			<b>Total</b>	<b>310 120</b>

Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

### 2.1.3 Coleta das informações

O Censo Demográfico 2010 contou com duas operações diferentes de coleta das informações: pré-coleta e coleta (IBGE, 2013). Veremos a seguir os objetivos de cada uma dessas operações, quando elas ocorreram e os locais onde elas foram realizadas.

### 2.1.3.1 Pré-coleta

A operação de pré-coleta foi realizada cerca de quatro meses antes do início da coleta do censo (Agosto/2010) e teve por objetivo a listagem dos endereços das unidades residenciais e não residenciais, a classificação do tipo de edificação e a coleta de características dos logradouros identificáveis visualmente, como pavimentação, identificação, arborização e iluminação pública (IBGE, 2010a)<sup>3</sup>. Devido às dificuldades logísticas e ao alto custo das operações de campo, esta etapa foi realizada exclusivamente nos setores censitários urbanos cujas áreas urbanizadas continham ocupação formal (IBGE, 2013). Nesta operação, foi realizada a atualização do registro dos endereços, inclusive com a atualização gráfica das vias. Também foram inseridos ou atualizados os códigos de quadra/face, de maneira que os mesmos pudessem ser associados posteriormente ao mapeamento censitário e permitir a localização geográfica desses segmentos de rua. A Tabela 3 mostra a quantidade de setores censitários povoados que fizeram parte da operação de pré-coleta e da operação de coleta.

### 2.1.3.2 Coleta

A operação de coleta teve início no dia 1º de agosto de 2010 e durou cerca de três meses. Seu objetivo foi a aplicação dos questionários que investigam as características dos domicílios do território nacional, bem como da população residente no país. O modelo de investigação do censo brasileiro contempla a utilização de dois questionários: um longo, aplicado em uma amostra de domicílios, e um curto, aplicado à totalidade de domicílios (IBGE, 2010a). Nas áreas rurais, no momento da aplicação do questionário, foram capturadas as coordenadas geográficas da unidade visitada, enquanto nas áreas urbanas foram capturados e/ou atualizados os endereços.

Tabela 3 - Quantidade de setores censitários povoados nas operações de pré-coleta e coleta, UF, 2010.

UF	Pré-coleta	Coleta		
		Urbano	Rural	Total
11 - Rondônia	1 384	1 392	954	2 346
12 - Acre	483	364	510	874
13 - Amazonas	3 533	3 573	2 068	5 641
14 - Roraima	487	492	332	824
15 - Pará	5 289	5 300	3 469	8 769
16 - Amapá	550	673	137	810
17 - Tocantins	1 211	1 235	866	2 101
21 - Maranhão	4 295	4 108	4 694	8 802
22 - Piauí	2 664	2 734	2 517	5 251
23 - Ceará	8 819	9 030	4 246	13 276
24 - Rio Grande do Norte	2 864	2 906	1 383	4 289
25 - Paraíba	3 587	3 614	1 934	5 548
26 - Pernambuco	7 368	8 501	3 878	12 379
27 - Alagoas	2 304	2 381	1 343	3 724
28 - Sergipe	1 995	2 017	1 280	3 297
29 - Bahia	12 898	13 921	9 861	23 792
31 - Minas Gerais	24 146	24 784	7 780	32 564
32 - Espírito Santo	4 773	5 110	1 270	6 380
33 - Rio de Janeiro	24 387	26 198	1 571	27 769
35 - São Paulo	56 781	60 484	5 612	66 096
41 - Paraná	12 370	12 883	4 582	17 465
42 - Santa Catarina	8 430	9 353	2 529	11 882
43 - Rio Grande do Sul	16 182	16 709	5 623	22 332
50 - Mato Grosso do Sul	2 986	3 097	1 110	4 207
51 - Mato Grosso	4 124	4 211	1 718	5 929
52 - Goiás	7 114	7 398	2 036	9 434
53 - Distrito Federal	4 160	4 100	249	4 349
<b>Brasil</b>	<b>225 184</b>	<b>236 714</b>	<b>73 406</b>	<b>310 120</b>

Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Além de recensear todos os moradores nos domicílios particulares ocupados e nos domicílios coletivos, a operação de coleta cadastra todas as unidades visitadas, residenciais e não residenciais, classificando-as de acordo com a sua ocupação, no caso das unidades residenciais, e de acordo com o seu uso, no caso das unidades não residenciais (IBGE, 2013). Estas informações também são armazenadas no CNEFE.

Durante a operação de coleta, eventualmente ocorrem inclusões de dados referentes a vias ou edificações que não existiam quando foi realizada a operação de pré-coleta. Nesta etapa também ocorrem atualizações de dados referentes, principalmente, à denominação dos logradouros (IBGE, 2013).

## **2.2 Dados**

### 2.2.1 Dados estatísticos

Os dados estatísticos utilizados neste trabalho são originários do questionário do universo do Censo Demográfico 2010. Foram utilizadas variáveis referentes às pessoas e aos domicílios investigados na pesquisa dos estados do Pará e de São Paulo. Para poder associar as informações do questionário à localização dos domicílios é necessário utilizar os dados coletados de forma individualizada, não sendo possível realizar tal tarefa a partir das bases disponibilizadas ao público, pois essas se apresentam agrupadas em unidades de área (setores censitários, distritos, subdistritos, municípios ou unidades da federação). Desta forma, os dados estatísticos utilizados neste trabalho são de acesso restrito e foram cedidos pelo IBGE em caráter extraordinário para elaboração desta tese.

Uma opção para minimizar o risco de quebra de confidencialidade no caso de disponibilização de dados agregados em unidades geográficas pequenas é a utilização de indicadores ou taxas ao invés de números absolutos (BALK;

GUZMÁN; SCHENSUL, 2013). Por exemplo, utilizar a Razão de Dependência<sup>11</sup> ou o Índice de Envelhecimento<sup>12</sup> em vez de idade, ou a Razão de Sexos<sup>13</sup> em vez do gênero. Neste trabalho, optamos por dar preferência à utilização de números absolutos, pois acreditamos que isso aumenta o potencial de utilização dos dados, oferecendo ao usuário a possibilidade de criar seus próprios indicadores, adequando-os ao estudo que está sendo realizado. Outra razão para esta opção está relacionada com a dificuldade que a utilização de taxas ou indicadores traz quando se realiza a agregação de células de um nível hierárquico para um nível superior (ver item 1.3.2.3) ou em unidades de análise definidas pela junção das próprias células, já que a soma dos valores numéricos dessas taxas não corresponde à taxa da nova unidade de análise.

Outra alternativa, que foi adotada nesta tese, é o agrupamento das variáveis em classes mais abrangentes, o que diminui o risco de quebra de confidencialidade. Esta opção foi adotada para as variáveis idade, cor, condição de ocupação do domicílio, abastecimento de água, esgotamento sanitário, coleta de resíduos sólidos, energia elétrica e rendimento domiciliar. Para estas variáveis, a quantidade de classes é menor do que a adotada na divulgação dos dados censitários em agregados de setores censitários. Por exemplo, a variável renda domiciliar apresenta nove classes no agregado de setores censitários e apenas três na grade estatística.

Os dados estatísticos foram processados no programa SPAW Statistics, versão 18, da empresa IBM. Uma das principais operações utilizadas foi a agregação de registros, tanto para a geração de classes, quanto para o

---

<sup>11</sup> Razão de dependência é a razão entre o segmento etário da população definido como economicamente dependente (os menores de 15 anos de idade e os de 60 e mais anos de idade) e o segmento etário potencialmente produtivo (entre 15 e 59 anos de idade), na população residente em determinado espaço geográfico, no ano considerado (RIPSA, 2008).

<sup>12</sup> Índice de envelhecimento é o número de pessoas de 60 e mais anos de idade, para cada 100 pessoas menores de 15 anos de idade, na população residente em determinado espaço geográfico, no ano considerado (RIPSA, 2008).

<sup>13</sup> Razão de sexos corresponde ao número de homens para cada grupo de 100 mulheres, na população residente em determinado espaço geográfico, no ano considerado (RIPSA, 2008).

agrupamento dos registros individuais em domicílios e desses em células da grade estatística.

As variáveis do questionário do universo selecionadas para serem disseminadas utilizando uma grade estatística são:

a. Referentes às pessoas

1. População residente;
2. Sexo: masculino, feminino;
3. Idade: até 9 anos, de 10 a 19 anos, de 20 a 64 anos, acima de 65 anos;
4. Cor ou raça: branca, não branca e
5. Alfabetização: sim, não.

b. Referentes aos domicílios

1. Condição de ocupação: próprio, alugado, cedido, outro;
2. Esgotamento sanitário: rede geral de esgoto ou águas pluviais, fossa séptica, fossa rudimentar, vala ou corpo d'água, outro;
3. Abastecimento de água: rede geral, poço ou nascente, corpo d'água, outra;
4. Coleta de resíduos sólidos: coleta por serviço de limpeza, queimado, enterrado, jogado em terreno baldio ou logradouro, jogado em corpo d'água, outro;
5. Energia elétrica: existe, não existe;
6. Rendimento domiciliar *per capita* em faixas de Salário Mínimo;
7. Espécie de unidade doméstica: unipessoal, nuclear, estendida, composta e
8. Domicílio particular permanente: número.

### 2.2.2 Dados vetoriais

Os dados vetoriais utilizados neste trabalho correspondem aos pontos de localização dos domicílios, às linhas vetoriais das faces de logradouros e às divisões de setores censitários. Os primeiros dados citados são oriundos do CNEFE, e os demais fazem parte do Mapeamento Censitário. Essas duas bases de dados, em sua versão completa, têm acesso restrito e foram cedidas pelo IBGE em caráter extraordinário para utilização nesta pesquisa de doutorado. Foram utilizados neste trabalho dados dos estados do Pará e de São Paulo.

Além dos dados citados acima, também foram utilizadas classificações de uso e cobertura das terras como informação auxiliar. Os dados para o estado do Pará são oriundos do projeto *TerraClass*, executado pelo Instituto Nacional de Pesquisas Espaciais - INPE e pela Empresa Brasileira de Pesquisa Agropecuária - EMBRAPA e são referentes à classificação de imagens obtidas para o ano de 2010 (EMBRAPA; INPE, [s.d.]). Para o estado de São Paulo foi utilizado o Mapa de Cobertura da Terra 2010 - escala 1:100.000, disponibilizado pela Secretaria do Meio Ambiente do Estado de São Paulo, e que também foi realizado com base em imagens de satélite do ano de 2010 (SMA-SP, 2013). Nos dois casos, as diversas classes de uso/cobertura foram agrupadas, sendo as classes relacionadas com características antrópicas classificadas como “povoada” e as classes relacionadas com características naturais como “não povoada”.

## 2.3 Grade Estatística

### 2.3.1 Definição geométrica e espacial

Podemos definir uma grade estatística como sendo um sistema de células regulares georreferenciadas que serve como repositório para dados estatísticos (TRAINOR, 2010). Esta grade deve ter algumas características

mínimas que facilitem a sua utilização pelos usuários, como fácil manipulação, estrutura hierárquica e sistema de codificação simples e que permita o reconhecimento da localização e da escala da célula (JRC, 2003). Pensando pelo lado do produtor de dados, a estrutura da grade deve exigir o mínimo possível de manipulação dos dados estatísticos e locais que serão utilizados na sua geração, de modo que a associação dos dados às células possa ser realizada de maneira o mais simples possível.

A primeira definição para o estabelecimento de uma grade estatística é a escolha do sistema de projeção cartográfica<sup>14</sup>. Alguns países definiram sua grade estatística seguindo os sistemas cartográficos locais, como, por exemplo, Finlândia, Estônia, Noruega e Áustria (EFGS, 2012); outros países adotaram o Sistema de Coordenadas Geográficas, como Japão (JAPAN STATISTICS BUREAU, s.d.). Os mapeamentos globais em grade identificados na revisão bibliográfica (ver Cap. 1, item 1.3.6) também utilizam o Sistema de Coordenadas Geográficas.

O Sistema de Coordenadas Geográficas, também conhecido como Projeção Geográfica, divide a superfície da terra utilizando meridianos e paralelos, que são usados como referência para a medida de distâncias angulares, denominadas longitude e latitude (IBGE, 1998). Este sistema tem seu uso difundido internacionalmente e, por isso, em termos de disseminação de dados, é uma boa opção. Após a popularização de aplicativos de visualização de mapas na internet, como o *Google Earth* e o *Google Maps*, e também com a crescente utilização de aparelhos de GPS, esse sistema se tornou amplamente conhecido, sendo de fácil interpretação mesmo para aqueles não especializados na leitura e interpretação de mapas. Uma desvantagem deste sistema é a geração de distorções nas medidas de distância e área causadas pelo formato não perfeitamente esférico da Terra<sup>15</sup>.

---

<sup>14</sup> Projeção cartográfica é um conjunto de métodos e relações matemáticas utilizadas para representar a superfície terrestre sobre um plano (IBGE, 1998).

<sup>15</sup> O raio da Terra no Equador corresponde a 6.378,38 km e nos polos a 6.359,90 km; o raio médio aproximado comumente utilizado é de 6.370 km.

A utilização de uma projeção que não preserva a área, fazendo com que a mesma varie ao longo da extensão do país, não é uma característica desejável para uma grade estatística, já que do ponto de vista técnico o ideal seria ter unidades com áreas iguais de modo a permitir uma representação espacial dos fenômenos sem distorções (GRASLAND; MADELIN, 2006). A diferença na área de uma célula no extremo sul do país e uma célula na altura do Equador é da ordem de 17%. Apesar de este valor ser considerável, ele é sistemático, e não podemos deixar de considerar que existem outros erros envolvidos na geração de uma grade estatística. Também devemos levar em consideração que os objetivos principais da grade – estabilidade espaço-temporal e integração de dados de fontes diversas - não são impactados pela propriedade de não preservação da área ao longo do território.

A opção pelo Sistema de Coordenadas Geográficas para a grade estatística proposta neste estudo se justifica pela grande difusão do seu uso, pela facilidade de integração com aplicações conhecidas e difundidas pela internet, e, principalmente, pelo fato de que os dados de localização capturados durante o Censo Demográfico 2010 e o Mapeamento Censitário estão nesta projeção.

A segunda definição, tão importante quanto a anterior, é a construção de uma estrutura hierárquica de células. Esse tipo de estrutura permite que se navegue por diferentes dimensões de células, o que conseqüentemente altera a escala de observação dos fenômenos e a escala de agregação dos dados, levando a mudanças também na escala da análise. Um sistema hierárquico também garante a aderência geométrica entre as células de um nível com o nível imediatamente superior.

A terceira definição que deve ser adotada na geração de uma grade estatística se relaciona com a sua nomenclatura. A solução de nomenclatura adotada pelo Japão consiste em aproveitar parcialmente a estrutura do sistema cartográfico existente no país para a geração de uma família de grades (ver Cap. 1, item 1.3.2.6, Quadro 1). O Sistema Cartográfico Brasileiro utiliza um sistema hierárquico que parte da Carta Internacional do Mundo ao Milionésimo, carta

adotada como padrão na Conferência Técnica das Nações Unidas realizada na Alemanha em 1962 (IBGE, 1998). Ao se dividir uma das folhas desta carta nas metades de sua latitude e longitude, serão geradas quatro folhas cuja escala será 1:500.000; desta folha serão geradas outras quatro folhas na escala 1:250.000, e assim sucessivamente até chegarmos à escala de 1:25.000. A partir desta escala, entramos no domínio das cartas chamadas cadastrais<sup>16</sup> e ainda não existem normas aprovadas no país para esta sistematização, existindo diferentes subdivisões utilizadas. O Quadro 3 mostra a escala e a dimensão das cartas do Sistema Cartográfico Nacional estendida até as escalas cadastrais. Além da determinação das dimensões, articulações e escala de cada um dos níveis, este sistema também determina a nomenclatura das cartas. Esta nomenclatura permite que cada carta seja identificada unicamente, além de ter características posicionais, ou seja, permite localizar a folha no território a partir do seu índice de nomenclatura. Este sistema atende vários dos pré-requisitos desejados para a sistematização das células de uma grade estatística, como possuir uma estrutura hierárquica e ter uma nomenclatura estabelecida, o que o torna uma boa opção para o sistema de grades proposto neste trabalho.

A determinação das dimensões das células da grade estatística é influenciada pelos critérios de confidencialidade estabelecidos para a disseminação dos dados censitários, sendo um fator essencial para atingir os objetivos de agregação dos dados censitários em unidades de pequena dimensão, de modo que se possa ter um aumento do poder analítico e uma melhor adequação a recortes espaciais diversos. Alguns países europeus, como Suécia, Finlândia e Estônia, utilizam células com 1 km e com subdivisão de 250 m (EFGS, 2012) para as suas grades estatísticas. O CIESIN utiliza células de 30" e 2,5° nos seus mapas globais de população (BALK *et al.*, 2004a), que em unidades métricas são equivalentes a 1 km e 5 km.

---

<sup>16</sup> O Mapeamento Cadastral compreende mapas em escala grande (maior que 1:25.000), geralmente utilizados para representar cidades e regiões metropolitanas com grande densidade de arruamentos e edificações. As escalas mais usuais são 1:1.000, 1:2.000, 1:5.000, 1:10.000 e 1:15.000 (IBGE, 1998).

Quadro 3 - Escala e dimensão de cartas, folhas topográficas e cadastrais.

Escala	Dimensão	
	grau, minuto, segundo	metro (aproximado)
1 : 1.000.000	6° x 4°	663 438 x 442 292
1 : 500.000	3° x 2°	331 719 x 221 146
1 : 250.000	1°30' x 1°	165 860 x 110 573
1 : 100.000	30' x 30'	55 287 x 55 287
1 : 50.000	15' x 15'	27 643 x 27 643
1: 25.000	7'30" x 7'30"	13 822 x 13 822
1 : 10.000	3'45" x 2'30"	6 911 x 4 608
1 : 2.000	37,5" x 37,5"	1 152 x 1 152
1 : 1.000	18,75" x 18,75"	576 x 576
1 : 500	9,375" x 9,375"	288 x 288
1 : 400	7,5" x 7,5"	230 x 230
1 : 200	3,75" x 3,75"	115 x 115

Fonte: Adaptado de IM [s.d.].

Com o intuito de conjugar o critério de confidencialidade, com a resolução espacial e a representação gráfica, foi realizado um estudo para as áreas rurais dos estados do Pará e de São Paulo que é apresentado no Apêndice A. A seleção desses dois estados se deve às diferenças nos seus padrões de ocupação do território, fato que enriquece as análises devido às diferenças no volume e distribuição espacial da população, como pode ser visto na Tabela 4.

Acrescenta-se a esses fatores o fato de que a avaliação de duas realidades distintas concorre para que as conclusões sejam mais representativas do conjunto do país. A opção por utilizar apenas as áreas rurais no estudo é devido ao fato de que a questão de confidencialidade é mais sensível em áreas com ocupação pouco densa ou esparsa, não ficando bem evidenciada em regiões densamente ocupadas. Outro fator determinante para a escolha desses dois estados foi a existência de projetos em desenvolvimento pela equipe

de professores do Instituto de Filosofia e Ciências Humanas - IFCH e pesquisadores do Núcleo de Estudos de População “Elza Berquó” – NEPO, como o projeto Desflorestamento da Amazônia e a estrutura das unidades domésticas, envolvendo o estado do Pará, o projeto Rede Clima, envolvendo a região costeira de São Paulo, e o projeto Geografia dos Riscos e Mudanças Ambientais: construção de metodologias para análise da vulnerabilidade, envolvendo o município de Limeira, SP.

Tabela 4 - Área e população por situação, Pará e São Paulo, 2010.

<b>Variável</b>	<b>Pará</b>	<b>São Paulo</b>
<b>Urbano</b>		
Setor (núm.)	5 300	60 484
Área (km <sup>2</sup> )	2 189	19 188
População (hab.)	5 191 559	39 585 251
Densidade (hab./km <sup>2</sup> )	2 378	2 063
<b>Rural</b>		
Setor (núm.)	3 469	5 612
Área (km <sup>2</sup> )	1 245 778	227 915
População (hab.)	2 389 482	1 676 907
Densidade (hab./km <sup>2</sup> )	2	7

Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

A primeira conclusão, obtida com o estudo apresentado no Apêndice A, está relacionada com a dimensão das células: elas devem ser preferencialmente menores do que os setores censitários para que se possa assegurar que haverá um ganho no poder analítico dos dados, tanto no que se refere à resolução espacial quanto no que se refere à identificação visual de padrões espaciais. Desta maneira, células com dimensão entre 1 e 2 km (37,5” a 75”) seriam as mais adequadas para o estado do Pará e para o estado de São Paulo, pois com estas dimensões um percentual pequeno de setores censitários (15% em São Paulo e 4% no Pará) não apresentaria perda de resolução espacial. Essas dimensões também proporcionam uma boa identificação visual de padrões espaciais como

pode ser visto nas figuras apresentadas no Apêndice A. Os padrões de distribuição espacial da população ficam mais evidentes com a utilização de células com dimensões de até 75" no Pará e 37,5" em São Paulo. Utilizando-se células com dimensões superiores a esses valores, esses padrões começam a não ficar tão claros e evidentes em análises visuais. Conjugando-se esta identificação de padrões visuais com a necessidade de adequação da grade estatística a recortes espaciais diversos, a melhor escolha seria por células de 37,5" (aproximadamente 1 km), pois, quanto menor a célula, melhor será esta adequação.

A segunda conclusão diz respeito à confidencialidade, sendo que o estudo mostrado no Apêndice A nos leva a crer que um aumento nas dimensões das células, apesar de conduzir a uma maior quantidade de dados liberados para divulgação, não resolve por completo o problema. Isso se deve ao fato de que quanto mais disperso o padrão de ocupação ou quanto menor a densidade populacional, menor a influência do tamanho da célula na quantidade de dados que teriam que ser omitidos para respeitar os critérios de confidencialidade adotados. Assim, o quesito confidencialidade também nos leva em direção a células com dimensões de 37,5" e a um critério de confidencialidade de três domicílios particulares permanentes ocupados como valor limite. Com esses valores, a quantidade de população localizada em células com dados liberados para disseminação pública é igual ou maior do que a quantidade localizada em células não liberadas.

Após analisar as informações referentes às características geométricas e espaciais que uma grade estatística ideal deveria apresentar e as características das grades já produzidas por outras instituições, optamos por gerar uma grade utilizando o sistema de coordenadas geográficas e seguindo o padrão do Sistema Cartográfico Nacional estendido com dimensões de 37,5" ou aproximadamente 1 km para as áreas rurais, com divisão em 25 células de 7,5" ou aproximadamente 230 m para as áreas urbanas. A grade segue a estruturação da Carta ao Milionésimo, tendo como limite norte a latitude 8°N e como limite oeste a longitude

78°W. Com relação à identificação das células, no caso daquelas com dimensão de 37,5", optou-se por adotar a nomenclatura do Sistema Cartográfico Nacional estendido até a escala 1:2.000. No caso das células com dimensão de 7,5" a identificação será feita com base na célula de nível hierárquico superior, acrescida de uma numeração sequencial (1 a 25). No que diz respeito à confidencialidade, o valor limite para liberação das variáveis consideradas sensíveis será de cinco domicílios particulares permanentes, enquanto as variáveis globais (total de domicílios particulares permanentes, total de população, população por sexo) serão liberadas sem restrição. Esse valor contradiz os resultados obtidos com o estudo empírico mostrado no Apêndice A, mas se justifica pelo fato de que os dados do Censo Demográfico 2010 já foram divulgados publicamente com este critério para os agregados de setores censitários (IBGE, 2013), não sendo apropriado suscitar a geração de eventuais divergências nos dados disponibilizados.

Os arquivos vetoriais da grade estatística foram gerados com o programa *ArcMap*, versão 10, desenvolvido e distribuído pela empresa ESRI. Foram gerados arquivos separados para cada folha da Carta ao Milionésimo, devido, principalmente, ao tamanho dos arquivos e aos recursos computacionais necessários para manipulação dos mesmos.

### 2.3.2 Abordagem híbrida

No item 2.1.1 foram discutidos os avanços tecnológicos adotados pelo Censo 2010 e, dentre esses avanços, a associação dos endereços das unidades visitadas aos dados de mapeamento e a captura de coordenadas geográficas são fundamentais para permitir a geração de grades estatísticas com a abordagem de agregação.

Nas áreas rurais, os pontos GPS das unidades visitadas que foram capturados durante a coleta de dados em campo podem ser diretamente

mapeados e relacionados aos dados estatísticos para, então, ser diretamente agregados às células da grade.

Nas áreas urbanas, vislumbram-se três caminhos. O primeiro deles considera o processamento dos dados vetoriais das vias, essas devidamente associadas aos endereços, para a geração de pontos que correspondem a cada domicílio visitado através da utilização de um algoritmo de geocodificação. Este algoritmo utiliza o desenho gráfico das vias e as tabelas contendo a identificação das vias e a numeração das edificações para transformar um endereço literal em um ponto de coordenadas conhecidas no espaço. A qualidade dos endereços é essencial para uma boa geocodificação, principalmente no que se refere à existência de numeração nas edificações, pois o algoritmo utiliza esta informação para a distribuição dos endereços ao longo da via. A geocodificação se baseia na interpolação métrica, tendo como base a numeração dos dois extremos da via e o seu comprimento, ou seja, com base na numeração fornecida para os pontos extremos da via os demais números são interpolados considerando-se a distância entre esses pontos. Desta maneira, endereços com baixa qualidade, como, por exemplo, contendo numeração equivocada, podem levar a uma geocodificação também equivocada. Não havendo numeração, a opção é alocar o endereço em um ponto qualquer do trecho da via, diminuindo a acurácia da espacialização e aumentando o tempo de processamento, tendo em vista que os domicílios sem numeração teriam que ser tratados antes de ser efetuada a geocodificação. Este tratamento se refere à inserção na base de dados de um número válido para a edificação, podendo ser, por exemplo, a média dos dois extremos do trecho de logradouro ou o valor de um desses extremos. No estado do Pará há cerca de 17% de endereços residenciais sem numeração (N = 1.490.170), enquanto em São Paulo o valor é de pouco mais de 5% (N = 14.112.542).

O segundo caminho para a localização espacial dos endereços em áreas urbanas é a utilização das quadras-faces dos logradouros como sendo a unidade mínima de dados e a aplicação de uma regra de ponderação com base no comprimento do trecho para distribuir os dados nas células. Esta opção não é

tão precisa quanto a anterior, no entanto, não necessita da numeração das edificações para ser realizada. A utilização das quadras-faces não é uma abordagem típica de agregação, pois a unidade adotada abriga diversos domicílios. Além disso, esses trechos de logradouros não se encontram necessariamente com toda a sua extensão dentro dos limites da unidade de área de agregação, sendo necessário utilizar alguma estratégia para redistribuir os dados associados aos trechos lineares no interior das células da grade. A ponderação linear é uma das técnicas mais simples que pode ser utilizada, apesar das incertezas associadas ao pressuposto de distribuição homogênea dos domicílios ao longo do trecho de via.

A terceira possibilidade é a conjugação dos dois caminhos anteriores, ou seja, utilizar a geocodificação para as edificações com numeração existente e utilizar a codificação de quadra/face para a localização das edificações sem numeração. Uma restrição a esta alternativa é a diferença nas incertezas envolvidas em cada um dos processos, o que pode levar a uma perda de acurácia no processo conjugado. A operacionalização deste processamento é mais complexa do que as anteriores, já que necessita ser dividida em etapas paralelas distintas devido à utilização de duas técnicas diferentes para o tratamento dos dados.

Os três caminhos apresentados para a espacialização dos endereços urbanos são tecnicamente viáveis, mas considerando-se a qualidade e a quantidade dos dados, o segundo se configura como sendo mais adequado, uma vez que a perda de acurácia é compensada pela agilidade no processamento. Observou-se que a maioria das faces apresenta extensão inferior a dimensão das células, estando, portanto, quase ou totalmente inseridas numa única célula. Nesses casos, o resultado obtido com o método que utiliza a codificação das quadra/faces é equivalente aquele obtido com a geocodificação, com a vantagem de requerer menor potência e tempo de máquina, além de não necessitar de divisão do processamento em etapas diferentes e complementares.

Com relação à abordagem para a associação dos dados censitários à grade, constatou-se a existência de uma quantidade significativa de registros sem dados de localização, fazendo com que a abordagem de agregação não seja suficiente para tratar a totalidade dos dados. Nas áreas urbanas existem duas fontes potenciais para a ausência de dados de localização. A primeira é a malha viária, que pode estar incompleta, e a segunda é a ausência de codificação desta malha viária, o que impede o seu relacionamento com os dados estatísticos. Estas duas situações se apresentam geralmente em áreas não urbanizadas, áreas urbanas isoladas e aglomerados rurais. No estado do Pará, cerca de 10% do total de setores censitários povoados se encontram nesta situação (899 setores de um total de 8.769); no estado de São Paulo, o percentual é menor, cerca de 5% (3.013 setores de um total de 66.096). No caso das áreas estritamente rurais, nem todas as edificações tiveram as suas coordenadas geográficas registradas, devido a razões operacionais e/ou técnicas. No Pará existem 2.612 setores censitários em zonas rurais (29,8% do total estadual) e em São Paulo existem 5.068 (7,7% do total no estado). Considerando as áreas urbanas e rurais, temos 40% do total de setores censitários povoados no Pará e 13% em São Paulo com potencial ausência de localização espacial. Esses setores correspondem a 99,8% da área total do estado do Pará e a 95% do estado de São Paulo, abrigando, respectivamente, 32% e 6,5% da população total estadual. A diferença entre os valores para os dois estados avaliados é devida principalmente à tipologia da ocupação do território, uma vez que o Pará apresenta, de uma maneira geral, setores rurais extensos e setores urbanos pequenos, ambos com baixa densidade populacional, enquanto São Paulo apresenta setores rurais e urbanos com menor extensão, mas com ocupação mais densa (ver Cap. 1, item 2.3.1, Tabela 4).

Diante desses números, constatamos que não é possível empregar apenas a abordagem de agregação para poder representar todos os dados do Censo 2010 em uma grade estatística, sendo necessária a utilização de uma abordagem híbrida, combinando agregação e desagregação. Nos locais onde a totalidade ou um grande percentual dos registros tem dados locacionais, foi

utilizada a abordagem de agregação; naquelas regiões onde isso não ocorreu, a abordagem de desagregação foi utilizada. Com a adoção desta estratégia ainda haverá um percentual de dados que não estará representado através da grade estatística, ou seja, o total de população registrado no Censo 2010 será maior do que o valor obtido com a grade estatística. No entanto, esta diferença pode ser considerada desprezível, e a metodologia se apresenta como viável diante das condições apresentadas.

Diante da decisão de utilização de uma abordagem híbrida, fez-se necessário determinar o limite para utilização de cada abordagem e o método de desagregação mais adequado. Para a determinação do limite de utilização de cada uma das abordagens foi realizada uma avaliação quantitativa da espacialização dos dados, ou seja, avaliou-se a quantidade de registros que tiveram a sua localização geográfica possível de ser realizada. Para isso, utilizou-se o setor censitário como unidade espacial e a quantidade de domicílios permanentes ocupados como variável de avaliação. A quantidade de domicílios foi calculada de duas maneiras distintas para cada setor censitário: somando o número de registros no banco de microdados e somando a quantidade de registros localizados espacialmente com utilização de relacionamentos entre a base geográfica e as bases de endereços e de dados coletados. Com esses dois resultados em mãos, calculou-se a diferença entre eles, a qual foi denominada “ausência de localização”. O mesmo cálculo foi feito para cada setor censitário, sendo registrada a ausência de localização relativa por setor. A ausência de localização relativa foi classificada em faixas e foi obtida a quantidade de setores e de domicílios pertencentes a cada uma dessas faixas. Esta avaliação foi realizada para todo o estado do Pará e apenas para o litoral do estado de São Paulo devido a questões operacionais relacionadas com o acesso aos dados no momento da realização das análises. O Gráfico 1 apresenta os resultados obtidos para as duas regiões.

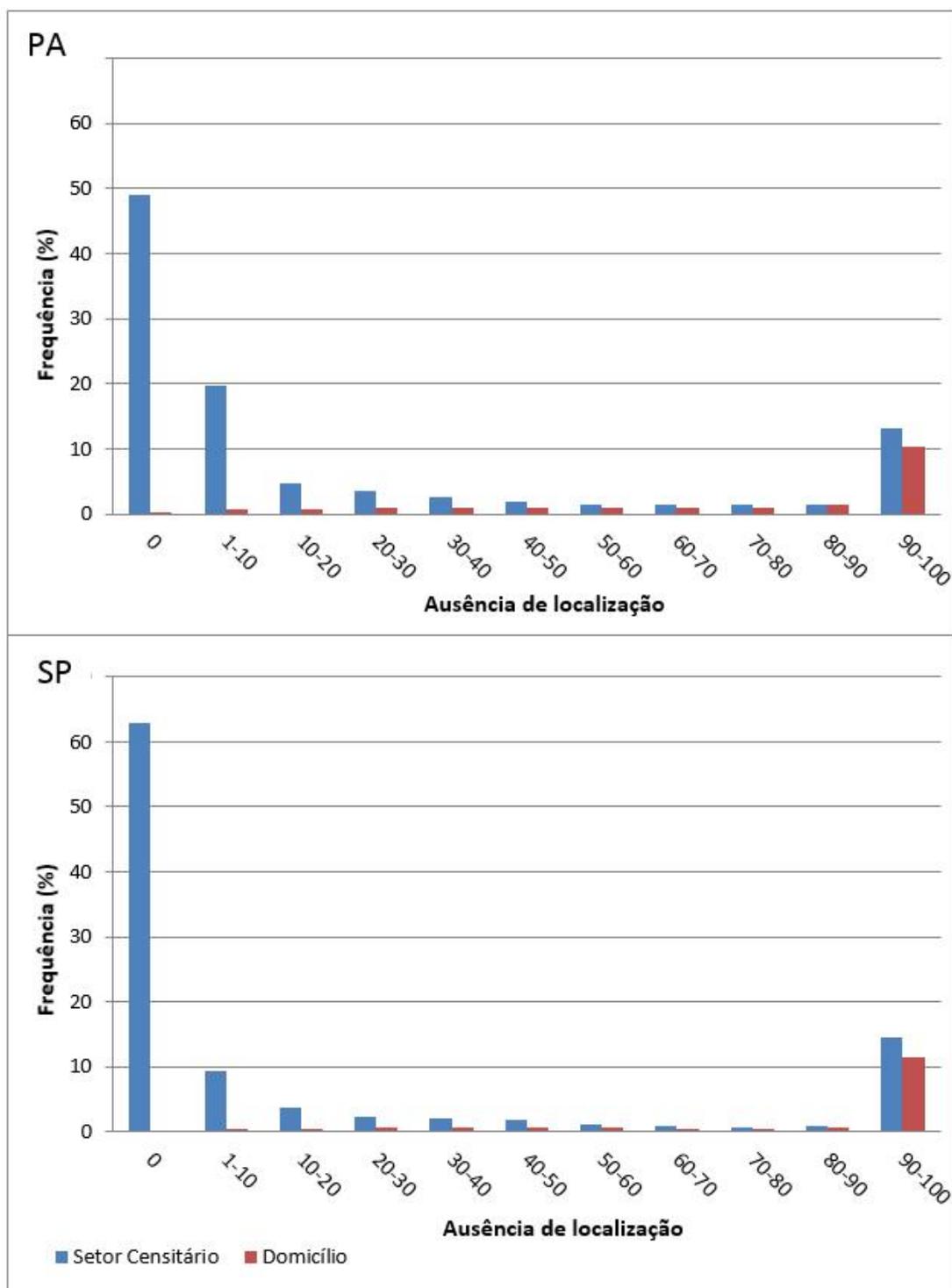
Analisando o Gráfico 1 podemos observar que a quantidade de setores censitários onde não se observa ausência de localização é maior no litoral de São

Paulo em comparação com o estado do Pará. No entanto, isso se inverte para a primeira faixa de ausência de localização (1 – 10%), com o valor obtido para o Pará sendo quase o dobro do valor obtido para o litoral de São Paulo. As demais faixas de ausência de localização apresentam valores semelhantes para as duas regiões. Com relação às quantidades de domicílios sem dados de localização, observamos que os valores são significativos apenas para a última faixa (90 – 100%), na qual os valores estão próximos de 10%. A soma dos percentuais desta variável para as demais faixas de ausência de localização é da ordem de 5% em São Paulo (litoral) e 8% no Pará.

Com base nesses gráficos, podemos concluir que a quantidade de dados censitários não considerados na grade estatística ao se adotar como limite para a determinação da abordagem um valor de ausência de localização de 90% é inferior a 10% nas duas regiões avaliadas, embora a quantidade de setores censitários seja de aproximadamente 25% em São Paulo e 40% no Pará.

A conjugação do tempo de processamento computacional necessário para tratar os dados (menor na metodologia de agregação), a necessidade de dados auxiliares para aplicação das técnicas de desagregação e as incertezas envolvidas na metodologia para geração de dados agregados em grades (maiores na metodologia de desagregação), leva a crer que a estratégia de agregação é a melhor opção. No entanto, devido às questões apresentadas e discutidas anteriormente relacionadas com a ausência de dados locais, a opção foi por uma abordagem híbrida, que mescla agregação e desagregação de acordo com a quantidade de dados não possíveis de localizar espacialmente. Como a quantidade de dados sem localização varia ao longo do território, optou-se por utilizar um valor médio como limite para a adoção de cada uma das abordagens. Desta maneira, nos setores censitários onde a ausência de localização é inferior a 50%, foi adotada a agregação de dados; a abordagem de desagregação foi adotada nos setores censitários onde a ausência de localização é superior a 50%. Com esta opção, mesmo havendo perda de dados, esta perda será inferior a 5%, que é um valor aceitável diante das circunstâncias apresentadas.

Gráfico 1 - Setores censitários e domicílios sem dados de localização por faixa de ausência de localização, PA e SP (litoral), 2010.



Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

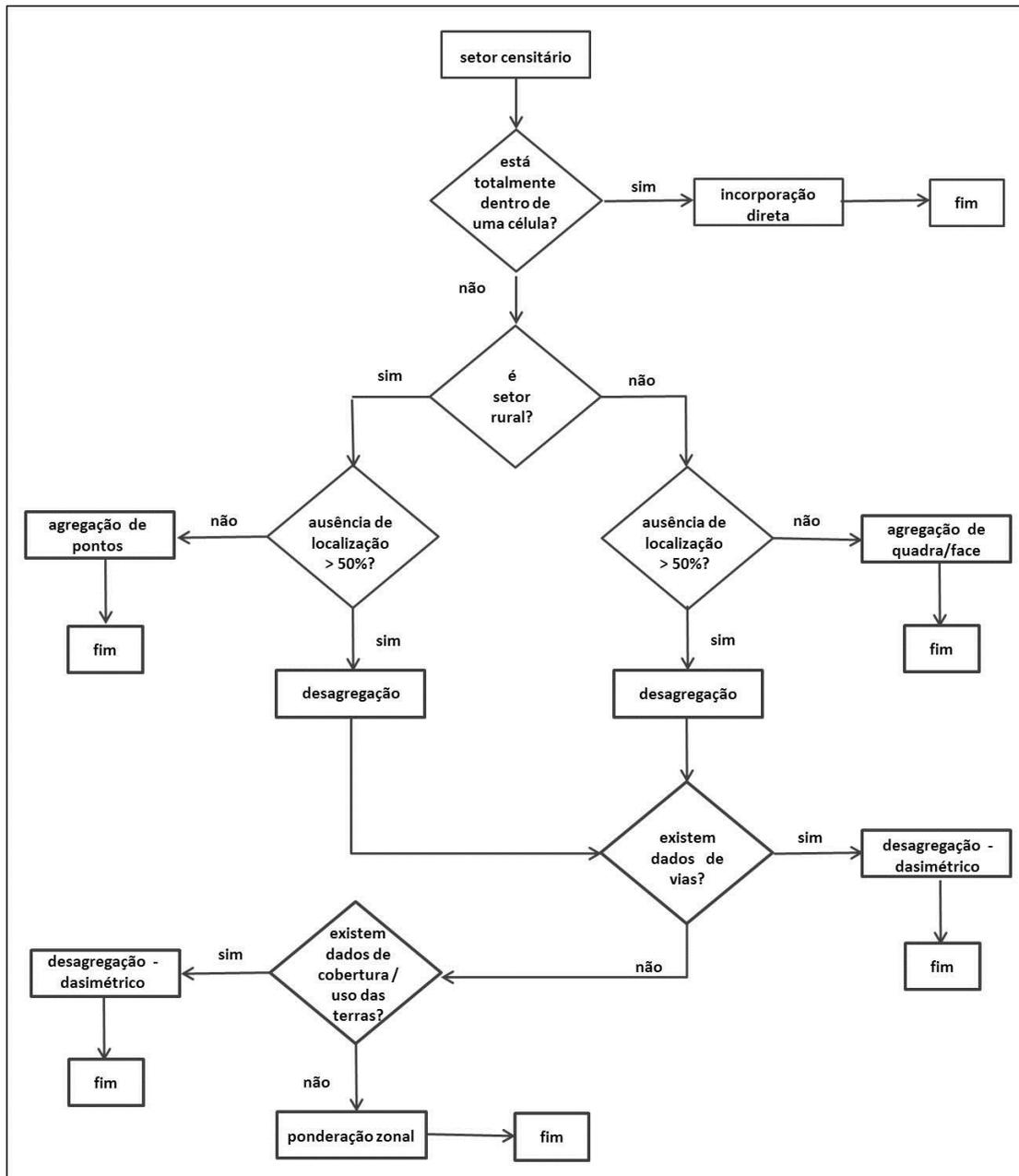
Para auxiliar a escolha do método de desagregação mais adequado, foram realizados dois estudos empíricos, cujos resultados ajudaram a esclarecer algumas questões relacionadas com as diferentes técnicas de desagregação. Esses estudos são apresentados no Apêndice B. A avaliação comparativa entre as técnicas de desagregação sugere que não existe uma técnica melhor do que a outra, pois esta decisão depende do objetivo da análise, da disponibilidade e qualidade de informações para auxiliar a distribuição da população e das características de ocupação da área de estudo. Avaliando os resultados obtidos com as técnicas de desagregação com aqueles obtidos com a técnica de agregação, que foi considerada como referência, e considerando tanto a acurácia do mapeamento quanto o valor estimado da população, foi constatado que o método dasimétrico com dados auxiliares de vias apresenta o melhor desempenho; na segunda posição está o método dasimétrico com dados auxiliares de classificação de uso/cobertura das terras; na última posição está o método de ponderação zonal, que apresenta os piores resultados.

Com relação aos dados auxiliares que foram utilizados na técnica de mapeamento dasimétrico, a primeira opção recai sobre dados detalhados do sistema viário, tanto na área urbana quanto na área rural. No caso da não existência desses dados, a classificação de uso/cobertura das terras a partir de imagens de satélite é a alternativa, não havendo aparentemente diferenças na utilização de uma classificação binária – apenas classe povoada e classe não povoada – ou de uma classificação mais detalhada que considera a densidade de população – três classes povoadas e uma classe não povoada, por exemplo. Na impossibilidade de aquisição de qualquer dado auxiliar, foi utilizado o método da ponderação zonal, mas com o reconhecimento das suas limitações e das incertezas envolvidas.

De posse das opções apresentadas para a escolha da abordagem e da técnica de desagregação foi criada uma árvore de decisão que sintetiza as opções de processamento para popular as células da grade estatística. Esta árvore de decisão, como pode ser visto na Figura 12, tem por unidade o setor censitário e

considera a área do setor com relação à célula da grade, a situação geográfica (rural ou urbana), o valor da ausência de localização espacial e a existência de dados auxiliares.

Figura 12 - Árvore de decisão para escolha da abordagem e da técnica de desagregação.



Fonte: Elaborado pelo autor.

### 2.3.3 Procedimentos

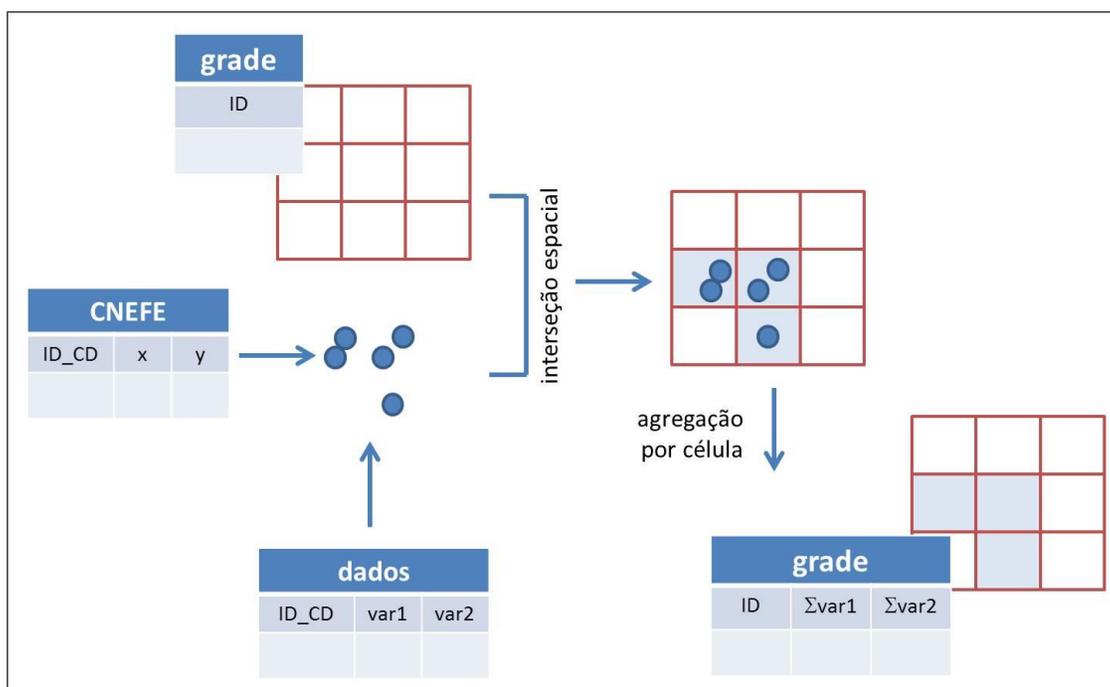
Devido às diferenças na estrutura dos dados vetoriais que representam os domicílios visitados durante o Censo Demográfico 2010 e às diferentes técnicas adotadas para popular a grade estatística, foi necessária a adoção de procedimentos diferentes para o processamento dos mesmos e, conseqüentemente, para a operacionalização da geração da grade estatística. Apesar desta diferença nos procedimentos, isso não significa que eles não possam acontecer numa mesma célula, sendo que nos locais em que isso acontece os dados são somados numa fase posterior aos procedimentos individuais. Nos próximos itens, são descritos com maiores detalhes os procedimentos adotados para popular a grade estatística nas diferentes situações conjugando abordagem e técnica.

#### 2.3.3.1 Agregação de pontos

Nas áreas estritamente rurais, devido à inexistência de endereçamento formal ou devido à fragilidade desse tipo de informação, foram capturadas as coordenadas geográficas dos domicílios visitados pelo censo. Essas coordenadas foram armazenadas em um banco de dados e possuem atributos que permitem a sua ligação ao questionário aplicado. A espacialização desses dados foi feita de maneira direta, bastando transformar as coordenadas armazenadas em formato textual em pontos. Depois deste procedimento, os dados estatísticos foram associados ao ponto com base em um código único identificador do domicílio. Em seguida, foi feita uma interseção espacial entre os pontos e os polígonos da grade, de modo que cada ponto tenha o código da célula da grade onde está localizado. Então, foi realizada uma operação espacial de dissolver, onde todos os pontos que apresentam o mesmo código de célula têm seus atributos somados, gerando uma tabela com registros unicamente identificados. Finalmente, esta tabela foi

associada à camada da grade utilizando como elemento de referência o código da célula. Desta maneira, todos os pontos inicialmente considerados tiveram seus atributos somados e agregados às células da grade estatística utilizando como referência a localização espacial. A Figura 13 apresenta esquematicamente o fluxo dos procedimentos adotados para a agregação de dados pontuais.

Figura 13 - Procedimentos para agregação de dados pontuais.



Fonte: Elaborado pelo autor.

### 2.3.3.2 Agregação de quadra/face

Nas áreas urbanas formais, os domicílios visitados pelo Censo Demográfico 2010 puderam ser espacializados através do endereço registrado durante a coleta de dados. Toda a área classificada como urbana pelo IBGE foi dividida em quarteirões ou quadras formadas por logradouros e representadas graficamente por um polígono. Por sua vez, cada um desses quarteirões foi dividido em faces que correspondem a cada um dos lados deste polígono. Tanto

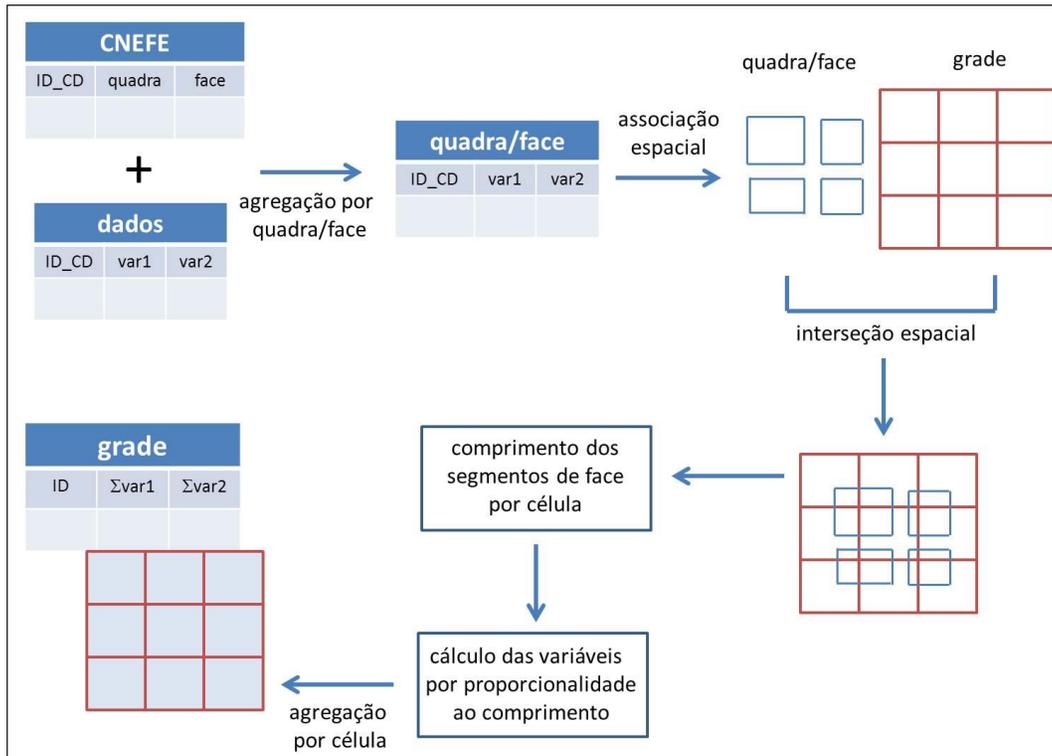
as faces quanto as quadras foram codificadas de maneira única dentro de cada setor censitário, permitindo que os domicílios ali localizados pudessem ser associados a um trecho de logradouro. Por sua vez, o mapeamento censitário tem a representação gráfica das vias urbanas divididas por faces e apresenta na tabela de atributos a sua codificação. Utilizando estes códigos presentes tanto no cadastro de endereços como no mapeamento censitário foi possível localizar espacialmente os domicílios das áreas urbanas que foram visitados pelo censo.

