

Járvis Campos

**Estimativas populacionais a partir de
dados orbitais de média resolução
espacial: aplicações em municípios da
Região Metropolitana de Belo Horizonte**

Belo Horizonte, MG
UFMG/Cedeplar
2017

Járvis Campos

Estimativas populacionais a partir de dados orbitais de média resolução espacial: aplicações em municípios da Região Metropolitana de Belo Horizonte

Tese apresentada ao curso de Doutorado em Demografia, do Centro de Desenvolvimento e Planejamento Regional, da Faculdade de Ciências Econômicas, da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do Título de Doutor em Demografia.

Orientador: Prof. José Irineu Rangel Rigotti
Co-orientador: Prof. Antônio Miguel Vieira Monteiro
Co-orientador: Prof^a Ilka Afonso Reis

Belo Horizonte, MG
Centro de Desenvolvimento e Planejamento Regional
Faculdade de Ciências Econômicas - UFMG
2017

Ficha Catalográfica

C198e
2017

Campos, Járvis.
Estimativas populacionais a partir de dados orbitais de média resolução espacial [manuscrito] : aplicações em municípios da Região Metropolitana de Belo Horizonte / Járvis Campos. – 2017. 238 f.: il., gráfs e tabs.

Orientador: José Irineu Rangel Rigotti.
Coorientador: Antônio Miguel Vieira Monteiro
Coorientador: Ilka Afonso Reis
Tese (doutorado) - Universidade Federal de Minas Gerais, Centro de Desenvolvimento e Planejamento Regional.
Inclui bibliografia (f. 231-238).

1. População - Pesquisa – Minas Gerais - Teses. 2. Sensoriamento remoto – Minas Gerais - Teses. 3. Demografia – Teses. I. Rigotti, Jose Irineu Rangel. II. Monteiro, Antonio Miguel Vieira. III. Reis, Ilka Afonso. IV. Universidade Federal de Minas Gerais. Centro de Desenvolvimento e Planejamento Regional. V. Título.

CDD: 304.6098151

Elaborada pela Biblioteca da FACE/UFMG – FPS 054 /2018

Folha de Aprovação

Este trabalho é dedicado à Camila,
às minhas filhas, Laís e Laura,
aos meus pais e à avó Margherida.

“Conhecereis a verdade, e a verdade vos libertará”

(Bíblia, João, 8, 32).

AGRADECIMENTOS

Agradeço à Deus pela oportunidade dada a mim de realizar o sonho de me tornar Demógrafo. Nessa longa caminhada, tive o privilégio de conviver e de ser rodeado por pessoas muito especiais, que contribuíram para o alcance deste objetivo e cujo sentimento de gratidão é enorme.

Camila, obrigado pelo amor expresso em companheirismo, empenho e compreensão, mas, principalmente, obrigado pelo desprendimento de ter feito do meu, o nosso sonho! Essa tese é dedicada a você e à nossa família.

Laís, você é o maior presente que recebi de Deus. Você veio ao mundo nessa época, e mesmo tão pequena, esteve sempre ao meu lado e sem saber dos desafios que eu enfrentara. Mas nem foi preciso. Nossos momentos juntos me deram a força e a motivação necessárias para vencer essa etapa.

Agradeço à minha mãe por ter sempre me apoiado e por toda a dedicação ao longo da minha vida. Essa conquista é também dedicada a você, mãe. Assim como minha mãe, sou muito grato ao meu pai e à minha avó Margherida por toda a minha formação e por todos os esforços dedicados. Faço essa homenagem póstuma ao papai e à vovó, que sempre estiveram comigo, tenho certeza! Essa conquista é de todos nós! À minha família, especialmente irmãos, sobrinhos e cunhados, obrigado por todo o carinho.

À Miriam, obrigado por toda a força e companheirismo dedicados nesse período. Sua presença foi fundamental para a superação de muitas dificuldades. À avó Nausa, pela superação, por ser essa guerreira e principalmente exemplo.

Ao Irineu, professor, orientador e amigo, agradeço imensamente por tudo o que fez por mim ao longo desses anos. Irineu sempre acreditou no meu trabalho, mesmo nos momentos difíceis, me abrindo muitas portas e tendo um papel fundamental na minha formação. Para além do profissional, fui beneficiado por momentos de grande aprendizado, através do exemplo de pessoa que o Irineu representa, para mim um mestre, com o qual tive a oportunidade de conviver. Sou muito grato por toda essa grande oportunidade que me foi dada ao longo desses anos. Sobre a tese, a semente deste trabalho teve origem num conselho dado por Irineu, de

buscar ir além das “fronteiras convencionais” da Demografia, e cujo caminho, embora árduo, foi fundamental para o meu crescimento profissional. Obrigado por tudo!

Ao professor e co-orientador Antônio Miguel, agradeço primeiramente por todo o apoio dedicado e por ter viabilizado o meu intercâmbio no INPE. O professor Miguel não apenas me recebeu muito bem em São José dos Campos, como permitiu que eu tivesse uma grande inserção e integração no doutorado em sensoriamento remoto no INPE, através da possibilidade de participação formal como aluno de disciplinas obrigatórias e optativas do curso, algo pouco comum e que fora fundamental para a minha formação e para a realização deste trabalho. Além disso, o professor Miguel foi sempre muito acessível e atencioso, tendo atuado de forma muito importante na definição do tema, além de ter tido grande participação nas correções e nos direcionamentos do trabalho. Obrigado, Miguel!

À professora e co-orientadora Ilka, eu gostaria de agradecer muito pela imensa ajuda dedicada ao longo do trabalho. De forma sempre muito gentil, solícita e muito paciente, a professora Ilka me ajudou enormemente com todas as etapas da tese. O apoio técnico na fase de modelagem foi determinante, sem o qual eu não teria conseguido. Contudo, se não bastasse todo o apoio nessa etapa de dados, a professora Ilka foi muito dedicada nas orientações de conceitos, e, especialmente, na correção minuciosa do trabalho. Ilka, obrigado por tudo!

Aos professores do Cedeplar eu gostaria de agradecer por toda a minha formação, sem dúvida marcada pela excelência e profissionalismo de todos os mestres. Além da sala de aula, tive a sorte de ter participado de muitos projetos, e cujas experiências foram muito importantes para minha formação como demógrafo.

À professora Laura eu gostaria de agradecer pelas várias oportunidades de trabalho ao longo do curso, em projetos tão especiais como os do Laboratório de Projeções, além do projeto da UNFPA, sobre o Afeganistão. Por outro lado, agradeço pelo enorme carinho no dia a dia, pela simplicidade, e por deixar as portas sempre abertas para o convívio e à amizade!

Ao professor José Alberto, muito obrigado por todas as parcerias de trabalho, pela amizade e convívio construídos ao longo desses anos, e, especialmente, por todo

apoio concedido nos momentos difíceis. José Alberto é a pessoa mais humana que conheci em toda minha vida. Me sinto honrado por ter vivido tantos bons momentos, e por ter aprendido tanto, ao lado do grande mestre.

Ao professor Alisson, obrigado por sempre ter acreditado no meu trabalho, e por todas as oportunidades e portas que me foram abertas, de forma sempre tão positiva, e que foram fundamentais para a minha formação. Tenho uma grande admiração pelo professor Alisson, além de gratidão por toda a atenção dedicada ao longo do curso. Muito obrigado.

Professor Gilvan, nesse mesmo sentimento eu lhe agradeço por todo o incentivo, pela força, especialmente numa fase mais avançada do doutorado, com grandes oportunidades e aprendizados, dos quais sou muito grato.

Obrigado professora Simone, pelo grande incentivo que me dera enquanto coordenadora, para a realização do meu intercâmbio no INPE, além de todo o apoio e convívio durante o curso.

Aos demais professores, Fausto, Dimitri, Eduardo, Paula, Bernardo, Cássio, Ana Paula, Adriana, muito obrigado por todo o convívio.

Aos professores Renato Moreira Hadad e Alexandre Diniz, quero dividir a minha satisfação pela participação de vocês na minha banca examinadora. Há dez anos tive esse privilégio na minha defesa de mestrado, e sou grato por vivenciar novamente um momento como este ao lado de vocês.

Agradeço à Amplo Engenharia, nas pessoas do Jackson e Luiz Cláudio, pela oportunidade de trabalho e pela enorme compreensão nos momentos de elaboração da tese.

Aos amigos cedeplarianos Marcelo, Márcio e Fátima, sou grato pelos grandes desafios que passamos juntos, durante o curso! Sem dúvida formamos uma bela equipe. Porém, agradeço principalmente à grande amizade construída com cada um, amizade que transcende ao campus da UFMG e que foi elemento importante nessa caminhada.

À Bruna, Juliana, Gabriela, Cristiano e Emerson, mesmo sem dividir as salas de aula no dia a dia da coorte de 2012, sou muito grato pelos momentos que passamos juntos! Vocês são amigos muito queridos, e pessoas que aprendi a admirar.

Aos funcionários do Cedeplar, nas pessoas da Cecília e do Sebastião, eu gostaria de agradecer por toda a competência e por todo o convívio, ao longo do tempo.

Ao Charles, Guilherme, Matheus Valle, Isabela, Tininha, Matheus Fritzsos, Pedro e demais colegas da Socioeconomia da Amplo Engenharia, obrigado pela imensa força nos momentos de elaboração de tese.

À Cláudia, Thaís, Wagner, Ana Márcia, Ricardo e demais amigos, muito obrigado pela força e pela torcida.

Ao CNPq eu agradeço pelo financiamento da bolsa de doutorado, que possibilitou a realização de todas as etapas da minha formação no doutorado.

Às demais pessoas que colaboraram e que torceram por mim, muito obrigado.

SUMÁRIO

1	INTRODUÇÃO E JUSTIFICATIVA	20
1.1	OBJETIVOS.....	33
2	INTRODUÇÃO ÀS ESTIMATIVAS POPULACIONAIS PARA PEQUENAS ÁREAS E OS CENSOS E ESTIMATIVAS DO IBGE	37
2.1	MÉTODOS TOP-DOWN.....	39
2.1.1	<i>Métodos de razão</i>	39
2.1.2	<i>Relação de Coortes</i>	42
2.2	MÉTODOS BOTTOM-UP	44
2.2.1	<i>Microsimulação</i>	44
2.2.2	<i>Modelos de difusão espacial</i>	46
2.3	ESTIMATIVAS POPULACIONAIS PARA PEQUENAS ÁREAS A PARTIR DE VARIÁVEIS SINTOMÁTICAS	49
2.3.1	<i>Distribuição Pro Rata</i>	51
2.3.2	<i>Distribuição Proporcional</i>	52
2.3.3	<i>Taxas Vitais</i>	52
2.3.4	<i>Razão Censitária</i>	53
2.3.5	<i>Diferença de Taxa</i>	55
2.3.6	<i>Método Composto</i>	55
2.3.7	<i>Correlação de Razões</i>	56
2.3.8	<i>Correlação de Taxa</i>	57
2.3.9	<i>Correlação de Diferença</i>	58
2.3.10	<i>Aplicações com foco na América Latina</i>	59
2.4	PROJEÇÕES OFICIAIS DO IBGE (REVISÕES 2008 E 2013).....	63
2.5	ESTIMATIVAS PÓS-CENSITÁRIAS OFICIAIS DO IBGE	66
2.6	AVALIAÇÃO DOS ERROS DAS ESTIMATIVAS.....	67
3	ESTIMATIVAS POPULACIONAIS PARA PEQUENAS ÁREAS E SENSORIAMENTO REMOTO: CONCEITOS E APLICAÇÕES	71
3.1	INTRODUÇÃO AOS CONCEITOS BÁSICOS DE SENSORIAMENTO REMOTO.....	71
3.1.1	<i>O Espectro Eletromagnético e as Resoluções em Sensoriamento Remoto</i>	72
3.1.2	<i>Cor, Comportamento Espectral de Alvos Urbanos e Classificação</i>	79
3.2	SATÉLITE LANDSAT 7 ETM+.....	84
3.3	DISTRIBUIÇÃO ESPACIAL, ESTIMATIVAS DE POPULAÇÃO E SENSORIAMENTO REMOTO	87
3.3.1	<i>Breve Introdução aos Métodos de Interpolação de Área que não utilizam de Informações Auxiliares</i>	88
3.3.2	<i>Métodos Dasimétricos, Estimativas Populacionais e Sensoriamento Remoto</i>	90
3.3.2.1	<i>Correlação entre População e Áreas Urbanas</i>	93
3.3.2.2	<i>Correlação entre População e Uso do Solo</i>	97

3.3.2.3	Contagem das Unidades Habitacionais.....	100
3.3.2.4	Correlação entre População e os Pixels das Imagens de Satélite.....	102
3.3.2.5	Métodos multivariados.....	110
3.3.2.6	A Grade Estatística do IBGE	112
4	PROPOSTA METODOLÓGICA PARA A CONSTRUÇÃO E AVALIAÇÃO DE ESTIMATIVAS POPULACIONAIS PÓS-CENSITÁRIAS A PARTIR DE IMAGENS ORBITAIS DE MÉDIA RESOLUÇÃO ESPACIAL.....	115
4.1	CRITÉRIOS PARA A SELEÇÃO DOS MUNICÍPIOS	117
4.2	BASES DE DADOS, MAPEAMENTO DASIMÉTRICO, ESTRUTURAÇÃO DOS RESULTADOS E OS MODELOS ADOTADOS PARA O CÁLCULO DAS ESTIMATIVAS.....	122
4.2.1	<i>Mapeamento Dasimétrico e Construção das Bases de Dados.....</i>	<i>122</i>
4.2.2	<i>Estruturação dos Resultados</i>	<i>125</i>
4.2.3	<i>Modelos Adotados para o Cálculo das Estimativas via Sensoriamento Remoto</i>	<i>130</i>
5	ESTIMATIVAS POPULACIONAIS A PARTIR DE IMAGENS ORBITAIS DE MÉDIA RESOLUÇÃO ESPACIAL: RESULTADOS PARA OS MUNICÍPIOS SELECIONADOS	136
5.1	CONTAGEM	136
5.1.1	<i>Mapeamento Dasimétrico.....</i>	<i>137</i>
5.1.2	<i>Modelos e Estimativas no nível dos Setores Censitários.....</i>	<i>146</i>
5.1.2.1	Validação Interna.....	148
5.1.2.2	Validação Externa e Análise Comparativa.....	152
5.1.3	<i>Modelos e Estimativas no nível dos Pixels</i>	<i>156</i>
5.1.3.1	Validação Interna.....	158
5.1.3.2	Validação Externa e Análise Comparativa.....	162
5.2	VESPASIANO	165
5.2.1	<i>Mapeamento Dasimétrico.....</i>	<i>166</i>
5.2.2	<i>Modelos e Estimativas no nível dos Setores Censitários.....</i>	<i>171</i>
5.2.2.1	Validação Interna.....	172
5.2.2.2	Validação Externa e Análise Comparativa.....	175
5.2.3	<i>Modelos e Estimativas no nível dos Pixels</i>	<i>178</i>
5.2.3.1	Validação Interna.....	179
5.2.3.2	Validação Externa e Análise Comparativa.....	183
5.3	SARZEDO	185
5.3.1	<i>Mapeamento Dasimétrico.....</i>	<i>186</i>
5.3.2	<i>Modelos e Estimativas no nível dos Pixels</i>	<i>190</i>
5.3.2.1	Validação Interna.....	191
5.3.2.2	Validação Externa e Análise Comparativa.....	195
5.4	BALDIM	198
5.4.1	<i>Mapeamento Dasimétrico.....</i>	<i>199</i>
5.4.2	<i>Modelos e Estimativas no nível dos Pixels</i>	<i>204</i>
5.4.2.1	Validação Interna.....	205
5.4.2.2	Validação Externa e Análise Comparativa.....	208

6 CONCLUSÃO.....	216
REFERÊNCIAS BIBLIOGRÁFICAS.....	231

LISTA DE ILUSTRAÇÕES

FIGURAS

FIGURA 1 - Espectro Eletromagnético: distribuição da intensidade da energia	73
FIGURA 2 – Espectros de reflectância e as posições das bandas do sensor LANDSAT 7.....	75
FIGURA 3 – Espalhamento atmosférico e esquema de captação da radiância por sensores de satélite.....	76
FIGURA 4 – Representação Conceitual do Modelo (Harvey, 2002b)	107
FIGURA 5 – Mapa de Localização dos Municípios Selecionados.....	118
FIGURA 6 - Diagrama de Fluxo de Dados.....	127
FIGURA 7 – Composições R3G4B5 de Imagens Landsat 7 ETM+ para os anos de 2000 (a), 2010 (b) e 2015 (c) – Município de Contagem	139
FIGURA 8 – Evolução das Manchas de Ocupação (2000, 2010 e 2015) a partir da classificação de Imagens Landsat 7 ETM+ – Município de Contagem	140
FIGURA 9 – Distribuição da População por Setores Censitários de 2000 (a), Setores Censitários de 2010 (b) e Grades Estatísticas de 2010 (c), a partir de Dados do Censo 2010 – Município de Contagem	143
FIGURA 10 – Mapeamento Dasimétrico por meio dos Pixels de Imagens Landsat 7 ETM+ e de Dados dos Setores Censitários do Censo 2000 (a), Setores Censitários do Censo 2010 (b), e da Grade Estatística de 2010 (c) – Município de Contagem.....	146
FIGURA 11 – Composições R3G4B5 de Imagens Landsat 7 ETM+ para os anos de 2000 (a), 2010 (b) e 2015 (c) – Município de Vespasiano.....	167
FIGURA 12 – Evolução das Manchas de Ocupação (2000, 2010 e 2015) a partir da classificação de Imagens Landsat 7 ETM+ – Município de Vespasiano.....	168
<i>FIGURA 13 – Distribuição da População por Setores Censitários de 2000 (a), Setores Censitários de 2010 (b) e Grades Estatísticas de 2010 (c), a partir de Dados do Censo 2010 – Município de Vespasiano.....</i>	<i>170</i>
FIGURA 14 – Mapeamento Dasimétrico por meio dos Pixels de Imagens Landsat 7 ETM+ e de Dados dos Setores Censitários do Censo 2000 (a), Setores Censitários do Censo 2010 (b), e da Grade Estatística de 2010 (c) – Município de Vespasiano	171
FIGURA 15 – Composições R3G4B5 de Imagens Landsat 7 ETM+ para os anos de 2000 (a), 2010 (b) e 2015 (c) – Município de Sarzedo	186

