



ESCOLA NACIONAL DE CIÊNCIAS ESTATÍSTICAS - ENCE
INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA
IBGE

Avaliação de Métodos Dasimétricos para Estimativa Populacional em Pequenas Áreas

Por

Vitor Oliveira França

Rio de Janeiro

Agosto de 2012

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA

ESCOLA NACIONAL DE CIÊNCIAS ESTATÍSTICAS

Vitor Oliveira França

**Avaliação de Métodos Dasimétricos para
Estimativa Populacional em Pequenas Áreas**

Dissertação submetida ao corpo docente do Programa de Mestrado em Estudos Populacionais e Pesquisas Sociais da Escola Nacional de Ciências Estatísticas, como parte dos requisitos necessários à obtenção do grau de Mestre.

Rio de Janeiro

Agosto de 2012

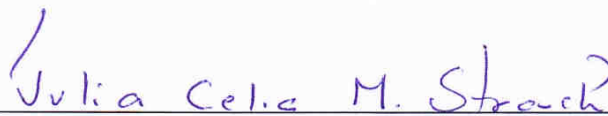


Vitor Oliveira França

**Avaliação de métodos dasimétricos para estimativa populacional
em pequenas áreas**

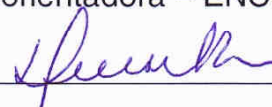
Dissertação apresentada ao Programa de Pós-Graduação em Estudos Populacionais e Pesquisas Sociais da Escola Nacional de Ciências Estatísticas do Instituto Brasileiro de Geografia e Estatística, como requisito parcial para a obtenção do Grau de Mestre.

Banca Examinadora:

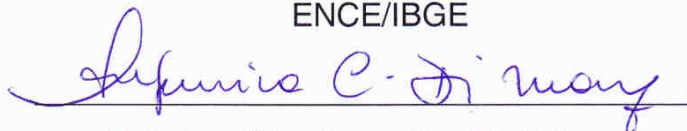


Dr^a. Julia Celia Mercedes Strauch

orientadora – ENCE/IBGE



Dr^a. Denise Britz do Nascimento Silva
ENCE/IBGE



Dr^a. Angélica Carvalho Di Maio
UFF

Rio de Janeiro, 29 de agosto de 2012.

F814a França, Vitor Oliveira

Avaliação de métodos dasimétricos para estimativa populacional em pequenas áreas / Vitor Oliveira França. – Rio de Janeiro, 2012.

103 f.

Inclui referências, glossário e apêndice.

Orientador: Prof. Dr. Julia Célia Mercedes Strauch.

Dissertação (Mestrado em Estudos Populacionais e Pesquisa Social) – Escola Nacional de Ciências Estatísticas.

1. População – Métodos estatísticos – Pesquisa - Teses. 2. População – Métodos estatísticos – Interpolação - Teses. 3. População – Métodos estatísticos – Teoria da estimativa - Teses. I. Strauch, Julia Célia Mercedesi. II. Escola Nacional de Ciências Estatísticas (Brasil). III. IBGE. IV. Título.

CDU: 519.2:314

*Dedico este trabalho às mulheres da minha vida:
minha mãe Maria,
minha esposa Angela
e minha filha Teresa*

AGRADECIMENTOS

Eis que chegou o momento de expressar meus sinceros agradecimentos aos que fizeram parte da minha trajetória até aqui. Primeiro agradeço a minha mãe que não poupou esforços para meu sucesso pessoal e profissional. Ao meu pai, que sempre apoiou meu desenvolvimento intelectual. À minha esposa, que não me deixou desanimar em nenhum momento da elaboração deste trabalho. À professora Julia Célia Mercedes Strauch pelo incentivo e paciência na orientação. A todos os professores do curso, que foram tão importantes na minha vida acadêmica e no desenvolvimento desta dissertação. Aos colegas de curso, que tornaram prazerosa esta caminhada.

Resumo

Informações detalhadas sobre a distribuição espacial das populações humanas são imprescindíveis para tomada de decisões em uma variedade de domínios. No entanto, dados censitários possuem limitações relacionadas às suas escalas espacial e temporal que dificultam sua utilização em algumas aplicações como, por exemplo, defesa civil, epidemiologia e gestão ambiental. A literatura de sensoriamento remoto e geoprocessamento mostra que algumas alternativas foram desenvolvidas para compensar estas limitações, dentre as quais se destaca o mapeamento dasimétrico, uma técnica cartográfica que utiliza informações auxiliares para interpolar variáveis demográficas. Dessa forma, este trabalho buscou investigar e aplicar técnicas dasimétricas para desagregar a variável população total do Censo Demográfico 2010, utilizando dados auxiliares de domínio público. Como área de estudo foi selecionada a Mesorregião Metropolitana de Belém. Foram selecionados três métodos, a saber: o algoritmo do Mapeamento Dasimétrico Inteligente (MENNIS, 2003), a ponderação por logradouros e a ponderação por endereços. Dentre os métodos avaliados, verificou-se que a ponderação por endereços, viabilizada no trabalho pelo Cadastro Nacional de Endereços para Fins Estatísticos (CNEFE), apresentou maior facilidade de aplicação e produziu estimativas mais exatas. Dessa forma, conclui-se que o CNEFE oferece novas oportunidades para espacialização de variáveis demográficas, de forma a facilitar a compatibilização de dados entre zoneamentos distintos e a comparação independente de unidade de área pré-determinada.

Palavras-Chave: Mapeamento dasimétrico, Interpolação zonal, População, Estimativa em pequenas áreas.

Abstract

Detailed information about spatial population distribution is essential for decision making in a variety of domains. However, census data often have limitations related to their spatial and temporal scales that hinder its use in some applications, like epidemiology, emergency planning and mitigation, and environmental management. Therefore, some alternatives have been proposed in the GIS and Remote Sensing literature to compensate these limitations, including dasymetric mapping, a cartographic technique to interpolate demographic data. Thus, this study aimed to investigate and apply dasymetric methods to disaggregate total population counts from 2010 Brazilian Demographic Census, using accessible public domain ancillary data. Metropolitan Region of Belém was chosen as the study area. Three methods have been selected, namely the *Intelligent Dasymetric Mapping* (MENNIS, 2003), street weighting and address weighting interpolation. Among the methods evaluated, it was found that the address weighting showed greater ease of application and produced more accurate estimates. The National Address Cadaster for Statistical Purposes (CNEFE), updated during the last Census and recently released in public domain, offers new opportunities for spatialization of demographic variables and can be used to facilitate integration of data spatially aggregated across in mismatching zonal systems

KEY-WORDS: Dasymetric mapping, Areal interpolation, Population, Small area estimation.

Sumário

1.	Introdução	1
1.1.	Motivação	1
1.2.	Objetivo Geral	4
1.3.	Objetivos Específicos	4
1.4.	Organização dos Capítulos	5
2.	SOBRE OS CENSOS E CONTAGENS DA POPULAÇÃO	6
2.1.	Características de um Levantamento Censitário	6
2.2.	O Censo Demográfico no Brasil	8
3.	MAPAS PARA REPRESENTAÇÃO DA POPULAÇÃO	13
3.1.	A Cartografia Temática e os Mapas Dasimétricos	13
3.2.	Modelos para representação digital da distribuição da população	27
4.	MÉTODOS DE ESTIMATIVA POPULACIONAL EM PEQUENAS ÁREAS	32
4.1.	Interpolação Zonal	32
4.1.1.	Interpolação zonal sem Informação auxiliar	34
4.1.2.	Mapeamento dasimétrico ou interpolação dasimétrica	38
4.1.3.	Outras abordagens e fontes de dados utilizados no mapeamento dasimétrico	44
4.1.4.	Avaliação da exatidão na interpolação dasimétrica	49
4.2.	Modelagem Estatística	49
5.	MÉTODOS DASIMÉTRICOS APLICADOS À MESORREGIÃO METROPOLITANA DE BELÉM	53
5.1.	Área de Estudo	53
5.2.	Materiais utilizados	54
5.3.	Fontes de Dados	56
5.3.1.	Censo Demográfico	57
5.3.2.	Mapas de uso e cobertura da Terra	58
5.3.3.	Base de Logradouros	60
5.4.	Métodos Avaliados	61
5.4.1.	Mapeamento dasimétrico inteligente	62
5.4.2.	Interpolação ponderada por logradouros	63
5.4.3.	Interpolação ponderada por de endereços	65
5.5.	Análise dos resultados	66

5.5.1.	O mapeamento dasimétrico inteligente	66
5.5.2.	A interpolação ponderada por logradouro nas áreas urbanas	70
5.5.1.	A interpolação ponderada por endereços nas áreas rurais	73
5.5.2.	Análise comparativa dos métodos	76
6.	CONCLUSÕES E TRABALHOS FUTUROS	78
	REFERÊNCIAS BIBLIOGRÁFICAS	81
	GLOSSÁRIO	89
	APENDICE A – DIAGRAMAS DE DISPERSÃO	92

Lista de Figuras

Figura 1: Ilustração do método dasimétrico.....	3
Figura 2: Diagrama de unidades político-administrativas e censitárias adotadas no Brasil.	11
Figura 3: Projeção de Mercator.....	i
Figura 4: Mapa Isogônico da Terra	15
Figura 5: Extrato do mapa corocromático de uso da terra criado por Thomas Milne	16
Figura 6: Reprodução da “Carte figurative de l’instruction populaire de la France”	17
Figura 7: Reprodução da “Carte philosophique figurant la population de la France”	17
Figura 8: Mapas coropléticos mostrando população e origem das prostitutas em Paris e na França....	18
Figura 9: Detalhe do mapa “Mostrando o número relativo de passageiros em diferentes direções por tipo de transporte público”.....	20
Figura 10: Mapa de Harness “Mostrando as variedades de sombreamento em comparação á densidade da população”	20
Figura 11: Detalhe do “Dasymetric Map of European Russia”	23
Figura 12: Mapa coroplético e dasimétrico	24
Figura 13: Reprodução do mapa coroplético “The Principal Local Activity of a Place”	25
Figura 14: Visualização da população noturna e diurna do LandScan USA	31
Figura 15: Classificação dos métodos de interpolação zonal	34
Figura 16: Ilustração das iterações do método picnofilático.	38
Figura 17 Fluxo básico para produção de um mapa dasimétrico através do IDM.	41
Figura 18: Métodos de amostragem do algoritmo IDM.....	42
Figura 19: Enquadramento geográfico da área de estudo.	54
Figura 20: Densidade populacional na área de estudo	56
Figura 21: Recorte da classificação do TerraClass na área de estudo.....	59
Figura 22: Extrato do OpenStreetMap para a área de estudo.	61
Figura 23: Fluxograma metodológico para avaliação do método dasimétrico inteligente.	63
Figura 24: Fluxograma metodológico para avaliação da ponderação por logradouros.....	64
Figura 25: Fluxograma metodológico para avaliação da ponderação por endereços.....	65
Figura 26: Mapa dasimétrico da Mesorregião Metropolitana de Belém.....	67
Figura 27: Distribuição espacial do erro por setor produzido com o algoritmo do IDM.	69
Figura 28: População total por setor censitário na área urbana de Belém e Ananindeua.....	70
Figura 29: População por setor censitário estimada na ponderação por logradouro.....	71
Figura 30: Distribuição espacial do erro por setor produzido na interpolação por logradouro.	71
Figura 31: População estimada por endereços rurais na área de estudo.....	73
Figura 32: População estimada por endereço rural residencial.	74
Figura 33: Distribuição espacial do erro por setor produzido na ponderação por endereço	75
Figura 34: Boxplots dos erros verificados em cada método.....	77

Lista de Gráficos

Gráfico 1: Número de citações do termo “Dasymetric” na base de dados do Google Acadêmico.	26
Gráfico 2: Distribuição do erro na interpolação dasimétrica inteligente	69
Gráfico 3: Distribuição do erro na interpolação por logradouro.	72
Gráfico 4: Histograma do erro bruto verificado na ponderação por endereço.	75

Lista de Tabelas

Tabela 1: Bases de dados demográficos modeladas na forma de superfícies.....	30
Tabela 2: Principais métodos de interpolação zonal descritos na literatura.....	48
Tabela 3: Estatísticas de resumo para o erro verificado no método de mapeamento dasimétrico inteligente.....	67
Tabela 4: Estatísticas de resumo para o erro verificado no método de mapeamento dasimétrico por logradouro.....	72
Tabela 5: Estatísticas de resumo para o erro verificado na ponderação por endereço.....	74
Tabela 6: Erro médio quadrático verificado em cada município para os métodos avaliados.....	76

Lista de Equações

Equação 1: Método dos centroides ponderados	35
Equação 2: Pesos para os centroides ponderados	35
Equação 3: Ponderação zonal simples	36
Equação 4: Simplificação da ponderação zonal simples	36
Equação 5: Interpolação picnofilática de Tobler	37
Equação 6: Interpolação inteligente do IDM (Mennis, 2003)	41
Equação 7: Densidade percentual estimada para o IDM	42
Equação 8: Ponderação Zonal Inteligente	43
Equação 9: Estimativa da densidade das classes não amostradas	43
Equação 10: Pesos para ponderação por rede viária	45
Equação 11: População estimada em cada segmento de logradouro	45
Equação 12: Ponderação por rede viária.	45
Equação 13: Ponderação por endereço	47
Equação 14: População média por endereço	47
Equação 15: População estimada por endereço	47
Equação 16: Crescimento alométrico das cidades	50

Lista de Siglas

ACMC – *Area Corrected Multi-Class*

CD2010 – Censo Demográfico 2010

CEDS - *Cadastral Expert Dasymmetric System*

CNEFE – Cadastro Nacional de Endereços para Fins Censitários

DMSP/OLS - *Defense Meteorological Satellites Program/Operational Linescan System*

GPS – *Global Positioning System*

GPW – *Global Population of the World*

IBGE – Instituto Brasileiro de Geografia e Estatística

IDM – *Intelligent Dasymmetric Mapping*

INPE – Instituto Nacional de Pesquisas Espaciais

LIDAR – *Light Detection and Ranging*

MAUP – *Modifiable Areal Unit Problem*

MMB – Mesorregião Metropolitana de Belém

NDVI – *Normalized Difference Vegetation Index*

OGRIP – *Ohio Geographically Referenced Information Program*

GPW - *Grided Population of the World*

SIG – Sistema de Informação Geográfica

SYMAP - *Synergraphic Mapping System*

TIGER – *Topologically Integrated Geographic Encoding and Referencing*

TIN – *Triangular Irregular Network*

1. Introdução

1.1. Motivação

Informações detalhadas sobre a estrutura e a distribuição espacial das populações humanas são imprescindíveis para tomada de decisões em uma variedade de domínios. A principal e mais completa fonte dessas informações são os recenseamentos de população ou de habitação, complexos levantamentos estatísticos realizados por quase todos os países em intervalos regulares.

Os recenseamentos são motivados pela necessidade de informações demográficas para planejamento de políticas públicas a médio e longo prazo, o que não impede que os dados obtidos sejam utilizados em uma ampla variedade de análises. Entretanto, variáveis censitárias possuem algumas limitações relacionadas à escala, tanto espacial quanto temporal, que comprometem sua utilização em algumas aplicações, como meio ambiente, defesa civil, saúde pública e *geomarketing*.

Em relação à escala espacial, os dados censitários representam totais ou médias de unidades de um zoneamento censitário. A agregação constitui uma forma de retirar a individualidade e a identidade das informações individuais, de forma a preservar a sua confidencialidade, além de reduzir o volume de dados. Em muitos países as unidades de agregação do censo são definidas de forma a facilitar a coleta. Dessa forma, a área da unidade censitária varia inversamente à densidade populacional. Em consequência, existe grande incerteza quanto à distribuição da população no interior das unidades do censo, sobretudo nas áreas rurais.

Zoneamentos censitários frequentemente não coincidem com outras regiões de análise, como unidades da paisagem, bacias hidrográficas, regiões administrativas de ensino e saúde etc. A alteração de limites das unidades censitárias ao longo do tempo também dificulta a realização de análises sobre a dinâmica sociodemográfica, pois os dados não podem ser diretamente comparados entre os censos.

Na maioria dos países, os censos são realizados com periodicidade decenal (UNSD, 2008), que é suficiente para atender à sua principal demanda, mas insuficiente para algumas aplicações críticas. Nos períodos intercensitários, a população pode ser

determinada através de contagens ou estimada a partir de projeções, porém com menor detalhamento espacial (LI e WENG, 2010). Normalmente, a população de referência dos censos é a população residencial (ou noturna), que não é adequada, por exemplo, para estimar a população diurna atingida por uma catástrofe (BHADURI *et al.*, 2007). Dessa forma, existe também grande incerteza quanto à distribuição da população fora do período de referência.

Em conjunto com a base cartográfica do zoneamento censitário, as variáveis demográficas podem ser estruturadas em bancos de dados geográficos, tratadas e analisadas por meio dos Sistemas de Informação Geográfica (SIGs), não apenas para representação cartográfica, como também para obter estimativas usadas em análises espaciais e no cálculo de taxas e indicadores. Devido à inerente agregação dos dados censitários, a técnica cartográfica mais utilizada para representar e analisar dados socioeconômicos é a técnica coroplética. Dados agregados em zonas coropléticas inevitavelmente introduzem viés em análises espaciais, pois a técnica assume o pressuposto da distribuição uniforme da variável no interior de cada unidade do zoneamento, o que raramente ocorre com a população. Quanto maior a unidade, maior o viés introduzido, maior a perda em detalhe e especificidade, maior a propagação de erros nos cálculos subsequentes. Dados agregados, e, por conseguinte, mapas coropléticos, estão fadados a *Falácia Ecológica* (ROBINSON, 1950 e 2009) e ao *Problema da Unidade de Área Modificável* (LAM, 1982), frequentemente citados na literatura.

Essas limitações podem ser contornadas pela geração de dados demográficos com resolução espacial e/ou temporal mais detalhada, que representem a distribuição populacional de forma mais realista. Nesse sentido, muitos métodos foram propostos para estimar população. Uma importante contribuição provém do geoprocessamento e do sensoriamento remoto. Na literatura específica, os métodos de estimativa populacional são divididos em duas grandes categorias: modelagem estatística (por regressão) e interpolação zonal (WU *et al.*, 2005).

Os métodos de regressão estatística buscam estabelecer uma relação entre a população e outras variáveis, sobretudo aquelas derivadas de sensoriamento remoto, para estimar a população em períodos intercensitários ou em áreas de difícil recenseamento. Já a interpolação zonal é utilizada para compatibilizar dados entre zoneamentos distintos ou para desagregar variáveis demográficas. Os métodos mais recentes utilizam informações

auxiliares. Mapeamento dasimétrico é um tipo específico de interpolação zonal, baseada em uma técnica cartográfica, que utiliza informações auxiliares para desagregar variáveis demográficas como população total e densidade populacional.

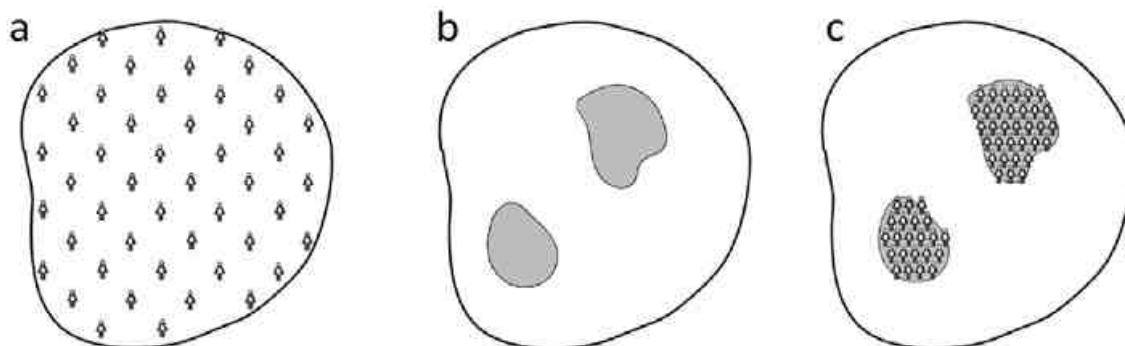


Figura 1: Ilustração do método dasimétrico. (a) população em zonas coropléticas conforme dados do censo; (b) informação auxiliar que diferencia regiões habitadas e inabitadas; (c) população redistribuída nas zonas dasimétricas. Fonte: Langford (2007).

Nas duas últimas décadas, diversas abordagens do método dasimétrico foram desenvolvidas, refinadas e comparadas em diferentes escalas. Classes de uso da terra derivadas de imagens de satélite constituem a informação tradicionalmente utilizada (MENNIS e HULTGREN, 2006; MOON e FARMER, 2001; WEBER, 2010), mas diferentes insumos geoespaciais foram propostos, incluindo luzes noturnas registradas por satélites meteorológicos (DMSP/OLS¹) (ELVIDGE *et al.*, 1999), bases de arruamento (REIBEL e BUFALINO, 2005), bases cadastrais de prefeituras (MAANTAY *et al.*, 2007), cadastros de endereços (ZANDBERGEN, 2011) e modelos digitais de superfície derivados de perfilamento laser (LIDAR) (LWIN e MURAYAMA, 2011). Algumas iniciativas foram realizadas para utilizar várias fontes de dados auxiliares (BHADURI *et al.*, 2007), em uma abordagem multidimensional.

Apesar de grande produção científica, ainda existem algumas barreiras que impedem o uso generalizado de métodos dasimétricos. Não existe consenso na literatura em relação ao melhor e poucos estudos abordam a exatidão proporcionada pelos diferentes métodos. Também são poucos e recentes os algoritmos disponibilizados para facilitar a aplicação desses métodos nos aplicativos SIG. No Brasil, apesar de relevantes trabalhos

¹ *Defense Meteorological Satellite Program* é um programa do Departamento de Defesa Norte Americano que projeta, constrói, lança e mantém vários satélites para o monitoramento meteorológico e oceanográfico. O sensor *Operational Linescan System* (OLS) é um radiômetro de varredura oscilatória que adquire imagens globais noturnas e diurnas em duas faixas do espectro eletromagnético - visível e infravermelho termal.

acerca do tema (AMARAL *et al.*, 2005, 2012; DUTENKEFER, 2010), não se tem conhecimento de uma iniciativa para produção sistemática de modelos detalhados da distribuição espaço-temporal da população, representativos do território nacional ou parte dele.

Vale lembrar que, ainda que seja utilizada quase que exclusivamente para representar e analisar a população residente ou a densidade populacional, a interpolação dasimétrica pode ser aplicada para desagregar qualquer variável quantitativa, desde que se tenha disponível um conjunto de dados auxiliares fortemente correlacionados com esta variável.

Desta forma, este trabalho busca contribuir para o estado da arte dos métodos dasimétricos, uma vez que a crescente oferta de dados geoespaciais e metodologias desenvolvidas tornam viável a representação mais detalhada da distribuição espacial da população.

1.2. Objetivo Geral

Este trabalho tem por objetivo aplicar e avaliar métodos dasimétricos para estimativa populacional na Mesorregião Metropolitana de Belém, a fim de elaborar uma metodologia para produção de mapas dasimétricos. Tendo em vista o ganho em exatidão que tais estimativas proporcionariam às análises espaciais que envolvem populações humanas em diversas áreas do conhecimento, os métodos aqui foram testados de modo que possam ser aplicados em outras áreas do território nacional, combinadas e aperfeiçoadas de acordo com a disponibilidade de dados auxiliares mais detalhados.

1.3. Objetivos Específicos

O objetivo geral do presente trabalho pode ser decomposto em objetivos mais específicos, que permitirão maior compreensão do tema e do que se pretende realizar. Para tanto, algumas etapas devem ser cumpridas no sentido de abordar os seguintes tópicos:

- i. Avaliar e selecionar bases de dados que possam ser utilizadas para detalhamento espacial ou temporal de variáveis demográficas;

- ii. Definir procedimentos para desagregação sistemática de dados demográficos;
- iii. Selecionar e aplicar métodos mais adequados para mapeamento dasimétrico em áreas urbanas e em áreas rurais;
- iv. Produzir mapas dasimétricos a partir de dados do Censo Demográfico 2010;

1.4. Organização dos Capítulos

A fim de apresentar conceitos e análises, de modo a atingir os objetivos aqui propostos, esta dissertação é desenvolvida em cinco capítulos. O Capítulo 2 contextualiza os censos e contagens da população. O Capítulo 3 apresenta uma revisão das bases teóricas que motivaram a pesquisa e a cartografia temática, para explicar a origem e evolução da representação da distribuição espacial da população, com destaque para a técnica dasimétrica, uma contribuição da escola russa de Geografia. O Capítulo 4 apresenta uma revisão da literatura sobre os métodos utilizados para estimar população em escalas diferentes daquelas disponibilizadas pelos censos e contagens. Neste capítulo são ainda abordadas iniciativas para estimar populações através de interpolação zonal. O Capítulo 5 apresenta o experimento realizado para avaliar os métodos dasimétricos selecionados e descreve os resultados obtidos com cada método. E finalmente, o Capítulo 6 apresenta as conclusões e considerações finais do trabalho.

2. SOBRE OS CENSOS E CONTAGENS DA POPULAÇÃO

A tomada de decisões com base empírica é um paradigma universalmente aceito de gestão eficiente e governança eficaz nas sociedades modernas em geral. Nesse contexto, como já foi dito, estrutura e distribuição espacial da população são informações imprescindíveis para tomada de decisões. A principal e mais completa fonte destas informações são os recenseamentos de população ou de habitação. Segundo as Nações Unidas (UNSD, 2008, p. 7), Um recenseamento de população (ou habitação) pode ser definido como *"o conjunto das operações com o propósito de coletar, agrupar e publicar dados demográficos, econômicos e sociais relativos a todos os habitantes (ou a todos os domicílios e seus ocupantes) de um país, ou de uma parte bem delimitada de um país, num momento específico"*.

Para melhor compreensão das operações envolvidas na construção de uma base territorial para a realização do Censo, neste capítulo são abordados os requisitos de um levantamento censitário e descritos alguns aspectos do censo demográfico no Brasil.

2.1. Características de um Levantamento Censitário

Realizar um censo é uma das mais complexas e massivas tarefas que uma nação pode realizar em tempo de paz. Requer mapear todo o país, mobilizar um exército de recenseadores treinados, realizar uma intensa campanha publicitária, visitar todos os domicílios para coletar informações individuais, compilar as respostas, analisar e divulgar os resultados. Mesmo com alto custo, divulgação demorada e frequência reduzida, e ainda que existam alternativas para coleta de informação, os censos ainda constituem o principal instrumento para obter dados sobre a população.

Contagens periódicas da população existem desde a era pré-cristã, como as que eram realizadas pelos impérios do Egito, Babilônia, China, Palestina e Roma com objetivos fiscais e militares. Esses levantamentos consideravam apenas os indivíduos relevantes para essas finalidades: proprietários, chefes de família ou homens em idade adequada ao alistamento militar. O primeiro censo moderno a contemplar todos os requisitos acima

listados foi o Censo Sueco de 1749 (HAKKERT, 1996). Na América, os Estados Unidos estabeleceram na sua Constituição de 1787 a base legal para a organização decenal de censos populacionais. A série foi iniciada em 1790 (HAKKERT, 1996).

A implantação dos censos em todos os países foi uma recomendação do Primeiro Congresso Internacional de Estatística, realizado em Bruxelas no ano de 1853. A Divisão de Estatísticas das Nações Unidas, desde 1958 vêm promovendo ativamente os censos, através da compilação e divulgação de princípios e recomendações pertinentes. Os requisitos essenciais de um levantamento censitário moderno incluem (UNSD, 2008, p 8):

- **Enumeração individual de todas as pessoas (ou domicílios):** condição necessária para a elaboração de quadros estatísticos detalhados. Deve-se evitar qualquer referência a finalidades tributárias ou militares. A coleta é geralmente realizada pelo recenseador, através de entrevistas individuais ou com o chefe do domicílio. Alguns países utilizam questionários distribuídos, inclusive pelo correio.
- **Universalidade da enumeração dentro do território definido:** seja a residência habitual (população residente, *de jure* ou de direito), seja o paradeiro na data de referência (população presente ou *de facto*), o critério é o mesmo para todos os habitantes. A universalidade não exclui a possibilidade de amostragem, desde que as informações básicas sejam coletadas para o conjunto da população. Em alguns casos, uma parte da população é excluída da contagem, seja por dificuldades operacionais ou por falta de interesse das autoridades.
- **Simultaneidade de todo o levantamento:** deve existir um período de referência pré-definido, como a meia noite de uma data determinada para outra. Embora a contagem possa levar semanas ou até meses, toda a enumeração deve referir-se a essa data de referência.
- **Periodicidade definida:** preferencialmente a cada cinco ou dez anos, pois nas análises demográficas a população normalmente é agregada em faixas etárias de cinco ou dez anos, de modo que intervalos não múltiplos de cinco dificultam a comparação entre censos sucessivos. A maioria dos países mantêm censos decenais.

- **Respaldo Legal:** deve haver legislação com especificação do escopo, dos fins, do orçamento, da administração, das garantias legais quanto ao sigilo da informação e das demais obrigações da entidade executora. Ao contrário de outros levantamentos estatísticos, o não fornecimento de informação à autoridade censitária está normalmente sujeito a sanções legais.

Um censo moderno é normalmente realizado em três etapas (HAKKERT, 1996, p. 21). O pré-recenseamento compreende as atividades para viabilizar o trabalho de campo, como o levantamento cartográfico, elaboração do questionário, plano tabular, calendário, além de recrutamento e treinamento. A consolidação da base territorial permite orientar a divisão do território em pequenas áreas para viabilizar a coleta do censo e outras pesquisas domiciliares. O recenseamento propriamente dito consiste na execução e supervisão do trabalho de campo. O pós-recenseamento envolve a revisão e crítica dos dados levantados, processamento e publicação.

Conforme a situação de cada país, o questionário é bastante variado. Quesitos considerados essenciais compreendem nome, idade, sexo, relação de parentesco com o chefe do domicílio ou da família, estado civil, ocupação, alfabetização, lugar de nascimento, residência habitual (*de jure*) ou lugar de enumeração (*de facto*). Também são frequentes quesitos sobre religião, raça ou etnia, língua nativa ou falada e renda. O nome do indivíduo é levantado com o propósito de identificação dos questionários.

Toda divulgação deve garantir a confidencialidade das informações. Para garantir essa confidencialidade, normalmente os dados são agregados em unidades de área hierárquicas. As unidades menores são caracterizadas apenas em termos das suas características demográficas principais como número de habitantes, e variáveis mais específicas, como renda, são compiladas em unidades mais abrangentes. A ameaça de violação das garantias de confidencialidade pode colocar em risco os institutos nacionais de estatística.