Diante das dificuldades relacionadas à numeração dos domicílios (ver item 2.3.2), a opção para espacializar os domicílios urbanos assumiu que a face era a unidade geográfica de origem dos dados e que o pressuposto de homogeneidade interna era verdadeiro, ou seja, a distribuição dos domicílios foi considerada uniforme na extensão da face. Com base neste pressuposto, a redistribuição dos dados nos casos em que a face não estiver totalmente inserida dentro de uma única célula foi feita utilizando-se a regra de proporcionalidade com base na extensão do trecho de logradouro em questão. A forma de operacionalizar esta regra foi através de uma operação de interseção espacial entre a face de logradouro e a grade estatística, de modo a alocar em cada célula da grade um percentual de dados equivalentes à extensão da face localizada no interior da célula. Apesar do pressuposto de homogeneidade da face não ser verdadeiro, ele pode ser considerado válido, uma vez que a extensão da quadra é pequena e que parte das faces não vai ser dividida por se encontrar totalmente inserida em uma única célula.

A Figura 14 ilustra graficamente os procedimentos adotados para a agregação de dados nas áreas urbanas. Inicialmente, o cadastro de endereços e os microdados do censo foram relacionados utilizando como chave de ligação o identificador único denominado “sequencial de endereço”. Em seguida, esta nova base de dados passou por uma operação de agregação, pela qual todos os registros com a mesma identificação de quadra/face foram somados. Esta base de dados de quadra/face foi, então, relacionada com a base de dados vetoriais das vias utilizando a codificação única de quadra/face que está presente nas duas

bases. Com esta operação, obteve-se uma base vetorial de faces associada às informações do censo, a qual pode ter seus atributos transportados para as células da grade estatística. Este transporte de dados foi feito com uma operação de interseção espacial que permite “quebrar” as faces nas interseções das células da grade, gerando segmentos de linha totalmente inseridos nas células. Após esta operação, calculou-se a extensão desses segmentos de face e, então, foi possível utilizar a regra de proporcionalidade para calcular os valores das variáveis em cada segmento em que a face foi dividida. Para finalizar, a operação de dissolver foi executada com base na identificação única da célula, de modo a totalizar as variáveis correspondentes aos diversos segmentos de diferentes faces que estão inseridos numa mesma célula. A conclusão do processamento foi feita com a associação do resultado da operação anterior à grade estatística utilizando como referência a identificação única das células.

Figura 14 - Procedimentos para agregação de dados lineares.



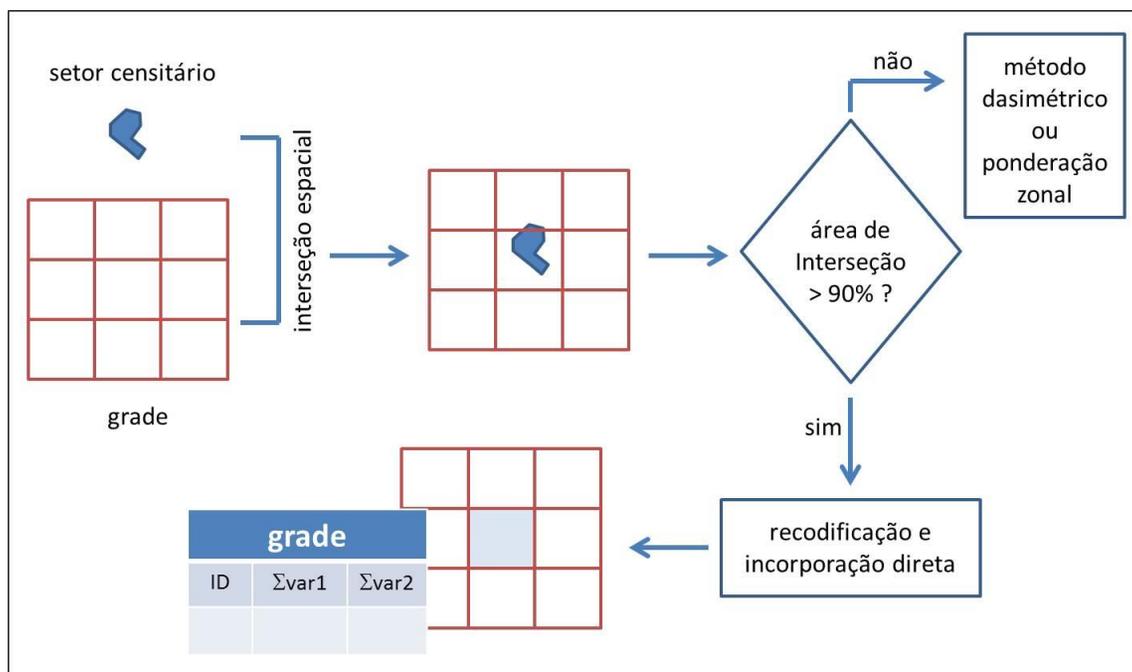
Fonte: Elaborado pelo autor.

A regra de proporcionalidade leva a geração de números não inteiros para as variáveis e, como regra geral, adotou-se que os valores numéricos não serão arredondados em nenhuma etapa intermediária, sendo esta operação realizada apenas no final do processamento. Esta regra teve o objetivo de evitar que os erros de arredondamento se multiplicassem e alterassem demasiadamente os resultados finais.

#### 2.3.3.3 Incorporação direta

Em alguns locais, principalmente em aglomerados rurais, os setores censitários têm dimensões pequenas quando comparado com as células da grade estatística. E por ser pequeno, este setor pode estar totalmente contido dentro de uma única célula. Nesses casos, não houve necessidade de operações espaciais para a transposição dos dados de uma unidade para outra, podendo ser feita uma incorporação direta. Na prática, isso significa que foi necessária apenas uma alteração da codificação dos dados. Nos casos em que o setor não estiver totalmente inserido em uma célula, adotou-se um critério de tolerância: se 90% ou mais da superfície do setor estiver dentro da célula, o setor foi considerado totalmente inserido na célula. Com isso, objetivamos aperfeiçoar a operação, evitando o emprego de recursos em tarefas complexas e que levam a resultados muito próximos. A Figura 15 ilustra a sequência de operações necessárias para realizar a incorporação direta de dados de um setor censitário em uma célula.

Figura 15 - Procedimentos para incorporação direta de dados de setores censitários em células.



Fonte: Elaborado pelo autor.

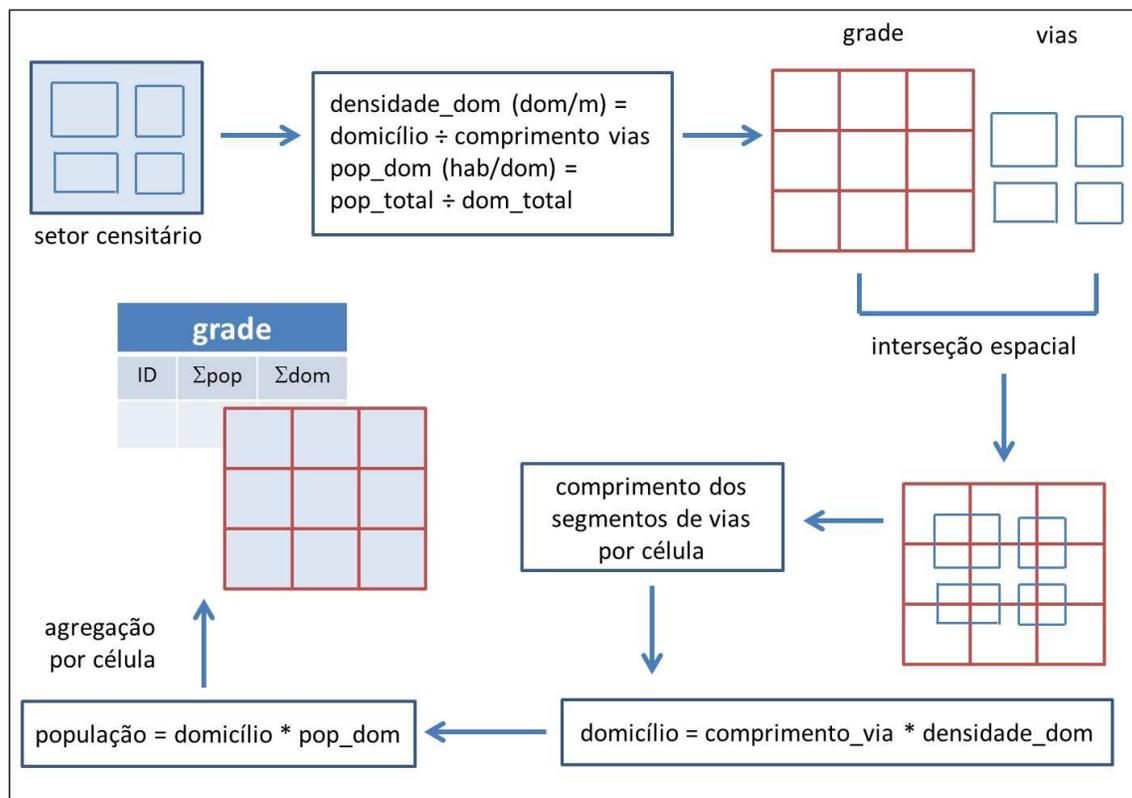
#### 2.3.3.4 Desagregação – método dasimétrico com vias

A metodologia de desagregação foi realizada tendo como unidade de origem os setores censitários e como unidade de destino as células da grade estatística. Os dados da malha viária foram utilizados como aproximação para a existência de residências e, conseqüentemente, de população.

Inicialmente, a camada vetorial de vias necessitou ser editada para que fossem eliminadas algumas feições que poderiam comprometer o cálculo da densidade populacional por extensão de via. Podemos citar como exemplo dessas feições os canteiros centrais, trevos, retornos, pistas de aeroporto e estruturas similares que não têm edificações alocadas ao longo de seu comprimento. A identificação dessas feições foi feita visualmente, havendo claramente um erro potencial envolvido nesta operação. No entanto, este tipo de erro não é significativo, uma vez que a operação de desagregação foi efetuada pontualmente

em regiões de pequena extensão. Após esta operação inicial de edição, teve início a execução de uma série de operações para desagregação dos dados, cujo fluxograma é apresentado na Figura 16.

Figura 16 - Procedimentos para desagregação de dados utilizando malha viária como dado auxiliar.



Fonte: Elaborado pelo autor.

O início das operações se deu com a totalização do comprimento das vias dentro da unidade de origem - o setor censitário. De posse do comprimento total de vias e da quantidade total de domicílios foi realizado o cálculo da densidade de domicílios por extensão de via. Esta densidade foi considerada homogênea tanto na unidade de origem quanto na de destino. Em cada célula inserida dentro de um único setor censitário, multiplicando-se a extensão de vias dentro do perímetro celular pela densidade calculada anteriormente, obtivemos o quantidade de domicílios nesta célula. No caso de a célula estar contida em dois

ou mais setores censitários, foi necessário efetuar a operação descrita acima para cada um dos setores separadamente e depois somar os resultados para obter o total de domicílios final da célula. Para cada setor censitário foi calculada a quantidade média de moradores por domicílio a partir da divisão da população residente total pela quantidade total de domicílios existentes naquele setor. Esta quantidade também foi considerada homogênea no setor censitário e na célula da grade e foi utilizada para calcular a população nas células a partir da quantidade de domicílios obtida anteriormente através da densidade de domicílios por comprimento de via.

#### 2.3.3.5 Desagregação – método dasimétrico com classificação de uso/cobertura das terras

A desagregação foi realizada tendo como unidade de origem os setores censitários e como unidade de destino as células da grade estatística. Os dados de classificação de uso/cobertura das terras com base em imagens orbitais foram utilizados como uma aproximação para a localização dos domicílios e, conseqüentemente, da população.

O método selecionado para classificação de uso/cobertura das terras foi o método binário, que implica a transformação das classes existentes em apenas duas: uma que contém população, e outra que não contém população. Este método é o mais simples descrito na literatura (LANGFORD *et al.*, 1991) e foi selecionado pela sua simplicidade e facilidade de execução, tendo em vista a necessidade de combinar classificações de origens diversas e que apresentam classes diferentes. As classes relacionadas com corpos d'água e cobertura vegetal nativa foram consideradas como não povoadas e aquelas relacionadas com áreas urbanas ou com agricultura/agropecuária foram consideradas povoadas. Apesar das incertezas envolvidas nesta simplificação da realidade, como a metodologia será utilizada de maneira complementar à metodologia de agregação e será aplicada a áreas pequenas, que abrigam um contingente

populacional também pequeno, consideramos que os erros porventura existentes são desprezíveis.

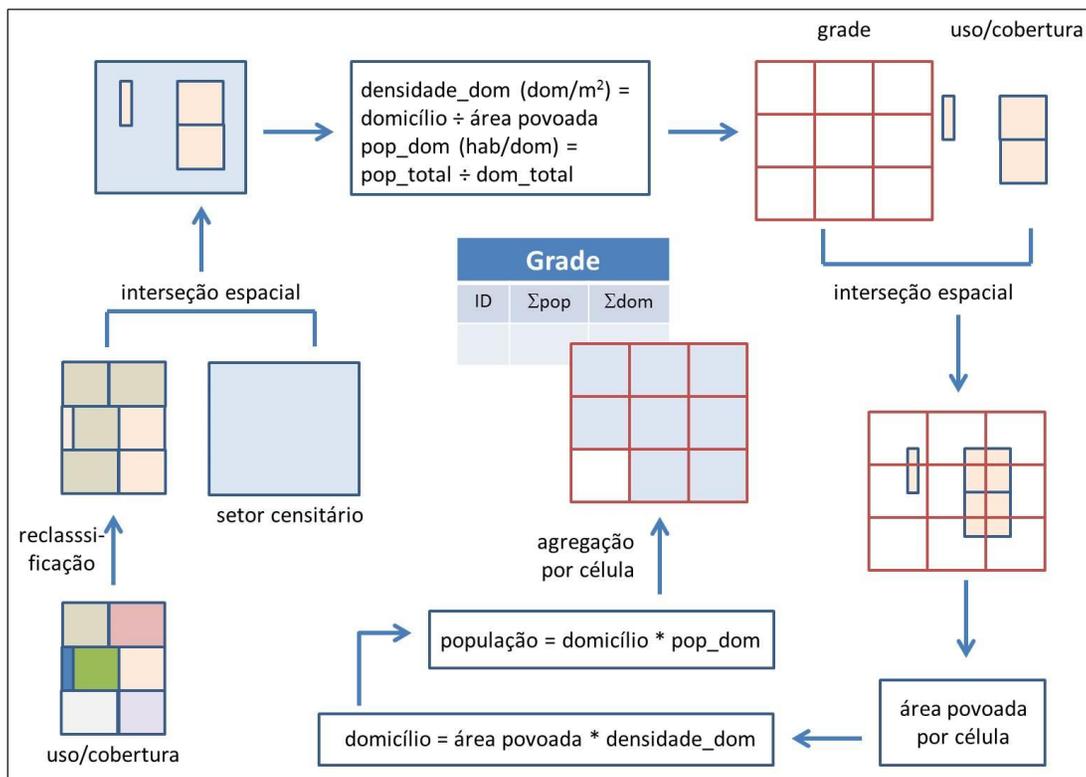
A Figura 17 mostra o fluxo de operações necessárias para o processamento desta técnica. A primeira operação foi a agregação das classes de uso/cobertura em classe povoada e não povoada e, em seguida, a determinação da área da classe povoada para o setor censitário considerado. Utilizando esta área e o total de domicílios no setor, calculamos a densidade de domicílios por unidade de área para o setor censitário em questão. Da mesma forma que no item anterior, esta densidade é considerada constante tanto na unidade de origem quanto na unidade de destino. Após a execução de uma interseção espacial entre o vetor da grade e os polígonos da classe povoada, foi calculada a área de cada um desses novos polígonos contidos nas células. Com estes dados foi calculada a quantidade de domicílios através da multiplicação da área do polígono povoado inserida na célula pela respectiva densidade de domicílios calculada para o setor censitário. O volume de população foi obtido multiplicando-se a quantidade de domicílios da célula pelo número de moradores por domicílio obtido para o setor censitário através da divisão da população residente total pelo número total de domicílios existentes no setor. Mais uma vez ressaltamos que o número de moradores por domicílios foi considerado como sendo um valor homogêneo dentro da unidade embora se saiba que isso não é verdadeiro.

#### 2.3.3.6 Ponderação zonal

Nos casos em que não há possibilidade de agregação de dados e também não existem dados auxiliares que possam auxiliar na metodologia de desagregação foi empregada a ponderação zonal simples para transformar os dados agregados em setores censitários para dados agregados em células da grade estatística. O parâmetro utilizado para realizar esta operação foi a área do setor censitário e o pressuposto de homogeneidade interna mais uma vez foi considerado verdadeiro. A quantidade de domicílios existentes no setor censitário

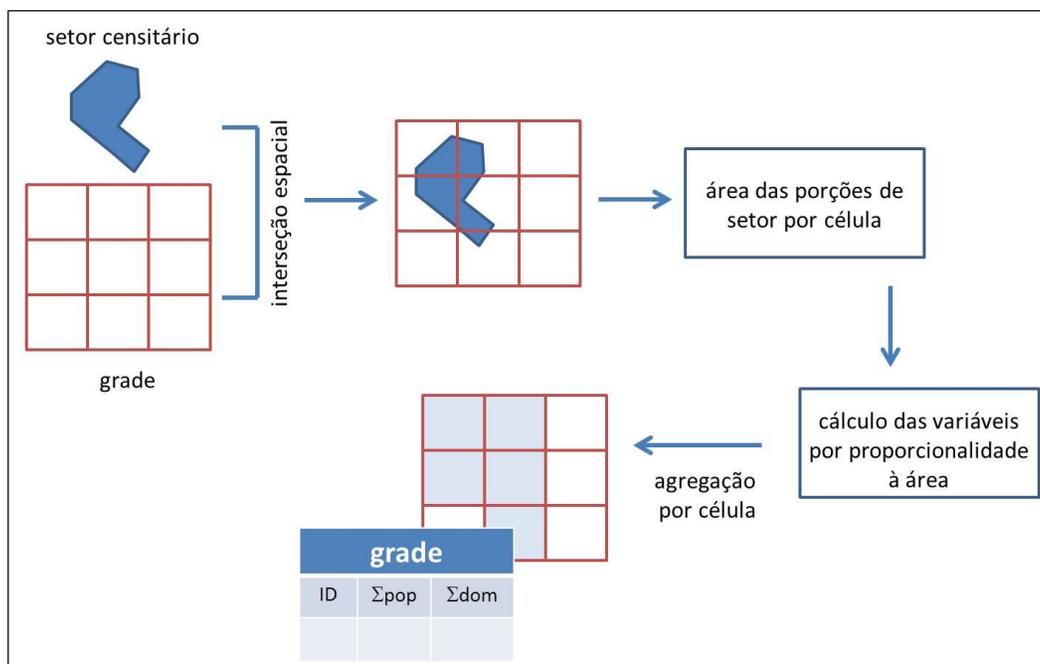
foi redistribuída para as células de acordo com o percentual de área do setor que estava inserido em cada uma das células. O volume de população foi calculado a partir da quantidade de moradores por domicílio existente no setor censitário. A sequência de operações descrita acima está esquematizada na Figura 18.

Figura 17 - Procedimentos para desagregação de dados utilizando classificação de uso/cobertura das terras como dado auxiliar.



Fonte: Elaborado pelo autor.

Figura 18 - Procedimentos para desagregação de dados utilizando ponderação zonal.



Fonte: Elaborado pelo autor.

### 2.3.3.7 Dados quantitativos

Dois procedimentos diferentes foram adotados para obter os valores das características da população e dos domicílios, de acordo com a abordagem adotada. Quando foi utilizada a agregação dos microdados, todas as variáveis foram obtidas a partir de uma simples soma dos dados censitários. No caso da agregação com base em quadra/face, todas as variáveis foram consideradas homogeneamente distribuídas ao longo da face, sendo feita uma ponderação pelo comprimento para obter as variáveis nas células da grade estatística. Depois deste procedimento, as variáveis foram somadas para chegar-se ao resultado final da célula.

Nos casos em que foi utilizada a desagregação, as variáveis “domicílio total” e “população residente” foram obtidas a partir dos cálculos de ponderação aplicados ao dado auxiliar utilizado e explicados nos itens 2.3.3.4, 2.3.3.5 e

2.3.3.6. Para calcular as demais variáveis, a proporção destas variáveis existente no setor censitário foi replicada para as células da grade, ou seja, calculou-se a variável por habitante ou por domicílio no setor censitário e multiplicou-se este valor pela quantidade de população ou de domicílios existente na célula.

Além dos dados relacionados com o censo de população e habitação, foi incluída uma variável para explicitar a abordagem utilizada para a obtenção dos dados em cada célula: agregação, desagregação ou misto (agregação e desagregação). Esta variável tem o objetivo de permitir que o usuário tome conhecimento das diferenças relacionadas com as incertezas que estão envolvidas na geração dos dados agregados.

### CAPÍTULO 3: RESULTADOS E ANÁLISES

*Errors using inadequate data are much less than those using no data at all.  
(Charles Babbage)*

Seguindo a metodologia apresentada no capítulo anterior foi gerada a grade estatística para os estados do Pará e São Paulo, a partir de dados do Censo Demográfico 2010. Este capítulo apresenta a grade estatística gerada e exhibe algumas análises quanto à qualidade do produto. Algumas dessas análises foram realizadas através de aplicações em Estudos de População que exploram diferentes aspectos da grade estatística, revelando as vantagens e os desafios desta forma de disseminação de dados censitários.

A Tabela 5 apresenta as quantidades de células povoadas e não povoadas, por dimensão. A quantidade de células de 230 x 230 m expressa indiretamente o grau de urbanização, uma vez que essas células foram geradas apenas para os locais onde houve interseção entre os setores censitários urbanos com as células de 1.000 x 1.000 m. Pode-se verificar que os dois estados apresentam quantidades relativas bem diferentes – 68,60% em São Paulo e 7,18% no Pará, números que refletem os contrastes entre os padrões de ocupação territorial existentes nesses estados. Outro contraste que pode ser notado a partir desses números é a extensão territorial dos dois estados, que é aproximadamente representada pela quantidade de células de 1 km<sup>2</sup>: o Pará se estende por mais de 1,4 milhões de km<sup>2</sup> enquanto São Paulo não atinge 300 mil km<sup>2</sup>. O estado do Pará apresenta um grande percentual de células não povoadas (90,08% de células de 1 x 1 km e 68,33% de células de 230 x 230 m) enquanto o estado de São Paulo apresenta um percentual menor, principalmente nas áreas rurais (62,23% de células de 1 x 1 km e 56,29% de células de 230 x 230 m).

Tabela 5 - Quantidade de células povoadas e não povoadas, por dimensão, Pará e São Paulo, 2010.

	dimensão (m)	Pará		São Paulo	
		N	%	N	%
Povoada	1 000	141 390	9,92	104 939	37,77
	230	34 927	31,67	265 238	43,71
Não povoada	1 000	1 283 718	90,08	172 902	62,23
	230	75 353	68,33	341 643	56,29
<b>Total</b>	<b>1 000</b>	<b>1 425 108</b>	<b>92,82</b>	<b>277 841</b>	<b>31,40</b>
	<b>230</b>	<b>110 280</b>	<b>7,18</b>	<b>606 881</b>	<b>68,60</b>

Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Os quantitativos por setor censitário de cada uma das técnicas utilizadas para o processamento dos dados são mostrados na Tabela 6. Podemos observar que a maioria dos setores censitários – mais de 80% - foi trabalhada utilizando a metodologia de agregação, enquanto menos de 20% dos setores utilizou a metodologia de desagregação.

Tabela 6 - Quantidade de setores censitários por abordagem e técnica para geração de grade estatística, Pará e São Paulo, 2010.

Abordagem	Técnica	Setor censitário			
		Pará		São Paulo	
		N	%	N	%
Agregação	ponto	2 154	24,56	3 000	4,54
	quadra-face	4 843	55,23	53 649	81,17
	incorporação direta	169	1,93	1 171	1,77
		<b>7 166</b>	<b>81,72</b>	<b>57 820</b>	<b>87,48</b>
Desagregação	dasimétrico: malha viária	1 100	12,54	1 989	3,01
	dasimétrico: uso/cobertura	450	5,13	3 384	5,12
	ponderação zonal	53	0,60	2 903	4,39
		<b>1 603</b>	<b>18,28</b>	<b>8 276</b>	<b>12,52</b>
	<b>Total</b>	<b>8 769</b>	<b>100,00</b>	<b>66 096</b>	<b>100,00</b>

Fonte: Elaborado pelo autor.

A pouca utilização da abordagem de agregação com a técnica de pontos no estado de São Paulo (4,54%) em comparação com o estado do Pará

(24,56%) pode ser explicada pela diferença na tipologia de ocupação rural nesses dois estados: o Pará apresenta uma quantidade bem mais expressiva de zonas rurais com ocupação esparsa do que o estado de São Paulo, onde as áreas urbanizadas estão muito mais presentes. A maior extensão da ocupação urbana em São Paulo explica o percentual maior de utilização da abordagem de agregação com a técnica de quadra-face - 81,17% em São Paulo e 55,23% no Pará. Por esta mesma razão, era esperado que a abordagem de desagregação utilizando a malha viária como dado auxiliar apresentasse uma quantidade relativa superior em São Paulo, mas isso não foi constatado, sendo que em apenas 3,01% dos setores censitários de São Paulo foi utilizada esta técnica, contra 12,54% no estado do Pará. A razão para esses números está na deficiência da cobertura da malha viária utilizada no processamento dos dados de São Paulo, o que também ocasionou os altos valores percentuais de utilização da técnica de ponderação zonal neste estado. Em São Paulo, foi observada uma grande quantidade de setores censitários de pequenas dimensões, sem mapeamento de malha viária disponível e com apenas uma classe de uso/cobertura das terras, geralmente, área urbana, tornando a ponderação zonal a única técnica possível para desagregação dos dados nestes locais.

A Tabela 7 mostra as quantidades de células povoadas de acordo com o tipo de abordagem adotada. A distribuição de frequências de cada abordagem é semelhante para os dois estados, ficando a agregação em primeiro lugar, em seguida a desagregação e, por fim, a abordagem mista. As diferenças entre essas quantidades nos dois estados avaliados podem ser atribuídas às especificidades da ocupação territorial.

Tabela 7 - Quantidade de células da grade estatística por abordagem, Pará e São Paulo, 2010.

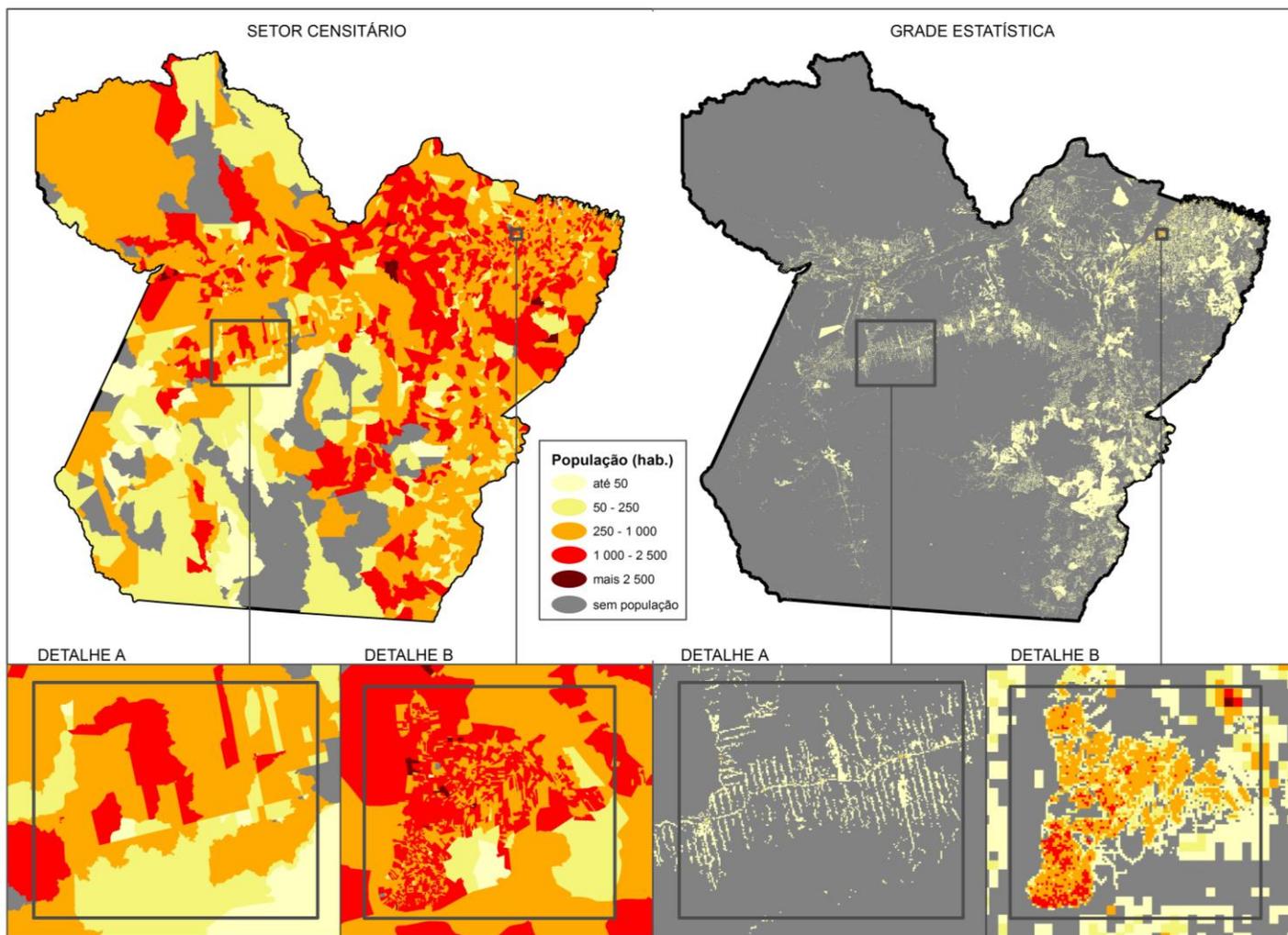
<b>Abordagem</b>	<b>Pará</b>		<b>São Paulo</b>	
	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>
Agregação	96 241	54,58	218 595	59,05
Desagregação	73 979	41,96	126 401	34,15
Mista	6 097	3,46	25 181	6,80
<b>Total</b>	<b>176 317</b>	<b>100,00</b>	<b>370 177</b>	<b>100,00</b>

Fonte: Elaborado pelo autor.

A seguir são apresentadas algumas análises realizadas considerando-se a resolução espacial, os valores numéricos obtidos para as variáveis censitárias e a questão do sigilo estatístico.

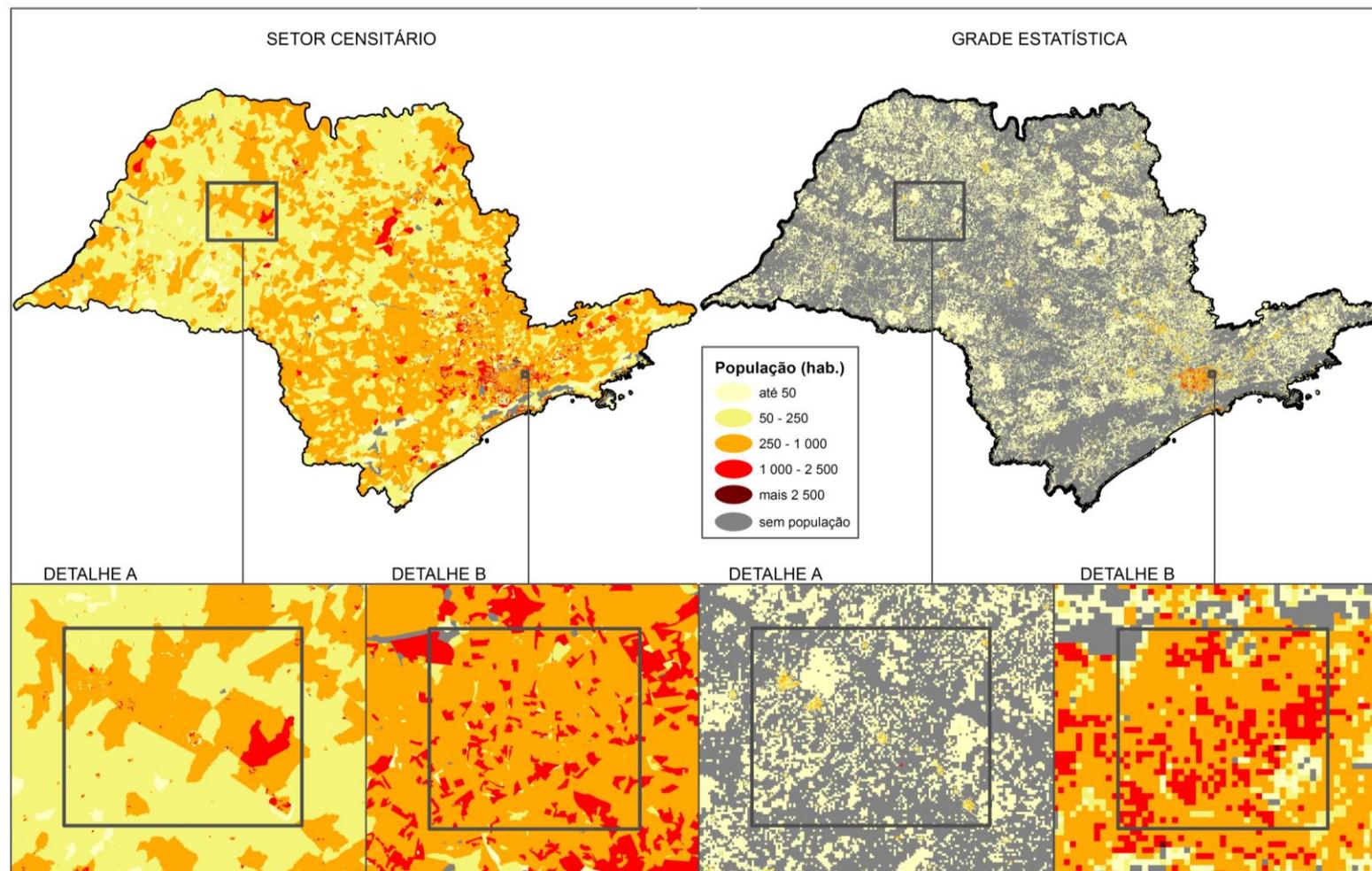
As Figuras 19 e 20 mostram a distribuição da população em São Paulo e Pará com a utilização da grade estatística e setores censitários como unidade espacial. Pode-se constatar visualmente a diferença na resolução espacial dos dados, havendo claramente um maior detalhamento da informação quando se utiliza a grade estatística, principalmente nas áreas rurais e nas áreas urbanas periféricas. Percebe-se, observando o detalhe A da Figura 19, que grandes vetores de ocupação, como rios e estradas, podem ser identificados pela distribuição espacial da população quando se utiliza a grade estatística, pois geralmente a população se aloja nas margens dessas feições, e a resolução espacial dos dados permite esta identificação. No estado de São Paulo (Figura 19), isso não fica tão evidente como no Pará, pois as áreas não ocupadas do território são menos extensas e mais espalhadas. No entanto, conseguimos perceber o alinhamento de áreas urbanas densamente povoadas ao longo dos eixos viários (detalhe A). Nas áreas urbanas dos dois estados, como mostrado no detalhe B, o ganho de resolução foi menor, principalmente nas áreas centrais das cidades, uma vez que os setores censitários urbanos apresentam pequenas dimensões.

Figura 19 - Distribuição da população utilizando setor censitário e grade estatística, Pará, 2010.



Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Figura 20 - Distribuição da população utilizando setor censitário e grade estatística, São Paulo, 2010.



Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

A diferença no volume total de população para os estados do Pará e São Paulo em relação ao total divulgado oficialmente por Unidade da Federação (IBGE, 2011a) foi respectivamente de cerca de 4% e 2%, valores que podem ser considerados muito bons tendo em vista a metodologia aplicada e os dados disponíveis. A existência desta diferença se deve às escolhas metodológicas feitas, notadamente a opção por considerar aceitável um valor para não localização espacial dos dados de até 50% (ver Cap. 2, item 2.3.2)

Em relação às variáveis de caracterização da população ou dos domicílios, não podemos nos esquecer de que foi aplicada uma regra para minimizar a possibilidade de identificação dos respondentes – a regra do sigilo estatístico. Esta regra faz com que parte dos dados - a parte que não atende às regras do sigilo - não seja divulgada publicamente. Por causa disso, não há como comparar os dados obtidos para essas variáveis utilizando a grade estatística e alguma outra unidade de disseminação tradicional.

A aplicação da regra de sigilo estatístico aos dados agregados foi traduzida em termos quantitativos na Tabela 8. Esta tabela apresenta as quantidades de unidades povoadas – células da grade estatística e setores censitários - com até quatro e com cinco ou mais domicílios particulares permanentes ocupados. No primeiro caso, temos a quantidade de unidades atingidas pelo sigilo estatístico; no segundo caso, temos as unidades que não foram atingidas por esta regra. Na mesma tabela, é apresentado o volume de população correspondente a essas unidades, ou seja, a quantidade de população que terá e que não terá suas características divulgadas devido às regras do sigilo estatístico.

Pode-se observar que o estado do Pará apresenta um percentual de células povoadas que abrigam até quatro domicílios permanentes ocupados superior ao estado de São Paulo – 78,12% e 55,60%. Conseqüentemente, a quantidade de população que não poderá ter suas características divulgadas devido ao sigilo estatístico também é maior no Pará do que em São Paulo – 9,85% e 2,35%. Esses números são superiores àqueles obtidos quando se adota o setor

censitário como unidade espacial de agregação dos dados (0,89% e 2,09%) e são decorrentes da adoção do mesmo critério de sigilo para unidades de dimensões diferentes. Ou seja, a aplicação do mesmo critério utilizado para setores censitários faz com que a divulgação em grade estatística tenha uma perda maior no volume de dados de características da população divulgado publicamente.

Tabela 8 - Unidades e população atingidas e não atingidas pelo critério de sigilo estatístico, Pará e São Paulo, 2010.

		Pará		São Paulo	
		N	%	N	%
célula povoada	até 4 domicílios particulares permanentes ocupados *	137 742	78,12	205 806	55,60
	5 e mais domicílios particulares permanentes ocupados **	38 575	21,88	164 371	44,40
	total	<b>176 317</b>	<b>100,00</b>	<b>370 177</b>	<b>100,00</b>
população em célula povoada	até 4 domicílios particulares permanentes ocupados *	718 913	9,85	950 109	2,35
	5 e mais domicílios particulares permanentes ocupados **	6 583 333	90,15	39 470 512	97,65
	total	<b>7 302 246</b>	<b>100,00</b>	<b>40 420 621</b>	<b>100,00</b>
setor censitário povoado	até 4 domicílios particulares permanentes ocupados *	78	0,89	1 379	2,09
	5 e mais domicílios particulares permanentes ocupados **	8 691	99,11	64 717	97,91
	total	<b>8 769</b>	<b>100,00</b>	<b>66 096</b>	<b>100,00</b>
população em setor censitário povoado	até 4 domicílios particulares permanentes ocupados *	8 343	0,11	102 458	0,25
	5 e mais domicílios particulares permanentes ocupados **	7 572 708	99,89	41 159 741	99,75
	total	<b>7 581 051</b>	<b>100,00</b>	<b>41 262 199</b>	<b>100,00</b>

Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

\* Até 4 domicílios permanentes ocupados = atingido pelo critério de sigilo estatístico.

\*\* 5 ou mais domicílios permanentes ocupados = não atingido pelo critério de sigilo estatístico.

Os dados censitários sob a forma de grade estatística foram utilizados em aplicações práticas com o propósito de demonstrar o ganho de poder analítico dos dados disponibilizados neste formato. Estas aplicações não foram realizadas com o propósito de obter uma análise demográfica, mas para gerar elementos que pudessem subsidiar uma discussão sobre as possibilidades analíticas oferecidas. Tendo em vista este objetivo, foram realizadas três aplicações focalizando regiões e temas diferentes, todas dentro do campo de Estudos de População. As áreas de estudo foram selecionadas de modo a auxiliar o entendimento de questões empíricas relacionadas com projetos em desenvolvimento pela equipe de professores do IFCH e pesquisadores do NEPO.

Quadro 4 - Sumário das aplicações utilizadas para analisar o potencial analítico dos dados censitários agregados em grade estatística.

	<b>Aplicação 1</b>	<b>Aplicação 2</b>	<b>Aplicação 3</b>
<b>Localização</b>	Pará	São Sebastião, SP	Limeira, SP
<b>Tema</b>	unidades de conservação ambiental	susceptibilidade a movimentos de massa e inundação	análise intraurbana
<b>Variável</b>	população total, sexo, renda domiciliar, alfabetização	população total, sexo, idade, renda domiciliar, ocupação do imóvel, tipo de unidade domiciliar	população total, sexo, idade, renda domiciliar, cor/raça, alfabetização
<b>Aspecto analítico</b>	adaptação a recortes geográficos diversos	integração de dados físicos/ambientais e sociodemográficos	análise espacial

Fonte: Elaborado pelo autor.

Para cada uma dessas aplicações foram selecionadas algumas variáveis a partir do conjunto de dados censitários disponibilizado. Essa escolha teve por base o assunto principal tratado e a adequação às discussões teóricas e metodológicas em que cada aplicação está envolvida. O Quadro 4 sumariza essas aplicações, mostrando o tema, a localização, as variáveis envolvidas e os aspectos analíticos analisados.

### **3.1 Aplicação 1: Adaptação a recortes geográficos diversos**

A falta de aderência geográfica entre as unidades espaciais utilizadas para disseminação de dados censitários e as áreas de estudo não delimitadas por unidades político-administrativas tradicionais ocasionam algumas dificuldades na realização de análise e estudos, como foi discutido no Capítulo 1. A utilização de dados censitários agregados em uma grade estatística é uma solução para este problema, uma vez que as células da grade são suficientemente pequenas para poderem se ajustar razoavelmente bem aos limites de qualquer recorte geográfico. Esta aplicação ilustra o caso de análise de volume populacional e características dos domicílios localizados em unidades de proteção ambiental do estado do Pará, demonstrando como a utilização da grade estatística possibilita a sumarização de dados para recortes geográficos diversos.

#### **3.1.1 Introdução**

No Brasil, as unidades de proteção ambiental são classificadas em dois grandes grupos: as de proteção integral e as uso sustentável (BRASIL, 2000; artigo 7, §§ 1 e 2). O objetivo principal das unidades do primeiro grupo é a preservação da natureza, sendo permitido apenas o uso indireto dos recursos naturais. No caso do segundo grupo, o objetivo é a conciliação entre conservação da natureza e utilização de recursos naturais. No que diz respeito à ocupação humana, apenas aquelas classificadas como sendo de uso sustentável têm permissão legal para abrigar população no seu interior. No entanto, a presença de população é uma realidade em unidades de pertencentes a ambos os grupos. Neste contexto, é necessário conhecer e acompanhar a dinâmica da população no interior e ao redor dessas áreas protegidas a fim de garantir a manutenção dos objetivos das unidades, para subsidiar planos de manejo adequados (D'ANTONA; BUENO; DAGNINO, 2013), permitir um bom planejamento de infraestruturas,

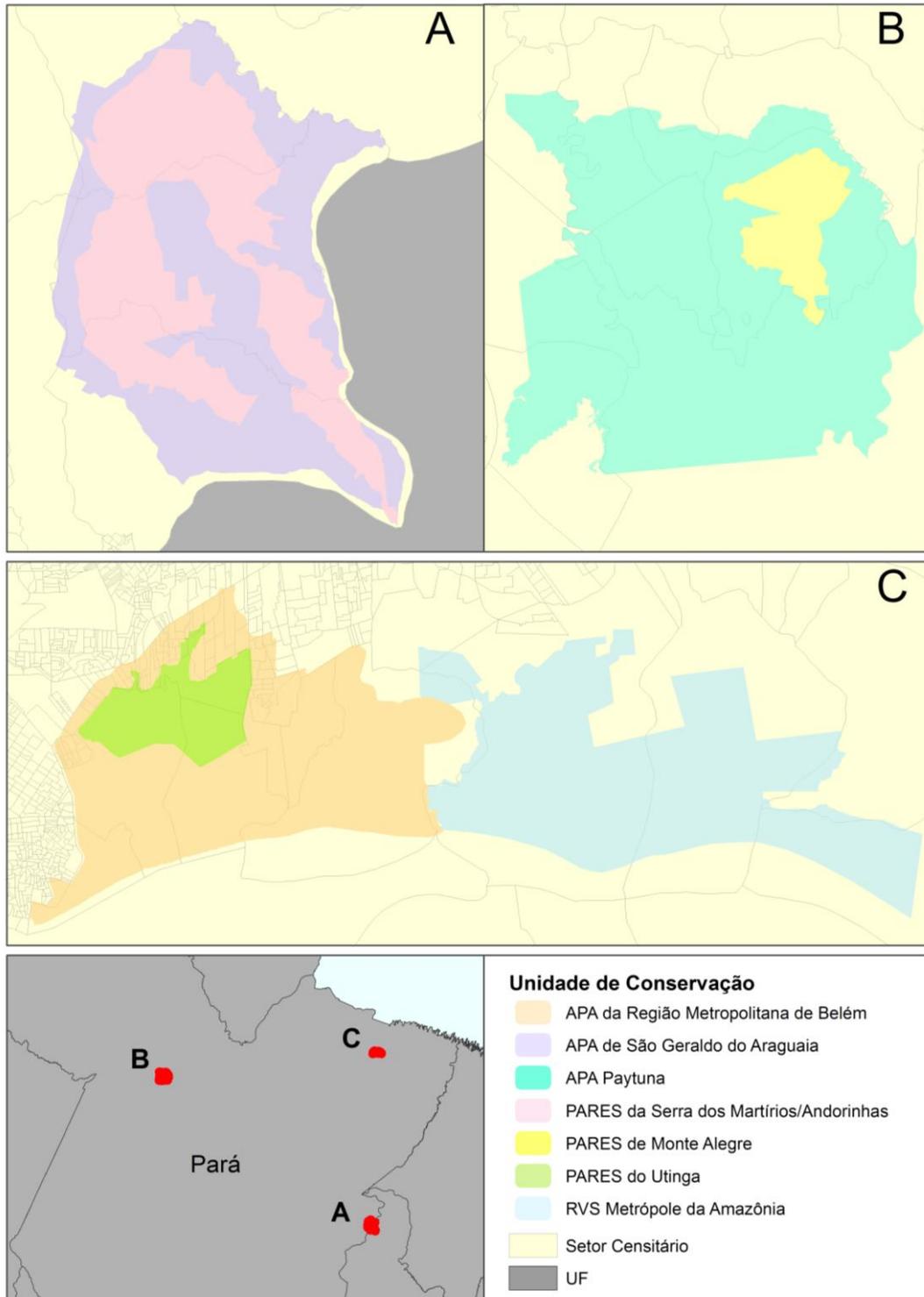
compreender e enfrentar o desmatamento e suas consequências. Infelizmente, apenas algumas unidades de conservação se apresentam isoladas em setores censitários, de modo que os dados coletados possam ser totalizados especificamente para essas unidades. Isto leva administradores e pesquisadores a lançarem mão de pesquisas locais específicas (*surveys*) para coletar dados socioeconômicos e demográficos. Embora essas pesquisas sejam relevantes, principalmente no que diz respeito às questões qualitativas, os altos custos envolvidos não justificam a sua realização para obtenção de dados básicos, uma vez que os mesmos poderiam ser supridos pelos dados do censo de população realizado pelo IBGE.

### 3.1.2 Materiais e métodos

A área de estudo é o estado do Pará, onde foram selecionadas sete unidades de conservação que não têm seus limites totalmente coincidentes com os limites dos setores censitários de 2010, conforme ilustra a Figura 21. As UCs selecionadas são todas administradas pelo estado e no Quadro 5 pode ser vista a categoria e o grupo aos quais elas pertencem.

A camada vetorial dos limites das unidades de conservação utilizada tem como fonte o Cadastro Nacional de Unidades de Conservação (MMA, [s.d.]) e foi a geometria utilizada para realizar as operações espaciais necessárias para obtenção dos totais das variáveis sociodemográficas. A Figura 21 mostra a localização dessas unidades.

Figura 21 - Localização das unidades de conservação selecionadas, Pará, 2010.



Fonte: MMA, Cadastro Nacional de Unidades de Conservação, [s.d.]. Elaborado pelo autor.

Quadro 5 - Unidades de Conservação Ambiental selecionadas, Pará.

<b>Grupo</b>	<b>Categoria</b>	<b>Nome</b>
Proteção Integral	Parque	de Monte Alegre
	Parque	da Serra dos Martírios/Andorinhas
	Parque	do Utinga
	Refúgio da Vida Silvestre	Metrópole da Amazônia
Uso Sustentável	Área de Proteção Ambiental	da Região Metropolitana de Belém
	Área de Proteção Ambiental	de São Geraldo do Araguaia
	Área de Proteção Ambiental	Paytuna

Fonte: MMA, Cadastro Nacional de Unidades de Conservação, [s.d.]. Elaborado pelo autor.

### 3.1.3 Resultados

A Tabela 9 mostra as quantidades de domicílios particulares permanentes e de população existente no interior das UCs avaliadas. Como esperado, as quantidades de domicílios e, conseqüentemente, de população residente, são maiores nas unidades de uso sustentável (APAs) em comparação com as unidades de proteção integral (PARES e RVS).

Tabela 9 - População residente, total e por sexo, e domicílios no interior das unidades de conservação selecionadas, 2010.