FIGURA 16 – Evolução das Manchas de Ocupação (2000, 2010 e 2015) a partir da classificação de Imagens Landsat 7 ETM+ – Município de Sarzedo	187
FIGURA 17 – Distribuição da População por Setores Censitários de 2000 (a), Setores Censitários de 2010 (b) e Grades Estatísticas de 2010 (c), a partir de Dados do Censo 2010 – Município de Sarzedo	189
FIGURA 18 – Mapeamento Dasimétrico por meio dos Pixels de Imagens Landsat 7 ETM+ e de Dados dos Setores Censitários do Censo 2000 (a), Setores Censitários do Censo 2010 (b), e da Grade Estatística de 2010 (c) – Município de Sarzedo.....	190
FIGURA 19 – Composições R3G4B5 de Imagens Landsat 7 ETM+ para os anos de 2000 (a), 2010 (b) e 2015 (c) – Município de Baldim.....	200
FIGURA 20 – Evolução das Manchas de Ocupação (2000, 2010 e 2015) a partir da classificação de Imagens Landsat 7 ETM+ – Município de Baldim.....	201
FIGURA 21 – Distribuição da População por Setores Censitários de 2000 (a), Setores Censitários de 2010 (b) e Grades Estatísticas de 2010 (c), a partir de Dados do Censo 2010 – Município de Baldim	202
FIGURA 22 – Mapeamento Dasimétrico por meio dos Pixels de Imagens Landsat 7 ETM+ e de Dados dos Setores Censitários do Censo 2000 (a), Setores Censitários do Censo 2010 (b), e da Grade Estatística de 2010 (c) – Município de Baldim	203

QUADROS

QUADRO 1 – Principais Características do Sensor ETM+.....	86
QUADRO 2 – Descrição das Variáveis utilizadas para a construção de modelos baseados em Harvey (2002a)	131
QUADRO 3 – Descrição das Variáveis utilizadas para a construção de modelos baseados em Harvey (2002b)	132

GRÁFICOS

GRÁFICO 1 – Erro Total dos Modelos baseados em Setores Censitários (2000 e 2010) - Contagem	150
GRÁFICO 2 – Erro Total dos Modelos baseados em Pixels, a partir de Setores Censitários (de 2000 e 2010) e Grade Estatística (de 2010) - Contagem	160
GRÁFICO 3 – Erro Total dos Modelos baseados em Setores Censitários (2000 e 2010) - Município de Vespasiano	174

GRÁFICO 4 – Erro Total dos Modelos baseados em Pixels, a partir de Setores Censitários (de 2000 e 2010) e Grade Estatística (de 2010) - Vespasiano	181
GRÁFICO 5 – Erro Total dos Modelos baseados em Pixels, a partir de Setores Censitários (de 2000 e 2010) e Grade Estatística (de 2010) - Sarzedo.....	194
GRÁFICO 6 – Erro Total dos Modelos baseados em Pixels, a partir de Setores Censitários (de 2000 e 2010) e Grade Estatística (de 2010) - Baldim	207
GRÁFICO 7 – Diferenças Percentuais das Estimativas via Sensoriamento Remoto para o ano de 2010 (baseados em modelos criados em 2000) e das Estimativas Oficiais de 2010, em relação à População do Censo 2010 - Baldim, Contagem, Vespasiano e Sarzedo	210
GRÁFICO 8 – Diferenças Percentuais das Estimativas via Sensoriamento Remoto (baseadas em modelos criados em 2000 e 2010) e de estimativas demográficas, para o ano de 2015, em relação às Estimativas Pós-Censitárias do IBGE de 2015	215

TABELAS

TABELA 1 – Matriz de Correlação entre Densidade e as Bandas 1 a 5 e 7: Modelo Baseado em Setores Censitários de 2000 –Contagem.....	147
TABELA 2 – Validação Interna dos Modelos no nível dos Setores Censitários (2000 e 2010) - Contagem	149
TABELA 3 – Validação Externa: Diferenças Percentuais da Estimativa via Sensoriamento Remoto para o ano de 2010 (baseado em modelo criado em 2000, no nível dos setores) e das Estimativas Oficiais de 2010, em relação à População do Censo 2010 - Contagem.....	152
TABELA 4 – Análise Comparativa: Diferenças Percentuais das Estimativas via Sensoriamento Remoto (baseados em modelos criados em 2000 e 2010, no nível dos setores) e de estimativas demográficas, para o ano de 2015, em relação às Estimativas Pós-Censitárias do IBG	154
TABELA 5 – Matriz de Correlação entre População e as Bandas 1 a 5 e 7: Modelo Baseado em Pixels a partir de Setores Censitários de 2000 (a) e a partir da Grade Estatística de 2010 (b) –Contagem	157
TABELA 6 – Validação Interna dos Modelos baseados em Pixels (2000 e 2010) - Contagem	159
TABELA 7 – Validação Externa: Diferenças Percentuais da Estimativa via Sensoriamento Remoto para o ano de 2010 (baseado em modelo criado em 2000, no nível dos pixels) e das Estimativas Oficiais de 2010, em relação à População do Censo 2010 - Contagem.....	163
TABELA 8 – Análise Comparativa: Diferenças Percentuais das Estimativas via Sensoriamento Remoto (baseados em modelos criados em 2000 e 2010, no nível dos pixels) e de estimativas demográficas, para o ano de 2015, em relação às Estimativas Pós-Censitárias do IBGE	164

TABELA 9 – Matriz de Correlação entre Densidade e as Bandas 1 a 5 e 7: Modelo Baseado em Setores Censitários de 2000 –Vespasiano	172
TABELA 10 – Validação Interna dos Modelos baseados em Setores Censitários (2000 e 2010) - Vespasiano	173
TABELA 11 – Validação Externa: Diferenças Percentuais da Estimativa via Sensoriamento Remoto para o ano de 2010 (baseado em modelo criado em 2000, no nível dos setores) e das Estimativas Oficiais de 2010, em relação à População do Censo 2010-Vespasiano e Contagem	176
TABELA 12 – Análise Comparativa: Dif. Perc. das Estimativas via Sens. Remoto (baseados em modelos criados em 2000 e 2010, no nível dos setores) e de estimativas demog., para o ano de 2015, em relação às Estimativas Pós-Cens. do IBGE de 2015-Vespasiano e Contagem	176
TABELA 13 – Matriz de Correlação entre População e as Bandas 1 a 5 e 7: Modelo Baseado em Pixels a partir de Setores Censitários de 2000 (a) e a partir da Grade Estatística de 2010 (b) –Vespasiano	178
TABELA 14 – Validação Interna dos Modelos baseados em Pixels (2000 e 2010) - Vespasiano	180
TABELA 15 – Validação Externa: Diferenças Percentuais da Estimativa via Sensoriamento Remoto para o ano de 2010 (baseado em modelo criado em 2000, no nível dos pixels) e das Estimativas Oficiais de 2010, em relação à População do Censo 2010-Vespasiano e Contagem	183
TABELA 16 – Análise Comparativa: Dif. Perc. das Estimativas via Sens. Remoto (baseados em modelos criados em 2000 e 2010, no nível dos pixels) e de estimativas demog., para o ano de 2015, em relação às Estimativas Pós-Cens. do IBGE de 2015-Vespasiano e Contagem	184
TABELA 17 – Matriz de Correlação entre População e as Bandas 1 a 5 e 7: Modelo Baseado em Pixels, a partir de Setores Censitários de 2000 (a) e Modelo Baseado em Pixels, a partir da Grade Estatística de 2010 – Sarzedo.....	191
TABELA 18 – Validação Interna dos Modelos baseados em Pixels (2000 e 2010) - Sarzedo	192
TABELA 19 – Validação Externa: Dif. Perc. da Estimativa via Sens. Remoto para o ano de 2010 (baseado em modelo criado em 2000, no nível dos pixels) e das Estimativas Oficiais de 2010, em relação à População do Censo 2010-Sarzedo, Contagem e Vespasiano	195
TABELA 20 – Análise Comparativa: Dif. Perc. das Estimativas via Sens. Remoto (baseados em modelos criados em 2000 e 2010) e de estimativas demog., para o ano de 2015, em relação às Estimativas Pós-Cens. do IBGE de 2015-Sarzedo, Contagem e Vespasiano	196
TABELA 21 – Matriz de Correlação entre População e as Bandas 1 a 5 e 7: Modelo Baseado em Pixels, a partir de Setores Censitários de 2000 (a) e Modelo Baseado em Pixels, a partir da Grade Estatística de 2010 – Baldim	204
TABELA 22 –Validação Interna dos Modelos baseados em Pixels (2000 e 2010) - Baldim.....	206

TABELA 23 – Validação Externa: Dif. Perc. das Estimativas via Sens. Remoto para o ano de 2010 (baseados em modelos criados em 2000) e das Estimativas Oficiais de 2010, em relação à População do Censo 2010-Baldim, Contagem, Vespasiano e Sarzedo.....	208
TABELA 24 – Análise Comparativa: Dif. Perc. das Estimativas via Sens. Remoto (baseadas em modelos criados em 2000 e 2010) e de estim. demog., para o ano de 2015, em relação às Estimativas Pós-Cens. do IBGE de 2015-Baldim, Contagem, Vespasiano e Sarzedo	212

RESUMO

Nas últimas décadas, tem se observado o aumento pela busca por informações mais detalhadas sobre a dinâmica populacional, diante à crescente demanda para o planejamento nas mais diferentes áreas do setor público, como educação, saúde, meio ambiente, economia, para o delineamento de programas sociais, bem como no setor privado, para a elaboração de pesquisas de mercado. Mais recentemente, a disseminação dos Sistemas de Informações Geográficas, somado à crescente disponibilização de imagens de sensoriamento remoto, tem contribuído para o desenvolvimento de metodologias para o campo das estimativas populacionais para pequenas áreas, com destaque para um conjunto de métodos que utilizam de imagens orbitais. Nesse contexto, o presente trabalho tem como objetivo avaliar a capacidade e a contribuição das imagens orbitais de média resolução espacial Landsat 7 ETM+ para a produção de estimativas populacionais pós-censitárias no nível municipal, a partir da construção de modelos baseados em setores censitários de 2000 e 2010, e da grade estatística do IBGE, de 2010. Para tanto, foram calculadas estimativas para quatro municípios da Região Metropolitana de Belo Horizonte: Contagem, Vespasiano, Sarzedo e Baldim. Primeiramente, foram construídos modelos que utilizam como variáveis explicativas a média das reflectâncias das bandas espectrais das imagens Landsat 7 ETM+, observado no nível dos setores censitários e das grades estatísticas, e a densidade populacional dos setores censitários como variável dependente. Num segundo momento, foram construídos modelos que utilizam as reflectâncias e a população distribuída no nível dos pixels das imagens, a partir de procedimento de regressão iterada. Todos os modelos foram testados através de procedimentos de validação interna (para o ano de construção dos modelos), de validação externa (para o ano de 2010), além de análises comparativas com estimativas pós-censitárias (para o ano de 2015). Para todos os municípios, a validação interna apresentou ótimos resultados (sempre abaixo de 6%, para todos os modelos), enquanto que, na validação externa, o método no nível dos pixels apresentou resultados consistentes para municípios com alta densidade populacional (0,7%), ao passo que o método no nível dos setores apresentou resultados satisfatórios para municípios com baixa densidade de população (1,9%). Ademais, foi possível identificar o potencial das imagens Landsat 7 ETM+ na produção de estimativas que indicam o decréscimo populacional. O objetivo dessas comparações foi analisar a eficácia das imagens orbitais de média resolução Landsat ETM+ na produção de estimativas pós-censitárias no nível municipal, bem como, de forma mais ampla, analisar a contribuição desse tipo de aplicação para o campo das estimativas para pequenas áreas e para a Demografia.

Palavras-chave: Estimativas Populacionais Pós-Censitárias, Estimativas Populacionais para Pequenas Áreas, Sensoriamento remoto, Mapeamento Dasimétrico.

ABSTRACT

In the last decades, there has been an increase in the search for more detailed information on population dynamics, given the growing demand for planning in the most different areas of the public sector, such as education, health, environment, economy, for the design of social programs, and, as well as, in the private sector, for the elaboration of market research. More recently, the dissemination of Geographic Information Systems, together with the increasing availability of remote sensing images, has contributed to the development of methodologies for the field of population estimates for small areas, with emphasis on a set of methods that use orbital images. In this context, the present work aims to evaluate the capacity and contribution of medium spatial resolution Landsat 7 ETM + orbital images, for the production of postcensal population estimates at the municipal level, based on the construction of models based on 2000 and 2010 census tracts, and the statistical grid of IBGE, 2010. Estimates were calculated for four municipalities in the Metropolitan Region of Belo Horizonte: Contagem, Vespasiano, Sarzedo and Baldim. Firstly, we constructed models that use as explanatory variables the average of the reflectances of the spectral bands of the Landsat 7 ETM + images of the census tracts and statistical grids and the population density of the census tracts as a dependent variable. In a second moment, we constructed models that use the reflectances and the distributed population at the level of the pixels of the images, from an iterated regression procedure. All models were tested through internal validation procedures (for the year of construction of the models), external validation (for the year 2010), and comparative analyzes of post-census estimates (for the year 2015). For all municipalities, the internal validation presented excellent results (always below 6%, for all models), while in the external validation the method at the pixel level presented consistent results for municipalities with high population density (0.7%), while the method at the sector level presented satisfactory results for municipalities with low population density (1,9%). In addition, it was possible to identify the potential of Landsat 7 ETM + images in the production of estimates that indicate a population decrease. The objective of these comparisons was to analyze the efficacy of Landsat ETM + medium resolution orbital images in the production of post-census estimates at the municipal level, as well as, more broadly, to analyze the contribution of this type of application to the field of estimation for small areas and Demography.

Keywords: Postcensal Population Estimates, Population Estimates for Small Areas, Remote Sensing, Dasymetric Mapping,

1 INTRODUÇÃO E JUSTIFICATIVA

Informações sobre o tamanho e a composição da população, e o seu ritmo de mudança têm sido objeto de interesse crescente do setor público, seja para o planejamento econômico, seja para a elaboração de políticas nas áreas da saúde, educação, ou mesmo para a análise do potencial social das populações. Até o presente momento, a fonte mais completa e confiável para a enumeração da população dos países e suas subdivisões geográficas são os censos, baseados nas entrevistas domiciliares.

Se, por um lado, a enumeração periódica dos censos em muitos países permite a captação do tamanho populacional e seu crescimento, bem como o registro dos nascimentos, das mortes e da migração, o intervalo temporal entre cada pesquisa censitária (geralmente 10 anos) normalmente não é adequado ao ritmo da mudança populacional. Ademais, as informações censitárias são limitadas, devido ao elevado custo das operações e diante da crescente exigência por informações mais detalhadas, com o aumento da complexidade das demandas sociais e econômicas dos governos (United Nations, 1952; Shryock e Siegel, 1973; Bryan, 2004; Smith e Morrison, 2005; Cabrera, 2011)¹.

Fontes de informações alternativas são utilizadas, tais como pesquisas por amostragem (embora igualmente dispendiosas), registros civis (especialmente para a enumeração de estatísticas vitais), censos escolares, e diversos cadastros governamentais, embora essas informações não sejam disponíveis para a maioria dos países (ou para suas subdivisões administrativas), além do problema dos sub-registros e da falta de qualidade e atualidade das informações (United Nations, 1952; Shryock e Siegel, 1973; Bryan, 2004; Smith e Morrison, 2005). Para contornar

¹ As ideias apresentadas nessa seção foram baseadas nesses autores.

essas limitações, métodos matemáticos que utilizam variáveis sintomáticas² são alternativas econômicas e acessíveis para a produção de estimativas populacionais, que devem ser compreendidas como aproximações da mudança populacional, fundamentais para o planejamento dos setores público e privado (Shryock e Siegel, 1973; Bryan, 2004; Smith e Morrison, 2005; Jannuzzi, 2007; Cabrera, 2011; Gonzáles e Torres, 2012; Silva e Santos, 2016).

As estimativas podem ser subdivididas em três tipos, que variam, principalmente, segundo o tempo de referência, além das bases de dados e de propostas metodológicas distintas, o que resulta, por consequência, em diferentes níveis de confiabilidade. O primeiro tipo são as estimativas intercensitárias, que correspondem a datas entre dois censos realizados, produzidas normalmente a partir de métodos de interpolação. O segundo são as estimativas pós-censitárias, referentes a datas correntes ou passadas após o último censo, a partir da utilização de técnicas de extrapolação, ou de métodos que incorporam indicadores sintomáticos. O terceiro tipo são as projeções, que compreendem uma variedade de métodos (de extrapolação e de corte-componentes), cujos resultados representam um conjunto particular de pressupostos sobre o futuro da população, definidos a partir da observação do comportamento das componentes demográficas (Shryock e Siegel, 1973; Smith *et al.*, 2002; Bryan, 2004; Smith e Morrison, 2005; Alho e Spencer, 2005; Cabrera, 2011).

Diante da grande variedade de fontes existentes, os dados a partir dos quais as populações são estimadas podem ser divididos em duas categorias: dados diretos, relativos aos censos demográficos e registros populacionais; e dados indiretos, utilizados para a produção de estimativas de certos parâmetros com base na informação indiretamente relacionada, ou sintomática, de seu valor real. Exemplos de dados indiretos são estatísticas de emprego, matrícula escolar, estatísticas de companhias de gás ou de energia elétrica ou mesmo de construção, registros de

² Variável sintomática é uma variável auxiliar fortemente correlacionada com o crescimento populacional de uma dada localidade, e que muda no tempo em concordância com as mudanças no volume da população. É normalmente utilizada para a produção de estimativas em áreas menores (Esquivel, 2001; Cabrera, 2011).

votos e métodos que utilizam informações de nascimentos e óbitos (Shryock e Siegel, 1973; Bryan, 2004; Cabrera, 2011).

Segundo Cabrera (2011), cada modelo que parte de dados indiretos utiliza dados e pressupostos diferentes, porém tendo em comum a suposição de que o volume da população da área menor guarda uma relação quantificável com as variáveis sintomáticas. São, portanto, estimativas e não projeções populacionais, pois não realizam suposições sobre o comportamento futuro de cada componente da dinâmica demográfica.