2.2. O Censo Demográfico no Brasil

O primeiro regulamento censitário do Brasil, instituído em 1846, determinava o intervalo de oito anos para execução dos censos. O primeiro censo seria realizado em julho

de 1852, mas fracassou devido à resistência popular, num levante que ficou conhecido como a Guerra dos Marimbondos. Um novo regulamento censitário foi estabelecido em 1870, aumentando o intervalo intercensitário para dez anos. *O Recenseamento da População do Império do Brasil*, realizado em 1872, é considerado o primeiro censo nacional efetivamente realizado, pois nele se explicitavam todos os critérios essenciais a este tipo de levantamento. Além disso, em oposição aos levantamentos anteriores, que tinham propósitos militares ou fiscais, neste censo se manifesta o propósito de se conhecer a população para a construção da nação Brasileira.

O *V Recenseamento Geral do Brasil*, realizado em 1940, foi organizado pelo recém-criado Instituto Brasileiro de Geografia e Estatística (IBGE), iniciando uma nova fase nas pesquisas populacionais do Brasil. Desde então o Censo Demográfico tem sido realizado rigorosamente a cada 10 anos, com exceção do censo de 1991, que atrasou devido à demora na liberação de verbas para contratação de recenseadores.

Desde o Censo Demográfico de 1950, adota-se o setor censitário, definido como a menor unidade territorial de coleta e controle cadastral, integralmente contida em área urbana ou rural de um mesmo distrito, cujo conjunto esgota a totalidade do Território Nacional (IBGE, 2011). O setor censitário possui dimensão e número de domicílios que permitem o levantamento por um único recenseador.

A partir de 1960, passou-se a utilizar a técnica de amostragem na coleta das informações, e desde então o Censo Demográfico utiliza dois questionários:

- Do universo - aplicado em todas as unidades domiciliares, contém perguntas referentes às características básicas dos domicílios e de seus moradores;
- Da amostra - aplicado apenas nos domicílios selecionados para a fração amostral, contém perguntas sobre os temas religião, cor ou raça, deficiência, migração, escolaridade, fecundidade, nupcialidade, trabalho e rendimento.

As variáveis demográficas obtidas a partir dos questionários da amostra não possuem significância estatística ao nível de setor censitário. Neste caso, a menor unidade de análise é a Área de Ponderação (APOND), definida como a unidade geográfica formada por um agrupamento de setores, para a aplicação dos procedimentos de calibração das estimativas com as informações conhecidas para a população como um todo, onde os resultados já

apresentam significância estatística e podem ser trabalhados (IBGE, 2012a). A partir destas unidades básicas os resultados podem ser tabulados e disponibilizados em diferentes escalas: bairros², subdistritos, distritos, municípios, regiões metropolitanas, microrregiões, mesorregiões, estados e grandes regiões (**Figura 2**).

As inovações tecnológicas possibilitam o aprimoramento nos processos produtivos e novas oportunidades na disseminação da informação censitária para a sociedade. A informatização proporcionou uma série de evoluções no Censo Demográfico. A partir de 1991, os resultados passaram a ser disseminados através de planilhas de dados agregados por setores censitários (para o universo) e microdados (para a amostra). Utilizou-se pela primeira vez disquetes e sistemas informatizados para consulta aos dados.

² A partir do Censo Demográfico 2010, somente para os municípios que possuem bairros estabelecidos em legislação.

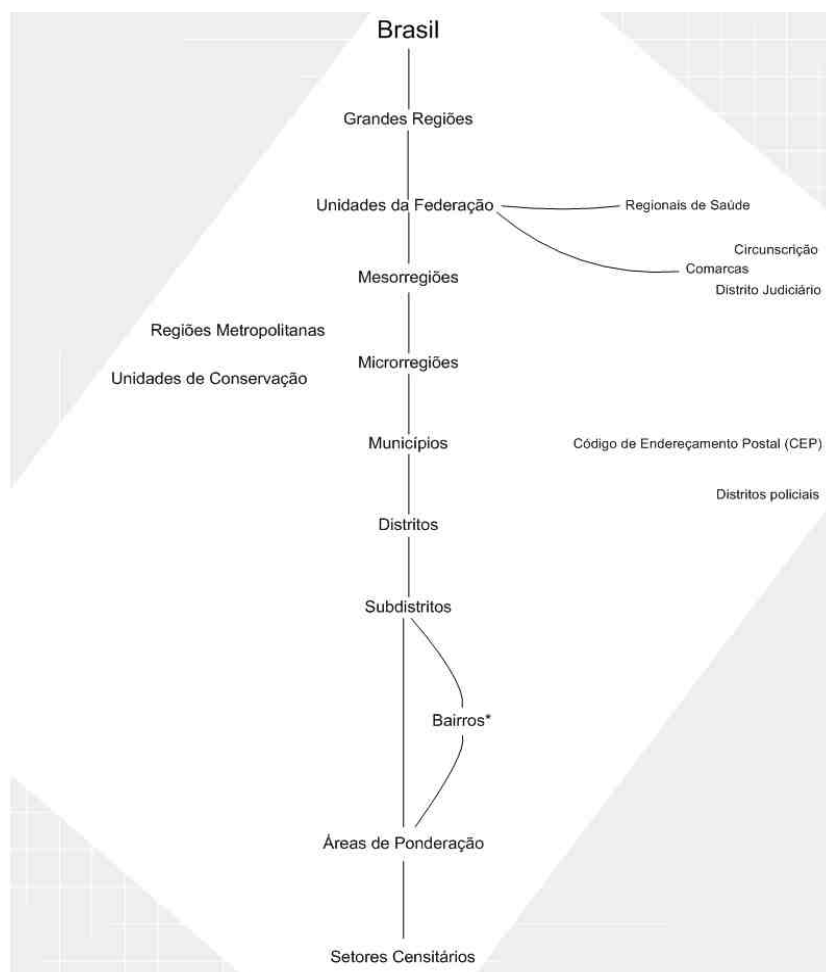


Figura 2: Diagrama de unidades político-administrativas e censitárias adotadas no Brasil.
Fonte: baseado em IBGE (2011).

Durante a pré-coleta do Censo 2000, o IBGE iniciou a elaboração de uma Base Territorial em meio digital, quando foi desenvolvido o *SisCart* (Sistema de Elaboração Semi- Automática de Mapas Municipais) em plataforma CAD (*Computer Aided Design*). Este censo foi o primeiro a disponibilizar uma malha de setores em formato vetorial. Naquela época adotou-se metodologia diferenciada para setores rurais e urbanos, que foi mantida até 2007. Como consequência, as duas malhas não foram integradas.

Para o Censo de 2010, o IBGE realizou um grande esforço para possibilitar a produção de dados demográficos georreferenciados. O Sistema de Mapeamento de Base Territorial (SISMAP) permitiu a consolidação de uma Base Territorial das malhas urbana e rural. As agências do IBGE atualizaram a cartografia dos municípios com até 20 mil habitantes e empresas de mapeamento foram contratadas para atualizar os demais. Os recenseadores utilizaram computadores de mão equipados com GPS para aplicação dos questionários, o que facilitou a orientação e supervisão da coleta.

Como resultado, a malha territorial do último censo é totalmente contínua. Os arquivos vetoriais podem ser integrados e analisados em conjunto com outros insumos cartográficos através dos SIGs. Outra inovação foi criação de um cadastro digital de endereços, a partir da digitalização das folhas de coleta do censo anterior e consolidação durante a coleta. Primeiro arquivo público do gênero no país, o Cadastro Nacional de Endereços para Fins Estatísticos (CNEFE), contém todas as unidades registradas pelos recenseadores. Nas áreas rurais, os registros contêm coordenadas geográficas das propriedades, estabelecimentos e domicílios. Foram inseridas na base as informações referentes à numeração das edificações existentes no início e final de cada trecho de logradouro.

Com cerca de 78 milhões de endereços, o CNEFE apoiará a realização das pesquisas domiciliares do IBGE nesta década, facilitando a seleção das amostras (domicílios que serão visitados na pesquisa) e o acompanhamento das operações de coleta. Além disso, o cadastro torna possível a incorporação de dados não georreferenciados provenientes de outras bases da administração pública, tais como: registro civil, sistemas de saúde e segurança pública.

O calendário de divulgação do Censo 2010 ainda prevê uma série de produtos georreferenciados, como arquivos de eixos de logradouros, quadras e faces (IBGE, 2012b). Muitas dessas informações estão sendo disseminadas através de *GeoServiços*, tecnologia que possibilita a interoperabilidade entre dados geoespaciais através da *Web*.

3. MAPAS PARA REPRESENTAÇÃO DA POPULAÇÃO

Neste capítulo é apresentada uma breve revisão histórica da cartografia temática até a origem dos mapas dasimétricos e a seguir são apresentados os modelos para representar a distribuição espacial da população em meio digital.

3.1. A Cartografia Temática e os Mapas Dasimétricos

Os primeiros empreendimentos cartográficos modernos estavam voltados para a produção de mapas destinados à navegação e exploração, permitindo ampliar nossa capacidade de observar e medir o mundo físico com elevados níveis de precisão e exatidão. A teoria geométrica e a concepção de instrumentos como a bússola e o sextante tornaram possível retratar precisamente os aspectos tangíveis do mundo físico como topografia, batimetria e temperatura (DORLING e FAIRBAIRN, 1997).

O grande avanço dessa cartografia deu-se na Europa durante o Renascimento (séculos XIV ao XVI), impulsionado pela intensificação do comércio entre Ocidente e Oriente e pelo surgimento das relações capitalistas. Houve uma grande demanda por mapas para navegação, bem como de meios para a respectiva orientação. Os primitivos mapas de Portolani, uma espécie de carta náutica do fim da Idade Média, eram pouco melhores que esboços a mão livre. Foram então aprimorados com a inclusão de “*uma rede de rosas-dos-ventosentrelaçadas, cujos raios demandavam a localidades e portos nos recortes costeiros*” (MARTINELLI, 2005). Esses mapas, acompanhados da bússola, forneceram a orientação necessária para o início das grandes navegações.

A invenção da imprensa no século XV teve grande influência no progresso da cartografia, reduzindo os erros introduzidos pelos copistas e barateando os custos, assim permitiu a disseminação dos mapas como mercadoria. A cartografia estabeleceu-se como ofício e ganhou uma definição: a exposição gráfica da informação geográfica (KISH, 1980 *apud* MARTINELLI, 2005).

Em 1569, o geógrafo e cartógrafo Gerhard Mercator (1512-1594) apresentou ao mundo seu planisfério, expressamente concebido para uso da navegação marítima, que constituiu um notável progresso na cartografia náutica (**Erro! Auto-referência de indicador não válida.**). Apesar da deformação nas proporções de área e distâncias, a projeção preservou a coerência de uma única característica conveniente à navegação – a direção. Mercator não deixou nenhuma explicação matemática sobre a sua projeção, o que foi realizado pelo inglês Edward Wright (1561-1615) em sua obra *Certaine Errors in Navigation* de 1599 (CROSBY, 1997).

O Ocidente do século XVI, mais especificamente a Europa renascentista, progrediu mais do que qualquer outra sociedade em sua capacidade de controlar o ambiente e projetar seu poderio, através da ciência e da tecnologia. Contribuiu para isso a cartografia, que se integrou ao processo capitalista de produção e os mapas confirmaram-se como armas do imperialismo para promoção da política colonial.

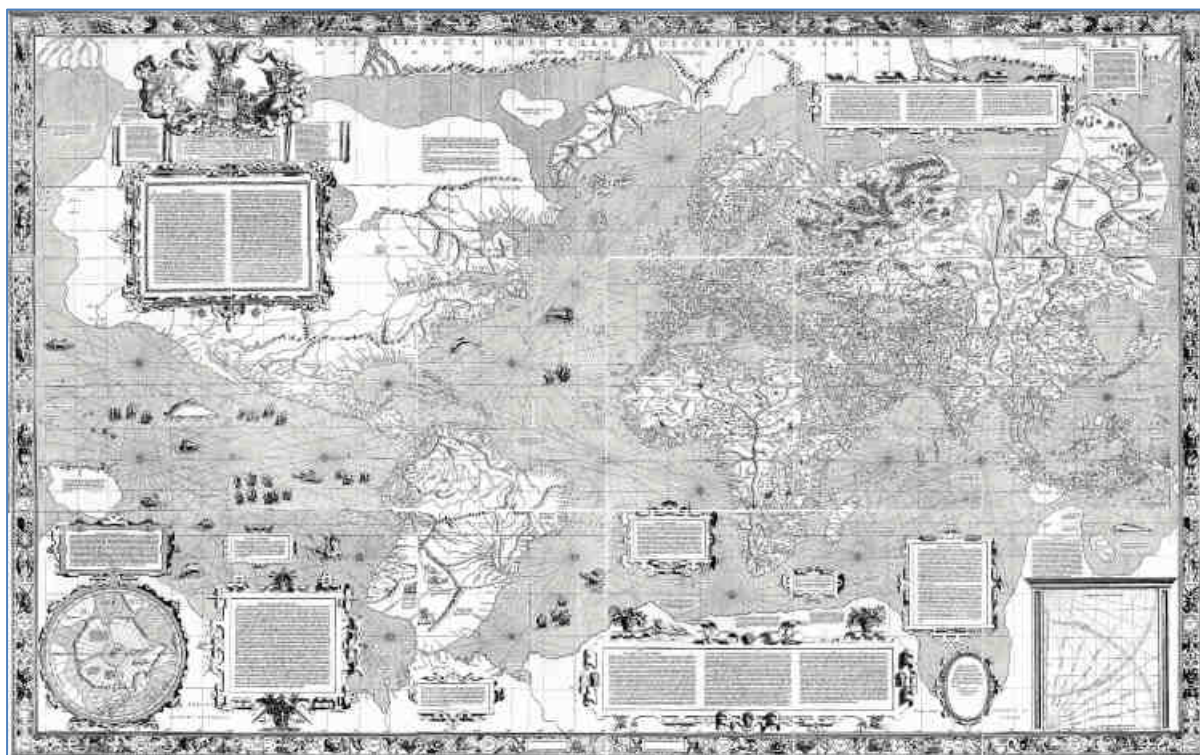


Figura 3: Projeção de Mercator utilizada na obra *Nova et Aucta Orbis Terrae Descriptio ad Usum Navigantium Emendate* (Nova e aumentada descrição da Terra corrigida para uso da navegação) de 1569.
Fonte: Wikipedia.

O marco inicial do estabelecimento de métodos para a cartografia temática foi o trabalho realizado pelo Astrônomo Real Britânico Edmond Halley (1656-1743), que em 1686 produziu o Mapa dos ventos oceânicos e das monções, tido como primeiro mapa temático. Em 1701, Halley criou o Mapa das declinações magnéticas do Oceano Atlântico

(**Figura 4**), inventando o método isarítmico ao traçar linhas de igual valor (isolinhas) que delimitavam regiões de valores constantes do campo magnético superficial da Terra (MACEACHREN, 1979).

O método isarítmico também foi explorado pelo geógrafo alemão Alexander Von Humboldt (1769-1859) que utilizou *isotermas*, em 1817, para representar o padrão de distribuição das temperaturas no globo, dando origem aos mapas climatológicos (FRIENDLY e DENIS, 2005).

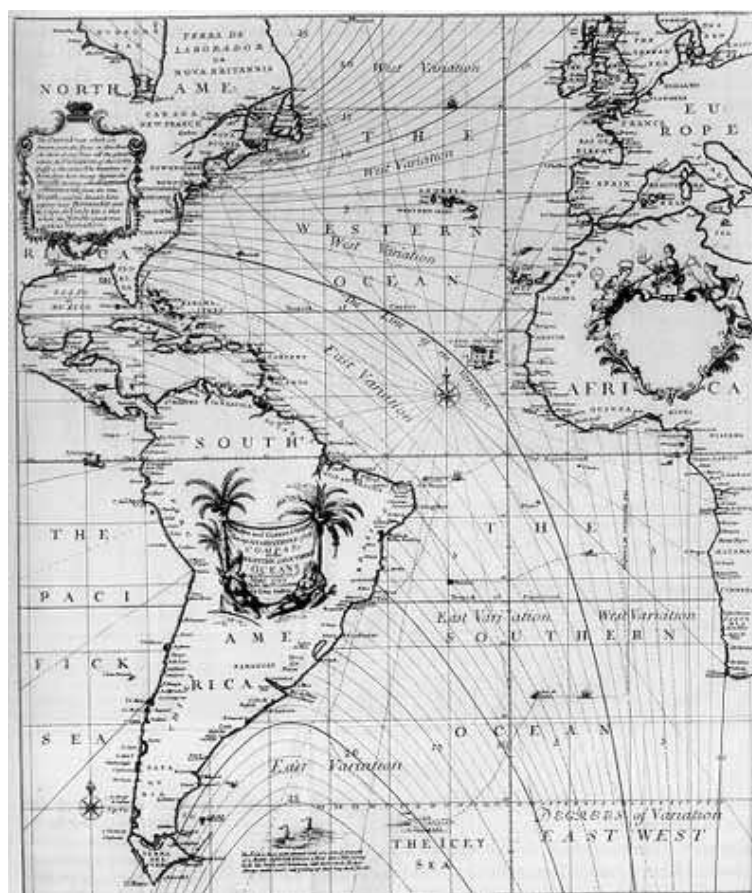


Figura 4: Mapa Isogônico da Terra (com linhas de igual declinação magnética). Criado em 1701 por Edmond Halley. Fonte: (FRIENDLY; DENIS, 2005).

Em 1782, o engenheiro francês Jean-Louis Dupain-Triel (1722-1805) sugeriu a utilização das curvas de nível para representar o relevo emerso sem sobrecarregar os mapas, e em 1798, com a inclusão de cores nas faixas de altitudes (tons escuros nas partes baixas do relevo e tons claros e nas partes altas), publicou o primeiro mapa *hipsométrico* conhecido. Já a invenção do método *corocromático* é atribuída a Thomas Milne, que em 1800 produziu um mapa da região metropolitana de Londres com 17 categorias de uso da

terra, codificadas por meio de cores e letras (**Figura 5**) (BULL, 1956 *apud* MARTINELLI, 2010).

Mapas que descrevem aspectos sociais, culturais e econômicos, os mapas temáticos ou estatísticos, surgem timidamente no início do século XIX. A quantificação da população começou a se consolidar na década de 1820, com a inclusão do dígito referente ao número de habitantes nos mapas, apresentados em conjunto com quadros estatísticos.

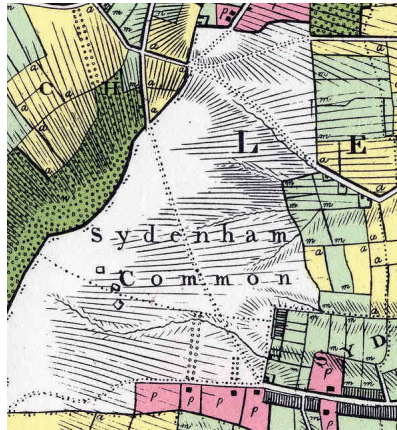


Figura 5: Extrato do mapa de uso da terra criado por Thomas Milne em 1800, utilizando a técnica corocromática. Fonte: (FRIENDLY; DENIS, 2005).

O matemático católico Pierre Charles François Dupin (1784-1873) apresentou um cartograma onde associou uma ordem visual de sombreamento contínuo, do claro para o escuro, a uma sequência de dados agrupados em classes, os níveis de analfabetismo nos departamentos administrativos da França de 1826 (**Figura 6**).

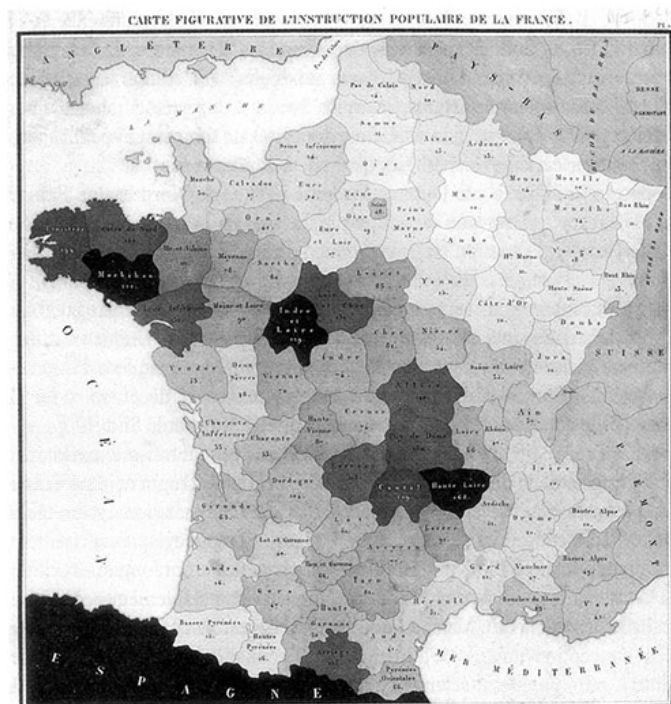


Figura 6: Reprodução da “Carte figurative de l’instruction populaire de la France” de autoria de Pierre Dupin, 1826. Fonte: (FRIENDLY e DENIS, 2005)

Este mapa (**Figura 6**) é considerado o primeiro a empregar o método *coroplético*, e o primeiro a retratar quantitativamente os aspectos sociais (FRIENDLY, 2008; MACEACHREN, 1979). Dupin evidenciou a relação entre a instrução e o desenvolvimento econômico, colocando frente a frente uma França já esclarecida contra outra ainda no obscurantismo. Assim, este cartograma constituiu uma imagem eloquente em favor da indústria e das doutrinas inglesas de civilização econômica e política. Nesta época, a cartografia temática se associava ao discurso do capitalismo liberal emergente (PALSKY, 1996 *apud* MARTINELLI, 2009).

O monge franciscano, Armand Joseph Frère de Montizon (1788-?) utilizou pontos para representar a distribuição espacial da população da França. No seu mapa, mostrado na **Figura 7** cada ponto representa 10.000 indivíduos.

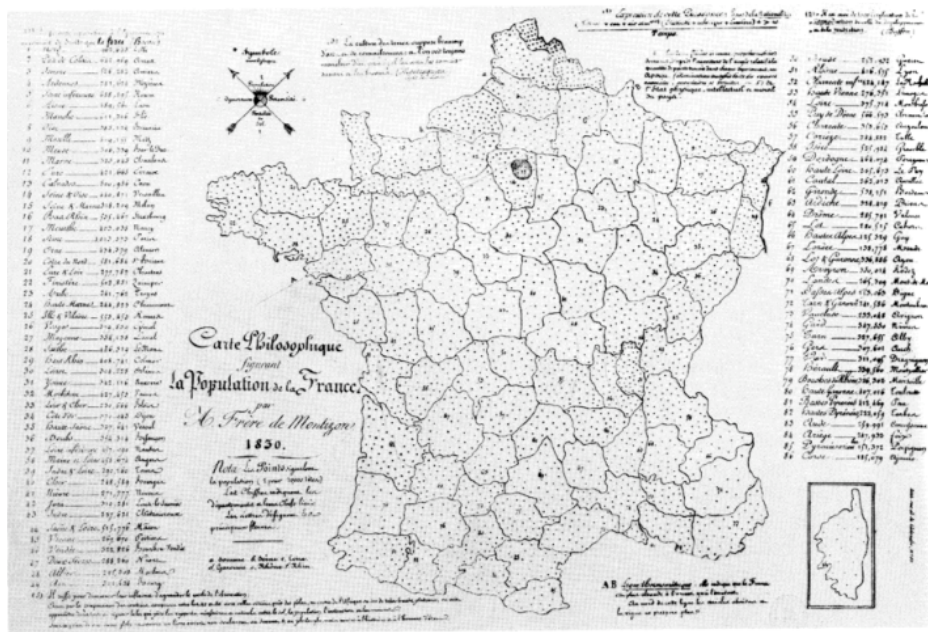


Figura 7: Reprodução da “Carte philosophique figurant la population de la France” (1830). Fonte: (FRIENDLY; DENIS, 2005)

Neste mapa (**Figura 7**) os pontos são regularmente espaçados conforme a população total de cada departamento administrativo. Pode ser considerado o primeiro exemplo de *Falácia ecológica*, uma vez que transmite ao leitor uma ideia de que a população está uniformemente distribuída no interior das unidades administrativas. Sua inovação foi

pouco divulgada. Mais tarde, em 1859, o mapa de densidade de pontos foi “reinventado” por Alexander von Mentzer (MARTINELLI, 2010).

Em um estudo sobre a prostituição em Paris, publicado em 1836, o médico sanitarista Alexandre Jean Baptiste Parent-Duchâtelet (1790–1835) realizou uma extensa tabulação de dados, analisou séries temporais e utilizou a técnica coroplética para apresentar seus resultados (**Figura 8**). O método coroplético difundiu-se para representação de valores relativos como índices e taxas (PALSKY, 1996 *apud* FRIENDLY, 2009).

O Estado moderno iniciado na Itália do século XVI tem seu ápice na Revolução Industrial no Reino Unido do século XVIII. Surge a necessidade de informações mais detalhadas sobre a população e fenômenos associados como saúde, criminalidade, educação, pobreza e economia. Quase todos os métodos utilizados nos gráficos estatísticos e mapas temáticos que conhecemos atualmente tiveram origem entre 1850 e 1900, período que (FRIENDLY, 2009) chamou de “a era dourada dos gráficos estatísticos”. Dados necessários à elaboração desses gráficos e mapas começaram a ser sistematicamente coletados e disponibilizados pelos institutos estatísticos nacionais e internacionais.

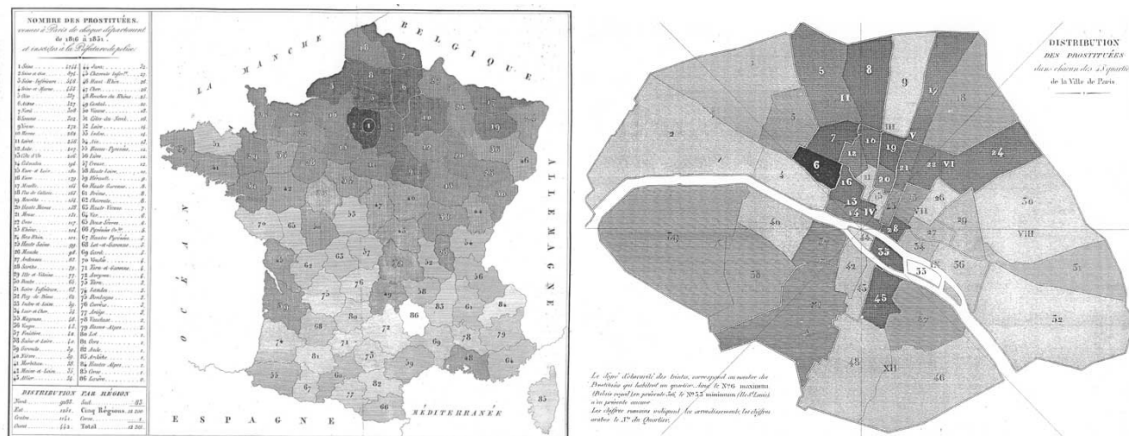


Figura 8: Mapas coropléticos mostrando população e origem das prostitutas em Paris (esquerda) e na França. Fonte: (PARENT-DUCHATELET, 1837)

A sistematização metodológica da cartografia temática foi discutida na terceira edição do Congresso Internacional de Estatística, realizada em Viena no ano de 1857. Os métodos de representação cartográfica foram organizados em correspondência às grandes categorias do conhecimento, e em resposta às questões – “o quê?”, “quanto?”, “onde?” e “quando?” (MARTINELLI, 2005).

O economista francês Pierre Émile Levasseur (1828-1911), filho do famoso cartógrafo Victor Levasseur, introduziu a representação estatística na cartografia da população. O *método Levasseur* estipulava oito classes de densidade demográfica, quatro acima e quatro abaixo da média, representadas mediante duas ordens visuais opostas. Émile também propôs a representação coroplética para números relativos, como a taxa de urbanização (MARTINELLI, 2005).

Considerada a pedra fundamental da Geografia Humana, a obra “*Antropo-geografia, ou a aplicação ampla da geografia à história*”, publicada em 1882 por Friedrich Ratzel (1844 - 1904), mostra a possibilidade de utilizar os mapas em estudos territoriais e demográficos. Conforme observado por Moraes (2005), sua pesquisa torna-se um instrumento poderoso de legitimação da política expansionista do Estado alemão recém-constituído.

O mapa de densidade populacional mundial produzido, em 1833, pelo geólogo e economista britânico George Poulett Scrope (1797-1876), utilizava uma técnica dasimétrica rudimentar para delimitar com linhas de espessura variável as regiões completamente povoadas, rarefeitas e inabitadas da Terra (SCROPE, 1833 *apud* MARTINELLI, 2005).

O engenheiro ferroviário Henry Drury Harness (1804-1883), publicou uma série de mapas temáticos para acompanhar um relatório sobre as ferrovias na Irlanda, nos quais quantificava o efetivo e o deslocamento da população. Foi o pioneiro ao utilizar o que hoje chamamos de mapas de fluxo (**Figura 9**), e também elaborou um mapa da densidade populacional da Irlanda (**Figura 10**) com dados do censo de 1831.



Figura 9: Detalhe do mapa “Mostrando o número relativo de passageiros em diferentes direções por tipo de transporte público” primeiro de mapa de fluxo. Utiliza a largura das linhas para representar um dado quantitativo. Fonte: (HARNESS, 1837 apud ROBINSON, 1955)³.