<b>Unidade de Conservação</b>	<b>População</b>			<b>Domicílio particular permanente</b>
	<b>Masc.</b>	<b>Fem.</b>	<b>Total</b>	
APA da Região Metropolitana de Belém	35 356	38 121	73 477	19 792
APA Paytuna	2 036	1 791	3 827	964
APA de São Geraldo do Araguaia	356	274	629	156
PARES da Serra dos Martírios/Andorinhas	46	35	81	23
PARES de Monte Alegre	55	59	114	24
PARES do Utinga	1 596	1 746	3 341	902
RVS Metrópole da Amazônia	106	110	216	49

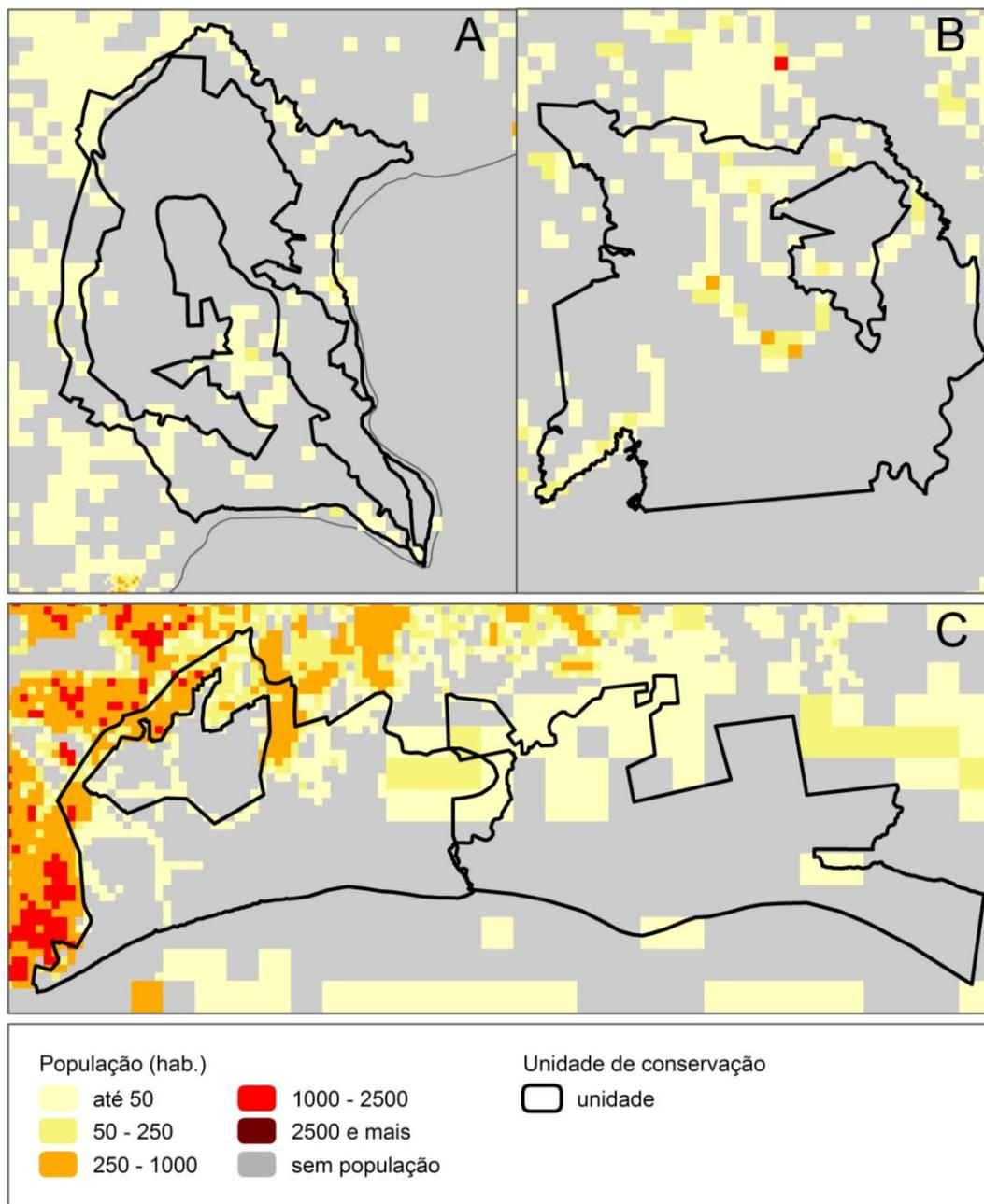
Fonte: MMA, Cadastro Nacional de Unidades de Conservação, [s.d.]; IBGE, Censo Demográfico 2010. Elaborado pelo autor.

A distribuição da população das unidades de conservação estudadas pode ser vista na Figura 22. Algumas categorias de unidades de conservação se localizam nas proximidades de áreas urbanas, o que influencia a quantidade de população nas proximidades dos seus limites. Podemos citar como exemplo desta situação a Área de Proteção Ambiental da Região Metropolitana de Belém (Figura 22, detalhe C). Em outras unidades, é possível perceber que o sistema viário leva a população para o interior da unidade (Figura 22, detalhe B), pois a ocupação se dá de forma linear.

Com relação às variáveis relacionadas com idade e renda domiciliar, a análise foi realizada apenas para as APAs da Região Metropolitana de Belém e Paytuna. Esta limitação foi adotada para garantir um resultado robusto, tendo em vista que as demais unidades têm poucos domicílios no seu interior, além do fato de que muitas células tiveram os seus dados omitidos devido ao critério de confidencialidade adotado. Para efeitos de comparação, essas variáveis também foram totalizadas para uma região abrangendo 20 km no entorno das unidades consideradas.

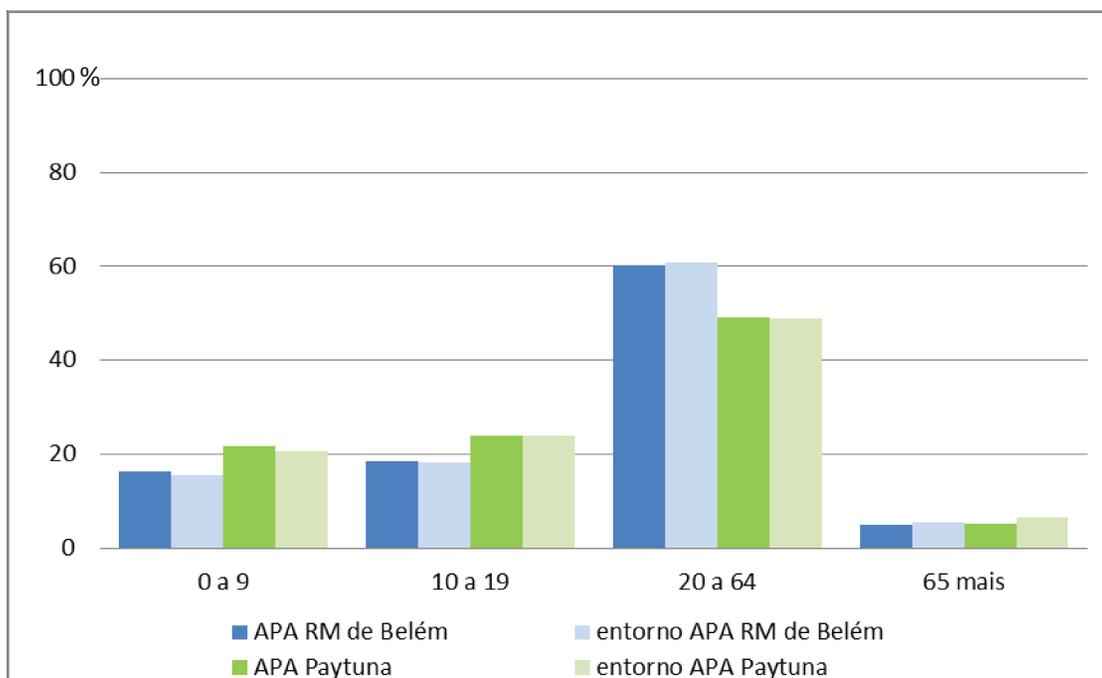
Os resultados referentes à distribuição etária podem ser vistos no Gráfico 2. Pode-se perceber que não há uma diferença significativa entre as quantidades relativas por faixa etária no interior e no entorno das unidades de conservação analisadas. Existe uma quantidade ligeiramente superior de crianças e jovens na APA Paytuna e entorno em comparação com a APA da Região Metropolitana de Belém e entorno, e, conseqüentemente, a quantidade de adultos é ligeiramente inferior. A quantidade de idosos nessas áreas é compatível com o padrão estadual.

Figura 22 - Distribuição da população nas unidades de conservação selecionadas, Pará, 2010.



Fonte: MMA, Cadastro Nacional de Unidades de Conservação, [s.d.], IBGE, Censo Demográfico 2010. Elaborado pelo autor.

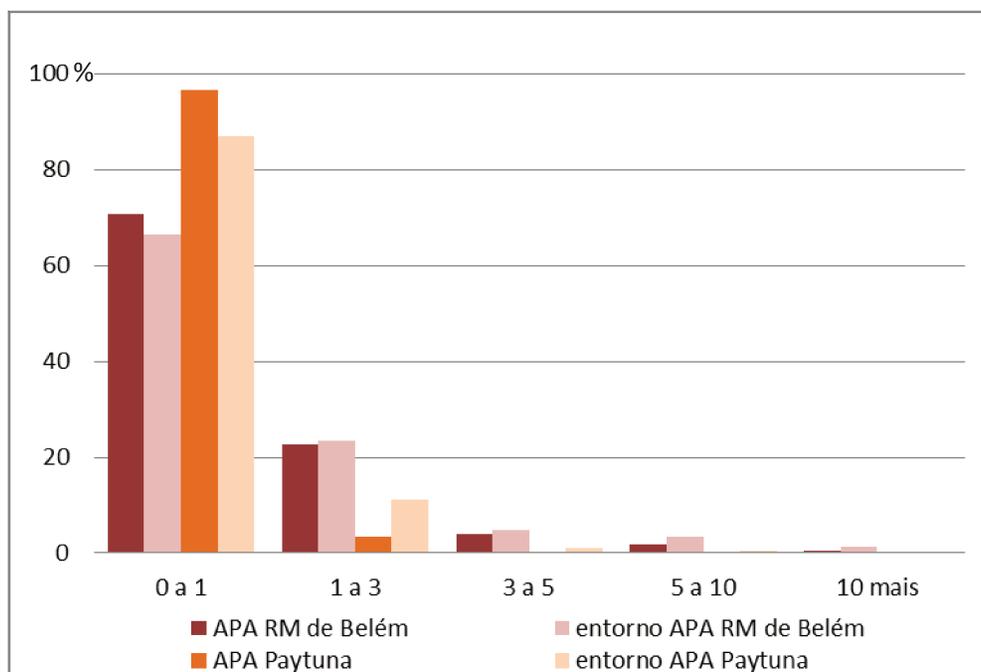
Gráfico 2 - Estrutura etária no interior e entorno de duas unidades de conservação, Pará, 2010.



Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Com relação à renda domiciliar média *per capita* em faixas de salários mínimos (SM), os resultados são mostrados no Gráfico 3. Pode-se observar que há um percentual ligeiramente superior de domicílios com renda média até 1 SM no interior das duas APAs em comparação com o entorno dessas unidades. Na faixa de 1 a 3 SM existe diferença apenas na APA Paytuna, que mostra um percentual menor no seu interior em comparação com a região de entorno. As demais faixas de renda apresentam valores insignificantes. Esses resultados nos levam a concluir que no interior dessas duas APAs reside uma população ligeiramente mais empobrecida do que no seu entorno.

Gráfico 3 - Renda domiciliar média *per capita* em salários mínimos no interior e entorno de duas unidades de conservação, Pará, 2010.



Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

### 3.1.4 Discussão

A utilização de dados censitários em forma de grade estatística oferece a oportunidade de obter dados para recortes geográficos para os quais comumente não há esta disponibilidade, como é o caso das unidades de conservação ambiental analisadas. Uma alternativa para obter informações básicas exclusivamente para essas áreas seria a realização de pesquisas locais, as quais demandam recursos financeiros e técnicos que podem tornar a sua concretização inviável. Outra alternativa seria a utilização dos setores censitários, caso em que o resultado seria pouco preciso por causa da não adequação dos limites das UCs aos dos setores.

Deve-se ressaltar que os números resultantes são aproximados, uma vez que a geometria das células da grade estatística se aproxima do recorte das

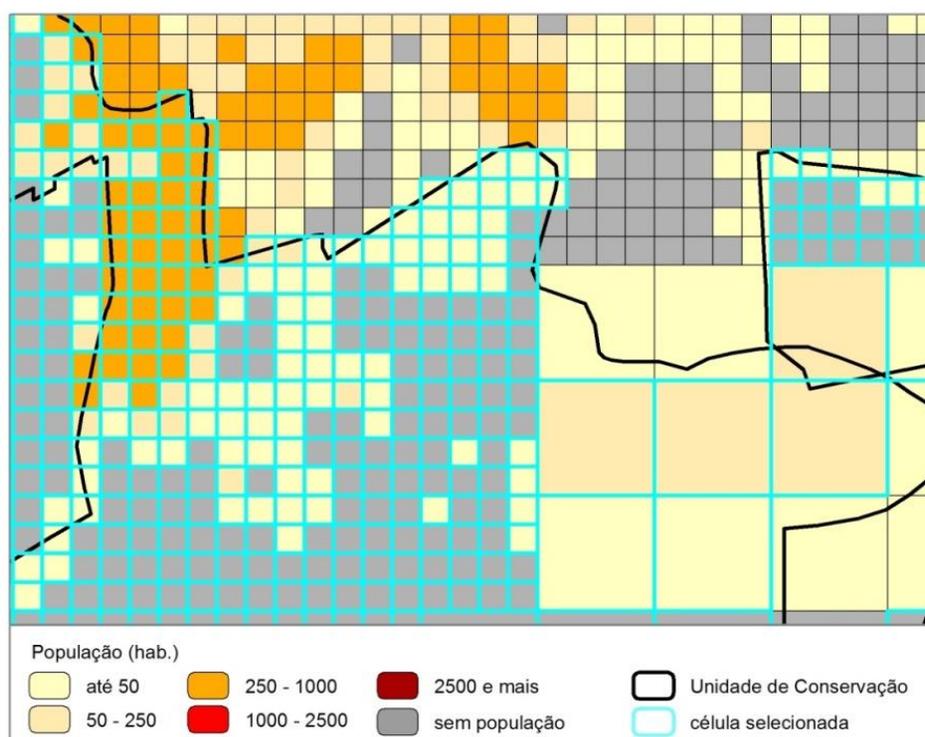
unidades de conservação, mas não é cartograficamente coincidente, como pode ser visto na Figura 23. O método de seleção espacial adotado para totalizar as variáveis analisada considera apenas as células que tem seu centroide localizado no interior da unidade de conservação (células com bordas azuis na Figura 23). Caso esta restrição não fosse adotada, as células que apenas tocam as unidades de conservação seriam consideradas, fazendo com que o total das variáveis fosse superestimado. Além disso, a população não está uniformemente distribuída dentro das células, podendo estar concentrada em uma porção da mesma que não necessariamente está localizada no interior da unidade. Destacamos que esta observação é válida para qualquer recorte geográfico, e não apenas para as unidades avaliadas neste caso de uso.

A questão dos dados omitidos por causa dos critérios de confidencialidade é agravada com a utilização da grade estatística, uma vez que o critério adotado parece não ser compatível com a dimensão da célula, fazendo com que uma grande quantidade de dados não possa ser disponibilizada publicamente. Desta maneira, torna-se necessário dispensar uma atenção especial aos dados obtidos para regiões pouco extensas ou com volume de população pequeno e disperso no território, uma vez que esses dados podem não corresponder à realidade. Para que as análises sejam robustas e consistentes, a quantidade de dados não omitidos deve ser suficientemente grande para poder ser considerada significativa, embora não haja um valor específico recomendado.

Mesmo com as limitações colocadas acima, a oportunidade de trabalhar com dados que mostram as características básicas da população e dos domicílios para áreas não coincidentes com as áreas comumente utilizadas para a disseminação de dados é única, não havendo outros dados disponíveis de forma sistemática e abrangente que permitam analisar essas variáveis com a mesma precisão. Embora a grade estatística ofereça poucas variáveis, elas fornecem subsídios para inúmeras análises, além de auxiliar na determinação de amostras para realização de *surveys*, que objetivam a obtenção de informações mais específicas e detalhadas. A conclusão que pudemos chegar com esta aplicação é que a grade

estatística atende ao propósito de oferecer dados compatíveis com qualquer recorte geográfico desejado, com qualidade espacial e numérica adequada a estudos de população.

Figura 23 - Adequação da grade estatística aos limites das UCs.



Fonte: MMA, Cadastro Nacional de Unidades de Conservação, [s.d.].  
Elaborado pelo autor.

### 3.2 Aplicação 2: Integração de dados

Esta aplicação ilustra a integração de dados ambientais físicos e dados demográficos, mostrando o desafio de utilizar dados de fontes diferentes, reunidos com o objetivo de fornecer elementos para estudos e análises da interação da dimensão humana e dos processos e sistemas ambientais, foco principal do campo do conhecimento denominado População & Ambiente (NEPO, [s.d.]). A área de estudo é o município de São Sebastião, localizado no litoral norte do

estado de São Paulo, onde é feita a integração de dados de suscetibilidade a movimentos de massa e inundação com dados demográficos relacionados com as algumas características da população e dos domicílios.

### 3.2.1 Introdução

No estado de São Paulo, no período compreendido entre os anos 2000 e 2011, ocorreram 2.917 acidentes relacionados com desastres naturais (movimentos de massa, inundações e outros) que levaram ao óbito de 348 pessoas e deixaram desalojadas ou desabrigadas mais de 120 mil (BROLLO *et al.*, 2011). Desse total, 62 ocorreram no litoral norte do estado, causando a morte de seis pessoas e afetando outras 1.796.

Diante dos prejuízos econômicos, físicos e humanos causados pelos desastres naturais, torna-se importante estruturar e colocar em operação alguns planos de prevenção, enfrentamento e redução dessas situações de risco. No estado de São Paulo, o Programa Estadual de Prevenção de Desastres Naturais e de Redução de Riscos Geológicos (SÃO PAULO, 2011) se divide em cinco etapas cíclicas complementares, a saber (BROLLO; TOMINAGA, 2012):

1. Diagnóstico: Tem o objetivo de conhecer o problema e avaliar seu controle e evolução, promovendo o diagnóstico atualizado dos perigos e riscos, inclusive com o estabelecimento de prioridades para mapeamento;
2. Planejamento e Ordenamento Territorial: Esta etapa procura evitar que o problema surja ou aumente, caso já exista, desenvolvendo estratégias de planejamento de uso e ocupação do solo, ordenamento territorial e planejamento ambiental;
3. Monitoramento e Fiscalização: Visa a evitar a ampliação das áreas de risco e a ocorrência de acidentes e a minimizar danos;

4. Redução, Mitigação e Erradicação: Deve promover a adoção de medidas corretivas para eliminar as situações de risco e reduzir as perdas;
5. Capacitação, Treinamento e Disseminação: Busca capacitar e treinar agentes e técnicos para atender às etapas do programa e disseminar informações que possam promover a ampliação de percepção de risco e a participação comunitária na busca de soluções.

Segundo Kobiyama *et al.* (2004), um dos instrumentos mais eficientes para saber onde estão localizadas as áreas vulneráveis é o mapeamento de risco, que serve como subsídio para o planejamento e tomada de decisão. A partir do mapa de risco é possível elaborar medidas preventivas, planificar as situações de emergência e estabelecer ações conjuntas entre a comunidade e o poder público, com o intuito de promover a defesa permanente contra os desastres naturais (PRUDENTE; REIS, 2010). No entanto, apesar de ser uma das bases essenciais para a diagnose do problema, apenas a dimensão física/ambiental está contemplada neste mapeamento; a dimensão humana não é considerada, impossibilitando uma análise completa e integrada da questão. Para suprir esta falha, propomos a associação do mapeamento de risco com dados demográficos e socioeconômicos, de modo a permitir a caracterização da população residente em áreas suscetíveis a desastres naturais, levando a um melhor planejamento e gerenciamento da situação. O objetivo desta aplicação é demonstrar a integração de dados físicos ambientais e dados censitários, ambos produzidos com propósitos distintos e por órgãos diferentes, mas compatibilizados espacialmente para permitir uma análise conjunta.

### 3.2.2 Materiais e métodos

A área selecionada para este estudo abrange o município de São Sebastião, situado no litoral norte do estado de São Paulo, cuja população total residente em 2010 era da ordem de 74.000 habitantes. A partir da primeira metade do

século XIX, com seu auge na década de 70, o município começou a experimentar uma intensa urbanização, devido principalmente a ações para melhorar a acessibilidade da região, incluindo a implantação de estradas. Com isso, houve um grande incremento na exploração turística local, estimulando o crescimento dos setores imobiliários e de construção civil e, conseqüentemente, ocasionando intervenções no meio físico que agravaram em intensidade e em frequência a ocorrência de movimentos de massa e inundações. O agravamento desses fenômenos despertou a preocupação em conhecer os processos geodinâmicos e as maneiras de prevenção e controle de acidentes que envolvem perdas de vidas e financeiras. Nesse contexto, surgiram os estudos que culminaram com a publicação da Carta de Risco a Movimentos de Massa e Inundação do Município de São Sebastião (1996).

O mapa de suscetibilidade, parte desses estudos, foi construído na escala 1:50.000 e teve origem na classificação das unidades básicas de compartimentação fisiográfica quanto à sua maior ou menor suscetibilidade aos processos potencializadores de risco, bem como na tipologia dos processos existentes e/ou esperados em uma dessas unidades. Esta classificação foi feita considerando-se as especificidades de cada compartimento fisiográfico, ou seja, nos planaltos e encostas foram analisados os processos de movimentos de massa, incluindo os diferentes tipos de escorregamentos e a erosão em sulcos e/ou ravinas; nas áreas de baixada, analisou-se a suscetibilidade à inundação; para as praias foi considerada a questão da erosão costeira. O Quadro 6 mostra resumidamente a classificação adotada e a Figura 24 mostra o resultado desta classificação em forma de mapa.

As variáveis demográficas utilizadas nesta integração de dados foram: volume total de população, sexo, idade, renda domiciliar média, forma de ocupação de imóvel e tipo de unidade doméstica.

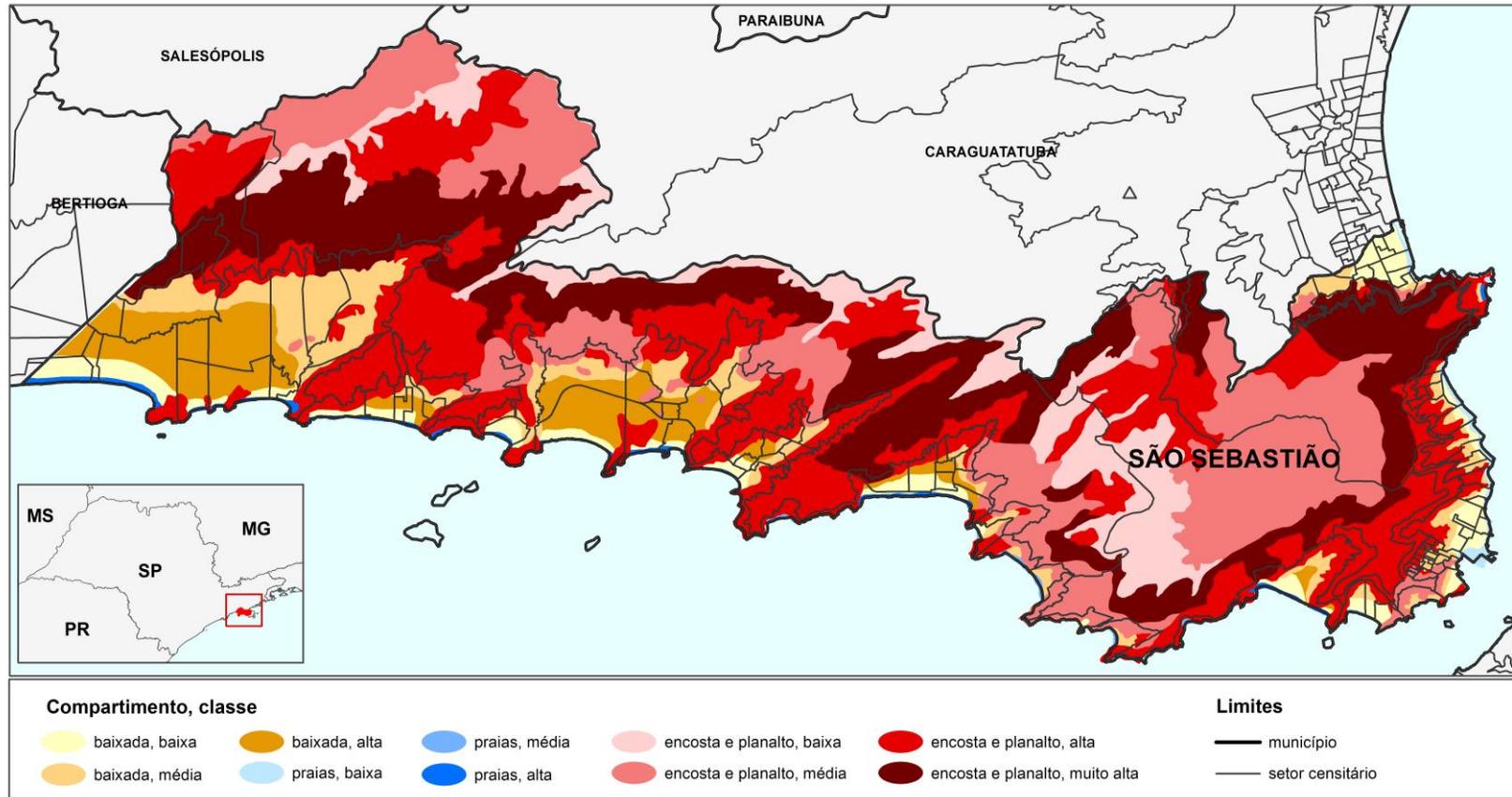
Os polígonos da classificação de suscetibilidade aos processos de risco foram sobrepostos às células da grade estatística permitindo a realização de uma operação espacial cujo resultado nos oferece o volume de população e suas características demográficas para cada associação compartimento-classe.

Quadro 6 - Classe de suscetibilidade por compartimento fisiográfico.

<b>Compartimento fisiográfico</b>	<b>Tipologia dos processos</b>	<b>Classe de suscetibilidade</b>
Encosta e planalto	I – Rastejo II – Escorregamentos de solo e em depósitos de tálus/colúvio	baixa
	III- Quedas, deslocamentos e tombamentos (rochas)	média
	IV – Rolamento de blocos e Matacões V – Corridas	alta
	VI – Processos erosivos do tipo sulcos e ravinas	muito alta
Baixada	VII - Inundação	baixa
		média
		alta
Praia	VIII – Erosão Costeira	baixa
		média
		alta

Fonte: Adaptado de Instituto Geológico, 1996.

Figura 24 - Mapa de suscetibilidade com divisão de setores censitários, São Sebastião, 2010.



Fonte: Instituto Geológico, 1996; IBGE, Censo Demográfico 2010. Elaborado pelo autor.

### 3.2.3 Resultados

Primeiramente, foram analisadas as variáveis censitárias de acordo com o compartimento fisiográfico onde os domicílios estão localizados. A Tabela 10 mostra a distribuição da população segundo o relevo presente no município.

Pode-se observar que aproximadamente dois terços da população total residem em áreas de baixada, enquanto um terço reside em áreas de encosta e planalto; a quantidade de população residente em praias é mínima, equivalendo a menos de 3% do total. A distribuição por sexo é equilibrada, havendo uma pequena preponderância de mulheres nas áreas de baixada e de praia e de homens nas encostas e planalto.

Tabela 10 - População por sexo de acordo com os compartimentos fisiográficos, São Sebastião, 2010.

compartimento fisiográfico	população total		homem		mulher	
	hab	%	hab	%	hab	%
encosta/planalto	22 510	30,86	11 382	31,28	11 129	30,45
baixada	48 499	66,50	24 081	66,18	24 417	66,81
praia	1 926	2,64	925	2,54	1 001	2,74
<b>Total</b>	<b>72 935</b>	<b>100,00</b>	<b>36 387</b>	<b>49,89</b>	<b>36 547</b>	<b>50,11</b>

Fonte: Instituto Geológico, 1996; IBGE, Censo Demográfico 2010. Elaborado pelo autor.

A distribuição etária da população de acordo com o relevo é mostrada na Tabela 11. Foram utilizadas quatro faixas etárias: criança, que abriga pessoas de 0 a 9 anos, jovem, que abriga pessoas de 10 a 19 anos, adultos, com pessoas de 20 a 64 anos e idosos com aquelas pessoas que apresentam mais de 65 anos de idade. Verifica-se que a distribuição é similar em todos os compartimentos fisiográficos, sendo ligeiramente diferente no caso das regiões de praia. Nesta região há, em comparação com os outros compartimentos de relevo, um percentual menor de crianças e jovens e, conseqüentemente, um percentual maior de adultos e idosos.

Tabela 11 - Distribuição da população por faixa etária de acordo com os compartimentos fisiográficos, São Sebastião, 2010.

compartimento fisiográfico	0 a 9 anos		10 a 19 anos		20 a 64 anos		65 anos e +	
	hab.	%	hab.	%	hab.	%	hab.	%
encosta/planalto	3 499	16,74	3 907	18,70	12 653	60,55	838	4,01
baixada	7 155	15,14	8 040	17,01	29 243	61,87	2 829	5,99
praia	170	10,01	248	14,61	1 102	64,90	178	10,48
<b>Total</b>	<b>10 824</b>	<b>15,49</b>	<b>12 195</b>	<b>17,46</b>	<b>42 998</b>	<b>61,55</b>	<b>3 845</b>	<b>5,50</b>

Fonte: Instituto Geológico, 1996; IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Cabe ressaltar que o total de população obtido pela soma dos totais por faixa etária na Tabela 11 é diferente do valor total apresentado na Tabela 10. Esta diferença é devido à supressão de informações para atendimento dos critérios de confidencialidade, o que na prática significa que a população que tem suas características divulgadas é menor do que a população recenseada.

Tabela 12 - Distribuição de renda domiciliar per capita em faixas de Salário Mínimo (SM) de acordo com os compartimentos fisiográficos, São Sebastião, 2010.

compartimento fisiográfico	sem renda		até 3 SM		de 3 a 5 SM		de 5 a 10 SM		mais 10 SM	
	dom	%	dom	%	dom	%	dom	%	dom	%
encosta/planalto	219	3,40	5 710	88,60	285	4,42	187	2,90	44	0,68
baixada	423	2,77	12 996	85,16	1 027	6,73	634	4,15	181	1,19
praia	6	0,97	437	70,48	90	14,52	64	10,32	23	3,71
<b>Total</b>	<b>648</b>	<b>2,90</b>	<b>19 143</b>	<b>85,74</b>	<b>1 402</b>	<b>6,28</b>	<b>885</b>	<b>3,96</b>	<b>248</b>	<b>1,11</b>

Fonte: Instituto Geológico, 1996; IBGE, Censo Demográfico 2010. Elaborado pelo autor.

A Tabela 12 mostra a distribuição da variável renda de acordo com a localização da residência. Os dados nos mostram que a maioria da população possui uma renda domiciliar *per capita* de até 3 salários mínimos, sendo esta classe de renda a maior, qualquer que seja o compartimento fisiográfico. Outra observação relevante diz respeito à alta concentração de domicílios com renda

domiciliar *per capita* de 3 até mais do que 10 SM nas áreas de praia - 28,53%. Isso indica que em relação ao total de domicílios em cada compartimento, existe relativamente uma maior quantidade de domicílios com faixa de renda alta localizados na praia.

Outra variável analisada foi o tipo de ocupação do imóvel, sendo consideradas quatro classes: próprio, alugado, cedido e outra condição. Mais uma vez não se percebem diferenças nos valores entre os diferentes compartimentos de relevo. Há predominância dos imóveis próprios, seguido pelos alugados, cedidos e outro tipo de relação. Observa-se que o percentual de imóveis cedidos nas regiões de praia é maior do que nas outras regiões, enquanto o percentual de domicílios próprios é menor. Os resultados podem ser vistos na Tabela 13.

Tabela 13 - Tipo de ocupação do imóvel de acordo com os compartimentos fisiográficos, São Sebastião, 2010.

compartimento fisiográfico	próprio		alugado		cedido		outro	
	dom.	%	dom.	%	dom.	%	dom.	%
encosta/planalto	4 461	69,31	1 432	22,25	531	8,25	12	0,19
baixada	9 355	61,37	3 798	24,92	2 031	13,32	59	0,39
praia	323	52,18	144	23,26	150	24,23	2	0,32
<b>Total</b>	<b>14 139</b>	<b>63,41</b>	<b>5 374</b>	<b>24,10</b>	<b>2 712</b>	<b>12,16</b>	<b>73</b>	<b>0,33</b>

Fonte: Instituto Geológico, 1996; IBGE, Censo Demográfico 2010. Elaborado pelo autor.

A última variável analisada por compartimento de relevo foi a espécie de unidade doméstica, podendo a mesma ser unipessoal, nuclear, estendida e composta. A unidade unipessoal é composta de apenas uma pessoa; na unidade nuclear está presente um dos pais, ou ambos, e filhos, caso existam; a unidade estendida abriga dois parentes que não se encaixem na definição de unidade nuclear; a unidade composta é constituída por pessoas sem relação de parentesco (IBGE, 2011b). Os resultados são mostrados na Tabela 14. Mais uma vez a distribuição dos valores da variável pelos compartimentos de relevo se

mostrou semelhante, não havendo distinções consideráveis entre os diferentes relevos do município.

Tabela 14 - Espécie de unidade doméstica de acordo com os compartimentos fisiográficos, São Sebastião, 2010.

compartimento fisiográfico	unipessoal		nuclear		estendida		composta	
	dom.	%	dom.	%	dom.	%	dom.	%
encosta/planalto	832	12,91	4 379	67,94	1 050	16,29	184	2,85
baixada	2 321	15,21	10 100	66,18	2 417	15,84	423	2,77
praias	112	18,06	409	65,97	78	12,58	21	3,39
<b>Total</b>	<b>3 265</b>	<b>14,62</b>	<b>14 888</b>	<b>66,68</b>	<b>3 545</b>	<b>15,88</b>	<b>628</b>	<b>2,81</b>

Fonte: Instituto Geológico, 1996; IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Também foram realizadas análises das variáveis demográficas considerando-se a classe de suscetibilidade aos processos potencializadores de risco em cada compartimento de relevo considerado. A Tabela 15 mostra os resultados obtidos para a quantidade de população residente em cada classe de suscetibilidade. Pode-se observar que o volume de população total do município está bem distribuído entre as classes de suscetibilidade baixa, média e alta, havendo uma discreta preponderância desta última. Já a classe de suscetibilidade muito alta apresenta um volume de população pouco significativo, sendo inferior a 2%. Com relação ao quantitativo de população nos compartimentos fisiográficos dentro de cada classe de suscetibilidade, os dados mostram que as áreas de baixada são mais populosas nos casos em que a classe de suscetibilidade é baixa ou média; nos casos em que a classe alta, as áreas de encosta e planalto apresentam uma população maior do que os demais compartimentos de relevo.

Tabela 15 - Distribuição da população por classe de suscetibilidade e compartimento fisiográfico, São Sebastião, 2010.

suscetibilidade	compartimento fisiográfico	população	
		hab.	%
muito alta	encosta e planalto	1 237	100,00
		<b>1 237</b>	<b>1,70</b>
alta	baixada	9 563	35,59
	encosta e planalto	16 828	62,62
	praia	481	1,79
		<b>26 872</b>	<b>36,84</b>
média	baixada	18 022	80,00
	encosta e planalto	4 413	19,59
	praia	92	0,41
		<b>22 527</b>	<b>30,89</b>
baixa	baixada	20 913	93,78
	encosta e planalto	33	0,15
	praia	1 353	6,07
		<b>22 299</b>	<b>30,57</b>
<b>Total</b>		<b>72 935</b>	<b>100,00</b>

Fonte: Instituto Geológico, 1996; IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Com relação à faixa etária da população, a distribuição pelas classes de suscetibilidade pode ser vista na Tabela 16. Observa-se que as faixas jovem e adulta têm uma distribuição bastante similar por todas as classes, no entanto, a faixa etária de idosos tem uma quantidade relativamente menor na classe alta, ao mesmo tempo que apresenta uma quantidade relativamente maior na classe baixa. Esses números podem ser explicados pela maior dificuldade dos idosos em enfrentar situações de risco, que potencialmente necessitam de maior vigor físico. A situação inversa ocorre no caso das crianças, que apresentam uma quantidade ligeiramente maior na classe alta e menor na classe baixa. Uma possível explicação para este fato é a ocorrência de mais nascimentos numa tentativa de repor as perdas ocorridas, embora não haja dados para corroborar esta especulação.

Tabela 16 - Faixa etária da população por classe de suscetibilidade e compartimento fisiográfico, São Sebastião, 2010.

suscetibi- lidade	compar- timento	0 a 9 anos		10 a 19 anos		20 a 64 anos		65 anos e +	
		hab.	%	hab.	%	hab.	%	hab.	%
muito alta	encosta e planalto	148	15,23	186	19,14	573	58,95	65	6,69
		<b>148</b>	<b>15,23</b>	<b>186</b>	<b>19,14</b>	<b>573</b>	<b>58,95</b>	<b>65</b>	<b>6,69</b>
alta	baixada	1 634	18,09	1 708	18,91	5 457	60,42	233	2,58
	encosta e planalto	2 679	16,94	2 918	18,45	9 591	60,64	628	3,97
	praias	44	12,64	57	16,38	227	65,23	20	5,75
		<b>4 357</b>	<b>17,29</b>	<b>4 683</b>	<b>18,59</b>	<b>15 275</b>	<b>60,62</b>	<b>881</b>	<b>3,50</b>
média	baixada	2 659	15,06	2 948	16,70	10 935	61,94	1 111	6,29
	encosta e planalto	665	16,32	796	19,53	2 471	60,62	144	3,53
	praias	7	12,28	5	8,77	41	71,93	4	7,02
		<b>3 331</b>	<b>15,29</b>	<b>3 749</b>	<b>17,21</b>	<b>13 447</b>	<b>61,72</b>	<b>1 259</b>	<b>5,78</b>
baixa	baixada	2 861	13,90	3 384	16,44	12 851	62,44	1 484	7,21
	encosta e planalto	6	18,75	7	21,88	18	56,25	1	3,13
	praias	119	9,21	187	14,47	833	64,47	153	11,84
		<b>2 986</b>	<b>13,63</b>	<b>3 578</b>	<b>16,33</b>	<b>13 702</b>	<b>62,55</b>	<b>1 638</b>	<b>7,48</b>
<b>Total</b>		<b>10 822</b>	<b>15,49</b>	<b>12 196</b>	<b>17,46</b>	<b>42 997</b>	<b>61,55</b>	<b>3 843</b>	<b>5,50</b>

Fonte: Instituto Geológico, 1996; IBGE, Censo Demográfico 2010. Elaborado pelo autor.

A Tabela 17 mostra a distribuição da renda domiciliar *per capita* de acordo com os compartimentos e as classes de suscetibilidade. Pode-se perceber que a faixa de renda de até 3 SM é a que apresenta maior quantidade de domicílios em números relativos para todas as classes de suscetibilidade. Entre os domicílios sem renda, a maior concentração está nas encostas e planaltos com suscetibilidade muito alta. Os domicílios com as maiores rendas (acima de 5 SM) estão localizados preferencialmente nas áreas de praia, com suscetibilidade média e baixa. Nas áreas de encosta e planalto com baixa suscetibilidade, a totalidade dos domicílios se apresenta com renda entre zero e três SM.

Tabela 17 - Distribuição de renda domiciliar *per capita* em faixas de Salário Mínimo (SM) por classe de suscetibilidade e compartimento fisiográfico, São Sebastião, 2010.

susctibi- lidade	comparti- mento	sem renda		até 3 SM		de 3 a 5 SM		de 5 a 10 SM		mais 10 SM	
		dom.	%	dom.	%	dom.	%	dom.	%	dom.	%
muito alta	encosta e planalto	15	5,02	259	86,62	17	5,69	6	2,01	2	0,67
		<b>15</b>	<b>5,02</b>	<b>259</b>	<b>86,62</b>	<b>17</b>	<b>5,69</b>	<b>6</b>	<b>2,01</b>	<b>2</b>	<b>0,67</b>
alta	baixada	93	3,29	2 580	91,36	91	3,22	47	1,66	13	0,46
	encosta e planalto	162	3,31	4306	87,90	227	4,63	164	3,35	40	0,82
	praias	2	1,68	109	91,60	5	4,20	2	1,68	1	0,84
		<b>257</b>	<b>3,28</b>	<b>6 995</b>	<b>89,20</b>	<b>323</b>	<b>4,12</b>	<b>213</b>	<b>2,72</b>	<b>54</b>	<b>0,69</b>
média	baixada	167	2,94	4791	84,38	384	6,76	256	4,51	80	1,41
	encosta e planalto	41	3,32	1135	91,90	40	3,24	17	1,38	2	0,16
	praias	0	0,00	18	75,00	2	8,33	3	12,50	1	4,17
		<b>208</b>	<b>3,00</b>	<b>5 944</b>	<b>85,69</b>	<b>426</b>	<b>6,14</b>	<b>276</b>	<b>3,98</b>	<b>83</b>	<b>1,20</b>
baixa	baixada	163	2,41	5 625	83,22	552	8,17	331	4,90	88	1,30
	encosta e planalto	1	10,00	9	90,00	0	0,00	0	0,00	0	0,00
	praias	4	0,84	310	64,99	83	17,40	59	12,37	21	4,40
		<b>168</b>	<b>2,32</b>	<b>5 944</b>	<b>82,03</b>	<b>635</b>	<b>8,76</b>	<b>390</b>	<b>5,38</b>	<b>109</b>	<b>1,50</b>
<b>Total</b>		<b>648</b>	<b>2,90</b>	<b>19 142</b>	<b>85,75</b>	<b>1 401</b>	<b>6,28</b>	<b>885</b>	<b>3,96</b>	<b>248</b>	<b>1,11</b>

Fonte: Instituto Geológico, 1996; IBGE, Censo Demográfico 2010. Elaborado pelo autor.

A Tabela 18 apresenta os dados relacionados com o tipo de ocupação do imóvel. Percebe-se que a maioria dos domicílios pertence aos tipos próprio e alugado nas áreas de baixada e encosta e planalto, sem diferenciação por classe de suscetibilidade. Já os domicílios localizados em praias são mais frequentemente do tipo cedido, principalmente nas regiões com classe de suscetibilidade alta e média. Nota-se uma ligeira predominância em imóveis próprios nas áreas de encosta e planalto em comparação com as regiões de baixada; já os imóveis alugados são mais frequentes nas baixadas do que nas encostas e planaltos. Não são perceptíveis diferenças relacionadas com a classe de suscetibilidade.

Tabela 18 - Tipo de ocupação do imóvel de acordo com a classe de suscetibilidade e o compartimento fisiográfico, São Sebastião, 2010.

suscetibi- lidade	compartimento	próprio		alugado		cedido		outro	
		dom.	%	dom.	%	dom.	%	dom.	%
muito alta	encosta e planalto	238	79,60	34	11,37	26	8,70	1	0,33
		<b>238</b>	<b>79,60</b>	<b>34</b>	<b>11,37</b>	<b>26</b>	<b>8,70</b>	<b>1</b>	<b>0,33</b>
alta	baixada	1 628	57,79	750	26,62	433	15,37	6	0,21
	encosta e planalto	3 329	68,05	1 124	22,98	429	8,77	10	0,20
	praias	48	40,34	14	11,76	57	47,90	0	0,00
		<b>5 005</b>	<b>63,94</b>	<b>1 888</b>	<b>24,12</b>	<b>919</b>	<b>11,74</b>	<b>16</b>	<b>0,20</b>
média	baixada	3 635	64,10	1437	25,34	560	9,87	39	0,69
	encosta e planalto	888	71,84	272	22,01	74	5,99	2	0,16
	praias	10	41,67	5	20,83	9	37,50	0	0,00
		<b>4 533</b>	<b>65,40</b>	<b>1 714</b>	<b>24,73</b>	<b>643</b>	<b>9,28</b>	<b>41</b>	<b>0,59</b>
baixa	baixada	4 092	60,57	1 611	23,85	1 039	15,38	14	0,21
	encosta e planalto	6	66,67	1	11,11	2	22,22	0	0,00
	praias	265	55,56	125	26,21	85	17,82	2	0,42
		<b>4 363</b>	<b>60,25</b>	<b>1 737</b>	<b>23,99</b>	<b>1 126</b>	<b>15,55</b>	<b>16</b>	<b>0,22</b>
		<b>14 139</b>	<b>63,40</b>	<b>5 373</b>	<b>24,09</b>	<b>2 714</b>	<b>12,17</b>	<b>74</b>	<b>0,33</b>

Fonte: Instituto Geológico, 1996; IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Para finalizar a apresentação de dados, a Tabela 19 mostra o tipo de unidade doméstica de acordo com a classe de suscetibilidade e o relevo. Todos os tipos de unidade estão bem distribuídos pelas classes de suscetibilidade, sendo que prevalece a unidade do tipo nuclear. Em regiões de praia com média suscetibilidade, chama a atenção a quantidade relativa de unidades domiciliares estendidas (8%) e unipessoais (28%), mas estes números podem ser resultantes do fato de haver uma pequena quantidade de domicílios localizados nesta área (25 domicílios em praias com classe de suscetibilidade média).

Tabela 19 - Espécie de unidade doméstica de acordo com a classe de suscetibilidade e o compartimento fisiográfico, São Sebastião, 2010.

susceptibilidade	compartimento	unipessoal		nuclear		estendida		composta	
		dom.	%	dom.	%	dom.	%	dom.	%
muito alta	encosta e planalto	42	14,00	197	65,67	53	17,67	8	2,67
		<b>42</b>	<b>14,00</b>	<b>197</b>	<b>65,67</b>	<b>53</b>	<b>17,67</b>	<b>8</b>	<b>2,67</b>
alta	baixada	422	15,49	1 759	64,57	455	16,70	88	3,23
	encosta e planalto	625	12,76	3 348	68,33	778	15,88	149	3,04
	praias	20	16,81	81	68,07	15	12,61	3	2,52
		<b>1 067</b>	<b>13,78</b>	<b>5 188</b>	<b>67,00</b>	<b>1 248</b>	<b>16,12</b>	<b>240</b>	<b>3,10</b>
média	baixada	843	14,85	3 800	66,92	885	15,59	150	2,64
	encosta e planalto	163	13,19	828	66,99	218	17,64	27	2,18
	praias	7	28,00	14	56,00	2	8,00	2	8,00
		<b>1 013</b>	<b>14,60</b>	<b>4 642</b>	<b>66,90</b>	<b>1 105</b>	<b>15,92</b>	<b>179</b>	<b>2,58</b>
baixa	baixada	1 056	15,63	4 440	65,70	1 077	15,94	185	2,74
	encosta e planalto	2	22,22	6	66,67	1	11,11	0	0,00
	praias	85	17,86	314	65,97	61	12,82	16	3,36
		<b>1 143</b>	<b>15,78</b>	<b>4 760</b>	<b>65,72</b>	<b>1 139</b>	<b>15,73</b>	<b>201</b>	<b>2,78</b>
<b>Total</b>		<b>3 265</b>	<b>14,69</b>	<b>14 787</b>	<b>66,53</b>	<b>3 545</b>	<b>15,95</b>	<b>628</b>	<b>2,83</b>

Fonte: Instituto Geológico, 1996; IBGE, Censo Demográfico 2010. Elaborado pelo autor.

### 3.2.4 Discussão

A utilização de dados censitários em forma de grade estatística permite a integração de dados de fontes distintas, favorecendo a realização de cruzamentos inéditos entre variáveis, o que dificilmente seria possível com a utilização de dados disseminados em unidades espaciais tradicionais. Este é o caso deste estudo em que foi possível analisar diversas variáveis sociodemográficas de acordo com variáveis físicas, como o relevo, ou variáveis compostas, como a classe de suscetibilidade aos processos potencializadores de riscos a movimentos de massa e inundações.

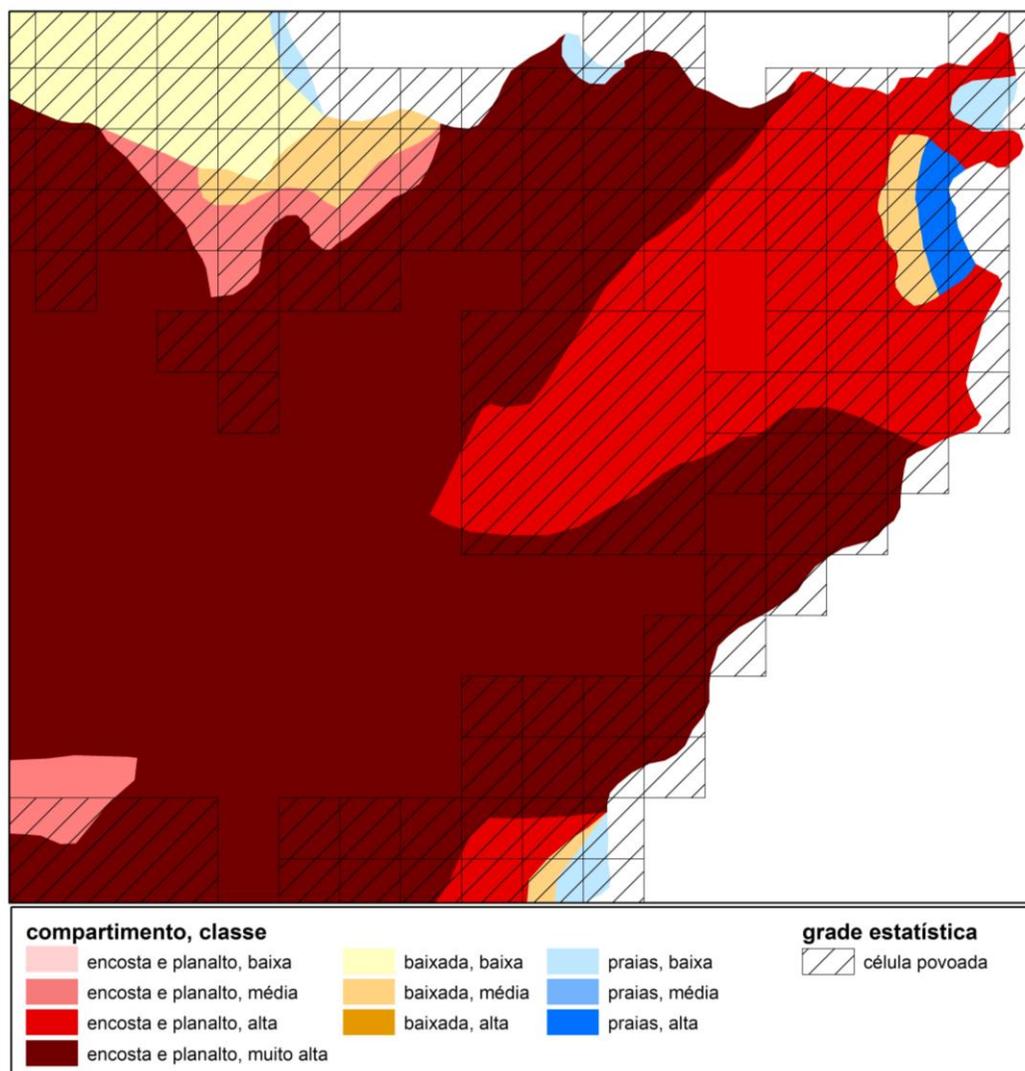
Esta integração de dados se mostrou promissora, uma vez que permite uma análise do ponto de vista social, econômico e demográfico de fenômenos relacionados com aspectos físicos e ambientais do território, como é o caso de acidentes naturais do tipo inundação ou deslizamentos e desmoronamentos.

Uma questão que deve ser analisada com cuidado é a escala dos dados, de modo que a falta de compatibilidade entre eles seja minimizada. Os levantamentos físicos ou ambientais foram feitos em uma determinada escala de trabalho e os dados censitários, apesar de terem sido levantados em uma escala bastante acurada - escala dos domicílios - foram agregados em unidades espaciais e perderam parte desta acuidade. A Figura 25 ilustra este problema de compatibilidade de dados e escalas. O mapeamento de suscetibilidade apresenta polígonos com área variando entre 20.000 e 20.000.000m<sup>2</sup>, enquanto a grade estatística é formada por células de 62.500m<sup>2</sup> e 1.000.000m<sup>2</sup>. A diferença entre os polígonos dos dois mapeamentos é muito grande, fazendo com que uma grande parte das áreas suscetíveis apresente uma área aproximadamente igual ou menor do que as menores células da grade. Na prática, isso faz com que poucas células estejam inteiramente dentro de apenas um polígono de suscetibilidade, sendo necessário utilizar métodos de seleção espacial (totalmente contido, com o centroide contido, com a maior parte contida) para a integração de dados. Todas essas técnicas acarretam incertezas, mas a principal fonte de incerteza não está no método utilizado, mas na diferença de escala entre os mapeamentos e que, conseqüentemente, gera incompatibilidades. A aproximação dos recortes espaciais também gera incertezas na totalização dos dados demográficos, levando a resultados imprecisos e que devem ser vistos como valores aproximados.