Órgãos governamentais responsáveis pela produção de estimativas determinam o tipo de informação que será produzida de acordo com o tipo de demanda, com a qualidade dos dados disponíveis, a quantidade de trabalho necessária para a produção das estimativas e os recursos disponíveis. Estimativas nacionais e para as primeiras subdivisões administrativas sobre a população total, por idade e sexo, são as mais importantes e frequentes, além de estimativas por subgrupos (ex.: étnicos, status marital, por níveis educacionais e de renda, etc.) e para áreas administrativas menores (Shryock e Siegel, 1973; Bryan, 2004; Smith e Morrison, 2005; Alho e Spencer, 2005; Cabrera, 2011).

Nos últimos anos, tem se observado o aumento na demanda por informações para pequenas áreas³, por parte dos setores público e privado. Unidades político-administrativas menores (como os municípios) são unidades de análise importantes para a investigação das condições e tendências econômicas, demográficas, ambientais, epidemiológicas, entre outras. São, portanto, fundamentais para o desenvolvimento de programas, para a alocação de fundos governamentais e para o delineamento e monitoramento de políticas públicas de um modo geral. No setor privado, estimativas para pequenas áreas são importantes na tomada de decisão

³ Segundo Smith e Morrison (2005), não há uma definição universalmente reconhecida para “pequenas áreas”, podendo-se referir a qualquer unidade administrativa abaixo do nível nacional e dos estados; e está normalmente associado à subpopulações para os quais não podem ser produzidas estimativas diretas com a precisão adequada. Segundo Gonzáles e Torres (2012), a expressão “pequenas áreas” deve ser avaliada para cada país, de acordo com a magnitude de sua população e as características de sua divisão territorial interna.

de negócios, escolha de mercados potenciais, previsão de vendas, entre outros (Smith e Morrison, 2005; Jannuzzi, 2007).

No Brasil, o Instituto Brasileiro de Geografia e Estatística (IBGE) produz, anualmente, estimativas no nível dos municípios, sendo utilizadas como informação base para o cálculo e repasse do Fundo de Participação dos Municípios. Porém, devido à falta de variáveis sintomáticas confiáveis e com série histórica, disponíveis para anos sucessivos, as estimativas do IBGE são calculadas a partir de projeções por Unidade da Federação (UF), que são, por sua vez, desagregadas para o nível municipal, o que limita as estimativas aos critérios e pressupostos adotados na construção das projeções por UF, bem como às limitações metodológicas das técnicas de desagregação que são utilizadas.

A disponibilidade de dados derivados dos programas espaciais, em particular aqueles de imageamento por satélite, chamaram a atenção para a possibilidade de seu uso no auxílio às estimativas populacionais há algum tempo, em geral, fora do campo da Demografia (Iisaka e Hegedus, 1982; Sutton, 1997). A disseminação e intensificação no uso dos Sistemas de Informações Geográficas (SIG's), o desenvolvimento de técnicas de análise espacial e o crescimento da disponibilidade de imagens de sensoriamento remoto, têm contribuído para o desenvolvimento de um conjunto de novas técnicas e metodologias com aplicações muito promissoras no campo das estimativas populacionais para pequenas áreas, a partir da utilização de imagens orbitais (Holt *et al.*, 2004; Li e Weng, 2005; Wu *et al.*, 2005; Lu *et al.*, 2006; Bhaduri *et al.*, 2007; Wu e Murray, 2007; Zandbergen e Ignizio, 2010).

Para o cálculo das estimativas via sensoriamento remoto, são construídos modelos de superfície ou dasimétricos, que correspondem à redistribuição geográfica de dados populacionais com um maior nível de detalhamento espacial, a partir da utilização de dados auxiliares, através do registro em células de resolução espacial fixa, o que permite a análise da distribuição espacial da população no espaço contínuo, de forma independente dos limites administrativos (Eicher e Brewer, 2001; Mennis, 2003; Wu *et al.*, 2005; Mennis e Hultgren, 2006; Langford, 2013).

De maneira bem geral, para a construção das estimativas via sensoriamento remoto, parte-se, primeiramente, de dados dos anos censitários, da distribuição

espacial mais desagregada da população obtida através do censo (por exemplo, os setores censitários, por sua vez restritos aos limites administrativos), que, por sua vez, é transformada em uma distribuição espacial da população mais refinada, a partir de dados auxiliares (como as imagens de satélite) e de métodos dasimétricos (Petrov, 2012). Maior resolução pode ser alcançada, pois é possível não apenas interpolar a distribuição da população para o espaço contínuo, a partir da observação das áreas ocupadas, como manchas urbanas em imagens de satélite (não permanecendo restrito, portanto, a limites político-administrativos), mas, também, por ser possível discriminar estruturas do espaço físico não ocupadas (como vegetação, rios, entre outros), muito embora seja importante observar as incertezas derivadas desse processo (Mennis, 2003). A partir do mapeamento dasimétrico realizado em um ano censitário, é construído um modelo estatístico, que é aplicado a outro mapa dasimétrico, correspondente a um ano pós-censitário (ou seja, o modelo é aplicado a dados de distribuição da população em um ano pós-censitário), obtendo-se, assim, a estimativa para o referido ano.

Este tipo de variável sintomática é relevante não apenas pela ausência de dados oficiais nos períodos intercensitários e pós-censitários, mas, também, pela ausência de dados indiretos confiáveis e com séries históricas disponíveis. Sabe-se que estimativas para pequenas áreas são susceptíveis a mudanças e grandes oscilações no ritmo de crescimento das populações em pequenos intervalos de tempo. Diante disso, a possibilidade de acompanhamento dessas mudanças, através de imagens em anos sucessivos, a flexibilidade em relação à restrição dos limites administrativos – em função das estimativas poderem ser produzidas para o espaço contínuo (com grande potencial para estudos temporais, pela possibilidade de padronização das unidades territoriais) – somado ao aumento da disponibilização dessas informações, definem o potencial das estimativas populacionais via sensoriamento remoto como fronteira na produção de estimativas para pequenas áreas (Mennis, 2003; Wu *et al.*, 2005; Langford, 2013; Zandbergen e Ignizio, 2010).

Frente à relevância desse campo do conhecimento para os setores público e privado, e na direção dos avanços recentes, com o desenvolvimento do sensoriamento remoto no campo das estimativas populacionais, o presente

trabalho tem como objetivo contribuir na avaliação da capacidade das imagens orbitais de média resolução espacial Landsat ETM+ na produção de estimativas populacionais pós-censitárias no nível municipal. Para tanto, estimativas via sensoriamento remoto foram calculadas para quatro municípios localizados na Região Metropolitana de Belo Horizonte - RMBH, a saber: Contagem, Vespasiano, Sarzedo e Baldim.

A escolha teve como critério a seleção de municípios com diferentes características em relação ao tamanho e ao crescimento populacional. Modelos de regressão construídos a partir de informações oriundas de imagens de satélite e de dados dos Censos de 2000 e 2010 foram utilizados para o cálculo de estimativas populacionais para os anos 2010 e 2015. As estimativas, por sua vez, foram comparadas com os dados do Censo 2010 e com estimativas pós-censitárias do Instituto Brasileiro de Geografia e Estatística – IBGE.

A partir da literatura clássica e das aplicações revisadas, tem-se como hipótese que as estimativas municipais, produzidas a partir de imagens Landsat ETM+, sejam eficazes (ou seja, tendem a apresentar baixos índices de erro) no caso dos municípios com elevada densidade populacional e ritmo de crescimento positivo. Por outro lado, espera-se constatar a redução da eficácia no caso de municípios com baixo contingente populacional, e, principalmente, com ritmo de crescimento negativo.

De forma geral, são diversos os elementos para além das resoluções espacial, espectral e/ou radiométrica das imagens que podem interferir na construção de estimativas confiáveis, como o nível de cobertura de nuvens, os procedimentos adotados nas etapas de processamento e classificação das imagens, além dos pressupostos e das limitações do método dasimétrico (de distribuição da população) e do modelo de estimação utilizados para o cálculo das estimativas. Contudo, a resolução espacial é elemento determinante na produção das estimativas, e a escolha da imagem (com uma dada resolução espacial) determina, em grande medida, os métodos de processamento e classificação, o modelo estatístico que será adotado e, ao final, a qualidade e a precisão das estimativas produzidas (Jensen e Cowen, 1999; Wu *et al.*, 2005; Jensen, 2009).

Nos últimos anos, avanços nos sensores remotos e na disseminação de imagens têm ampliado os estudos e as aplicações de sensoriamento remoto para o cálculo de estimativas, a partir do desenvolvimento de novos métodos e, principalmente, a partir da utilização de imagens de sensores remotos de maior resolução espacial (Liu e Clarke, 2002; Souza, 2004; Tomás, 2010; Zandbergen, 2010; Langford, 2013; Xie *et al.*, 2015). Porém, atualmente, grande parte das imagens de alta resolução espacial (com resolução do pixel a partir de 5 metros, aproximadamente) não são disponibilizadas gratuitamente, e, em relação àquelas cujo acesso é livre não se dispõe de uma série histórica que compreenda o período de dois censos, elemento fundamental para a validação do modelo proposto (pois, usualmente, são definidos os parâmetros num primeiro censo, enquanto no censo seguinte é atestada a qualidade do modelo).

Outra limitação das imagens de alta resolução é a aplicação para áreas mais extensas, pois, a menor área do campo de visada de cada cena torna praticamente inviável, na atualidade, a construção de grandes mosaicos, em relação à quantidade e ao tempo de processamento das imagens (exceção das grandes empresas do setor privado, como a Google, Amazon, etc). Como um dos objetivos do setor público é a produção sistemática de estimativas para grandes extensões de área (estados ou mesmo para um país), o custo e o tempo de processamento das imagens de alta resolução na atualidade seriam limitadores desse tipo de imagem na produção de estimativas oficiais, caso não exista uma infraestrutura computacional pública e compartilhada para o tratamento desta informação, cuja manutenção envolve também altos custos.

É verdade que essa afirmativa é datada, pois os avanços tecnológicos tendem, num ritmo acelerado, a superar as restrições de processamento e a redução dos custos. De toda forma, a demanda pela produção de estimativas para pequenas áreas, replicáveis a grandes extensões de áreas⁴, justificam a escolha por imagens de acesso gratuito, com série histórica disponível e consolidada em relação à qualidade das estimativas calculadas.

⁴ Sobre a discussão da replicabilidade e do custo benefício dos métodos de estimativas via sensoriamento remoto, ver Langford (2007; 2013) e Jensen e Cowen (1999).

As imagens Landsat (NASA, 2017) “*represents the world's longest continuously acquired collection of space-based moderate-resolution land remote sensing data*” (USGS, 2017). Com resolução espacial de 30 metros e disponibilizadas gratuitamente⁵, apresenta um campo de visada maior, o que reduz o tempo de processamento e aumenta a capacidade de sua utilização para áreas mais extensas, além de dispor de uma longa série histórica (cobrem, por exemplo, o período correspondente aos últimos dois censos brasileiros), o que possibilita a construção e a validação de modelos para o cálculo de estimativas.

No caso das imagens Landsat, apesar da existência de uma ampla literatura internacional sobre sua utilização para o cálculo de estimativas para pequenas áreas (Yuan *et al.*, 1997; Lo e Welch, 1977; Lisaka e Hegedus, 1982; Holt *et al.*, 2004; Li e Weng, 2005; Lu *et al.*, 2006; Wu e Murray, 2007), no Brasil ainda são poucos os estudos que as utilizam para esse tipo de aplicação (exceções são os trabalhos de Forest, 1978; Reis, 2005; Durand *et al.*, 2007). Nesse sentido, o presente trabalho busca contribuir para a produção de estimativas populacionais pós-censitárias para pequenas áreas, a partir de imagens de média resolução espacial.

No presente trabalho, modelos foram construídos separadamente para cada um dos quatro municípios, a partir dos dados do Censo de 2000, e os modelos foram validados internamente (a partir de amostras das bases de dados para o mesmo ano), e externamente, através da comparação com a população enumerada pelo Censo 2010. Além do Censo 2010, as estimativas via sensoriamento remoto, construídas a partir dos modelos de 2000 e calculadas para o ano de 2010, foram comparadas com projeções do IBGE (revisões 2008 e 2013) para o ano de 2010. Na realidade, a projeção de 2008 do IBGE foi produzida para o agregado do País, sendo que para as Unidades da Federação foram calculadas a partir de métodos de desagregação; enquanto que a revisão de 2013 do IBGE projetou diretamente os Estados. Ambas as projeções foram desagregadas para o nível dos municípios,

⁵ Maiores detalhes sobre o Programa de Satélites Landsat serão comentados no Capítulo 4. Ver também os sites: www.inpe.br e www.usgs.com.

para efeito de comparação com as estimativas via sensoriamento remoto, produzidas para o ano de 2010.

A comparação das estimativas por imagens de satélite com outros modelos de projeção populacional não tem como objetivo definir qual o “melhor método”, pois não é objetivo interpretar os resultados dos métodos sob o conceito de “assertividade”. Todos os métodos possuem pressupostos e limitações (que serão discutidos), não havendo um que seja, de maneira generalizada, superior aos demais. Não obstante, a comparação entre os métodos propostos neste trabalho pode ser útil para a análise da eficácia das estimativas via sensoriamento remoto, pois, resultados muito similares (ou muito díspares) podem ser avaliados no contexto do custo da adoção de cada método, fornecendo elementos para a avaliação quanto à eficácia das imagens de satélite na produção de estimativas populacionais.

Num segundo momento, foram produzidas estimativas pós-censitárias municipais (a partir de imagens Landsat 7 ETM+) para o ano de 2015, a partir de modelos de regressão baseados em imagens de satélites e dados populacionais de 2000 e de 2010. Sem dúvida, as estimativas via sensoriamento remoto baseadas em dados de 2000 (e calculadas para o ano de 2010) foram importantes na validação externa, e, conseqüentemente, na análise da eficácia das estimativas via sensoriamento remoto (a partir da comparação com o Censo 2010). Porém, faz-se importante calcular estimativas para um ano pós-censitário, e o ano de 2015 foi definido para a análise das estimativas via sensoriamento remoto, produzidas a partir de modelos baseados em dados de 2000 e de 2010.

Nesse contexto, as estimativas para o ano de 2015 foram comparadas com as estimativas pós-censitárias do IBGE, que constituem, por sua vez, as estimativas que se dispõe para o todo o território nacional, sendo uma fonte de referência oficial no campo das estimativas pós-censitárias. Além das estimativas oficiais do IBGE, foram utilizadas, para efeito de comparação, estimativas calculadas a partir dos métodos de extrapolação simples, e de taxas vitais e de razão censitária (os dois últimos baseados em variáveis sintomáticas, como serão detalhadas no Capítulo 2).

Embora não haja um valor “correto” (ou seja, um dado observado de população) para o ano de 2015, a comparação entre as diferentes fontes de estimativas pode contribuir para a análise da eficácia das estimativas via sensoriamento remoto. As estimativas por sensoriamento remoto também serão estudadas através de dados referentes a dois anos distintos (2000 e 2010), com o objetivo de contribuir para a discussão quanto aos possíveis diferenciais que podem ser encontrados entre estimativas calculadas por dados de anos diferentes (sob a hipótese de que as estimativas produzidas a partir de dados de 2010 apresentam melhores resultados, se comparado às estimativas obtidas a partir de dados de 2000). A discussão sobre o tamanho da diferença percentual entre as estimativas – tanto entre as via sensoriamento remoto e as demográficas como entre as próprias estimativas calculadas via imagens de satélite – pode contribuir na discussão sobre o custo associado à produção através de cada modelo, especialmente quando as diferenças são pequenas entre os valores encontrados.

Portanto, a estrutura dos resultados da tese consiste, de forma geral, na construção de modelos baseados em dados de 2000, que geram, por sua vez, estimativas populacionais para 2010 e 2015. As estimativas para 2010 são comparadas com o Censo 2010 (etapa denominada como de validação externa), e com a população de 2010 das projeções de 2008 e 2013, do IBGE. Num segundo momento, modelos construídos a partir de dados de 2000 e de 2010 serão utilizados para a elaboração de estimativas para o ano de 2015. Essas estimativas via sensoriamento remoto para 2015 serão comparadas com estimativas pós-censitárias do IBGE e de outras três fontes (pelos métodos de Extrapolação Simples, Taxas Vitais e Razão Censitária). Ressalta-se, ainda, que todos os modelos passam por uma etapa de validação interna, a partir da seleção de uma amostra da base de dados do mesmo ano em que o modelo foi construído.

Dentre as estimativas pós-censitárias via sensoriamento remoto para o ano de 2015, os melhores modelos (e estimativas) construídos devem partir, por hipótese, dos dados do último censo realizado (no caso, o Censo 2010). No caso dos modelos baseados em dados de 2010, eles foram subdivididos conforme a fonte de dado populacional: um conjunto de modelos foi construído a partir dos dados de setores censitários, enquanto outro conjunto de modelos foi a partir da grade

estatística do IBGE (2015). As diferenças encontradas auxiliam na análise da eficácia das estimativas segundo a fonte de dados utilizada, num contexto em que as grades estatísticas são mais desagregadas espacialmente, se comparado aos setores censitários; tendo o potencial de fornecer, por hipótese, melhores resultados quando comparados a estimativas de modelos criados a partir dos setores.

Tem-se como objetivo, portanto, analisar as diferenças entre as estimativas, à luz dos pressupostos e fonte de dados de cada modelo, com o intuito de contribuir com a literatura existente na avaliação das vantagens e limitações (enfim, da eficácia) das imagens orbitais de média resolução espacial na produção de estimativas populacionais pós-censitárias, em nível municipal, e de forma mais ampla, para o campo da Demografia.

É importante ressaltar que a comparação entre as estimativas (em 2010 e 2015) não tem como objetivo avaliar qual método seria necessariamente o mais indicado. Em primeiro lugar, porque cada método tem suas próprias vantagens e limitações, e a escolha do método deve ser sempre pautada pela adequação entre as características de um dado método aos objetivos que se propõe (Smith *et al.*, 2002; George *et al.*, 2004 Bryan, 2004). Em segundo, há uma grande quantidade de métodos e de variáveis sintomáticas existentes para a construção de estimativas, e que não serão consideradas neste trabalho.