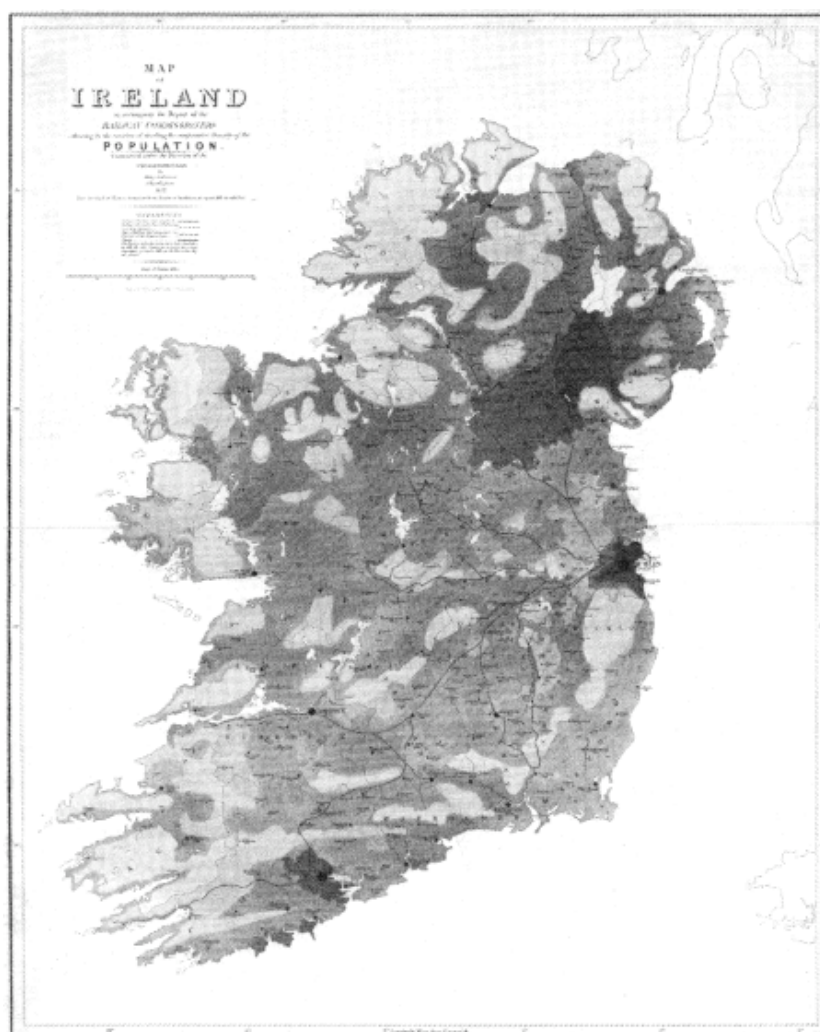


Figura 10: Mapa de Harness “Mostrando as variedades de sombreamento em comparação à densidade da população. Fonte: (HARNESS, 1837 apud ROBINSON, 1955).

³ Atlas to accompany the second report of the railway commissioners, Ireland 1837.

Neste mapa (**Figura 10**), Harness (1837) respeitou os limites administrativos, mas suprimiu áreas não habitadas. Os limites entre as classes não foram bem delimitados e seu mapa também caracteriza a técnica dasimétrica, embora esta não fosse sua intenção (FRIENDLY e DENIS, 2005 e ROBINSON, 1955).

O geógrafo e anarquista francês Elisée Reclus (1830-1905) propôs a representação da densidade demográfica por quadrículas, bem como a representação da população urbana por círculos proporcionais, técnicas que foram aplicadas em sua obra “*Nouvelle Géographie Universelle*” (RECLUS, 1888 *apud* MARTINELLI, 2005).

A origem do termo e do método dasimétrico é citada em versões conflitantes na literatura, mas foi definitivamente esclarecida por Andrey Petrov (2008). De acordo com o autor, foi o cartógrafo russo Benjamin (*Veniamin Petrovich*) Semenov-Tian-Shansky (1870-1942), filho do famoso geógrafo e explorador Peter (*Petr Petrovich*) Semenov-Tyan-Shansky⁴ (1827-1914), quem descreveu a técnica em um texto encaminhado à “Sociedade Russa de Geografia” em 1911 (DE GEER, 1926). Muitos autores citam a origem russa do mapeamento dasimétrico mais falham em creditar corretamente o autor, atribuindo o invento ao seu pai ou mesmo citando apenas o sobrenome (BIELECKA, 2005; MAANTAY *et. al.*, 2007).

Benjamin Semenov-Tian-Shansky traduziu do idioma Russo as palavras “medida” e “densidade” para o Grego e efetuou uma transliteração de volta ao Russo, estabelecendo o termo “ ” (“*dazimetricheskiy*”) e definiu o mapa dasimétrico como aquele em que “*a densidade da população, independentemente dos limites administrativos, é distribuída na ‘realidade’, isto é, por pontos naturais de concentração e de rarefação*”(DUTENKEFER, 2010; PETROV, 2008).

A popularização do termo ocorreu com a publicação, entre 1922 e 1925, de parte da carta “*Dasymetric Map of European Russia*” (**Figura 11**). Nas décadas de 1920 e 1930 os cartógrafos russos utilizaram o termo dasimétrico como sinônimo de mapas populacionais. Posteriormente, o uso deste termo desaparece rapidamente, provavelmente devido ao cancelamento do projeto e a repressão à escola de Benjamin Semenov-Tian-Shanski pelo

⁴ Petrov(2008) esclarece que existiram cinco renomados acadêmicos russos com sobrenome Semenov-Tian-Shansky. Peter Semenov-Tyan-Shansky, pai de Benjamin, estudou em Berlim com Alexander Von Humboldt e Carl Ritter, precursores da geografia moderna.

regime stalinista em 1936. Ao todo, foram publicadas 47 das 148 folhas planejadas (PETROV, 2008).

O trabalho de Benjamin Semenov-Tian-Shansky foi resenhado por De Geer (1926) na revista “*Geographical Review*”:

“Em primeira instância o mapa distingue áreas inabitadas e esparsamente habitadas daquelas densamente povoadas. A densidade correspondente a 10 habitantes por verst quadrado foi escolhida como limite inferior da área cultivada, ou limite de população zonal contínua. O limite foi traçado desenhando círculos com um raio de um verst sobre todos os lugares preenchidos no chamado padrão cartográfico da Rússia de 10 verst mapa (1:420,000) e, suavizando os arcos coalescentes nas margens externas das áreas assim delimitadas. [Se círculos adjacentes estivessem separados por menos de dois verst, eles eram unidos. Dentro destas zonas de concentração populacional, o mapa mostra a densidade média da população.] A seleção de um raio de um verst [1,067 km] é justificada pelo fato de que os campos pertencentes a aldeias russas estão geralmente dentro de um verst do centro da vila. Pequenos grupos de casas podem ser omitidos, mas aldeias no interior das linhas são incluídas dentro da curva mais suave, mesmo separados por distâncias de dois ou três versts. Admite-se que o raio e as modificações das curvas são relativamente arbitrários, mas como o método é utilizado de maneira uniforme em toda a Rússia, espera-se os resultados sejam, pelo menos, comparáveis.”(DE GEER, 1926, p. 326 apud PETROV, 2008, p. 134. Tradução nossa)



Figura 11: Recorte do "Dasymeric Map of European Russia". Fonte: PREOBRAZHENSKEY *apud* PETROV (2008).

Dois anos depois, o próprio Benjamin publicou na mesma revista um artigo onde introduz o termo e faz uma breve discussão de seu mapa dasimétrico (SEMENOV-TIAN-SHANSKY, 1928 *apud* PETROV, 2008). Mais tarde, o geógrafo norte-americano John Kirtland Wright (1891-1969), publicou um artigo também na "Geographical Review", onde apresentou mapas de densidade populacional de *Cape Cod, Massachusetts* (WRIGHT, 1936) e exaltou as virtudes do mapa dasimétrico sobre o mapa coroplético (Figura 12). Wright declarou que dasimétrico significa "medição da densidade", indicou que o termo era de origem russa, mas não mencionou Semenov-Tian-Shansky. A ele também é atribuído o termo "coroplético" (valor por área), ainda que esses mapas já estivessem disseminados. Alguns autores (MOON e FARMER, 2001; WU *et al.*, 2005)

citam incorretamente que Wright foi o primeiro autor a publicar um artigo sobre mapeamento dasimétrico em um periódico em inglês.

Estes mapas dasimétricos tradicionais utilizavam o conhecimento especialista e cartas topográficas para auxiliar a representação cartográfica da população. A técnica nunca foi normatizada e nunca alcançou a relevância de outras técnicas da cartografia temática. A razão para esta relativa falta de popularidade pode ser atribuída à subjetividade e dificuldade envolvidas na elaboração dos mapas dasimétricos.

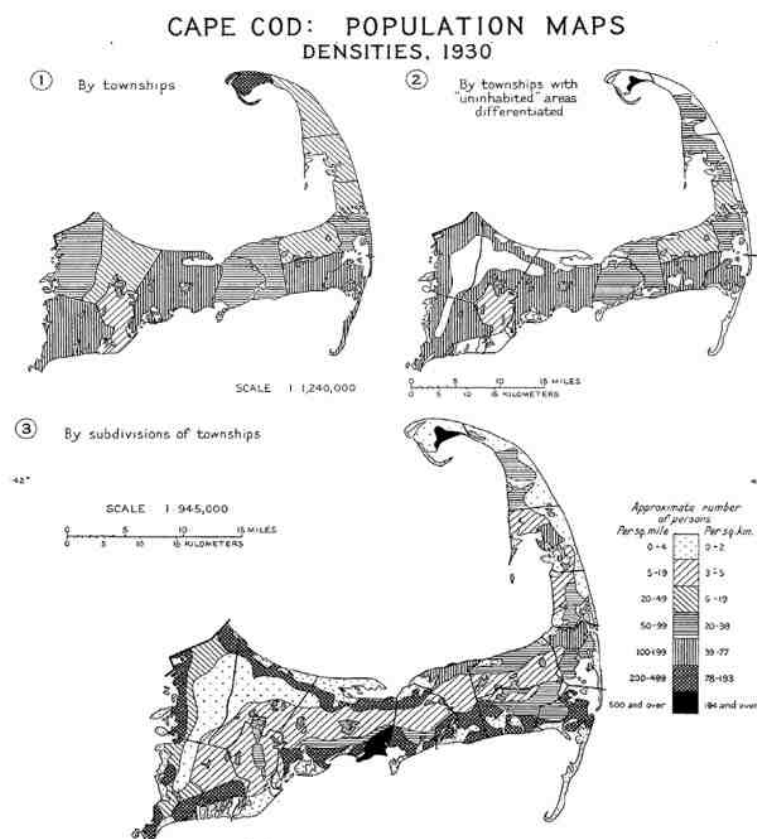


Figura 12: Mapa coroplético (1), de uso da terra (2) e dasimétrico (3) de Cape Cod, Massachusetts.
Fonte: WRIGHT (1936)

Na década de 1960 começam a surgir os primeiros computadores de grande porte (mainframes) e novas possibilidades de manipulação da informação espacial. O primeiro aplicativo para cartografia automatizada, o SYMAP (*Synergraphic Mapping System*), abriu caminho para o desenvolvimento das tecnologias de computação gráfica. Suas maiores potencialidades residiam no algoritmo para interpolação espacial de superfícies e na

produção em larga escala de mapas coropléticos “*brutos, porém informativos*” (GOODCHILD, 1988). Este sistema foi largamente utilizado para produção de mapas temáticos com dados censitários (**Figura 13**).

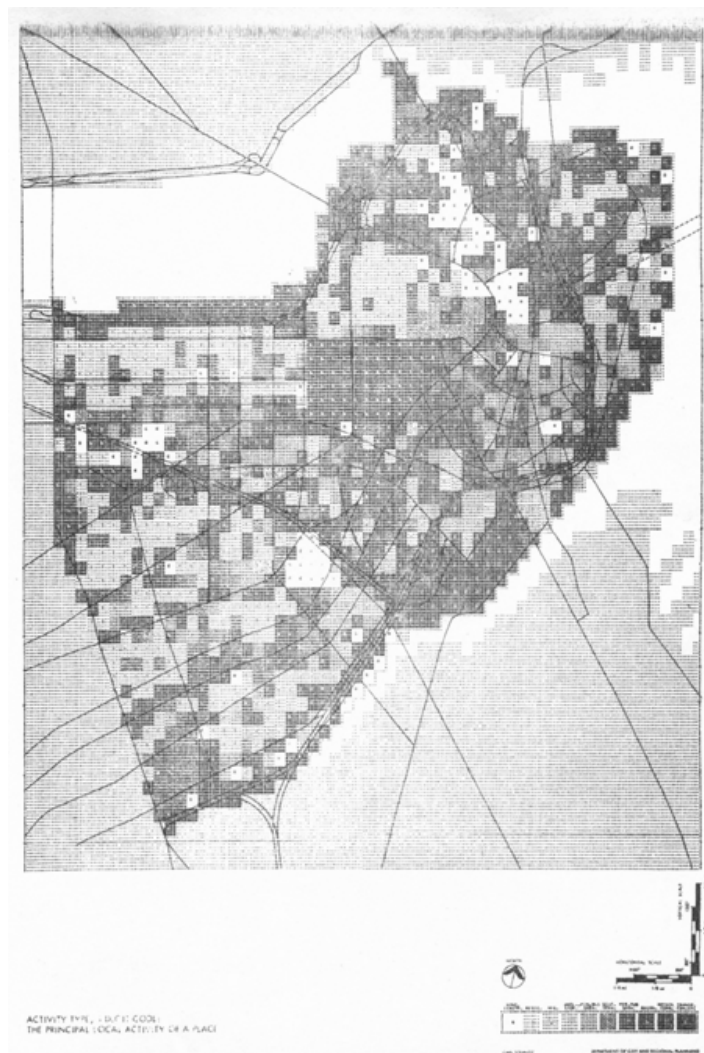


Figura 13: Reprodução do mapa coroplético “*The Principal Local Activity of a Place*”, produzido com o SYMAP por Carl Steinitz. O software utilizava caracteres ASCII para saída gráfica em impressoras de impacto em linha, único dispositivo de alta velocidade disponível á época. Fonte: (LEWIS, 2010)

Surgiram assim os primeiros Sistemas de Informações Geográficas (SIGs) ainda restritos aos *mainframes*. A partir da década de 1980, com o barateamento das estações de trabalho gráficas, além do surgimento dos computadores pessoais e dos Sistemas Gerenciadores de Bancos de Dados (SGBDs), ocorre uma grande difusão do processamento de dados geográficos em meio digital, ou geoprocessamento, termo mais difundido em nosso idioma.

A técnica coroplética tornou-se o padrão para divulgação de mapas populacionais, sobretudo devido à praticidade e facilidade de comunicação. Os mapas dasimétricos permaneceram pouco acessíveis aos usuários de SIGs, em parte devido ao alto custo e dificuldade no processamento dos dados auxiliares. Somente a partir da década de 1990, com evolução das plataformas de sensoriamento remoto, ocorre o resgate da técnica dasimétrica, que se torna uma das áreas mais férteis da cartografia digital (PETROV, 2008). Para ilustrar este fato, o **Gráfico 1** mostra o número de citações por ano do termo “*dasymetric*”, a partir do ano de 1990. No Capítulo 3 descreve-se como se deu esse resgate.

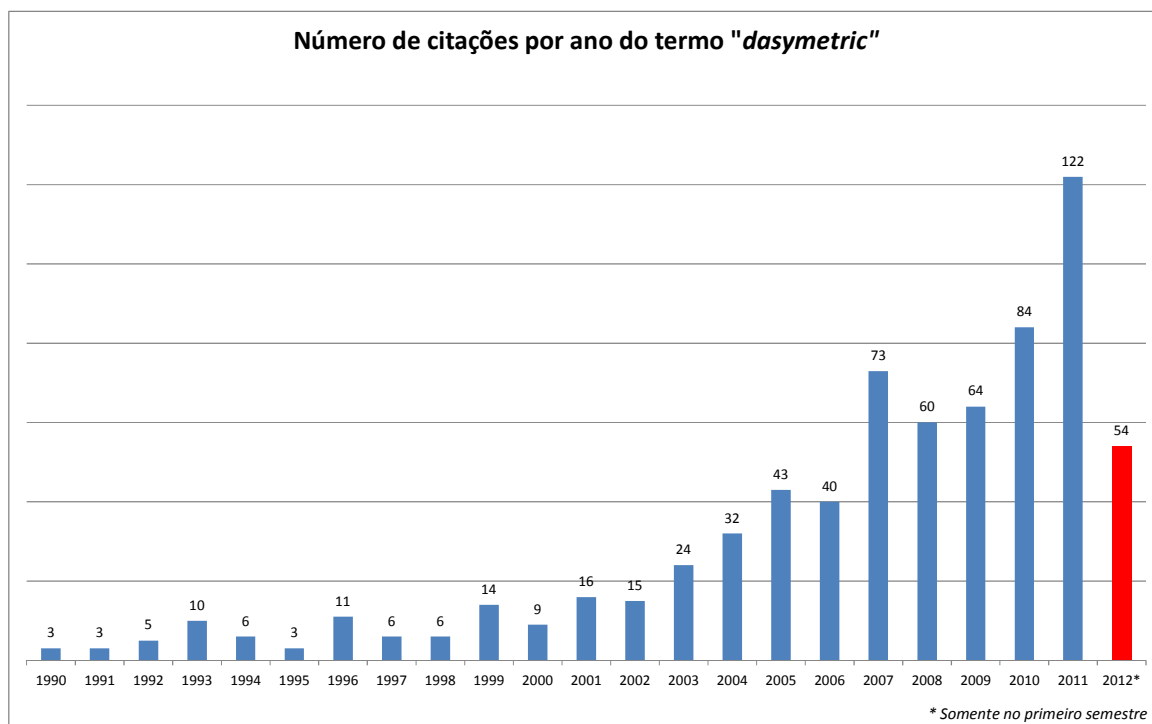


Gráfico 1: Número de citações do termo “*Dasymetric*” na base de dados do *Google Acadêmico*.

3.2. Modelos para representação digital da distribuição da população

Representar e analisar objetos do mundo real em ambiente computacional requer a adoção de um modelo, que pode ser definido como “uma construção artificial em que partes de um domínio (universo de origem) são representadas em outro domínio (universo de destino)” (WORBOYS, 1994). É uma simplificação da realidade, que inevitavelmente terá alguma limitação.

De acordo com GOODCHILD (1992), os dados geográficos podem ser modelados segundo duas visões complementares: campos e objetos. A visão de campos enxerga o mundo como uma superfície contínua, sobre a qual variam os fenômenos a serem observados segundo diferentes distribuições. Cada fenômeno é visto como uma camada contínua, não existindo nenhuma posição no espaço geográfico que não esteja associada a algum valor correspondente à variável representada. Na visão de objetos, o mundo é visto como uma superfície ocupada por objetos identificáveis, com posição, geometria e características próprias. Essas visões podem ser mapeadas nas estruturas de dados matricial ou vetorial, de acordo com a situação. Normalmente, os “*geo-campos*” são representados através de uma grade regular (estrutura matricial ou “*raster*”) e os “*geo-objetos*” são mais bem representados no formato vetorial.

Dessa forma, dados demográficos podem ser modelados de acordo com três abordagens: (a) manter os dados associados ao indivíduo ou domicílio e suas coordenadas; (b) agregar os dados em unidades de área de um zoneamento do território ou; (c) considerar a população como um fenômeno contínuo no espaço (MARTIN, 1996). As duas primeiras são realizadas através da estrutura vetorial e a última através da estrutura raster. Doravante as denominaremos abordagens pontual (a), zonal (b) e superficial (c).

A estratégia utilizada para representação espacial dos dados demográficos depende da escala da análise, da disponibilidade de dados e da finalidade da pesquisa. A disponibilidade depende não apenas da existência do dado, mas também das condições para a divulgação das informações. A abordagem pontual proporciona a maior exatidão espacial possível. Nos recenseamentos populacionais modernos, cada domicílio contado tem seu endereço armazenado junto ao questionário. Em tese, esta lista de endereços pode ser “geocodificada” e, conforme mencionado por MARTIN (1998), um censo constitui

uma representação pontual extremamente detalhada da população e suas características. Entretanto, a agregação das informações é uma forma de controlar o volume de dados e garantir a confidencialidade das informações individuais. Por isso, os dados censitários são predominantemente divulgados na forma de agregados de unidades geográficas hierárquicas, as “geografias censitárias” (MARTIN, 1998a).

Ao se adotar a abordagem zonal, assume-se implicitamente que o fenômeno estudado está uniformemente distribuído no interior das unidades do zoneamento adotado (MOON e FARMER, 2001). Este fato raramente ocorre na prática, e assim introduz-se viés nos dados, resultando em perda de detalhe e especificidade. Quanto maior a unidade, maior o viés introduzido (WU *et al.*, 2005). Este viés é responsável pela propagação de erros no cálculo de taxas e indicadores, quando as unidades de análise são menores que as originais, como ocorre em estudos de saúde, criminalidade, avaliação de riscos e impactos ambientais, entre outros (MAANTAY *et al.*, 2007).

Além disso, as geografias censitárias frequentemente não coincidem com outras regiões de análise, tais como paisagens, bacias hidrográficas e unidades administrativas de saúde, educação e segurança (MOON e FARMER, 2001). Esta “incongruência espacial” ainda constitui um obstáculo à transdisciplinariedade, sobretudo nos estudos ambientais (DEICHMANN *et al.*, 2001; VOSS *et al.*, 1999). Como as unidades são delimitadas para facilitar a coleta de dados, resultam em polígonos pequenos para as regiões urbanas e grandes nas regiões rurais. Nos mapas temáticos produzidos, a delimitação das áreas não transmite significado algum, e assim valoriza-se o tamanho dos polígonos ao invés da informação apresentada (LANGFORD e UNWIN, 1994).

Dados zonais inevitavelmente são afetados pelo “Problema da Unidade de Área Modificável” (MAUP – do inglês *Modifiable Areal Unit Problem*), conforme descrito por Openshaw e Taylor (1984). Estes autores estabeleceram que “... *as estimativas obtidas dentro de um sistema de unidades de área são influenciadas pelas diversas maneiras que estas unidades podem ser agrupadas e, portanto, ao se modificarem os limites e escala de agregação, se afeta significativamente o resultado*”. Por conseguinte, muitas vezes é difícil distinguir se os resultados das análises de dados censitários refletem alguma realidade sobre a população ou se são função da unidade de área utilizada na análise (MENNIS, 2002).

O termo “falácia ecológica” utilizado nas ciências sociais e na epidemiologia consiste estabelecer inferências sobre comportamentos individuais a partir de dados observados em grupos (agregados). ROBINSON (1950, 2009) demonstrou com dados do censo norte-americano de 1930 que o coeficiente de correlação entre duas variáveis geralmente difere quando observadas no nível individual e no nível ecológico (agregado). O “*viés agregacional*” tende a aumentar a correlação entre variáveis e diminuir flutuações (OPENSHAW, 1984).

Uma alternativa para contornar as limitações da abordagem zonal consiste em considerar a população como um fenômeno contínuo no espaço e representá-la através de grades regulares ou superfícies estatísticas. As grades são repositórios de dados espaciais geralmente com dimensões fixas, que apresentam algumas vantagens como estabilidade temporal, escalabilidade, análise independente de área pré-fixada e, a possibilidade da aplicação de técnicas analíticas específicas para superfícies (LANGFORD e UNWIN, 1994; MENNIS, 2003). Utilizar o modelo superficial permite a seleção da unidade de análise com base em considerações teóricas, sem a limitação das geografias censitárias. O pesquisador pode estabelecer limites de acordo com sua necessidade, o que é particularmente interessante para analisar fenômenos que são inapropriadamente representados em dados agregados.

Nesse contexto, os autores Bracken e Martin (1989) utilizaram pela primeira vez os SIGs para produzir superfícies populacionais do Reino Unido, a partir de dados agregados na menor unidade espacial disponível do censo britânico de 1981. Em seguida, adotaram a mesma abordagem para compatibilizar dados dos censos de 1981 e 1991 (BRACKEN e MARTIN, 1995). Desde então, algumas iniciativas surgiram para disponibilizar dados demográficos através de superfícies populacionais (Tabela 1).

Grande progresso tem sido obtido por pesquisadores e organizações na criação de bases de dados populacionais na forma de superfícies. Tobler *et al.*(1997) desenvolveram a primeira versão do *Gridded Population of the World* (GPW), atualmente um projeto do *Center for International Earth Science Network* (CIESIN) da Universidade de Columbia, nos Estados Unidos da América (EUA), assim como o *Global Rural Urban Mapping Project* (GRUMP) que incorpora a extensão dos assentamentos rurais e urbanos ao GPW versão 3 (GPWv3), com resolução espacial de 2,5’ por 2,5’ (minutos) (CIESIN, 2004).

O projeto *LandScan Global Population Databases*, desenvolvido pelo *Oak Ridge National Laboratory* (ORNL), fornece superfícies da população mundial com resolução espacial 1 km, a mais detalhada até o momento (BHADURI *et al.*, 2007). Já o *LandScan USA* fornece superfícies de população noturna (residencial), diurna ou ambiente (média entre diurna e noturna) com resolução espacial de aproximadamente 90 metros para a região continental dos Estados Unidos (BHADURI *et al.*, 2007), conforme ilustrado na **Figura 14**. Os dados do programa são fornecidos gratuitamente para órgãos de governo dos Estados Unidos e comercializados para demais usuários. Os valores de licenciamento são avaliados caso a caso, conforme informação publicada no sítio do projeto (ORNL, 2012).

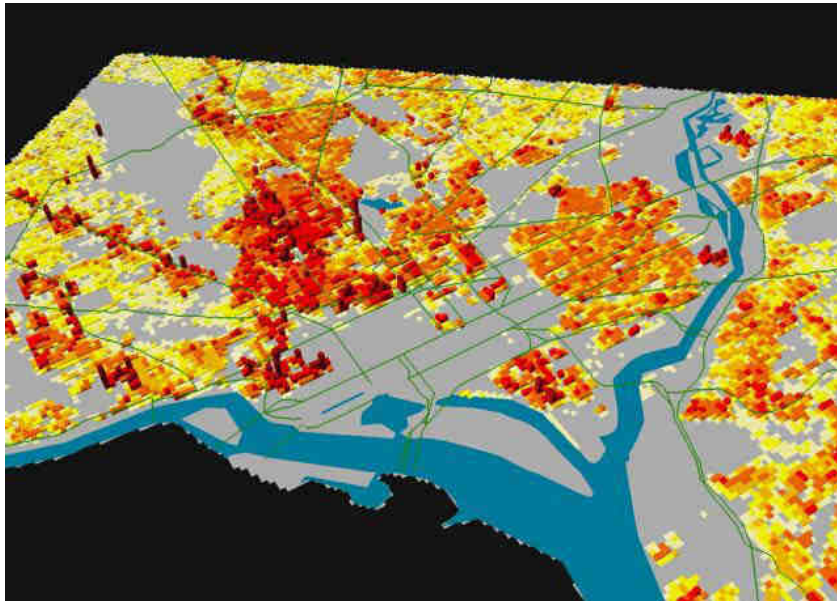
Tabela 1: Bases de dados demográficos modeladas na forma de superfícies.

Base de Dados	Cobertura	Resolução Espacial	População	Método	Ano de Referência	Política de Distribuição
UNEP	África	5 km	Residente	Interpolação inteligente	1990	Domínio Público
GPW2.0	Global	5 km	Residente	Ponderação zonal	1995	Domínio Público
GPW3.0	Global	5 km	Residente	Ponderação zonal	2000	Domínio Público
GPW3.0UR	Global	1 km	Residente	Interpolação dasimétrica	2003	Domínio Público
LandScan Global	Global	1 km	Ambiente	Interpolação multidimensional	1998	Restrita /Comercial
LandScan USA	Estados Unidos	90 m	Residente e diurna	Interpolação multidimensional	2000	Restrita
Gridded Population Database of China	China	1 km	Residente	Interpolação multidimensional	1995 e 2000	Restrita

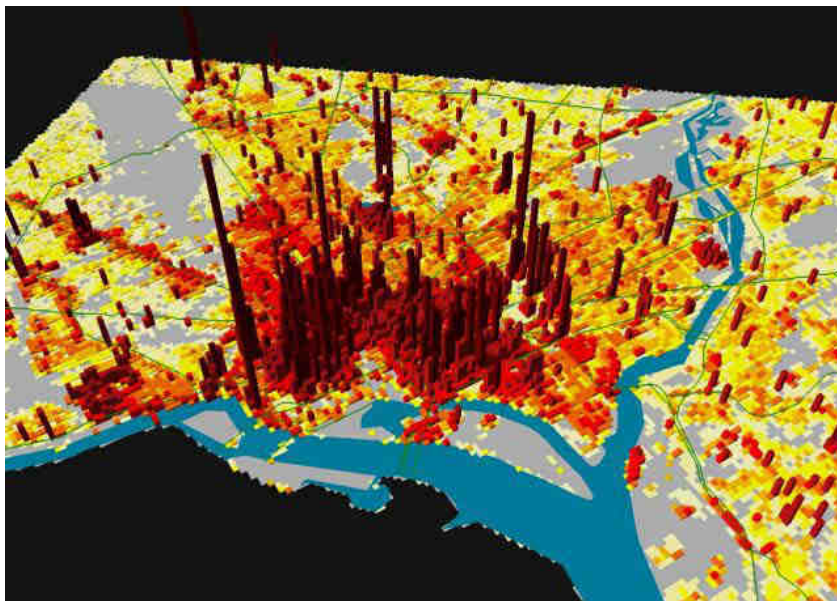
Fonte: UNEP (URL: <http://grid2.cr.usgs.gov/datasets/datalist-php3>) (Deichmann 1996), GPW2.0 (URL:

<http://sedac.ciesin.columbia.edu/plue/gpw/>) (Deichmann et al. 2001), GPW3.0 (Balke Yetman 2004), GPW3.0UR (Balk e Yetman 2004), LandScan (URL: <http://www.ornl.gov/sci/gist/landscan/>) (Dobson et al. 2000,2003 Bhadhuri, 2007) e GPDC (YANG *et al.*, 2009) Adaptado de Hay et al. 2005

As metodologias adotadas nesses projetos empregam técnicas que evoluíram desde a ponderação zonal simples (sem uso de dados auxiliares) utilizada nas versões iniciais da GPW, até o interpolador dasimétrico multivariado adotado no programa *LandScan*, baseado em uma variedade de informações auxiliares, como relevo, hidrografia, malha viária, luzes noturnas. A avaliação destas superfícies, quanto à utilidade em aplicações como epidemiologia e defesa civil foram realizadas em algumas pesquisas (HALL *et al.*, 2012; HAY *et al.*, 2005; SABESAN *et al.*, 2007).



(a)



(b)

Figura 14: Visualização da população noturna (a) e diurna (b) da cidade de Washington a partir de dados do Programa *LandScan USA* do *Oak Ridge National Laboratory*, com resolução espacial de 90 metros. Fonte: (BHADURI *et al.*, 2007)

4. MÉTODOS DE ESTIMATIVA POPULACIONAL EM PEQUENAS ÁREAS

Muitas áreas de aplicação da informação demográfica requerem dados em escala espacial e/ou temporal mais detalhada do que as disponibilizadas pelos censos. Essas informações podem ser estimadas através de modelos demográficos. Entretanto, conforme observado por Qiu *et. al.* (2010), o emprego de tais modelos envolve considerável conhecimento especialista em análise demográfica e, a qualidade das estimativas depende fortemente da qualidade dos dados e do conhecimento do desempenho do modelo em situações anteriores, nem sempre disponíveis. Portanto, os modelos demográficos para estimativa em pequenas áreas possuem maior aplicação na área de projeções populacionais.