Outro aspecto a ser destacado é a natureza dinâmica dos estudos e/ou avaliações do meio físico, em que os dados e as informações devem ser constantemente atualizados e retrabalhados, visando a acompanhar as alterações ambientais e socioeconômicas que interferem no tema estudado.

Pode-se afirmar que mesmo com as incertezas presentes nos resultados a oportunidade de realizar análises que integram dados censitários a dados físicos é extremamente valiosa, enriquecendo os estudos no campo de População e Ambiente.

Figura 25 - Compatibilização de escalas.



Fonte: Instituto Geológico, 1996; IBGE, Censo Demográfico 2010.

### 3.3 Aplicação 3: Análise espacial

Para estudar a realidade interna dos municípios, é necessário ter dados em escala intraurbana, que é uma escala onde dificilmente existem dados secundários disponíveis quando se trata de cidades pequenas e médias. Apesar dos setores censitários serem os candidatos naturais à unidade geográfica desses estudos, a sua falta de estabilidade espaço-temporal prejudica a realização de um

acompanhamento das medidas ao longo do tempo. Além disso, algumas técnicas de análise espacial apresentam um resultado mais robusto quando os dados se apresentam agregados em unidades com forma e tamanho regular (ESRI, 2009). Isso torna a utilização de grades regulares associada a dados censitários uma alternativa analiticamente vantajosa. Neste estudo, essas duas unidades espaciais – setor censitário e grade estatística - foram utilizadas objetivando realizar comparações entre os resultados obtidos.

### 3.3.1 Introdução

A realidade desigual das cidades precisa ser compreendida para que sejam formuladas políticas públicas que possam modificá-la. No entanto, esta heterogeneidade é reconhecida apenas de maneira parcial, pois é simples identificar as áreas ricas e pobres de uma cidade, mas as diferenças das características da população, associadas à sua distribuição espacial são mais sutis e difíceis de serem percebidas (BUENO; MARANDOLA JR; D'ANTONA, 2012). Os esforços voltados para a compreensão do espaço urbano requerem a integração de dois elementos básicos (RAMOS, 2002): conceitos e teorias que caracterizam os fenômenos urbanos e métodos quantitativos de representação e análise desses fenômenos. Este caso de uso se dedica a empregar métodos quantitativos para analisar a formação de agrupamentos espaciais a partir de características demográficas da população.

### 3.3.2 Materiais e métodos

O município de Limeira, no interior do estado de São Paulo, foi a área selecionada para este estudo. Desde a década de 1970, até os dias de hoje este município passa por um processo de expansão da sua mancha urbana

(D'ANTONA *et al.*, 2012), fato que ressalta a importância da realização de estudos intraurbanos.

A análise de padrões espaciais faz uso da Primeira Lei da Geografia de Tobler (TOBLER, 1970), que anuncia que coisas espacialmente próximas estão mais relacionadas entre si do que coisas distantes. Esta lei se operacionaliza na propriedade da autocorrelação espacial, que avalia a tendência das variáveis associadas a uma unidade espacial de serem similares àsquelas associadas às unidades vizinhas. Inicialmente, utiliza-se o Índice Global de Moran para avaliar se a distribuição espacial das variáveis estudadas segue um padrão aleatório, agrupado ou disperso e também para avaliar a distância que deve ser empregada na próxima etapa da análise.

O Índice Local de Moran apresenta a localização espacial dos agrupamentos, uma vez que é importante a integração de metodologias exploratórias tradicionais com métodos visuais, como os mapas (HAINING; WISE, 1997). Valores positivos do Índice Local de Moran indicam que existe um grupo de unidades com valores similares ou *cluster*, enquanto valores negativos indicam que existe um grupo de unidades com valores discrepantes ou *outlier*.

A área de estudo selecionada é formada pelo conjunto de setores urbanos contíguos do tipo 1 (área urbanizada de cidades e vilas) (IBGE, 2003) e pelas células da grade estatística com dimensões de 230 x 230 m que interceptam esses setores. Apesar de existirem algumas pequenas diferenças, as duas áreas de estudo são equivalentes, como pode ser visto na Figura 26.

O limite máximo para a distância de análise foi estabelecido em 10 km, que é a dimensão do lado de um quadrado envolvendo a área de estudo (Figura 26). Esta distância também é o valor máximo para o qual é possível obter resultado para o Índice de Moran Global utilizando setores censitários, pois para distâncias maiores os valores das variáveis são praticamente iguais e a variância é nula, o que impede a geração de resultados.

Figura 26 - Delimitação das áreas de estudo para a análise do índice Global e Local de Moran, Limeira, 2010.



Fonte: IBGE, Censo Demográfico 2010 e Malha Municipal 2010.  
Elaborado pelo autor.

A análise foi realizada para variáveis censitárias que caracterizam a população na sua forma mais básica: densidade populacional, que expressa o volume da população, razão de sexos, que mostra a razão entre homens e mulheres, e razão entre faixas etárias, mostrando se existe uma concentração no volume de crianças (idade até 9 anos) ou de idosos (idade superior a 65 anos) em relação à população jovem e adulta (idade entre 10 e 64 anos). Outras variáveis que caracterizam a população e que estão relacionadas com questões de segregação também foram utilizadas, como raça/cor (quantidade de brancos na

população total), renda domiciliar (valor médio mensal domiciliar *per capita*) e alfabetização (razão entre não alfabetizados e alfabetizados).

Todas as análises espaciais foram executada no programa de geoprocessamento ArcMap, versão 10, com as ferramentas do grupo de Estatística Espacial.

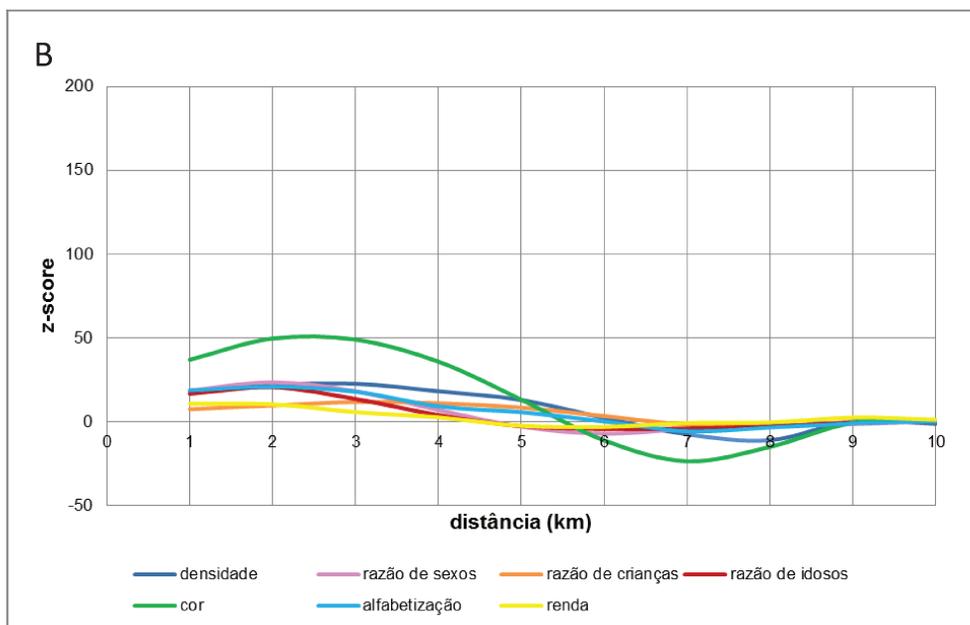
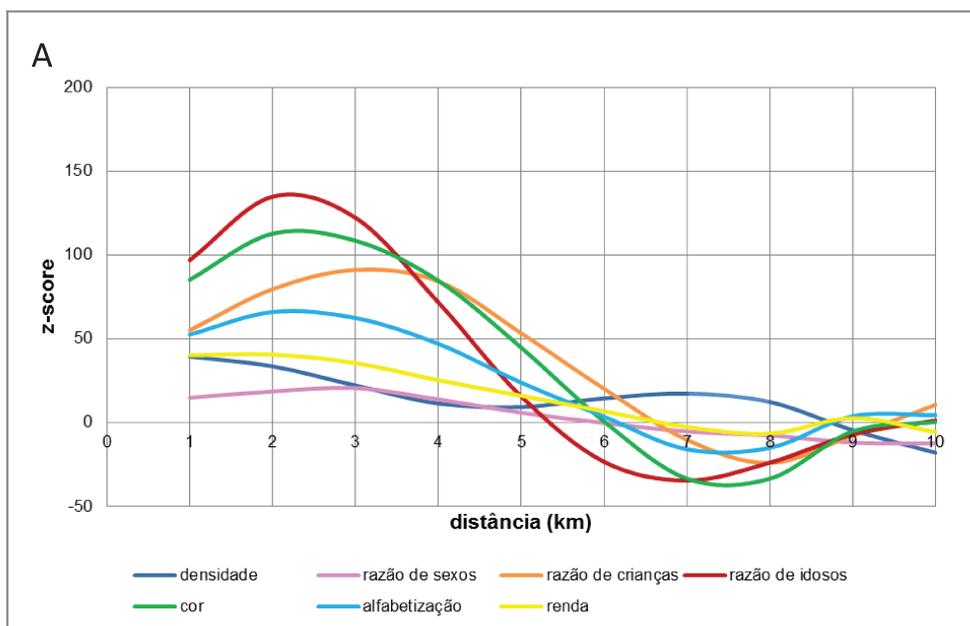
### 3.3.3 Resultados

Os resultados da análise do Índice Global de Moran para diferentes distâncias utilizando a grade estatística e os setores censitários como unidade espacial podem ser vistos no Gráfico 4. Valores positivos indicam que existe um padrão de agregação, enquanto valores negativos indicam um padrão de dispersão. Valores próximos de zero indicam a não existência de um padrão espacial de dados, significando uma distribuição aleatória de dados. Os dois gráficos mostram que com uma distância de 2.000 metros as variáveis cor, alfabetização e razão de idosos apresentaram autocorrelação espacial máxima ou agrupamento máximo, que é indicado por um pico no gráfico. A variável razão de crianças apresentou autocorrelação máxima com distância de 3.000 m tanto com a utilização da grade estatística quanto com setores censitários.

A conclusão a que se pode chegar analisando os dois gráficos é que o padrão espacial para essas variáveis é similar, variando a sua amplitude em virtude da diferença entre as dimensões das unidades de análises.

As variáveis renda, razão de sexos e densidade não apresentaram coincidência na distância onde ocorre autocorrelação máxima utilizando as duas unidades espaciais e, portanto, não é possível afirmar que exista um padrão de distribuição espacial. É provável que esta distribuição seja determinada pelo arranjo dos dados, ou seja, causado pelo MAUP (ver Cap. 1, item 1.2.3).

Gráfico 4 - Autocorrelação espacial (z-score) de acordo com a distância, Limeira, 2010.



Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

A – Grade estatística    B – Setor censitário

Com base nos resultados obtidos para o Índice Global de Moran e com base em conhecimentos locais da área de estudo (BUENO; D'ANTONA, 2012), adotou-se o valor de 2.000 m para a distância da banda para a análise do Índice Local de Moran. Os mapas resultantes desta análise são mostrados nas figuras 27 a 33. Do lado esquerdo da figura é mostrado o resultado quando foi utilizada a grade estatística como unidade espacial de análise e do lado direito, quando foi utilizado o setor censitário. Quando uma unidade tem um valor alto para a variável analisada e é circundada dentro do limite do valor de banda por unidades de valor também alto, o resultado é um agrupamento de valores altos ou *cluster High-High*. Quando acontece o contrário, ou seja, unidades de valores baixos são circundadas por unidades de valores também baixos, temos um agrupamento de valores baixos ou *cluster Low-Low*. Quando uma unidade apresenta um valor alto e é circundada por unidades de valor baixo, ou vice-versa, temos grupos de valores discrepantes ou *outliers (High-Low e Low-High)*. Finalmente, quando uma unidade apresenta valores médios e também é circundada por unidades de valores semelhantes, temos áreas onde não há um padrão espacial significativo.

A Figura 27 mostra uma concentração de brancos no coração da área urbana e uma concentração de não brancos nas áreas periféricas ao sul, leste e oeste. Pode-se dizer que o padrão espacial para esta variável é idêntico para as duas unidades de análise.

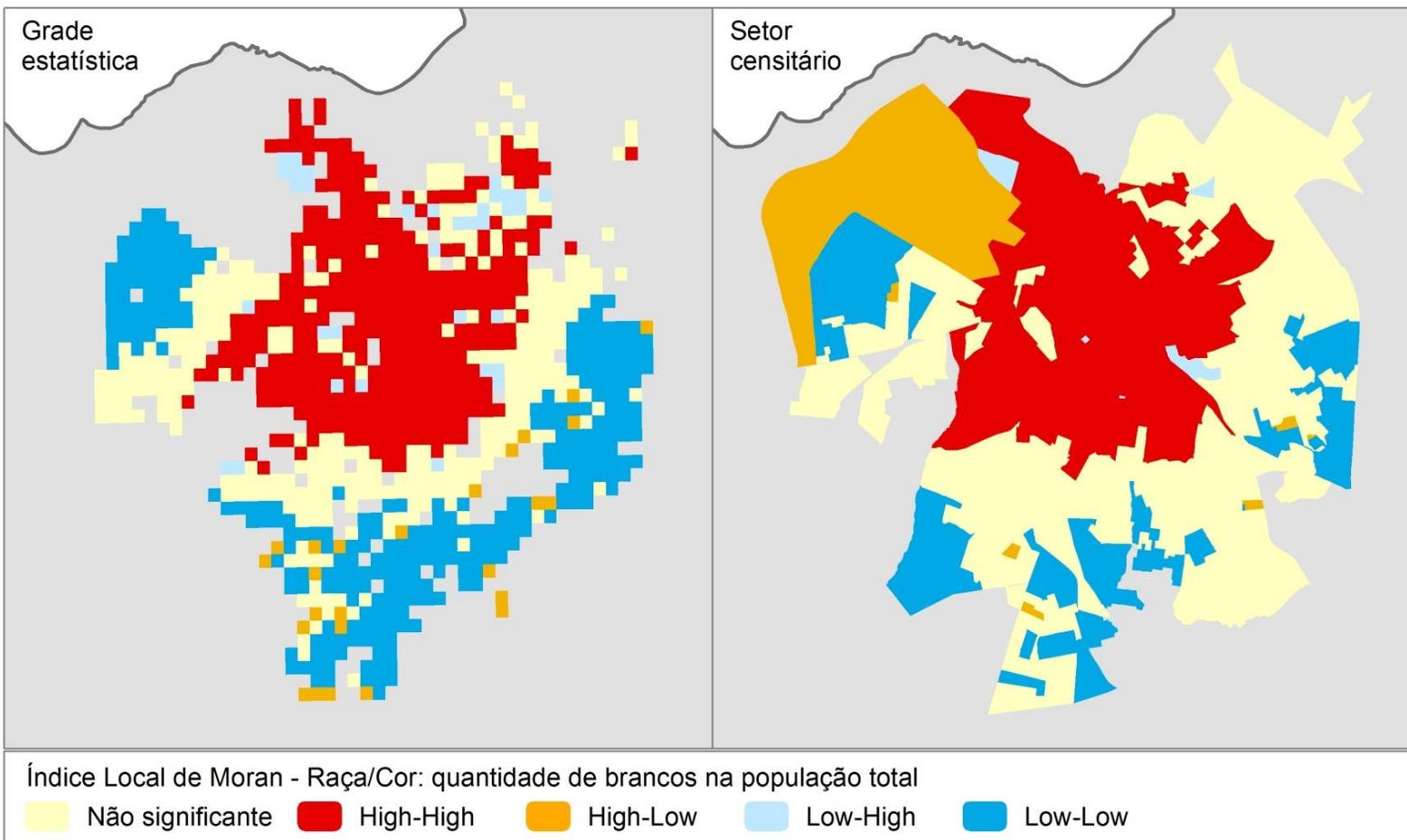
Na Figura 28 temos o mapeamento dos agrupamentos da variável alfabetização, mostrando a relação entre as quantidades de indivíduos não alfabetizados e indivíduos alfabetizados. Para esta variável, o padrão espacial também é o mesmo para as duas unidades estudadas, mostrando que os não alfabetizados se concentram na periferia da área urbana (centro-sul, centro-leste e norte-oeste) enquanto os alfabetizados se concentram na região urbana central.

Para as duas variáveis relacionadas com a estrutura etária mostradas nas Figuras 29 e 30, pode-se observar que o padrão espacial é similar, apresentando algumas coincidências, mas que não pode ser considerado idêntico. No caso da razão entre crianças e jovens + adultos, há uma baixa concentração

de crianças na área urbana central para as duas unidades espaciais. No entanto, as altas concentrações de crianças surgem na região sul-oeste nos dois casos e ao longo do arco sul-leste e na região noroeste apenas quando se utiliza a grade estatística. Isso sugere que não se têm garantias de que o padrão visualizado seja real e que o mesmo pode ser resultado da forma de disposição e agregação dos dados. Na Figura 30, podemos ver um resultado inverso ao da Figura 29, já que a variável em questão é a razão entre a quantidade de idosos e a quantidade de jovens + adultos. Um resultado coincidente com relação às duas unidades de análise estudadas ocorreu apenas na área central, a qual apresenta uma grande concentração de idosos. A baixa concentração de idosos no arco sul-oeste quando se utiliza a grade estatística pode ser resultado do arranjo dos dados e não de um padrão espacial determinado pela proximidade entre as unidades.

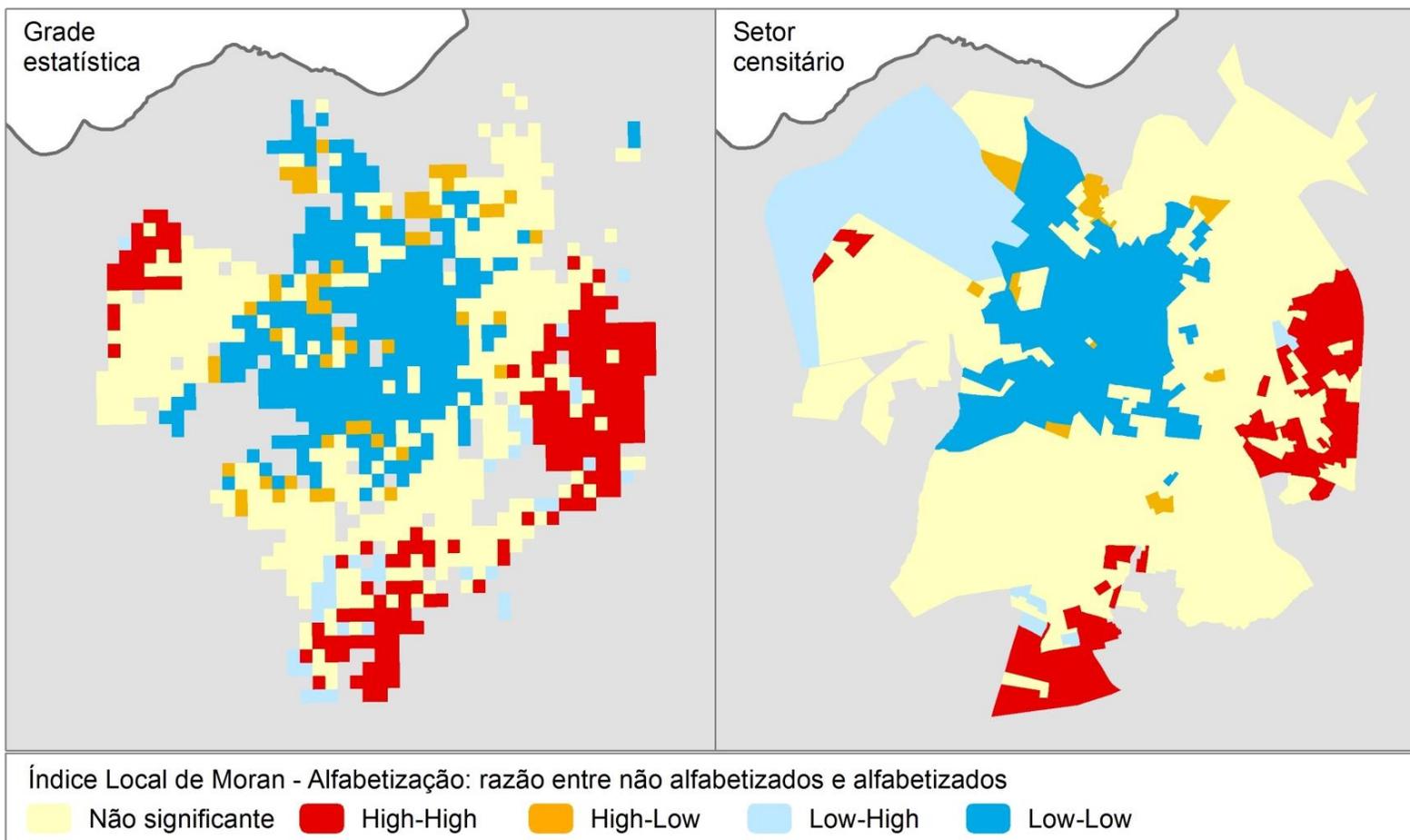
As Figuras 31 a 33 mostram resultados bastante distintos para as variáveis renda, razão de sexos e densidade quando se utiliza a grade estatística ou os setores censitários como unidade espacial. Esses resultados para o Índice Local de Moran são condizentes com aqueles obtidos para o índice Global de Moran apresentados no Gráfico 4 e indicam que não se pode assegurar a existência de um padrão espacial de distribuição para essas variáveis.

Figura 27 - Índice Local de Moran – Raça/Cor, Limeira, 2010.



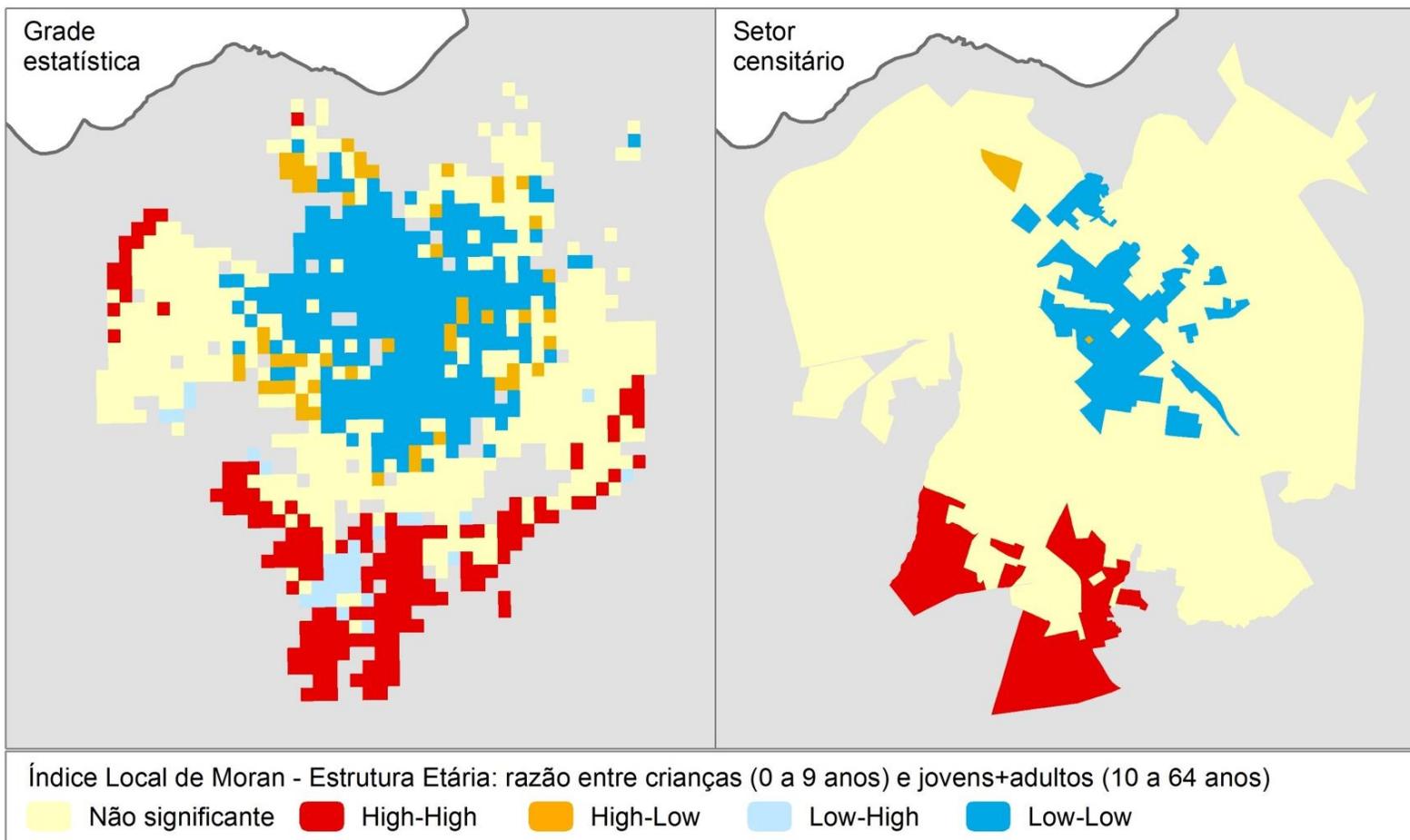
Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Figura 28 - Índice Local de Moran – Alfabetização, Limeira, 2010.



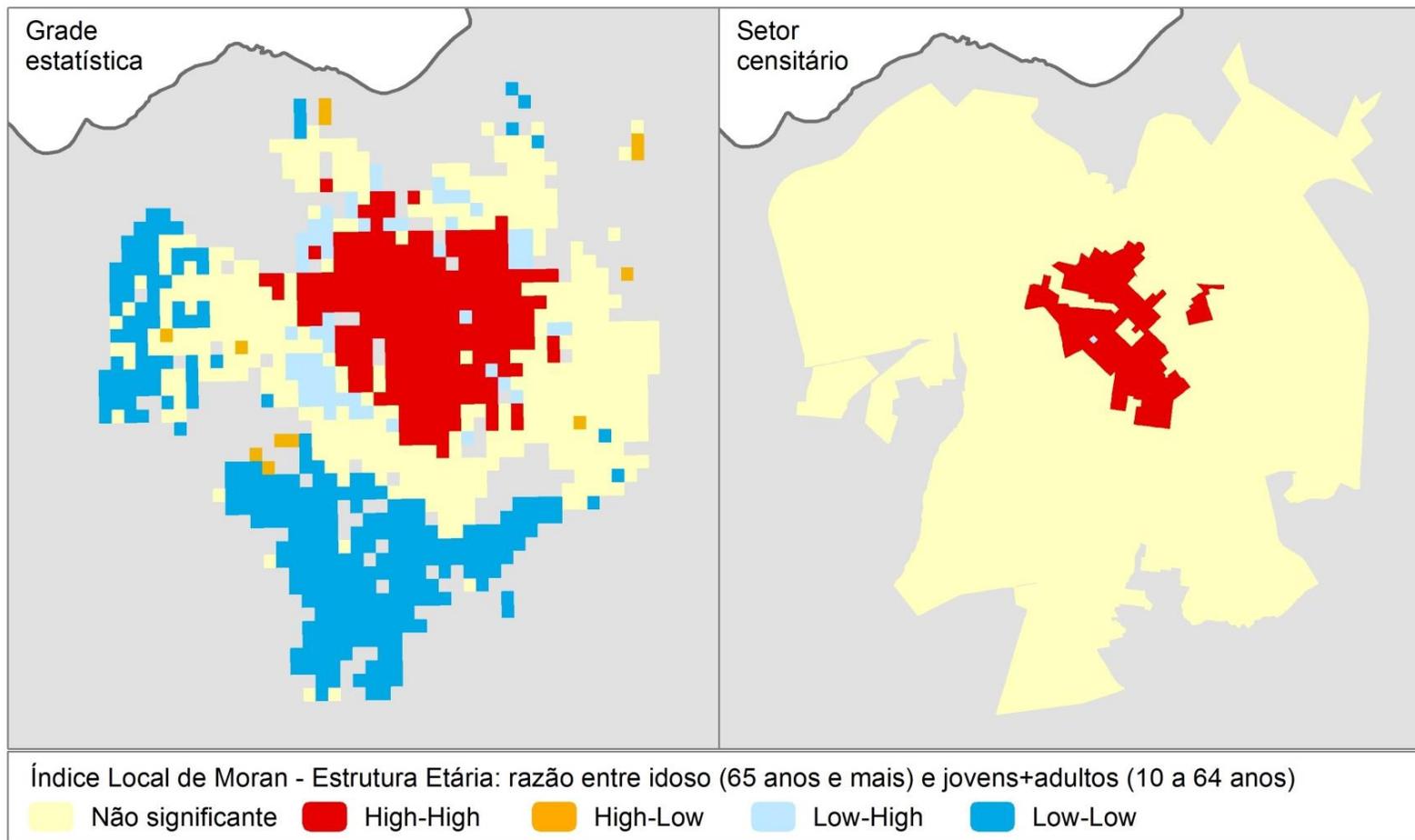
Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Figura 29 - Índice Local de Moran - Razão entre crianças e jovens + adultos, Limeira, 2010.



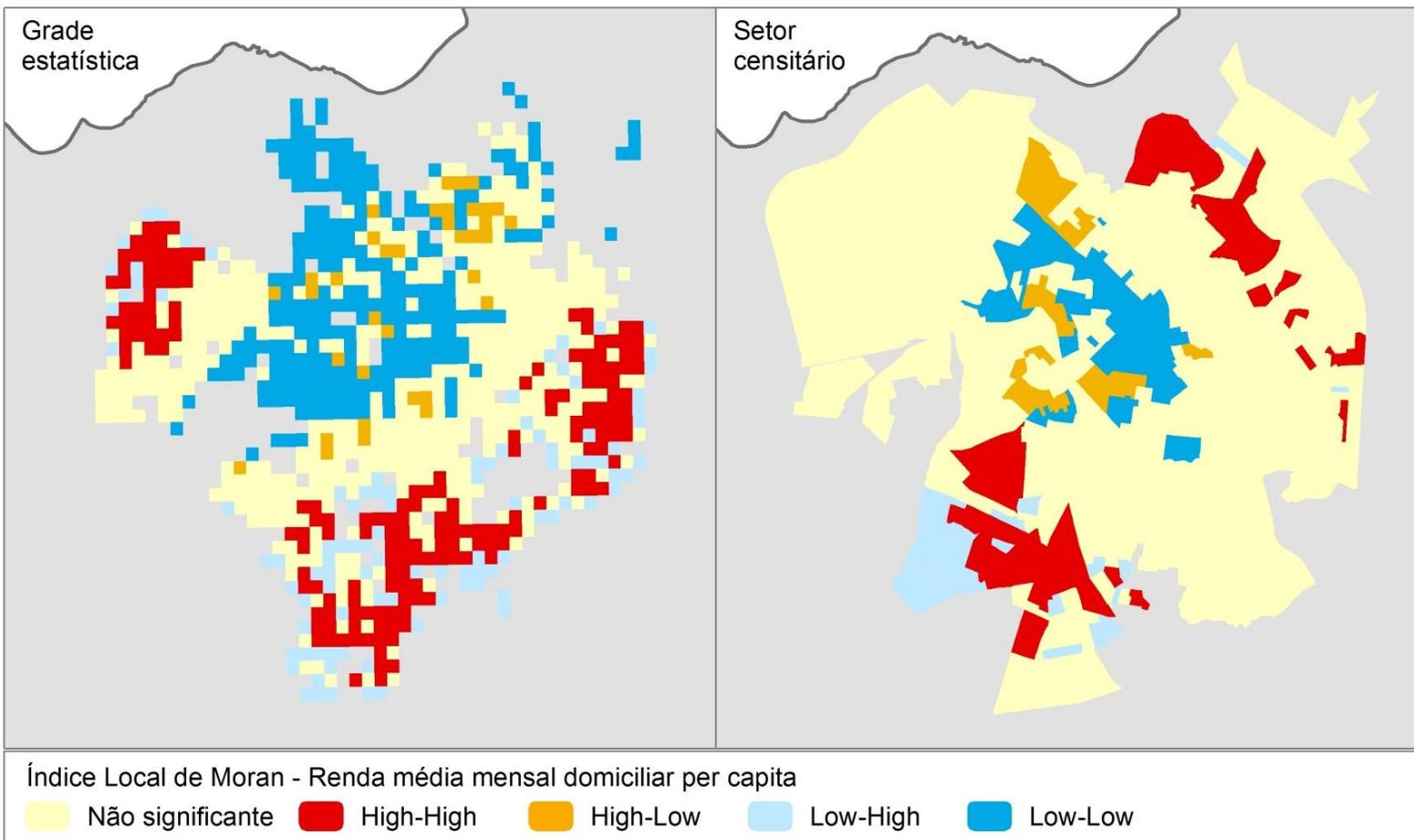
Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Figura 30 - Índice Local de Moran - Razão de idosos e jovens + adultos, Limeira, 2010.



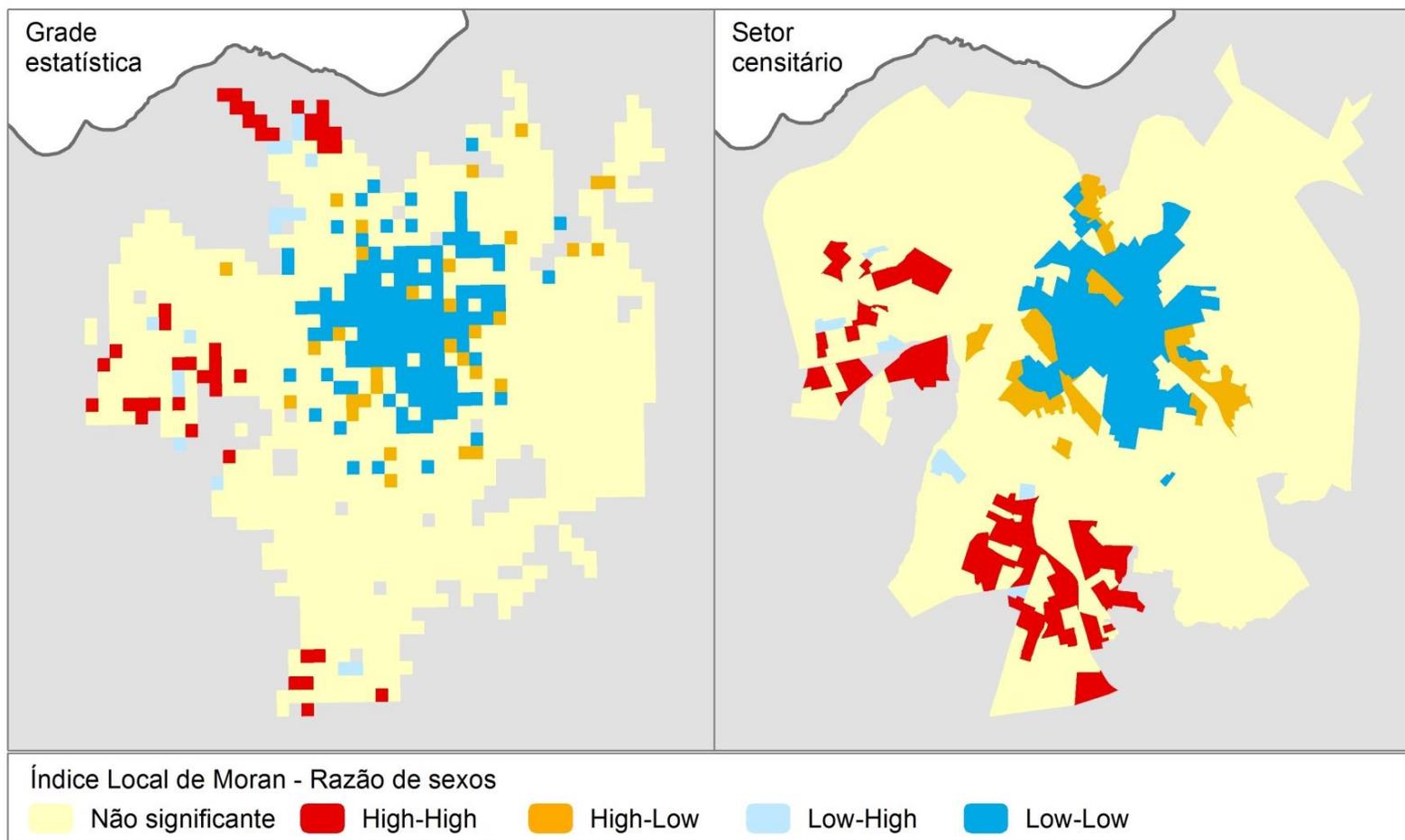
Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Figura 31 - Índice Local de Moran – Renda média mensal domiciliar *per capita*, Limeira, 2010.



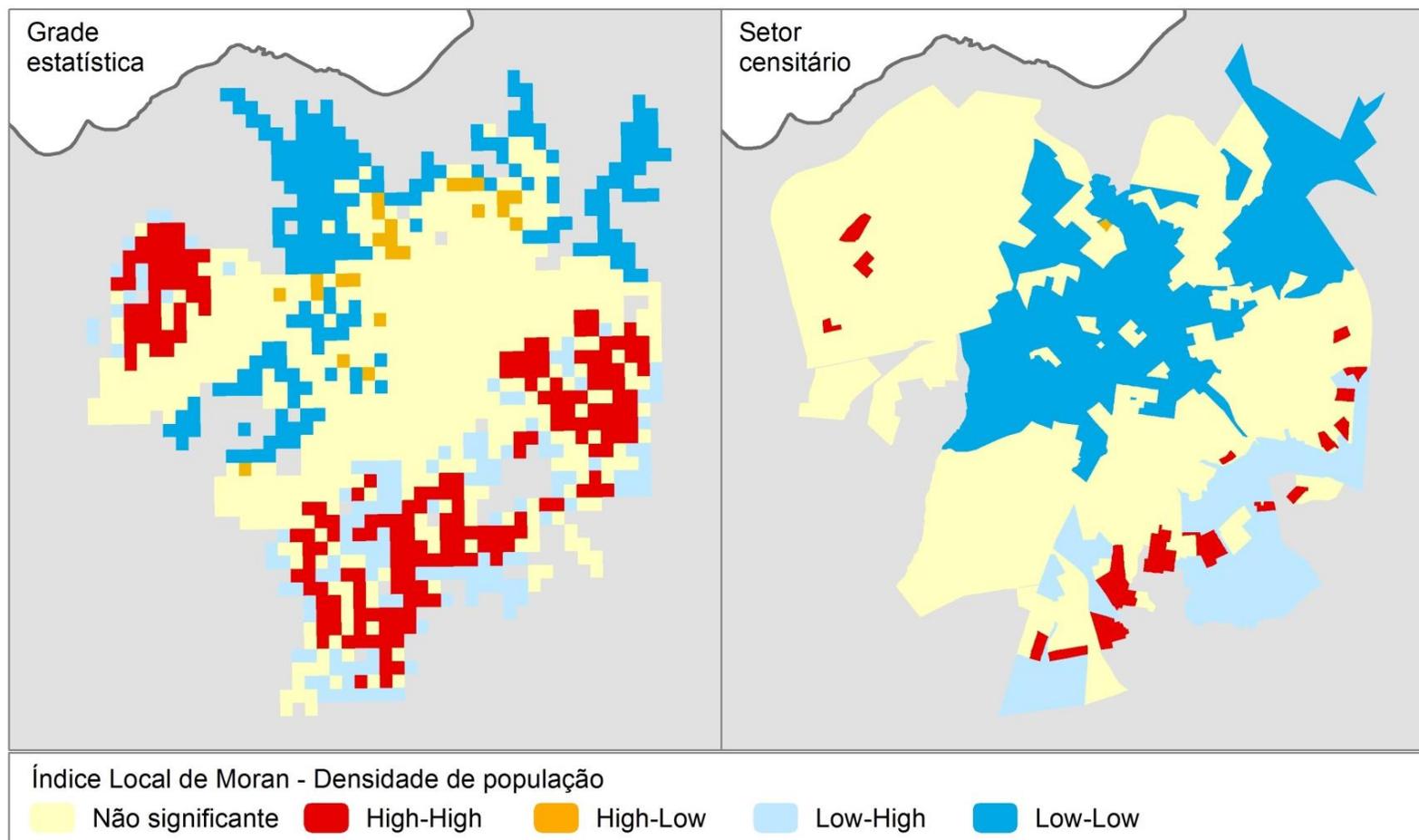
Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Figura 32 - Índice Local de Moran - Razão de sexos, Limeira, 2010.



Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Figura 33 - Índice Local de Moran – Densidade de população, Limeira, 2010.



Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

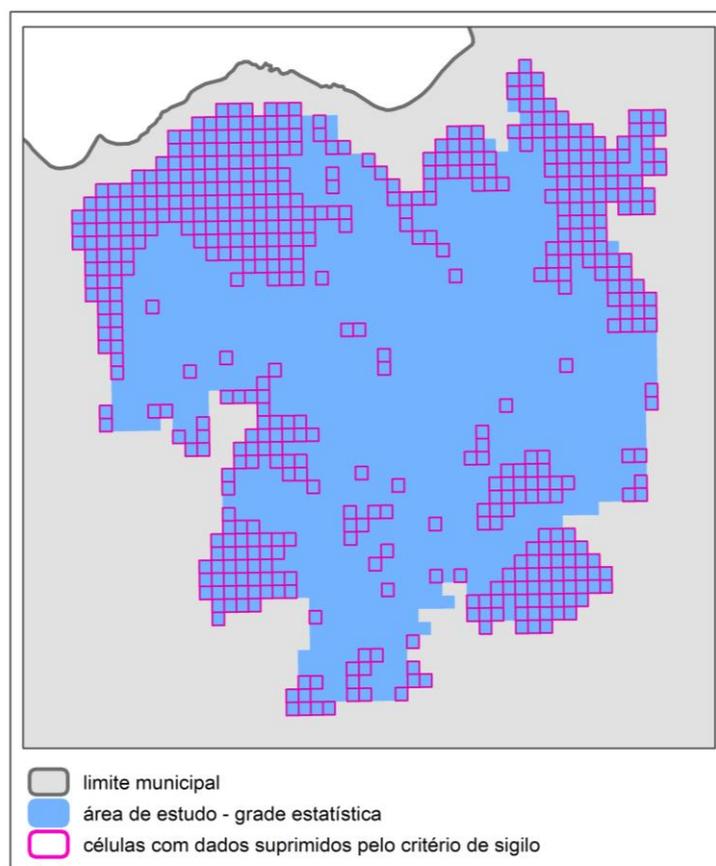
### 3.3.4 Discussão

A execução de análises espaciais utilizando unidades geográficas representando escalas diferentes é proposta na bibliografia como sendo uma metodologia para explorar mais profundamente os dados e assim avaliar se os resultados são decorrentes ou na da ocorrência de MAUP (ver Cap. 3, item 3.2.4). Nesta aplicação, analisando os resultados obtidos para os Índices Global e Local de Moran, pode-se afirmar que as variáveis raça/cor, alfabetização e idade apresentam um padrão de distribuição espacial similar para as duas unidades espaciais estudadas, indicando que este padrão não é determinado pelo arranjo espacial dos dados, mas pelos próprios dados. O mesmo não acontece com as variáveis renda, razão de sexos e densidade, onde, para cada unidade de análise, os valores máximos de autocorrelação ocorrem a distâncias diferentes e os agrupamentos espaciais apresentam padrões também diferentes. Esses resultados são uma indicação de que a distribuição espacial dos dados não está relacionada com algum fator subjacente relacionado com a distância, mas com a forma de agregação dos dados propriamente dita. Ou seja, alterando a forma e o tamanho da unidade de agregação (célula ou setor censitário), o padrão espacial de distribuição também se altera, já que o mesmo é resultado do arranjo na disposição dos dados. Desta maneira, pode-se concluir que a execução de análises espaciais utilizando dados agregados em unidades geográficas diferentes é uma forma de se garantir um resultado mais robusto. As variáveis que claramente não apresentam MAUP – raça/cor, alfabetização e idade – podem ter seus resultados explorados, buscando-se avaliar quais são os fenômenos subjacentes que determinam o padrão espacial encontrado. As demais variáveis - densidade, razão de sexos e renda - por apresentarem resultados diferentes para as unidades estudadas, não devem ter os resultados explorados, já que os mesmos aparentemente foram gerados pela forma de agregação.

A existência de uma grande quantidade de células sem dados relacionados com as características das pessoas e dos domicílios devido aos critérios

de sigilo estatístico adotado não compromete esse tipo de análise, uma vez que as células sem dados não são consideradas no processamento dos índices de Moran. Vale ressaltar que os dados omitidos não são representados matematicamente como sendo nulos ou iguais a zero, mas como dados não existentes (*NULL*). A Figura 34 mostra a região central urbana de Limeira coberta pela grade estatística com as células com dados suprimidos pelo critério de sigilo sendo destacadas. Comparando-se esta figura com os resultados apresentados nas Figuras 27 a 33 pode-se notar que as áreas em destaque aparecem sem resultados, indicando que essas células não foram consideradas no cálculo do Índice Local de Moran.

Figura 34 - Grade estatística: células com dados suprimidos pelo critério de sigilo, Limeira, 2010.



Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

A variável densidade demográfica, além de não ter assegurada a existência de um padrão espacial de distribuição, deve ter seu resultado analisado em separado devido à própria natureza da variável. Sendo uma densidade, o seu cálculo está diretamente relacionado com a área, e conseqüentemente, com a unidade espacial dos dados. Assim, por utilizar unidades diferentes, era de se esperar que os resultados fossem distintos. Os setores censitários se apresentam com forma e dimensões diversas, enquanto as células da grade estatística são regulares e uniformes. Além disso, existe uma grande diferença entre as áreas dessas duas unidades, sendo as células, de uma maneira geral, menores do que os setores, principalmente nas áreas periféricas ao núcleo urbano central. Esta diferença na dimensão nos leva a um resultado mais detalhado com a utilização da grade, permitindo a observação de variações no valor da densidade que são obscurecidos pela dimensão dos setores censitários. Além dessas variações, a utilização da grade estatística permite a observação de espaços sem população permeados nos espaços povoados, enriquecendo sobremaneira as análises.

## CONSIDERAÇÕES FINAIS

*A tarefa não é tanto ver aquilo que ninguém viu,  
mas pensar o que ninguém ainda pensou  
acerca daquilo que todo mundo vê.  
(Arthur Schopenhauer)*

No primeiro capítulo foram mostrados os desafios que dificultam a incorporação efetiva do espaço aos estudos e análises no campo das Ciências Sociais e mais especificamente da Demografia. Espera-se que com a utilização da grade estatística haja um rompimento da estrutura formal aninhada da geografia dos censos no que diz respeito à disseminação e utilização de dados. Isso possibilitará que os usuários façam uso de unidades sem um significado previamente embutido e, ao mesmo tempo, mais bem ajustadas aos seus objetos de estudo, o que provavelmente levará ao afloramento dos fenômenos que ali ocorrem. Então, possivelmente as conclusões advindas destas análises serão um passo em direção ao estabelecimento dos relacionamentos entre as dinâmicas social e demográfica e a dinâmica geoespacial.

No que diz respeito à geração da grade estatística, considera-se que a tarefa foi bem sucedida e que o produto final apresenta qualidade comparável às unidades de agregação tradicionais de disseminação de dados censitários (setores censitários e unidades político-administrativas). Embora o método de agregação de dados tenha a sua utilização restrita aos institutos nacionais de estatística por envolver a manipulação dos microdados não desidentificados dos censos, o que é permitido apenas aos produtores dos dados, ele se configura como sendo o método mais preciso. No entanto, o emprego exclusivo deste método em todo o território brasileiro não se mostrou viável devido à falta de atributos de localização geográfica em uma parcela dos dados. Para viabilizar a geração da grade estatística foi necessário empregar uma metodologia híbrida: parte agregação e parte desagregação. O uso desta metodologia híbrida acarreta uma incerteza maior aos resultados, mas não prejudica a sua utilização, desde

que o usuário dos dados tenha ciência das limitações existentes. A inclusão de uma variável que indica o método utilizado para agregar os dados em cada célula (agregação, desagregação ou misto) foi a maneira encontrada para explicitar essas incertezas e limitações.

Diante da perspectiva de um contínuo aprimoramento das metodologias de coleta de dados de pesquisas domiciliares, sobretudo através da incorporação de geotecnologias cada vez mais avançadas, o desenvolvimento de um trabalho constante e sistemático de atualização e complementação da base de dados de endereços e de vias, através, por exemplo, do estabelecimento de parcerias com governos locais e concessionárias de serviços, poderia levar a uma utilização cada vez menor das técnicas de desagregação de dados e, conseqüentemente, a obtenção de uma grade com dados quase que exclusivamente gerados por agregação, ou seja, com um grau de certeza maior.

Sabe-se que a qualidade dos resultados obtidos está diretamente relacionada com a qualidade dos dados utilizados. Assim, as incertezas existentes na base de dados censitários também estarão presentes na grade estatística e a qualidade dos dados auxiliares é determinante para a qualidade dos resultados obtidos através de métodos dasimétricos, uma vez que eles servem como indicadores de presença de população. A escala dos dados auxiliares, a metodologia utilizada para sua obtenção e o objetivo do levantamento, principalmente no caso de classificação de uso/cobertura das terras, influenciam sobremaneira o resultado, sendo que quanto mais detalhados e precisos forem os dados auxiliares, melhor será o resultado obtido com as técnicas dasimétricas. No entanto, é preciso levar em consideração a disponibilidade desses dados. Por exemplo, dados tridimensionais obtidos com sensores de *laser* podem gerar classificações de uso/cobertura com características volumétricas, sendo, portanto, mais precisos do que classificações obtidas utilizando imagens tradicionais de sensoriamento remoto. Por outro lado, as oportunidades de aquisição das imagens de satélite são maiores do que as de imagem de radar, seja devido aos custos envolvidos ou à própria disponibilidade dos dados. Além disso, o

processamento das imagens de *laser* requer recursos técnicos e humanos mais especializados do que o processamento de imagens de satélite. Desta maneira, a precisão advinda da utilização de dados e técnicas muito especializadas nem sempre é condizente com a necessidade de gerar dados para áreas extensas e com aplicação de poucos recursos, havendo a necessidade de se buscar o equilíbrio entre disponibilidade e qualidade. Isso justifica o emprego de técnicas simples e dados menos precisos na abordagem de desagregação de dados utilizando classificações de uso das terras como dado auxiliar, tendo a vista o objetivo futuro de geração de uma grade estatística abrangendo todo o território brasileiro.