Todavia, a comparação das estimativas pós-censitárias, via imagens orbitais de média resolução espacial, com estimativas convencionais e com informações censitárias e estimativas oficiais permitirá identificar o grau de similaridade entre os resultados, permitindo avaliar as vantagens e limitações de cada modelo, especialmente aqueles baseados em imagens orbitais de média resolução espacial.

As estimativas via sensoriamento remoto, por serem baseadas em variável sintomática, têm, por exemplo, a capacidade de identificar mudanças no ritmo de crescimento populacional, num curto intervalo de tempo (fenômeno comum em pequenas áreas). Já as estimativas pós-censitárias do IBGE, elaboradas a partir de

projeção populacional, não captam tais mudanças, o que pode ser considerado uma vantagem das estimativas que utilizam de imagens orbitais.

Por outro lado, estimativas via sensoriamento remoto de média resolução não são, por hipótese, adequadas para municípios que apresentam crescimento populacional negativo, dado que o crescimento populacional é obtido através da expansão da área ocupada, e a diminuição da população (um surto emigratório, por exemplo) não se refletiria, no curto prazo, na conversão do uso do solo de uma área construída para uma área não construída (como será discutido posteriormente). Trata-se sem dúvida de uma grande limitação, a considerar a possibilidade de replicação do modelo para um conjunto de municípios.

Aliás, na literatura sobre estimativas via sensoriamento remoto, os municípios estudados apresentam normalmente um padrão de crescimento populacional positivo e acelerado, num contexto de expansão contínua das manchas urbanas. Para a avaliação da eficácia das estimativas produzidas via imagens orbitais de média resolução espacial, buscou-se a seleção de municípios que apresentassem diferentes comportamentos em relação à taxa de crescimento, mas, principalmente, municípios que apresentassem diferentes tamanhos e níveis de densidade populacionais.

Foram definidos três municípios que apresentaram taxas de crescimento positiva e com diferentes tamanhos populacionais, o que contribui para a análise da eficácia das estimativas. Isso ocorre, pois, partindo da taxa de crescimento positiva (que representa o comportamento ideal para os modelos, que se ajustam bem aos municípios que ganham população), os erros das estimativas passam a ser resultado das diferenças entre o tamanho e o nível de densidade dos municípios.

Entretanto, ainda que não seja adequada a aplicação em municípios que apresentaram crescimento populacional negativo no último decênio, foi selecionado um município com essa característica. Não se trata apenas de hipótese, de fato, é um pressuposto desses modelos que a ausência de crescimento da mancha de ocupação não fornece qualquer tipo de informação quanto ao ritmo de decréscimo da população. Contudo, tem-se como objetivo identificar o erro nesses casos, inclusive se for considerado que os modelos não são baseados

meramente em função do tamanho da mancha de ocupação, mas das reflectâncias (que, em resumo, corresponde à fração refletida dos alvos de superfície à radiação solar, tema que será aprofundado posteriormente). Mesmo no caso das reflectâncias, por hipótese, uma edificação não deverá fornecer uma resposta espectral diferente pelo fato de ter ficado desocupada no período entre a construção do modelo e o cálculo da estimativa (em função de uma emigração, por exemplo). Ainda assim, mesmo que a ineficácia desses modelos em municípios com taxa de crescimento negativa seja um pressuposto, pode ser considerada uma contribuição avaliar qual a medida de erros nesses casos.

Ainda sobre a contribuição do trabalho, vale ressaltar que, além da seleção de municípios com diferentes tamanhos e ritmos de crescimento populacional, a metodologia proposta inova (tanto na literatura nacional e internacional) ao realizar validações externas e ao produzir estimativas via sensoriamento remoto para diferentes anos, e compará-las entre si e com um conjunto de estimativas demográficas.

Em relação à definição da área de estudo, a escolha por municípios da RMBH se justifica, primeiramente, devido ao maior conhecimento do autor deste trabalho em relação às tendências demográficas dessa área de estudo, bem como pela tradição do Cedeplar em estudos sobre essa região metropolitana. O maior conhecimento da realidade dos municípios selecionados deve ser considerado como uma vantagem, não apenas na análise qualitativa dos resultados, mas, também, na fase de construção da base de dados de sensoriamento remoto, especialmente na etapa de classificação das imagens de satélite, para a delimitação das áreas ocupadas.

Contudo, diante à diversidade de municípios existentes no País, com diferentes morfologias e formas de ocupação do território – exemplo para os municípios Amazônicos, que apresentam baixa densidade populacional, e cuja distribuição da população dificulta o mapeamento dasimétrico da população, e, conseqüentemente, a produção de estimativas – este trabalho não pretende realizar validações externas “universais” e tampouco esgotar o tema. De fato, a escolha por municípios de um único estado limita generalizações, muito embora a escassez já mencionada de trabalhos sobre o tema, especialmente no Brasil, contribui para o desenvolvimento desse campo do conhecimento.

Um aspecto importante sobre as estimativas e projeções demográficas municipais que serão consideradas para a comparação com as imagens via sensoriamento remoto, é que essas estimativas partem de projeções elaboradas para o agregado do País ou das Unidades da Federação. Em síntese, a qualidade das projeções está relacionada à qualidade dos registros civis (United Nations, 1956; Shryock e Siegel, 1973; Bryan, 2004; Smith e Morrison, 2005; O'Neil *et al.*, 2001), e a RMBH apresenta uma das melhores coberturas sobre as estatísticas vitais do País, o que resulta em estimativas e projeções mais confiáveis⁶, por hipótese. Essa vantagem, naturalmente, contribui para uma melhor avaliação das estimativas via sensoriamento remoto.

Por outro lado, estudos recentes apontam para problemas de cobertura no Censo de 2010, o que representa uma limitação para a validação de quaisquer estimativas intercensitárias ou pós-censitárias que venham a serem desenvolvidas. Além disso, em pequenas áreas, a qualidade das estimativas depende não apenas da fecundidade e da mortalidade, mas em grande medida da migração. Essa componente apresenta grandes oscilações no nível dos municípios, determinando, em muitos casos, mudanças abruptas no ritmo de crescimento populacional, em pequenos intervalos de tempo (Bryan, 2004; Smith e Morrison, 2005). Em todo caso, apesar das restrições, é válido o argumento da qualidade dos registros civis na definição da unidade de análise.

O próximo item apresenta os objetivos (geral e específicos) do trabalho e a organização dos capítulos da tese.

1.1 Objetivos

O presente trabalho tem como objetivo geral avaliar a capacidade e a contribuição das imagens orbitais de média resolução espacial Landsat 7 ETM+ para a produção de estimativas populacionais pós-censitárias no nível municipal, a partir de validações internas e externas dos modelos construídos, e da comparação das

⁶ Mesmo tendo de se considerar outros fatores igualmente importantes, como o modelo de projeção escolhido e os pressupostos definidos, sobre as tendências futuras das componentes demográficas.

estimativas produzidas via sensoriamento remoto com dados censitários e estimativas baseadas em métodos demográficos. Para tanto, serão produzidas estimativas para quatro municípios selecionados da RMBH, para os anos de 2000, 2010 e 2015. São objetivos específicos:

- Construir modelos para a produção de estimativas populacionais municipais, através de imagens orbitais de média resolução espacial Landsat 7 ETM+, a partir de dados de setores censitários de 2000 e 2010, e de dados da grade estatística do IBGE, de 2010.
- Avaliar a qualidade dos modelos construídos a partir de imagens de satélite, através de procedimentos de validação interna, a partir dos dados dos Censos de 2000 e 2010.
- Analisar a qualidade dos modelos construídos a partir de imagens de satélite, através de procedimentos de validação externa. Para tanto, serão utilizados dados do Censo 2010, levando-se em consideração a discussão sobre o problema de subcobertura censitária.
- Comparar as estimativas produzidas para o ano de 2010 (baseada em dados de 2000), com as projeções do IBGE (revisões 2008 e 2013), desagregadas ao nível municipal; com o objetivo de contribuir para a análise da eficácia das estimativas calculadas por imagens orbitais de média resolução espacial.
- Calcular estimativas pós-censitárias para o ano de 2015, através de imagens Landsat 7 ETM+ e de modelos construídos (a partir de dados de setores censitários dos Censos de 2000 e de 2010 e da grade estatística de 2010) para os municípios selecionados. Comparar os resultados com estimativas demográficas de outras fontes, a saber: estimativas pós-censitárias oficiais do IBGE; extrapolação simples; taxas vitais; e razão censitária; com o objetivo de analisar a eficácia das imagens orbitais de média resolução Landsat ETM+ na produção de estimativas populacionais pós-censitárias, em nível municipal.

- Avaliar as vantagens, limitações e a eficácia das estimativas calculadas via imagens orbitais de média resolução espacial Landsat ETM+ para o campo das estimativas populacionais em pequenas áreas, com a proposição de agendas futuras para esse campo do conhecimento.

O Capítulo 2 introduz o tema de estimativas para pequenas áreas, a partir da apresentação dos principais métodos demográficos utilizados na produção de estimativas municipais. Este tema é relevante, primeiramente, porque alguns desses métodos foram utilizados para efeito de comparação com as estimativas via sensoriamento remoto. Contudo, o mais importante é que o conhecimento sobre as vantagens e limitações dos principais métodos demográficos utilizados para a produção de estimativas para pequenas áreas contribui para a compreensão das vantagens (ou, de forma mais ampla, da eficácia) das estimativas baseadas em imagens orbitais.

Na segunda parte do Capítulo 2 são analisadas as principais características das projeções do IBGE (revisões 2008 e 2013) e das estimativas pós-censitárias municipais do IBGE, que serão, por sua vez, utilizadas na comparação com as estimativas via sensoriamento remoto. Ao final do Capítulo 2, são introduzidos os principais métodos utilizados para a avaliação dos erros das estimativas, dos quais alguns serão utilizados para a avaliação dos resultados.

O Capítulo 3 introduz o sensoriamento remoto neste trabalho, primeiramente através de uma síntese dos principais conceitos de Processamento Digital de Imagens, necessários para a compreensão das discussões que serão empreendidas na sequência (em relação aos métodos dasimétricos e às estimativas via imagens orbitais). Em seguida, o capítulo introduz os principais métodos de distribuição espacial da população, com foco nos métodos dasimétricos e sua relação com os modelos de estimativas populacionais via sensoriamento remoto. Este tema é importante, pois nessa revisão já são apresentados os modelos que foram adotados para a construção da proposta metodológica e dos resultados desse trabalho.

O Capítulo 4 descreve a proposta metodológica para a construção e avaliação das estimativas populacionais pós-censitárias municipais, calculadas a partir de

imagens orbitais Landsat 7 ETM+. O capítulo inicia com os critérios considerados para a seleção dos municípios utilizados na pesquisa. Num segundo momento, apresentam-se as bases de dados e o método dasimétrico adotado para o mapeamento da população e para o cálculo das estimativas. De posse dos conhecimentos relativos ao modelo proposto (discutido no Capítulo 3), das bases de dados e do método dasimétrico empregado, o Capítulo 4 apresenta a proposta metodológica e a estruturação dos resultados. Ao final, é apresentada uma síntese dos modelos adotados para o cálculo das estimativas via imagens orbitais.

O Capítulo 5 corresponde aos resultados do trabalho, que se baseiam fundamentalmente na comparação de estimativas elaboradas via sensoriamento remoto com: dados censitários de 2010; a população projetada pelo IBGE (revisões 2008 e 2013), desagregada para o nível dos municípios, para o ano de 2010; estimativas demográficas (baseadas em variáveis sintomáticas); e estimativas pós-censitárias oficiais. A partir dos resultados apresentados, espera-se contribuir com a literatura existente na avaliação da eficácia das imagens orbitais de média resolução espacial (Landsat 7 ETM+), na produção de estimativas pós-censitárias calculadas para municípios com diferentes características demográficas.

O Capítulo 6 traz as conclusões, onde se discutem as contribuições do presente trabalho para o campo das estimativas populacionais para pequenas áreas, além de propostas e indicações de agendas futuras para esse campo do conhecimento.

2 INTRODUÇÃO ÀS ESTIMATIVAS POPULACIONAIS PARA PEQUENAS ÁREAS E OS CENSOS E ESTIMATIVAS DO IBGE

A demanda pela investigação das tendências econômicas e demográficas da população tem crescido nos últimos anos, e, nesse contexto, projeções para pequenas áreas tem recebido especial atenção. No setor público, esse campo tem sido importante para a elaboração de políticas públicas, nas mais diferentes áreas, como, por exemplo, para o planejamento de construção de novas escolas, estradas, hospitais, para os sistemas de transporte público, para a análise da expansão do setor elétrico, bem como para a avaliação das consequências socioambientais do crescimento populacional (Smith *et al.*, 2002; Smith e Morrison, 2005). No setor privado, projeções para pequenas áreas são normalmente utilizadas para a tomada de decisão de negócios, escolha de mercados potenciais e previsão de vendas (Smith e Morrison, 2005).

São várias as definições para “pequenas áreas”, não havendo um consenso na literatura. Smith e Morrison (2005) afirmam que as pequenas áreas “*vary in size from less than an acre to thousands of square miles, and from a mere handful of residents (or none at all) to many millions*”. González e Torres (2012) afirmam que a expressão deve ser avaliada para cada país, de acordo com o tamanho da população e das características de sua divisão territorial interna, sendo que, no caso da América Latina, vários países “*cuentan con provincias o estados cuyas poblaciones permiten la proyección por el método de los componentes (...) para el área mayor y luego estimar las poblaciones de las subáreas con métodos matemáticos*” (González e Torres, 2012).

Para este trabalho, pequenas áreas são definidas como unidades cujo nível de desagregação espacial impossibilita a produção de estimativas populacionais intercensitárias ou pós-censitárias confiáveis, através de métodos demográficos convencionais (como métodos de extrapolação ou de corte componentes), seja pela falta de qualidade dos registros civis, seja pela oscilação da dinâmica

demográfica (especialmente da migração) em pequenos domínios (como é o caso dos municípios).

O tamanho das populações em pequenas áreas e a disponibilidade de dados criam problemas metodológicos, não verificados (ao menos não na mesma dimensão) em projeções para áreas maiores (Smith *et al.*, 2002). Além da qualidade e disponibilidade das fontes de dados, a unidade geográfica para a qual as projeções são construídas também constitui num problema para áreas menores, devido a mudanças nos limites, como é o caso da emancipação de novos municípios (Smith *et al.*, 2002). Do ponto de vista demográfico, a menor dimensão das áreas aumenta as oscilações (ou a variabilidade aleatória) das componentes – fecundidade, mortalidade e, principalmente, migração – em função do número limitado de eventos, somado aos problemas de sub registro, aumentando os desafios na produção de estimativas confiáveis, em especial nos países em desenvolvimento (Waldvogel, 1998; Smith *et al.*, 2002; George *et al.*, 2004; Cabrera, 2011; Santos e Barbieri, 2015).

Em projeções populacionais para áreas menores, é possível distinguir os métodos em duas abordagens gerais: os métodos *top down* e *bottom-up*. Nos métodos *top down*, são utilizados métodos de alocação para decomposição da população da área maior, através de uma “distribuição proporcional ou interpolativa para áreas menores” (Muniz, 2013). Dentre os principais modelos para a elaboração de projeções para pequenas áreas figuram o $A_i B_i$ (Madeira e Simões, 1972), e o Relação de Coortes (Duchesne, 1989). Entretanto, esses modelos não captam as inter-relações existentes nas pequenas áreas – por exemplo, não se considera as especificidades das trocas migratórias entre as áreas menores – o que constitui uma limitação importante, pois, parte considerável do crescimento em pequenas áreas pode estar associada à componente migração (Machado, 1993). Os métodos *bottom-up*, por sua vez, consistem na construção de projeções separadamente, para cada região, sendo que o total da área maior é obtido a partir da agregação de todas as projeções das áreas menores; privilegiando, assim, as características regionais (Machado, 1993).

A seção 2.1 apresenta primeiramente os métodos *top-down*, consolidados na literatura sobre projeções e utilizados pelo IBGE para a elaboração das estimativas

e projeções oficiais no Brasil (sendo objeto de estudo deste trabalho, na comparação com as estimativas via imagens de satélite). Na seção 2.2 são introduzidos os métodos *bottom-up* para a construção de pequenas áreas. Os métodos *bottom-up* constituem-se, basicamente, nos modelos de microssimulação e nos modelos de difusão espacial, e serão brevemente revisados. Na sequência, a seção 2.3 introduz os principais métodos que utilizam variáveis sintomáticas para a produção de estimativas para pequenas áreas, tema relacionado diretamente ao presente trabalho.

Na segunda parte do capítulo, são discutidos os pressupostos e características das projeções (revisão 2008 e 2013) e das estimativas pós-censitárias, do IBGE. Essas fontes de dados serão utilizadas para a comparação e validação externa dos modelos de estimativas populacionais baseados em imagens Landsat ETM. Ao final, o capítulo apresenta os principais métodos de avaliação dos erros de projeção, dos quais alguns serão utilizados para a avaliação das estimativas produzidas neste trabalho.

2.1 Métodos Top-down

2.1.1 Métodos de razão

O “*ratio method*”, ou projeção da participação no crescimento é considerado um método específico de extrapolação, sendo utilizado quando a área da população a ser projetada é parte de uma área maior, para o qual projeções estão disponíveis (United Nations, 1956; Shryock e Siegel, 1973). Uma vantagem dos *ratio methods* é que a soma das populações das áreas menores é igual ao total da população projetada na área maior, não sendo necessária a utilização de técnicas de padronização para a correção do nível entre as projeções (Santos e Barbieri, 2015). Normalmente aplicados em projeções para pequenas áreas, os *ratio methods* podem ser divididos em: *constant-share*; *shift-share* e *share-of-growth* (George *et al.*, 2004).