Como alternativa, diversos métodos de estimativa populacional foram propostos na literatura de geoprocessamento e sensoriamento remoto. Em uma extensa revisão dos métodos até então descritos, Wu *et.al.*, (2005) os classificaram em duas grandes categorias: interpolação zonal e modelagem estatística.

A interpolação zonal é tradicionalmente utilizada como solução para a transferência de valores de uma variável entre diferentes conjuntos de unidades espaciais, e, para tanto pode ou não fazer uso de dados auxiliares. Já a modelagem estatística busca estabelecer uma estimativa com base na relação entre população e outras variáveis, utilizando obrigatoriamente informações auxiliares.

4.1. Interpolação Zonal

Em matemática, denomina-se interpolação o método que permite construir um novo conjunto de dados a partir de um conjunto discreto de dados pontuais previamente conhecidos, normalmente obtidos a partir de amostragem. Métodos de interpolação espacial são essenciais para a estimativa de variáveis biofísicas e socioeconômicas em diversas disciplinas. São especificamente desenvolvidos para determinado tipo de dado ou variável (LI e HEAP, 2008).

Interpolação zonal é um tipo de interpolação espacial aplicado para estimar valores de uma ou mais variáveis em um conjunto de unidades espaciais (zonas de destino) a partir de valores existentes em outro conjunto de unidades (zonas de origem) na mesma área de estudo (LAM, 1982). É utilizada para tratamento e análise de dados de diferentes fontes, coletados ou agregados em zoneamentos distintos, assim como para a desagregação espacial, ou seja, para decomposição de uma variável de zonas maiores em zonas menores. Como exemplo, sejam as seguintes situações, para as quais a interpolação zonal pode prover resultados mais exatos na análise de dados demográficos:

- Uma empresa possui dados referentes à suas unidades operacionais e precisa conhecer o total de pessoas que vivem sob a área de influência destas unidades.
- Um órgão de Defesa Civil necessita estimar o número de pessoas potencialmente atingidas por uma catástrofe natural, como uma enchente.
- Um cientista social quer analisar uma série temporal de dados agregados por setores censitários, mas as malhas de setores possuem diferentes delimitações entre os censos.
- Um órgão de planejamento necessita calcular o tempo de deslocamento da população aos serviços de saúde ou educação.

Existem muitos métodos de interpolação zonal e cada um deles está baseado pressupostos assumidos em relação à distribuição espacial do fenômeno analisado. A escolha de um método deve ser realizada observando-se o tipo de dado utilizado, a facilidade de implementação, a exatidão desejada, o esforço computacional e a disponibilidade de dados e tempo (HAWLEY e MOELLERING, 2005). Nenhum método garante resultados totalmente precisos, pois sempre há alguma limitação nas hipóteses assumidas. A exatidão da estimativa depende de fatores como a qualidade dos dados originais, os pressupostos assumidos sobre a distribuição original dos objetos e a relação espacial imposta no processo de interpolação (LAM, 1982).

Os métodos de interpolação zonal podem ser classificados por diferentes critérios. Rase (2001) propôs uma classificação baseada em cinco categorias de acordo com o propósito da operação: agregação, desagregação, compatibilização, suavização e para criar superfícies contínuas a partir de pontos. Entretanto, a classificação mais adotada na

literatura foi proposta por Wu *et al.*, (2005). Segundo esses autores, os métodos distinguem-se inicialmente entre os que não utilizam e os que utilizam dados auxiliares. Os métodos que utilizam dados auxiliares são denominados dasimétricos (**Figura 15**).

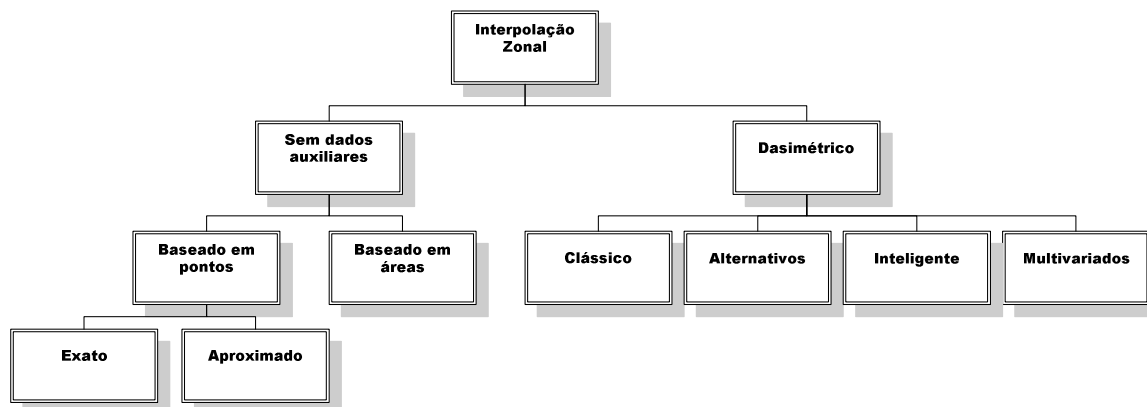


Figura 15: Classificação dos métodos de interpolação zonal. Fonte: Adaptado de (SILVA, 2009; WU *et al.*, 2005)

4.1.1. Interpolação zonal sem Informação auxiliar

Os métodos sem uso de informações auxiliares foram os primeiros desenvolvidos em função da necessidade de manipulação de dados em sistemas computacionais. Podem ser baseados em pontos ou em áreas. A interpolação zonal baseada em pontos não utiliza integralmente as zonas originais. Uma etapa preliminar consiste em transformá-las em um conjunto de pontos representativos, como centroides. Em seguida, cria-se uma grade regular e estimam-se os valores dos pontos da vizinhança com base nos valores conhecidos. Os métodos exatos priorizam a preservação dos valores originais, enquanto outros priorizam o ajuste se uma função global para a área, sem preservar obrigatoriamente os valores originais (SILVA, 2009; WU *et al.*, 2005).

Os métodos baseados em pontos apresentam sérias deficiências no trato de variáveis socioeconômicas. A seleção arbitrária de pontos representativos não considera a complexa geometria das unidades de enumeração. Além disso, esses métodos assumem pressupostos em relação à superfície interpolada que muitas vezes não se adequam ao complexo fenômeno geográfico subjacente, como no caso dos métodos ponderados pela distância que descrevem uma variável através de uma função determinística (WU *et al.*, 2005). Dentre as várias técnicas de interpolação baseadas em ponto disponíveis, a krigagem é o melhor

estimador linear não enviesado, caso a distribuição da população possa ser descrita pelo semi-variograma (MORAL, 2003 *apud* SILVA, 2009)

Para compensar os problemas do mapeamento censitário do Reino Unido, Martin (1989) desenvolveu uma técnica conhecida como método dos centroides para produzir superfícies de densidade populacional com um algoritmo *kernel*, que realiza uma ponderação baseada na função de decaimento da distância entre o centroide da zona de origem e célula da grade. O algoritmo de Martin foi amplamente usado no Reino Unido para especializar dados censitários. A formulação básica do seu modelo é dada por:

$$\hat{P}_i = \sum_{j=1}^C P_j \cdot W_{ij}$$

Equação 1: Método dos centroides ponderados

Onde:

\hat{P}_i = população estimada para a célula da grade i ;

P_j = a população associada ao centroide j ;

C = número de centroides na área mapeada,

W_{ij} = peso associado á distância entre i e j , especificado por uma função de decaimento da distância f_j que considera a dispersão dos centróides na janela kernel, determinado por:

$$W_{ij} = f_j \left(\frac{d_{ij}}{\sum_{l=1}^k [\bar{d}_{jl}]} \right)$$

Equação 2: Pesos para os centroides ponderados

Onde:

f_j = é a função de decaimento da distância, apropriadamente determinada;

\bar{d}_{jl} = é a distância do centroide j aos centroides l ($l=1,2,..;k$) no interior da janela centrada em j ;

d_{ij} = é a distância entre o ;

Uma propriedade desejável dos métodos de interpolação zonal é a preservação de volumes, ou propriedade picnofilática (TOBLER, 1979). Quando o método atende a esta propriedade, a soma dos valores estimados nas zonas de destino se mantem igual ao valor da zona original correspondente. A preservação de volume garante que “*pessoas não sejam*

destruídas ou fabricadas no processo de interpolação” (LANGFORD e UNWIN, 1994), e, portanto, constitui uma característica essencial para obtenção de estimativas populacionais confiáveis.

Os métodos baseados em área são mais eficientes na preservação de volume, ou seja, o somatório do valor da variável nas zonas de destino mantém-se igual ao valor da variável na zona de origem que as contém. A ponderação zonal simples, também conhecida como método “*overlay*” (LAM, 1983), estima uma variável com base na relação espacial das zonas de origem e destino, de acordo com a **Equação 3**. Cada zona de origem contribui com uma fração correspondente ao percentual da área de interseção com as zonas de destino correspondentes (GOODCHILD e LAM, 1980). Seja a zona coroplética de origem *o* e a zona de destino *d*:

$$\hat{P}_d = \sum_{o=1}^n P_o \left[\frac{A_{ond}}{A_o} \right]$$

Equação 3: Ponderação zonal simples

Onde:

\hat{P}_d = valor estimado da variável (população) na zona de destino;

P_o o valor da variável em *o*

A_{ond} = a área de interseção entre *o* e *d*;

A_o é a área da zona de origem; *n* o número *o* que intersectam *d*

Quando cada zona *d* intersecta apenas uma zona de *o*, a equação pode ser simplificada para:

$$\hat{P}_d = \frac{P_o A_d}{A_o}$$

Equação 4: Simplificação da ponderação zonal simples

A maioria dos aplicativos SIGs possui algoritmos para interpolação zonal baseado na ponderação zonal simples, que tem como vantagens a facilidade de execução e a utilização integral das zonas de origem, mas o inconveniente de assumir uma distribuição uniforme da variável no interior das mesmas. A principal deficiência da ponderação zonal simples é o fato de assumir distribuição homogênea da variável no interior das zonas de origem, o que raramente acontece na realidade geográfica.

A interpolação picnofilática, proposta por Waldo Tobler (1979), assume a existência de uma função suave de densidade, e considera o efeito de zonas de origem adjacentes (LAM, 1983). Seu método superpõe uma grade regular ao zoneamento e atribui a cada célula desta grade (referenciadas por x e y) o valor da variável z na zona de origem dividido pelo número de células no seu interior. O valor obtido é então suavizado pela média de seus vizinhos, comparado com os valores z originais e ajustado para atender a condição picnofilática (ou preservação de volume) definida da seguinte forma:

$$H_i = \iint_{R_i} Z(x,y) dx dy$$

Equação 5: Interpolação picnofilática de Tobler

Onde:

R_i é a *iésima* região (célula da grade);

H_i é o valor inteiro não negativo da variável ;

$Z(x,y)$ é a função de densidade.

Este procedimento iterativo continua até que não exista diferença significativa entre os valores de z original e estimado dentro das zonas de origem, ou até que não ocorram alterações significativas nos valores das células em relação à interação anterior. O método assume distribuição heterogênea da variável z na zona de origem e apresenta as vantagens de preservar o somatório da variável dentro da zona de origem e considerar também os valores de áreas adjacentes. A geração de superfícies de densidade a partir do interpolador picnofilático de Tobler promove a suavização dos limites entre as unidades, semelhante ao efeito da aplicação de um filtro de mediana (**Figura 16**).

Tobler (1979) ilustrou a aplicação de seu método picnofilático utilizando a seguinte analogia:

“Imagine que o histograma bivariado apresentado em perspectiva é constituído por blocos de argila. Cada estado é representado por um bloco, com cores diferentes e massa proporcional a sua população. Deseja-se então esculpir esses blocos até que constituam uma superfície totalmente suavizada, mas sem remover ou adicionar qualquer quantidade de argila, sem permitir que nenhuma partícula mova-se de um estado para outro” (TOBLER, 1979 p. 521 tradução nossa).

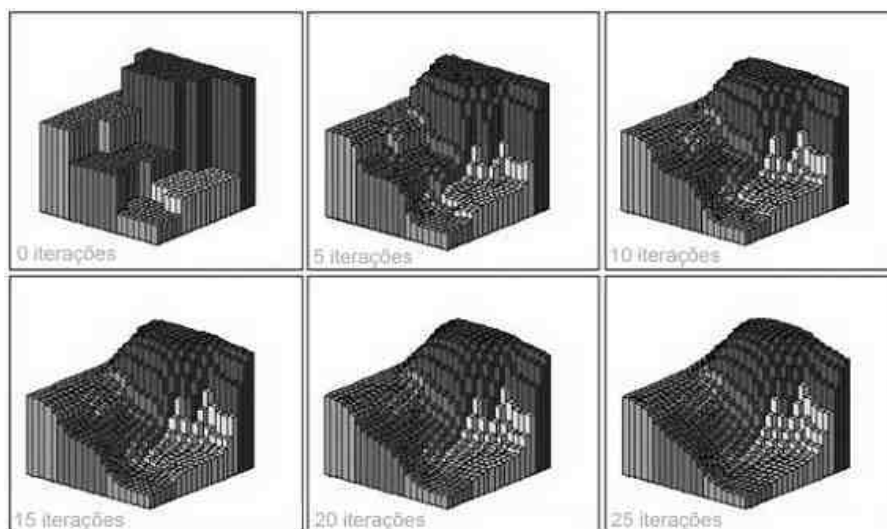


Figura 16: Ilustração das iterações do método picnofilático. Fonte: AMARAL (2003)

O método picnofilático foi revisitado em diversos trabalhos. Por exemplo, Rase (2001) incorporou as Redes Triangulares Irregulares (TIN – do inglês *Triangular Irregular Network*) ao método. Como os TINs utilizam os mesmos vértices que delimitam as zonas de origem, evitam o erro que ocorre na conversão destas para grades regulares.

4.1.2. Mapeamento dasimétrico ou interpolação dasimétrica

Graças a difusão dos SIGs e a disponibilidade de dados de sensoriamento remoto, os princípios do método dasimétrico de Semenov-Tian-Shansky (1928) puderam ser aplicados ao problema da interpolação zonal com dados digitais. Um mapa dasimétrico tem zonas mais homogêneas, derivadas da intersecção entre as zonas de origem e auxiliares, permitindo ao leitor uma melhor compreensão da distribuição da variável quando comparado aos mapas coropléticos. A interpolação dasimétrica possui diversas abordagens e pode ser realizada com diferentes fontes de informação auxiliar. Os métodos convencionais são baseados na informação sobre o uso e cobertura da terra.

O método dasimétrico binário é o mais simples de todos (LANGFORD *et al.*, 1991). Consiste em separar as classes de uso da terra em relação à presença ou ausência de população para redistribuí-la somente entre as classes habitáveis. Fischer e Langford (1995) observaram, através da simulação de Monte Carlo, que esse

método apresentou melhor desempenho quando comparado com a ponderação zonal simples e outros três métodos baseados em regressão estatística. O método binário é bastante utilizado, muitas vezes combinado com outros métodos. Sua maior deficiência é a incapacidade de considerar as variações de densidade populacional dentro das áreas habitadas.

Eicher e Brewer (2001) avaliaram o método binário e propuseram duas outras abordagens, o método das três classes e o método da variável limitante. O primeiro é uma extensão do método binário que redistribui percentualmente a população de uma zona de origem entre as classes de uso da terra nela contidas, por exemplo: 70% para áreas urbanas, 20% para áreas agrícolas, 10% para floresta e 0% para massa d'água. Sua deficiência consiste em não considerar a área relativa das classes de uso/cobertura da terra dentro de cada zona de origem, o que influencia significativamente as razões relativas entre as densidades resultantes.

O método da variável limitante estabelece limiares de densidade populacional para algumas das classes e, no caso a densidade de alguma zona de origem exceder este limiar, o excesso é redistribuído para as classes que não tiveram limiar definido. O método da variável limitante possui a vantagem de considerar a área relativa de cada classe auxiliar dentro da zona de origem.

Eicher e Brewer (2001) verificaram através da raiz do erro quadrático médio (ou RMS, do inglês *Root Mean Square*) que o método da variável limitante produziu resultados significativamente mais exatos que o método binário e método das três classes. Os autores também verificaram diferenças nos resultados em relação ao tipo de estrutura de dados utilizada (vetorial ou matricial). Os mesmos métodos apresentaram erros menores quando desenvolvidos usando uma estrutura vetorial, embora a diferença não fosse significativa.

A aplicação correta, tanto do método das três classes quanto da variável limitante, requer que percentagens (para método das três classes) ou densidades (para método da variável limitante) sejam adequadamente especificadas. Eicher e Brewer (2001) definiram subjetivamente as percentagens em sua demonstração do método de três classes, mas analisaram empiricamente os dados para especificar limiares na abordagem da variável limitante. Nesse último, as densidades populacionais das zonas censitárias que estavam

completamente dentro de uma categoria específica de cobertura da terra foram usadas para definir a densidade máxima para essa categoria.

MENNIS (2003) aperfeiçoou o método das três classes adicionando a amostragem empírica para estimar percentagens adequadas e ajustando as contagens finais com base na área relativa de cada classe auxiliar na zona de origem. Seu método, alcunhado de *Mapeamento Dasimétrico Inteligente* (IDM), utiliza uma estimativa do percentual da densidade populacional em cada classe auxiliar, que pode ser pré-definida pelo analista ou obtida através de amostragem.

Em seguida, Mennis e Hultgren (2005) desenvolveram o algoritmo IDM, em linguagem *Visual Basic for Applications* para o ambiente de geoprocessamento do software *ArcGIS*. O código permite que o usuário especifique os parâmetros de amostragem de acordo com critérios de localização das zonas de origem e destino. Além disso, o usuário pode optar por ignorar a amostragem, e atribuir um valor pré-determinado para uma classe (ex. 0% para massa d'água).

Na última década, muitos artigos foram publicados com referência ao método multiclasse proposto por Mennis (2003). A partir do primeiro código em domínio público, outros *scripts* foram publicados em diferentes linguagens de programação, tanto para estrutura de dados vetorial (SLEETER e GOULD, 2007; WEBER, 2010) como para matricial (HULTGREN, 2004). Através desses *scripts*, as etapas envolvidas na produção do mapa dasimétrico são automatizadas com o módulo de geoprocessamento do *ArcGIS*. Funções como *Intersect*, *Combine* e *Calculate Field* são sequenciadas para realização dos procedimentos. As etapas básicas envolvidas na confecção de um mapa dasimétrico através IDM podem ser resumidas de acordo com o diagrama apresentado na **Figura 17**.

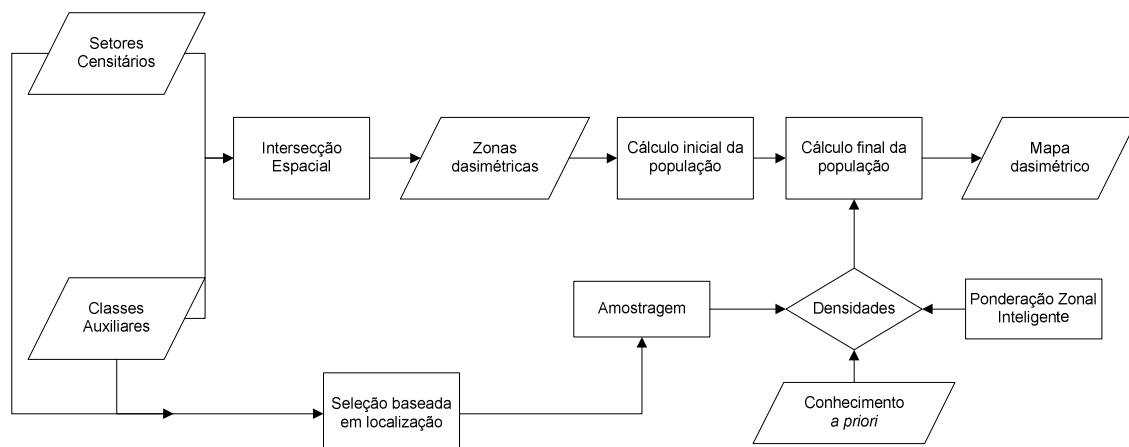


Figura 17. Fluxo básico para produção de um mapa dasimétrico através do IDM.

Zonas dasimétricas são criadas pela intersecção espacial entre setores censitários e classes auxiliares. Em seguida calcula-se a população das zonas dasimétricas, em duas etapas. Uma estimativa preliminar é obtida multiplicando-se a área da zona pela densidade da classe auxiliar, que pode ser definida *a priori* ou estimada por amostragem. Os valores obtidos são ajustados para garantir que a soma da população das zonas dasimétricas seja igual a população da zona original (propriedade picnofilática), além de manter a proporção da contagem inicial.

A premissa básica do IDM é expressa da seguinte forma:

$$\hat{P}_s = P_o \left(\frac{A_t \bar{D}_c}{\sum (A_t \bar{D}_c)} \right)$$

Equação 6: Interpolação inteligente do IDM (Mennis, 2003)

Onde P_o é a população observada, \hat{P}_s é a população estimada, s é a zona coroplética, t é a intersecção entre s e a classe auxiliar c ($t = o \cap c$), \bar{D}_c é a densidade estimada para a classe c , que pode ser arbitrariamente definida ou estimada amostrando-se as zonas s parcial ou completamente contidas em uma classe c .

Para cada classe de uso e cobertura da terra, o IDM realiza um procedimento de amostragem para selecionar zonas censitárias que possam ser representativas dessa categoria e a densidade populacional é estimada pela Equação 7. Existem três opções disponíveis para seleção de amostras, com base na relação espacial entre as unidades (MENNIS; HULTGREN, 2005, p. 181), conforme ilustrado na **Figura 18**:

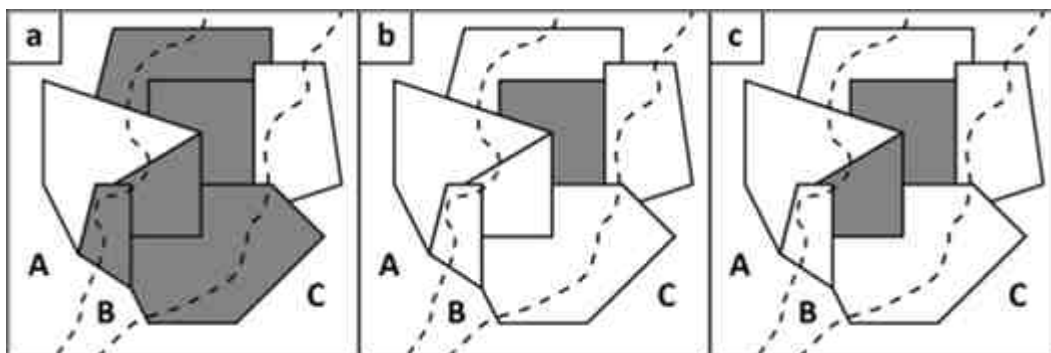


Figura 18: Métodos de amostragem do algoritmo IDM. As zonas censitárias em cinza selecionadas para a classe auxiliar *B*, em cada uma das opções acima. Fonte: (WEBER, 2010).

- Centroide contido:* A unidade censitária é considerada representativa caso seu centroide esteja contido no interior da classe auxiliar
- Inteiramente contido:* A unidade censitária é considerada representativa caso esteja inteiramente contida na classe auxiliar.
- Percentual de cobertura:* A unidade do censo é considerada representativa caso esteja parcialmente contida na classe auxiliar. O usuário especifica um limiar para definir a seleção, por exemplo, 90%.

Além disso, o usuário pode optar por ignorar o procedimento de amostragem e especificar valores de densidade para uma ou mais classes auxiliares, baseado no conhecimento *a priori* da área de estudo. Dessa forma, é possível também realizar uma filtragem binária, uma vez que se pode definir densidade zero para as classes que não possuem ocupação, como hidrografia.

Uma vez que uma zona coroplética tenha sido selecionada como representativa de uma classe *c*, a densidade dessa classe é estimada pela **Equação 7**:

$$\hat{D}_c = \frac{\sum_{s=1}^m P_{cs}}{\sum_{s=1}^m A_{cs}}$$

Equação 7: Densidade percentual estimada para o IDM

Onde *m* é número de zonas coropléticas associadas com a classe auxiliar *c*.

É possível que, para uma ou mais classes auxiliares, nenhuma zona censitária representativa seja selecionada na amostragem. Quanto mais seletivo o critério de amostragem selecionado ou quanto maior o número de classes, maior a probabilidade que isso ocorra. Nesse caso, a população dessa classe (denotada por *u*) é temporariamente

calculada pelo que Mennis e Hultgren (2006) chamaram de *Ponderação Zonal Inteligente* (Equação 8).

$$\hat{P}_{d(u)} = \left[P_o - \sum_{z(k) \in z} (\hat{D}_k A_{z(k)}) \right] \left[\frac{A_{d(u)}}{\sum_{z(u) \in z} A_{d(u)}} \right]$$

Equação 8:
Ponderação
Zonal Inteligente

Onde:

$\hat{P}_{d(u)}$ = população estimada para a classe auxiliar u

\hat{D}_k = densidade estimada da classe amostrada k ;

$A_{d(k)}$ = área da zona dasimétrica associada com a classe k ; e

$A_{d(u)}$ = área da zona dasimétrica associada á classe não amostrada u ;

Uma vez calculada a população inicial, a densidade estimada da classe u é estimada pela seguinte fórmula:

$$\hat{D}_u = \frac{\sum_{z(u)=1}^p \hat{P}_{d(u)}}{\sum_{z(u)=1}^p A_{d(u)}}$$

Equação 9: Estimativa
da densidade das
classes não
amostradas

Onde:

\hat{D}_u = densidade estimada da classe u ; e

p = número de zonas dasimétricas associadas com a classe u .

Esse procedimento é realizado para todas as classes não amostradas e a população é recalculada para todas as zonas no mapa, considerando as novas densidades estimadas para as classes não amostradas. Caso não sejam obtidas amostras para nenhuma classe auxiliar, e nenhuma densidade seja informada pelo usuário, o algoritmo do IDM redistribui as contagens através da ponderação zonal simples (Equação 3), que não caracteriza uma interpolação dasimétrica.

O IDM foi utilizado em diversos trabalhos. Por exemplo, Sleeter (2004) o utilizou para produzir um mapa dasimétrico da região da Baía de São Francisco, na Califórnia. Wang (2012) utilizou para melhorar a visualização cartográfica da oferta de empregos em uma região dos EUA. Além disso, o IDM tem sido integrado com outros métodos e fontes de dados numa abordagem “multidimensional”, como nos projetos *LandScan USA*

(BHADURI *et al.*, 2007) e no *Gridded Population Database of China* (YANG *et al.*, 2009).

Uma importante contribuição ao IDM foi dada por Eric Weber (2010), que organizou o código em módulos e incorporou uma função para arredondar as contagens dasimétricas para valores inteiros e preservando a propriedade picnofilática. Além disso, incluiu um módulo para testar e comparar mapas produzidos com diferentes parâmetros de amostragem. Como não havia limite quanto ao número de classes utilizadas, o autor concluiu que o método de Mennis seria mais bem denominado como método Multiclasse Corrigido pela Área (do inglês, *Area Corrected MultiClass* - ACMC).

4.1.3. Outras abordagens e fontes de dados utilizados no mapeamento dasimétrico

Os métodos dasimétricos convencionais utilizam classificações de uso e cobertura da terra. Dobson *et al.* (2000) verificaram que mesmo com baixa resolução espacial (~250m) essa informação pode ser considerada um bom indicador da distribuição espacial da população. Em muitos dos artigos publicados são utilizadas bases nacionais ou regionais de uso e cobertura da terra, que possibilitam a produção de dados dasimétricos em escala cartográfica de semidetalhe ou de reconhecimento (geralmente de 1:100.000 a 1:250.000) para amplas áreas do território.

Entretanto, a classificação de imagens orbitais de média resolução espacial (15-30m) é pouco eficiente na distinção dos diferentes padrões de uso da terra em áreas urbanas. Até mesmo imagens de alta resolução espacial (na ordem de um metro) apresentam dificuldades para interpretação do uso dado a essas áreas, como residencial, industrial, comercial e áreas desenvolvidas não habitadas (MAANTAY *et al.*, 2007).

Como consequência, os métodos convencionais são eficientes para redistribuir população em áreas rurais, mas pouco eficientes em áreas urbanas. Dessa forma, outros tipos de informação auxiliar, assim como outras abordagens metodológicas, foram aplicados ao problema da interpolação dasimétrica.

Para contornar deficiências dos métodos dasimétricos em áreas urbanas, os autores Maantay, Maroko e Herrmann (2007) desenvolveram um sistema dasimétrico cadastral (CEDS - *Cadastral Expert Dasymetric System*), para desagregar dados censitários ao nível

de lote fiscal na cidade de Nova Iorque. Esse tipo de informação geralmente é obtido e atualizado por prefeituras e governos locais como forma de possibilitar a cobrança de impostos.

Reibel e Bufalino (2005) desenvolveram um algoritmo ponderado pela densidade da malha viária (**Equação 10**) a partir de arquivos vetoriais com eixos de logradouros TIGER (*Topologically Integrated Geographic Encoding and Referencing*) fornecidos pelo *U.S. Census Bureau*. O algoritmo proporcionou redução de erro (em relação á ponderação zonal) na ordem de 70% para a variável total de domicílios, e 20% para a variável população total. De acordo com os autores, a população pode ser estimada em qualquer segmento de logradouro i a partir das seguintes equações:

$$W_i = \frac{L_i}{L_o}$$

Equação 10: Pesos para ponderação por rede viária

$$P_i = W_i \times P_o$$

Equação 11: População estimada em cada segmento de logradouro

Onde,

W_i é o peso atribuído ao segmento de logradouro i ;

L_i é a extensão do segmento i ;

L_o é a soma da extensão de todos os segmentos de rua da unidade; e

P_i é a população no segmento.

Também se pode estimar população em outro zoneamento do espaço através da Equação 12:

$$W_{ond} = \frac{L_{ond}}{\sum L_o}$$

Equação 12: Ponderação por rede viária.

Onde:

W_{ond} = peso atribuído a cada fragmento da interseção entre as zonas de origem e destino o e d ;

L_{ond} é a extensão de cada segmento de rua na zona de interseção entre o e d ;

L_o a extensão de cada segmento de logradouro na zona o .