Com relação aos benefícios obtidos com a utilização da grade estatística, devemos retornar às questões anunciadas na introdução desta tese para verificar se elas foram atendidas. A primeira questão está relacionada com a adequação da grade a unidades que não são coincidentes com unidades político-administrativas. Observou-se nos resultados obtidos nas aplicações realizadas neste estudo que a grade se adequa a quaisquer recortes geográficos, pois, por apresentar pequenas dimensões, as células da grade estatística podem ser vistas como blocos capazes de compor superfícies irregulares. A alternativa de utilizar os setores censitários, por exemplo, leva a resultados imprecisos, enquanto a opção de realizar levantamentos para obter dados primários requer investimentos que nem sempre tornam esta alternativa viável. A adequação da grade estatística a um recorte territorial qualquer pode ser vista como uma possibilidade de superação de uma das dificuldades presentes nos estudos de população e mais especificamente no campo de População e Ambiente e que está relacionada com a integração entre a dimensão física e ambiental e a dimensão humana. A aderência entre a grade estatística e uma área irregular qualquer não é perfeita, e, dependendo da escala, pode ser bastante imprecisa. No entanto, isso não invalida a sua utilização e permite obter dados para a realização de análises de forma rápida e simples. Além disso, não se pode esquecer que caso a grade estatística seja adotada pelo IBGE como um produto regular de disseminação de dados censitários, a sua

frequência será a mesma dos demais produtos e a sua qualidade será reconhecida.

A segunda questão faz referência à comparabilidade temporal das unidades de disseminação de dados censitários. Uma das principais finalidades de se adotar uma grade regular como unidade de disseminação de dados é a sua não alteração ao longo do tempo, ou seja, uma vez determinadas as características geográficas desta estrutura, ela deve se manter fixa e inalterada. Da maneira como a grade estatística foi estabelecida nesta tese, com células de aproximadamente 1 km nas áreas rurais e 230 m nas áreas urbanas, a estrutura se mantém fixa apenas para a maior dimensão. Como o Brasil ainda apresenta uma grande dinâmica territorial, à medida que o tempo passa, surgem novas áreas urbanas, e as áreas existentes se expandem. Isso impede que a estrutura espacial da grade se mantenha fixa, considerando-se as células de 230 m, pois nos locais onde surgiram essas áreas urbanas é necessário que seja feita a divisão das células de 1 km em células menores. Já no caso das células de 1 km, a estrutura se mantém fixa ao longo do tempo, pois as células menores podem ser agregadas de modo a manter a estrutura inalterada. A geração de uma estrutura com células de dimensões diferentes traz ganhos analíticos, discretizando os dados em áreas mais densas e minimizando a quantidade de dados omitidos devido ao sigilo estatístico, embora a questão da não alteração da estrutura espacial seja parcialmente comprometida. Infelizmente, não foi possível mostrar aplicações explorando a comparabilidade temporal da grade estatística por falta de dados, uma vez que a metodologia adotada pressupõe a existência de microdados georreferenciados ou com possibilidade de georreferenciamento, o que é possível apenas a partir do Censo Demográfico 2010. Mas, mesmo sem ter uma demonstração prática, é possível vislumbrar que a estabilidade das unidades espaciais facilita a realização de análises comparativas ao longo do tempo, e, em alguns casos, pode-se dizer que até mesmo permite que isso seja efetivamente possível.

A terceira questão está relacionada com a integração de dados que utilizam unidades espaciais diferentes e suas conclusões são similares às aquelas apresentadas para a primeira questão, uma vez que a componente principal é a mesma, qual seja a adaptação da grade a um recorte espacial diferente daqueles tradicionalmente utilizados para disseminar os dados censitários. Por ter uma estrutura espacial composta por unidades de pequenas dimensões, a grade estatística se molda, embora não perfeitamente, a qualquer recorte geográfico, inclusive aqueles que apresentam perímetro irregular. Esta adaptação é tão melhor quanto mais parecidas forem as escalas dos dados ou quando a escala da grade é mais detalhada do que a escala da camada de dados a ser integrada. Quando a escala da camada de dados que se deseja integrar aos dados censitários for mais detalhada do que a grade estatística, a adaptação cartográfica é bastante imprecisa, gerando resultados com um grande grau de incerteza. Quando a camada a ser integrada é derivada de imagens de satélite, como é o caso de informações físicas e ambientais, a adequação cartográfica fica ainda mais fácil de ser feita. Isso se deve ao fato de que tanto as unidades básicas das imagens - os pixels - quanto as células da grade apresentam um formato regular que facilita a integração dos dados. Mesmo nos casos em que as dimensões dos pixels e das células são diferentes, a adequação é mais simples devido à possibilidade de união ou divisão dessas unidades em outras maiores ou menores.

Levando em consideração essas três questões, pode-se concluir que a disseminação de dados censitários através de uma grade estatística oferece um ganho significativo de potencial analítico, seja com relação à comparabilidade espacial e/ou temporal.

Além dos aspectos discutidos acima, podem ser acrescentadas algumas outras vantagens que ajudam a corroborar os benefícios da grade estatística, como por exemplo, a sua utilização em modelos de simulação. Diversas técnicas de modelagem, como os Modelos Baseados em Agentes e os Modelos de Autômatos Celulares, têm sido bastante empregadas atualmente em

simulações de fenômenos urbanos e sociais e a estrutura espacial em grade facilita sobremaneira a utilização de dados sociodemográficos censitários como parte integrante do conjunto de dados de entrada para a execução dessas análises. Os dados em grade também têm sido utilizados para a projeção de população espacialmente explícita (JONES, 2014), que, por sua vez são utilizados para a simulação de cenários futuros de uma variedade de fenômenos que estão relacionados direta ou indiretamente com o volume populacional. Como não existem dados disponibilizados e publicados oficialmente pelo IBGE sob a forma de grade, os pesquisadores que necessitam utilizar dados neste formato específico acabam tendo que gerar esses insumos, muito embora esse não seja o objetivo principal das pesquisas. No caso dessa forma de disseminação ser adotada pelo instituto produtor dos censos, isso não será mais necessário, o que dará uma maior confiabilidade aos resultados e ampliará as possibilidades de aplicação e/ou replicação dessas análises, já que os dados serão igualmente gerados para todo o país.

Um ponto que não pode deixar de ser mencionado é o fato de que as duas técnicas empregadas nesta tese - agregação e desagregação - permitem a identificação de vazios de população, embora a primeira apresente um nível de certeza maior do que a segunda. Em um país de dimensões continentais como o Brasil e que apresenta grandes e variados processos de transformação e ocupação do território, tão importante quanto saber onde a população se localiza é saber onde ela não se localiza. Isso traz a possibilidade de se estabelecer um planejamento visando uma ocupação sustentável para as áreas não ocupadas ou recém-ocupadas antes que se dê uma ocupação caótica e desordenada.

A questão do sigilo estatístico pode ser apontada como um desafio enfrentado na geração da grade estatística. Este ponto é bastante sensível para todos os produtores de dados e devido ao fato de não existir um padrão internacional que sirva como referência e não haver uma metodologia que permita avaliar o quanto se perde em potencial analítico ao se aplicar esta ou aquela regra de sigilo, este assunto fica definido muito mais pelo bom-senso do que pela

técnica propriamente dita. O que se pode observar nos casos avaliados nesta tese para os estados de São Paulo e Pará foi que a quantidade de células com dados omitidos é grande, embora o volume de população residente nestas células seja pequeno diante do volume total de população recenseada. Apesar de ser esperado que a quantidade de dados omitidos na grade estatística fosse superior à omitida no caso dos setores censitários, devido principalmente às diferenças nas dimensões dessas unidades, a grande diferença encontrada nos leva a crer que deva ser buscado um critério que minimize a quantidade de dados omitidos e/ou uma metodologia que assegure o sigilo ao mesmo tempo em que cause pouco ou nenhum impacto no resultado final das análises efetuadas com esses dados.

Finalmente, as recomendações para trabalhos futuros que deem continuidade ao trabalho iniciado nesta tese estão relacionadas com a automatização dos procedimentos adotados de forma a aperfeiçoar e agilizar a geração da grade, o aprimoramento da estimativa das variáveis sociodemográficas quando são utilizados métodos de desagregação, a investigação de técnicas que permitam a utilização dos microdados na abordagem de desagregação e a investigação de métodos para a associação de dados de pesquisas amostrais a uma grade estatística.



## REFERÊNCIAS

AFRIPOP PROJECT. **AfriPop**. 2009. Disponível em: <<http://www.afripop.org>>. Acesso em: 15 jun. 2013.

AMERIPPOP PROJECT. **AmeriPop**. 2012. Disponível em: <<http://www.ameripop.org/>>. Acesso em: 15 jun. 2013.

ANSELIN, Luc; FLORAX, Raymond; REY, Sergio J. **Advances in Spatial Econometrics: Methodology, Tools and Applications**. Berlin: Springer-Verlag, 2004.

ARIAS, Santa. Rethinking space: an outsider's view of the spatial turn. **GeoJournal**, vol. 75, n.1, p. 29-41, 2010.

ASIAPOP PROJECT. **AsiaPop**. 2011. Disponível em: <<http://www.asiapop.org/>>. Acesso em: 15 jun. 2013.

BACKER, Lars. **The Geostat project: part of an international effort to build an ISS?** In: EUROPEAN FORUM FOR GEOSTATISTICS WORKSHOP, Bled, Slovenia, 2008.

BALK, Deborah; YETMAN, Gregory. **The Global Distribution of Population: Evaluating the Gains in Resolution Refinement**. Palisades, NY: CIESIN, Columbia University, 2004.

BALK, Deborah; POZZI, Francesca; YETMAN, Gregory; DEICHMANN, Uwe; NELSON, Andy. **The distribution of people and the dimension of place: Methodologies to improve the global estimation of urban extents**. Palisades, NY: CIESIN, Columbia University, 2004a.

BALK, Deborah; PULLUM, Thomas; STOREYGARD, Adam; GREENWELL, Fern; NEUMAN, Melissa. Spatial analysis of childhood mortality in West Africa. **Population, Space and Place**, vol. 10, n. 3, p. 175-216, 2004b.

BHADURI, Budhendra; BRIGHT, Edward; COLEMAN, Phillip; URBAN, Marie L. LandScan USA: A High Resolution Geospatial and Temporal Modeling Approach for Population Distribution and Dynamics. **GeoJournal**, vol. 69, p. 103-117, 2007.

BARCELLOS, Christovam; MACHADO, Jorge M. Huet. A organização espacial condiciona as relações entre ambiente e saúde: o exemplo da exposição ao mercúrio em uma fábrica de lâmpadas fluorescentes. **Ciência & Saúde Coletiva**, vol. 3, n. 2, p.103-113, 1998.

BRASIL . **Lei nº 9.985, de 18 de julho de 2000.** Institui o Sistema Nacional de Unidades de Conservação da Natureza do Brasil (SNUC). Brasília, 2000.

BRACKEN, Ian; MARTIN, David. The generation of spatial population distributions from census centroid data. **Environment and Planning A**, vol. 21, n. 4, p. 537-543, 1989.

BRINEGAR, Sarah J.; POPICK, Stephen J. A comparative analysis of small area population estimation methods. **Cartography and Geographic Information Science**, vol. 37, n. 4, p. 273-84, 2010.

BROLLO, Maria José; TOMINAGA, Lídia K. **Desastres Naturais e Riscos Geológicos no Estado de São Paulo: Cenário de Referência – 2012.** São Paulo: Coordenadoria Estadual de Defesa Civil, 2012.

BROLLO, Maria José; FERREIRA, Cláudio J.; TOMINAGA, Lídia K.; VEDOVELLO, R.; FERNANDES DA SILVA, Paulo César; ANDRADE, Eduardo; GUEDES, Antonio Carlos M. **Situação dos desastres e riscos no estado de São Paulo e instrumentos de gerenciamento.** In: CONGRESSO BRASILEIRO DE GEOLOGIA DE ENGENHARIA E AMBIENTAL, São Paulo, 2011.

BUENO, Maria do Carmo D.; PACHECO, Carlos Thadeu; PEREIRA, Clodomir; LIMA, Leonilde A.; LEITE, Leandro A.; MATTOS, Mário Henrique M.; MORAES, Raphael S. **Tecnologias geoespaciais e censos - a experiência de Cabo Verde.** In: SIMPÓSIO DE GEOTECNOLOGIAS DO PANTANAL, II, Corumbá, MS, p. 71-79, 2009.

\_\_\_\_\_; DAGNINO, Ricardo. População em Unidades de Conservação da Amazônia Legal: estimativas a partir da Contagem Populacional 2007. In: D'ANTONA, Alvaro O.; CARMO, Roberto L.(Org.). **Dinâmicas demográficas e ambiente.** Campinas: NEPO/UNICAMP, 2011.

\_\_\_\_\_; D'ANTONA, Álvaro O. **Utilização de grades regulares para análises espaciais intramunicipais de variáveis demográficas – Testes para Limeira – SP, 2010.** In: ENCONTRO NACIONAL DE ESTUDOS POPULACIONAIS, XVIII, ABEP, Águas de Lindóia, SP, 2012.

\_\_\_\_\_; MARANDOLA JUNIOR, Eduardo J.; D'ANTONA, Álvaro O. Desigualdades intraurbanas em Limeira: implicações para o planejamento e para as políticas públicas. In: BAENIGER, Rosana; PERES, Roberta G.; D'ANTONA, Álvaro O.; ETULAIN, Carlos R. (Org.). **Por dentro do Estado de São Paulo**, vol. 5, Região de Limeira. Campinas: NEPO/UNICAMP, 2012.

BULMER, Martin. **The Chicago School of Sociology: Institutionalization, Diversity and the Rise of Sociological Research.** Chicago: University of Chicago Press, 1984.

CÂMARA, Gilberto; DAVIS, Clodoveu; MONTEIRO, Miguel V. **Introdução à Ciência da Geoinformação**. São José dos Campos: DPI/INPE, 2001.

\_\_\_\_\_. Representação computacional de dados geográficos. In CASANOVA, Marco; CÂMARA, Gilberto; DAVIS, Clodoveu; QUEIROZ, Gilberto R. (Ed.). **Bancos de Dados Geográficos**. Curitiba: MundoGEO, 2005. Disponível em: <<http://www.dpi.inpe.br/livros/bdados/>>. Acesso em: 10 out. 2013.

CARVALHO, Renata Marzzano de; NASCIMENTO, Luiz Fernando Costa. Spatial distribution of dengue in the city of Cruzeiro, São Paulo State, Brazil: use of geoprocessing tools. **Revista do Instituto de Medicina Tropical de São Paulo**, São Paulo, vol. 54, n. 5, 2012.

CARVALHO, Marília Sá; SANTOS, Reinaldo S. Análise de dados espaciais em saúde pública: métodos, problemas, perspectivas. **Cadernos de Saúde Pública**, vol. 21, n. 2, p. 361-378, 2005.

\_\_\_\_\_; CÂMARA, Gilberto; CRUZ, Oswaldo G.; CORREA, Virginia. Análise Espacial de Áreas. In: DRUCK, Suzana; CARVALHO, Marília Sá; CAMARA, Gilberto; MONTEIRO, Antonio Miguel V. (Ed.). **Análise Espacial de Dados Geográficos**. Brasília: EMBRAPA, 2004.

CASTRO, Marcia C. Spatial Demography: An Opportunity to Improve Policy Making at Diverse Decision Levels. **Population Research and Policy Review**, vol. 26, p. 477-509, 2007.

CENTER FOR INTERNATIONAL EARTH SCIENCE INFORMATION NETWORK - CIESIN; COLUMBIA UNIVERSITY; CENTRO INTERNACIONAL DE AGRICULTURA TROPICAL – CIAT. **GPW and GRUMP: A Brief Background, Comparison, and History**. [s.d.] Disponível em: <<http://sedac.ciesin.columbia.edu/data/collection/gpw-v3/about-us>>. Acesso em: 6 out. 2013.

CLARKE, John Innes; RHIND, David W. **Human Dimensions of Global Environmental Change**. Programme Report 3. International Social Science Council e UNESCO. 1992.

CHRISTOFOLETTI, Antonio. As características da nova geografia. **Geografia**, vol. 1, n.1, 1976.

COCKINGS, Samantha; MARTIN, David. Zone design for environment and health studies using pre-aggregated data. **Social Science & Medicine**, vol. 60, p. 2729-2742, 2005.

CUNHA, José Marcos P. da; JAKOB, Alberto A. E.; HOGAN, Daniel J. e CARMO, Roberto L. do. A vulnerabilidade social no contexto metropolitano: o caso de Campinas. In: CUNHA, José M. P. da. (Org.) **Novas metrópoles paulistas: população, vulnerabilidade e segregação**. Campinas: NEPO/UNICAMP, 2006. p. 143-168.

D'ANTONA, Álvaro O.; MELLO, Allan Yu I. ; BUENO, Maria do Carmo D.; RIZARDI, Luis; MARCONDES, Talita. A expansão urbana de Limeira-SP entre 1970 e 2010. In: BAENIGER, Rosana; PERES, Roberta G.; D'ANTONA, Álvaro O.; ETULAIN, Carlos R. (Org.) **Por dentro do Estado de São Paulo**, vol. 5, Região de Limeira. Campinas: NEPO/UNICAMP, 2012. p. 73-86.

\_\_\_\_\_; BUENO, Maria do Carmo D.; DAGNINO, Ricardo S. Estimativa de população em unidade de conservação na Amazônia Legal Brasileira – uma aplicação de grades regulares a partir da Contagem 2007. **Revista Brasileira de Estudos de População**, vol. 30, n. 2, p. 401-428, 2003.

DAHMS, Harry. Rethorizing global space in sociology – Towards a new kind of discipline. In: WARF, Barney; ARIAS, Santa (Ed.). **The Spatial Turn: Interdisciplinary Perspectives**. Londres e Nova York: Routledge, 2009.

DARK, Shawna J.; BRAM, Danielle. The modifiable areal unit problem (MAUP) in physical geography. **Progress in Physical Geography**, vol. 31, n.5, p. 471-479, 2007.

de SHERBININ, Alex; BALK, Deborah; YAGER, Karina; JAITEH, Malanding; POZZI, Francesca; GIRI Chandra; WANNBO, Antoinette. **A CIESIN thematic guide to social science applications of remote sensing**. Nova York: Center for International Earth Science Information Network (CIESIN), 2002.

DEICHMANN, Uwe. **A Review of Spatial Population Database Design and Modelling**. Technical Report TR-96-3. National Center for Geographic Information and Analysis (NCGIA), University of California, Santa Barbara, USA, 1996.

DOBSON, Jerome E.; BRIGHT, Edward A.; COLEMAN, Philip R.; DURFEE, Richard C.; WORLEY, Brian A. LandScan: A global population database for estimating populations at risk. **Photogrammetric Engineering & Remote Sensing**, vol. 66, n. 7, p. 849-857, 2000.

DONNAY, Jean-Paul; UNWIN, David J. Modelling geographical distributions in urban areas. In DONNAY, Jean-Paul; BARNESLEY, Mike J.; LONGLEY, Paul A. (Ed.) **Remote sensing and urban analysis**. London: Taylor & Francis, 2001. p. 205-224.

EFGS – EUROPEAN FORUM FOR GEOSTATISTICS. **GEOSTAT 1A – Representing Census data in a European population grid - Final Report.** 2012.

EICHER, Cory L; BREWER, Cynthia A. Dasymetric mapping and areal interpolation: implementation and evaluation. **Cartography and Geographic Information Science**, vol. 28, n. 2, p. 125-138, 2001.

EMBRAPA – EMPRESA BRASILEIRA DE PESQUISAS AGROPECUÁRIAS; INPE – INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. **Levantamento de informações de uso e cobertura da terra na Amazônia - 2010.** 2010. Disponível em: <[http://www.inpe.br/cra/projetos\\_pesquisas/sumario\\_terraclass\\_2010.pdf](http://www.inpe.br/cra/projetos_pesquisas/sumario_terraclass_2010.pdf)>. Acesso em: 18 set. 2013.

ENTWISLE, Barbara. Putting People into Place. **Demography**, vol. 44, n. 4, p. 687-703, 2007.

ESRI. ArcGIS Desktop Help. 2009.

EWING, Reid; PENDALL, Rolf; CHEN, Don. **Measuring Sprawl and its Impacts.** Volume 1. Smart Growth America, 2002.

EXETER, Daniel J.; BOYLE, Paul; FENG, Zhiqiang; FLOWERDEW, Robin; SCHIERLOH, Nick. The creation of Consistent Areas Through Time (CATTs) in Scotland, 1981–2001. **Population Trends**, n. 119, 2005.

FLOWERDEW, Robin. How serious is the Modifiable Areal Unit Problem for analysis of English census data? **Population Trends**, vol. 145, p. 106, 2011.

FONSECA, Rinaldo B.; DAVANZO, Aurea M.Q.; NEGREIROS, Rovena M.C. (Org.). **Livro verde: desafios para a gestão da região metropolitana de Campinas.** Campinas: UNICAMP, 2002.

FOTHERINGHAM, Stewart. Trends in quantitative methods II: stressing the computational. **Progress in Human Geography**, vol. 22, n. 2, p. 283-292, 1998.

FRANÇA, Vitor Oliveira. **Avaliação de Métodos Dasimétricos para Estimativa Populacional em Pequenas Áreas.** Dissertação (Mestrado) - Escola Nacional de Ciências Estatísticas, Rio de Janeiro, 2012.

GEHLKE, C. E.; BIEHL, Katherine. Certain Effects of Grouping upon the Size of the Correlation Coefficient in Census Tract Material. **Journal of the American Statistical Association**, sup. 29, p. 169-170, 1934.

GOODCHILD, Michael F.; LAM, Nina S. Areal Interpolation: a variant of the traditional spatial problem. **Geoprocessing**, vol. 1, p. 297-312, 1980.

GOODCHILD, Michael F. **Research Initiative 1: Accuracy of Spatial Databases. Final Report.** National Center for Geographic Information and Analysis, University of California, Santa Barbara, 1992.

\_\_\_\_\_; ANSELIN, Luc; DEICHMANN, Uwe. A framework for the areal interpolation of socioeconomic data. **Environment and Planning A**, vol. 25, p. 383-397, 1993.

\_\_\_\_\_. What is Geographic Information Science? In: **NCGIA Core Curriculum in GIScience**. 1997. Disponível em: <<http://www.ncgia.ucsb.edu/giscc/units/u002/u002.html>>. Acesso em: 18 dez. 2012.

\_\_\_\_\_; JANELLE, Donald G. Thinking Spatially in the Social Sciences. In: GOODCHILD, Michael F.; JANELLE, Donald G. (Ed.) **Spatially Integrated Social Science**. New York: Oxford University Press, 2004.

GOTWAY, Carol A.; YOUNG, Linda J. Combining Incompatible Spatial Data. **Journal of the American Statistical Association**, vol. 97, n. 458, p. 632-648, 2002.

\_\_\_\_\_. **Combining Incompatible Spatial Data: An Introductory Overview of Statistical Issues and Methods**. In: WORKSHOP FOR THE SAMSI PROGRAM ON MULTISCALE MODEL DEVELOPMENT AND CONTROL DESIGN, Durham, North Carolina, 2004.

\_\_\_\_\_. A Geostatistical Approach to Linking Geographically Aggregated Data From Different Sources. **Journal of Computational and Graphical Statistics**, vol. 16, n. 1, p. 115–135, 2007.

GRASLAND, Claude; MADELIN, Malika. **The Modifiable Areas Unit Problem – Final Report**. ESPON - European Spatial Planning Observation Network, 2006.

GUTMANN, Myron P.; STERN, Paul C. **Putting People on the Map: Protecting Confidentiality with Linked Social-Spatial Data**. Washington, D.C: National Academy Press, 2007.

\_\_\_\_\_; WITKOWSKI, Kristine; COLYER, Corey; O´ROURKE, JoAnne M.; MCNALLY, James. Providing Spatial Data for Secondary Analysis: Issues and Current Practices Relating to Confidentiality. **Population Research and Policy Review**, vol. 27, n. 6, p. 639-665, 2008.

GUZMÁN, José Miguel; SCHENSUL, Daniel; ZHANG, Sainan. Understanding Vulnerability and Adaptation Using Census Data. In: MARTINE, George; SCHENSUL, Daniel (Ed.). **The Demography of Adaptation to Climate Change**. New York, London and Mexico City: UNFPA, IIED e El Colegio de México, 2013.

HAGERSTRAND, Torsten. **Innovation Diffusion as a Spatial Process**. Tradução de Allan Pred. Chicago: Chicago University Press, 1967.

HAINING, Robert; WISE, Stephen. **Exploratory Spatial Data Analysis**. In: NCGIA CORE CURRICULUM IN GISCIENCE, 1997. Disponível em: <<http://www.ncgia.ucsb.edu/giscc/units/u128/u128.html>>. Acesso em: 18 dez. 2012.

SRIDHARAN, Harini; QIU, Fang. A Spatially Disaggregated Areal Interpolation Model Using Light Detection and Ranging-Derived Building Volumes. **Geographical Analysis**, vol. 45, p. 238–258, 2013.

HOGAN, Daniel J. Crescimento demográfico e Meio Ambiente. **Revista Brasileira de Estudos de População**, vol. 8, n. 1- 2, p. 61-77, 1991.

\_\_\_\_\_. Crescimento populacional, padrões de assentamento e o ambiente físico. In: **Ciências sociais hoje**. Rio de Janeiro: ANPOCS, 1992.

\_\_\_\_\_. Indicadores Sócio-demográficos de Sustentabilidade. In: HOGAN, Daniel J.; BAENINGER, Rosana.; CUNHA, Jose Marcos P.; CARMO, Roberto L. (Org.). **Migração e ambiente nas aglomerações urbanas**. Campinas: NEPO/UNICAMP, 2001.

\_\_\_\_\_. População e Meio Ambiente: a emergência de um novo campo de estudos. In: HOGAN, Daniel J. (Org.). **Dinâmica populacional e mudança ambiental: cenários para o desenvolvimento brasileiro**. Campinas: UNICAMP, 2007.

\_\_\_\_\_; D'ANTONA, Alvaro O.; CARMO, Roberto L. Dinâmica demográfica recente da Amazônia. In: BATISTELA, Mateus; MORAN, Emilio F.; ALVES, Diógenes S. (Org.). **Amazônia: natureza e sociedade em transformação**. São Paulo: EDUSP, 2008.

HOLT, D.; STEEL, D.G.; TRANMER, M.; WRIGLEY, N. Aggregation and ecological effects in geographically based data. **Geographical Analysis**, vol. 28, n. 3, p. 244-261, 1996.

HOLT, James B.; LO, C.P.; HODLER, Thomas W. Dasymetric Estimation of Population Density and Areal Interpolation of Census Data. **Cartography and Geographic Information Science**, vol. 31, n. 2, p. 103-121, 2004.

HONEYCUTT, D.; WOJCIK, J. **Development of a population density surface for the conterminous United States**. GIS/LIS Proceedings, Anaheim, vol. 1, p. 484-496, 1990.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Noções Básicas de Cartografia**, 1998. Disponível em: <[ftp://geoftp.ibge.gov.br/documentos/cartografia/nocoas\\_basicas\\_cartografia.pdf](ftp://geoftp.ibge.gov.br/documentos/cartografia/nocoas_basicas_cartografia.pdf)>. Acesso em: 05 ago. 2012.

\_\_\_\_\_. **Metodologia do Censo Demográfico 2000**. Série Relatórios Metodológicos, vol. 25. Rio de Janeiro, 2003.

\_\_\_\_\_. **Censo 2010 – Síntese das Etapas da Pesquisa**. Rio de Janeiro, 2010a. Disponível em: <[http://censo2010.ibge.gov.br/images/pdf/censo2010/sintese/sintese\\_censo2010\\_portugues.pdf](http://censo2010.ibge.gov.br/images/pdf/censo2010/sintese/sintese_censo2010_portugues.pdf)>. Acesso em: 23 mar. 2011.

\_\_\_\_\_. **Censo Demográfico 2010 – Manual do Recenseador CD – 1.09**. Rio de Janeiro, 2010b.

\_\_\_\_\_. **Base de informações do Censo Demográfico 2010: Resultados do Universo por Setor Censitário**. Rio de Janeiro, 2011a. Disponível em: <[ftp://ftp.ibge.gov.br/Censos/Censo\\_Demografico\\_2010/Resultados\\_do\\_Universo/Agregados\\_por\\_Setores\\_Censitarios/](ftp://ftp.ibge.gov.br/Censos/Censo_Demografico_2010/Resultados_do_Universo/Agregados_por_Setores_Censitarios/)>. Acesso em: 12 mai. 2012.

\_\_\_\_\_. **Indicadores Sociais Municipais**. Rio de Janeiro, 2011b. Disponível em: <[http://www.ibge.gov.br/english/estatistica/populacao/censo2010/indicadores\\_sociais\\_municipais/indicadores\\_sociais\\_municipais.pdf](http://www.ibge.gov.br/english/estatistica/populacao/censo2010/indicadores_sociais_municipais/indicadores_sociais_municipais.pdf)>. Acesso em: 12 abr. 2014.

\_\_\_\_\_. **Metodologia do Censo Demográfico 2010**. Série Relatórios Metodológicos, vol. 41. Rio de Janeiro, 2013.

INSTITUTO GEOLÓGICO. **Carta de risco a movimentos de massa e inundação do município de São Sebastião**. 1996.

INSPIRE – INFRASTRUCTURE FOR SPATIAL INFORMATION IN THE EUROPEAN COMMUNITY. **INSPIRE Specification on Geographical Grid Systems**. Guidelines, version 3.0.1, 2010. Disponível em: <[http://inspire.jrc.ec.europa.eu/documents/Data\\_Specifications/INSPIRE\\_Specification\\_GGS\\_v3.0.1.pdf](http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_Specification_GGS_v3.0.1.pdf)>. Acesso em: 17 Jun. 2013.

IM - INSTITUTO DE MATEMÁTICA, UFRJ. **Ciências Cartográficas – Sistema Cartográfico Nacional**. [s.d.] Disponível em: <<http://www2.sorocaba.unesp.br/professor/robertow/arquivos/Sistema%20Cartografico%20Nacional.pdf>>. Acesso em: 23 mai. 2011.

JAMES, Wesley L.; COSSMAN, Ronald E.; COSSMAN, Jeralynn S.; CAMPBELL, Carol; BLANCHARD, Troy. A brief visual primer for the mapping of mortality trend data. **International Journal of Health Geographics**, vol. 3, n. 7, 2004.

JAPAN STATISTICS BUREAU. **Grid Square Statistics**. [s.d.] Disponível em: <<http://www.stat.go.jp/english/data/mesh/index.htm>>. Acesso em: 18 jun. 2012.

JELINSKI, Dennis E.; WU, Jianguo. The modifiable areal unit problem and implications for landscape ecology. **Landscape Ecology**, vol. 11, n. 3, p. 129-140, 1996.

JOHNSON, Kenneth M.; VOSS, Paul R.; HAMMER, Roger B.; FUGUITT, Glenn V.; MCNIVEN, Scott. Temporal and spatial variation in age-specific net migration in the United States. **Demography**, vol. 42, n. 4, p. 791–812, 2005.

JONES, Bryan. Assessment of a gravity-based approach to constructing future spatial population scenarios. **Journal of Population Research**, vol. 31, n. 1, p. 71-95, 2014.

JOINT RESEARCH CENTER – JRC; Institute for Environment and Sustainability. **Short Proceedings of the First Workshop on European Reference Grids**, ISPRA, Outubro, 2003. Disponível em: <[http://eusoils.jrc.ec.europa.eu/projects/alpsis/Docs/ref\\_grid\\_sh\\_proc\\_draft.pdf](http://eusoils.jrc.ec.europa.eu/projects/alpsis/Docs/ref_grid_sh_proc_draft.pdf)>. Acesso em: 19 ago. 2011.

JUDSON, Dean H. Information integration for constructing social statistics: history, theory and ideas towards a research programme. **Journal of the Royal Statistical Society: Series A (Statistics in Society)**, vol. 170, n. 2, p. 483–501, 2007.

KAMPEL, Silvana A. **Geoinformação para estudos demográficos: representação espacial de dados de população na Amazônia Brasileira**. Tese (Doutorado) - Departamento de Engenharia de Transportes, Escola Politécnica da Universidade de São Paulo, São Paulo, 2004.

KAMINGER, Ingrid. **Assuring confidentiality with grid data**. In: GIS FOR STATISTICS, Luxemburgo, 2007.

KARR, Alan F.; KOHNEN, Christine N.; OGANIAN, A.; REITER, J.P.; SANIL, A.P.. A Framework for Evaluating the Utility of Data Altered to Protect Confidentiality. **The American Statistician**, vol. 60, n. 3, p. 224-232, 2006.

KOBIYAMA, Masato; CHECCHIA, Tatiane; SILVA, Roberto V.; SCHRÖDER, Paulo Henrique; GRANDO, Ângela; REGINATTO, Gisele M. P. **Papel da comunidade e da universidade no gerenciamento de desastres naturais**. In: SIMPÓSIO BRASILEIRO DE DESASTRES NATURAIS, I, Florianópolis, 2004.

KOSINSKI, Leszek A. Demography and Geography. In: PAVLIK, Z. **Position of Demography Among Other Disciplines**. Prague: Charles University, 2000.

KYRIAKIDIS, Phaedon C.; YOO, Eun-Hye. Geostatistical prediction/simulation of point values from areal data. In: MARTIN, David (Ed.). **Proceedings of the 7th International Conference on Geocomputation**, Southampton, UK, 2003.

\_\_\_\_\_. A geostatistical framework for area-to-point spatial interpolation. **Geographical Analysis**, vol. 36, n. 3, p. 259-289, 2004.

\_\_\_\_\_; SCHNEIDER, P.; GOODCHILD, Michael F. Fast geostatistical areal interpolation. In: XIE, Yichun; BROWN, Daniel G. (Ed.). **Proceedings of the 8th International Conference on Geocomputation**, Ann Arbor, Michigan, 2005.

\_\_\_\_\_. Popular cartographic areal interpolation methods viewed from a geostatistical perspective. In: RUAS, Anne. **Proceedings of the 25th International Cartographic Conference**, Paris, France, 2011.

LAM, Nina Siu-Ngan. Spatial Interpolation Methods: A Review. **The American Cartographer**, vol. 10, n. 2, p.129-150, 1983.

LANGFORD, Mitchel; MAGUIRE, D. J; UNWIN, David J. The areal interpolation problem: estimating population using remote sensing in a GIS framework. In: MASSER, Ian; BLAKEMORE, Michael (Ed.). **Handling geographical information: Methodology and potential applications**. London: Longman, 1991.

\_\_\_\_\_; UNWIN, David J. Generating and mapping population density surfaces within a geographical information system. **The Cartographic Journal**, vol. 31, n. 1, p.21-26, 1994.

\_\_\_\_\_. Obtaining population estimates in non-census reporting zones: An evaluation of the 3-class dasymetric method. **Computers, Environment and Urban Systems**, vol. 30, p. 161-180, 2006.

LINARD, Catherine; ALEGANA, Victor A.; NOOR, Abdisalam M.; SNOW, Robert W.; TATEM, Andrew J. A high resolution spatial population database of Somalia for disease risk mapping. **International Journal of Health Geographics**, vol. 9, n. 45, 2010.

\_\_\_\_\_; TATEM, Andrew J. Large-scale spatial population databases in infectious disease research. **International Journal of Health Geographics**, vol. 11, n. 7, 2012.

\_\_\_\_\_; GILBERT, Marius; SNOW, Robert W.; NOOR, Abdisalam M.; TATEM, Andrew J. Population distribution, settlement patterns and accessibility across Africa in 2010. **PLoS ONE**, vol. 7, n. 2, e31743, 2012.

LINHARES, Elizabeth K. Delimitação de áreas para a gestão ambiental: uma contribuição ao estudo de bacias hidrográficas urbanas. In: ENCONTRO NACIONAL DE ESTUDOS POPULACIONAIS, XIX, Caxambu, MG, 2004.

LIVERMAN, Diana; MORAN, Emilio F.; RINDFUSS, Ronald R.; STERN, Paul C. (Ed.). **People and Pixels: Linking Remote Sensing and Social Science**. Washington, D.C.: National Academy Press, 1998.

LIU, XiaoHang. **Estimation of the Spatial Distribution of Urban Population Using High Spatial Resolution Satellite Imagery**. Tese (Doutorado) - University of California, Santa Barbara, 2003.

LOGAN, John R.; ZHANG, Weiwei; XU, Hongwei. Applying spatial thinking in social science research. **GeoJournal**, vol. 75, n. 1, p.15-27, 2010.

LUTZ, Wolfgang; PRSKAWETZ, Alexia; SANDERSON, Warren C. Population and Environment: Methods of Analysis. **Population and Development Review**, vol. 28, 2002.

MAANTAY, Juliana A. Vulnerabilities and risks in population and environment studies. **Population and Environment**, vol. 28, p. 83-112, 2006.

\_\_\_\_\_; MAROKO, Andrew R.; HERRMANN, Christopher. Mapping Population Distribution in the Urban Environment: The Cadastral-based Expert Dasymeric System (CEDS). **Cartography and Geographic Information Science**, vol. 34, n. 2, p. 77-102, 2007.

MARTIN, David. Optimizing census geography: the separation of collection and output geographies. **International Journal of Geographical Information Science**, vol. 12, n. 7, p. 673–685, 1998.

\_\_\_\_\_. **Census 2001: making the best of zonal geographies**. In: THE CENSUS OF POPULATION: 2000 AND BEYOND, Manchester, 2000.

\_\_\_\_\_; NOLAN, A.; TRANMER, M. The application of zone design methodology to the 2001 UK Census. **Environment and Planning A**, vol. 33, n. 11, p. 1949-1962, 2001.

\_\_\_\_\_. Developing the automated zoning procedure to reconcile incompatible zoning systems. **International Journal of Geographical Information Science**, vol. 17, n. 2, p. 181-196, 2003.

MARTINE, George. O lugar do espaço na equação população/meio ambiente. **Revista Brasileira de Estudos de População**, vol. 24, n. 2, p. 181-190, 2007.

MARTINE, George; SCHENSUL, Daniel (Ed.). **The Demography of Adaptation to Climate Change**. New York, London and Mexico City: UNFPA, IIED e El Colegio de México, 2013.

MATTHEWS, Stephen A. **GIS and Spatial Demography**. GIS Resource Document 03-63, Pennsylvania State University, 2003.

\_\_\_\_\_; JANELLE, Donald G.; GOODCHILD, Michael F. **Future Directions in Spatial Demography Specialist Meeting - Final Report**. 2011.

MCHARG, Ian L. **Design with Nature**. New York: John Wiley & Sons, Inc., 25<sup>th</sup> anniversary edition, 1969.

MENNIS, Jeremy. Using Geographic Information Systems to Create and Analyze Statistical Surfaces of Population and Risk for Environmental Justice Analysis. **Social Science Quarterly**, vol. 38, n. 1, p. 281-297, 2002.

MMA - MINISTÉRIO DO MEIO AMBIENTE. **Cadastro Nacional de Unidades de Conservação**. [s.d.] Disponível em:  
<<http://www.mma.gov.br/sitio/index.php?ido=conteudo.monta&idEstrutura=119>>. Acesso em: 20 dez. 2011.

MIRANDA, Evaristo E. A floresta Urbanizada. **Revista Brasileira de Ecologia do Século XXI**, vol. 41, p. 9-10, 1999.

MORTON, Thomas A.; YUAN, Fei. Analysis of population dynamics using satellite remote sensing and US census data. **Geocarto International**, vol. 24, p. 143-63, 2009.

MOUW, Ted. **Visions of The Future: New Directions in Population Research**. In: POPULATION ASSOCIATION OF AMERICA ANNUAL MEETING, Los Angeles, USA, 2000.

MROZINSKI, Richard D. Jr.; CROMLEY, Robert G. Singly- and doubly-constrained methods of areal interpolation for vector-based GIS. **Transactions in GIS**, vol. 3, p. 285-301, 1999.

MULLER, Erika P. L.; CUBAS, Márcia Regina; BASTOS, Laudelino C. Georreferenciamento como instrumento de gestão em unidade de saúde da família. **Revista Brasileira de Enfermagem**, vol. 63, n. 6, 2010.

NAÇÕES UNIDAS. **Handbook on geographic information systems and digital mapping**. Studies in Methods, Series F, N. 79. New York: Nações Unidas, 2000.

NAÇÕES UNIDAS. Comissão Econômica das Nações Unidas para a Europa. **Register-based statistics in the Nordic countries - Review of best practices with focus on population and social statistics.** New York e Genebra: Nações Unidas, 2007a.

\_\_\_\_\_. Divisão de Estatística. **Report of the Expert Group Meeting on Contemporary Practices in Census Mapping and Use of Geographical Information Systems**, 2007b. Disponível em: <[http://unstats.un.org/unsd/demographic/meetings/egm/CensusMappingEGM07/docs/Final\\_Report.pdf](http://unstats.un.org/unsd/demographic/meetings/egm/CensusMappingEGM07/docs/Final_Report.pdf)>. Acesso em: 14 jun. 2010.

\_\_\_\_\_. Divisão de Estatística. **Principles and Recommendations for Population and Housing Census.** Statistical Papers, Série M, N. 67/Rev. 2. New York: Nações Unidas, 2008.

\_\_\_\_\_. Divisão de Estatística. **Handbook on geospatial infrastructure in support of census activities.** Studies in Methods, Série F, N. 103. New York: Nações Unidas, 2009.

\_\_\_\_\_. Divisão de Estatística. **Overview of national experiences for Population and Housing Census 2010 Round**, 2013. Disponível em: <<http://unstats.un.org/unsd/censuskb20/Attachment494.aspx>>. Acesso em: 15 jul. 2013.

\_\_\_\_\_, Programa Ambiental. **Global Resource Information Database.** [s.d.]. Disponível em: <<http://na.unep.net/siouxfalls/datasets/datalist.php>>. Acesso em: 11 out. 2013.

NAOKI, Makita; MASAYUKI, Terada. **Small Area Statistics on Population: Japan's Experience and Challenge.** In: REGIONAL TRAINING WORKSHOP ON USE OF POPULATION AND HOUSING CENSUS DATA FOR LOCAL DEVELOPMENT PLANNING, Chiba, Japão, 2012. Disponível em: <[http://www.unsiap.or.jp/ematerial/ematerial\\_other/HC\\_30\\_010\\_PHC\\_National\\_Planning/HC\\_30\\_010\\_f009\\_Small\\_Area\\_Statistics.pdf](http://www.unsiap.or.jp/ematerial/ematerial_other/HC_30_010_PHC_National_Planning/HC_30_010_f009_Small_Area_Statistics.pdf)>. Acesso em: 23 jun. 2013.

NATIONAL INSTITUTE OF CHILD HEALTH AND HUMAN DEVELOPMENT - NICHD. **Goals and Opportunities: 2002-2006**, 2002. Disponível em: <[https://www.nichd.nih.gov/publications/pubs/dbsb\\_goals\\_02to06.cfm](https://www.nichd.nih.gov/publications/pubs/dbsb_goals_02to06.cfm)>. Acesso em: 12 mai. 2011.

\_\_\_\_\_. **Demographic and Behavioral Sciences (DBS) Branch Long-Range Planning 2006-2007: Highlights from a Panel Discussion**, 2007. Disponível em: <[http://www.nichd.nih.gov/publications/pubs/upload/dbsb\\_panel\\_highlights\\_2006\\_2007.pdf](http://www.nichd.nih.gov/publications/pubs/upload/dbsb_panel_highlights_2006_2007.pdf)>. Acesso em: 02 jun. 2011.

NÚCLEO DE ESTUDOS DE POPULAÇÃO ELZA BERQUÓ – NEPO. **População e Ambiente**. [s.d.]. Disponível em: <<http://www.nepo.unicamp.br/pesquisa/linhaPesq7.html>>. Acesso em: 28 fev. 2014.

NORMAN, Paul; REES, Philip; BOYLE, Paul. Achieving Data Compatibility over Space and Time: Creating Consistent Geographical Zones. **International Journal of Population Geography**, vol. 9, p. 365-386, 2003.

OFFICE FOR NATIONAL STATISTICS - ONS. **Changes to Output Areas and Super Output Areas in England and Wales, 2001 to 2011**, 2012. Disponível em: <<http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/census/output-area--oas-/index.html>>. Acesso em: 23 ago. 2013.

OJIMA, Ricardo. **Análise comparativa da dispersão urbana nas aglomerações urbanas brasileiras: elementos teóricos e metodológicos para o planejamento urbano e ambiental**. Tese (Doutorado) – Instituto de Filosofia e Ciências Humanas, UNICAMP, 2007.

\_\_\_\_\_; MARTINE, George. Resgates sobre População e Ambiente: breve análise da Dinâmica Demográfica e a Urbanização nos Biomas Brasileiros. **Ideias**, n. 5, p. 55-70, 2012.

OLEA, Ricardo A. (Ed.) **Geostatistical Glossary and Multilingual Dictionary**. New York: Oxford University Press, 1991.

OLIVEIRA, Evangelina X. G.; TRAVASSOS, Cláudia; CARVALHO, Marília Sá. Acesso à internação hospitalar nos municípios brasileiros em 2000: territórios do Sistema Único de Saúde. **Cadernos de Saúde Pública**, Rio de Janeiro, vol. 20, suppl. 2, 2004.

\_\_\_\_\_; MELO, Enirtes C. P.; PINHEIRO, Rejane, S.; NORONHA, Cláudio, P.; CARVALHO, Marília S. Acesso à assistência oncológica: mapeamento dos fluxos origem-destino das internações e dos atendimentos ambulatoriais. O caso do câncer de mama. **Cadernos de Saúde Pública**, vol. 27, n. 2, 2011.

OPENSHAW, Stan. A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modelling. **Transactions of the Institute of British Geographers, New Series**, vol. 2, n. 4, p. 459-472, 1977.

\_\_\_\_\_; TAYLOR, Paul. A Million or so Correlation Coefficients. In: WRIGLEY, N. (Ed.) **Statistical Methods in the Spatial Sciences**. London: Pion, 1979.

\_\_\_\_\_. The Modifiable Areal Unit Problem. In: WRIGLEY, N., BENNETT, R. (Ed.). **Quantitative Geography: A British View**. London: Routledge & Kegan Paul, 1981, p. 60–69.

OPENSHAW, Stan. The modifiable areal unit problem. **Concepts and Techniques in Modern Geography**, n. 38. Norwich: Geo Books, 1984.

\_\_\_\_\_.; RAO, L. Algorithms for reengineering 1991 Census geography. **Environment and Planning A**, vol. 27, 425-446, 1995.

\_\_\_\_\_. Developing GIS relevant zone based spatial analysis methods. In: LONGLEY, Paul A. and BATTY, M. (Ed.) **Spatial Analysis: Modeling in a GIS Environment**. Cambridge, U.K.: GeoInformation International, 1996.

\_\_\_\_\_.; ALVANIDES, Seraphim. Applying geocomputation to the analysis of spatial distributions. In: LONGLEY, Paul A., GOODCHILD, Michael, MAGUIRE, David J., RHIND, David W. (Ed.) **Geographical Information Systems: Principles, Techniques, Applications and Management**. Chichester: Wiley, 1999.

OWENS, Peter M.; TITUS-ERNSTOFF, Linda; GIBSON, Lucinda; BEACH, Michael L.; BEAUREGARD, Sandy; DALTON, Madeline A. Smart density: A more accurate method of measuring rural residential density for health-related research. **International Journal of Health Geographics**, vol. 9, n. 8, 2010.

PANDIT, Kavita; BAGCHI-SEN, Sharmistha. The spatial dynamics of U.S. fertility, 1970–1990. **Growth and Change**, vol. 24, n. 2, p. 229–246, 1993.

PARK, Robert; BURGESS, Ernest W.; MCKENZIE, Roderick D. **The City**. Chicago: University of Chicago Press, 1925.

PEÑA, Landy. L. S. Métodos para el análisis espacial. Una aplicación al estudio de la geografía de la pobreza. In: CONGRESO DE LA ASOCIACION LATINOAMERICANA DE POBLACION, II, Guadalajara, México, 2006.

PLUMEJEAUD, Christine ; PRUD'HOMME, Julie ; DAVOINE, Paule-Annick; GENSEL, Jérôme. Transferring Indicators into Different Partitions of Geographic Space. In: TANIAR, David; GERVASI, Osvaldo; MURGANTE, Beniamino; PARDEDE, Eric; APDUHAN, Bernady O. (Ed.). **ICCSA'10 Proceedings of the 2010 international conference on Computational Science and Its Applications - Volume Part I**. Berlin, Heidelberg: Springer-Verlag, 2010.

PRUDENTE, Cristiane N.; REIS, Rui Bran J. Mapeamento das áreas de riscos de desastres naturais no estado de Minas Gerais. In: ENCONTRO NACIONAL DE ESTUDOS POPULACIONAIS, XVII, Caxambu, MG, 2010.

QIU, Fang; SRIDHARAN, Harini. **A spatially disaggregated areal-interpolation model using LiDAR derived building volumes**. In: ANNUAL CONFERENCE OF AMERICAN ASSOCIATION OF GEOGRAPHERS, Los Angeles, CA, 2013.

RAMOS, Frederico R. **Análise Espacial de Estruturas Intra-Urbanas: O Caso de São Paulo**. Dissertação (Mestrado) - Instituto Nacional de Pesquisas Espaciais, 2002.

RASE, Wolf-Dieter. Volume-Preserving Interpolation of a Smooth Surface from Polygon-Related Data. **Journal of Geographical Systems**, vol. 3, n. 2, p. 199-213, 2001.

RASE, Daniel. Dealing with the modifiable areal unit problem: Spatial transformation methods for the analysis of geographic data. In: EUROPEAN ENVIRONMENT AGENCY. **Towards agri-environmental indicators: Integrating statistical and administrative data with land cover information**. Copenhagen: EEA, 2001, p. 25-38.

REDE INTEGRADA DE INFORMAÇÕES PARA A SAÚDE - RIPSA. **Indicadores básicos para a saúde no Brasil: conceitos e aplicações**. 2ª ed. Brasília: Organização Pan-americana de Saúde (OPAS), 2008.

REIBEL, Micahel; BUFALINO, Michael E. Street-weighted interpolation techniques for demographic count estimation in incompatible zone systems. **Environment and Planning A**, vol. 37, n. 1, p. 127-139, 2005.

\_\_\_\_\_; AGRAWAL, Aditya. Areal Interpolation of Population Counts Using Pre-classified Land Cover Data. **Population Research and Policy Review**, vol. 26, p. 619-633, 2007.

RINDFUSS, Ronald R.; STERN, Paul C. Linking Remote Sensing and Social Science: The Need and the Challenges. In: LIVERMAN, Diana; MORAN, Emilio F.; RINDFUSS, Ronald R.; STERN, Paul C. (Ed.). **People and Pixels: Linking Remote Sensing and Social Science**. Washington, D.C.: National Academy Press, 1998.

ROBINSON, W. Ecological correlations and the behavior of individuals. **American Sociological Review**, vol. 15, p. 351-57, 1950.

ROSENBERG, Mark W. Medical or Health Geography? Population, Peoples and Places. **International Journal of Population Geography**, vol. 4, p. 212-226, 1998.

RUMSEY, Abby Smith. **Scholarly Communication Institute 7: Spatial Technologies and the Humanities**. Full Report. University of Virginia, 2009.

RUSANEM, Jarmo; NAUKKARINEN, Arvo; COLPAERT, Alfred. Square Kilometer Grid System: An efficient Database in Rural Studies. **Geography Research Forum**, vol. 13, p.129-138, 1993.

RUSANEM, Jarmo; MUILU, Toivo; COLPAERT, Alfred; NAUKKARINEN, Arvo. Finnish socio-economic grid data, GIS and the hidden geography of unemployment. **Tijdschrift voor Economische en Sociale Geographie**, vol. 92, n. 2, p. 139-147, 2001.

SÃO PAULO. **Decreto nº 57.512**, de 11 de Novembro de 2011. Institui o Programa Estadual de Prevenção de Desastres Naturais e de Redução de Riscos Geológicos e dá providências correlatas. São Paulo: Diário Oficial do Estado de São Paulo, 2011.