No método *constant-share*, a parcela da área menor em relação à área maior é mantida constante ao longo do tempo, sendo normalmente a parcela observada no

ano de início da projeção (ou ano de lançamento). Este método requer dados históricos de apenas uma data, sendo útil em situações de falta de qualidade dos dados ou ainda em áreas cujas unidades geográficas (administrativas) estejam em processo de mudança, o que dificulta a construção de séries históricas (George *et al.*, 2004).

$$P_{it} = (P_{it} / P_{jt})(P_{jt}) \quad (2.1)$$

Onde: P_{it} é a população projetada para a área pequena (i) no ano alvo; P_{it} é a população da área menor no ano de lançamento; P_{jt} é a população da área maior (j) no ano de lançamento; e P_{jt} é a projeção da área maior no ano alvo.

A principal limitação é que as projeções das áreas menores crescem exatamente no mesmo ritmo da área maior, o que não costuma ser um bom pressuposto, principalmente ao considerar a instabilidade da dinâmica migratória em áreas menores, e a importância dessa componente na definição do ritmo de crescimento, e conseqüentemente do tamanho populacional, dessas populações menores.

Diferentemente do método *constant-share*, o método *shift-share* aceita o pressuposto de mudança populacional das áreas menores. Segundo Santos e Barbieri (2015), a técnica mede a mudança na participação das pequenas áreas entre o ano base e o ano inicial de projeção, a partir da multiplicação de um fator que indica o horizonte de projeção (George *et al.*, 2004). A partir desse indicador de mudança na participação, agrega-se a distribuição inicial da projeção e multiplica-se novamente pela projeção da área maior (Santos e Barbieri, 2015).

$$P_{it} = (P_{jt})[(P_{it} / P_{jt}) + (z / y)((P_{it} / P_{jt}) - (P_{ib} / P_{jb}))] \quad (2.2)$$

Onde: i é a área menor; j é a área maior; z é o número de anos no horizonte de projeção; y é o número de anos no período base; b, l e t correspondem, respectivamente, aos anos base, de lançamento e ano alvo (futuro).

Apesar da sensibilidade do modelo a mudanças populacionais da área maior, George *et al.* (2004) afirma que, em horizontes mais longos de projeção, esse modelo pode estimar perdas substanciais de população, em áreas que apresentam crescimento muito lento ou crescimento negativo. Da mesma maneira, o método pode gerar projeções extremamente altas em áreas que apresentam crescimento acelerado. Portanto, assim como nos demais métodos de extrapolação, deve-se ter

cautela com períodos muito longos de projeção, bem como em áreas menores que apresentam crescimento acelerado (George *et al.*, 2004).

O método *share-of-growth* (ou *apportionment method*), conhecido no Brasil como AiBi ou método dos coeficientes, foi utilizado pela primeira vez no País por Madeira e Simões (1972), no qual projetaram a população urbana e rural para as Unidades da Federação, entre 1960 e 1980. O AiBi é utilizado pelo IBGE na produção das estimativas e projeções oficiais dos municípios brasileiros.

O método AiBi projeta a população da pequena área a partir de sua contribuição no crescimento absoluto da população esperada na área maior. Ao contrário das técnicas de extrapolação, esse método considera que o crescimento das áreas menores possui relação linear com o crescimento da área maior, o que, segundo George *et al.* (2004), produz geralmente projeções mais razoáveis, em relação aos demais métodos de extrapolação por razão. Segundo González e Torres (2012), para a aplicação do método é preciso dispor da projeção da população da área maior, e a população de cada uma das áreas menores, dos dois últimos censos.

$$P_{it} = P_{il} + \left[\left(\frac{P_{il} - P_{ib}}{P_{jl} - P_{jb}} \right) (P_{jt} - P_{jl}) \right] \quad (2.3)$$

$$P_{it} = A_i P_{jt} + B_i \quad (2.4)$$

Onde: i é a área menor; j é a área maior; b, l e t correspondem, respectivamente, aos anos base, de lançamento e ano alvo (futuro); A_i é um coeficiente de proporcionalidade entre o crescimento da área menor e da área maior, e que pode ser expresso como: $A_i = (P_{il} - P_{ib}) / (P_{jl} - P_{jb})$; B_i é um coeficiente linear de correção, e que pode ser expresso como: $B_i = P_{il} - A_i P_{jl}$ (Waldvogel, 1998)

Em situações improváveis, como municípios que tendem a apresentar valores negativos de população ao longo do período de projeção, é possível corrigir arbitrariamente os parâmetros A_i e B_i , como, por exemplo, em Waldvogel (1998).

Segundo Santos e Barbieri (2015), projeções que utilizam o método AiBi são mais precisas nos casos em que o erro da grande área é pequeno; ou quando o crescimento da pequena e da grande área apresentam uma relação aproximadamente constante; quando a quantidade de pequenas áreas é menor, tendo cada área uma maior participação no crescimento da área maior; quando as

relações são constantes entre os dois níveis, e nos casos em que não há grandes diferenças de tamanho populacional entre as pequenas áreas.

Os mesmos autores calcularam medidas de erro para cinco técnicas de extrapolação matemática (frequentemente utilizadas para projeções em pequenas áreas), para as Microrregiões de Minas Gerais em 2010. Os autores mostraram que o método AiBi (*share of growth*) apresentou o menor erro entre os métodos escolhidos, seguido pela partição constante (*constant-share*) e a extrapolação linear; enquanto os métodos *shift-share* e extrapolação exponencial apresentaram os piores indicadores de erro em relação às demais técnicas. Os autores concluem que o método AiBi produz resultados mais aceitáveis, por não ser determinado diretamente pelo tamanho da população, mas sim pela participação das pequenas áreas no crescimento da área maior (Santos e Barbieri, 2015).

2.1.2 Relação de Coortes

O método de relação de coortes parte do pressuposto de que, num período base, o crescimento populacional de uma área menor apresenta relação linear com o mesmo grupo populacional (ou coorte) da área maior. Segundo González e Torres (2012) “*partiendo de una proyección por componentes de la población del área mayor, se calculan las tasas de crecimiento de sus cohortes y se aplican a las cohortes de las poblaciones menores, diferenciando entre cada área menor*” (González e Torres, 2012).

O método de Relação de Coortes considera a estrutura da população e algumas mudanças nas variáveis demográficas, além de manter a correspondência entre a soma das projeções das áreas menores e as projeções conhecidas da área maior. A grande diferença em relação ao método AiBi é que o método de Relação de coortes incorpora o comportamento das estruturas etárias no cômputo das projeções, e não apenas a da população total, como no AiBi (Jannuzzi, 2007).

A limitação do modelo consiste em seu uso para grupos populacionais muito pequenos, que apresentam grande instabilidade, além de informações insuficientes por sexo e idade (Duchesne, 1989; CEDEPLAR, 2014; Santos e Barbieri, 2015).

As equações 2.5 a 2.7 mostram o cálculo para a quinquênios de projeção e grupos etários quinquenais, respectivamente para os menores de 5 anos, para a população entre 5 e 80 anos e para a população com 80 anos ou mais.

$${}_5N_0^{t+5} = B^{t,t+5} \times P_b^{t,t+5} \times K_b^{t-5,t} \quad (2.5)$$

$${}_5N_{x+5}^{t+5} = {}_5N_x^t \times CR_x^{t,t+5} \times K_x^{t,t+5} \quad (2.6)$$

$$N_{80+}^{t+5} = N_{75+}^t \times CR_{75+}^{t,t+5} \times K_{75+}^{t,t+5} \quad (2.7)$$

Onde:

${}_5N_0^{t+5}$ é a população do grupo quinquenal entre 0 e 4 anos, da área menor, no momento t+5;

${}_5N_{x+5}^{t+5}$ é a população do grupo quinquenal de idade x+5, x+10 anos, da área menor, no momento t+5;

N_{80+}^{t+5} é a população com 80 anos ou mais de idade, na área menor, no momento t+5;

$B^{t,t+5}$ é o total de nascimentos ocorridos na área menor, entre os períodos t e t+5;

$P_b^{t,t+5}$ é a relação de sobrevivência ao nascimento dos nascidos vivos do quinquênio, da área maior, no período t, t+5;

$K_b^{t-5,t}$ é o quociente entre as relações de sobrevivência, ao nascimento, da área menor em relação à área maior, observado no período anterior (t-5, t).

${}_5N_x^t$ é a população inicial do grupo quinquenal de idade x, x+5 da área menor, no momento t;

$CR_x^{t,t+5}$ é a relação de sobrevivência, na projeção da área maior, correspondente ao grupo quinquenal de idade x, x+5 anos, no momento t, que alcança as idades x+5, x+10, no momento t+5;

$K_x^{t,t+5}$ é o coeficiente entre as relações de sobrevivência, da área menor em relação à área maior, correspondentes ao grupo quinquenal de idade x, x+5 anos, no momento t e que alcança as idades x+5, x+10 anos, no momento t+5.

Calculado o fator K, deve ser realizado um conjunto de modificações para a correção de valores extremos que podem eventualmente ocorrer para alguns municípios. Normalmente, a correção do fator K é feita de modo subjetivo, e, nesse contexto, Assunção (2002) propõe uma metodologia estocástica, baseada em estatística bayesiana, para obter melhores estimativas para os fatores de crescimento diferenciado. Para a estimação dos fatores K, o autor adotou distribuições *a priori* para os parâmetros desconhecidos, através de métodos de

cadeia de Markov via Monte Carlo, para a obtenção da distribuição a posteriori. Considerou-se a média a posteriori das distribuições como o estimador pontual dos parâmetros. Para tanto, utiliza de dados do Censo 1991 e da Contagem 1996 para a projeção da população do ano 2000.

Jardim (2003) estimou a população dos municípios do Rio Grande do Sul nos anos posteriores à 2000, a partir dos métodos de Correlação (que será discutido mais adiante) de Razões e de Relação de Coortes (método de Duchesne), além da integração dos dois métodos, através da média dos resultados, ponderada pelo tempo (pelo ano da estimativa). Para o método de Correlação de Razões, Jardim (2003) utilizou taxas vitais (nascidos vivos e óbitos), matrícula escolar para séries do Ensino Fundamental, número de eleitores e razões entre dados censitários de 1991 e 2000. Dentre os resultados, o autor destaca a diminuição do erro das estimativas quando se considera a média dos dois métodos.

2.2 Métodos Bottom-up

2.2.1 Microssimulação

Nas últimas décadas, o aumento da capacidade computacional tem permitido o desenvolvimento de metodologias para projeções populacionais em níveis baixíssimos de desagregação geográfica, através de modelos de microssimulação. Esses modelos simulam eventos do ciclo de vida e mantém registros detalhados de transições por estados demográficos – como casamentos, divórcios, nascimentos – para cada indivíduo de uma amostra selecionada (Imhoff e Post, 1998; O’neil *et al.*, 2001; Willekens, 2006).

Os modelos de microssimulação partem de experimentos aleatórios repetidos, calculados a partir de probabilidades médias de uma amostra da população, o que possibilita a análise estocástica dos resultados (Imhoff e Post, 1998; O’neil *et al.*, 2001; Willekens, 2006; Ledent e Zeng, 2010). A principal vantagem é a possibilidade de incorporação de diversos “estados” ou atributos dos indivíduos, algo de difícil reprodução em modelos tradicionais (de macro simulação), como os de coorte-componentes ou multiestado. Contudo, deve-se considerar que a grande

quantidade de variáveis explicativas resulta em relações complexas, além de aumentar a aleatoriedade e as medidas de erro dos modelos (Willekens, 2006).

Outras limitações podem ser destacadas, como a natureza dos experimentos aleatórios de Monte Carlo, no qual diferentes execuções do mesmo modelo de entrada (a partir de uma mesma amostra da população) produzem diferentes conjuntos de resultados (Imhoff e Post, 1998). Essa característica estocástica de variação aleatória pode ser reduzida aumentando o tamanho da amostra ou tomando a média de um conjunto de rodagens, embora não possa ser completamente solucionada. Outro problema diz respeito à população de partida dos modelos, que são amostras da população total, sujeitas a erros de amostragem clássicos, principalmente para pequenos subgrupos (Imhoff e Post, 1998). Vale ressaltar ainda a maior complexidade na rodagem dos modelos de micro simulação, tanto em relação à maior quantidade de variáveis como em termos de operacionalização e processamento computacional.

Segundo Ledent e Zeng (2010), nos últimos anos esforços tem sido realizados para a elaboração de projeções, a partir da integração de modelos de simulação macro e micro, para a modelagem de projeções multiestados. O European Multi-Institutional MicMac Project, liderado por Frans Willekens, combina abordagens de macro e micro simulação para a elaboração de projeções em estados multidimensionais (definida por idade, sexo, saúde e participação da força de trabalho), bem como o tempo gasto em cada um desses estados (Ledent e Zeng, 2010; Willekens, 2006). Enquanto a macro simulação envolve a construção de coortes biográficas no nível da população, a partir de grandes bases de dados (como os censos), a micro simulação parte de dados de amostragem para a produção de biografias individuais mais detalhadas, que são, então, agrupadas para determinar as características da população. Em seguida, os dados de macro e micro simulação são integrados, a partir da utilização de taxas de transição, estimadas a partir dos dados da pesquisa, o que torna a dinâmica entre os dois níveis internamente consistentes (Willekens, 2006).

2.2.2 Modelos de difusão espacial

A partir do final da década de 1980, a crescente disponibilização de dados e a expansão da computação e dos Sistemas de Informação Geográfica (SIG), permitiram avanços importantes na representação espacial de modelos dinâmicos de sistemas urbanos (Baker, 1989; Batty *et al.*, 1997; Almeida, 2004), que foram, por sua vez, a base para o desenvolvimento de novos métodos de projeção populacional, em especial para pequenas áreas.

Na perspectiva do espaço, o crescimento populacional está normalmente associado à expansão de áreas densamente povoadas, em direção a áreas periféricas, e os “modelos de difusão espacial” refletem essa dinâmica de espraiamento de um dado fenômeno particular, no tempo e no espaço (Batty *et al.*, 1997; Almeida, 2004; Smith *et al.*, 2002). Introduzidos na década de 1950 por Hägerstrand, os modelos de difusão espacial são modelos matemáticos que buscam reproduzir a dinâmica de sistemas urbanos, a partir da incorporação de medidas de distância, e que serviram de base para o surgimento dos modelos dinâmicos, a partir da década de 1970.

Esses modelos dinâmicos⁷, entre os quais se destacam os modelos apoiados em autômatos celulares, são sistemas baseados em grades regulares, com uma variável discreta em cada localidade (ou célula), nos quais os processos de mudanças (baseados em regras de transição) são representados em tempos discretos e em cada célula, como resultado do que aconteceu nas células vizinhas, num momento anterior do tempo (White e Engelen, 1993; Batty *et al.*, 1997; Soares-Filho *et al.*, 2002; Almeida, 2004; Sante *et al.*, 2010, Umbelino, 2012). Segundo Sante *et al.* (2010), a habilidade dos autômatos celulares em simular o crescimento urbano é baseada no pressuposto de que o desenvolvimento urbano passado afeta o padrão futuro através das interações locais entre os usos da terra. Assim, os autômatos celulares podem ser considerados como uma categoria de modelos

⁷ Segundo Baker (1989), os modelos dinâmicos analisam e reproduzem, em ambiente computacional, os padrões espaciais resultantes das mudanças na paisagem. Segundo Almeida (2004), um modelo é dinâmico se apresenta uma dimensão temporal explícita, se o estado (ou situação) atual depende de estados anteriores e se as entradas e saídas se modificam com o tempo.

dinâmicos espacialmente explícitos⁸ (ou modelos dinâmicos espaciais), por serem modelos que procuram descrever e prever a evolução de atributos ambientais em subunidades de área com distintas localização e configuração (Baker, 1989; Soares-Filho *et al.*, 2002; Almeida, 2004).

Os autômatos celulares ganharam proeminência no âmbito urbano nas décadas de 1960 e 1970. Em 1960, Segundo Batty *et al.* (1997), Chapin e colegas da Carolina do Norte foram pioneiros na utilização de modelos espaciais articulados em células, e que consideravam a mudança de estado das células, em função dos efeitos de vizinhança. No contexto da Geografia Quantitativa, Arthur Burks e Waldo Tobler tiveram papel importante no desenvolvimento de modelos baseados em células, para aplicação aos sistemas urbanos (Batty *et al.*, 1997; Sante *et al.*, 2010). Tobler propôs a utilização de modelos celulares em sistemas geográficos, e na década de 1980, Couclelis (influenciado por Tobler) avançou nas aplicações sobre mudanças urbanas, período marcado pela utilização em larga escala dos modelos dinâmicos. Na década de 1990, consolidaram-se avanços nos modelos urbanos via autômatos celulares, a partir da incorporação de dimensões ambientais, socioeconômicas e políticas, além do desenvolvimento na articulação analítica de fatores micro e macroespaciais (White e Engelen, 1993; Batty *et al.*, 1997; Sante *et al.*, 2010).

Os modelos de autômatos celulares são muito úteis na investigação de dinâmicas de mudanças urbanas (como mudanças no uso e ocupação do solo), e, conseqüentemente, para o planejamento urbano, seja pela sua característica intrinsecamente espacial, seja pela capacidade em analisar fenômenos espaciais complexos – como simulações do crescimento urbano – a partir de regras (ou interações locais) bastante simples (White e Engelen, 1993; Soares-Filho *et al.*, 2002; Almeida, 2004; Sante *et al.*, 2010). Essas características tornam esses modelos bastante promissores no campo das projeções para pequenas áreas, pois, diferentemente dos modelos tradicionais (que são baseados em séries históricas de uma dada localidade ou área), os modelos dinâmicos se baseiam nas

⁸ Segundo Baker (1989), os modelos de simulação espacialmente explícitos podem ser utilizados para testar hipóteses sobre a evolução na paisagem, e sobre diversos cenários, como, por exemplo, a partir de diferentes estruturas socioeconômicas, políticas e ambientais.

tendências históricas das áreas vizinhas (Batty et al., 1997; Smith et al., 2002; Sante et al., 2010). Além disso, esses modelos têm o potencial de predição de mudanças mais abruptas no ritmo de crescimento de uma dada área, fenômeno comum em áreas muito pequenas (Smith et al., 2002). Todavia, os métodos de projeção que utilizam Sistemas de Informações Geográficas são complexos e caros (Smith et al., 2002), o que dificulta a utilização de forma sistemática para grandes extensões de áreas. Todavia, os avanços tecnológicos e computacionais (em relação ao aumento da capacidade de processamento) tornam promissor, para um futuro próximo, a replicação de métodos de projeções populacionais via SIG para grandes áreas.