Como utiliza somente dados vetoriais, o método ponderado pela densidade viária é acessível a usuários dos SIGs, como analistas de planejamento e cientistas sociais. Entretanto, o modelo apresentou melhor desempenho em áreas urbanas densamente povoadas, mas mostrou-se pouco eficiente em áreas rurais (REIBEL e BUFALINO, 2005; TAPP, 2010).

Recentemente, alguns estudos exploraram o uso de cadastros digitais de endereços para interpolação dasimétrica (BAKER *et al.*, 2012; TAPP, 2010; ZANDBERGEN e IGNIZIO, 2010; ZANDBERGEN, 2011). Em todo o mundo, muitos governos locais investem neste tipo de cadastro, representativos da localização de todas as estruturas habitacionais, para facilitar operacionalização de serviços à população, como atendimento à emergências. Além disso, existem também os cadastros de endereços para facilitar atividades censitárias e pesquisas amostrais. Exemplos de países que possuem esse tipo de cadastro incluem a Austrália (*Geocoded National Address File from the Public Sector Mapping Agencies*), Os Estados Unidos da América (*Master Address File from The Census Bureau*), a Holanda (*Address Coordinates of the Netherlands from the Dutch Land Registry*), o Reino Unido (*ADDRESS-POINT dataset from Ordinance Survey*) e o Brasil, que recentemente publicou os dados do Cadastro Nacional de Endereços para Fins Estatísticos (CNEFE) atualizado no último Censo Demográfico. Cabe ressaltar que nem todos os cadastros citados são disponibilizados ao público, como no Brasil.

Tapp (2010) utilizou uma base de endereços do serviço de atendimento de emergências (911) para interpolar a população de três condados do estado norte-americano da Carolina do Norte e verificou que este método produziu resultados mais exatos que outros métodos avaliados, o da variável limitante e a ponderação por logradouros. O resultado foi ainda melhor quando, combinou endereços com dados cadastrais produzidos pelos condados e estimou a população por lotes fiscais.

Zandbergen (2011) utilizou um cadastro de endereços de alta resolução do *Ohio Geographically Referenced Information Program* para interpolar a população em uma área composta por 90 condados do estado norte-americano de *Ohio* e obteve erro médio absoluto de 4,9%, menor em relação á interpolação com uso da terra (11,6%), logradouros (13,3%), luzes noturnas (18,6%) e ponderação zonal simples (21,2%), entre outros métodos avaliados.

A interpolação com cadastro de endereços é computacionalmente mais simples que qualquer outro método dasimétrico. A formulação básica é dada por:

$$W_o = 1/N_o$$

Equação 13: Ponderação por endereço

$$P_i = W_o \times P_o$$

Equação 14: População média por endereço

$$P_d = N_d \times P_i$$

Equação 15: População estimada por endereço

Onde:

W_o = peso atribuído a cada endereço na zona de origem;

N_o = número de endereços no setor;

N_d = número de endereços na zona de destino;

P_i = é a população média por endereço i ;

P_o = população total no setor;

P_d = população na zona de destino

Tais cadastros fornecem uma informação muito mais detalhada que outros tipos informação auxiliar. Entretanto, esses conjuntos de dados, são relativamente novos e há poucas avaliações da sua qualidade e desempenho.

Uma nova abordagem foi utilizada por Bajat *et al.* (2011), que aplicaram conceitos de ecologia e biogeografia para avaliar a dinâmica de populações humanas. Os autores utilizaram uma análise de Nicho Ecológico através de um modelo de máxima entropia (software *MaxEnt*), utilizado para prever a ocorrência de espécies, além de dados ambientais de domínio público para estimar índices de adequabilidade de habitat e de alteração populacional em uma região da Sérvia

Uma adaptação “multivariada” do método dasimétrico foi utilizada no projeto *LandScan*, desenvolvido pelo *Oak Ridge National Laboratory*, nos Estados Unidos da América (BHADURI *et al.*, 2007). Múltiplas fontes de dados auxiliares de resolução mais detalhada que os polígonos censitários, tais como variáveis de relevo, rede viária, luzes noturnas, mapas de uso e cobertura da terra, são utilizadas num processo iterativo e incremental.

Um modelo multivariado também foi desenvolvido por Amaral *et al.* (2012) para desagregar a população no interior de setores censitários em uma superfície potencial matricial. Tendo como área de estudo o município de Marabá/PA, os autores selecionaram variáveis ambientais indicativas da presença humana – distância de estradas, distância dos rios, distância dos núcleos urbanos, percentual de cobertura florestal, e declividade - e desenvolveram um sistema de inferência baseado na Lógica Nebulosa.

Os principais métodos de interpolação zonal descritos na literatura, com ou sem utilização de informação auxiliar, encontram-se resumidos na **Tabela 1**. Todavia, neste trabalho serão avaliados os seguintes métodos dasimétricos: inteligente de Mennis (2003), ponderação por logradouros e ponderação por endereços.

Tabela 2: Principais métodos de interpolação zonal descritos na literatura.

Autor	Ano	Método	Área de Estudo	Informação Auxiliar	Avaliação da exatidão
Semenov-Tian_Shansky	1926	Dasimétrico	Parte Europeia da Rússia	Mapas topográficos	Não
Wright	1936	Dasimétrico	Cape Cod, MA, EUA	Mapas topográficos	Não
Goodchild e Lam	1980	Ponderação Zonal Simples	Londres, Reino Unido	-	Não
Martin	1989	Centroides ponderados	Reino Unido	-	Não
Fisher e Langford	1996	Binário	Charnwood, Leicester, Oadby e Wigston, UK	LANDSAT TM	Não
Eicher e Brewer	2001	Três classes e variável limitante	Columbia, West Virginia e Virginia, EUA	Classes de uso da e cobertura da terra na escala 1:7.500.000	Sim
Riebel e Buffalino	2005	Ponderação por logradouros	condado de Los Angeles, Califórnia, EUA	Eixos de vias do TIGER, fornecidos pelo US Census Bureau	Sim
Mennis	2003	Inteligente	região de <i>Front Range</i> , Colorado, EUA	Classes de uso da terra	Sim
Weber	2010	Multi-area Multi-classe	estado do Kansas, EUA	Classes de uso da terra	Sim
Bajat <i>et al.</i>	2011		Pomoravlje, Sérvia	Dados abertos	Não
Zandbergen	2011	Ponderação por endereços	Condados de Ohio, EUA	Cadastro de Endereços fornecidos pelo estado de Ohio	Sim

4.1.4. Avaliação da exatidão na interpolação dasimétrica

É senso comum que o resultado da interpolação dasimétrica é mais exato que o dado original agregado. No entanto, tal como outros produtos cartográficos os mapas dasimétricos não estão isentos de erros e incertezas. A avaliação da exatidão não é uma tarefa simples, pois em uma situação real de aplicação da técnica, deseja-se desagregar os dados a partir da menor unidade existente, e por consequência, não existem dados de referência para comparar com os resultados.

Para avaliar o desempenho de cada técnica, trabalhos anteriores (EICHER e BREWER, 2001; MENNIS e HULTGREN, 2005; TAPP, 2010) utilizaram dois níveis hierárquicos de agregação das variáveis censitárias. Inicialmente, mapas dasimétricos foram criados a partir dos dados agregados em uma unidade espacial maior e comparados com a realidade conhecida na unidade menor. Assim, os resultados podem ser avaliados através de estatísticas como o erro médio absoluto, coeficiente de variação e o Erro Médio Quadrático.

Pode-se recorrer à abordagem em dois níveis e assumir a hipótese que o método com melhor desempenho no maior nível de agregação também será melhor no nível mais detalhado (WEBER, 2010).

4.2. Modelagem Estatística

Os modelos de regressão estatística para estimar população começaram a ser desenvolvidos a partir de meados da década de 1960 como uma alternativa para contornar o alto custo e baixa frequência dos censos decenais. Geralmente, esses modelos buscam inferir uma relação entre a população (variável dependente) e dados de sensoriamento remoto (variáveis explicativas ou independentes) para estimar a população em uma de área difícil acesso ou no período intercensitário. Os dados censitários são utilizados somente para treinar o modelo.

As primeiras tentativas de estimar população através de Sensoriamento Remoto foram feitas utilizando-se fotografias aéreas e espaciais. Inspirado na Lei biológica do crescimento alométrico, Nordbeck (1965 *apud* WU *et. al.*, 2005) concluiu que a área construída em assentamentos urbanos era proporcional a sua população elevada a uma

potência. Esta relação foi testada por Waldo Tobler (1969) que utilizou imagens obtidas no voo espacial tripulado *Gemini* para estimar a relação entre população e área construída de diversas cidades no mundo (considerando apenas a circunferência), e obteve coeficientes de correlação (R^2) melhores que 0,87 entre o raio (r) e população. Além disso, o autor observou coeficientes semelhantes nos Estados Unidos, Suécia e Canadá, porém muito diferentes de Japão e Oriente Médio em função do padrão compacto de adensamento na ocupação.

Mais tarde, a disponibilidade de dados do programa LANDSAT possibilitou uma série de estudos de correlação entre população e áreas urbanas. Lo e Welch (1977, *apud* WU *et al.*, 2005) obtiveram coeficientes melhores que 0,75 entre população e área construída extraída de imagens do sensor *Multispectral Scanner* (MSS) em 10 cidades na China continental, e elaboraram a equação do crescimento alométrico das cidades (**Equação 16**), que descreve que o crescimento relativo da população (P) é proporcional ao crescimento da área de uso residencial (A):

$$P = a \times A^b$$

Equação 16: Crescimento alométrico das cidades

Welch e Zuppko (1980, *apud* WU *et al.*, 2005) sugeriram que as imagens noturnas do sensor meteorológico OLS (*Operational Linescan System*) do programa DMSP (*Defense Meteorological Satellite Program*), desenvolvido para observações de nuvens noturnas, fossem utilizadas para mapear a distribuição espacial das atividades humanas. A metodologia criada por Elvidge *et al.* (1997) para identificar as luzes estáveis das cidades, vilas e polos industriais a partir destes dados possibilitou uma série de estudos que correlacionam população com luzes noturnas. Sutton (1997) obteve coeficientes de correlação entre volume de luzes noturnas derivadas DMSP e população variando de 0,92 a 0,98 em cidades do continente norte-americano. Elvidge (*et al.*, 1999) encontraram relações lineares, em escala logarítmica, ao compararem a área iluminada na imagem DMSP/OLS com os valores de população ($R^2= 0,85$), produto interno bruto ($R^2= 0,97$) e o consumo de energia elétrica ($R^2= 0,96$). Os autores concluíram que a imagem DMSP pode ser utilizada para definir e atualizar a distribuição espacial da população em escala global, desde que seja efetuada a devida calibração. No Brasil, Amaral (2003) obteve coeficientes melhores que ($R^2= 0,96$) para cidades do estado do Pará.

Em sua maioria, os modelos de regressão utilizam como variáveis independentes as médias de radiância das bandas espectrais (HARVEY, 2002a; IISAKA e HEGEDUS, 1982), quantidade de pixels em categorias de uso e cobertura da terra (BIELECKA, 2005; LANGFORD e UNWIN, 1994), medidas de textura e variabilidade, além de índices e razões entre bandas espectrais de sensores remotos (HARVEY, 2002b). Esses estudos demonstraram considerável correlação entre densidade populacional em pequenas áreas e variáveis agregadas extraídas de imagens multiespectrais, com coeficientes de correlação entre (R^2) 0,6 e 0,9 para o conjunto de treinamento.

Outra abordagem utilizada é análise de correlação entre população e unidades domiciliares. A população total de uma área pode ser estimada através da multiplicação do número de domicílios pela média do número de pessoas por habitação. Pode-se também categorizar os tipos de domicílios e aplicar uma razão diferente para cada categoria. Esta razão pode ser obtida por amostragem ou calculada a partir dos dados do censo. Considerando que cada casa é ocupada por um domicílio, o número total de domicílios em uma área pode ser estimado a partir de fotografias aéreas e imagens de satélite.

Em sua revisão, Wu *et al* (2005) esclareceram que:

“Green (1956) propôs a contagem de domicílios a partir de fotografias e Porter (1957) aplicou esta metodologia na Libéria, determinando a razão de pessoas por domicílio a partir de observações de campo” ... “Hsu (1971) aplicou a mesma metodologia para estimar a população intercensitária na região metropolitana de Atlanta, mas determinou a razão a partir de dados censo norte-americano”... “Para obter estimativas mais precisas, Lo e Chan (1980) utilizaram uma metodologia de levantamento de campo para calcular razão pessoa/domicílio para vários tipos de habitação. Já Lo (1989) tentou definir uma metodologia para determinar automaticamente a densidade de edificações residenciais em uma grade regular (raster) a partir de imagens aéreas e orbitais.” (Tradução nossa).

No passado, os pesquisadores recorreram à extração manual de edificações residenciais de imagens aéreas. Com oferta de imagens multiespectrais de altíssima resolução espacial, como aquelas obtidas pelos satélites IKONOS, QuickBird, Geoeye-1 e Worldview-1, e a evolução das técnicas de extração automática de feições, o processo foi facilitado.

Alguns autores incorporaram a regressão estatística ao mapeamento dasimétrico. Langford *et al.* (1991) desenvolveram uma técnica que utiliza a regressão estatística para realizar uma “interpolação inteligente”. Baseado em assinaturas espectrais das classes de uso da terra, determina-se o número de pixels de cada classe dentro de cada zona de origem. Uma análise de regressão é então realizada e os parâmetros obtidos são utilizados para estimar população nas zonas de destino.

Bielecka (2005) utilizou modelos estatísticos para produzir um mapa dasimétrico da região nordeste da Polônia com classes de uso e cobertura da terra do projeto CORINE (BOSSARD *et al.*, 2000), que abrange toda a Europa. A análise de regressão foi utilizada para determinar a relação entre as classes e a densidade populacional.

5. MÉTODOS DASIMÉTRICOS APLICADOS À MESORREGIÃO METROPOLITANA DE BELÉM

Esta pesquisa buscou aplicar a interpolação dasimétrica para estimativa populacional em pequenas áreas. Embora seja um tema com intensa produção acadêmica recente, conforme apresentado no Capítulo 4, aplicações práticas são ainda incipientes, com exceção de poucos projetos em escala global e nacional. Dessa forma, este trabalho visou a seleção de algoritmos e a avaliação de dados disponíveis, de modo fornecer os subsídios para a elaboração de uma folha metodológica que estabelece os princípios e procedimentos para produção sistemática de mapas dasimétricos e superfícies populacionais.

5.1. Área de Estudo

A Mesorregião Metropolitana de Belém (MMB), conforme classificação do IBGE, é composta por onze municípios e 2.432 setores censitários, dos quais 228 são rurais e 2.204 urbanos. Trata-se de uma região situada entre os paralelos $0^{\circ} 59' 17''$ S e $1^{\circ} 50' 28''$ S e os meridianos $47^{\circ} 58' 19''$ W Gr e $48^{\circ} 50' 15''$ W Gr, com uma área de 6.785 km^2 , e população de 2.433.983 habitantes, segundo o Censo Demográfico 2010 (IBGE, 2010)

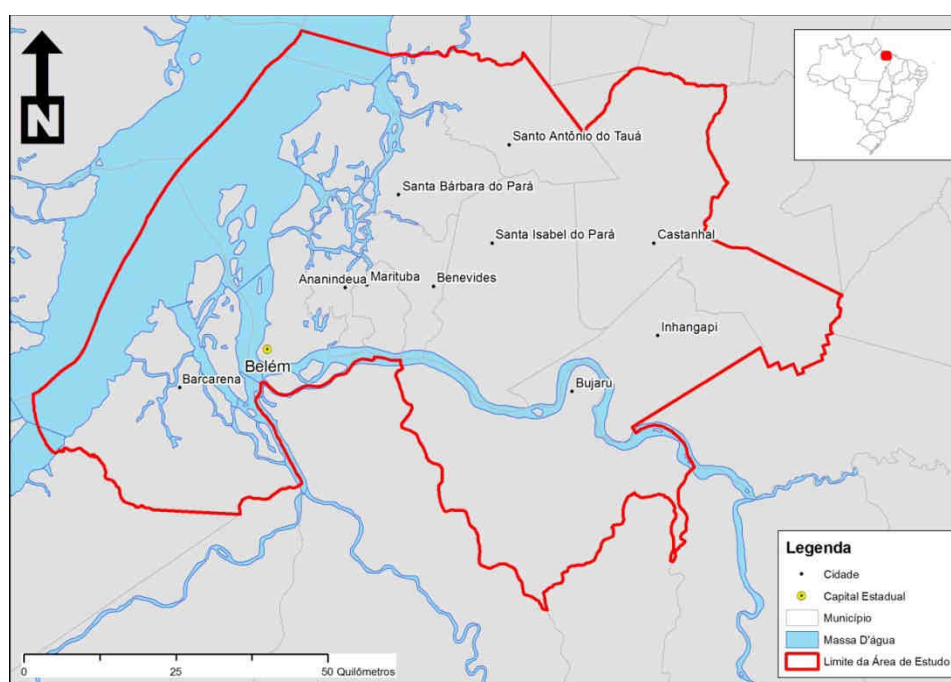


Figura 19: Enquadramento geográfico da área de estudo.

Os municípios de Belém, Ananindeua, Marituba, Santa Bárbara, Benevides e Santa Izabel do Pará constituem atualmente a Região Metropolitana de Belém, instituída por Lei Complementar Federal em 1973, originalmente com dois municípios, e alterada em 1996 e em 2010. Os demais municípios são predominantemente rurais, com pequenos enclaves urbanos.

A escolha da MMB para área de estudo nesta dissertação foi motivada por se tratar de uma região com padrões de ocupação humana diversificados, incluindo áreas de intenso adensamento populacional e ocupações rurais isoladas, bem como a presença de vazios demográficos (**Figura 20**). Além disso, a região possui setores censitários bastante heterogêneos em relação à área, o que proporciona uma boa oportunidade para comparar o desempenho dos métodos dasimétricos selecionados. Além disso, a existência de grandes extensões de massa d'água é uma condição que introduz viés no cálculo de razões e taxas por unidade de área. Nessas condições, o mapeamento dasimétrico oferece uma alternativa mais adequada. Outro fator decisivo para definição desta área de estudo foi a existência de uma base de dados de uso e cobertura da terra para toda Amazônia Legal.

5.2. Materiais utilizados

O tratamento e análise dos dados foram realizados com o auxílio de alguns softwares, abaixo explicitados com seus respectivos usos:

- *PostgreSQL* com cartuxo espacial *PostGIS* – Sistema Gerenciador de Banco de Dados (SGBD) utilizado para armazenar, tratar e consultar dados do universo, base territorial e cadastro de endereços do Censo Demográfico.
- *SAFE Feature Manipulation Engine* (FME) – ferramenta de ETL espacial (*Extract Load and Transform*) utilizada para facilitar a transformação e tratamento dos dados geoespaciais. Este software foi utilizado para auxiliar a carga de dados do censo na base construída no *Postgre/PostGIS*.

- *ArcGIS 9.3* – Aplicativo SIG utilizado para implementar e automatizar os procedimentos de interpolação dasimétrica, através do componente *Model Builder*. Além disso, foi utilizado para confecção dos mapas e gráficos.
- *PythonWIN* – Ambiente de programação para edição e execução dos algoritmos em linguagem *Python*.
- Algoritmo do *Intelligent Daysimetric Mapping* – Código desenvolvido por Hultgren (2006) para estrutura matricial disponibilizado na página do autor.

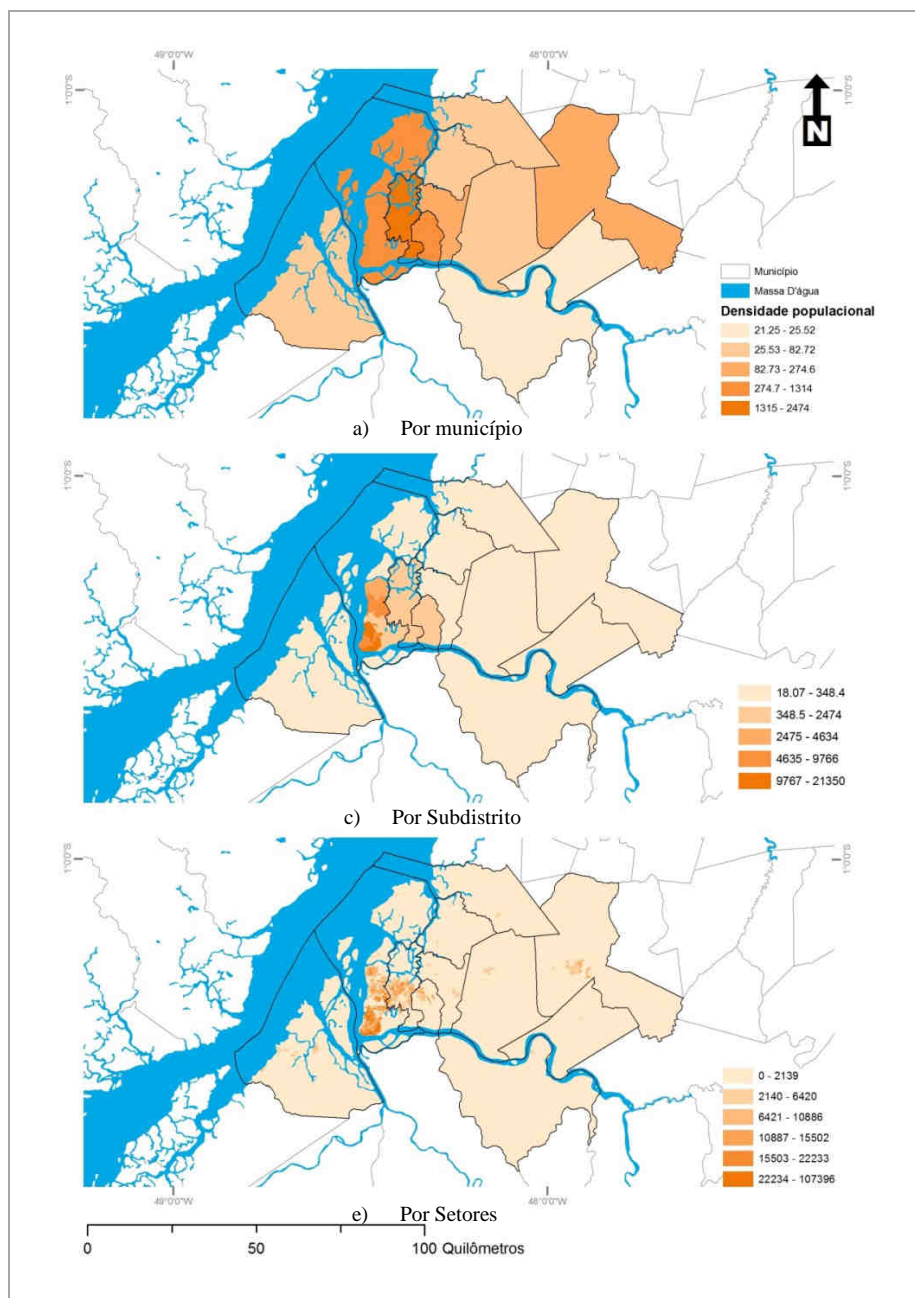


Figura 20: Densidade populacional (habitantes por km²) na área de estudo por : *a)* município, *b)* subdistrito e por *c)* setor censitário. No primeiro e segundo é usado como escala de classificação para a representação o método de *natural break* e no terceiro por desvio padrão. Fonte: IBGE (2012a)

5.3. Fontes de Dados

Tendo em vista a dificuldade de encontrar dados auxiliares que oferecessem uma ampla cobertura para a região foram selecionados os seguintes conjuntos de dados:

- Dados do Censo Demográfico de 2010, incluindo a Base Territorial, Dados do Universo por Setores Censitários e o Cadastro Nacional de Endereços para Fins Estatísticos;
- Mapa de uso da terra elaborado pelo projeto TerraClass que foi utilizado em conjunto com os dados do Censo para a elaboração do mapa dasimétrico;
- Base de Logradouros do *OpenStreetMap* para efetuar uma interpolação por logradouro nas áreas urbanas; e

Assim, nesta seção são apresentados os conjuntos de dados utilizados, as variáveis empregadas e os distintos tratamentos efetuados.

5.3.1. Censo Demográfico

Foram usados três conjuntos de dados do Censo Demográfico de 2010 (CD2010), a saber: a Base Territorial, os Dados do Universo por Setores Censitários e o Cadastro Nacional de Endereços para Fins Estatísticos (CNEFE). Estes dados foram integrados em um banco de dados *PostgreSQL 9.0* habilitado com suporte espacial fornecido pelo módulo *PostGIS 2.0*.

A base territorial do CD2010 compreende todos os setores censitários do Brasil, separados por Unidades da Federação em estrutura de dados vetorial no formato *shapefile* (ESRI, 1998). Os arquivos se encontram em sistema de coordenadas geodésicas e *datum* horizontal SIRGAS 2000, obtidos no sítio da Diretoria de Geociências do IBGE (<ftp://geofp.ibge.gov.br/>). Estes dados em conjunto com os respectivos memoriais descritivos de setores foram importados para o banco de dados *PostGIS* com auxílio de um fluxo de transformação e carga de dados elaborado com o software *Safe FME*. Após a importação, realizou-se uma validação topológica que removeu áreas de auto interseção e vazios entre os setores. Também foi executado um ajuste espacial, de forma que as arestas da malha vetorial de setores coincidisse com os eixos de logradouros. Para isso foi empregada uma tolerância máxima de 25 metros que permitisse o deslocamento dos vértices dos setores.

Os resultados do Universo do Censo Demográfico 2010 por setor censitário foram disponibilizados pelo IBGE através de conjuntos de planilhas divididas de acordo com as

variáveis de domicílio, pessoa e responsável. Para a finalidade deste trabalho, apenas três planilhas referentes ao estado do Pará foram importadas para o banco *PostgreSQL*, sendo duas planilhas de domicílio e uma planilha com descrição das unidades territoriais. O procedimento de importação incluiu a conversão das colunas de algumas variáveis para tipo de dados numérico, a fim de facilitar análises subsequentes.

Através de uma consulta em linguagem SQL (*Structured Query Language*) a malha vetorial de setores censitários foi recuperada e realizada uma união (*join*) com as variáveis selecionadas do conjunto de dados do Universo. O campo *geocódigo* existente nas duas bases foi utilizado como chave para realização do “*join*”. Verificou-se que, dos 2.432 setores da base territorial na área de estudo, não foi possível recuperar os atributos correspondentes aos dados do universo em apenas oito setores. O resultado foi então convertido para o formato *shapefile* na projeção *Universal Transversa de Mercator* (UTM) fuso 22, no sistema de referencia geodésico SIRGAS 2000, assim como os demais dados utilizados no trabalho.

Os arquivos do CNEFE também foram carregados no banco *PostgreSQL*. Utilizou-se o *layout* fornecido em conjunto com os microdados para identificar os atributos dos endereços. Após o carregamento, foi necessário o seguinte tratamento efetuado em SQL no banco *PostGis*: a transformação da variável que se encontrava em formato texto para numérico e, a seguir, a transformação deste para uma representação geométrica a ser armazenada na tabela como atributo. Uma característica dos dados divulgados do CNEFE é que apenas os endereços rurais possuem campos de latitude e longitude preenchidos. Para os setores urbanos, a localização dos endereços pode ser recuperada através de *geocoding* requerendo uma base georreferenciada e confiável de logradouros com os número dos endereços dos domicílios (número de porta). Tendo em vista a dificuldade de acesso a este tipo de dados, este procedimento não foi realizado neste trabalho.

5.3.2. Mapas de uso e cobertura da Terra

Os mapas de uso e cobertura da terra utilizados neste trabalho, ilustrado na **Figura 21**, foram adquiridos através do projeto *TerraClass*, que mapeou na escala 1:100.000 as áreas desflorestadas da Amazônia Legal. O projeto é uma parceria do Instituto Nacional de Pesquisas Espaciais (INPE) com a Empresa Brasileira de Pesquisa Agropecuária

(Embrapa). O *TerraClass* mapeou os 18% do da floresta amazônica já desmatados, gerando uma classificação dos diversos usos e situações em cada um dos nove estados da região. As análises foram feitas a partir de informações resultantes do PRODES (Projeto de Monitoramento do Desflorestamento na Amazônia Legal). Os dados divulgados são referentes à classificação de imagens adquiridas no ano de 2008.

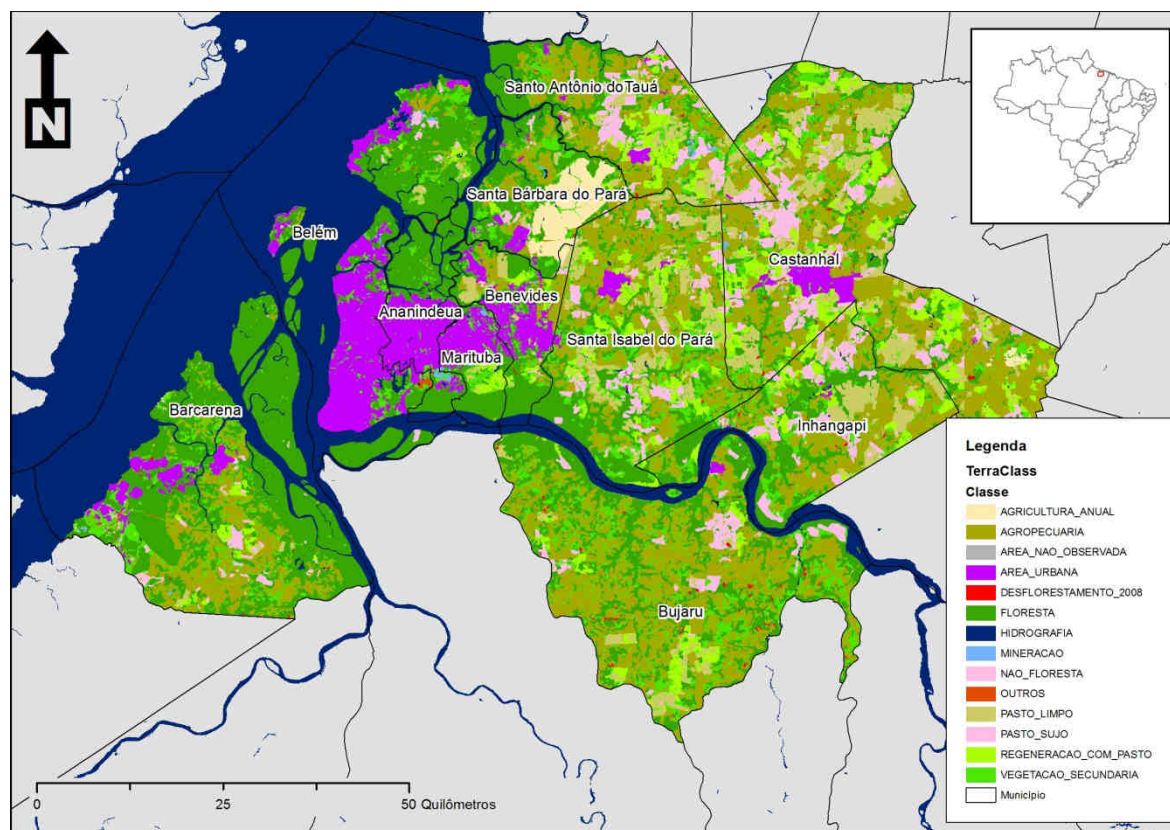


Figura 21: Recorte da classificação do *TerraClass* na área de estudo.