SCHUURMAN, Nadine; LESZCZYBSKI, Agnieszka; FIEDLER, Rob; GRUND, Darrin; BELL, Nathaniel. Building an Integrated Cadastral Fabric for Higher Resolution Socioeconomic Spatial Data Analysis. In: RIEDL, Andreas; KAINZ, Wolfgang; ELMES, Gregory A. (Ed.). **Progress in Spatial Data Handling: 12<sup>th</sup> International Symposium on Spatial Data Handling**. Berlin, Heidelberg, New York: Springer, 2006.

SMA - SECRETARIA DE MEIO AMBIENTE DO ESTADO DE SÃO PAULO, Coordenadoria de Planejamento Ambiental. **Mapa de Cobertura da Terra do Estado de São Paulo – 2010 - escala 1:100.000**. 2013. Disponível em: <[http://www.ambiente.sp.gov.br/cpla/files/2013/10/Ficha\\_Tecnica\\_Cobertura\\_Terra\\_ESP.pdf](http://www.ambiente.sp.gov.br/cpla/files/2013/10/Ficha_Tecnica_Cobertura_Terra_ESP.pdf)>. Acesso em: 18 set. 2013.

SEHLIN, Johnny. **Production of grid statistics at Statistics Sweden**. In: EUROPEAN FORUM FOR GEOSTATISTICS WORKSHOP, Lisboa, Portugal, 2011.

SOJA, Edward. Taking space personally. In: WARF, Barney; ARIAS, Santa (Ed.). **The Spatial Turn: Interdisciplinary Perspectives**. Londres e Nova York: Routledge, 2009.

SOUZA, Gustavo O. C.; TORRES, Haroldo G. O estudo da metrópole e o uso de informações georreferenciadas. **São Paulo em Perspectiva**, vol. 17, n. 3-4, p. 35-44, 2003.

SNOW, John. **On the Mode of Communication of Cholera**. London: John Churchill, 1855.

STATISTICS FINLAND. **Grid Database 2012**, 2012. Disponível em: <[http://tilastokeskus.fi/tup/ruututietokanta/griddb2012\\_description.pdf](http://tilastokeskus.fi/tup/ruututietokanta/griddb2012_description.pdf)>. Acesso em: 25 jun. 2013.

STEEL, D.G.; HOLT, D. Analyzing and adjusting aggregation effects: The ecological fallacy revisited. **International Statistical Review**, vol. 64, n. 1, p. 39-60, 1996.

STODDARD Steven T.; FORSHEY Brett M.; MORRISON Amy C.; PAZ SOLDAN Valerie A.; VAZQUEZ-PROKOPEC Gonzalo M.; ASTETE Helvio; REINER Robert C.; VILCARROMERO Stalin; ELDER John P.; HALSEY Eric S.; KOCHEL Tadeusz J.; KITRON Uriel; SCOTT Thomas W. House-to-house human movement drives dengue virus transmission. **PNAS**, vol. 110, n. 3, p. 994-999, 2013.

STEINNOCHER Klaus, KAMINGER Ingrid, KOSTL Mario, WEICHSELBAUM Jürgen. **Gridded Population – new data sets for an improved disaggregation approach**. In: EUROPEAN FORUM FOR GEOSTATISTICS WORKSHOP, Tallin, Estonia, 2010.

STERN, Paul C.; YOUNG, Oran R.; DRUCKMAN, Daniel (Ed.). **Global Environmental Change: Understanding the Human Dimensions**. Committee on the Human Dimensions of Global Change, National Research Council. Washington, D.C.: National Academy Press, 1992.

STRAND, Geir-Harald; BLOCH, Vilni V. H. **Statistical grids for Norway**. Statistics Norway, Department of Economic Statistics, 2009.

SUTTON, Paul C. Modeling population density with night-time satellite imagery and GIS. **Computers, Environment and Urban Systems**, vol. 21, n. 3-4, p. 227-244, 1997.

SUI, Daniel. GIS, Environmental Equity Analysis, and the Modifiable Areal Unit Problem. In: CRAGLIA, Massimo; ONSRUD, Harlan (Ed.) **Geographic Information Research: Trans-Atlantic Perspectives**. London, UK: Taylor and Francis, 1999.

SWEENEY, Stuart. **Enabling Spatial Demography: Concepts, Tools, and Resources**. In: DEMSEM SEMINAR, University of Madison, Wisconsin, 2002.

TAMMILEHTO-LUODE, Marja, BACKER, Lars; ROGSTAT, Lars. **Grid data and area delimitation by definition. Towards a better European territorial statistical system**. In: CONFERENCE OF EUROPEAN STATICIANS, Suíça, 2000.

\_\_\_\_\_. **Opportunities and challenges of grid-based statistics**. In: WORLD STATISTICS CONGRESS OF THE INTERNATIONAL STATISTICAL INSTITUTE, Irlanda, 2011.

TAMMISTO, Rina. **Merging national population grids (bottom-up approach) into a European dataset**. In: GIS FOR STATISTICS. Luxembourg, 2007.

TAPP, Anna F. Areal interpolation and dasymetric mapping methods using local ancillary data sources. **Cartography and Geographic Information Science**, vol. 37, p. 215-28, 2010.

TATEM, Andrew J.; NOOR, Abdisalam M.; VON HAGEN, Craig; DI GREGORIO, Antonio; HAY, Simon I. High Resolution Population Maps for Low Income Nations: Combining Land Cover and Census in East Africa. **PLoS ONE**, vol. 2, n. 12, e1298, 2007.

\_\_\_\_\_; QIU, Youliang; SMITH, David L.; SABOT, Oliver; ALI, Abdullah S.; MOONEN, Bruno. The use of mobile phone data for the estimation of the travel patterns and imported Plasmodium falciparum rates among Zanzibar residents. **Malaria Journal**, vol. 8, n. 287, 2009.

\_\_\_\_\_; ADAMO, Susana; BHARTI, Nita; BURGERT, Clara R.; CASTRO, Marcia; DORELIEN, Audrey; FINK, Gunter; LINARD, Catherine; JOHN, Mendelsohn; MONTANA, Livia; MONTGOMERY, Mark R.; NELSON, Andrew; NOOR, Abdisalan M.; PINDOLIA, Deepa; YETMAN, Gregory; BALK, Deborah. Mapping populations at risk: improving spatial demographic data for infectious disease modeling and metric derivation. **Population Health Metrics**, vol. 10, n. 8, 2012.

TAYLOR, D. R. Fraser. **Global Geographic Information Management: Some Institutional and Data Sharing Issues in Integrating Geospatial and Statistical Data**. In: SECOND PREPARATORY MEETING OF THE PROPOSED UNITED NATIONS COMMITTEE ON GLOBAL GEOGRAPHIC INFORMATION MANAGEMENT, New York, 2010.

TRAINOR, Tim. **Combining variable spatial data with grids to improve data visualization**. In: CONFERENCE OF EUROPEAN STATICIANS, Paris, 2010.

TOBLER, Waldo R.. A computer movie simulating urban growth in the Detroit region. **Economic Geography**, vol. 46, n. 2, p. 234-240, 1970.

\_\_\_\_\_; DEICHMANN, Uwe; GOTTSEGEN, Jon; MALOY, Kelly. **The global demography project**. Technical Report TR-95-6. National Center for Geographic Information and Analysis. Santa Barbara: Department of Geography, University of California, 1995.

\_\_\_\_\_; DEICHMANN, Uwe; GOTTSEGEN, Jon; MALOY, Kelly. World Population in a Grid of Spherical Quadrilaterals. **International Journal of Population Geography**, vol. 3, p. 203–225, 1997.

TURNER, Andy; OPENSHAW, Stan. **Disaggregative Spatial Interpolation**. In: GISRUK, Glamorgan, País de Gales, 2001.

VILLAÇA, Flávio. **Efeitos do Espaço sobre o Social na Metrópole Brasileira**. In: ENCONTRO NACIONAL DA ANPUR, VII, Recife, 1997.

VOSS, Paul R. Demography as a Spatial Social Science. **Population Research and Policy Review**, vol. 26, p. 457-476, 2007.

WACHTER, Kenneth W. Spatial Demography Special Feature: Spatial Demography. **PNAS**, vol. 102, n. 43, 2005.

WARF, Barney; ARIAS, Santa. Introduction: the reinsertion of space into the social sciences and humanities. In: WARF, Barney; ARIAS, Santa (Ed.). **The Spatial Turn: Interdisciplinary Perspectives**. Londres e Nova York: Routledge, 2009.

WEEKS, John R. The Role of Spatial Analysis in Demographic Research. In: GOODCHILD, Michael F.; JANELLE, Donald G. (Ed.). **Spatially Integrated Social Science**. New York: Oxford University Press, 2004.

\_\_\_\_\_; GETIS, Arthur; HILL, Allan G.; GADALLA, M. Saad; RASHED, Tarek. The fertility transition in Egypt: Intraurban patterns in Cairo. **Annals of the Association of American Geographers**, vol. 94, n. 1, p. 74-93, 2004.

WRIGHT, John K. A method of mapping densities of population: With Cape Cod as an example. **Geographical Review**, vol. 26, n. 1, p. 103-110, 1936.

WONG, D. W. S. Aggregation effects in geo-referenced data. In: GRIFFITHS, Daniel A. (Ed.). **Advanced Spatial Statistics**. Boca Raton, Florida: CRC Press, 1996.

WRIGLEY, N. Revisiting the modifiable areal unit problem and the ecological fallacy. In CLIFF, Andrew; GOULD, Peter; HOARE, Anthony e THRIFT, Nigel (Ed.). **Diffusing geography: Essays for Peter Haggett**. New York: Wiley-Blackwell, 1995.

WU, Shuo-sheng; QIU, Xiaomin; WANG, Le. Population Estimation Methods in GIS and Remote Sensing: A Review. **GIScience and Remote Sensing**, vol. 42, n. 1, p. 58-74, 2005.

XIE, Yichun. The Overlaid Network Algorithms for Areal Interpolation Problem. **Computers, Environment and Urban Systems**, vol. 19, n. 4, p. 287-306, 1995.

XIE, Zhixiao. A framework for interpolating the population surface at the residential housing-unit level. **GIScience and Remote Sensing**, vol. 43, p. 233–251, 2006.

YOO, Eun-Hye; KYRIAKIDIS, Phaedon C. Area-to-point kriging with inequality type data. **Journal of Geographical Systems**, vol. 8, n. 4, p. 357-390, 2006.

YUAN, Yew; SMITH, Richard M.; LIMP, W. Fredrick. Remodelling census population with spatial information from Landsat TM imagery. **Computers Environment and Urban Systems**, vol. 21, p. 245–258, 1997.

YOUNG, Linda J.; GOTWAY, Carol A. Linking spatial data from different sources: the effects of change of support. **Stochastic Environmental Research and Risk Assessment**, vol. 21, n. 5, 2007.

ZANDBERGEN, Paul A. Dasymetric Mapping Using High Resolution Address Point Datasets. **Transactions in GIS**, vol. 15(s1), p. 5-27, 2011.

\_\_\_\_\_; IGNIZIO, Drew A. Comparison of Dasymetric Mapping Techniques for Small-Area Population Estimates. **Cartography and Geographic Information Science**, vol. 37, n. 3, p. 199-214, 2010.

ZELINSKY, Wilbur. **A prologue to population geography**. Englewood Cliffs, New Jersey: Prentice-Hall, 1966.

ZIMMERMAN, Dale L.; JIE, Li. The effects of local street network characteristics on the positional accuracy of automated geocoding for geographic health studies. **International Journal of Health Geographics**, vol. 9, p. 8, 2010.



## APÊNDICES

### APÊNDICE A - Avaliação das dimensões das células

#### 1 Introdução

Com o objetivo de conjugar o critério de confidencialidade e a necessidade de ter unidades espaciais com as menores dimensões possíveis, foi realizado um estudo para avaliar a dimensão ideal para as células de uma grade estatística. Foram selecionadas duas Unidades da Federação para este estudo: São Paulo e Pará. Esses dois estados foram selecionados por apresentarem características bastante distintas com relação à distribuição espacial da população e, por isso, poderem ajudar na determinação da dimensão de célula a ser utilizada em todo o território nacional. Foram utilizados os dados do Censo Demográfico 2010, incluindo as coordenadas geográficas dos domicílios. Com base nesses dados, foi possível agregar em células as informações referentes ao total de população e total de domicílios particulares permanentes, o que foi feito para os diferentes tamanhos de células que serão avaliados.

Primeiramente, foi selecionada uma dimensão mínima para a célula com base nos valores utilizados por outros países (EFGS, 2012) e nas dimensões de cartas existentes no Sistema Cartográfico Nacional (ver Cap. 2, Quadro 3). O valor selecionado corresponde à carta na escala 1:2.000 e é aproximadamente 1.150 m ou 37,5". A partir desse valor, a dimensão da célula foi multiplicada e dividida duas vezes para atingir os demais valores. As dimensões de células utilizadas foram 7,5", 18,75", 37,5", 75" e 2,5', equivalentes a aproximadamente 230, 580, 1.150, 2.300 e 4.600 m.

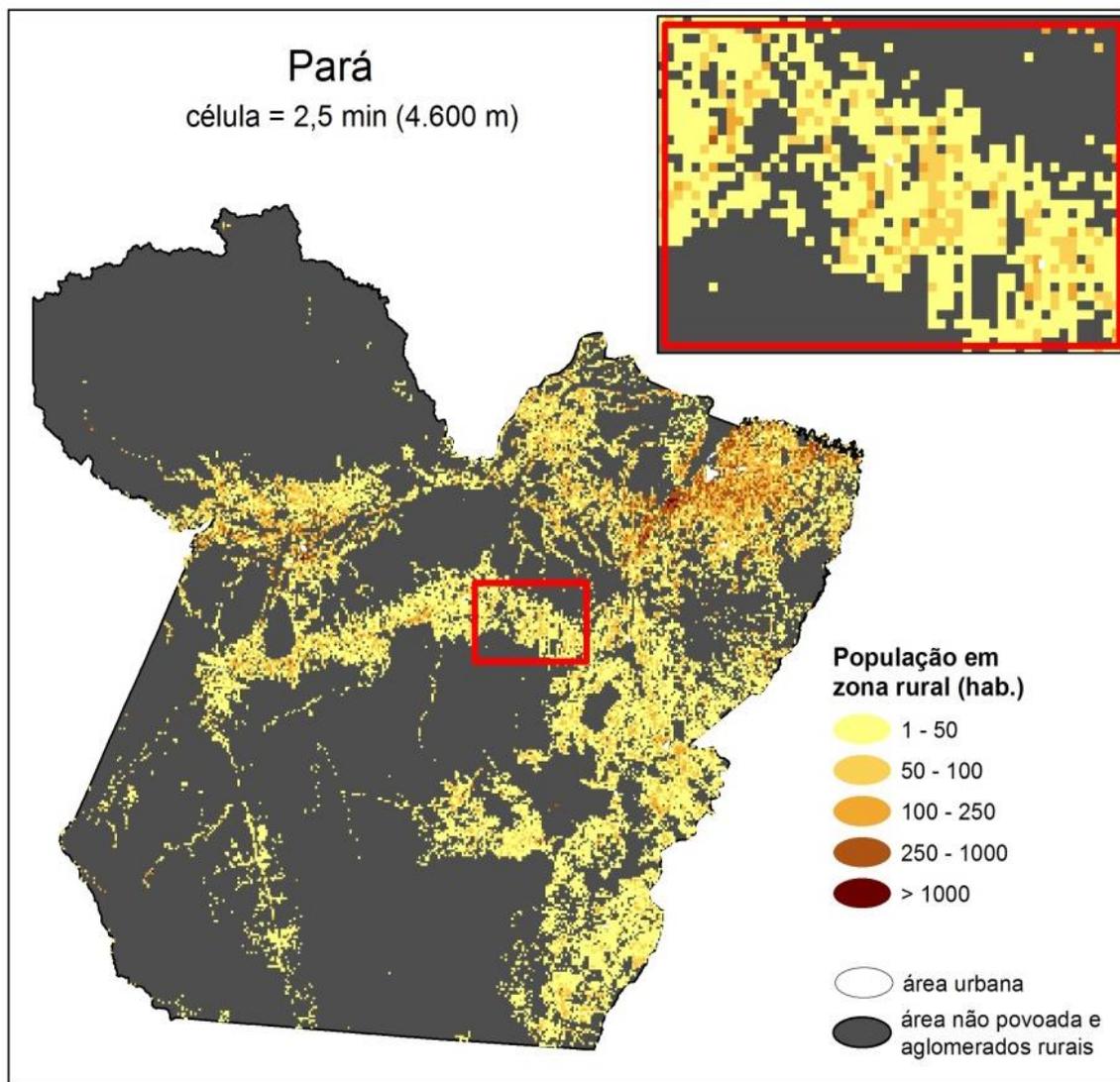
Na primeira análise, foi feita uma interpretação visual dos padrões da variável total de população utilizando mapas temáticos construídos com as células de diferentes dimensões. O objetivo desta análise é avaliar as perdas na

identificação visual de padrões quando são utilizadas unidades espaciais com diferentes dimensões. A segunda análise teve o objetivo de avaliar a resolução espacial, verificando se haveria perda ou ganho neste indicador de acordo com o tamanho da célula. A terceira análise enfatizou a questão da confidencialidade, avaliando a quantidade de dados que deveria ser omitida na distribuição ao usuário final.

## **2 Representação espacial**

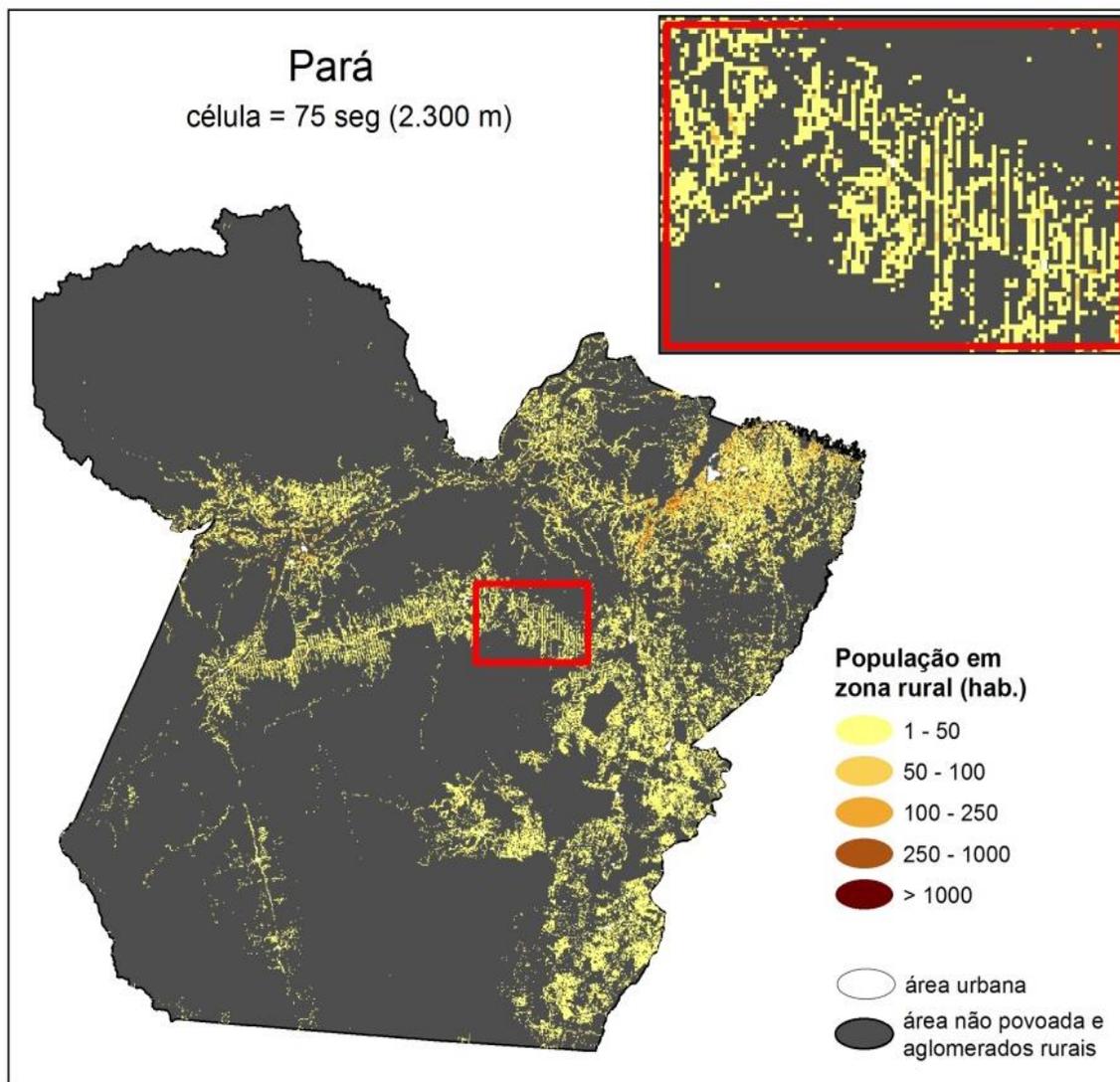
As Figuras 1 a 10 mostram mapas de distribuição da população para as duas regiões de estudo, os estados do Pará e de São Paulo, cada um desses mapas utilizando uma dimensão diferente para as células da grade estatística. No caso do Pará, podemos observar que os mapas com dimensões de células menores ou iguais a 1' permitem que sejam identificados visualmente os principais vetores de ocupação da região, quais sejam os rios e estradas; já com células de aproximadamente 4 km de dimensão esta identificação visual não é possível. No caso de São Paulo, como a ocupação segue um padrão mais homogêneo em todo o estado, a possibilidade de identificação visual de padrões espaciais é dificultada. No entanto, a utilização de células com dimensões iguais ou menores do que 30'' permite a visualização de corredores de ocupação, como pode ser visto nos detalhes dos mapas apresentados.

Figura 1 - População em zona rural com células de 2,5', Pará, 2010.



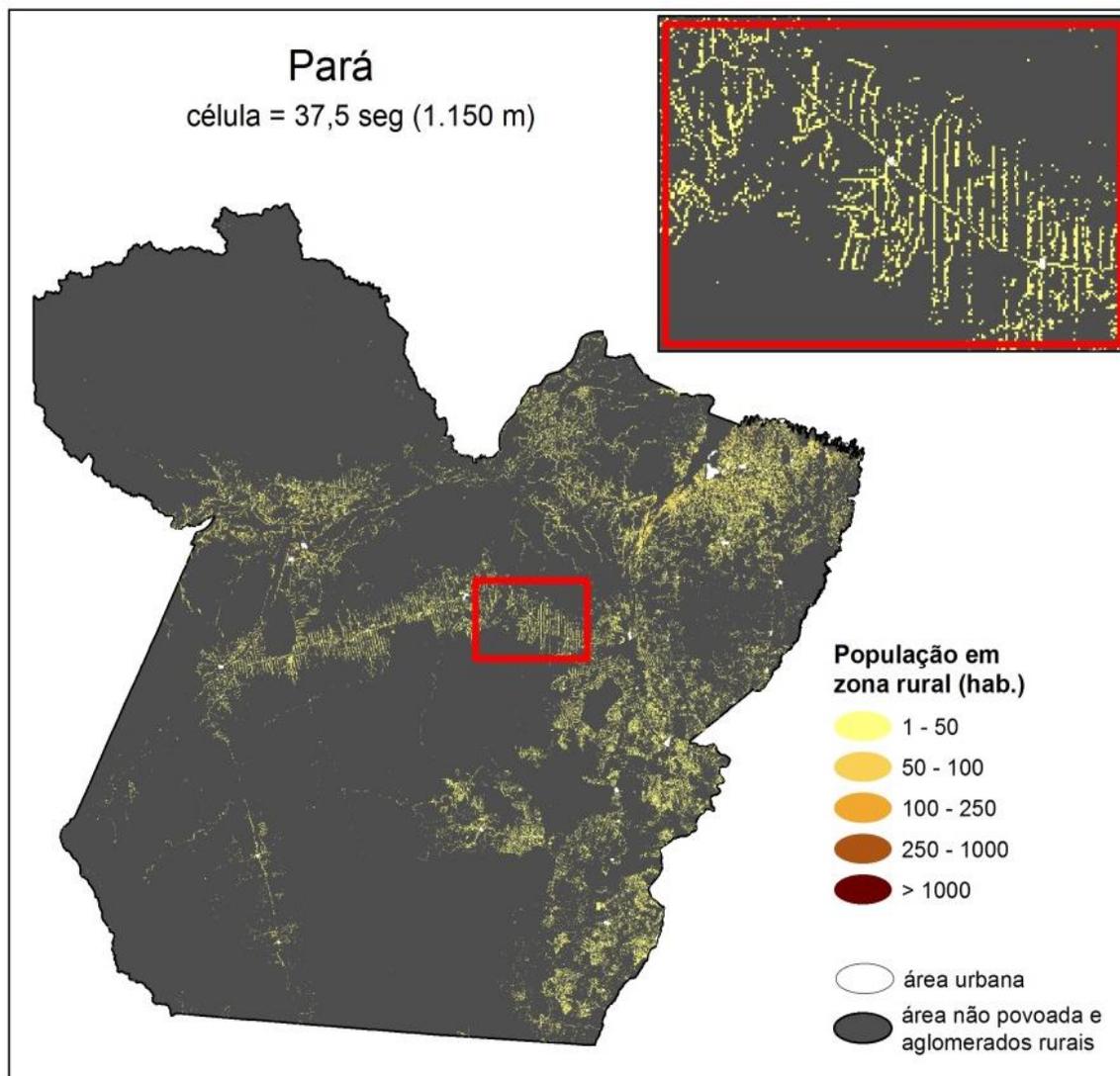
Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Figura 2 - População em zona rural com células de 75", Pará, 2010.



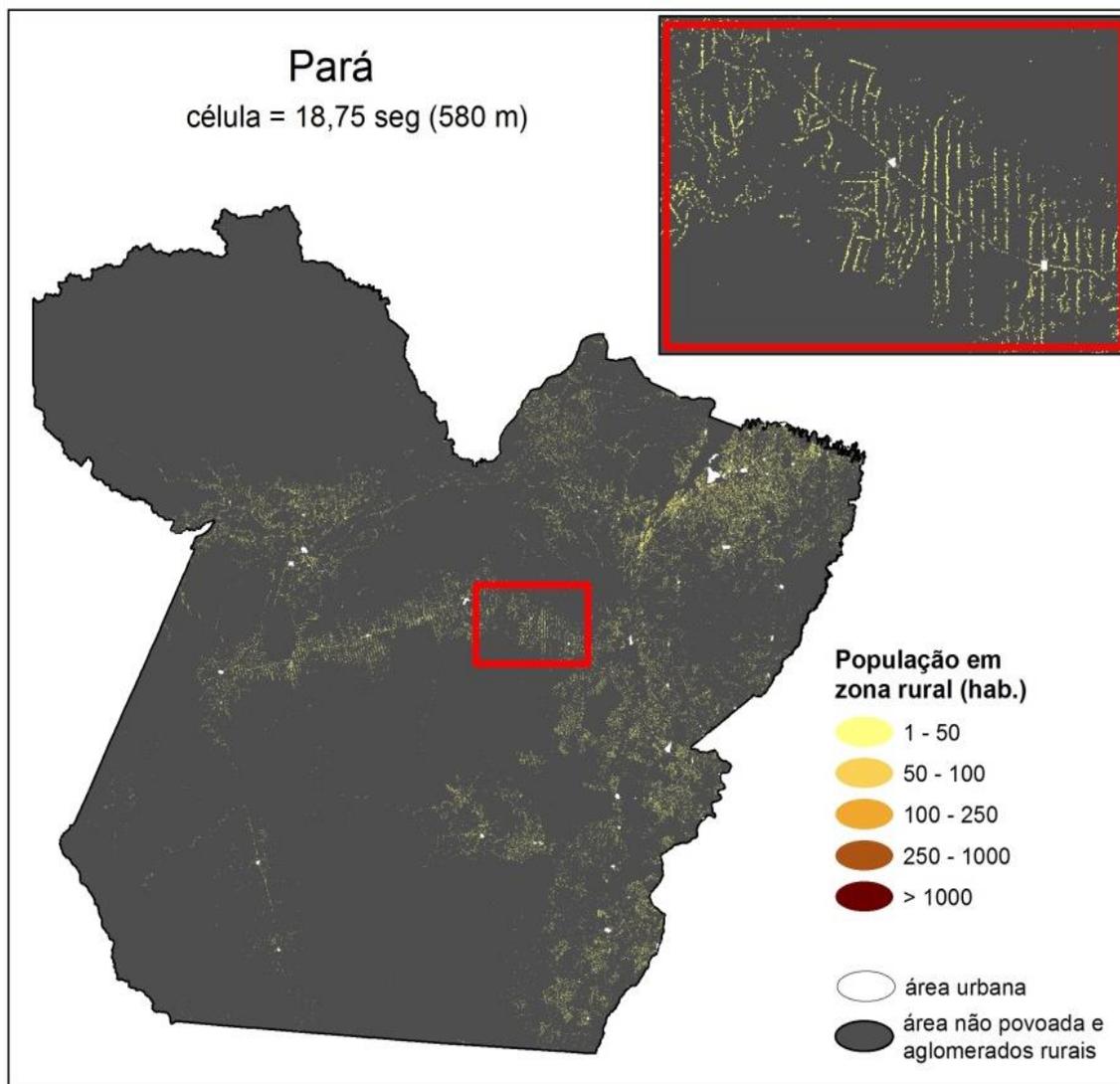
Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Figura 3 - População em zona rural com células de 37,5", Pará, 2010.



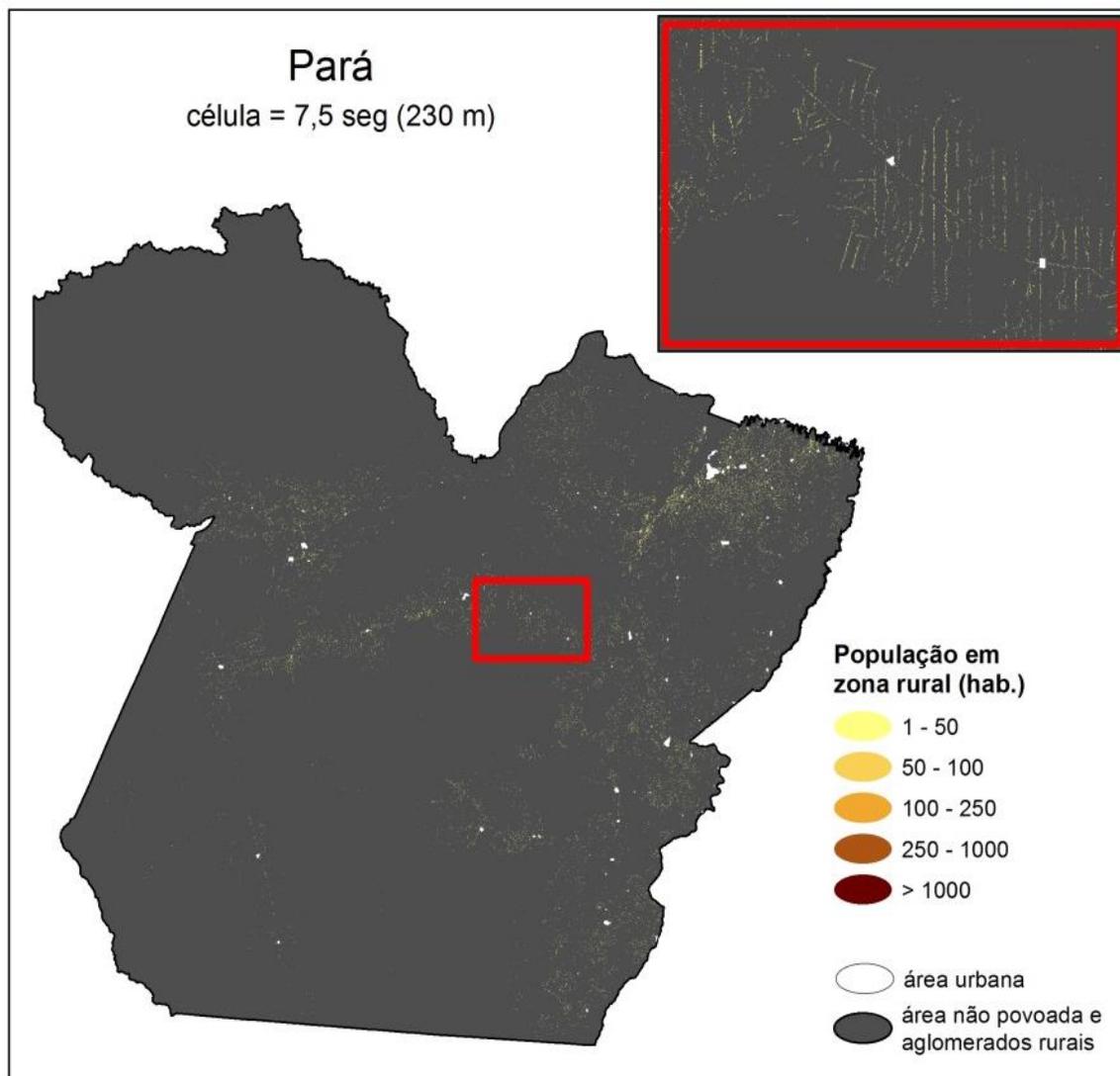
Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Figura 4 - População em zona rural com células de 18,75", Pará, 2010.



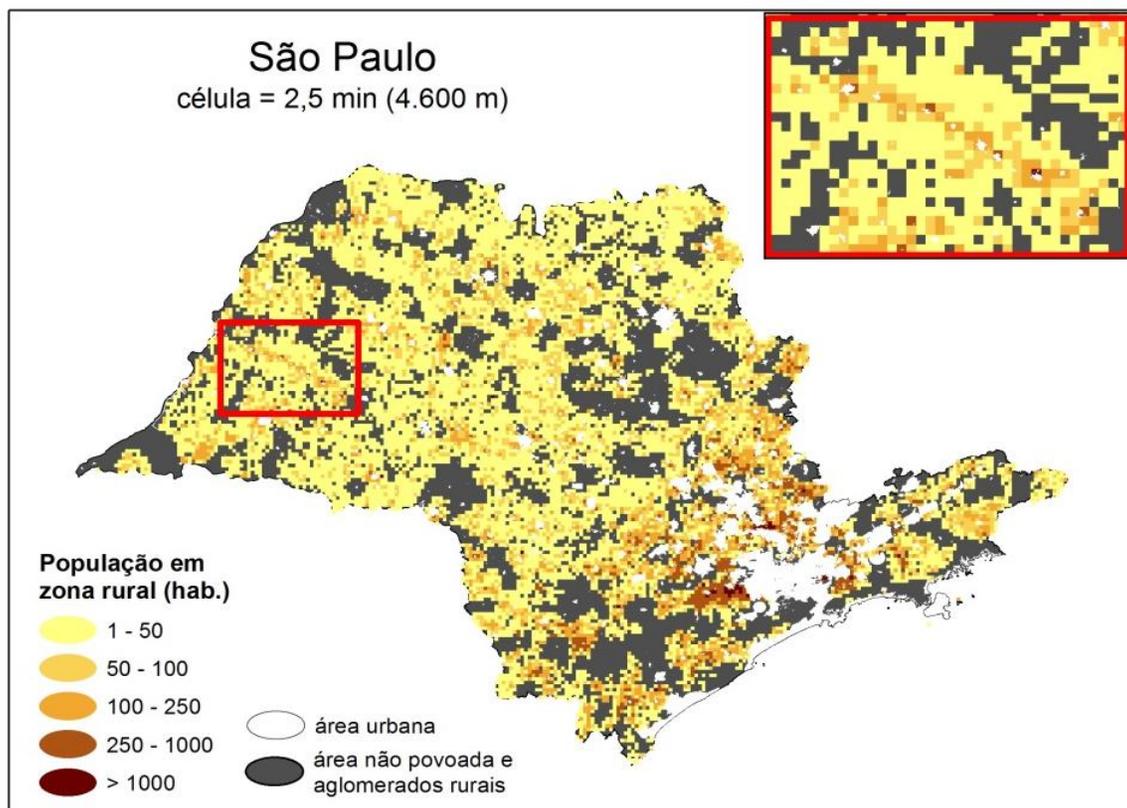
Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Figura 5 - População em zona rural com células de 7,5", Pará, 2010.



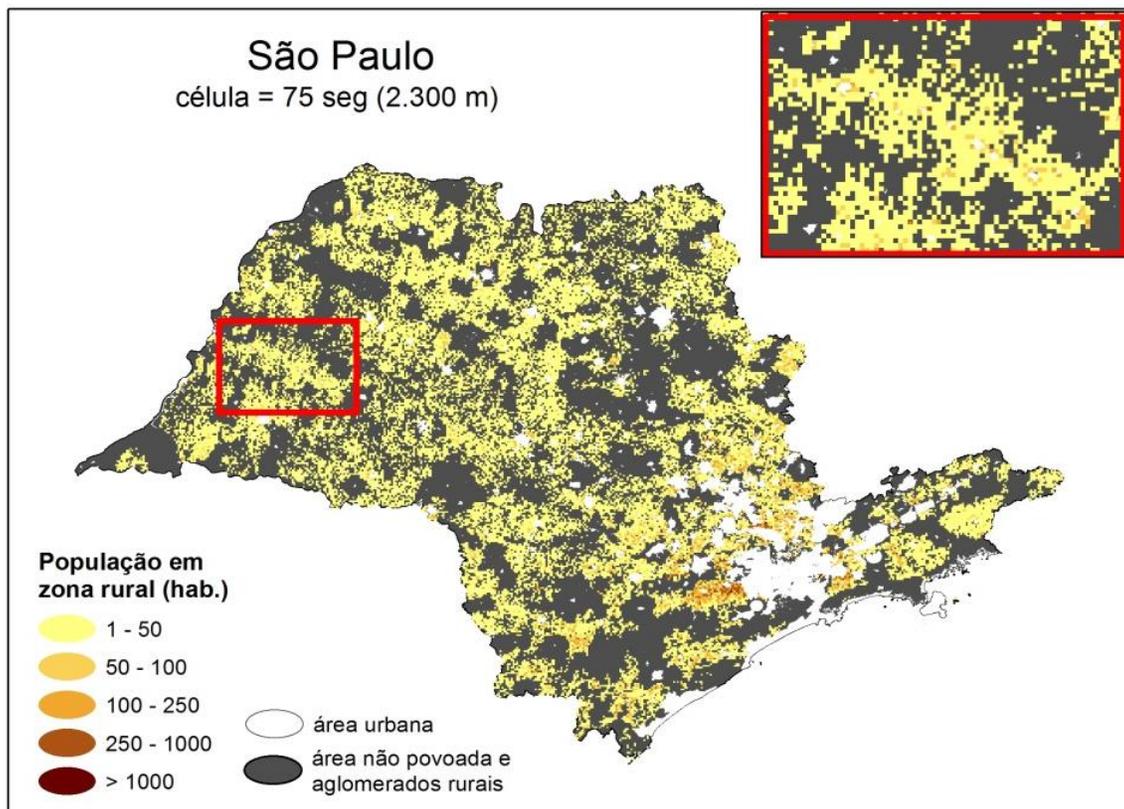
Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Figura 6 - População em zona rural com células de 2,5', São Paulo, 2010.



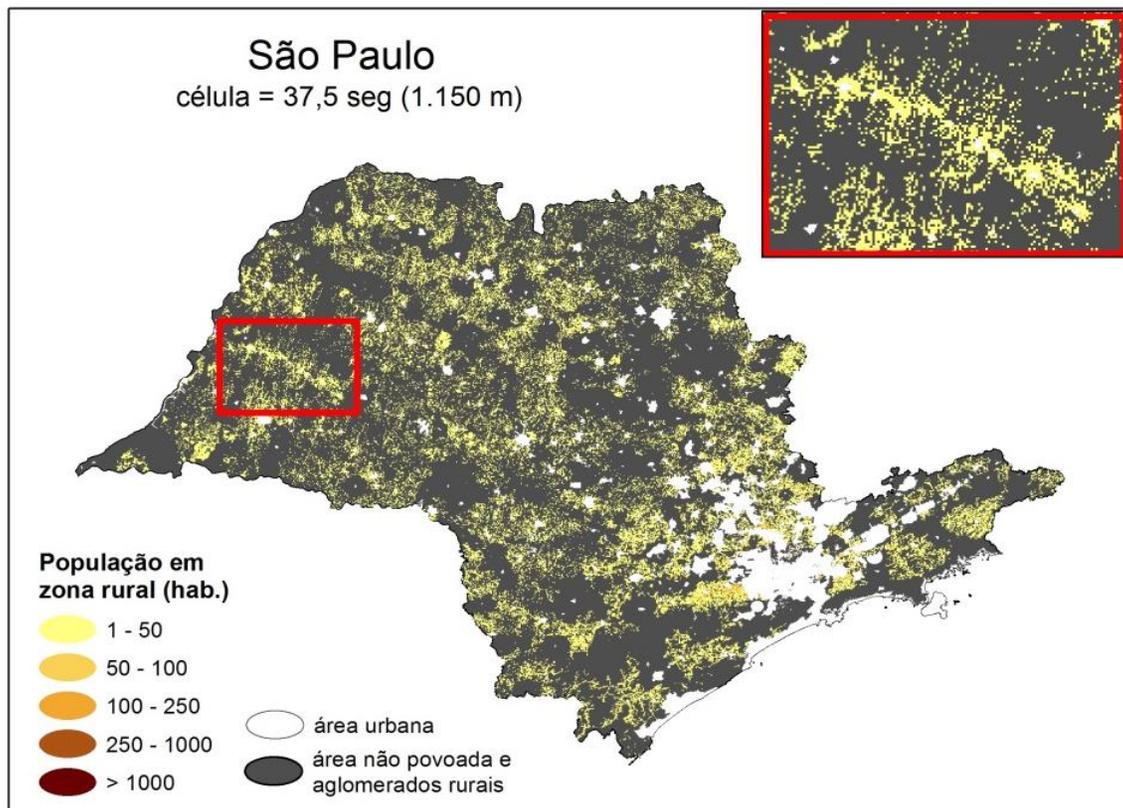
Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Figura 7 - População em zona rural com células de 75", São Paulo, 2010.



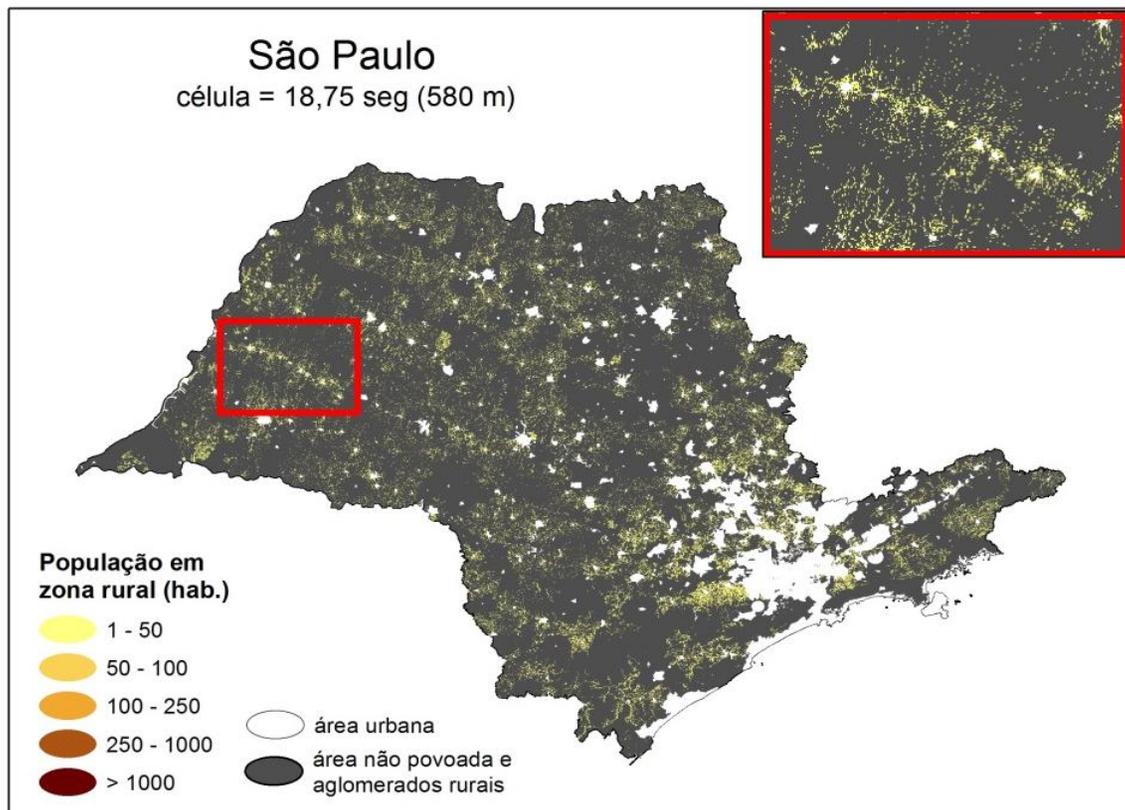
Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Figura 8 - População em zona rural com células de 37,5", São Paulo, 2010.



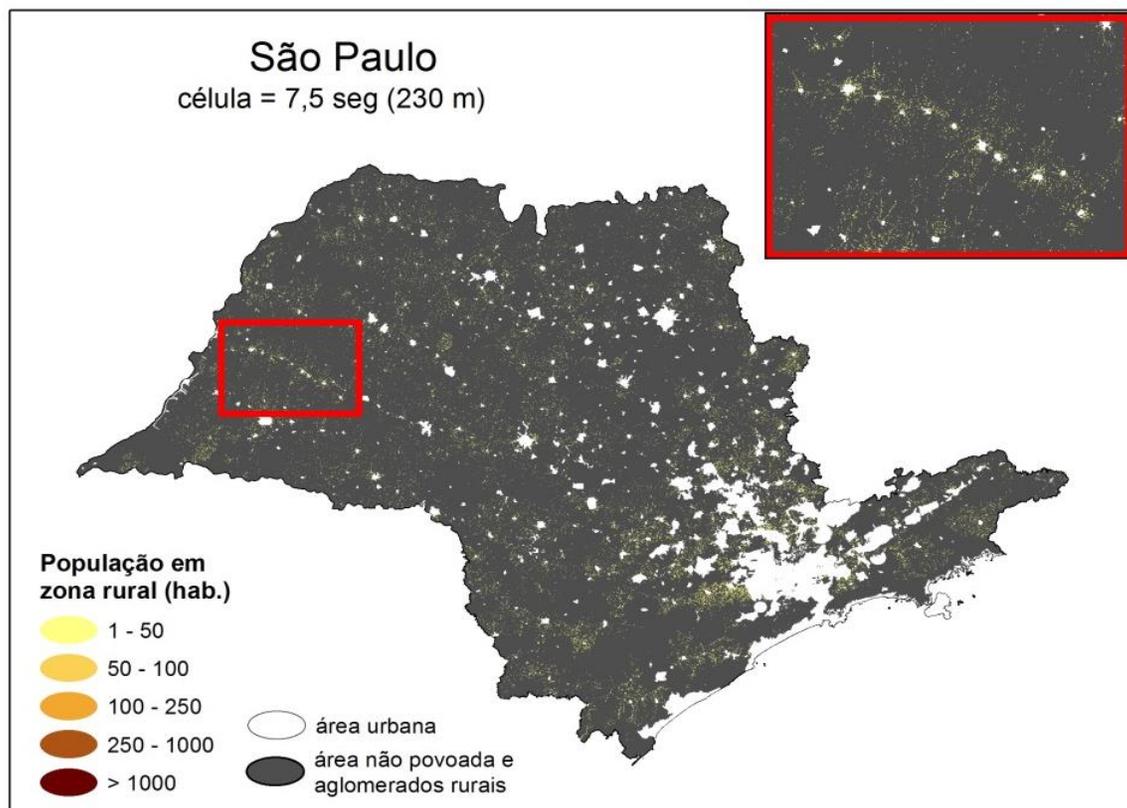
Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Figura 9 - População em zona rural com células de 18,75", São Paulo, 2010.



Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Figura 10 - População em zona rural com células de 7,5'', São Paulo, 2010.



Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

### 3 Resolução espacial

A Tabela 1 mostra algumas métricas relacionadas à área dos setores censitários correspondentes às zonas rurais dessas duas Unidades da Federação, incluindo a resolução espacial média. Observando os dados apresentados na Tabela 20 é possível verificar que o estado do Pará apresenta uma variação muito grande na área dos seus setores censitários rurais, fato que pode ser comprovado pelo valor do desvio padrão. Além disso, o Pará possui setores censitários rurais extremamente extensos, de acordo com o valor máximo apresentado na tabela. Com relação à resolução espacial média podemos observar que ela se apresenta bem menor para o estado de São Paulo do que para o estado do Pará. Observando o Gráfico 1 nós podemos ver que os dois estados apresentam uma

predominância de setores censitários em zonas rurais com grandes extensões, embora isso seja bem mais expressivo no estado do Pará.

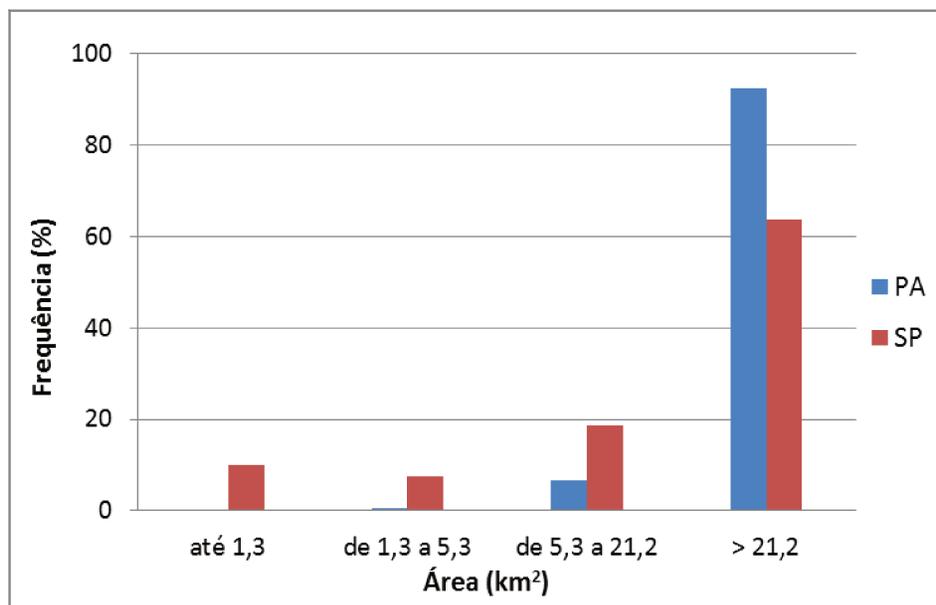
Tabela 1 - Métricas para setores censitários de zonas rurais, Pará e São Paulo, 2010.