Segundo Sante *et al.* (2010), os primeiros modelos de dinâmica urbana focaram na simulação de estruturas urbanas simples. Mais recentemente, tem se observado o desenvolvimento de modelos de autômatos celulares que consideram o relaxamento das regras de transição, como a transição entre células não regulares, a incorporação de probabilidade, a utilização de metodologias de inteligência artificial e lógica *fuzzy* nas regras de transição; além da incorporação de outros fatores, para além dos efeitos de vizinhança, como distância de arruamentos e de diversas infraestruturas urbanas, fatores ambientais, etc. (Sante *et al.*, 2010).

Na literatura nacional, pesquisadores do INPE tem tradição na utilização de modelos de autômatos celulares para a análise de mudanças no uso e ocupação do solo. Almeida (2004) simulou mudanças de uso do solo urbano em Bauru e Piracicaba (SP). Umbelino (2012) simulou cenários de expansão da mancha urbana e de adensamento urbano para a Região Metropolitana de Belo Horizonte (RMBH). Soares Filho *et al.* (2002), por sua vez, foi pioneiro ao desenvolver o software “Dinâmica”, inicialmente idealizado para a modelagem espacialmente explícita da dinâmica da paisagem Amazônica (em especial para a modelagem de pequenas propriedades). O software Dinâmica disponibiliza funções de transição baseados em vizinhanças multiescalares, além de simulações estocásticas e a possibilidade de aplicação de regressões logísticas para o cálculo das probabilidades de transição da dinâmica espacial.

Embora a discussão realizada nessa seção tenha relação com Sistemas de Informações Geográficas e sensoriamento remoto, é importante ressaltar que este

tipo de aplicação diz respeito à elaboração de projeções populacionais, por meio de simulações. Os métodos que serão explorados neste trabalho têm como objetivo a produção de estimativas (de anos anteriores), sendo que as imagens de satélite são utilizadas como informações sintomáticas da mudança populacional. Portanto, são conceitos e finalidades diferentes. Todavia, fez-se necessária a discussão sobre os modelos de difusão espacial, para ratificar essa distinção existente entre estes modelos e as estimativas produzidas via sensoriamento remoto, objeto de estudo da tese. Ademais, trata-se de uma revisão bibliográfica importante para a compreensão da relação entre o campo das estimativas e projeções populacionais com o campo da análise espacial e do sensoriamento remoto.

Diferentemente dos métodos *top down*, que desagregam projeções populacionais para áreas menores, e dos métodos *bottom up*, que produzem projeções populacionais a partir de simulações, há um conjunto de métodos que utilizam de variáveis sintomáticas da mudança populacional para a produção de estimativas em períodos pós-censitários, e que serão discutidas no próximo item.

2.3 Estimativas Populacionais para Pequenas Áreas a partir de Variáveis Sintomáticas

A crescente demanda por estimativas populacionais para pequenas áreas, especialmente em períodos intercensitários, tem resultado no desenvolvimento de modelos matemáticos que utilizam de variáveis sintomáticas, que associam as fontes de informação ao volume e à mudança populacional, como registros administrativos disponíveis para áreas menores, permitindo a atualização da população em períodos não censitários.

As variáveis sintomáticas têm como requisitos a existência de registros permanentes (disponibilidade temporal), disponibilidade geográfica (para as áreas no qual se deseja produzir as estimativas), além da alta correlação entre o tamanho e a evolução da população (González e Torres, 2012). Em relação às fontes de dados, são comuns o uso de registros civis (dados vitais), bem como registros referente à serviços básicos (como água e eletricidade), além de estatísticas de matrícula escolar, registros eleitorais, dados de habitação, entre outros (González

e Torres, 2012). As variáveis sintomáticas são, também, “*de gran utilidad para examinar la validez de aquellas proyecciones y señalar correcciones focalizadas*” (Álvarez, 2001).

Dentre as limitações, González e Torres (2012) apontam a omissão diferencial dos registros por áreas e ao longo do tempo, alterações nas divisões administrativas, além de mudanças nos programas sociais e bases de dados, dificultando as análises temporais. Esquivel (2001) ressalta ainda que a exatidão de um método não pode ser generalizada, devendo-se evitar o uso “mecânico” dos métodos desenvolvidos, diante à diferença da qualidade das informações para diferentes áreas e ao longo do tempo. Essa afirmativa é fundamental para a definição dos propósitos deste trabalho. Como já fora comentado, a avaliação da eficácia das estimativas por imagens orbitais de média resolução, a partir dos municípios e métodos selecionados, tem como objetivo contribuir com este campo do conhecimento, a partir de uma nova abordagem (que inclui validações externas e análises temporais); contudo, não se pretende generalizar os resultados encontrados, especialmente se considerar que algumas características das imagens orbitais se modificam ao longo do tempo (tema que será discutido mais adiante).

Segundo González e Torres (2012), as variáveis sintomáticas podem ser divididas em dois grandes grupos: informações de fluxo e de estoque. As informações de fluxo representam medidas da mudança populacional, especialmente das componentes demográficas (fecundidade, mortalidade e migração), como registros de nascimentos, ou ainda dados de matrícula escolar. Já os dados de estoque estão relacionados ao tamanho total da população, como é o caso das imagens de satélite, muito úteis para a produção de estimativas de pequenas populações (González e Torres, 2012).

Em relação às aplicações, González e Torres (2012) afirmam que os modelos baseados em variáveis sintomáticas são pouco utilizados, especialmente no caso dos órgãos oficiais:

“las estimaciones que emplean variables sintomáticas no están generalizadas em âmbitos oficiales, ya sea por desconocimiento de las técnicas, falta de capacitación del personal y también por la baja frecuencia con que se efectúan estimaciones para áreas pequeñas” (González e Torres, 2012, p. 106).

Descreve-se a seguir os principais métodos utilizados na literatura, pois, o conhecimento sobre os métodos utilizados para a produção de estimativas para pequenas áreas via variáveis sintomáticas contribui para a qualificação da análise da eficácia das estimativas produzidas via sensoriamento remoto (a partir da discussão sobre as vantagens e limitações de cada modelo existente).

2.3.1 Distribuição Pro Rata

A distribuição pro rata supõe que a razão entre a população das áreas menores e da área maior é igual à razão correspondente a uma variável sintomática (Álvarez, 2001; Jardim, 2001; Cabrera, 2011; González e Torres, 2012). Para o cálculo é necessário que a variável sintomática esteja disponível para o ano-base e para o ano em que se deseja estimar, tanto para as áreas menores como para a área maior (ou total).

$$P_{i,t+n} = \frac{S_{i,t+n}}{S_{T,t+n}} \times P_{T,t+n} \quad (2.8)$$

Onde: t é o ano do último censo; n é o número de anos transcorridos desde o último censo; $P_{i,t+n}$ é a população da área i no momento t+n; $S_{i,t+n}$ é o valor da variável sintomática da área i no momento t+n; $S_{T,t+n}$ é o valor da variável sintomática da área maior no momento t+n; $P_{T,t+n}$ é a população projetada da área maior no momento t+n.

Dentre as vantagens, destaca-se a possibilidade de se aplicar uma única variável sintomática, e para apenas um ano-calendário. Dentre as desvantagens, ressalta-se que as estimativas são afetadas quando o pressuposto de constância não se sustenta (Álvarez, 2001; Cabrera, 2011; González e Torres, 2012).

2.3.2 Distribuição Proporcional

Neste método, parte-se do pressuposto que a população varia na mesma proporção da variável sintomática. Segundo González e Torres (2012), a diferença em relação ao método pro rata é que, no método de distribuição proporcional, considera-se um ajuste a partir das mudanças ocorridas na variável sintomática e na população da área maior, no período considerado. Jardim (2001) afirma que o ajuste impede que as estimativas sejam influenciadas pelas mudanças na qualidade das variáveis sintomáticas (desde que estas apresentem a mesma tendência para todas as áreas menores), e permite a consistência das estimativas com a população total. Para o cálculo deste método são necessárias informações das variáveis sintomáticas em dois momentos no tempo (um ano inicial, normalmente um censo e o ano de interesse que se deseja estimar) [Álvarez, 2001; Jardim, 2001; Cabrera, 2001; González e Torres, 2012].

$$P_{i,t+n} = \frac{S_{i,t+n}}{S_{T,t+n}} \times P_{i,t} \times F_a \quad (2.9)$$

$$F_a = \frac{P_{T,t+n}}{\sum \left(P_{i,t} \times \frac{S_{i,t+n}}{S_{i,t}} \right)} \quad (2.10)$$

Onde: F_a é o fator de ajuste da fórmula para que a soma de $P_{i,t+n}$ seja igual a PT_{t+n} .

Segundo Cabrera (2011), este método tem a vantagem de permitir a utilização de variáveis cuja distribuição não seja semelhante à população, desde que a tendência de variação da informação sintomática seja um bom indicador das mudanças no tamanho da população.

2.3.3 Taxas Vitais

O método das Taxas Vitais necessita de informações sobre o número de nascimentos e mortes de cada área menor no ano base e no ano considerado para estimação, bem como a população da área maior e as taxas brutas de natalidade e de mortalidade (Álvarez, 2001; Cabrera, 2011; González e Torres, 2012). Proposto por Donald Bogue em 1950 (apud Jardim, 2001), o método parte do pressuposto de que há uma relação inversa entre as variações anuais das taxas de

fecundidade e mortalidade das áreas menores, e que o desvio proporcional destas taxas, em relação às taxas da área maior, permanece constante ao longo do tempo (González e Torres, 2012). As estimativas para as áreas menores são calculadas a partir da razão entre as taxas das áreas menores e da área maior.

$$P_{i,t+n} = a \times \left(\frac{B_{i,t+n}}{b_{i,t+n}} \right) + (1-a) \times \left(\frac{D_{i,t+n}}{d_{i,t+n}} \right) \quad (2.11)$$

$$b_{i,t+n} = \frac{B_{i,t} / P_{i,t}}{B_{T,t} / P_{T,t}} \times \left(\frac{B_{T,t+n}}{P_{T,t+n}} \right) \quad (2.12)$$

$$d_{i,t+n} = \frac{D_{i,t} / P_{i,t}}{D_{T,t} / P_{T,t}} \times \left(\frac{D_{T,t+n}}{P_{T,t+n}} \right) \quad (2.13)$$

Onde: a é o coeficiente de ponderação (0,5 para um peso igual entre as estimativas derivadas de nascimentos e mortes); B são os filhos nascidos vivos; b é a taxa bruta de natalidade; D são as mortes; d é a taxa bruta de mortalidade.

2.3.4 Razão Censitária

Este método parte de taxas de ocorrência de uma dada variável sintomática, sob o pressuposto de que a população da área menor tem uma relação proporcional à área maior, no período entre o último censo e o ano que se pretende estimar (Esquivel, 2001; Cabrera, 2011; González e Torres, 2012). Segundo Esquivel (2001) e González e Torres (2012), primeiramente é calculada a taxa de ocorrência de cada variável sintomática para a área menor, durante o último período censitário (equação 2.14). Num segundo momento, a taxa da variável sintomática no momento $t+n$ é aproximada por meio de um parâmetro ϕ , que representa a razão da mudança na variável sintomática entre o ano censitário t e o ano $t+n$ (equação 3.15) [Esquivel, 2001; Cabrera, 2011; González e Torres, 2012].

$$r_{i,t} = \frac{S_{i,t}}{P_{i,t}} \quad (2.14) \quad r_{i,t+n} = \phi \times r_{i,t} \quad (2.15)$$

Onde: $r_{i,t}$ é a taxa de ocorrência do indicador sintomático S para a localidade i no ano do último censo; $S_{i,t}$ é a informação da variável sintomática S observada na localidade i no ano do último censo; $P_{i,t}$ é a população da localidade i no último censo.

Para se estimar o valor de ϕ , deve-se obter a informação da área maior (T) que contemple as áreas menores, sobre a qual se podem fazer estimativas independentes da população no momento t+n. Assim, são calculadas as taxas de ocorrência de S na área T, no momento do censo (t) e no momento t+n, a partir da equação 2.16 (González e Torres, 2012). Considera-se que a razão da mudança na variável sintomática para a área maior aproxima-se do valor ϕ (equação 2.16), enquanto a equação 2.17 define o cálculo de $r_{i,t+n}$; e, por fim, a equação 2.18 a estimativa populacional para a localidade i no momento t+n ($P_{i,t+n}$).

$$\phi = \frac{r_{T,t+n}}{r_{T,t}} \quad (2.16)$$

$$r_{i,t+n} = \phi \times r_{i,t} = \frac{r_{T,t+n}}{r_{T,t}} \times r_{i,t} \quad (2.17)$$

$$P_{i,t+n} = \frac{S_{i,t+n}}{r_{i,t+n}} \quad (2.18)$$

Segundo Esquivel (2001), as diferentes variáveis sintomáticas que podem ser utilizadas normalmente resultam em mais de uma estimativa demográfica, sendo que a estimativa final pode ser obtida a partir da média simples ou ponderada das estimativas individuais.

2.3.5 Diferença de Taxa

O método da Diferença de Taxa é baseado no método da Razão Censitária, cuja diferença está na forma de se estimar a taxa de ocorrência da variável sintomática para as áreas menores, sendo que o método supõe que o crescimento da taxa para uma área menor é igual ao crescimento da área maior (Esquivel, 2001; Cabrera, 2011; González e Torres, 2012). A equação abaixo apresenta o cálculo da taxa de ocorrência.

$$r_{i,t+n} = r_{i,t} + (r_{T,t+n} - r_{T,t}) \quad (2.19)$$

2.3.6 Método Composto

O método composto consiste na produção de estimativas desagregadas por sexo e idade, a partir de diferentes variáveis sintomáticas, no qual cada uma representa intervalos etários específicos, e que, ao final, são somadas para a obtenção da população total (Jardim, 2001; Cabrera, 2011; González e Torres, 2012). Assim, os nascidos vivos são utilizados para se estimar a população entre 0 e 6 anos, a matrícula escolar primária para se estimar a população entre 7 e 14 anos; o número de eleitores para se estimar a população entre 15 e 49 anos; e o número de mortes para se estimar a população com 50 anos ou mais (Jardim, 2001; González e Torres, 2012). A equação abaixo mostra o cálculo para a população entre 0 e 6 anos. Ressalta-se que a forma de cálculo é a mesma para os demais grupos etários.

$$P_{i,t+n}(0-6) = \frac{P_{i,t}(0-6) \times \frac{B_{i,t+n}}{B_{i,t}}}{P_{T,t}(0-6) \times \frac{B_{T,t+n}}{B_{T,t}}} \times P_{T,t+n}(0-6) \quad (2.20)$$

2.3.7 Correlação de Razões

Os métodos de correlação partem de um conjunto de equações baseadas em modelos de regressão, cujo objetivo é identificar as correlações existentes entre cada uma das variáveis sintomáticas do crescimento populacional e o próprio crescimento da população, a partir de modelos de regressão linear (Álvarez, 2001; González e Torres, 2012; Cabrera, 2011; Silva e Santos, 2016). Para tanto, estima-se um parâmetro $Y_{i,t}$ que expressa a relação da área maior e a área menor a ser projetada, podendo ser também compreendida como a proporção da população da área i sobre o total, no momento t e no momento 0. Assim, Y é o parâmetro para a projeção da população, baseado na diferença de razões (Álvarez, 2001; González e Torres, 2012).

$$Y_{i,t} = a_0 + a_1 * X1_{i,t} + \dots + a_j * Xn_{i,t} + u_i \quad (2.21)$$

Onde: a_j é o parâmetro de correlação entre a variável sintomática j e o parâmetro de crescimento; $X_{i,t}$ é qualquer variável sintomática ajustada pela diferença de razões. $X_{i,t}$ também pode ser compreendida como a razão entre a proporção de uma dada variável sintomática da área i sobre o total, no momento t e no momento 0; u_i é um termo de erro.

No método de correlação de razões, a evolução da população é definida pela correlação entre a razão de dois coeficientes de proporção em tempos distintos e essa mesma razão de proporções de um conjunto de covariáveis, como mostram as equações 3.22 e 3.23 (Álvarez, 2001; González e Torres, 2012).

$$Y_{i,t} = \frac{P_{i,t} / P_{T,t}}{P_{i,0} / P_{T,0}} \quad (2.22)$$

$$X_{j,i,t} = \frac{S_{j,i,t} / S_{j,T,t}}{S_{j,i,0} / S_{j,T,0}} \quad (2.23)$$

$Y_{i,t}$ é a razão entre a proporção da população da área i sobre a total, no momento t e no momento 0; $X_{j,i,t}$ é a razão entre a proporção da

variável sintomática j da área i sobre o total, no momento t e no momento 0 .

A população da área i (ou seja, da área menor) no momento $t+n$ é estimada a partir do modelo de regressão múltipla, baseado no momento 0 , no momento t (normalmente os últimos dois censos), na estimativa posterior de $Y_{i,t+n}$ com base no período $t, t+n$, e numa projeção da área total T no momento $t+n$ (Álvarez, 2001; González e Torres, 2012; Silva e Santos, 2016).

$$P_{i,t+n} = Y_{i,t+n} \times \frac{P_{i,t}}{P_{T,t}} \times P_{T,t+n} \quad (2.24)$$

Segundo González e Torres (2012), a efetividade do método depende da qualidade da projeção da área maior e da estabilidade da relação entre a variável dependente e as variáveis independentes, ao longo do período de estimação. Dentre as vantagens, destaca-se a possibilidade de se trabalhar com mais de uma variável sintomática (González e Torres, 2012). Jardim (2003) destaca o potencial desse método em lidar com o problema de incompatibilidade territorial, resultante da mudança das unidades geográficas, como no caso da emancipação de municípios. Segundo Jardim (2003), o método de Correlação de Razões não necessita de informações de uma subárea específica para o qual se deseja obter estimativa. Na definição dos parâmetros do modelo, embora possam ser considerados municípios que perderam área geográfica no período, as estimativas para os anos posteriores podem ser obtidas para todos os municípios, inclusive para aqueles criados após o ano base “desde que as variáveis sintomáticas utilizadas no modelo sejam referentes a essa divisão territorial” (Jardim, 2003).