Os arquivos vetoriais em formato *Shapefile*, recortados segundo a grade de ponto/órbita do LANDSAT podem ser obtidos na página http://www.inpe.br/cra/projetos_pesquisas/terraclass.php. Para este trabalho, foram utilizados os arquivos das cenas 223/061 e 224/061 as quais foram unidas em um mosaico da região. A seguir, o mosaico foi recortado conforme o limite da área de estudo e efetuado uma generalização para suavizar o efeito escada na imagem proveniente da conversão *raster-vetor* do mapa de uso da terra.

Os dados contemplam 15 classes de uso e cobertura da terra, delimitadas por diferentes metodologias de classificação de imagens LANDSAT TM. Para o mapeamento da vegetação secundária foi aplicado um Modelo Linear de Mistura Espectral. O

mapeamento das pastagens foi efetuado através da interpretação analógica. Para mapeamento da agricultura anual foi realizada uma classificação espectro-temporal de imagens do índice de vegetação do satélite MODIS e detalhamento com imagens LANDSAT. Para as demais modalidades de uso da terra foi realizada a segmentação de imagens LANDSAT e classificação supervisionada.

Estes dados foram utilizados neste trabalho no mapeamento dasimétrico inteligente (MENNIS e HULTGREN, 2005). As 15 classes de uso e cobertura da terra propostas pelo *TerraClass* foram agrupadas em quatro para aumentar a probabilidade de seleção de amostra para estimativa da densidade das classes. Estas quatro classes são: *inabitada*, *agropecuária*, *vegetação* e *área urbana*.

5.3.3. Base de Logradouros

O grande volume de informações espaciais proporcionada por receptores GPS associados a computadores portáteis e dispositivos conectados a Internet possibilitou o surgimento da chamada *Informação Geográfica Voluntariada* (GOODCHILD, 2007). Uma das principais iniciativas do gênero é o *OpenStreetMap*, um projeto de construção de uma base cartográfica colaborativa a partir de levantamentos realizados por usuários em todo o mundo. Fundado em 2005 na *University College of London*, o projeto possui uma rede de aproximadamente 650.000 colaboradores cadastrados até julho de 2012.

O conjunto de dados compilado pelo projeto é disponibilizado livremente⁵. Assim, os usuários podem copiar distribuir, transmitir ou adaptar os mapas, desde que forneçam os créditos. Caso os dados sejam modificados, devem ser distribuídos sob a mesma licença. Como a maior parte do levantamento é realizada com GPS de navegação, a base está sujeita a erros que variam na ordem de 3 a 15 metros, em média. No Brasil, os grandes centros urbanos do país já possuem informação detalhada da rede, conforme pode ser observado na **Figura 22** para as cidades de Belém e Ananindeua.

⁵ Licença *Creative Commons Attribution-ShareAlike 2.0 CC BY-SA*).

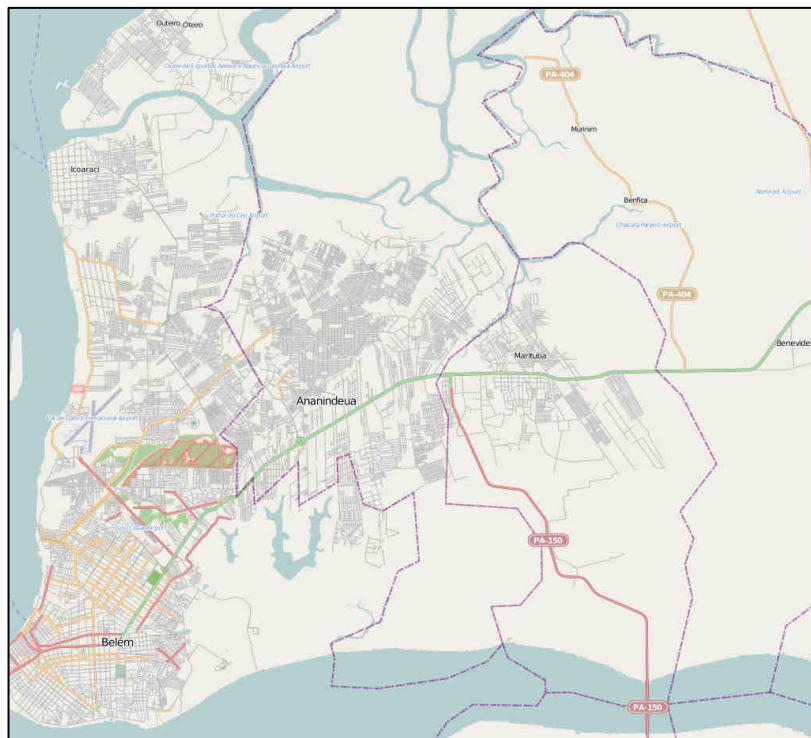


Figura 22: Extrato do *OpenStreetMap* para a área de estudo.
Fonte: (OSM, 2012)

5.4. Métodos Avaliados

Para atingir o objetivo proposto, o presente trabalho avaliou a aplicação de três métodos de interpolação zonal para desagregar a variável população residente a partir do conjunto de dados do Universo do Censo 2010, a saber: *i*) o mapeamento dasimétrico inteligente (MENNIS e HUTGREN, 2005); *ii*) a ponderação por logradouros e a *iii*) ponderação por endereços.

Para todos os métodos avaliados, a exceção do mapeamento dasimétrico inteligente, foram elaborados algoritmos para automação dos procedimentos no aplicativo *ArcGIS*. Nesse sentido, foi utilizada a ferramenta *ModelBuilder* do ambiente que proporcionou a construção visual do modelo e a geração do código elaborado na linguagem de programação *Python*.

Para efeito de comparação dos resultados obtidos em cada uma das técnicas, utilizou-se a mesma abordagem adotada em trabalhos anteriores (EICHER; BREWER, 2001; TAPP, 2010; WEBER, 2010). As contagens originalmente agrupadas por setor censitário foram agregadas em uma unidade hierárquica maior (subdistritos ou bairros, quando

existentes) e este conjunto de unidades foi utilizado como zona de origem. A população por setores censitários (zona de destino) foi estimada através dos métodos selecionados e comparada com o valor conhecido no setor.

5.4.1. Mapeamento dasimétrico inteligente

O método inteligente proposto por Mennis (2003) (**Equação 6**) é frequentemente citado e tem sido utilizado em situações práticas, como na produção de bases demográficas matriciais. Em muitos desses trabalhos, a informação auxiliar é derivada de bases nacionais ou regionais de uso e cobertura da terra, que possibilitam a produção de dados dasimétricos em escala cartográfica de semi-detulhe ou de reconhecimento (geralmente de 1:100.000 a 1:250.000) para amplas áreas do território.

Para viabilizar o mapeamento dasimétrico inteligente neste trabalho, foi utilizado o algoritmo do Mapeamento Dasimétrico Inteligente desenvolvido por Mennis e Hultgren (2005) para a estrutura de dados vetorial aplicado ao nível de subdistrito. Neste algoritmo o operador define os pesos (densidades) das classes auxiliares ou especifica os parâmetros para seleção de amostras que podem ser o centroide ou percentual de cobertura. Deste modo, as áreas definidas como inabitadas, no caso os corpos d'água, receberam peso igual a 0 (zero), e as demais classes tiveram as densidades definidas por amostragem.

Na execução do algoritmo foram informados os dados de entrada, um arquivo com a população total agregada por subdistrito e outro com a classificação de uso da terra “recortada” por setores censitários, além do método e o percentual de cobertura para seleção de amostras (95%). A partir do mapa dasimétrico obtido, a população nos setores foi estimada pelo reagrupamento dos valores obtidos no interior dos setores. Dessa forma, foi possível avaliar a exatidão dos resultados através da comparação com a população conhecida no setor. Além da aplicação do método ao nível de subdistrito foi produzido um mapa dasimétrico a partir dos dados agregados por setor censitário, a fim de possibilitar a análise visual e comparação com resultados obtidos com os outros métodos.

Para melhor compreensão, a **Figura 23** apresenta o fluxograma com toda a sequência de operações realizadas desde a aquisição do mapa de uso da terra do *TerraClass*, o ajuste das classes e dos setores até a elaboração dos mapas dasimétricos a nível de subdistrito e

setores e a comparação dos resultados obtidos para efetuar a avaliação do método de amostragem usado no mapeamento dasimétrico inteligente.

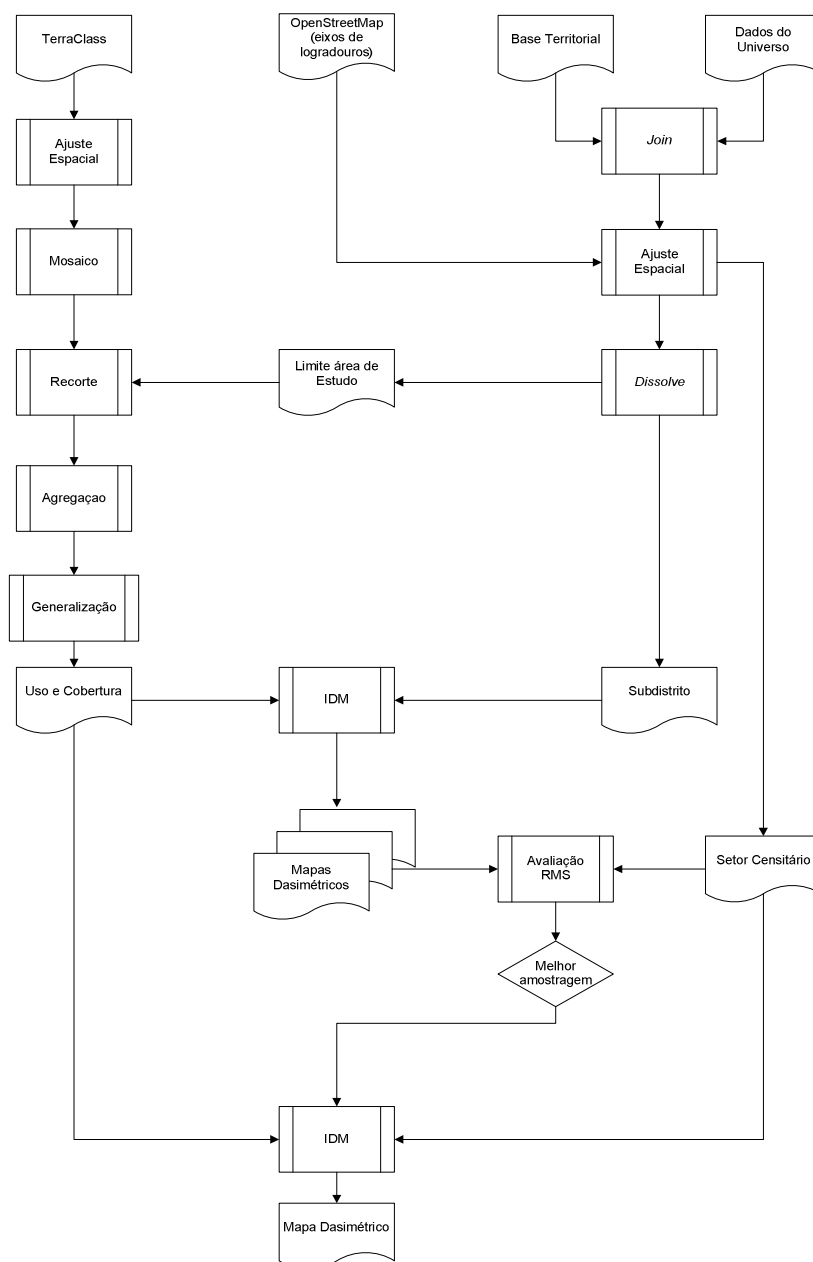


Figura 23: Fluxograma metodológico para avaliação do método dasimétrico inteligente e produção do mapa dasimétrico.

5.4.2. Interpolação ponderada por logradouros

O método de interpolação zonal desenvolvido por Reibel e Bufallino (2005) (Equação 11) utiliza bases de estradas e arruamentos para transferir valores entre zoneamentos distintos. Cada segmento de logradouro recebe um peso de acordo com a

proporção entre sua extensão e a soma da extensão de todos os segmentos contidos na zona de origem.

Para avaliar o método, foram utilizados os eixos de logradouros do *OpenStreetMap* e a contagem populacional agregada por subdistrito. Dessa forma, foi possível calcular os pesos e estimar a população em cada segmento de logradouro. Em seguida, a população estimada foi agregada por setores censitários e comparada com o valor conhecido. A sequência de operações realizadas para efetuar a ponderação por logradouros encontra-se resumida no fluxograma apresentado na **Figura 24**.

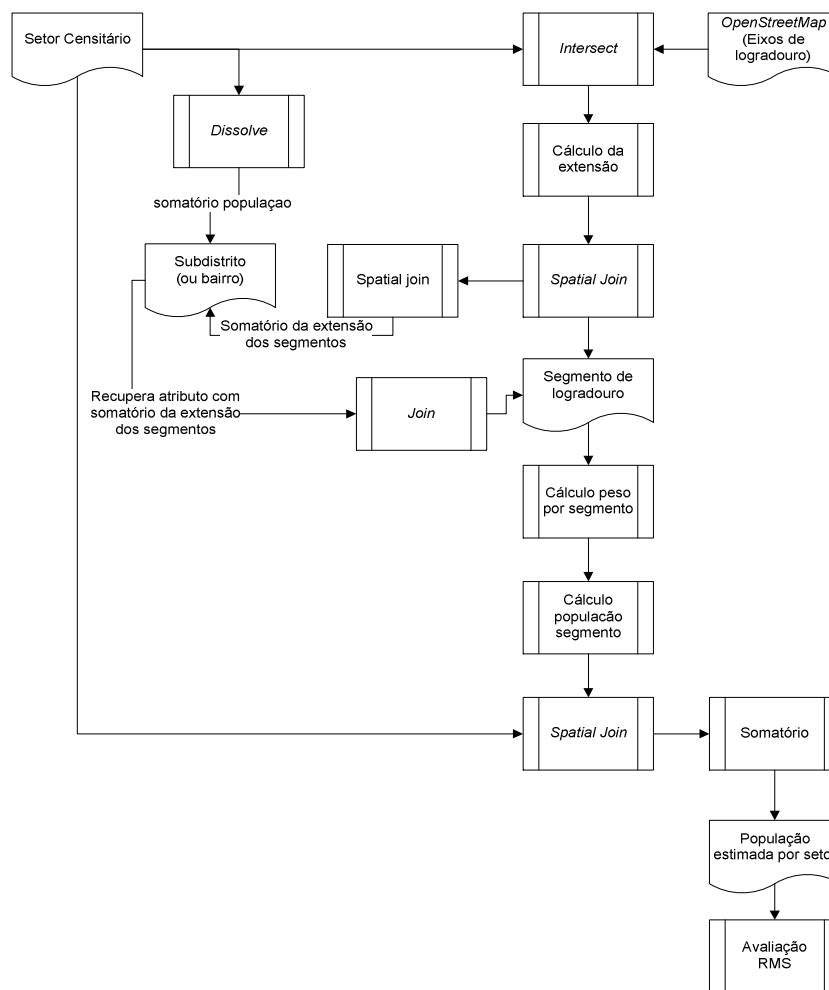


Figura 24: Fluxograma metodológico para avaliação da ponderação por logradouros.

A primeira etapa consiste em sobrepor três camadas de mapas vetoriais: zonas de origem, zonas de destino e eixos de logradouros. Todos os eixos são segmentados no limite da intersecção entre as zonas de origem e destino para evitar que uma feição seja contabilizada mais de uma vez. Em seguida, atribui-se um peso a cada segmento de logradouro. Este peso é calculado pela razão entre a sua extensão e o somatório da

extensão de todos os segmentos no setor da zona de origem (Equação 11). A população no segmento pode ser estimada através da multiplicação do peso pela população original no subdistrito (Equação 12), e na zona de destino (setor) pelo somatório da população estimada nos segmentos nela contidos (Equação 13).

5.4.3. Interpolação ponderada por de endereços

Este método é similar à ponderação por logradouros, porém computacionalmente ainda mais simples uma vez que associa um peso a cada endereço (**Equação 13**). Para melhor compreensão, a **Figura 25** apresenta o fluxograma da sequência de operações efetuadas.

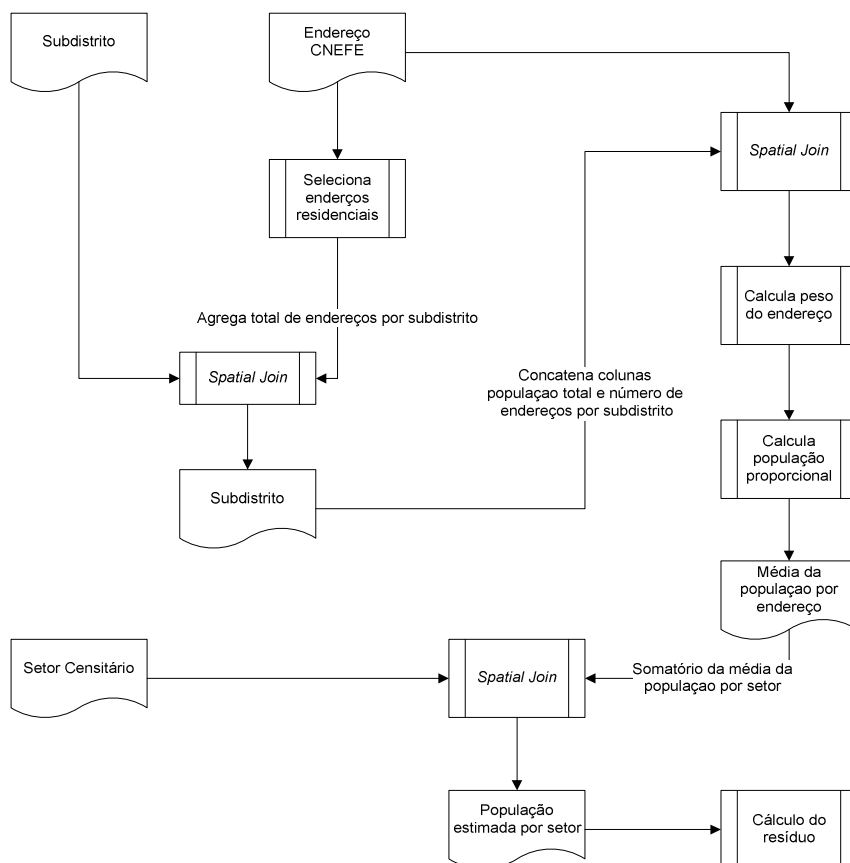


Figura 25: Procedimentos realizados para avaliação da ponderação por endereços.

Este método possui a vantagem de utilizar a localização conhecida dos domicílios. Uma vez definidos os pesos e estimada a população por endereço, pode-se utilizar o conjunto de dados para interpolar a variável desejada em qualquer unidade de destino.

Para viabilizar o procedimento na área de estudo, foram utilizados os endereços do CNEFE. Através de uma consulta à base armazenada no banco de dados *PostGIS*, foram selecionados os endereços residenciais dos setores rurais na área de estudo georreferenciados, totalizando 23.156 feições pontuais. Com uma sequência de operações espaciais e cálculo de atributos, foram obtidos os pesos (**Equações 13 e 14**). Em seguida, calculou-se o número médio de habitantes para cada ponto.

5.5. Análise dos resultados

Para melhor compreensão a análise dos resultados é efetuada em três etapas. A primeira análise visou à elaboração de um mapa dasimétrico de toda área de estudo, através do algoritmo do IDM. Este método foi o único que pôde ser aplicado em toda a extensão da área de estudo, pois, a classificação de uso e cobertura da terra possui cobertura completa para todos os setores da Mesorregião Metropolitana de Belém.

A segunda análise objetivou efetuar a estimativa populacional nas áreas urbanas, aplicando o método de ponderação por logradouro. Esse método foi aplicado à área urbana dos municípios de Belém e Ananindeua, uma vez que nesses municípios todos os setores possuíam ao menos uma rua digitalizada na base do *OpenStreetMap*.

A terceira análise objetivou o mapeamento de áreas rurais utilizando os endereços do CNEFE. Ressalta-se que nos dados divulgados deste cadastro somente os endereços rurais são georreferenciados, ou seja, apresentam os campos de latitude e longitude preenchidos. Além disto, verificou-se a existência de dados incompletos. Dos 228 setores rurais, apenas 125 possuíam endereços georreferenciados. Dessa forma, somente estes setores foram utilizados na análise.

5.5.1. O mapeamento dasimétrico inteligente

O mapa dasimétrico da Mesorregião Metropolitana de Belém, ilustrado na **Figura 26**, foi produzido com o algoritmo do IDM descrito no fluxograma da **Figura 23**. A densidade da classe “inabitada” foi predefinida com valor zero, a densidade populacional na classe vegetação e urbana foi estimada por amostragem e a classe agropecuária foi

estimada pela ponderação zonal simples. Estas duas últimas foram definidas pelo algoritmo enquanto a primeira foi informada como parâmetro.

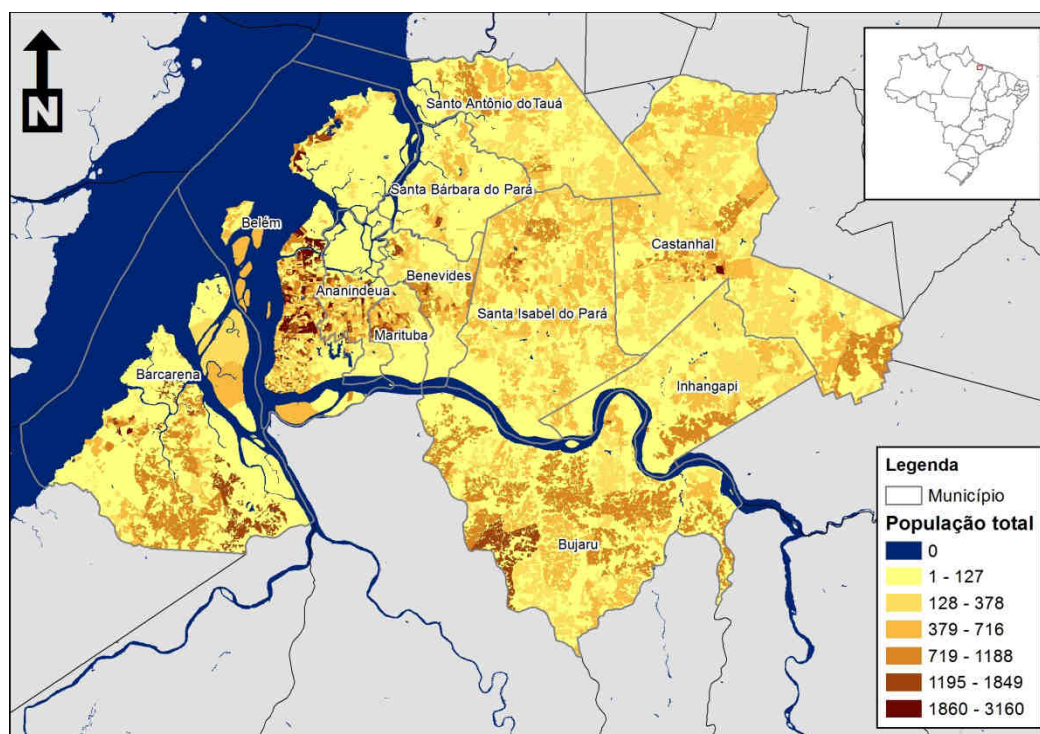


Figura 26: Mapa dasimétrico da Mesorregião Metropolitana de Belém produzido com o algoritmo do IDM a partir dos dados agregados por subdistrito.

Utilizou-se o mapa produzido para reagrupar a população por setor censitário e comparar com a contagem do Censo. O erro bruto, ou verdadeiro, foi obtido subtraindo-se a população estimada da população conhecida no setor.

A distribuição do erro verificado nos setores rurais e urbanos na MMB e por município é apresentada na **Tabela 3**. Valores negativos indicam que a população foi superestimada. Pode-se verificar que os valores apresentaram distribuição bastante assimétrica.

Neste método os municípios de Ananindeua, Castanhal e Marituba apresentaram o maior desvio padrão nos setores urbanos enquanto os municípios de Barcarema e Benevides apresentaram nos setores rurais.

Tabela 3: Estatísticas de resumo para o erro verificado no método de mapeamento dasimétrico inteligente.

Mesorregião Metropolitana de	Tipo	Setores	Mínimo	Máximo	Média	Desvio Padrão
------------------------------	------	---------	--------	--------	-------	---------------

Belém	Rural	228	-7.119,55	1.304,35	12,86	869,47
	Urbano	2204	-19.971,91	1.746,68	-20,85	1.005,83
Municípios	Tipo	Setores	Mínimo	Máximo	Média	Desvio Padrão
Ananideua	Rural	2	-397,44	0,00	-198,72	281,03
	Urbano	488	-19.971,91	1.031,61	-19,90	1.464,50
Barcarena	Rural	70	-7.119,55	1.304,35	-30,09	1.339,13
	Urbano	34	-2.695,85	625,39	28,86	628,93
Belém	Rural	17	-770,88	508,00	74,77	341,65
	Urbano	1303	-6.421,93	1.574,80	-22,18	635,97
Benevides	Rural	26	-4.461,33	861,14	91,59	1.081,39
	Urbano	31	-3.429,95	742,15	-80,33	919,52
Bujarú	Rural	21	-410,15	568,41	37,57	249,67
	Urbano	8	-580,13	240,92	-95,42	283,31
Castanhal	Rural	28	-608,17	505,75	-34,00	320,23
	Urbano	164	-11.588,79	696,84	-6,75	1.320,21
Inhangapi	Rural	10	-240,68	377,09	109,52	204,42
	Urbano	5	-152,66	198,28	4,79	128,17
Marituba	Rural	3	-193,70	210,69	24,98	204,20
	Urbano	107	-11.472,28	802,00	-25,74	1.680,49
Santa Bárbara do Pará	Rural	16	-1.620,58	792,04	65,99	645,82
	Urbano	6	-1.138,37	800,86	-217,79	752,20
Santa Isabel do Pará	Rural	20	-694,91	547,68	11,66	351,62
	Urbano	41	-1.961,57	1.746,68	8,65	733,55
Santo Antônio do Tauá	Rural	15	-936,91	584,43	-34,11	456,62
	Urbano	17	-2.623,58	968,64	-16,55	801,98
TOTAL		2432	-19.971,91	1.746,68	-17,69	993,53

O dado de uso da terra aplicado neste método apresenta baixa resolução espacial uma vez que é resultado do processo de classificação de imagens LANDSAT. Isto não permitiu detalhar a distribuição da população nas áreas urbanas.

Para melhor compreensão, o **Gráfico 2** apresenta a distribuição do erro no método dasimétrico inteligente e a **Figura 27** apresenta o mapa da região com a distribuição espacial do erro em termos de desvio padrão. A distribuição do erro forte assimetria negativa.

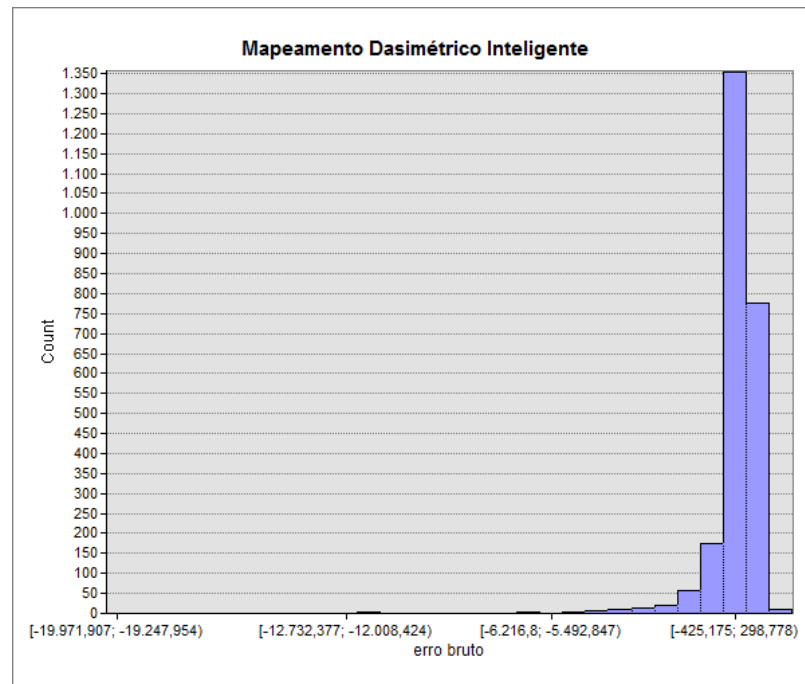


Gráfico 2: Distribuição do erro na interpolação dasimétrica inteligente

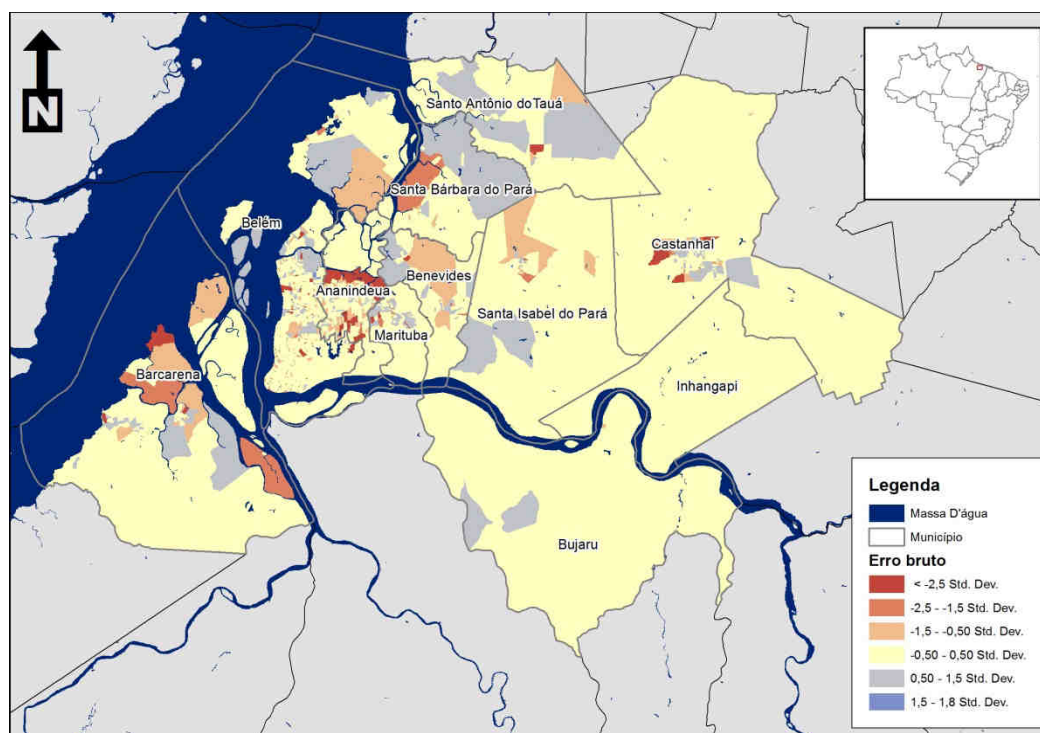


Figura 27: Distribuição espacial do erro por setor produzido com o algoritmo do IDM.