Setores Censitários de zonas rurais (situação 8)	Pará	São Paulo
Quantidade (num)	2 612	5 068
Área máxima (km <sup>2</sup> )	62 622,913	399,367
Área mínima (km <sup>2</sup> )	0,053	0,000447
Área média (km <sup>2</sup> )	417,689	44,358
Soma da área (km <sup>2</sup> )	1 091 003,790	224 806,742
Desvio Padrão da área	1 622,874	45,086
Resolução espacial média	20,437	6,660

Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Observando a Tabela 1 e o Gráfico 1 e comparando com os valores das áreas das células em estudo (0,0529, 0,3364, 1,3225, 5,29 e 21,16 km<sup>2</sup>) podemos dizer que para o Pará a adoção de células com quaisquer das dimensões analisadas representa um ganho em termos de resolução espacial, pois mais de 90% dos setores têm área maior do que a área da maior célula analisada (21,2 km<sup>2</sup>). Para São Paulo, cerca de 35% dos setores censitários de zonas rurais apresentam área menor do que 21,2 km<sup>2</sup>, o que significa que para esses setores haveria uma perda na granularidade espacial caso se adotassem células com esta dimensão. No caso da adoção de células com área menor do que 5,3 km<sup>2</sup>, aproximadamente 18% dos setores de zonas rurais de São Paulo perderiam em termos de resolução espacial; já com a adoção de células de até 1,3 km<sup>2</sup> de área haveria uma perda de resolução espacial em cerca de 10% dos setores de zonas rurais de São Paulo.

Gráfico 1 - Frequência de áreas dos setores censitários em zonas rurais, Pará e São Paulo, 2010.



Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

#### 4 Confidencialidade

A regra mais comumente utilizada para que os dados possam ser disponibilizados ao público com a devida segurança da manutenção do sigilo estatístico é a supressão de informações consideradas sensíveis utilizando um valor limite. No caso da disseminação de dados censitários em setores censitários o valor adotado é de cinco domicílios particulares permanentes. Na prática, quando existem poucos domicílios particulares permanentes numa determinada unidade de agregação de dados, apenas as características básicas daquela unidade são disponibilizadas ao público, como população total e número de domicílios. As demais características não são divulgadas, pois podem levar a uma identificação indireta do respondente, principalmente quando estas características são combinadas entre si. Neste estudo foram considerados os valores de três, cinco e dez domicílios particulares permanentes nas análises relacionadas com a

determinação da menor dimensão de células tendo em vista a questão da confidencialidade.

Como há uma quantidade significativa de dados sem localização espacial nas regiões consideradas para esta avaliação de confidencialidade e isso poderia enviesar as análises, foram selecionadas oito áreas (quatro em cada estado) onde a localização espacial dos dados é total, não havendo perda de dados por ausência de espacialização.

As regiões de estudo foram avaliadas com relação a diferentes dimensões das células e diferentes valores para o limite de confidencialidade, e os resultados podem ser vistos na Tabela 2. Observando os resultados obtidos podemos constatar que o aumento do tamanho da célula não conduz necessariamente a um aumento da quantidade relativa de células liberadas para divulgação. Este fato é evidente em quase todas as áreas de estudo, qualquer que seja o limite utilizado como critério de confidencialidade. Uma possível explicação para este comportamento está nos arranjos de distribuição da população no território e na densidade populacional, uma vez que ambos influenciam as quantidades avaliadas. Quando a densidade populacional é baixa, a tendência é que a quantidade de células liberadas tenda a se estabilizar para dimensões maiores de células; quando a densidade populacional é alta, a tendência é que a quantidade de células liberadas diminua até certo ponto, inflexione e volte a crescer. A quantidade de células vazias é igualmente influenciada pelas características de ocupação do território, uma vez que os dados mostram que este valor cai mais abruptamente quando a densidade populacional é maior. Esses efeitos podem ser vistos nos Gráficos 2, 3 e 4, para critérios de confidencialidade de três, cinco e dez domicílios particulares permanentes, respectivamente.

Com relação à quantidade de população que seria possível de ser liberada para o público sem restrições nas suas características, podemos ver ao observarmos os gráficos mostrados nos Gráficos 5, 6, e 7 que existe um ponto em que as curvas de população liberada e população não liberada se cruzam. A parte à direita deste ponto, onde há uma quantidade maior de população liberada do

que não liberada, se configura como sendo a região onde devem estar localizadas as dimensões ideais da célula. Quando o limite de confidencialidade é de dez domicílios, praticamente os tamanhos de célula avaliados não atendem e seria necessário ter células maiores do que as avaliadas neste estudo. Para cinco domicílios como limite do critério de confidencialidade, a maioria dos gráficos aponta para células com dimensão adequada entre 30'' e 60''. Finalmente, o limite de três domicílios conduz a células de 30'' como sendo a dimensão mais adequada.

A partir dos resultados obtidos para as oito áreas de estudo localizadas nas duas regiões selecionadas, a opção que parece ser a mais adequada, ou seja, que conjuga a menor dimensão de célula com a maior quantidade de informações liberadas sem restrição, é utilizar células de dimensões próximas de 30'' (cerca de 1 km) com um limite de três domicílios particulares permanentes como critério de confidencialidade.

## **5 Conclusão**

Como o objetivo desta tese é gerar uma grade estatística que seja representativa para todo o país, a situação ótima seria realizar esta análise para todo o território de modo a poder avaliar melhor o impacto da dimensão da célula na quantidade de dados liberados ao público e a questão da resolução espacial da célula em relação aos setores censitários. No entanto, devido às dimensões continentais do nosso país, esta tarefa seria trabalhosa e demorada por demais, o que nos levou a utilizar apenas duas regiões para as análises, mas as mesmas apresentando características bastante distintas de modo que os resultados pudessem ser considerados válidos para todo o país.

Avaliando a representação espacial, a melhor opção seria utilizar células com dimensões menores do que 4 km, pois com células deste tamanho já não é possível identificar com clareza padrões espaciais de ocupação. Considerando a resolução espacial a melhor opção seria por utilizar células

apresentando 1 ou 2 km de lado, pois com estas dimensões um percentual de setores censitários inferior a 15,2% no caso de São Paulo e 0,6% no caso do Pará apresentaria perda de granularidade. Acrescentando a questão da confidencialidade, novamente a melhor opção seria utilizar células de 1 km de lado, pois apesar de haver uma grande quantidade de dados passíveis de restrição à divulgação, o aumento das dimensões das células não oferece uma melhora significativa. Reforçando o objetivo de adequação da grade estatística a recortes espaciais diversos, a escolha final é por células de 1 km de lado, pois quanto menor a célula, melhor será esta adequação. A quantidade de domicílios particulares permanentes utilizada como valor limite para liberação de características da população e dos domicílios ao público que se mostrou mais adequada à dimensão de célula selecionada é três. Embora não exista um padrão a ser seguido, este valor é o que mais se aproxima dos valores utilizados com mais frequência pelos países europeus (10 indivíduos).

Tabela 2 - Quantidade relativa de células e volume de população de acordo com a dimensão da célula e o valor limite de domicílios particulares permanentes, zonas rurais de Pará e São Paulo, 2010. (continua)

dimensão da célula		valor limite de domicílios particulares permanentes		3		5		10	
		vazia	liberada	não liberada	liberada	não liberada	liberada	não liberada	
Pará – área 1									
7,5"	célula	99,95	99,95	0,05	99,95	0,05	99,95	0,05	
	população		40,49	59,51	33,94	66,06	26,48	73,52	
18,75"	célula	99,82	99,84	0,16	99,83	0,17	99,83	0,17	
	população		45,61	54,39	36,88	63,12	31,76	68,24	
37,5"	célula	99,43	99,53	0,47	99,46	0,54	99,45	0,55	
	população		57,71	42,29	41,93	58,07	34,04	65,96	
75"	célula	98,28	98,84	1,16	98,56	1,44	98,38	1,62	
	população		72,28	27,72	56,30	43,70	37,71	62,29	
2,5'	célula	95,24	97,32	2,68	96,62	3,38	95,88	4,12	
	população		83,68	16,32	72,56	27,44	52,69	47,31	
Pará – área 2									
7,5"	célula	99,87	99,88	0,12	99,87	0,13	99,87	0,13	
	população		42,88	57,12	30,26	69,74	23,43	76,57	
18,75"	célula	99,54	99,60	0,40	99,55	0,45	99,54	0,46	
	população		52,04	47,96	36,39	63,61	29,67	70,33	
37,5"	célula	98,45	98,77	1,23	98,54	1,46	98,48	1,52	
	população		60,21	39,79	41,61	58,39	32,68	67,32	
75"	célula	95,51	97,12	2,88	96,20	3,80	95,67	4,33	
	população		76,90	23,10	58,55	41,45	38,34	61,66	
2,5'	célula	88,44	93,98	6,02	91,93	8,07	89,69	10,31	
	população		86,16	13,84	75,34	24,66	53,55	46,45	

Tabela 2 - Quantidade relativa de células e volume de população de acordo com a dimensão da célula e o valor limite de domicílios particulares permanentes, zonas rurais de Pará e São Paulo, 2010. (continua).

dimensão da célula		valor limite de domicílios particulares permanentes		3		5		10	
		vazia	liberada	não liberada	liberada	não liberada	liberada	liberada	não liberada
Pará – área 3									
7,5"	célula	98,61	98,87	1,13	98,72	1,28	98,65	1,35	
	população		57,86	42,14	43,05	56,95	28,99	71,01	
18,75"	célula	95,62	96,87	3,13	96,20	3,80	95,82	4,18	
	população		70,00	30,00	54,31	45,69	37,25	62,75	
37,5"	célula	87,73	93,14	6,86	90,52	9,48	88,73	11,27	
	população		84,10	15,90	69,63	30,37	51,14	48,86	
75"	célula	71,21	89,81	10,19	84,43	15,57	76,74	23,26	
	população		94,11	5,89	86,47	13,53	66,84	33,16	
2,5'	célula	49,63	90,11	9,89	85,82	14,18	77,05	22,95	
	população		98,72	1,28	97,24	2,76	91,48	8,52	
Pará – área 4									
7,5"	célula	98,84	99,02	0,98	98,91	1,09	98,85	1,15	
	população		45,15	54,85	26,49	73,51	4,62	95,38	
18,75"	célula	96,06	96,93	3,07	96,40	3,60	96,08	3,92	
	população		53,98	46,02	33,07	66,93	4,64	95,36	
37,5"	célula	87,94	92,07	7,93	89,85	10,15	88,18	11,82	
	população		68,12	31,88	46,84	53,16	12,91	87,09	
75"	célula	68,56	85,19	14,81	77,70	22,30	70,84	29,16	
	população		83,96	16,04	64,54	35,46	29,11	70,89	
2,5'	célula	42,30	86,23	13,77	77,70	22,30	59,67	40,33	
	população		96,25	3,75	90,65	9,35	67,27	32,73	

Tabela 2 - Quantidade relativa de células e volume de população de acordo com a dimensão da célula e o valor limite de domicílios particulares permanentes, zonas rurais de Pará e São Paulo, 2010. (continua)

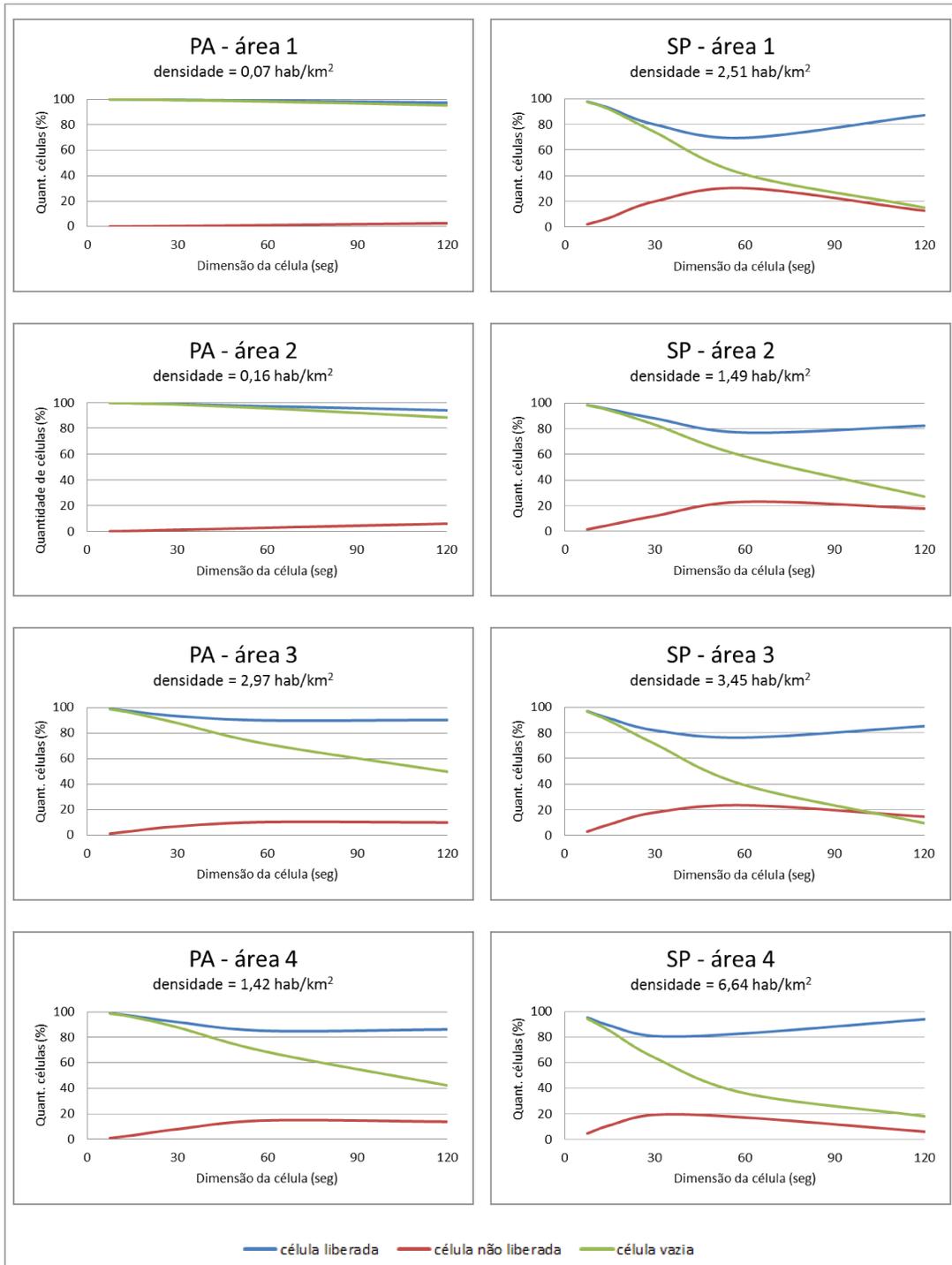
		valor limite de domicílios particulares permanentes		3		5		10	
dimensão da célula		vazia	liberada	não liberada	liberada	não liberada	liberada	não liberada	
São Paulo – área 1									
7,5"	célula	97,56	97,74	2,26	97,60	2,40	97,57	2,43	
	população		28,43	71,57	17,15	82,85	11,11	88,89	
18,75"	célula	91,59	92,63	7,37	91,84	8,16	91,65	8,35	
	população		38,63	61,37	22,11	77,89	13,17	86,83	
37,5"	célula	74,10	79,97	20,03	75,75	24,25	74,36	25,64	
	população		53,96	46,04	30,78	69,22	15,39	84,61	
75"	célula	41,14	69,57	30,43	55,02	44,98	43,90	56,10	
	população		79,80	20,20	57,01	42,99	26,63	73,37	
2'	célula	15,24	87,20	12,80	72,26	27,74	43,60	56,40	
	população		99,97	0,03	99,90	0,10	99,62	0,38	
São Paulo – área 2									
7,5"	célula	98,30	98,43	1,57	98,32	1,68	98,30	1,70	
	população		22,73	77,27	8,94	91,06	6,09	93,91	
18,75"	célula	94,13	94,98	5,02	94,24	5,76	94,16	5,84	
	população		34,64	65,36	10,60	89,40	6,17	93,83	
37,5"	célula	83,12	88,06	11,94	84,72	15,28	83,28	16,72	
	população		58,57	41,43	28,88	71,12	8,10	91,90	
75"	célula	58,47	76,94	23,06	68,71	31,29	61,06	38,94	
	população		79,24	20,76	59,10	40,90	26,62	73,38	
2,5'	célula	26,69	82,20	17,80	66,95	33,05	44,07	55,93	
	população		95,83	4,17	86,97	13,03	59,64	40,36	

Tabela 2 - Quantidade relativa de células e volume de população de acordo com a dimensão da célula e o valor limite de domicílios particulares permanentes, zonas rurais de Pará e São Paulo, 2010. (conclusão)

dimensão da célula		valor limite de domicílios particulares permanentes		3		5		10	
		vazia	liberada	não liberada	liberada	não liberada	liberada	não liberada	
São Paulo – área 3									
7,5"	célula	96,58	96,84	3,16	96,61	3,39	96,58	3,42	
	população		18,83	81,17	4,05	95,95	0,84	99,16	
18,75"	célula	89,22	91,20	8,80	89,60	10,40	89,26	10,74	
	população		38,01	61,99	11,95	88,05	2,53	97,47	
37,5"	célula	71,51	82,00	18,00	76,11	23,89	72,28	27,72	
	população		65,37	34,63	41,26	58,74	11,69	88,31	
1'	célula	39,33	76,34	23,66	63,13	36,87	46,73	53,27	
	população		86,13	13,87	72,51	27,49	39,47	60,53	
2'	célula	9,69	85,20	14,80	77,04	22,96	52,04	47,96	
	população		95,73	4,27	93,03	6,97	78,71	21,29	
São Paulo – área 4									
7,5"	célula	94,37	95,34	4,66	94,74	5,26	94,42	5,58	
	população		47,95	52,05	29,23	70,77	11,21	88,79	
18,75"	célula	84,84	89,00	11,00	86,92	13,08	85,27	14,73	
	população		65,48	34,52	49,55	50,45	21,57	78,43	
37,5"	célula	63,84	80,78	19,22	72,66	27,34	67,55	32,45	
	população		82,07	17,93	66,13	33,87	43,37	56,63	
1'	célula	36,08	82,91	17,09	67,72	32,28	51,90	48,10	
	população		95,23	4,77	87,36	12,64	69,92	30,08	
2'	célula	18,00	94,00	6,00	88,00	12,00	70,00	30,00	
	população		99,21	0,79	98,19	1,81	91,98	8,02	

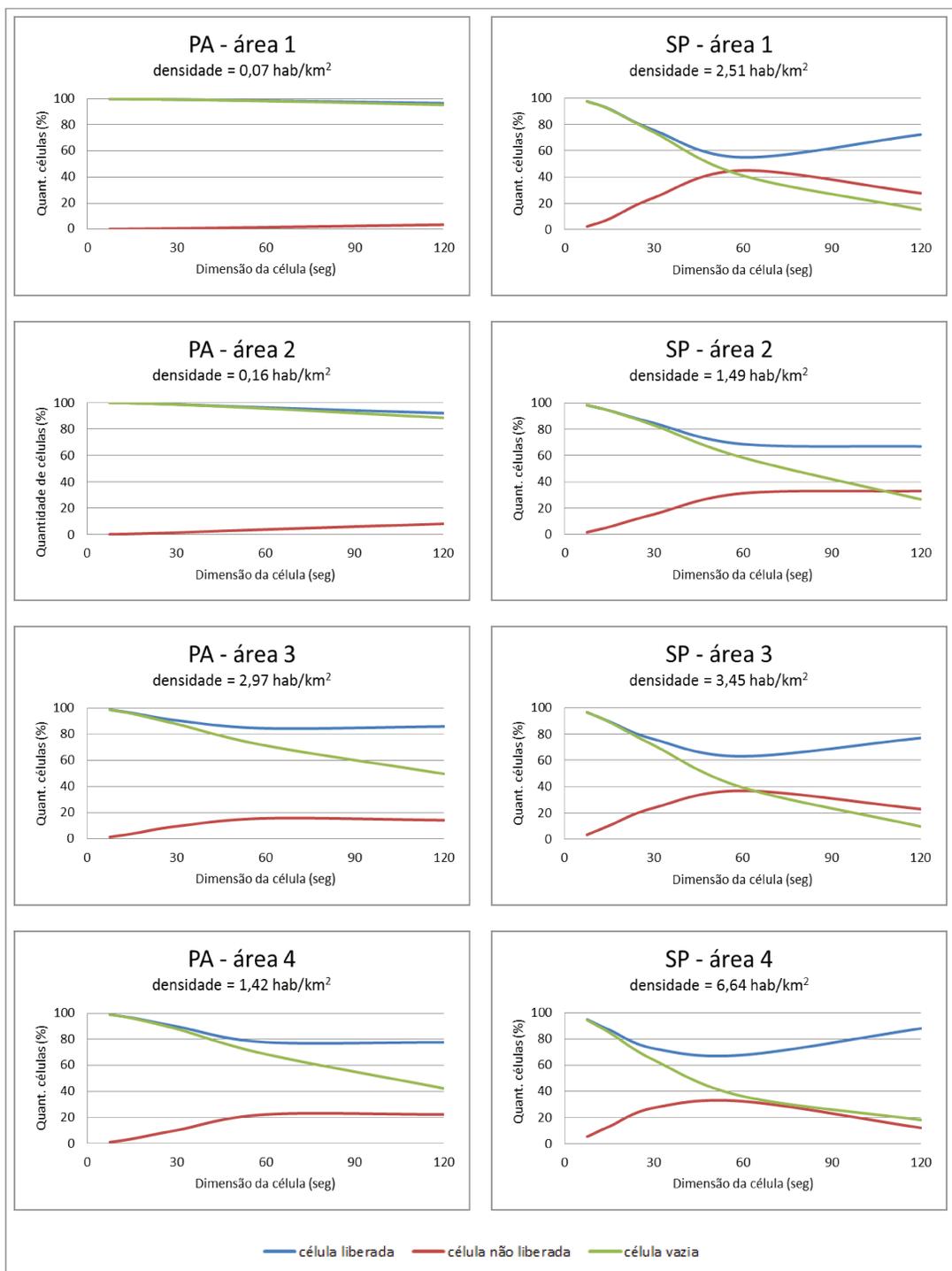
Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Gráfico 2 - Quantidade de células vazias, com dados liberados e não liberados, de acordo com a dimensão da célula para um limite de confidencialidade de três domicílios particulares permanentes, zonas rurais de Pará e São Paulo, 2010.



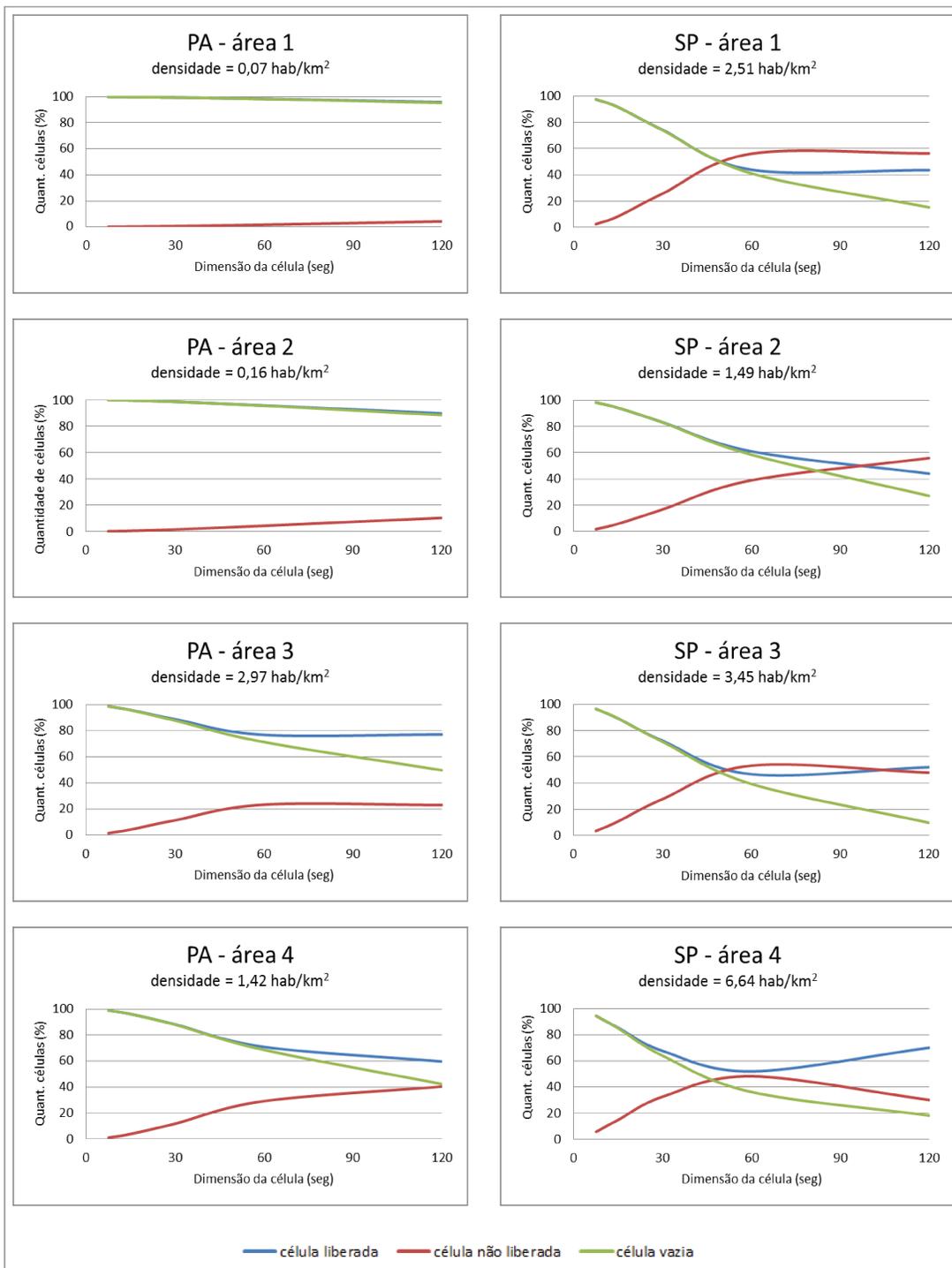
Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Gráfico 3 - Quantidade de células vazias, com dados liberados e não liberados, de acordo com a dimensão da célula para um limite de confidencialidade de cinco domicílios particulares permanentes, zonas rurais de Pará e São Paulo, 2010.



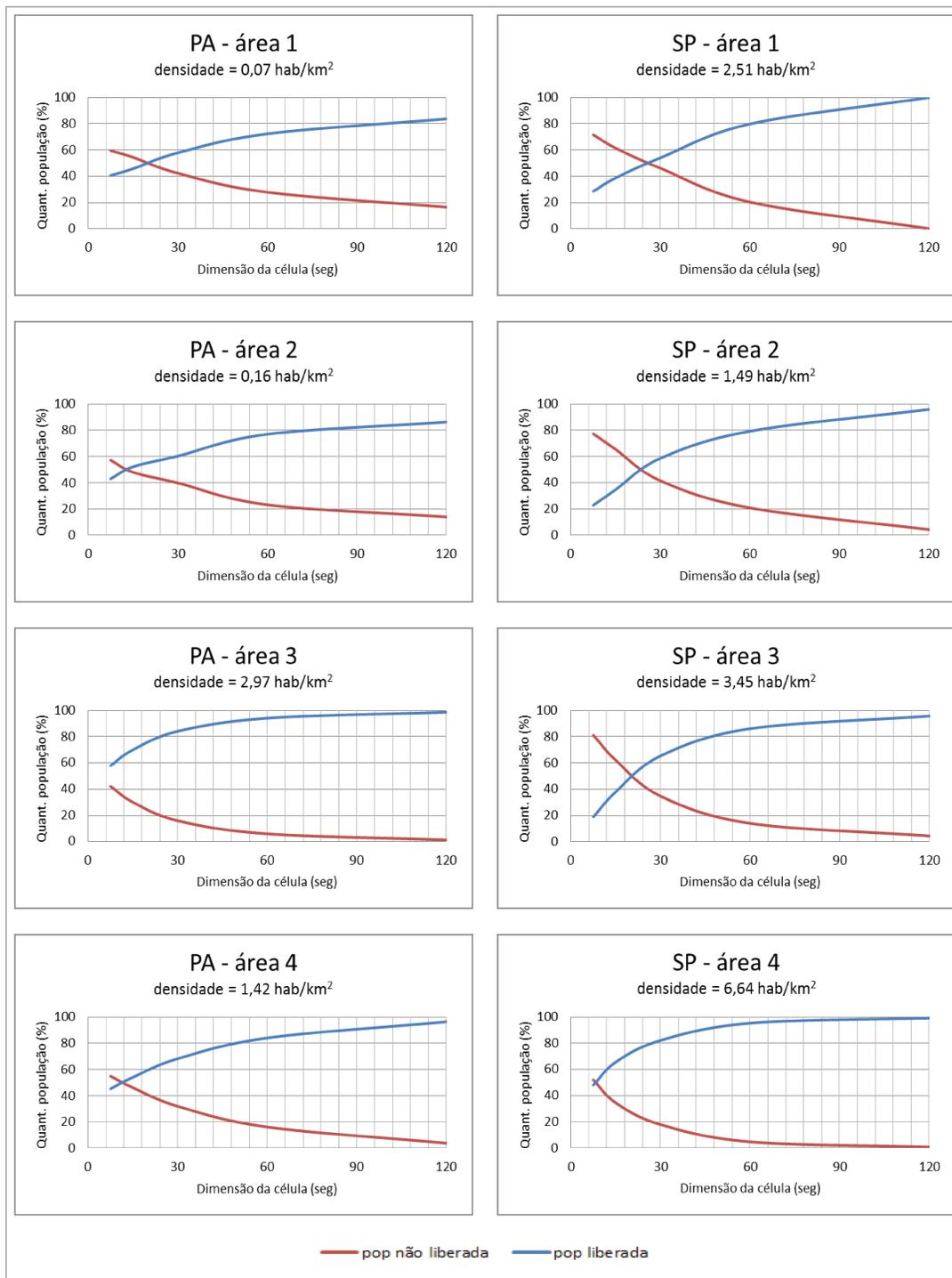
Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Gráfico 4 - Quantidade de células vazias, com dados liberados e não liberados, de acordo com a dimensão da célula para um limite de confidencialidade de dez domicílios particulares permanentes, zonas rurais de Pará e São Paulo, 2010.



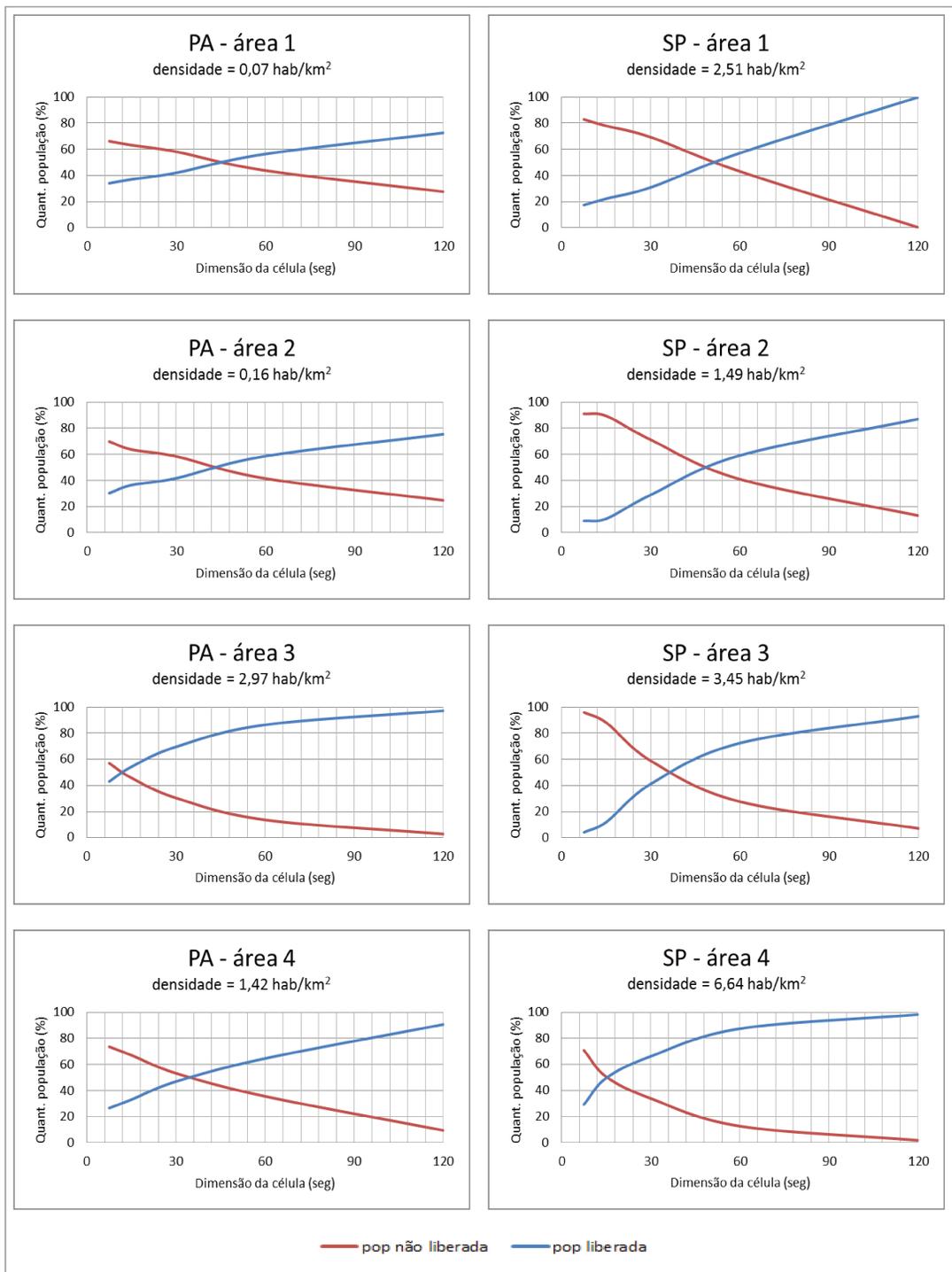
Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Gráfico 5 - População em células com dados liberados e não liberados de acordo com o tamanho da célula para um limite de confidencialidade de três domicílios particulares permanentes, zonas rurais de Pará e São Paulo, 2010.



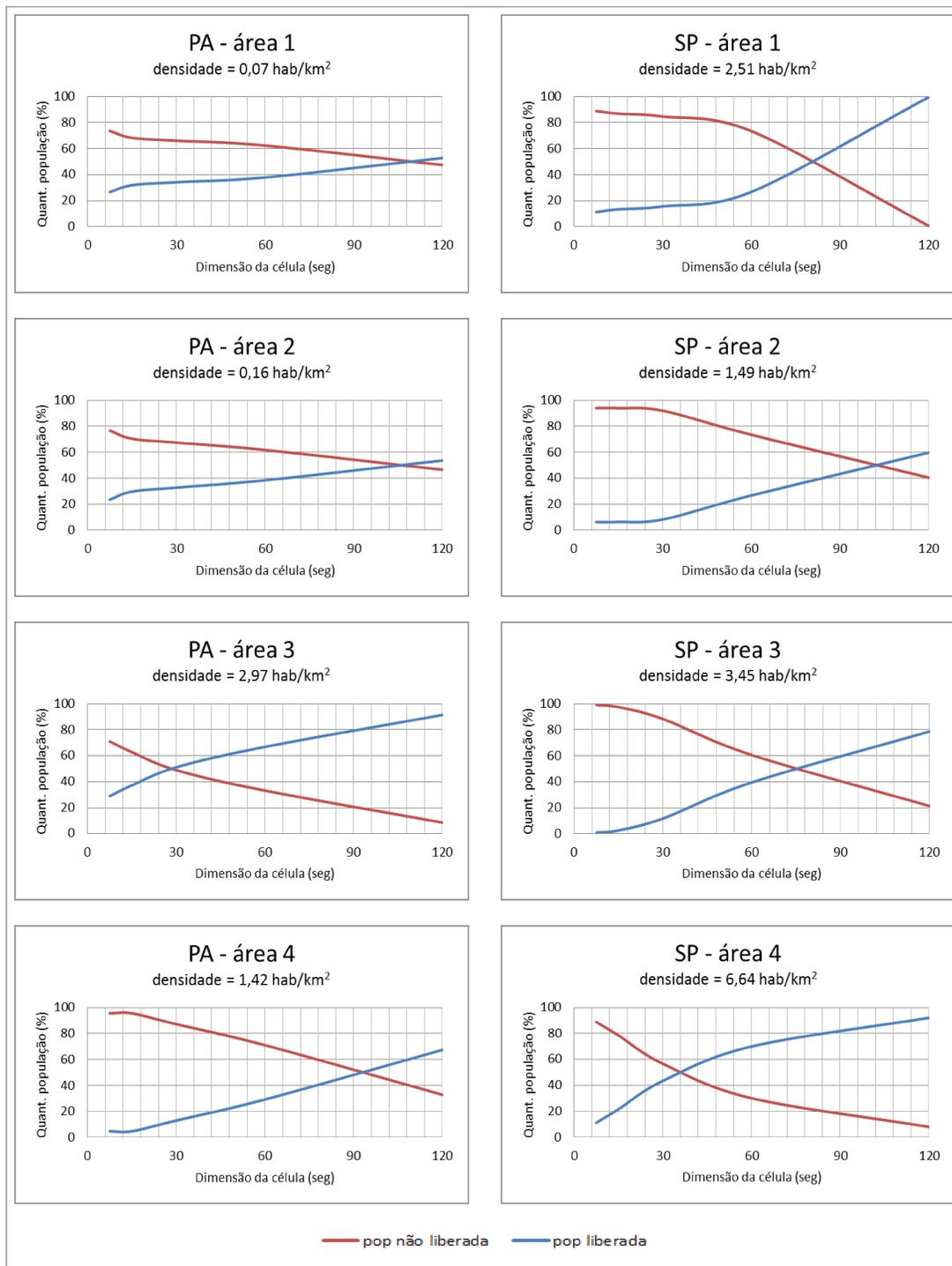
Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Gráfico 6 - População em células com dados liberados e não liberados de acordo com o tamanho da célula para um limite de confidencialidade de cinco domicílios particulares permanentes, zonas rurais de Pará e São Paulo, 2010.



Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.

Gráfico 7 - População em células com dados liberados e não liberados de acordo com o tamanho da célula para um limite de confidencialidade de dez domicílios particulares permanentes, zonas rurais de Pará e São Paulo, 2010.



Fonte: IBGE, Censo Demográfico 2010. Elaborado pelo autor.



## **APÊNDICE B - Avaliação de métodos de desagregação**

B1 – The Brazilian Population Grid: a hybrid approach.

Trabalho apresentado no European Forum for Geostatistics – Sofia Conference, realizado em Sofia, Bulgária, de 23 a 25 de Outubro de 2013.

Autores: Maria do Carmo Dias Bueno, David Martin e Álvaro de Oliveira D'Antona.



## The Brazilian Population Grid: a hybrid approach

**Maria do Carmo BUENO**

*IBGE, Brazil, mbueno@ibge.gov.br*

**David MARTIN**

*University of Southampton, UK*

**Alvaro D'ANTONA**

*University of Campinas, Brazil*

### ABSTRACT

There is a growing interest in the generation of statistical population grids due to their stability through time and ease of integration with different spatial data sources. The conventional means of producing these estimates may be divided into bottom-up (aggregation) and top-down (disaggregation) approaches, depending on the national data environment. This paper introduces a hybrid model proposed for creation of a population grid for Brazil by the national statistical agency using data from Census 2010. It has been necessary to develop this novel hybrid methodology due to the diverse data environments found in urban and rural settings. Two regions in the state of Para were selected as study areas to evaluate several methods in order to find the most suitable combination. Results of both the aggregation and disaggregation methods are available for the study region, making possible comparison of the results obtained using different methods. The analysis suggests that each set of conditions requires a different treatment and highlights the importance of good metadata. The insights gained from this analysis have potential application in many countries facing similar data challenges.

**KEYWORDS** population grid, Brazil, aggregation, disaggregation

### INTRODUCTION

A Census produces essential information for national, state and municipal policy-making, including service planning (educational, health, social and utilities), emergency services related to natural disasters and numerous analyses (poverty, labour force, marketing, epidemiological). The growth of geospatial technologies has led to much wider use of this statistical information, increasing the demand for detailed and geographically disaggregated information. Although these types of data are traditionally produced for



irregular geographical units, there are many benefits to the production of data on a regular geographical grid. The latter offers particular advantages of stability over time and relative ease of integration with spatial data from other sources.

In Nordic countries the population grid is a regular product offered by statistical agencies for example, since 1970 in Finland and 1980 in Sweden. This is only possible due to the underlying point-based statistical system, permitting aggregation from a georeferenced building-code system to the cells of the grid (UN 2007). This approach to generation of grid-based data is known as the aggregation approach. In many other countries where this approach is not adopted – most often due to the absence of a suitably detailed georeferencing basis, researchers need to use some alternative spatial and/or statistical method to reallocate census data from irregular units into a population grid, termed the disaggregation approach. There are a variety of disaggregation methods, some of which use ancillary data. Examples without ancillary data include areal weighting, pycnophylactic interpolation (Tobler 1979) and kernel estimation (Bracken and Martin 1989). Examples using ancillary data include dasymetric mapping utilising land use classification derived from remotely sensed images (Eicher and Brewer 2001; Mennis 2003; Holt et al. 2004; Langford 2006), a network vector layer (Reibel and Bufalino 2005) or point addresses (Zandbergen 2011).

This paper addresses the challenge of producing a population grid for Brazil, in a context in which the data available to the national statistical agency varies greatly between urban and rural areas. It is proposed that generation of the population grid requires development of a novel hybrid model which combines both aggregation and disaggregation approaches. The following section describes the Brazilian context, proposed methods and the study area. The third section presents the study results and evaluation of the performance of the different methods. The final section presents conclusions and recommendations regarding population grid generation in this challenging context.

## **METHODS AND DATA**

The Brazilian 2010 Census had two significant advances that deserve mention: integrated digital census mapping covering both urban and rural areas, and an address list combined with use of handheld computers with GPS. The first advance has allowed the georeferencing of dwellings in urban areas via addresses, and the second has allowed the capture of point locations of buildings in rural areas via GIS (IBGE 2010). These two technological advances permit for the first time aggregation of census data into grid cells. However, analysis of an initial sample revealed a significant number of enumeration areas with missing locational data, preventing direct aggregation. In urban areas and rural conglomerates (villages and small settlements) the spatial location is based on street block face codification. This presents two potential sources of missing locational data: there may



be no street network map, or the network may be missing block face codes. In the state of Para we have around 10% of enumeration areas that are possibly in this situation. In rural areas with a sparse settlement pattern, enumerators captured the building GPS points at the time of enumeration, but these may be missing due to operational and technical failures. In the state of Para around 3% of enumeration areas has locational data missing whilst about 3% has incomplete locational data. Under these circumstances it is not possible to simply apply an aggregation approach to the generation of a national population grid. Rather, it will be necessary to develop a hybrid approach which combines both aggregation and disaggregation approaches according to the local data context. It remains, however, to determine the most appropriate disaggregation method to be used in these circumstances.

Two study areas have been chosen in the state of Para in northern Brazil. The results of both aggregation and disaggregation are available for these areas, permitting a comparative analysis. The two areas have similar features, with a large rural portion and a small urban zone. Area 1 encompasses part of the municipality of Santarem (300,000 inhabitants) and Area 2 encompasses part of the municipality of Altamira (100,000 inhabitants). The rural part of both study areas is a mix of forest and agro-pasture. Area 1 presents a settlement pattern strongly related to the road network, while in Area 2 settlement pattern is more diverse and sparse. The urban part of Area 1 is more densely populated than that of Area 2.

Aggregated microdata from Census 2010 is here used as reference data the aggregation method differs between urban and rural areas. In rural areas, the grid cell result is the simple summation of the population count at each GPS point inside each grid cell. In urban areas, the block face is the smallest geographical unit and a linear weighting method is used to reallocate the population count from each block face into grid cells. For the grid cells that are partially urban and partially rural both results are summed. From now on this combined method will be referred as aggregation (AGG). The grids used here are based on a geographic projection with approximately square cells with sides around 1 km in rural areas and 250 m in urban areas. Four disaggregation methods are evaluated, each based on population count by enumeration area from Census 2010.

1. WEIGHT. Areal weighting based on 2010 Census data. It assumes that the distribution is homogenous within source (enumeration area) and target (grid cell) areas.

2. IMAGE. Dasymetric method using binary land use classification derived from 2009-10 Landsat 30m image data. Land use class "impervious surface" has been considered as populated and classes related to vegetation and water as unpopulated. Some known non-residential impervious features (e.g. airports) have been deleted from the information layer.

3. STREET. Dasymetric method using edited 1:5000 road network from 2010 IBGE Census Mapping. It is only available in urban areas.



4. POINT. Dasymetric method using 2007 IBGE Population Count residential building points. It is only available in rural areas.

The first evaluation assesses the populated and unpopulated cells correctly and incorrectly identified by disaggregation, compared to aggregation. These are termed omission and commission errors and are tabulated for each study area and disaggregation method..

The second evaluation concerns the population values estimated by each method. Linear regression is employed and selected goodness of fit statistics reported, although there is not space for these to be fully tabulated here.

The third evaluation is related to the difference between population estimated by the disaggregation models and the population count resulting from AGG. Cell values were grouped into classes and then the difference was computed. The formula used is:

$$D = (POP_{model} - POP_{AGG}) \div (POP_{model} + POP_{AGG})$$

POP<sub>AGG</sub> is the population value obtained from the aggregation method and POP<sub>model</sub> is the population value estimated by the disaggregation models. This measure has been chosen because it is able to describe the direction as well as the magnitude of the error. Negative values occur when predictions are smaller than observations.

## RESULTS AND EVALUATION

Results are considered separately for urban and rural areas due to the different methods available in each context. As noted above, some areas are missing the information required for the aggregation approach and these are therefore excluded from the analysis. The total number of urban cells in the analysis is 1,347 in Area 1, and 542 in Area 2; the total number of rural cells is 4,424 in Area 1 and 2,656 in Area 2.

The tabulation of omission and commission errors for urban and rural areas respectively is presented in Tables 1 and 2. Omission errors correspond to cells not recognized as belonging to a class whilst commission errors are related to the incorrect identification of the class. Map accuracy concerns the probability that the classification is correct. In relation to urban areas (Table 1), all the methods perform better in more densely populated areas. The areal weighting presents the worst results and performs particularly poorly in Area 2, as it is not able to identify unpopulated places. The two dasymetric methods have similar overall accuracy in both areas, but the IMAGE model is less accurate in identifying unpopulated places in Area 1. Inspection of the mapped results (not shown here) suggests that this may relate to poorer performance in urban areas with plenty of open spaces and lower population density, but with a significant built street network.



Table 1 – Omission and commission errors (%) and map accuracy (%) in urban areas

		WEIGHT			IMAGE			STREET		
		Om	Com	Acc	Om	Com	Acc	Om	Com	Acc
AREA 1	Populated	0.00	28.86	100.00	3.34	13.61	96.66	0.21	8.16	99.79
	Unpopulated	100.00	0.00	0.00	37.63	11.68	62.37	21.91	0.66	78.09
	Overall Accuracy			<b>71.14</b>			<b>86.79</b>			<b>93.54</b>
AREA 2	Populated	0.00	54.80	100.00	11.43	21.66	88.57	4.90	17.67	95.10
	Unpopulated	100.00	0.00	0.00	20.20	10.57	79.80	16.84	4.63	83.16
	Overall Accuracy			<b>45.20</b>			<b>83.76</b>			<b>88.56</b>

In rural areas (Table 2) the WEIGHT model performs poorly in identifying populated places, and therefore has a low overall mapping accuracy. This is due to the very low population density and scattered pattern of human settlements in rural areas. The other methods produced good overall results, but IMAGE shows a map accuracy less than in urban areas. The POINT model shows a very low accuracy in identifying populated places in Area 1 due to the poor quality of the point layer, previously noted. All dasymetric methods identify unpopulated places better than populated places – reflecting the far greater number of unpopulated cells.

Table 2 – Omission and commission errors (%) and map accuracy (%) in rural areas

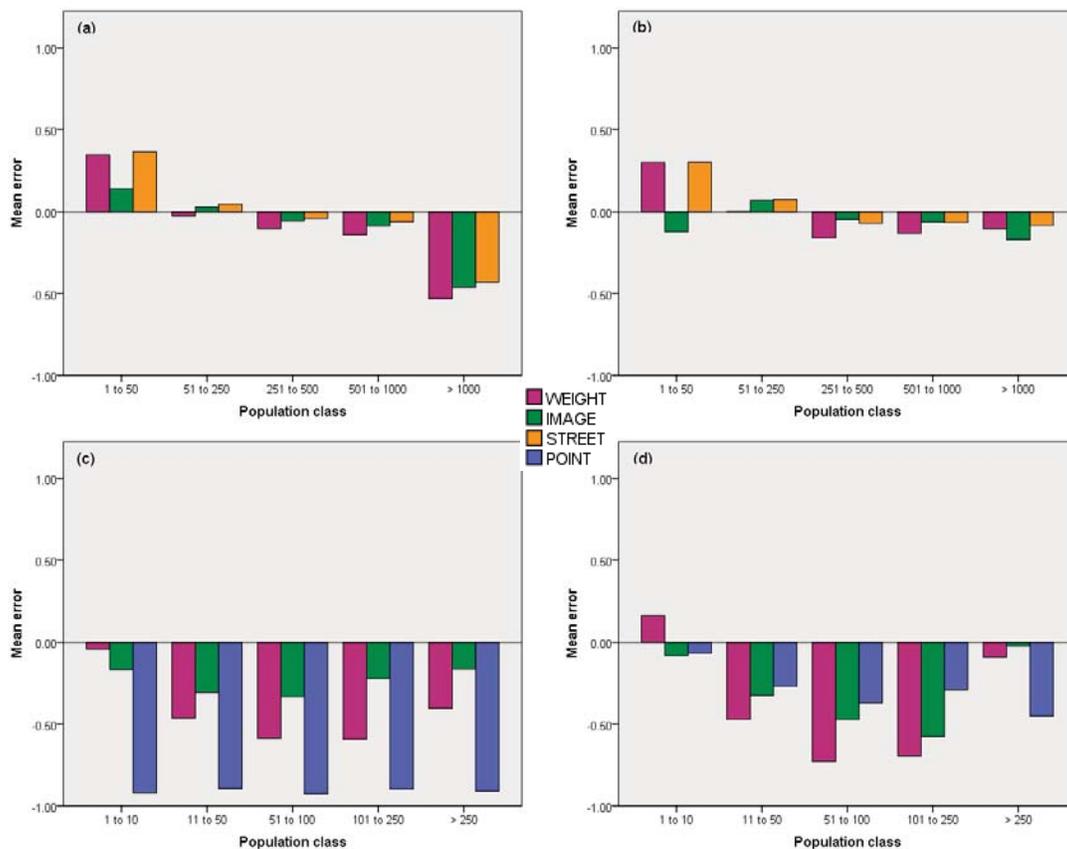
		WEIGHT			IMAGE			POINT		
		Om	Com	Acc	Om	Com	Acc	Om	Com	Acc
AREA 1	Populated	0.00	82.62	100.00	22.76	56.23	77.24	92.85	36.05	7.15
	Unpopulated	100.00	0.00	0.00	20.88	5.71	79.12	0.85	16.46	99.15
	Overall Accuracy			<b>17.38</b>			<b>78.80</b>			<b>83.16</b>
AREA 2	Populated	0.00	70.07	100.00	37.99	38.22	62.01	34.47	25.25	65.53
	Unpopulated	100.00	0.00	0.00	16.39	16.25	83.61	9.46	13.99	90.54
	Overall Accuracy			<b>29.93</b>			<b>77.15</b>			<b>83.06</b>

Turning to the regression models for urban areas, there are not great differences between the two study areas or methods analysed.  $R^2$  measures the proportion of the variability in the dependent variable explained by regression. In Area 1, the WEIGHT model explains 85.90% of the variance of the values of population count, IMAGE explains 90.80%, and STREET explains 94.10%. In Area 2, the WEIGHT model explains 86.00%, IMAGE explains 92.30%, and STREET explains 94.50%. The models can be ordered by ascending accuracy: WEIGHT, IMAGE, and STREET. All F ratios are statistically significant at the 0.01 level. We speculate that the strong performance of the STREET model might be due to a circularity effect, as both aggregation and disaggregation methods in urban areas use



the network vector as georeferencing layer and ancillary data respectively. However, the F ratio values for STREET are notably greater than for IMAGE and WEIGHT.