2.3.8 Correlação de Taxa

Segundo Esquivel (2001), no método de Correlação de Razões, os coeficientes do modelo são calculados tendo como referência o período entre os censos, normalmente decenais. No entanto, o período de tempo transcorrido entre o ano-base e as estimativas é menor, o que pode afetar a estabilidade dos coeficientes e das estimativas produzidas. Nesse sentido, Esquivel (2001) propõe uma modificação do método de Correlação de Razões, baseada na aproximação

exponencial da mudança, calculada a partir do logaritmo natural de cada razão, e da divisão pelo tempo transcorrido entre os censos (Álvarez, 2001; Esquivel, 2001; Cabrera, 2011; González e Torres, 2012).

$$Y^*_{i,t} = a_0 + a_1 * X^*_{1,i,t} + \dots + a_j * X^*_{n,i,t} + u_i \quad (2.25)$$

$$Y^*_{i,t} = \frac{\ln(Y_{i,t})}{k} \quad (2.26)$$

$$X^*_{i,t} = \frac{\ln(X_{i,t})}{k} \quad (2.27)$$

Segundo González e Torres (2012), o método de correlação de taxa possui a mesma tendência do método de Correlação de Razões, dado que a diferença entre os métodos está somente na transformação logarítmica. Ademais, o método de Correlação de Razões apresenta maiores inconsistências, dado que, no método de correlação da taxa, o coeficiente Y_i é anualizado a partir da linearização dos coeficientes, podendo ser aplicado em diferentes horizontes de projeção (Silva e Santos, 2016).

2.3.9 Correlação de Diferença

Assim como no método anterior, a Correlação de Diferença trata-se de uma modificação do método de Correlação de Razões, sendo a principal diferença a forma de cálculo da variação da população, baseadas em diferenças ao invés de razões (Álvarez, 2001; Cabrera, 2011; González e Torres, 2012).

$$W_{i,t} = a_t + a_i * Z_{i,n} + \dots + a_j * Z_{j,n} + e \quad (2.28)$$

$$W_{i,t} = \frac{P_{i,t}}{P_{T,t}} - \frac{P_{i,0}}{P_{T,0}} \quad (2.29)$$

$$Z_{j,i,t} = \frac{S_{i,t}}{S_{T,t}} - \frac{S_{i,0}}{S_{T,0}} ; (j=1,\dots,n) \quad (2.30)$$

Onde: $W_{i,t}$ é a diferença entre a proporção da população da área i no momento t e o momento 0 ; $Z_{j,t}$ é a diferença entre a proporção de cada variável sintomática da área i no momento t e no momento 0 .

Segundo González e Torres (2012), a população da área i no momento $t+n$ é estimada a partir de um modelo de regressão múltipla, que se baseia no momento 0 , no momento t , na estimativa posterior de $W_{i,t+n}$ com base no período $t, t+n$, e numa projeção da área total T no momento $t+n$ (González e Torres, 2012).

$$P_{i,t+n} = \left(W_{i,t+n} \times \frac{P_{i,t}}{P_{T,t}} \right) \times P_{T,t+n} \quad (2.31)$$

González e Torres (2012) afirmam que o método apresenta as mesmas vantagens e desvantagens do método de Correlação de Razões. Contudo, ressalta-se que as variações de população calculadas pelo método da Correlação de Diferença resultam em correlações maiores entre as variáveis populacionais e sintomáticas, além de que as estimativas costumam ser menos afetadas por mudanças temporais nas variáveis (González e Torres, 2012).

2.3.10 Aplicações com foco na América Latina

Nas últimas duas décadas tem aumentado os estudos que buscam o desenvolvimento de metodologias para a produção de estimativas para pequenas áreas, dentre as quais aquelas que consideram o uso de variáveis sintomáticas. Dentre esses esforços destaca-se o CELADE, que vem elaborando estudos e desenvolvendo metodologias para melhorar as estimativas populacionais oriundas de extrapolações matemáticas, sendo que alguns estudos exploraram metodologias baseados em indicadores indiretos do tamanho das populações.

Sobre esses esforços, Álvarez (2001), por exemplo, avalia um conjunto de métodos que utilizam variáveis sintomáticas para a construção de estimativas, tendo como área de estudo Departamentos da Argentina. Foram utilizados registros vitais e registros escolares (da educação primária) para os anos de 1991 e 1996, primeiramente para a comparação com o censo do país, e, num segundo momento, para a comparação das estimativas pós-censitárias com projeções oficiais disponíveis para as áreas em estudo. Álvarez (2001) conclui que os métodos de

estimativas baseados em modelos de correlação linear são mais eficazes que aqueles gerados a partir de distribuições, como os métodos Pro Rata, de Distribuição Proporcional e das Taxas Vitais, sendo que, dentre os métodos de correlação, os que apresentaram as melhores estimativas foram os de Correlação de Razões e Correlação de Taxa. Já em relação às variáveis sintomáticas selecionadas, a matrícula escolar foi a que apresentou maior poder de predição.

Esquivel (2001) considerou a não realização do censo de 1990 na Costa Rica para investigar técnicas para a produção de estimativas populacionais, a partir de variáveis sintomáticas da mudança populacional. Para tanto, selecionou sete fontes de dados: matrícula escolar; nascimentos e mortes; dados de construção civil; registros de seguro social; registros de atendimento de centros de saúde; e número de eleitores. O número de eleitores e os nascimentos foram as variáveis que apresentaram correlações mais elevadas com a população, principalmente a partir de técnicas que utilizam regressão linear, sendo que o método de Correlação de Razões e Correlação de Taxa apresentaram as estimativas mais precisas. Contudo, como era esperado, quanto menor o tamanho populacional (e a área em estudo), menor era a precisão das estimativas.

Cabrera (2011) utiliza dados dos Censos de 1996 e de 2004 do Uruguai como referência para a análise da qualidade de estimativas produzidas para departamentos, a partir dos métodos Pro Rata, Distribuição Proporcional, Taxas Vitais, Razão Censitária e Correlação de Diferença. Para tanto, utiliza dados vitais, matrícula escolar, dados de consulta hospitalar, de tarifa de energia elétrica e dados de seguridade social. Para cada modelo foram criadas matrizes de correlação, que mostram alta correlação das variáveis independentes com a variável dependente. Os métodos de Razão Censitária e de Diferença de Taxas apresentaram as melhores estimativas, com erro médio relativo de, respectivamente, 2,5% e 2,4%.

No Brasil, Brito, Cavenaghi e Jannuzzi (2010) compararam estimativas para municípios do Estado do Rio de Janeiro, a partir da aplicação dos métodos AiBi, Relação de Coortes e Correlação de Razões, tendo como base o Censo 2000 e a Contagem de 2007. A projeção realizada pelo método AiBi apresentou maior tendência a ficar subestimada, enquanto o método de Correlação de Razões mostrou os melhores resultados, em parte por incorporar informações novas e mais

recentes. Em relação à comparação com a Contagem de 2007, observou-se que as estimativas (a partir dos três métodos aplicados) ficaram com forte tendência à superestimação, o que indicou a subestimação da Contagem Populacional de 2007.

Jannuzzi (2007) utiliza dados de registros administrativos (histórico de nascimentos e óbitos), a partir de um sistema de equações diferenciais e combinado com opiniões de especialistas sobre as tendências demográficas, para a projeção em áreas intramunicipais (no caso distritos) do município de São Paulo, para o período entre 2005 e 2010. O modelo, denominado de ProjPeq, utiliza indicadores relacionados à atratividade residencial e à morfologia urbana (ao uso do solo), tais como custos de moradia, disponibilidade de terrenos, índices de verticalização, proximidade de locais de maior oferta de empregos, restrições ambientais (como áreas de preservação), características dos sistemas de transportes, dentre outros, para a produção de estimativas.

O modelo é composto, portanto, de dois coeficientes principais, sendo o primeiro coeficiente relacionado ao crescimento vegetativo da população e o segundo à atratividade migratória (ou residencial). O modelo demanda, portanto, de um conjunto grande de informações, com séries históricas disponíveis, e o modelo utiliza de algoritmos recursivos, que requer um valor inicial para os coeficientes de atratividade migratória ou residencial, “que podem ser estimados a partir do comportamento demográfico na década anterior e da relação dos mesmos com os fatores físico-territoriais mencionados” (Jannuzzi, 2007). Ao aplicar para os distritos de São Paulo, observou-se que o modelo ProjPeq apresentou melhores resultados, quando comparado com projeções oficiais, elaboradas pela prefeitura do município. A projeção da prefeitura foi 4,3% superior aos resultados do ProjPeq, enquanto que a projeção da Fundação Seade foi superior em 0,5%, e a projeção por AiBi foi praticamente igual, com diferença mínima em relação ao ProjPeq.

Mais recentemente, Silva e Santos (2016) analisaram um conjunto de métodos utilizados para a projeção de pequenas áreas, a partir de municípios do Estado do Paraná, no período de 2000 a 2010. Para tanto, utilizaram os censos de 2000 e 2010, a Contagem de 2007, além de um conjunto de variáveis sintomáticas, como matrículas do censo escolar, número de veículos registrados nos municípios, número de eleitores cadastrados, emprego formal nos municípios, nascidos vivos

e óbitos registrados nos municípios. Ao comparar os resultados do método AiBi com métodos baseados em variáveis sintomáticas, os autores concluíram que os últimos apresentaram estimativas muito mais próximas da população efetivamente enumerada no Censo 2010. Tendo como exemplo a raiz quadrática média do erro relativo, a correlação de taxas (3,9%), a correlação de razões (4%) e a correlação de diferenças (4,2%) apresentaram resultados superiores ao método AiBi (10,7%).

Os tópicos do Capítulo 3 até aqui permitiram analisar uma grande quantidade de métodos demográficos que podem ser utilizados para a produção de estimativas para pequenas áreas, desde métodos de extrapolação simples, passando por métodos de alocação para decomposição da população de áreas maiores para áreas menores, a métodos que utilizam de modelos de regressão, cujas variáveis independentes são sintomáticas da mudança populacional.

O método de relação de coortes apresenta vantagens, como a possibilidade de projeção da população por sexo e idade, além da menor quantidade de informação necessária para a produção de estimativas, se comparado com métodos que utilizam de variáveis sintomáticas (como os de Correlação de Razões, por exemplo). Por outro lado, Jardim (2003) reconhece que as estimativas baseadas em variáveis sintomáticas se baseiam em informações mais recentes, se comparado às projeções. A análise dos métodos até aqui apresentados permite afirmar, por exemplo, que as estimativas baseadas em relação de coortes são limitadas aos erros da projeção populacional das áreas maiores, dentre os quais o fato de representar tão somente a síntese da avaliação prospectiva (ou da extrapolação) das componentes demográficas. Métodos que utilizam variáveis sintomáticas se baseiam em mudanças recentes e observadas, o que, se fortemente correlacionada à mudança populacional, tem a capacidade de identificar tendências não captadas pelos métodos de extrapolação, especialmente em pequenas áreas, marcadas por rupturas nas tendências observadas do crescimento populacional. Soma-se ainda ao bom nível de ajuste para intervalos temporais mais curtos, como é o caso dos períodos intercensitários.

O próximo item apresenta brevemente as projeções e estimativas oficiais disponíveis no Brasil. Estes são elementos importantes, pois, como vem sendo colocado ao longo dos capítulos, a avaliação da eficácia das estimativas por

imagens orbitais de média resolução espacial será medida a partir da análise das informações censitárias, bem como das projeções e estimativas oficiais. Com o objetivo de fundamentar essa análise, os próximos itens (2.4 a 2.6) discutem os aspectos mais gerais das últimas duas projeções (revisões 2008 e 2013) e das estimativas pós-censitárias, do IBGE.

2.4 Projeções Oficiais do IBGE (Revisões 2008 e 2013)

A projeção do IBGE (Revisão 2008) será utilizada neste trabalho para a comparação das projeções para o ano 2010 com as estimativas calculadas via sensoriamento remoto, para o mesmo ano. Diferentemente da Revisão de 2013, a Revisão de 2008 não contou com os dados do Censo 2010, de modo que a população estimada para este ano corresponde a uma informação projetada e não corrigida, o que torna a comparação importante na compreensão da eficácia dos métodos de projeção convencionais, em comparação àqueles que utilizam de variáveis sintomáticas.

Porém, a Revisão 2008 restringe-se ao País e Unidades da Federação, e no ano de 2010 não são calculadas as estimativas municipais pelo IBGE, por ser um ano censitário. Nesse contexto, a projeção para o ano de 2010 do Estado de Minas Gerais será desagregada para o nível municipal, através da técnica AiBi. Essa comparação será uma das formas de avaliação das estimativas calculadas por imagens orbitais de média resolução espacial.

Já a Revisão de 2013 será utilizada tanto para a avaliação das estimativas via sensoriamento remoto, para o ano de 2010, como para a avaliação das estimativas via sensoriamento remoto calculadas para o ano de 2015. Para o ano de 2010 – que corresponde ao primeiro ano projetado da Revisão 2013, e que implica numa análise diferente, se comparado à projeção da Revisão 2008 – as estimativas de Minas Gerais serão desagregadas para os municípios selecionados (através da técnica AiBi). Para o ano de 2015, serão consideradas as estimativas pós-censitárias oficiais, dado que o próprio IBGE gera as estimativas municipais no período pós-censo, a partir da projeção das Unidades da Federação (obtida na Revisão 2013). Nesse contexto, faz-se importante compreender as principais

características dessas duas revisões de projeção do IBGE, o que permitirá qualificar a análise da avaliação das estimativas produzidas via sensoriamento remoto.

A projeção do IBGE (Revisão 2008) é gerada pelo método das Componentes, para o Brasil como um todo, para o período entre 1980 e 2050, e para as Unidades da Federação, até 2030. Já a projeção do IBGE (Revisão 2013) é feita pelo método das Componentes, para a população brasileira, de 2000 até 2060, bem como para as Unidades da Federação, entre 2000 e 2030. Além do ano base da Revisão 2013 ser o ano 2000 (enquanto a Revisão 2008 adota 1980 como o ano de partida), a grande diferença entre as duas projeções é o método pelo qual as Unidades da Federação são projetadas. Enquanto a Revisão 2008 apenas desagrega a populacional nacional projetada para o nível dos estados através do método AiBi, a Revisão 2013 projeta efetivamente as Unidades da Federação, através do método das Componentes (IBGE, 2008; 2013).

Na Revisão de 2013, adotou-se um procedimento, denominado “Conciliação Censitária”, que consistiu na avaliação das populações dos Censos de 1991, 2000 e 2010, através de análises demográficas, com o objetivo de “ajustar a estrutura por sexo e idade da população enumerada no Censo Demográfico 2000 para o Brasil” (IBGE, 2013). Ainda segundo o IBGE (2013), com esse procedimento, espera-se obter coerência entre as informações dos censos e os eventos demográficos captados por outras fontes (especialmente registro civil), “de tal modo que se cumpra o explicitado na equação do balanço demográfico” (IBGE, 2013). Os problemas referem-se a erros de declaração da idade e diferenças de cobertura entre os censos (especialmente a sub enumeração de crianças). Segundo o IBGE (2013), para as Unidades da Federação foi utilizada a estrutura de cobertura fornecida pela Pesquisa de Avaliação do Censo 2000 (Oliveira *et al.*, 2003).

Segundo o IBGE (2013), as populações de 1990 e 2010 foram projetadas e retroprojetadas para o ano 2000, a partir de estimativas de fecundidade e mortalidade, com o objetivo de compará-las com a população observada, do Censo 2000. A partir dessas análises foram definidas a estrutura ajustada, por sexo e idade, da população de 2000 para o Brasil, no qual se destacam a sub enumeração das crianças entre 0 e 9 anos, dos adultos jovens entre 25 e 29 anos (especialmente

homens), a sobreenumeração de idosos, em decorrência dos erros de declaração de idade e a sub enumeração dos homens (IBGE, 2013).

A população entre 0 e 4 anos foi estimada a partir de uma média entre as estimativas de nascimentos (entre 1995 e 2000) e a população entre 10 a 14 anos retroprojetada (IBGE, 2013). A população entre 5 e 9 anos também foi calculada através de média, mas apenas para as mulheres, dado que a razão de sexo em 2010 para o grupo entre 15 e 19 anos já apresenta sub enumeração (de homens maior que a de mulheres) [IBGE, 2013].

A conciliação demográfica realizada na Revisão de 2013 resultou no ajuste de 2,1% da população total do Brasil em 2000. Esse fator foi utilizado como parâmetro para o ajuste das populações de partida das Unidades da Federação, ponderado pelo padrão de omissão censitária calculado para cada estado, pela Pesquisa de Avaliação do Censo 2000 (IBGE, 2013).

Já na Revisão de 2008 foram empregados métodos similares à Revisão de 2013, como é o caso dos procedimentos de conciliação para a população base (a partir de dados dos Censos de 1970 a 2000, além da Contagem 2007) e o método das Componentes para a projeção da população brasileira. Porém, o IBGE definiu que a população de partida seria a população do Censo 1980, e não a população corrigida pela conciliação demográfica (que reajustaria a população do Censo 1980 em 1,8%), em função da coerência entre os censos de 1970 a 2000. Ademais, as projeções por Unidades da Federação não foram calculadas pelo método das Componentes, mas estimadas pelo método AiBi, a partir da desagregação da projeção do País (e da mesma forma os municípios).

Ainda sobre a Revisão de 2008, vale mencionar que a fecundidade foi estimada a partir de um ajuste logístico, que levou em consideração dados dos Censos 1991 e 2000, e das PNAD's de 2002 e 2006, o que resultou numa TFT limite de 1,5 filhos por mulher. Interessante observar que, embora a PNAD de 2007 tenha apresentado uma TFT de 1,95 por mulher, a TFT projetada para o ano de 2010 foi de 1,76 para o Brasil, nível este adequado às discussões recentes sobre fecundidade do Brasil para aquele ano.

No caso da fecundidade da projeção Revisão 2013 do IBGE, foram utilizados para a projeção dos níveis as TFT's estimadas através do método da Razão P/F de Brass, a partir da aplicação de dados dos censos de 2000 e 2010. A projeção da fecundidade levou em consideração a redução acelerada dos níveis de fecundidade, observado no País a partir da década de 1970. Enquanto na década de 1960 a fecundidade girava em torno de 6,2 filhos por mulher, em 2013, a TFT girava em torno de 1,7, já com sinais de envelhecimento do padrão etário da curva de fecundidade (IBGE, 2013)⁹.