5.5.2. A interpolação ponderada por logradouro nas áreas urbanas

Para esta aplicação foram selecionados os municípios de Belém e Ananindeua. Os procedimentos estão ilustrados no fluxograma da **Figura 24**. Os eixos de logradouros foram segmentados na divisa dos setores e, a seguir, foi calculada a extensão de cada segmento, que é sumarizada pela zona de origem da interpolação, no caso, bairro ou subdistrito. A adoção de diferentes unidades como zona de origem para o experimento foi necessária porque somente o município de Belém possui bairros regulamentados e representados no Censo Demográfico 2010. Os pesos de cada seguimento são calculados pela razão entre a sua extensão e a soma da extensão de todos os segmentos no bairro ou subdistrito. Este peso é então utilizado para estimar a população por segmento de rua.

A partir da população por seguimento de logradouro pôde-se estimar a população por setores através de um procedimento conhecido como *Spatial join*. A **Figura 29** apresenta a população estimada pelo método e exemplos de logradouros com a população estimada por seguimento. Para avaliar o desempenho, a população estimada por segmento de rua foi agregada por setor censitário e comparada com a contagem do Censo (**Figura 28**). O resultado é ilustrado na **Figura 30** e na **Tabela 4**.

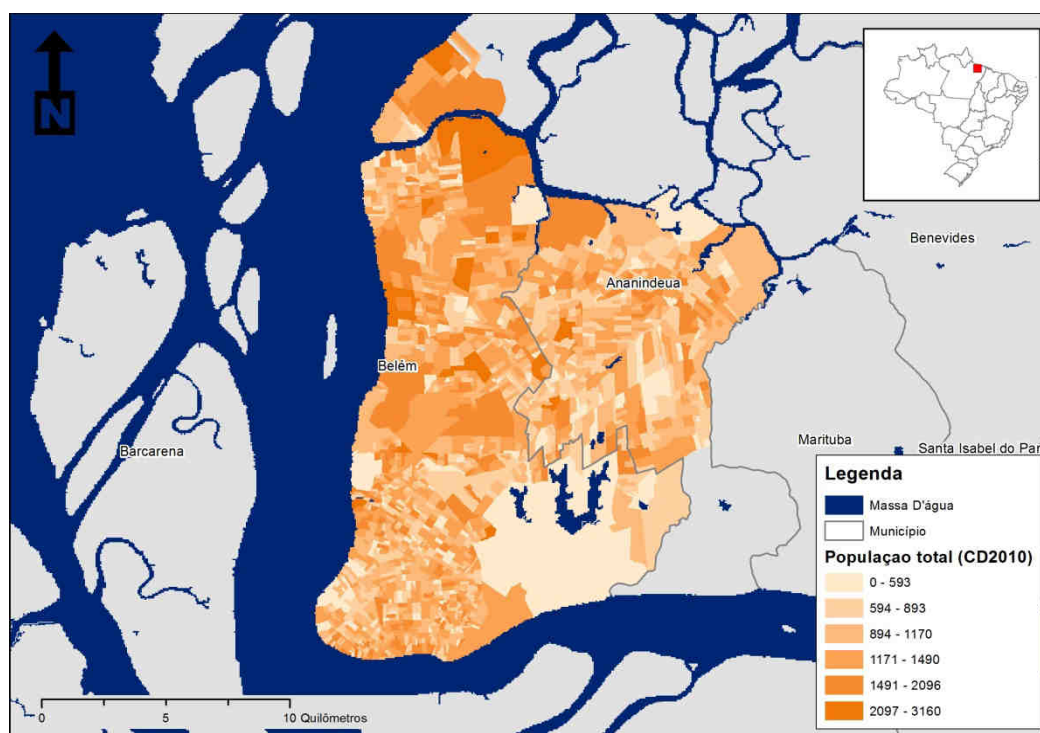


Figura 28: População total por setor censitário na área urbana de Belém e Ananindeua.

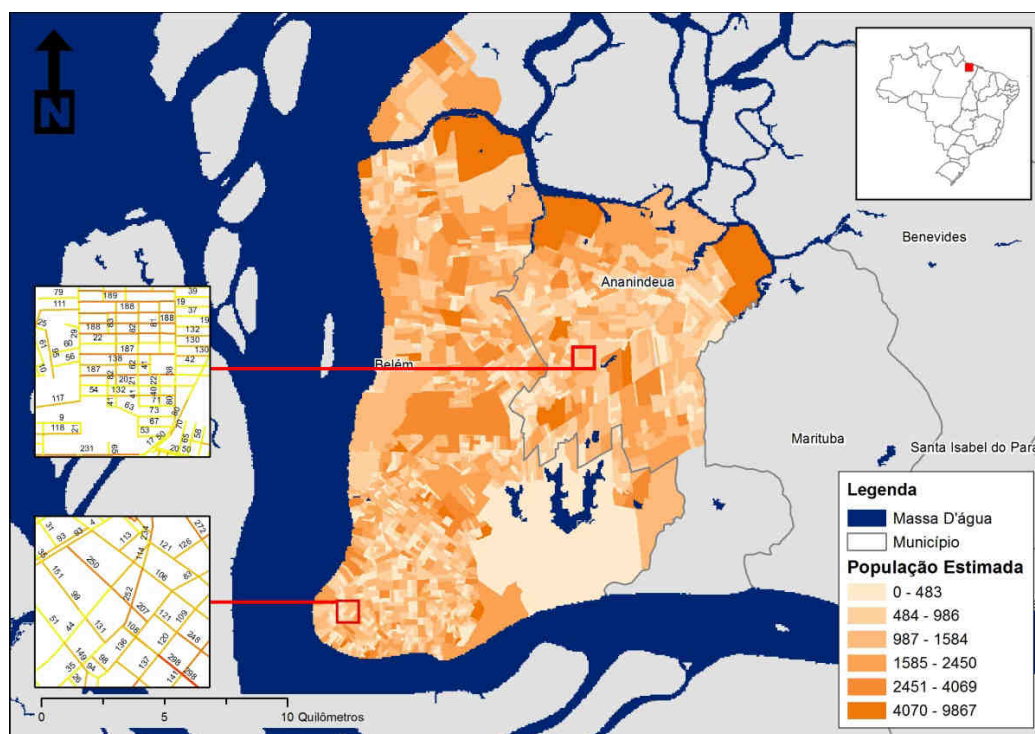


Figura 29: População por setor censitário estimada na ponderação por logradouro e detalhe da população estimada por logradouro.

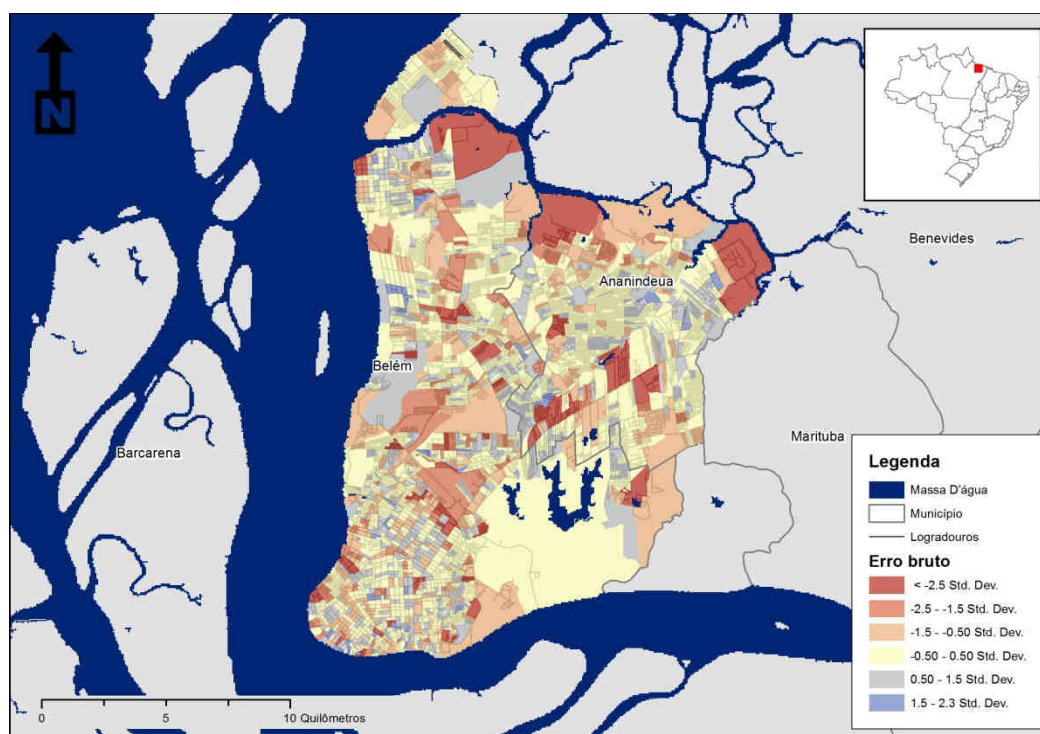


Figura 30: Distribuição espacial do erro por setor produzido na interpolação dasimétrica por logradouro.

Na **Figura 30**, observa-se que setores em vermelho tiveram população superestimada, provavelmente pela presença de ruas não residenciais. De forma inversa, os setores em azul tiveram população subestimada em função da existência de poucos logradouros.

Os resultados apresentados na **Tabela 4** demonstram que o que o desvio padrão foi menor e que as médias ficaram mais próximas à zero. No **Gráfico 3** observa-se uma distribuição assimétrica do erro, porém menor do que a observada com o método dasimétrico inteligente.

Tabela 4: Estatísticas de resumo para o erro verificado no método de mapeamento dasimétrico por logradouro

Município	Erro verdadeiro				
	Setores urbanos	Mínimo	Máximo	Média	Desvio Padrão
Ananideua	485,00	-8.888,60	1.549,00	-1,77	810,23
Belém	1.270,00	-4.521,79	1.790,00	-0,47	742,81
TOTAL	1.775,00	-8.888,60	1.790,00	-0,83	761,59

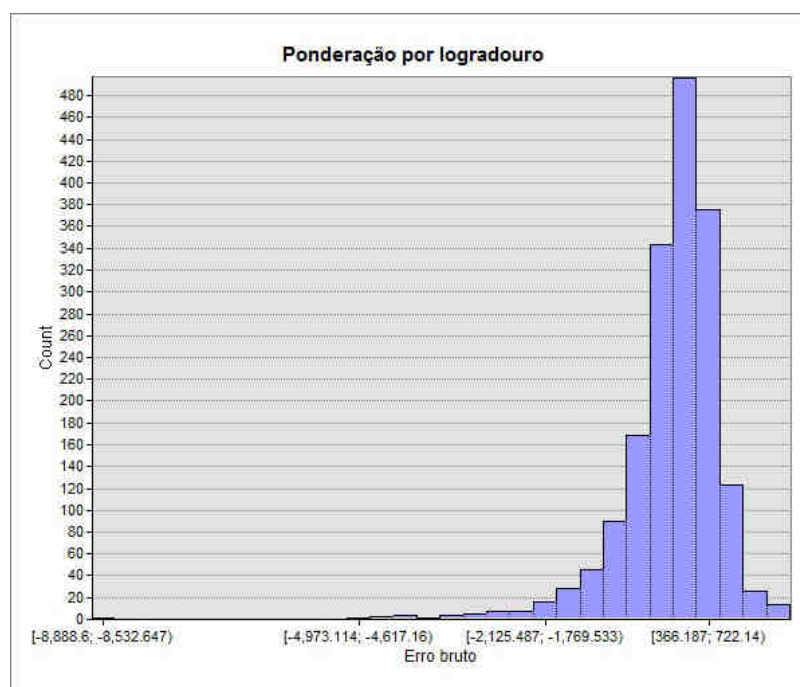


Gráfico 3: Distribuição do erro na interpolação por logradouro.

5.5.1. A interpolação ponderada por endereços nas áreas rurais

A interpolação ponderada por endereço foi aplicada aos setores rurais onde havia endereços residenciais do CNEFE georreferenciados, conforme apresentado na **Figura 31**. Na região, o município de Bujarú foi o único que apresentou todos os endereços rurais georreferenciados.

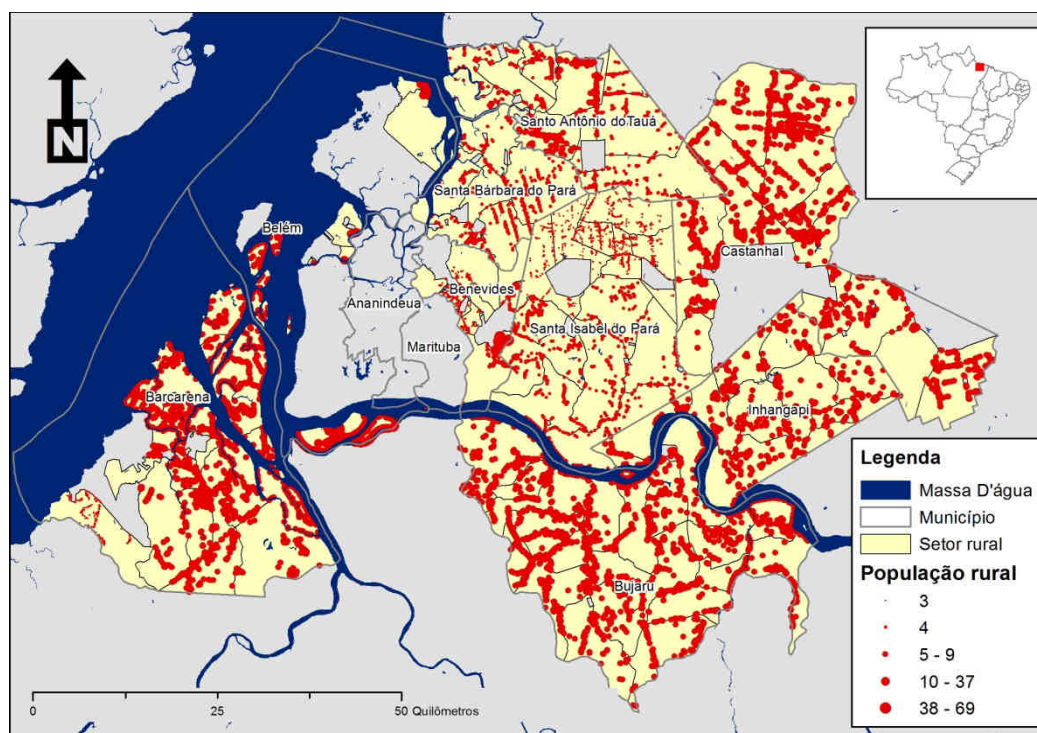


Figura 31: Mapa de pontos com população estimada por endereços rurais do CNEFE na MMB

Os endereços foram agregados por subdistritos conforme ilustrado no fluxograma da **Figura 25** e a seguir foram calculados os pesos e população média por endereço. Com uma operação conhecida como *dissolve*, obteve-se um arquivo de feições pontuais que representam o número médio de habitantes para cada combinação única de logradouro, número de logradouro, latitude e longitude. A estimativa da população por setor rural foi calculada pelo somatório destes valores nos pontos integralmente contidos em cada setor. O mapa da população estimada pelo método é apresentado na **Figura 32**.

Para avaliar o desempenho desta técnica, a população estimada por setor rural foi e comparada com a contagem do Censo de 2010 nos setores censitários rurais. O resultado desta comparação é ilustrado na **Tabela 5** onde se verifica que, com exceção dos municípios de Barcarena e Castanhal, a diferença entre população conhecida e estimada

apresentou média igual a zero. Na **Figura 33** é apresentada a distribuição espacial do erro verificado em toda a área de estudo

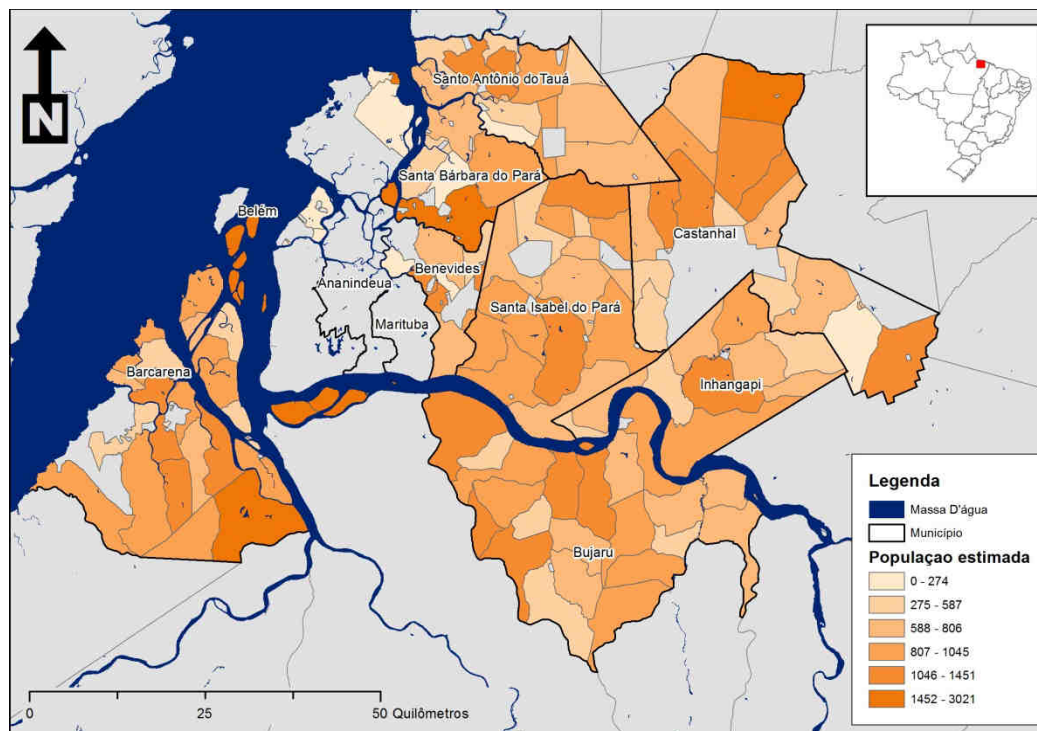


Figura 32: Mapa da população estimada por endereço rural residencial.

Tabela 5: Estatísticas de resumo para o erro verificado na ponderação por endereço.

Município	Setores rurais	Erro verdadeiro			Desvio Padrão
		Mínimo	Máximo	Média	
Barcarena	24	-3.81,04	1.430,00	105,13	408,11
Belém	10	-2.023,18	2.044,89	0	1.213,14
Benevides	9	-369,88	269,60	0	203,85
Bujarú	21	-162,95	226,72	0	124,97
Castanhal	19	-618,81	1.110,00	58,42	414,50
Inhangapi	10	-192,11	198,05	0	145,73
Santa Bárbara do Pará	6	-395,38	627,84	0	352,30
Santa Isabel do Pará	14	-193,52	289,95	0	142,35
Santo Antônio do Tauá	12	-212,78	215,10	0	132,10
TOTAL	125	-2.023,18	-2.044,89	29,06	422,42

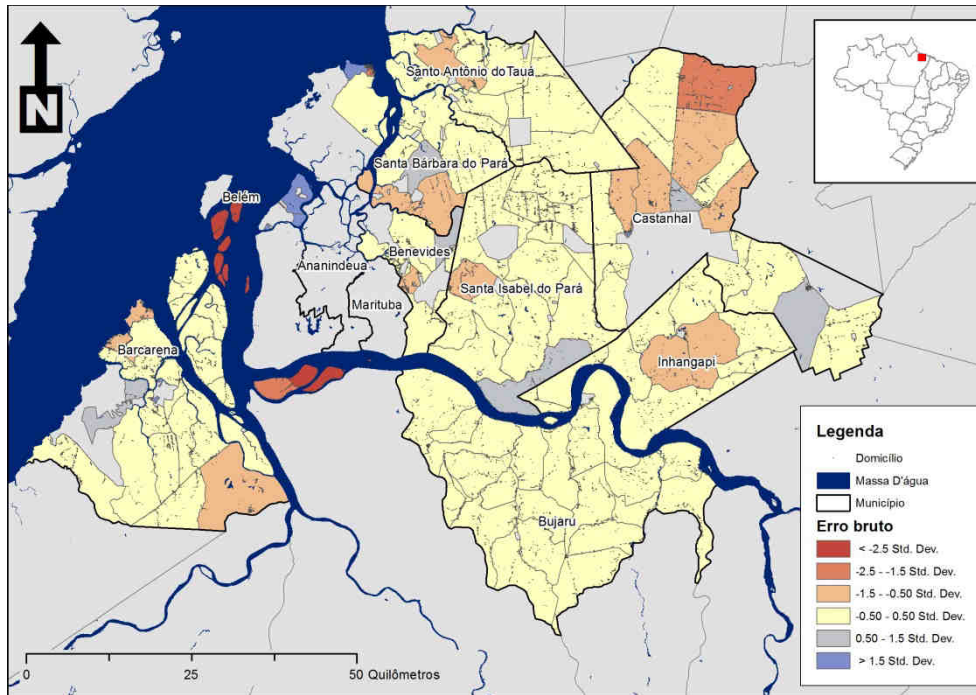


Figura 33: Distribuição espacial do erro por setor produzido na interpolação por endereço nas áreas rurais.

No histograma apresentado no **Gráfico 4**, podem-se observar uma distribuição do erro em torno da média (29,04), com coeficiente de assimetria igual a 0,27 e *kurtosis* igual a 11,84.

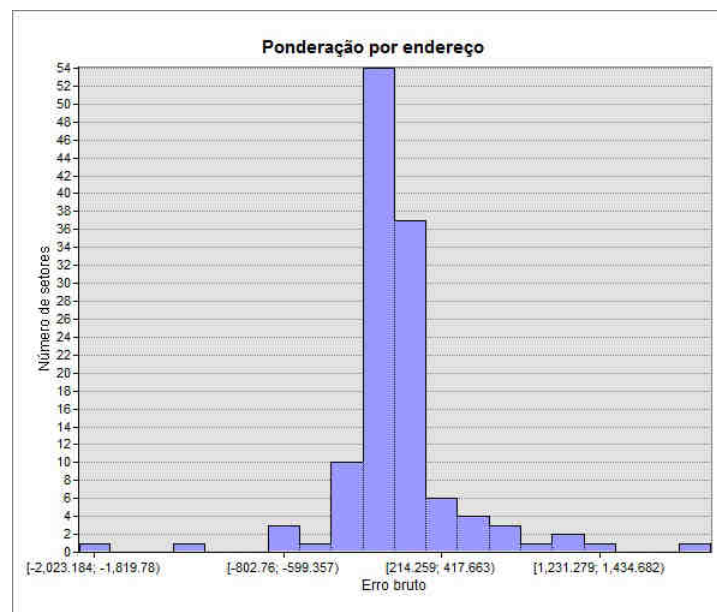


Gráfico 4: Histograma do erro bruto verificado na ponderação por endereço.

5.5.2. Análise comparativa dos métodos

Os valores de desvio padrão para cada município e tipo de setor na área de estudo são apresentados na Tabela 6.

Tabela 6: Erro absoluto verificado em cada município para os métodos avaliados.

Município	Setores	Dasimétrico Inteligente	Ponderação por Logradouro	Ponderação por Endereço
Ananideua	490	1.461,60	805,80	
rural	2	281,03		
urbano	488	1.464,50	805,80	
Barcarena	104	1.152,74	0,00	206,94
rural	70	1.339,13		206,94
urbano	34	628,93	0,00	
Belém	1320	633,07	696,14	1757,82
rural	17	341,65		1757,82
urbano	1303	635,97	696,14	
Benevides	57	991,20	0,00	217,92
rural	26	1.081,39		217,92
urbano	31	919,52	0,00	
Bujaru	29	261,25	0,00	124,97
rural	21	249,67		124,97
urbano	8	283,31	0,00	
Castanhal	192	1.225,57	0,00	336,54
rural	28	320,23		336,54
urbano	164	1.320,21	0,00	
Inhangapi	15	184,85	0,00	145,73
rural	10	204,42		145,73
urbano	5	128,17	0,00	
Marituba	110	1.657,45	0,00	
rural	3	204,20		
urbano	107	1.680,49	0,00	
Santa Bárbara do Pará	22	670,35	0,00	352,30
rural	16	645,82		352,30
urbano	6	752,20	0,00	
Santa Isabel do Pará	61	630,78	0,00	142,35
rural	20	351,62		142,35
urbano	41	733,55	0,00	
Santo Antônio do Tauá	32	663,62	0,00	132,10
rural	15	473,75		132,10
urbano	17	801,98	0,00	
TOTAL	2432	993,94	655,72	351,53

Características desejáveis nos métodos de interpolação dasimétrica compreendem uma distribuição simétrica dos erros em torno da média, que deve ser próxima a zero, desvio padrão reduzido, e poucos *outliers*. Nos três métodos utilizados neste trabalho o que melhor apresentou essas características foi a ponderação por endereço, conforme ilustrado através de *boxplots* na **Figura 34**.

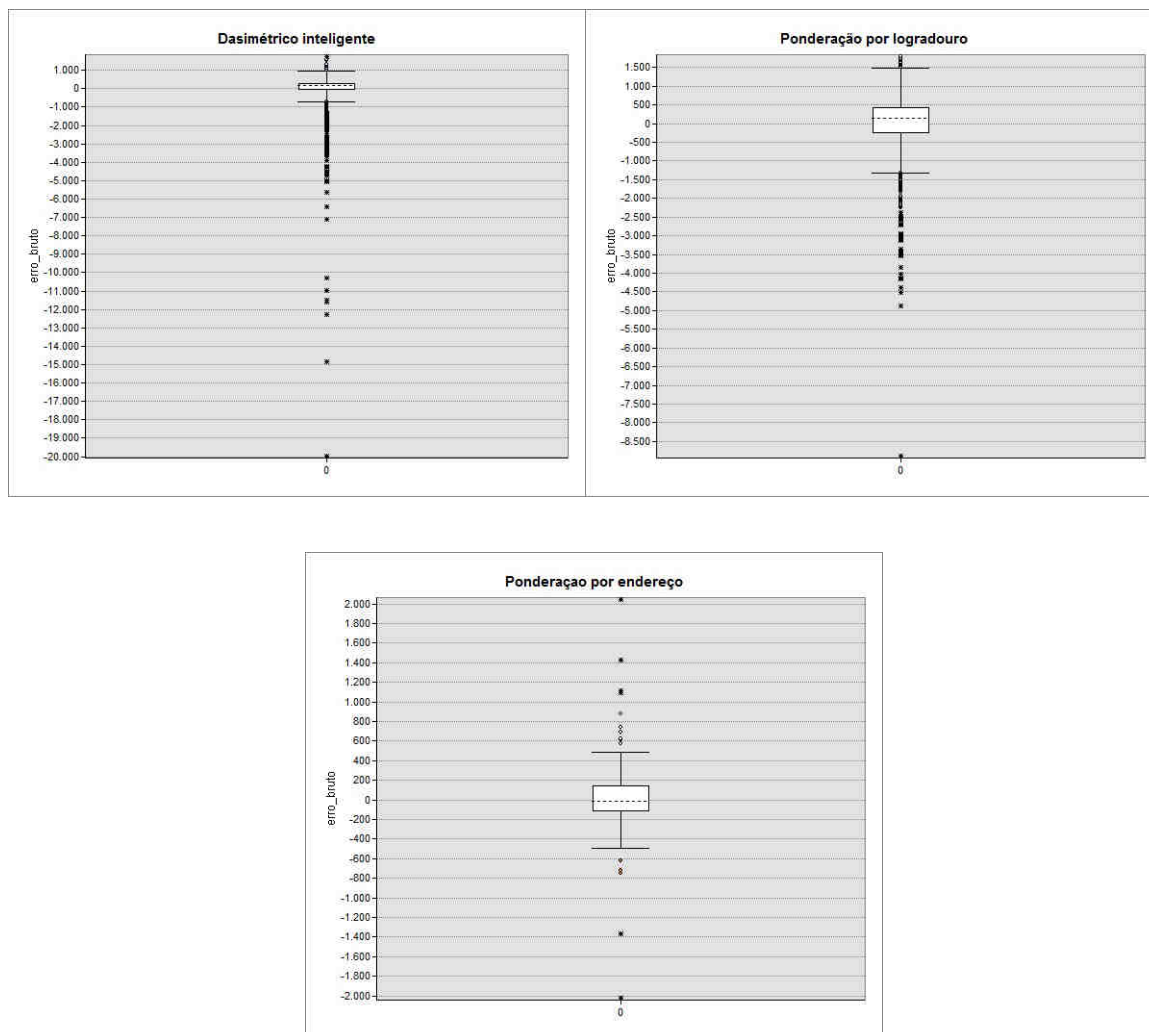


Figura 34: *Boxplots* dos erros verificados nos métodos avaliados.