Figure 1 - Mean error difference between observed and estimated populations



Turning to rural areas, the  $R^2$  are much lower. For the POINT model in Area 1 this reflects known poor data quality.  $R^2$  values for WEIGHT in Area 1 are 38.5% and for POINT in Areas 1 and 2 are 1.5% and 36.3% respectively. All other models explain more than 60% of the variability in the AGG population values. An analysis of the F ratios suggests that IMAGE is the best model in Areas 1 and 2, although in Area 2 the WEIGHT model has also a good fit. There are potentially errors in the POINT model due to the distribution of 2010 population count on a 2007 point layer.

Figure 1(a) and (b) show that overall in urban areas the models underestimate the population count when population is greater than 250 and overestimate it when it is lower than this. An exception is population class 1 to 50, where the IMAGE model overestimates



population in Area 1 and underestimates in Area 2. In general we can say that all the methods underestimate population in rural areas (Figures 1(c) and 1(d)). The models can be ordered by descending errors: WEIGHT, IMAGE, and STREET. The first population class (1 to 10) in Area 2 has small positive errors for WEIGHT and small negative errors for IMAGE and POINT, which probably reflects the very large number of truly vacant cells. The POINT method has an atypical behaviour in Area 1 due to the missing data. The last population class (> 250) in rural areas refers to very small numbers of cells.

## CONCLUSION

The analysis presented here clearly shows that disaggregation methods can perform well in places where aggregation is not possible. However, no one method is best suited for use in all contexts. The choice will need to be determined by the characteristics of the application region, data availability and quality and the purpose of the analysis. With regard to the latter, it is important to consider whether the location (presence/absence) of population or the overall accuracy of the estimated counts is most important. For a national statistical agency, it is important to adopt a strategy that meets the analytical needs of many different users. Different models are likely to perform better in urban and rural areas but, more particularly, model performance is sensitive to density and settlement pattern. The best performance is achieved in dense urban areas and dasymetric methods consistently perform better than simple areal weighting. The choice of dasymetric method needs to take account of completeness, date, resolution/scale, format and availability of ancillary data. The quality of the metadata available on potential ancillary data sources can be critical in helping to inform these decisions. Further the output population grid should contain as much metadata as possible to inform the user about the method used and the uncertainties involved. We conclude that where countries face internal diversity in collected census data, a hybrid approach presents a viable means of generating a national population grid but that further research is needed on the optimal way of performing the choice of dasymetric disaggregation method based on the ancillary data available.

## ACKNOWLEDGEMENTS

This study was performed as part of PhD research by Maria do Carmo Dias Bueno, funded by Instituto Brasileiro de Geografia e Estatística and by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (process number 17235-12-0). 2010 Census microdata and mapping were provided by IBGE solely for use in this research.



## REFERENCES

- Bracken I, Martin D (1989). The Generation of Spatial Population Distributions from Census Centroid Data Source. *Environment and Planning A* 21(4): 537-543.
- Eicher C L, Brewer C A (2001). Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation. *Cartography and Geographic Information Science* 28(2): 125-138.
- Holt, J B, Lo, C P, Hodler, T W (2004). Dasymetric Estimation of Population Density and Areal Interpolation of Census Data. *Cartography and Geographic Information Science* 31(2): 103-121.
- IBGE – Brazilian Institute of Geography and Statistics (2010). 2010 Census – Summary of Survey Steps.  
<[http://censo2010.ibge.gov.br/images/pdf/censo2010/sintese/sintese\\_censo2010\\_ingles.pdf](http://censo2010.ibge.gov.br/images/pdf/censo2010/sintese/sintese_censo2010_ingles.pdf)>. Accessed 25 July 2013.
- Langford, M (2006). Obtaining population estimates in non-census reporting zones: An evaluation of the 3-class dasymetric method. *Computers, Environment and Urban Systems* 30: 161–180
- Mennis, J (2003). Generating Surface Models of Population Using Dasymetric Mapping. *Professional Geographer* 55(1): 31-42.
- Reibel, M, Bufalino, M E (2005). A test of street weighted areal interpolation using geographic information systems. *Environment and Planning A* 37: 127–139.
- Tobler, W R (1979). Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association* 74: 519-530.
- UNITED NATIONS, United Nations Economic Commission for Europe (2007). Register-based statistics in the Nordic countries - Review of best practices with focus on population and social statistics. New York and Geneva: United Nations.  
<[http://www.unece.org/fileadmin/DAM/stats/publications/Register\\_based\\_statistics\\_in\\_Nordic\\_countries.pdf](http://www.unece.org/fileadmin/DAM/stats/publications/Register_based_statistics_in_Nordic_countries.pdf)>. Accessed 25 July 2013.
- Zandbergen, P A (2011). Dasymetric Mapping Using High Resolution Address Point Datasets. *Transactions in GIS* 15(s1): 5–27.



B2 – Comparação entre métodos para distribuição espacial da população.

Trabalho publicado na Revista Espinhaço (n. 3, vol. 1, 2014).

Autores: Maria do Carmo Dias Bueno e Álvaro de Oliveira D'Antona.

# Avaliação de métodos de desagregação para geração de grades de população

Maria do Carmo Dias Bueno<sup>1\*</sup>  
Álvaro de Oliveira D'Antona<sup>2</sup>

<sup>1</sup> Mestre em Engenharia Civil pela Universidade Federal do Rio de Janeiro (UFRJ), Mestre em Geomática pela UFRJ, Doutoranda em Demografia pelo Instituto de Filosofia e Ciências Humanas (UNICAMPI).

<sup>2</sup> Mestre em Antropologia pela Universidade Estadual de Campinas (UNICAMPI), Doutor em Ciências Sociais pela UNICAMPI, Doutor em População e Ambiente pela Indiana University, Estados Unidos.

---

**Resumo** O objetivo deste trabalho é avaliar métodos de desagregação de dados para gerar grades de população, utilizando como referência uma grade gerada através da agregação de microdados censitários, o que permite obter resultados mais precisos acerca do desempenho desses métodos. Para realizar a avaliação foram selecionadas três metodologias. A primeira utiliza uma matriz contendo os erros e acertos na identificação de áreas povoadas e não povoadas, permitindo avaliar a acurácia da distribuição espacial da população. A segunda utiliza uma regressão linear cujos coeficientes permitem avaliar a qualidade da adequação aos dados de referência. Finalmente, utilizamos uma fórmula para calcular a diferença entre os valores de população, indicando se há uma subestimação ou superestimação. Os resultados sugerem que a escolha da metodologia mais adequada depende dos objetivos a que se destina a grade de população, bem como da disponibilidade e qualidade dos dados auxiliares, além das características da área de estudo.

**Palavras-chave:** distribuição da população; grades de população; agregação; desagregação.

---

## 1. Introdução

Os censos de População e Habitação produzem informações essenciais para a criação e monitoração de políticas públicas nos níveis nacional, estadual e municipal, e também para a tomada de decisão relacionada com investimentos. Os dados obtidos com o censo são utilizados para o planejamento de serviços para a população, como educação, saúde, serviços sociais e serviços públicos (água, esgoto, energia, gás), para o planejamento de serviços de emergência (desastres naturais) e diversos estudos e análises (mapeamento da pobreza, estudos epidemiológicos, análise de mercados). Esses dados também são utilizados para fornecer referências para as projeções populacionais que são utilizadas pelo governo federal na definição das cotas do Fundo de Participação dos Estados e dos Municípios (IBGE, 2010).

Um dos benefícios das geotecnologias aplicadas à área das estatísticas oficiais é a função de integração de dados oferecida pelos Sistemas de Informação Geográfica – SIGs, facilitando o relacionamento de informações oriundas de diferentes áreas. Isso levou, sem sombra de dúvidas, a uma utilização mais abrangente das informações estatísticas e também a uma maior exigência, por parte dos usuários, de dados geograficamente mais desagregados.

Existem dois problemas relacionados com a utilização de dados agregados em unidades geográficas que são bastante conhecidos e citados na literatura. O primeiro deles diz respeito a não coincidência entre as unidades para as quais

os dados censitários são disponibilizados e as unidades para as quais os dados são necessários. Essas últimas podem ser unidades baseadas em características sociais, ambientais ou mesmo unidades artificiais geradas em meio digital, como áreas de influência ou *buffers*. Referências sobre esta questão podem ser vistas, por exemplo, em Clarke e Rhind (1992), Hogan (1992), Deichmann (1996), Ojima e Martine (2012) e Martine e Schensul (2013). O segundo problema está relacionado à estabilidade temporal dessas unidades geográficas, uma vez que tanto as unidades político-administrativas como as unidades operacionais estatísticas sofrem alterações ao longo do tempo, fato este que dificulta a realização de análises temporais (EXETER et al., 2005; NORMAN, REES; BOYLE, 2003).

A utilização de unidades geográficas baseadas em células regulares dispostas num sistema de grade parece ser a solução para esses dois problemas, pois as células têm pequenas dimensões podendo ser consideradas como tijolos que se juntam para formar qualquer recorte espacial desejado além de não se alterarem ao longo do tempo, já que são unidades artificiais e arbitrárias e, uma vez determinadas, não necessitam sofrer alterações (MARTIN, 2000; TAMMILEHTO-LUODE et al., 2000; RUSANEM et al., 2001; TAMMILEHTO-LUODE, 2011).

Existem duas abordagens metodológicas principais para a geração de dados de população em grades: a abordagem *bottom-up* ou de agregação e a *top-down* ou de desagregação. A primeira abordagem utiliza os microdados censitários associados aos seus respectivos atributos de localização para a agregação em células. A geração de

---

\*bueno.mariadocarmo@gmail.com

grades de população utilizando metodologias que tem por base esta abordagem é realizada desde meados da década de 90 em países do norte europeu, como Finlândia e Suécia (TAMMILEHTO-LUODE; BACKER; ROGSTAT, 2010). Isso se tornou possível porque esses países desenvolveram um sistema estatístico com base em pontos codificados que representam as edificações, o que permite o relacionamento entre as diversas bases de registros que são utilizadas para a geração das estatísticas oficiais (NAÇÕES UNIDAS, 2007). Esta metodologia também tem sido utilizada mais recentemente por países que realizam o censo na sua forma tradicional de coleta de dados em campo e que utilizam geotecnologias para a localização espacial dos dados coletados, como, por exemplo, a Estônia (EFGS, 2012).

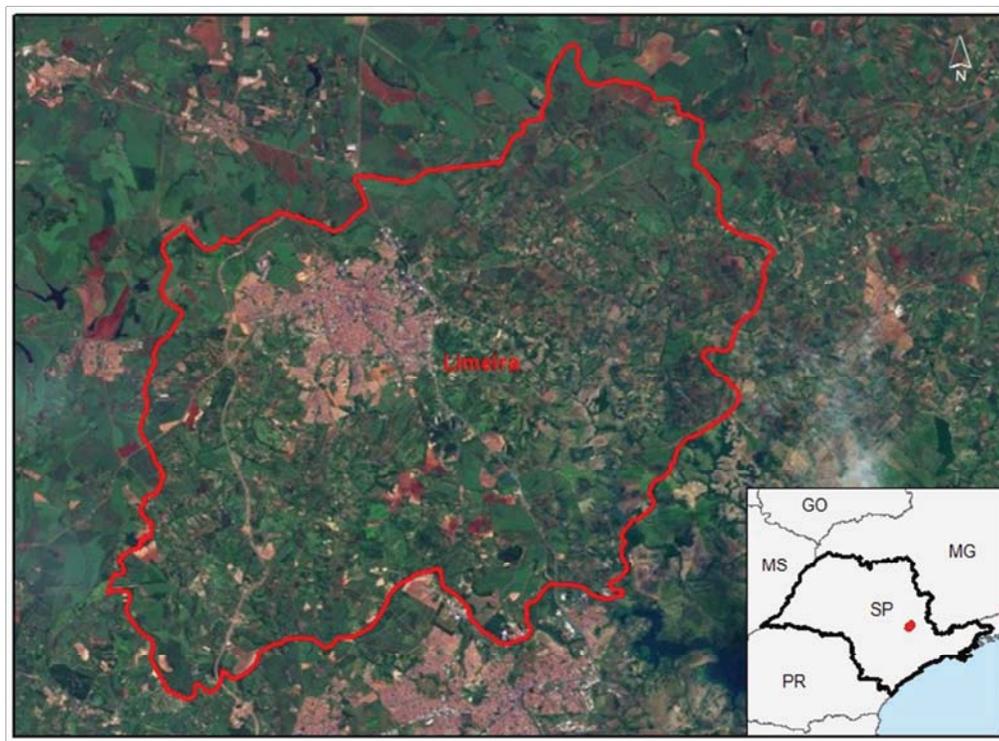
Nos países onde a abordagem de agregação não é adotada pela agência nacional de estatística, os pesquisadores utilizam métodos espaciais e/ou estatísticos para realocar os dados disponibilizados em unidades irregulares para células regulares com o objetivo de enfrentar os problemas citados anteriormente. Esta abordagem é conhecida como desagregação e envolve diferentes metodologias, algumas com a utilização de dados auxiliares e outras não. Entre os métodos que não utilizam dados auxiliares podemos citar a ponderação zonal, a interpolação picnofilática (TOBLER, 1979) e a estimação *kernel* (BRACKEN; MARTIN, 1989). Entre os métodos que usam dados auxiliares, o mapeamento dasimétrico é o mais conhecido, sendo que diferentes dados podem auxiliar na estimativa de população, como classificação de uso e cobertura das terras derivada de imagens de satélite (EICHER; BREWER, 2001; MENNIS, 2003), malha de sistema viário (XIE, 1995; REIBEL;

BUFALINO, 2005) ou endereços geocodificados (ZANDBERGEN, 2011).

A avaliação da acurácia da abordagem de desagregação geralmente é realizada utilizando unidades geográficas para onde existem dados censitários publicados, sendo poucos os casos em que é realizada uma comparação com grades de população desenvolvidas com a abordagem de agregação, já que esses dados são regularmente produzidos por poucas agências de estatística. Um exemplo deste tipo de avaliação pode ser visto em Steinnocher et al. (2011). O objetivo deste artigo é avaliar a qualidade de quatro métodos de desagregação em comparação com os resultados obtidos com a agregação de microdados censitários espacializados e agregados em células regulares. Os materiais e métodos utilizados no desenvolvimento do trabalho e a discussão dos resultados serão mostrados a seguir.

## 2. Materiais e Métodos

A FIG. 1 mostra a área de estudo selecionada para a análise das diferentes metodologias de espacialização da população. A área se localiza no estado de São Paulo, abrangendo totalmente o município de Limeira, tendo cerca de 600 km<sup>2</sup> e população de 276.022 habitantes, segundo o Censo Demográfico 2010 (IBGE, 2011b). O município de Limeira está inserido na área de interesse do Laboratório de Geografia de Riscos e Resiliência do Centro de Ciências Humanas e Sociais Aplicadas da Faculdade de Campinas e, por este motivo, tem sido objeto de diversos estudos e análises.



**Figura 1:** Localização da área de estudo.

Fonte: Malha Municipal Digital 2010, IBGE; ESRI Basemap, s/d.

O compromisso com a confidencialidade dos dados estatísticos, a necessidade de trabalhar com uma resolução espacial melhor do que a dos setores censitários, a adequação simultânea a estudos no nível nacional e a estudos locais e a comparabilidade internacional levaram a optar por utilizar células com dimensões de 1 x 1 km nas áreas rurais e 250 x 250 m nas áreas urbanas.

A abordagem de agregação foi realizada com base nos microdados do Censo Demográfico 2010 e do Cadastro Nacional de Endereços para Fins Estatísticos - CNEFE e é utilizada como referência para a avaliação dos demais métodos. Nas áreas rurais foram utilizadas diretamente as coordenadas das edificações coletadas pela pesquisa, enquanto nas áreas urbanas os endereços das unidades domiciliares visitadas passaram por um processo de geocodificação para serem transformados em pontos. Foi observado uma perda na localização espacial da ordem de 1,5% dos domicílios visitados pelo Censo 2010 no município de Limeira (N = 84.531) devido à falta de informações que permitisse a geocodificação nas áreas urbanas e à falta de coordenadas coletadas nas zonas rurais.

Todos os métodos de desagregação utilizam os dados do Censo Demográfico 2010 por setor censitário (IBGE, 2011a; 2011b) e o primeiro deles é a ponderação zonal. Este método é o mais simples entre os que se propõem a realocar a população entre duas unidades geográficas diferentes (PLUMEJEAUD et al., 2010). Neste caso, a unidade de origem é o setor censitário e a unidade de destino é a grade de células. O método pressupõe que a distribuição da população é homogênea na unidade geográfica de origem e de destino, de modo que a população da unidade de destino possa ser calculada com base na densidade populacional encontrada na unidade de origem.

O segundo método de desagregação é um método dasimétrico binário (LANGFORD; UNWIN, 1994) com utilização de uma classificação de uso das terras feita com base em imagens do satélite CBERS (pixel de 30 metros) do ano de 2010. A classe “superfície impermeável” foi considerada como sendo povoada enquanto as demais classes, relativas à vegetação e água, foram consideradas não povoadas.

O terceiro método de desagregação é também um método dasimétrico que utiliza o mesmo dado auxiliar do método anterior, sendo que a classe povoada foi reclassificada em alta densidade residencial, baixa densidade residencial e área não urbana habitada. A distribuição de população foi realizada com o auxílio de uma extensão para o programa ArcMap, denominada *Dasymetric Mapping Extension* – DME (SLEETER; GOULD, 2008).

O quarto método de desagregação também é um método dasimétrico, agora utilizando uma malha viária como dado auxiliar. Nas áreas urbanas foi utilizada a malha de ruas do mapeamento censitário do IBGE, com escala aproximada de 1:5.000, enquanto nas áreas rurais foi utilizada a malha viária do IGC, com escala de 1:50.000. Algumas edições foram necessárias nestas camadas vetoriais para eliminar feições que poderiam prejudicar o cálculo da densidade populacional por extensão de via, como canteiros centrais de avenidas e estradas, retornos e outras feições similares.

Foram selecionadas três metodologias para avaliação dos resultados. A primeira delas é uma técnica importada da área de Sensoriamento Remoto que faz uso de uma matriz

com os erros e acertos da classificação e tem por objetivo avaliar a acurácia com que o modelo de realocação de população identifica as regiões povoadas e não povoadas. A técnica é baseada em uma matriz contendo as quantidades de células classificadas de forma correta e incorreta com base em amostras de campo (dados de referência) e a medida final é expressa sob a forma de uma distribuição de percentuais de células classificadas corretamente (SMITS et al, 1999). Também é apresentada outra medida derivada desta mesma matriz – o Índice *Kappa* - que considera no seu cálculo todos os elementos da matriz, o que torna esta medida mais robusta do que a anterior. Utilizou-se uma classificação da qualidade do mapeamento com base neste índice proposta por Landis e Koch (1977).

A segunda metodologia utiliza uma regressão linear ajustada à origem que foi aplicada a todos os métodos de desagregação para avaliar o seu ajuste aos dados de referência. Para realizar a comparação entre os diferentes métodos foram utilizados gráficos de dispersão, coeficiente de correlação, coeficiente de determinação ( $R^2$ ) e taxa de adequação do modelo (*F ratio*). O gráfico de dispersão nos permite uma avaliação visual entre os valores estimados e os valores de referência, enquanto o coeficiente de correlação indica a força e a direção do relacionamento entre a população estimada e a de referência. O coeficiente de determinação nos dá a proporção da variância da população estimada com relação à população de referência, enquanto a taxa de adequação do modelo permite avaliar o quanto o modelo proporcionou uma melhora na estimativa da população em comparação com os níveis de incerteza do mesmo.

Finalmente, utilizou-se uma medida para avaliar o erro entre a população estimada pela desagregação e a população de referência. A fórmula utilizada para este cálculo é:

$$D = \frac{(POP_{\text{modelo}} - POP_{\text{REF}})}{(POP_{\text{modelo}} + POP_{\text{REF}})}$$

onde D é a diferença normalizada,  $POP_{\text{modelo}}$  é a população do método de desagregação empregado e  $POP_{\text{REF}}$  é a população de referência obtida com o método de agregação. Os dados de população foram agrupados em cinco classes (1 a 49, 50 a 199, 200 a 499, 500 a 999 e 1000 e mais) e a média da diferença normalizada para cada classe e para cada método de desagregação é apresentado em forma de gráfico. Esta medida foi selecionada por ser capaz de descrever a direção da diferença entre os valores de população, de modo que valores negativos e positivos ocorrem quando os valores estimados são, respectivamente, menores ou maiores do que os valores de referência, indicando uma subestimação ou uma superestimação. Valores próximos de zero para esta medida indicam que os valores do modelo são praticamente iguais aos valores de referência.

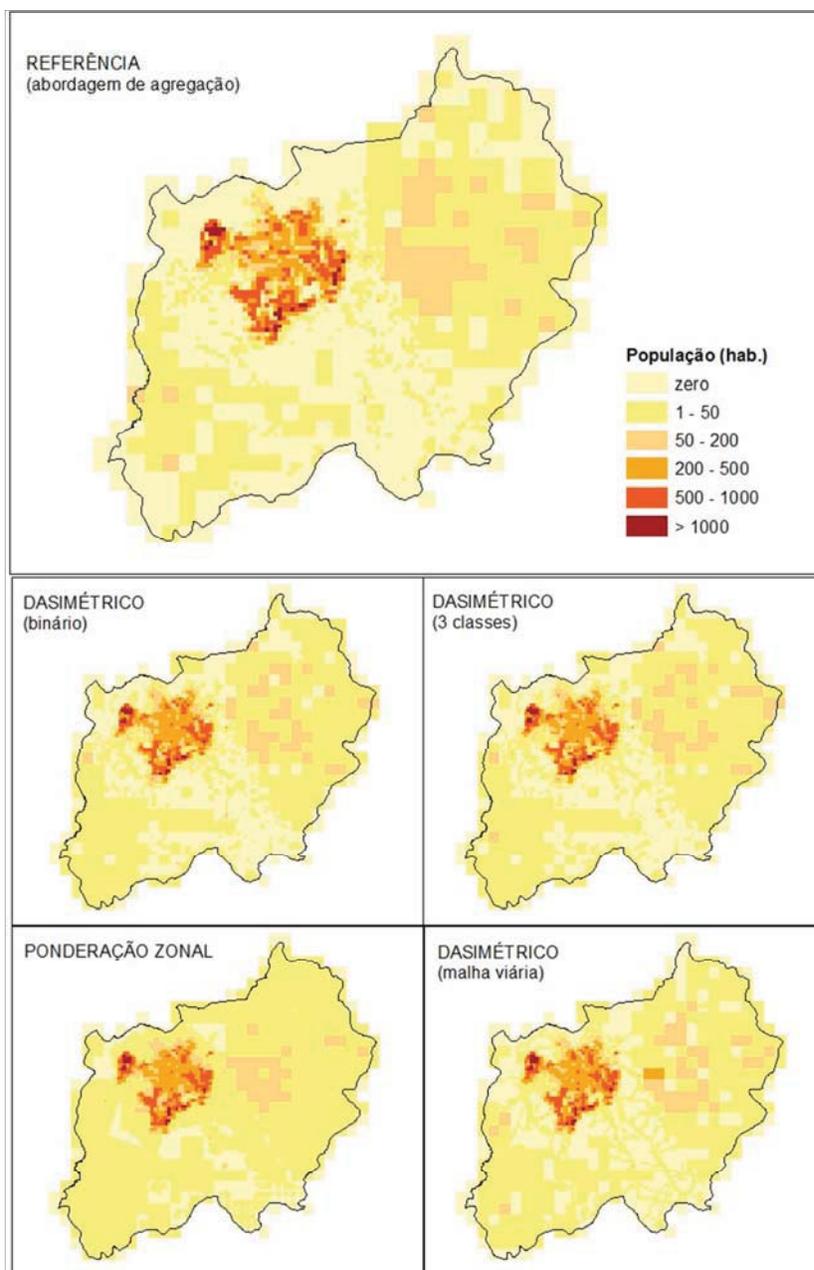
### 3. Resultados e Discussão

A FIG. 2 mostra sob a forma de mapas os resultados obtidos para a distribuição espacial da população utilizando os métodos avaliados e o método de referência. Os resultados são graficamente similares, com pequenas diferenças visíveis na área rural do município. Entretanto, para avaliar melhor as diferenças entre os métodos é

necessário utilizar medidas mais objetivas, conforme veremos a seguir.

A primeira avaliação está relacionada com a quantidade de células povoadas e não povoadas identificadas corretamente pelos métodos de desagregação em comparação com os dados de agregação que servem de referência para o estudo. Na TAB. 1 podemos observar os valores obtidos para a acurácia do mapeamento de cada classe (povoada e não povoada) e a acurácia geral, além do Índice *Kappa*. De acordo com os dados apresentados, podemos dizer que o método dasimétrico binário e o dasimétrico de 3 classes não apresentam diferenças significativas entre si, mostrando uma acurácia que vai de

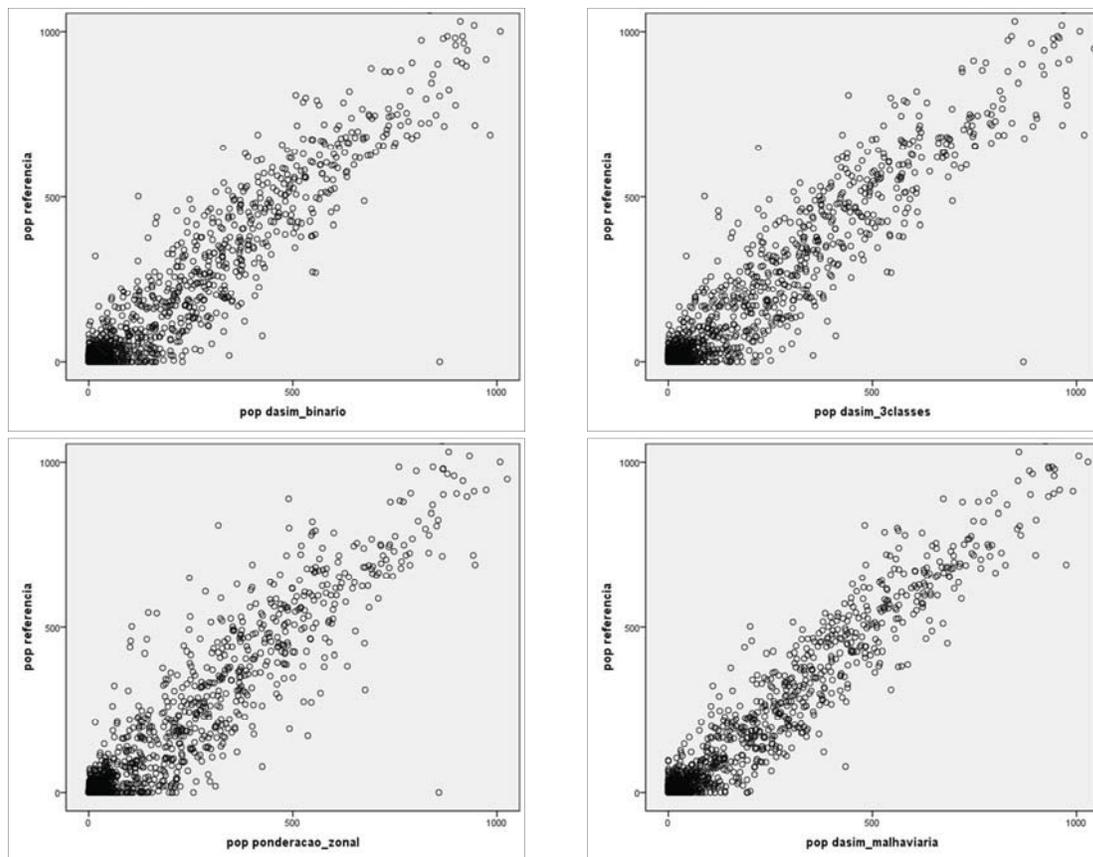
razoável à boa. O método de ponderação zonal revelou-se o pior método entre aqueles utilizados neste estudo, obtendo uma classificação de acurácia variando entre péssima e ruim. O método dasimétrico com a utilização de malha viária como dado auxiliar alcançou a melhor classificação, ficando entre razoável e boa. Todos os métodos dasimétricos apresentaram uma capacidade geral de classificação das classes povoada e não povoada superiores a 60%, sendo que este percentual é inferior nas áreas rurais, indicando que os dados auxiliares não conseguem representar de maneira eficiente a distribuição da população em áreas com ocupação dispersa no território.



**Figura 2:** Mapas de distribuição da população de acordo com o método.  
Fonte: Censo Demográfico 2010, IBGE.

Localização	Método de desagregação	Classe Povoada		Classe Não povoada		Geral	
		Acurácia	<i>Kappa</i>	Acurácia	<i>Kappa</i>	Acurácia	<i>Kappa</i>
Rural N = 338	Dasim. Binário	65,38	0,625	34,43	0,230	70,71	0,336 (razoável)
	Dasim. 3 classes	65,84	0,636	35,76	0,242	71,30	0,351 (razoável)
	Ponderação zonal	59,94	1,000	0,73	0,004	60,06	0,009 (ruim)
	Dasim. Malha viária	62,85	0,505	31,85	0,200	68,34	0,286 (razoável)
Urbano N = 3122	Dasim. Binário	51,25	0,644	57,83	0,320	70,79	0,423 (boa)
	Dasim. 3 classes	51,26	0,639	58,16	0,323	70,94	0,423 (boa)
	Ponderação zonal	36,85	0,791	0,61	0,002	37,09	0,004 (péssima)
	Dasim. Malha viária	52,81	0,693	58,62	0,332	71,72	0,449 (boa)
Geral N = 3460	Dasim. Binário	53,13	0,650	56,31	0,326	70,78	0,435 (boa)
	Dasim. 3 classes	53,19	0,654	56,70	0,322	70,98	0,432 (boa)
	Ponderação zonal	39,11	0,817	0,62	0,002	39,33	0,004 (péssima)
	Dasim. Malha viária	54,14	0,688	56,79	0,329	71,39	0,445 (boa)

**Tabela 1:** Acurácia do mapeamento de acordo com a localização e o método.  
Fonte: Elaboração Própria.



**Figura 3:** Gráficos de dispersão de acordo com o método.  
 Fonte: Censo Demográfico 2010, IBGE.

O método de ponderação zonal não tem habilidade para identificar áreas não povoadas, uma vez que ele pressupõe uma distribuição homogênea da população na região considerada e os valores baixos, mas não nulos, para a acurácia da classe “Não Povoadas” são explicados pela quantidade de casas decimais utilizadas para representar os valores de população (apenas duas). Especula-se que o valor relativamente alto da acurácia geral para áreas rurais deste método (60,06%) é devido ao número pequeno de células consideradas na análise e também às características locais, pois em estudos realizados para outras áreas isso não se confirma (BUENO; MARTIN; D’ANTONA, 2013).

A FIG. 3 mostra os gráficos de dispersão construídos com os dados obtidos da regressão linear para toda a área de estudo. Visualmente o melhor resultado parece ser o do método dasimétrico com a utilização de vias como dado auxiliar, pois os pontos se apresentam menos dispersos em relação a uma linha de tendência. Além disso, o gráfico desse modelo dasimétrico exibe uma menor quantidade de pontos ao longo do eixo X, que representam valores estimados de população nulos para valores reais de população não nulos. Com relação à quantidade de pontos

ao longo do eixo Y, que representam valores estimados de população não nulos para valores reais nulos, aparentemente a quantidade é semelhante em todos os quatro gráficos.

A TAB. 2 apresenta os principais coeficientes obtidos com uma regressão linear ajustada à origem aplicada aos quatro modelos de desagregação da população. O coeficiente de correlação indica a existência de uma correlação forte entre os dados analisados para a área urbana e média para a área rural, mas prevalecendo uma correlação forte quando analisamos o município integralmente. Esta conclusão se repete para o coeficiente de determinação, onde para a área urbana os modelos aplicados conseguem explicar mais de 90% da variância dos valores de população; para a área rural temos valores mais baixos, variando entre 46 e 71%. No caso da análise para toda a área de estudo, os valores permanecem altos, acima de 90%. Com relação ao *F ratio*, o método dasimétrico com malha viária se destaca em relação aos demais, indicando um melhor ajuste do modelo na área urbana e na área total. De uma maneira geral, podemos dizer que todos os métodos apresentam resultados satisfatórios para as áreas urbanas, mas não tão bons para as áreas rurais.

Localização	Modelo	Coef. correlação	Coef. determinação	Fratio **
Rural N = 338	Dasim. binário	0,539	0,463	290,553
	Dasim. 3 classes	0,546	0,468	296,370
	Ponderação zonal	0,767	0,712	833,010
	Dasim. Malha viária	0,656	0,579	463,869
Urbano N = 3 122	Dasim. binário	0,952	0,918	35 024,880
	Dasim. 3 classes	0,950	0,915	33 589,578
	Ponderação zonal	0,945	0,906	30 238,245
	Dasim. Malha viária	0,966	0,941	49 803,529
Geral N = 3 460	Dasim. binário	0,952	0,917	38 157,222
	Dasim. 3 classes	0,950	0,914	36 638,838
	Ponderação zonal	0,945	0,906	33 334,743
	Dasim. Malha viária	0,965	0,940	54 154,504

**Tabela 2:** Sumário da regressão linear\* de acordo com a localização e o método.  
\* Regressão linear ajustada à origem. \*\* Valores estatisticamente significantes (*p value* = 0).

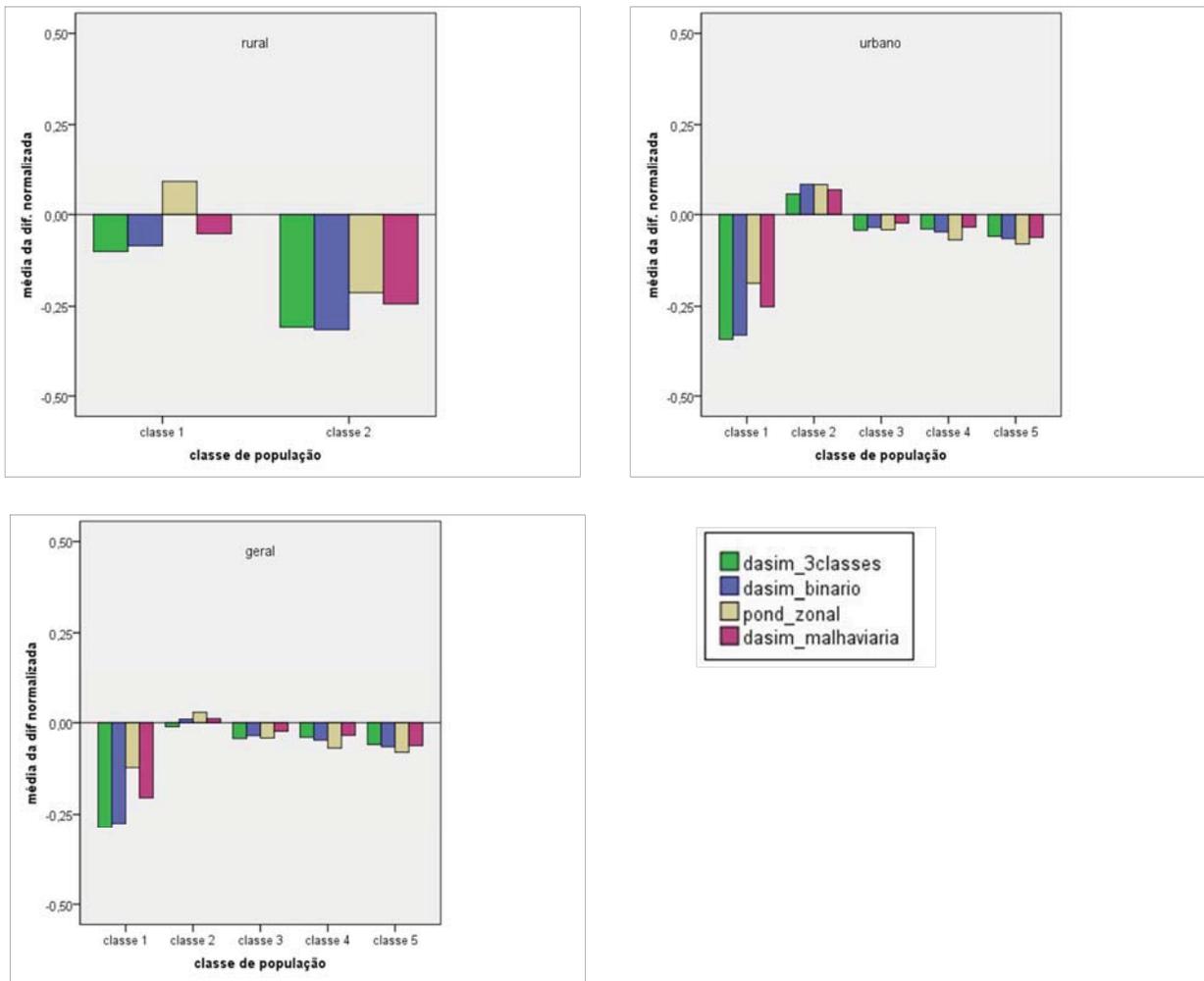


Figura 4: Média da Diferença normalizada por classe de população de acordo com a localização e o método.

Pode-se observar que foram obtidos valores comparativamente altos para todos os coeficientes com o método de ponderação zonal em área rural. Como já mencionado anteriormente, acredita-se que isso se deva a fatores locais, pois este método não costuma apresentar resultados muito bons em áreas rurais, apesar de apresentar resultados satisfatórios para áreas urbanas, principalmente as mais densas.

Observando a FIG. 4 pode-se dizer que, de uma maneira geral, os modelos de desagregação subestimam os valores de população. Tem-se como exceção a classe 2 de população (50 – 200) na área urbana. As maiores diferenças médias surgem para os menores valores de população, ou seja, para aqueles que pertencem à classe 1 (1 – 50 habitantes). A classe 2, com valores entre 200 e 500 habitantes, apresenta diferenças negativas na área rural e positivas na área urbana, o que nos leva a uma diferença média geral próxima de zero para todos os modelos. Para as classes com maior quantidade de população o método de ponderação zonal apresenta as maiores diferenças; em seguida, estão os modelos dasimétricos com utilização de imagens classificadas com valores bem próximos entre si; por último, está o modelo dasimétrico com vias.

Comparando os resultados obtidos para as diferenças médias por classe de população com aqueles obtidos para

outras áreas no Brasil (BUENO; MARTIN; D'ANTONA, 2013), pode-se observar que não há um padrão visível de subestimação ou superestimação de acordo com as classes de população, no entanto, são necessários estudos mais aprofundados para avaliar mais profundamente esta questão.

#### 4. Conclusões

A avaliação entre os diferentes métodos de desagregação realizada neste estudo sugere que não existe um método melhor do que o outro, pois esta decisão depende do objetivo da análise, da disponibilidade de informações para auxiliar a distribuição da população e das características de ocupação da área de estudo. Assim, nos casos em que o objetivo principal da grade é determinar a presença ou ausência de população em detrimento da quantidade, os três métodos dasimétricos apresentados levam a um resultado satisfatório. Neste caso, o método da ponderação zonal seria descartado por não apresentar bons resultados. Quando além da presença ou ausência de população, a determinação da quantidade de população for também importante, os resultados apontam o método dasimétrico com malha viária como informação auxiliar como sendo o mais adequado, no entanto, os demais métodos também apresentaram bons resultados.

Analisando sob o ponto de vista da disponibilidade de informações auxiliares como suporte aos modelos de distribuição espacial da população, a primeira opção recai sobre dados detalhados relativos ao sistema viário, tanto na área urbana como na área rural. No caso da não existência desses dados, imagens de satélite classificadas quanto ao uso das terras são uma boa opção, não havendo aparentemente diferenças na utilização de uma classificação binária – classe povoada e classe não povoada – ou de uma classificação mais detalhada, considerando a densidade de população – três classes povoadas e uma classe não povoada. Na impossibilidade de aquisição de qualquer dado auxiliar, o método da ponderação zonal pode ser utilizado, mas com conhecimento das suas limitações e das incertezas envolvidas.

A localização da área de estudo também pode orientar na escolha do método para desagregação da população. De uma maneira geral, todos os métodos apresentam resultados melhores para áreas urbanas do que para áreas rurais, indicando que é necessário o desenvolvimento de metodologias específicas para estas áreas ou a utilização de dados auxiliares mais representativos da ocupação rural.

Como conclusão geral, pode-se dizer que os métodos de desagregação representam uma solução viável quando não há a disponibilidade de dados obtidos por meio de agregação direta dos dados coletados, devendo-se, no entanto, ser consideradas nas análises as incertezas proporcionadas pela metodologia adotada.

## Agradecimentos

Este estudo foi apoiado pelo Instituto Brasileiro de Geografia e Estatística - IBGE e pela Coordenação de Aperfeiçoamento de Nível Superior – CAPES (processo 17235-12-0). Os microdados e informações cartográficas detalhadas do Censo Demográfico 2010 foram cedidos pelo IBGE em caráter excepcional.

## REFERÊNCIAS

- [1] BRACKEN, I.; MARTIN, . The Generation of Spatial Population Distributions from Census Centroid Data Source. **Environment and Planning A**, v. 21, n. 4, p. 537-543, 1989.
- [2] BUENO, M. C. D.; MARTIN, D.; D'ANTONA, A. O. The Brazilian Population Grid: a hybrid approach. In: **European Forum for Geostatistics – SOFIA CONFERENCE**, 2013, Bulgária. Disponível em <[http://www.nsi.bg/efgs2013/data/uploads/presentations/DAY2\\_WS1\\_5\\_Paper\\_BUENO\\_ok.pdf](http://www.nsi.bg/efgs2013/data/uploads/presentations/DAY2_WS1_5_Paper_BUENO_ok.pdf)>. Acesso em outubro de 2013.
- [3] CLARKE, J. I.; RHIND, D. W. **Human Dimensions of Global Environmental Change**. International Social Science Council e UNESCO: 1992.
- [4] DEICHMANN, U. **A Review of Spatial Population Database Design and Modelling**. National Center for Geographic Information and Analysis (NCGIA). 1996.
- [5] EICHER, C. L.; BREWER, C. A. Dasymeric Mapping and Areal Interpolation: Implementation and Evaluation. **Cartography and Geographic Information Science**, v. 28, n. 2, p.125-138, 2001.
- [6] EFGS – European Forum for Geostatistics. **GEOSTAT 1A - Representing Census data in a European population grid**, Final Report: 2012.
- [7] EXETER, D. J.; BOYLE, P.; FENG, Z.; FLOWERDEW, R.; SCHIERLOH, N.. The creation of Consistent Areas Through Time (CATTs) in Scotland, 1981-2001. **Population Trends**, v. 119, n.1, 2005.
- [8] HOGAN, D. J. Crescimento populacional, padrões de assentamento e o ambiente físico. In: ANPOCS (ed.). **Ciências sociais hoje**. Rio de Janeiro: ANPOCS, 1992.
- [9] IBGE – Instituto Brasileiro de Geografia e Estatística. **Censo 2010 – Síntese das Etapas da Pesquisa**, 2010. Disponível em <[http://censo2010.ibge.gov.br/images/pdf/censo2010/sintese/sintese\\_censo2010\\_portugues.pdf](http://censo2010.ibge.gov.br/images/pdf/censo2010/sintese/sintese_censo2010_portugues.pdf)>. Acesso em março de 2011.
- [10] \_\_\_\_\_. **Censo Demográfico 2010 - Malha de Setores Censitários**, 2011a. Disponível em <[ftp://geofp.ibge.gov.br/malhas\\_digitais/censo\\_2010/](ftp://geofp.ibge.gov.br/malhas_digitais/censo_2010/)>. Acesso em outubro de 2011.
- [11] \_\_\_\_\_. **Base de informações do Censo Demográfico 2010: Resultados do Universo por Setor Censitário**, 2011b. Disponível em <[ftp://ftp.ibge.gov.br/Censos/Censo\\_Demografico\\_2010/Resultados\\_do\\_Universo/Agregados\\_por\\_Setores\\_Censitarios/](ftp://ftp.ibge.gov.br/Censos/Censo_Demografico_2010/Resultados_do_Universo/Agregados_por_Setores_Censitarios/)>. Acesso em maio de 2012.
- [12] LANDIS, J.R.; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, v. 33, n. 1, p.159–174, 1977.
- [13] LANGFORD, M.; UNWIN, D. J. Generating and mapping population density surfaces within a geographical information system. **The Cartographic Journal**, v. 31, n. 1, p.21-26,1994.
- [14] MARTIN, D. Census 2001: making the best of zonal geographies. In: **The census of population: 2000 AND BEYOND**, Reino Unido: 2000.
- [15] MARTINE, G.; SCHENSUL, D. (eds.). **The Demography of Adaptation to Climate Change**. New York, London and Mexico City: UNFPA, IIED e El Colegio de México, 2013.
- [16] MENNIS, J. Generating Surface Models of Population Using Dasymeric Mapping. **Professional Geographer**, v. 55, n. 1, p. 31-42, 2003.
- [17] NAÇÕES UNIDAS, United Nations Economic Commission for Europe. **Register-based statistics in the Nordic countries - Review of best practices with focus on population and social statistics**. New York and Geneva: United Nations, 2007. Disponível em <<http://www.unece.org/fileadmin/DAM/stats/pu>>

- blications/Register\_based\_statistics\_in\_Nordic\_countries.pdf>. Acesso em julho de 2013.
- [18] NORMAN, P.; REES, P.; BOYLE, P. Achieving Data Compatibility over Space and Time: Creating Consistent Geographical Zones. **International Journal of Population Geography**, v. 9, n.1,p. 365-386, 2003.
- [19] OJIMA, R.; MARTINE, G. Resgates sobre População e Ambiente: breve análise da Dinâmica Demográfica e a Urbanização nos Biomas Brasileiros. **Ideias**, n. 5, p. 55-70, 2012.
- [20] PLUMEJEAUD, C.; PRUD'HOMME, J.; DAVOINE, P.-A.; GENSEL, J. Transferring Indicators into Different Partitions of Geographic Space. In: TANIAR, D.; GERVASI, O.; MURGANTE, B.; PARDEDE, E.; APDUHAN, B. O. (eds.) **ICCSA'10 Proceedings of the 2010 international conference on Computational Science and Its Applications - Volume Part I**. Berlin, Heidelberg: Springer-Verlag, 2010.
- [21] REIBEL, M.; BUFALINO, M. E. A test of street weighted areal interpolation using geographic information systems. **Environment and Planning A**, v. 37, n.1,p. 127–139, 2005.
- [22] RUSANEM, J.; MUILU, T.; COLPAERT, A.; NAUKKARINEN, A. Finnish socio-economic grid data, GIS and the hidden geography of unemployment. **Tijdschrift voor Economische en Sociale Geographie**, v. 92, n. 2, p. 139-147, 2001.
- [23] SMITS, P. C.; DELLEPIANE, S. G.; SCHOWENGERT, R. A. Quality assessment of image classification algorithms for land-cover mapping: a review and a proposal for a cost based approach. **International Journal of Remote Sensing**, v. 20, n.8, p. 1461-1486, 1999.
- [24] SLEETER, R.; GOULD, M. **Geographic Information System Software to Remodel Population Data Using Dasymetric Mapping Methods**. Techniques and Methods 11–C2. U.S. Department of the Interior e U.S. Geological Survey, 2008.
- [25] STEINNOCHER K., KAMINGER I., KOSTL M., WEICHSELBAUM J.. Gridded Population – new data sets for an improved disaggregation approach. In: **European Forum for Geostatistics Workshop**, 2010, Estonia. Disponível em <<http://www.efgs.info/workshops/efgs-2010-tallinn-estonia>> . Acesso em janeiro de 2011.
- [26] TAMMILEHTO-LUODE, M., BACKER, L.; ROGSTAT, L. Grid data and area delimitation by definition. Towards a better European territorial statistical system. In: **Conference of European Statisticians**, Suíça, 2000.
- [27] TAMMILEHTO-LUODE, M. Opportunities and challenges of grid-based statistics. In: **World Statistics Congress of the International Statistical Institute**, 2011, Irlanda.
- [28] TOBLER, W. R. Smooth pycnophylactic interpolation for geographical regions. **Journal of the American Statistical Association**, v. 74,n.1, p. 519-530, 1979.
- [29] XIE, Y. The overlaid network algorithms for areal interpolation problem. **Computer, Environment and Urban Systems**, v. 19, n. 4, p. 287-306, 1995.
- [30] ZANDBERGEN, P. A. Dasymetric Mapping Using High Resolution Address Point Datasets. **Transactions in GIS**, v. 15 (s1), p. 5–27, 2011.

# Evaluation of disaggregation methods to generate population grids

Maria do Carmo Dias Bueno<sup>1</sup>  
Álvaro de Oliveira D'Antona<sup>2</sup>

<sup>1</sup> Civil Engineer, Master in Geomatics UERJ, Rio de Janeiro (RJ), PhD candidate in Demography at the Institute of Philosophy and Human Sciences - UNICAMP, Campinas (SP), technologist of the Brazilian Institute of Geography and Statistics.

<sup>2</sup> Bachelor of Economics, Master in Anthropology, Doctor of Social Sciences (Population Studies), a research collaborator at the Center for Population Studies (NEPO-Unicamp), professor in the School of Applied Sciences (FCA-Unicamp) and the Program postgraduate Demography (IFCH-Unicamp).

---

**Abstract** The purpose of this work is to evaluate disaggregation methods used to generate a population grid, using as reference a population grid built with the aggregation of census micro-data, which allows more accurate results about the performance of these methods. Three methods were selected to conduct the evaluation. The first uses an array of errors and successes in identifying populated and non-populated areas, allowing an evaluation of the accuracy of the spatial distribution. The second method uses coefficients reported by a linear regression to evaluate the data fit. Finally, it is used a formula to calculate the difference between the population values, showing whether there is underestimation or overestimation. The results suggest that the choice of the most appropriate method depends on the purpose of the study, the quality and availability of the ancillary data as well as the features of the interest area.

**Keyword:** population distribution; population grid; disaggregation; Aggregation.

---

## Informações sobre os autores

### Maria do Carmo Dias Bueno (IBGE)

Endereço para correspondência: Centro de Documentação e Disseminação de Informações, Rua General Canabarro, 706 – sala 206 – Maracanã – CEP 20271-205 – Rio de Janeiro, RJ.

E-mail: bueno.mariadocarmo@gmail.com.

Link para o currículo lattes: <http://lattes.cnpq.br/7979542180039260>.

### Álvaro de Oliveira D'Antona (FCA/Unicamp)

Endereço para correspondência: Rua Pedro Zaccharia, 1300, Cidade Universitária – CEP 13484-350 - Limeira, SP .

E-mail: alvaro.dantona@fca.unicamp.br.

Link para o currículo lattes: <http://lattes.cnpq.br/1771971577733548>.

**Artigo Recebido em:** 14-03-2014

**Artigo Aprovado em:** 25-05-2014

∴