6. CONCLUSÕES E TRABALHOS FUTUROS

O trabalho desenvolvido para esta dissertação teve como objetivo a avaliação de métodos dasimétricos para estimativa populacional em pequenas áreas. Para alcançar esse objetivo foram apresentadas as bases teóricas da cartografia temática, com ênfase nos métodos para representar a população, e revistos os principais métodos de interpolação zonal, com dados auxiliares. Para isso foram elaborados três esquemas metodológicos para produção e avaliação dos métodos estudados, a saber: *i)* o mapeamento dasimétrico inteligente (MENNIS e HUTGREN, 2005); *ii)* a ponderação por logradouros e a *iii)* ponderação por endereços.

O método dasimétrico inteligente combina o conhecimento especialista (usando valores de densidade predefinidos), com uma estimativa estatística (amostragem empírica) para estabelecer uma relação funcional entre as classes auxiliares de uso e cobertura da terra e a superfície estatística subjacente analisada. Além disso, a informação obtida por esses procedimentos possibilita estimar densidades para as classes cuja densidade não é conhecida nem pode ser estimada por amostragem. Outra característica do método é que ele oferece uma variedade de opções para parametrização da amostragem. Dessa forma, é possível criar um conjunto de mapas cujo desempenho pode ser comparado estatisticamente. Entretanto ressalta-se que este não foi o objetivo deste trabalho. A grande vantagem desse método é que ele é facilmente operacionalizado através de algoritmos em domínio público, tanto para a estrutura de dados vetorial quanto para estrutura matricial, que podem livremente ser modificados e adaptados.

Os resultados verificados com o método inteligente estão em conformidade com a escala cartográfica do dado auxiliar utilizado. O mapas de uso e cobertura da terra do projeto *TerraClass* são derivados de imagens de média resolução espacial e não possibilitam a distinção adequada dos tipos de uso da terra em áreas urbanas. Entretanto, pôde-se verificar que o algoritmo proporcionou um detalhamento da distribuição espacial da população que melhora a qualidade do dado produzido em relação aos mapas coropléticos. Nos setores rurais, onde a densidade populacional é baixa, o método apresentou melhor desempenho, proporcionando uma visualização cartográfica mais

intuitiva e coerente com a realidade. Além disso, o método contribui para atenuar os efeitos do problema da unidade de área modificável (MAUP). Em situações em que não existem dados de melhor resolução, o método tem grande utilidade.

Em função da eficiência verificada, acredita-se que o método pode ser uma alternativa ao método das áreas mínimas comparáveis, que ao contrário do primeiro, agrega os dados em grandes unidades espaciais para poder comparar séries temporais de dados demográficos. Em virtude disso, sugere-se a utilização do método inteligente para desagregar e compatibilizar dados de censos anteriores. A existência de longas séries temporais de imagens de sensoriamento remoto como a dos programas de observação da terra norte americano LANDSAT e do indiano IRS (*Indian Remote Sensing*), que recentemente foram disponibilizadas em domínio público, pode viabilizar a solução para este problema recorrente na análise de séries temporais. O acervo destes programas possui cerca de três décadas de imagens, que certamente viabilizaria a produção de superfícies dasimétricas referentes aos censos de 1990 e 2000. Projetos semelhantes existem nos Estados Unidos e na China, conforme descrito no capítulo 3.

Merece destaque a iniciativa do projeto *TerraClass*, uma parceria do INPE e EMBRAPA, que disponibilizou esta base de dados de uso e cobertura da terra para toda a Amazônia legal. Muitos dos artigos publicados sobre mapeamento dasimétrico utilizam bases de dados semelhantes, como o *CORINE Land Cover* da União Europeia e o *National Land Cover Data Set* dos Estados Unidos, desenvolvidas para atender uma variedade de aplicações. As imagens LANDSAT utilizadas no projeto foram obtidas no ano de 2008, portanto dois anos antecedente aos dados do censo. Esta diferença pode subestimar áreas que experimentaram crescimento populacional intenso. Entretanto, acredita-se que em função da escala espacial reduzida do modelo produzido, essa diferença não é significativa.

O método ponderado por logradouros utiliza somente dados vetoriais. É computacionalmente mais simples e mais acessível aos usuários não especializados dos SIGs. Trabalhos anteriores (BRINEGAR e POPICK, 2010; REIBEL e BUFALINO, 2005) verificaram que o método é mais eficiente em áreas urbanas do que em áreas rurais, onde a densidade populacional não apresenta necessariamente alta correlação com a densidade viária. No presente trabalho, o dado auxiliar utilizado para avaliar este método compreende os eixos de logradouros obtidos através do projeto *OpenStreetMap*. Embora a maioria das ruas já esteja mapeada nos grandes centros urbanos do Brasil, a cobertura ainda é

deficiente para áreas rurais e pequenas cidades. Este fato impediu que o método fosse aplicado em toda a área de estudo. Dessa forma, optou-se por avaliar o método somente nas áreas urbanas dos municípios de Belém e Ananideua. Os resultados obtidos indicam que o método é mais adequado do que a interpolação com classes de uso e cobertura da terra.

A ponderação por endereços é uma abordagem relativamente recente na literatura sobre mapeamento dasimétrico. Os poucos artigos publicados (BAKER *et al.*, 2012; TAPP, 2010; ZANDBERGEN, 2011) indicam que este tipo de dado auxiliar possui desempenho superior aos demais dados utilizados na interpolação de variáveis censitárias. Os resultados obtidos neste trabalho corroboram com estas indicações. A grande vantagem é que este dado permite localizar com maior exatidão a posição espacial de estruturas residenciais e não residenciais, e dessa forma estimar a população em pontos discretos.

Este último exercício, que apresentou resultados mais exatos, somente pode ser efetuado graças ao esforço do IBGE em atender à demanda da sociedade por informações georreferenciadas, disponibilizando o Cadastro Nacional de Endereços para Fins Estatísticos (CNEFE) ao público. O CNEFE é um produto que permite identificar os endereços visitados durante a operação do Censo Demográfico de 2010. Nem todos os endereços do cadastro estão georreferenciados, o que não inviabiliza sua utilização para o mapeamento dasimétrico uma vez que as coordenadas podem ser obtidas através de um processo conhecido como *geocoding*, que recupera coordenadas de endereços a partir de outras bases georreferenciadas. Este exercício não foi incluído na dissertação.

Acredita-se que o trabalho realizado nessa dissertação seja a primeira tentativa de utilizar o método no Brasil assim como o primeiro a utilizar um cadastro censitário com endereços visitados durante a coleta.

A principal contribuição deste trabalho foi a de sistematizar o conhecimento sobre os métodos dasimétricos e avaliar fontes de dados ainda não utilizadas para estimar população em pequenas áreas. Para isto foi elaborado um esquema metodológico que permite aplicar os métodos avaliados em outras áreas do território nacional, uma vez que os procedimentos para tratamento de dados foram definidos e ajustados através de rotinas de tratamento e transformação dados espaciais.

REFERÊNCIAS BIBLIOGRÁFICAS

AMARAL, S. Geoinformação para estudos demográficos: Representação Espacial de Dados de População na Amazônia Brasileira. 2003. Universidade de São Paulo, São Paulo, Brasil, 2003.

AMARAL, S.; GAVLAK, A. A.; ESCADA, M. I. S.; MONTEIRO, A. M. V. Using remote sensing and census tract data to improve representation of population spatial distribution: case studies in the Brazilian Amazon. *Population & Environment*, p. 1-29, 2012.

AMARAL, S.; MONTEIRO, A. M.V; CÂMARA, G.; QUINTANILHA, J. A. Estimativa da distribuição espacial da população na Amazônia brasileira com o uso de Sensoriamento remoto orbital: Contribuições das imagens CCD/CBERS no caso do município de Marabá-PA. 2005.

BAJAT, BRANISLAV; HENGL, TOMISLAV; KILIBARDA, MILAN; KRUNIC, NIKOLA. Mapping population change index in Southern Serbia (1961-2027) as a function of environmental factors. *Computers, Environment and Urban Systems*, v. 35, n. 1, p. 35-44, jan. 2011.

BAKER, J.; ALCANTARA, A.; RUAN, X.; WATKINS, K. The impact of incomplete geocoding on small area population estimates. *Journal of Population Research*, p. 1-22, 2012.

BHADURI, B.; BRIGHT, E.; COLEMAN, P.; URBAN, M. L. LandScan USA: A high-resolution geospatial and temporal modeling approach for population distribution and dynamics. *GeoJournal*, v. 69, n. 1, p. 103-117, 2007.

BIELECKA, E. A Dasymetric Population Density Map of Poland. 2005, [S.l: s.n.], 2005. p. 9-15.

BOSSARD, M.; FERANEC, J.; OTAHEL, J.; AGENCY, EUROPEAN ENVIRONMENT. CORINE land cover technical guide: Addendum 2000. [S.l.]: European Environment Agency Copenhagen Denmark, 2000.

BRACKEN, I.; MARTIN, D. Linkage of the 1981 and 1991 UK Censuses using surface modelling concepts. *Environment and Planning A*, v. 27, p. 379-379, 1995.

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

BRINEGAR, S. J.; POPICK, S. J. A comparative analysis of small area population estimation methods. *Cartography and Geographic Information Science*, v. 37, n. 4, p. 273-284, 2010.

CIESIN, CENTER FOR INTERNATIONAL EARTH SCIENCE INFORMATION NETWORK. Gridded Population of the World - and the Global Rural Urban

Mapping Project. Disponível em: <<http://sedac.ciesin.columbia.edu/gpw/>>. Acesso em: 30 jun. 2012.

CROSBY, A. W. The measure of reality: quantification and Western society, 1250-1600. [S.l.]: Cambridge Univ Pr, 1997.

D'ANTONA, ÁLVARO DE OLIVEIRA; DAGNINO, RICARDO SAMAPIO; BUENO, MARIA DO CARMO DIAS. Geotecnologias e gestão de políticas públicas: uso de dados demográficos. In: BAENINGER, ROSANA (Org.). População e Cidades - Subsídios para o planejamento e para as políticas sociais. [S.l.]: Núcleo de Estudos de População-Nepo/Unicamp, 2010. p. 99-116.

DE GEER, S. A Population Density Map of European Russia. Geographical Review, v. 18, n. 4, p. 616-640, 1926.

DEICHMANN, U.; BALK, D.; YETMAN, G. Transforming population data for interdisciplinary usages: from census to grid. Washington (DC): Center for International Earth Science Information Network, 2001.

DOBSON, J. E; BRIGHT, E. A; COLEMAN, P. R; DURFEE, R. C; WORLEY, B. A. LandScan: a global population database for estimating populations at risk. Photogrammetric Engineering and Remote Sensing, v. 66, n. 7, p. 849-857, 2000.

DORLING, D.; FAIRBAIRN, D. Mapping: ways of representing the world. [S.l.]: Prentice Hall, 1997.

DUTENKEFER, EDUARDO. Representações do espaço geográfico: mapas dasimétricos, anamorfozes e modelização gráfica. 2010. Dissertação – USP, Faculdade de Filosofia, Letras e Ciências Humanas, São Paulo, Brasil, 2010. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/8/8136/tde-25022011-115539/pt-br.php>>.

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

ELVIDGE, C. D; BAUGH, K. E; DIETZ, J. B; BLAND, T.; SUTTON, P. C; KROEHL, H. W. Radiance calibration of DMSP-OLS low-light imaging data of human settlements. Remote Sensing of Environment, v. 68, n. 1, p. 77-88, 1999.

ELVIDGE, C. D; BAUGH, K. E; KIHN, E. A; KROEHL, H. W; DAVIS, E. R; DAVIS, C. W. Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. International Journal of Remote Sensing, v. 18, n. 6, p. 1373-1379, 1997.

ESRI, URL. ESRI Shapefile Technical Description. [S.l.]: URL: <http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf> [29th March 2005], 1998.

FISHER, P. F; LANGFORD, M. Modelling the errors in areal interpolation between zonal systems by Monte Carlo simulation. Environment and Planning A, v. 27, p. 211-211, 1995.

FRIENDLY, M. The golden age of statistical graphics. *Statistical Science*, p. 502-535, 2008.

FRIENDLY, M.; DENIS, D. J. Milestones in the history of thematic cartography, statistical graphics, and data visualization. Retrieved August, v. 7, p. 2005, 2005.

GOODCHILD, M. F. Geographical data modeling. *Computers & Geosciences*, v. 18, n. 4, p. 401-408, 1992.

GREEN, N. E. Aerial photographic analysis of residential neighborhoods: an evaluation of data accuracy. *Soc. F.*, v. 35, p. 142, 1956.

HAKKERT, R. Fontes de dados demográficos. Belo Horizonte: ABEP, p. 1870-1950, 1996.

HALL, O.; STROH, E.; PAYA, F. From Census to Grids: Comparing Gridded Population of the World with Swedish Census Records. *Open Geography Journal*, v. 5, p. 1-5, 2012.

HARNESS, H. D. Atlas to accompany the second report of the railway commissioners, Ireland. Dublin: Irish Railway Commission, 1837.

HAWLEY, K.; MOELLERING, H. A comparative analysis of areal interpolation methods. *Cartography and Geographic Information Science*, v. 32, n. 4, p. 411-423, 2005.

HAY, S. I.; NOOR, A. M.; NELSON, A.; TATEM, A. J. The accuracy of human population maps for public health application. *Tropical Medicine & International Health*, v. 10, n. 10, p. 1073-1086, 2005.

HSU, S. Y. Population estimation. *Photogrammetric Engineering*, v. 37, n. 5, p. 449-454, 1971.

HULTGREN, T. Raster-based automated dasymetric mapping. *Proceedings: UCGIS Assembly 2004*, 2004.

IBGE, INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Base de informações do Censo Demográfico 2010: Resultados do Universo por setor censitário. Documentação do Arquivo. . [S.l: s.n.]. . Acesso em: 20 dez. 2011. , dez. 2011

IBGE, INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Censo Demográfico 2010 - Resultados gerais da amostra. [S.l: s.n.], 2012a.

IBGE, INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. IBGE: Instituto Brasileiro de Geografia e Estatística. Disponível em: <<http://www.ibge.gov.br/home/estatistica/populacao/censo2010/calendario.shtm>>. Acesso em: 29 out. 2012b.

KISH, G. La carte: image des civilisations. Paris: Seuil, 1980.

LAM, N. S. An evaluation of areal interpolation methods. 1982, [S.l: s.n.], 1982.

LAM, N. S.N. Spatial interpolation methods: a review. *Cartography and Geographic Information Science*, v. 10, n. 2, p. 129-150, 1983.

LANGFORD, M.; MAGUIRE, D. J; UNWIN, D. J. The areal interpolation problem: estimating population using remote sensing in a GIS framework. *Handling geographical information: Methodology and potential applications*, p. 55-77, 1991.

LANGFORD, M.; UNWIN, D. J. Generating and mapping population density surfaces within a geographical information system. *Cartographic Journal*, The, v. 31, n. 1, p. 21-26, 1994.

LEWIS, DANIEL. Carl Steinitz, Symap and Place. *Volunteered Geographic Information*. [S.l: s.n.]. Disponível em: <<http://danieljlewis.org/2010/08/11/carl-steinitz-symap-and-place/>>. Acesso em: 27 maio 2012. , ago. 2010

LI, G.; WENG, Q. Fine-scale population estimation: how Landsat ETM+ imagery can improve population distribution mapping. *Canadian Journal of Remote Sensing*, v. 36, n. 3, p. 155-165, 2010.

LI, JIN; HEAP, ANDREW D. A review of spatial interpolation methods for environmental scientists. Canberra: Geoscience Australia, 2008. Disponível em: <http://www.ga.gov.au/image_cache/GA12526.pdf>. Acesso em: 2 jun. 2011. (, 23).

LO, C. P.; WELCH, R. Chinese urban population estimates. *Annals of the Association of American Geographers*, v. 67, n. 2, p. 246-253, 2005.

LWIN, K. K; MURAYAMA, Y. Estimation of Building Population from LIDAR Derived Digital Volume Model. *Spatial Analysis and Modeling in Geographical Transformation Process*. [S.l: s.n.], 2011. p. 87-98. Disponível em: <<http://www.springerlink.com/content/q106l606w7445543/>>.

MAANTAY, J. A; MAROKO, A. R; HERRMANN, C. Mapping population distribution in the urban environment: the cadastral-based expert dasymetric system (CEDs). *Cartography and Geographic Information Science*, v. 34, n. 2, p. 77-102, 2007.

MACEACHREN, A. M. The evolution of thematic cartography/a research methodology and historical review. *Cartographica: The International Journal for Geographic Information and Geovisualization*, v. 16, n. 1, p. 17-33, 1979.

MARTIN, D. 2001 Census output areas: from concept to prototype. *trends*, 1998a.

MARTIN, D. An assessment of surface and zonal models of population. *International Journal of Geographical Information Science*, v. 10, n. 8, p. 973-989, 1996.

MARTIN, D. Mapping population data from zone centroid locations. *Transactions of the Institute of British Geographers*, v. 14, n. 1, p. 90-97, 1989.

MARTIN, D. Optimizing census geography: the separation of collection and output geographies. *International Journal of Geographical Information Science*, v. 12, n. 7, p. 673-685, 1998b.

MARTINELLI, M. Os mapas da geografia. 2005, [S.l: s.n.], 2005.

MARTINELLI, M. Relevo do Estado de São Paulo. *Confins. Revue franco-brésilienne de géographie/Revista franco-brasileira de geografia*, n. 7, 2009.

MARTINELLI, M. Um Breve Apanhado Sobre a Breve História da Cartografia Temática. In: SIMPÓSIO IBEROAMERICANO DE HISTÓRIA DA CARTOGRAFIA, 2010, São Paulo. Anais... São Paulo: [s.n.], 2010.

MENNIS, J. Generating surface models of population using dasymetric mapping. *Professional Geographer*, v. 55, n. 1, p. 31-42, 2003.

MENNIS, J. Using geographic information systems to create and analyze statistical surfaces of population and risk for environmental justice analysis. *Social Science Quarterly*, v. 83, n. 1, p. 281-297, 2002.

MENNIS, J.; HULTGREN, T. Dasymetric mapping for disaggregating coarse resolution population data. 2005, [S.l: s.n.], 2005. p. 9-16.

MENNIS, JEREMY; HULTGREN, TORRIN. Intelligent dasymetric mapping and its application to areal interpolation. *Intelligent dasymetric mapping and its application to areal interpolation*, v. 33, n. 3, 2006 Disponível em: <http://astro.temple.edu/~jmennis/pubs/mennis_cagis06.pdf>.

MOON, Z. K; FARMER, F. L. Population density surface: a new approach to an old problem. *Society & Natural Resources*, v. 14, n. 1, p. 39-51, 2001.

MORAES, A. C.R. Geografia: pequena história crítica. [S.l.]: Annablume, 2005.

MORAL, F. J. La representación gráfica de las variables regionalizadas: Geoestadística lineal. [S.l.]: Universidad de Extremadura, Servicio de Publicaciones, 2003.

NORDBECK, S. The law of allometric growth. [S.l.]: Department of Geography, University of Michigan, 1965.

OPENSHAW, S. Ecological fallacies and the analysis of areal census data. *Environment and Planning A*, v. 16, n. 1, p. 17-31, 1984.

OPENSHAW, S.; TAYLOR, P. The modifiable areal unit problem. [S.l.]: N, 1984. v. 38.

ORNL, OAK RIDGE NATIONAL LABORATORY. LandScan Home. Disponível em: <<http://www.ornl.gov/sci/landscan/>>. Acesso em: 30 jun. 2012.

OSM, OPENSTREETMAP. OpenStreetMap. Disponível em: <<http://www.openstreetmap.org/>>. Acesso em: 17 ago. 2012.

PALSKY, G. Des chiffres et des cartes: naissance et développement de la cartographie quantitative française au XIXe siècle. [S.l.]: Comité des travaux historiques et scientifiques, 1996. v. 19.

PARENT-DUCHATELET, A. J.B. De la prostitution dans la ville de Paris: considérée sous le rapport de l'hygiène publique, de la morale et de l'administration; ouvrage appuyé de documents statistiques... [S.l.]: Établissement encyclographique, 1837.

PETROV, A. N. Setting the Record Straight: On the Russian Origins of Dasymetric Mapping. *Cartographica: The International Journal for Geographic Information and Geovisualization*, v. 43, n. 2, p. 133-136, 2008.

PORTER, P. W. Population distribution and land use in Liberia. Unpublished Ph. D. Thesis, University of London, London, 1957.

QIU, F.; SRIDHARAN, H.; CHUN, Y. Spatial autoregressive model for population estimation at the census block level using LIDAR-derived building volume information. *Cartography and Geographic Information Science*, v. 37, n. 3, p. 239-257, 2010.

RASE, W. D. Volume-preserving interpolation of a smooth surface from polygon-related data. *Journal of Geographical Systems*, v. 3, n. 2, p. 199-213, 2001.

RECLUS, E. Nouvelle géographie universelle: la terre et les hommes. [S.l.]: Hachette et cie., 1888. v. 13.

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

ROBINSON, A. H. The 1837 maps of Henry Drury Harness. *The Geographical Journal*, v. 121, n. 4, p. 440-450, 1955.

ROBINSON, W. S. Ecological correlations and the behavior of individuals. *International journal of epidemiology*, Reprinted with permission, v. 38, n. 2, p. 337, 2009.

ROBINSON, W. S. Ecological Correlations and the Behavior of Individuals. *American Sociological Review*, v. 15, n. 351-357, 1950.

SABESAN, A.; ABERCROMBIE, K.; GANGULY, A. R.; BHADURI, B.; BRIGHT, E. A.; COLEMAN, P. R. Metrics for the comparative analysis of geospatial datasets with applications to high-resolution grid-based population data. *GeoJournal*, v. 69, n. 1, p. 81-91, 2007.

SCROPE, G. P. Principles of political economy. [S.l.]: Longman, Rees, Orme, Brown, Green, & Longman, 1833.

SEMENOV-TIAN-SHANSKY, B. Russia: Territory and Population: A Perspective on the 1926 Census. *Geographical Review*, p. 616-640, 1928.

SILVA, F. B. Modelação cartográfica e ordenamento do território: Um ensaio metodológico de cartografia dasimétrica aplicado à região Oeste e Vale do Tejo. 2009. Dissertação de Mestrado – 2009.

SLEETER, R. Dasymetric mapping techniques for the San Francisco Bay region, California. 2004, [S.l.: s.n.], 2004. p. 7–10.

SLEETER, R.; GOULD, M. Geographic information system software to remodel population data using dasymetric mapping methods. US Geological Survey, Techniques and Methods, p. 11-C2, 2007.

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

TAPP, A. F. Areal interpolation and dasymetric mapping methods using local ancillary data sources. Cartography and Geographic Information Science, v. 37, n. 3, p. 215-228, 2010.

TOBLER, W.; DEICHMANN, U.; GOTTSEGEN, J.; MALOY, K. World population in a grid of spherical quadrilaterals. International Journal of Population Geography, v. 3, n. 3, p. 203-225, 1997.

TOBLER, W. R. Satellite confirmation of settlement size coefficients. Area, v. 1, n. 3, p. 30-34, 1969.

TOBLER, W. R. Smooth pycnophylactic interpolation for geographical regions. Journal of the American Statistical Association, p. 519-530, 1979.

UNSD, UNITED NATIONS STATISTICAL DIVISION. Principles and recommendations for population and housing censuses. [S.l.]: United Nations Pubns, 2008.

VOSS, P. R.; LONG, D. D.; HAMMER, R. B.; ECOLOGY, UNIVERSITY OF WISCONSIN--MADISON. CENTER FOR DEMOGRAPHY AND. When census geography doesn't work: Using ancillary information to improve the spatial interpolation of demographic data. [S.l.]: Citeseer, 1999.

WANG, S. Job density and employment subcenters in the four US metropolitan areas. 2012. NORTHERN ILLINOIS UNIVERSITY, 2012.

WEBER, E. M. DasyTools: A user-friendly dasymetric mapping toolkit with automated accuracy assessments. 2010.

WELCH, R.; ZUPKO, S. Urbanized area energy utilization patterns from DMSP data. Photogrammetric Engineering and Remote Sensing, v. 46, n. 2, p. 201-207, 1980.

WORBOYS, M. F. A unified model for spatial and temporal information. The Computer Journal, v. 37, n. 1, p. 26-34, 1994.

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

WU, S.; QIU, X.; WANG, L. Population estimation methods in GIS and remote sensing: a review. *GIScience & Remote Sensing*, v. 42, n. 1, p. 80-96, 2005.

YANG, X.; HUANG, Y.; DONG, P.; JIANG, D.; LIU, H. An updating system for the gridded population database of China based on remote sensing, GIS and spatial database technologies. *Sensors*, v. 9, n. 2, p. 1128-1140, 2009.

ZANDBERGEN, P. A. Dasymetric Mapping Using High Resolution Address Point Datasets. *Transactions in GIS*, v. 15, p. 5-27, 2011.

ZANDBERGEN, P. A.; IGNIZIO, D. A. Comparison of Dasymetric Mapping Techniques for Small-Area Population Estimates. *Cartography and Geographic Information Science*, v. 37, n. 3, p. 199-214, 2010.

GLOSSÁRIO

ATLAS

Conjunto ou coleção de mapas ou cartas geográficas reunidos em um volume.

BANDA ESPECTRAL

Porção do espectro eletromagnético, limitada por dois comprimentos de onda, duas frequências ou dois números de onda.

CARTA

É a representação no plano, em escala média ou grande, dos aspectos artificiais e naturais de uma determinada área de uma superfície planetária, subdividida em folhas delimitadas por linhas convencionais - paralelos e meridianos - com a finalidade de possibilitar a avaliação de pormenores, com grau de precisão compatível com a escala.

CARTOGRAMA

É um esquema representativo de uma superfície ou parte dela, sobre a qual são apresentadas informações quantitativas e qualitativas, de eventos geográficos, cartográficos e socioeconômicos.

COROCROMÁTICO

Os mapas corocromáticos apresentam dados geográficos e utilizam diferenças de cor na implantação zonal. Este método deve ser empregado sempre que for preciso mostrar diferenças nominais em dados qualitativos, sem que haja ordem ou hierarquia.

COROPLÉTICO

Método de representação cartográfica por meio da primitiva gráfica área para representação de valores absolutos e valores relativos, sendo mais recomendado para estes últimos. Introduzido no início do século XIX, tem sido amplamente utilizado por cartógrafos e geógrafos por ser de fácil assimilação para o usuário. Os símbolos criados a partir de "polígonos" coincidem com as regiões onde foram coletados os dados, o que dá a impressão de que há uniformidade de dados dentro de cada uma das regiões e que as quebras ocorrem sempre nos limites destas áreas. Assim, cada unidade de interesse passa a ser destacada de forma a representar diferentes magnitudes de um determinado atributo, ou seja, proporcionalmente ao nível de medida da variável estatística que está sendo retratada no mapa. A legenda de um mapa coroplético é dividida em classes conforme os métodos de classificação utilizados (intervalos iguais, quantis, médias, quebras-máximas, quebras-naturais, fisher-jenks e etc.) e regras próprias de utilização da variável visual. As variáveis visuais mais utilizadas em mapas coropléticos são os valores de cor, variando sua intensidade conforme a seqüência de valores apresentados nas classes estabelecidas. Vários são os fatores a serem considerados para a seleção de um esquema de cores como o tipo de dados (unipolar, bipolar e balanceado), o tipo de saída de dados (impressão em papel, saída em monitor vga e etc), razões estéticas, uso do mapa e o tipo de audiência deste.

DASIMÉTRICO

O mapeamento dasimétrico busca estabelecer limites mais realistas que os mapas coropléticos para a distribuição da variável mapeada. Para tanto, utiliza informação

adicional. Por exemplo, o mapeamento da densidade demográfica por meio de classes de uso e cobertura da terra. Toda representação dasimétrica é, em determinado sentido, arbitrária, pois encontra-se dependente da natureza e escala das informações adicionais utilizadas.

GEOSSERVIÇO

Termo adotado pela Infraestrutura Nacional de Dados Espaciais (INDE) para designação dos *Web Services* padronizados pelo *Open Geospatial Consortium* (OGC) para integração e intercâmbio de dados entre Sistemas de Informação Geográfica.

MAPA ISOPLÉTICO OU ISARÍTMICO

Mapa temático que utiliza isolinhas (linhas formadas pelos pontos que possuem o mesmo valor) para representação de um fenômeno de valores contínuos. Também conhecido como mapa de contornos, é uma representação planimétrica para uma superfície em três dimensões.

MAPA HIPSOMÉTRICO

Mapa temático que representa as classes de altitude através de cores hipsométricas.

PROPRIEDADE PICNOFILÁTICA

Também conhecida como preservação de volume, é uma propriedade desejável aos métodos de interpolação zonal, que garante que o valor total dos dados coletados nas zonas de origem mantém-se inalterado após processos de transformação de dados entre diferentes recortes espaciais.

PLANISFÉRIO

Representação de toda a superfície esférica da Terra sobre uma folha plana. Também conhecido como planisfério.

ROSA DOS VENTOS

(pré-AO 1990: rosa-dos-ventos) é uma imagem que representa as quatro direções fundamentais e suas intermediárias. Corresponde à volta completa do horizonte e surgiu da necessidade de indicar exatamente uma direção que nem mesmo os pontos intermediários determinariam, pois um mínimo desvio inicial torna-se cada vez maior, à medida que vai aumentando a distância. Assim, praticamente todos os pontos na linha do horizonte podem ser localizados com exatidão. A utilização de rosas dos ventos é extremamente comum em todos os sistemas de navegação antigos e atuais. Seu desenho em forma de estrela tem a finalidade única de facilitar a visualização com o balanço da embarcação.

SISTEMA DE INFORMAÇÃO GEOGRÁFICA (SIG)

Também conhecido por sua sigla em Inglês: GIS (*Geographic Information System*), é Conjunto de procedimentos, manual ou automatizado, utilizados no sentido do armazenamento, e manipulação de informação georreferenciada.”

SISTEMAS GERENCIADORES DE BANCOS DE DADOS (SGBD)

Do inglês Data Base Management System (DBMS) - é o conjunto de programas de computador (*softwares*) responsáveis pelo gerenciamento de uma base de dados. Seu principal objetivo é retirar da aplicação cliente a responsabilidade de gerenciar o acesso, a manipulação e a organização dos dados. O SGBD disponibiliza uma interface para que seus clientes possam incluir, alterar ou consultar dados previamente armazenados. Em bancos de dados relacionais a interface é constituída pelas APIs (*Application Programming*

Interface) ou drivers do SGBD, que executam comandos na linguagem SQL (*Structured Query Language*).

APENDICE A – DIAGRAMAS DE DISPERSÃO