2.5 Estimativas Pós-Censitárias Oficiais do IBGE

A metodologia adotada para a produção das estimativas intercensitárias do IBGE é realizada a partir do método AiBi (Madeira e Simões, 1972), detalhado no item 2.1.1 deste trabalho. Vale ressaltar que esse método se baseia na tendência de crescimento populacional observado dos municípios, entre dois censos consecutivos, em relação à mesma tendência de uma área geográfica maior (no caso, os estados). O método utiliza a projeção populacional do IBGE para as Unidades da Federação – elaborada, por sua vez, a partir do método das componentes demográficas – e utiliza de métodos matemáticos que relacionam a dinâmica demográfica da área maior aos municípios. Assim, “se a tendência de crescimento populacional do município entre os censos for positiva, a estimativa populacional será maior que a verificada no último levantamento censitário” (IBGE, 2002), caso contrário, a estimativa apresentará valor inferior ao último censo.

Para os municípios com população inferior a 100 mil habitantes, obtém-se, primeiramente, para cada estado, os quartis das populações segundo o tamanho dos municípios (no ano censitário), e os quartis das taxas médias geométricas anuais de crescimento, observadas no período intercensitário.

A partir do cruzamento dos quartis de população e taxa de crescimento, e separando os municípios com taxas de crescimento positivas daqueles com taxas

⁹ Para a discussão sobre sub-cobertura do Censo 2010 e sobreestimativa dos níveis de fecundidade da projeção Revisão 2013 do IBGE, ver UNESCO (2017).

negativas, formam-se grupos de municípios dentro de cada Unidade da Federação, que apresentam tamanho de população e taxas de crescimento similares (IBGE, 2002).

Em seguida, estimam-se as populações residentes totais desses grupos, em 1º de julho do ano t , considerando-se como área maior o estado (a exceção dos municípios já estimados, acima de 100 mil habitantes), e como áreas menores os grupos formados. De posse da proporção (ou do peso) de cada município dentro do respectivo grupo, em relação à população do ano censitário, aplica-se a mesma proporção ao total estimado para o seu grupo em 1º de julho do ano t . São obtidas, assim, as populações residentes estimadas para o mesmo ano t (data acima), para os municípios com população inferior a 100 mil habitantes.

Discutidas as principais características dos censos, projeções e estimativas do IBGE, o próximo item apresenta os principais métodos existentes para a avaliação dos erros das estimativas e projeções, dos quais alguns serão utilizados no decorrer deste trabalho.

2.6 Avaliação dos Erros das Estimativas

Na literatura sobre estimativas, medidas de acurácia são normalmente utilizadas para a avaliação do nível de acerto das projeções, quando estas são utilizadas como previsão do comportamento futuro da população. Contudo, embora menos recorrentes, as medidas de erro também podem ser empregadas para analisar se um dado modelo foi bem ajustado aos dados observados, durante um horizonte de projeção (Smith *et al.*, 2002; Bryan, 2004).

Smith *et al.* (2002) definem como medida de erro a diferença entre a população projetada (F), de uma dada área e um ano t , e a população real (A) da mesma área e ano. Os erros de projeção podem ser expressos simplesmente como diferença absoluta, ou ainda percentual, quando o objetivo, por exemplo, é avaliar comparativamente os erros entre diferentes regiões. A equação abaixo apresenta o Erro Percentual.

$$PE_t = [(F_t - A_t) / A_t] \times 100 \quad (2.32)$$

O Erro Médio, por sua vez, leva em consideração a direção do erro, sendo os erros positivos e negativos compensados (onde E corresponde ao erro no ano t).

$$\text{Erro Médio (EM)} = \sum \frac{E_t}{n} \quad (2.33)$$

O Erro Médio Absoluto ignora a direção do erro, de modo que os erros positivos e negativos não se anulam. Segundo Smith *et al.* (2002), essa medida mostra a média da diferença entre a projeção e a população observada, independentemente se a projeção foi muito alta ou muito baixa.

$$\text{Erro Médio Absoluto (EMA)} = \sum \frac{|E_t|}{n} \quad (2.34)$$

As medidas acima são baseadas em diferenças numéricas. As próximas medidas levam em consideração o erro percentual entre a estimativa e o dado observado.

A Média Algébrica do Erro Percentual (conhecida como Média do Erro Percentual, ou pela sigla em inglês, *MALPE*), é uma medida na qual os valores positivos e negativos podem se anular (em função da ausência de módulo, no somatório dos erros percentuais). A *MALPE* positiva reflete uma tendência para projeções muito altas, enquanto que a negativa, reflete projeções muito baixas (Smith *et al.*, 2002; Bryan, 2004).

$$\text{Média Algébrica do Erro Percentual (MALPE)} = \frac{\sum PE_t}{n} \quad (2.35)$$

Muito utilizada para a avaliação da acurácia das projeções, a Média Absoluta do Erro Percentual (*MAPE*) é uma medida na qual os valores positivos e negativos não se anulam. A *MAPE* mostra a média da diferença percentual entre as estimativas e a população observada, independentemente de a projeção ter sido muito alta ou muito baixa (Smith *et al.*, 2002; Bryan, 2004).

$$\text{Média Absoluta do Erro Percentual (MAPE)} = \frac{\sum |PE_t|}{n} \quad (2.36)$$

Segundo Smith *et al.* (2002) e Bryan (2004), em alguns casos é importante utilizar medidas de erro que forneçam um peso maior para erros maiores, em comparação a erros de menor escala, pois, erros maiores tendem a comprometer desproporcionalmente a acurácia das estimativas. O Erro Quadrático Médio (EQM, equação 2.37) e a Raiz Quadrática Média dos Erros (RQME, equação 2.38) tratam do erro absoluto, enquanto a Raiz Quadrática Média dos Erros Relativos (RQMÉR, equação 2.39), muito utilizado na literatura, considera os erros percentuais (Smith *et al.*, 2002).

$$\text{Erro Quadrático Médio (EQM)} = \frac{\sum (E_t)^2}{n} \quad (2.37)$$

$$\text{Raiz Quadrática Média dos Erros (RQME)} = \sqrt{\frac{\sum (E_t)^2}{n}} \quad (2.38)$$

$$\text{Raiz Quadrática Média dos Erros Relativos (RQMÉR)} = \sqrt{\frac{\sum (PE_t)^2}{n}} \quad (2.39)$$

Segundo Smith *et al.* (2002), existem ainda outras medidas de erro que focam na mediana da distribuição ou no erro percentil, muito úteis para a construção de intervalos de predição (Smith *et al.*, 2002). De toda forma, as medidas apresentadas anteriormente são as mais utilizadas na literatura e servem de parâmetro para a definição do método de avaliação de erros do presente trabalho.

Segundo Smith *et al.* (2002), erros entre 8% e 14% são aceitáveis para um horizonte de 10 anos, enquanto que, para um horizonte de projeção entre 10 e 20 anos, considera-se como aceitáveis erros de projeção de até 16%. Por sua vez, Wilson e Rowe (2011) consideram 5% de erros nas estimativas como de baixa magnitude, 10% como erros aceitáveis, enquanto que acima de 20% considera-se que as estimativas apresentam graves problemas.

Este tópico encerra a revisão sobre os modelos demográficos, e suas diversas aplicações, no campo das estimativas populacionais. O próximo capítulo apresenta e discute um conjunto de metodologias muito promissoras para o campo das

estimativas para pequenas áreas, que tem sido desenvolvido nos últimos anos, a partir da utilização de imagens orbitais e de métodos e técnicas de sensoriamento remoto.

3 ESTIMATIVAS POPULACIONAIS PARA PEQUENAS ÁREAS E SENSORIAMENTO REMOTO: CONCEITOS E APLICAÇÕES

Este capítulo tem como objetivo principal apresentar uma revisão sobre os principais métodos de estimativas populacionais para pequenas áreas, a partir da utilização de sensoriamento remoto. No entanto, faz-se importante primeiramente introduzir os principais conceitos de sensoriamento remoto orbital e as possibilidades do processamento digital das imagens geradas por satélites de observação da Terra, para a compreensão das discussões que serão empreendidas na sequência do capítulo (especialmente para um público não especializado na área do sensoriamento remoto). A partir dessa introdução, o capítulo apresenta os principais métodos de distribuição espacial da população, com foco nos métodos dasimétricos e sua relação com os modelos de estimativas populacionais via sensoriamento remoto. Ainda neste capítulo, no transcorrer da revisão sobre os principais modelos que utilizam imagens orbitais para a produção de estimativas, são apresentados e discutidos os modelos adotados para a construção da proposta metodológica e dos resultados deste trabalho.

3.1 Introdução aos Conceitos Básicos de Sensoriamento Remoto

Segundo Jensen (2009), sensoriamento remoto pode ser definido como o processo de capturar informação sobre algum elemento sem estar em contato físico com ele. Os satélites carregam consigo sensores que observam os elementos da superfície da terra, podendo ser sensores ativos (que emitem radiação para captação das informações do alvo) ou passivos (que captam a radiação eletromagnética, refletida pelos alvos de superfície). Os sensores auxiliam na predição da temperatura, no acompanhamento de culturas agrícolas, no planejamento urbano, além de predizer fenômenos, tais como terremotos, inundações, furacões, entre outras aplicações (Jensen, 2009).

As técnicas de sensoriamento remoto possibilitam a análise da superfície terrestre nas diversas regiões do espectro eletromagnético, sendo que os processos podem ser divididos em três grandes etapas: pré-processamento de imagens (que inclui a restauração da imagem para correção de degradações geométricas e radiométricas do sensor, no processo de formação das imagens) [INPE, 2001]; o processamento de imagens, que consiste na extração de informação das imagens de sensores remotos (através de técnicas de reconhecimento de padrões, como da segmentação e dos métodos de classificação); e a etapa de pós-processamento, que consiste na análise da qualidade das informações extraídas (INPE, 2001; Jensen, 2009; Meneses e Almeida, 2012).

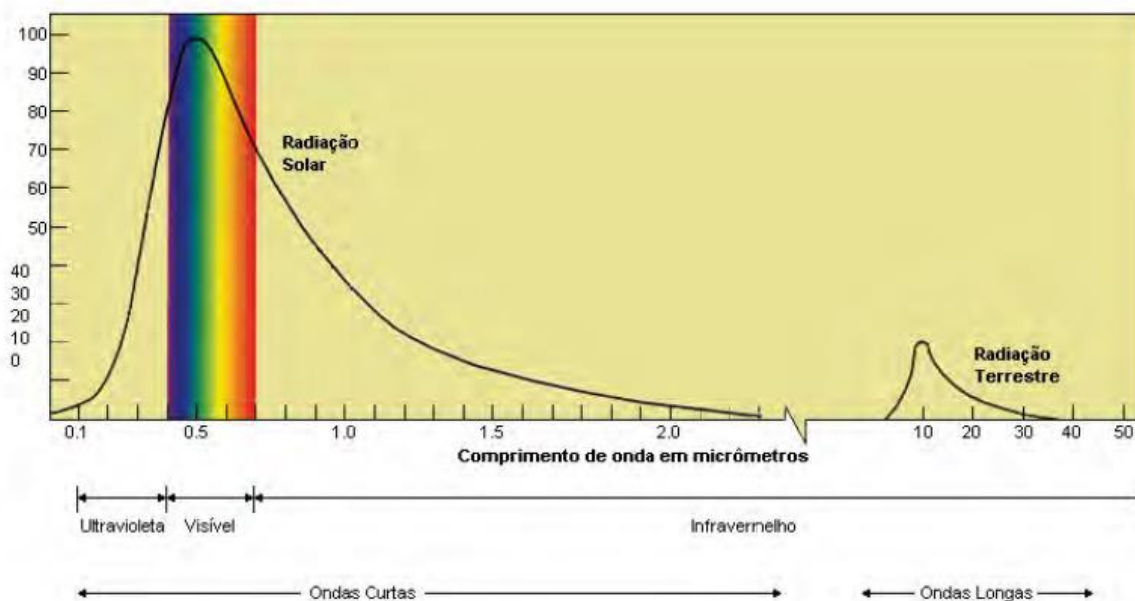
Os itens 3.1.1 e 3.1.2 introduzem alguns conceitos básicos de sensoriamento remoto, necessários para a compreensão dos resultados que serão apresentados neste trabalho.

3.1.1 O Espectro Eletromagnético e as Resoluções em Sensoriamento Remoto

São diversas as formas de radiação eletromagnética, como a luz, ondas de rádio, micro-ondas, radiação infravermelha, raios ultravioleta, raios-X e raios gama (Jensen, 2009). Essas formas, também conhecidas como espectro eletromagnético, são similares por se moverem à velocidade da luz, embora sejam diferentes em relação ao comprimento de onda; o que, por sua vez, está relacionado com a quantidade de energia que a onda eletromagnética carrega consigo (quanto menor o comprimento de onda, maior a energia) [Gonzales e Woods, 2000; INPE, 2001; Jensen, 2009; Meneses e Almeida, 2012].

A Figura 1 apresenta a distribuição da intensidade da energia, emitida pelas fontes naturais de radiação eletromagnética, medidas no topo da atmosfera (imagem extraída de Meneses e Almeida, 2012). Na porção direita do gráfico, (referente às ondas longas) estão localizadas ondas de rádio, enquanto que, na porção esquerda situam-se as ondas curtas, que são milhões de vezes menores do que as ondas da luz visível (INPE, 2001, Jensen, 2009, Meneses e Almeida, 2012).

FIGURA 1 - Espectro Eletromagnético: distribuição da intensidade da energia



Fonte: Reproduzido de Meneses e Almeida, 2012.

A maioria dos sensores remotos captam intervalos do espectro eletromagnético correspondentes ao visível e ao infravermelho próximo. O espectro visível está localizado entre 400 e 700 nanômetros¹⁰, enquanto que o infravermelho se estende da região visível à aproximadamente um milímetro, e inclui a radiação termal.

Os objetos ou materiais possuem diferentes características espectrais, ou comportamentos espectrais, definidos a partir de valores de reflectância específicos, em função da variação de suas propriedades químicas e físicas. A Figura 1 apresenta o esquema de captação da radiação, onde se observa que apenas uma fração da energia eletromagnética (eixo x do gráfico) é captada pelo olho humano. Para uma melhor compreensão da Figura 1, devem ser descritos ao menos três conceitos básicos de sensoriamento remoto: irradiância, radiância e reflectância.

Jensen (2009) define irradiância como “a quantidade de fluxo radiante incidente sobre uma superfície por unidade de área de superfície”. A radiação

¹⁰ Nanômetro corresponde a uma unidade da faixa do espectro eletromagnético.

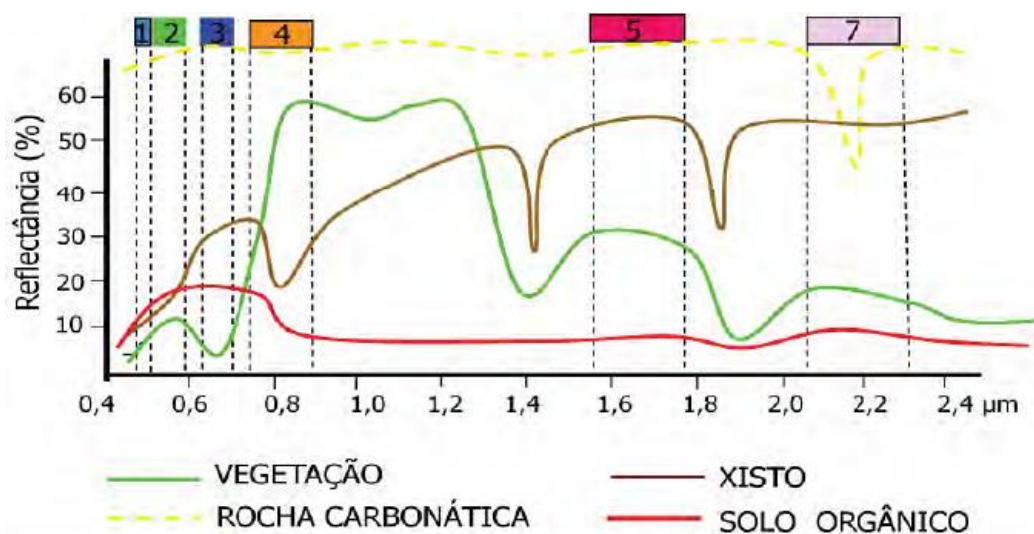
eletromagnética do sol propaga-se pelo espaço e interage com superfícies (ou seja, a irradiância), podendo ser refletida, absorvida ou reemitida. A porção absorvida é normalmente transformada em calor, enquanto que a parte refletida espalha-se pelo espaço.

O fator que mede a capacidade de um objeto de refletir a energia radiante é a reflectância. A reflectância é a proporção entre o fluxo que é refletido e a radiação eletromagnética incidente numa superfície. De forma similar, Jensen (2009) define reflectância como a “razão adimensional entre o fluxo radiante refletido por uma superfície e o fluxo radiante incidente nela”. O fluxo radiante refletido é a *radiância*, enquanto que o fluxo radiante incidente é a irradiância, e a reflectância é dada pela razão entre a radiância e a irradiância.

Souza (2012) afirma que através de medidas de reflectância de superfície é possível construir gráficos sobre o comportamento padrão dos principais objetos da superfície terrestres, ao longo do espectro eletromagnético. Souza (2012) afirma ainda que, ao conhecer o comportamento espectral das coberturas terrestres, o pesquisador pode extrair informações de interesse da imagem, a partir de diferentes procedimentos metodológicos, como será discutido adiante.

A Figura 2 (extraída de Meneses e Almeida, 2012) mostra o espectro eletromagnético (no eixo X) e os níveis de reflectância (no eixo Y), para diferentes usos do solo. Os números (1, 2, 3, 4, 5 e 7) exemplificam a faixa espectral de cada banda do satélite LANDSAT 7 ETM+ (satélite que fornece os dados de sensoriamento remoto utilizados neste trabalho).

FIGURA 2 – Espectros de reflectância e as posições das bandas do sensor LANDSAT 7



Fonte: Meneses e Almeida, 2012.

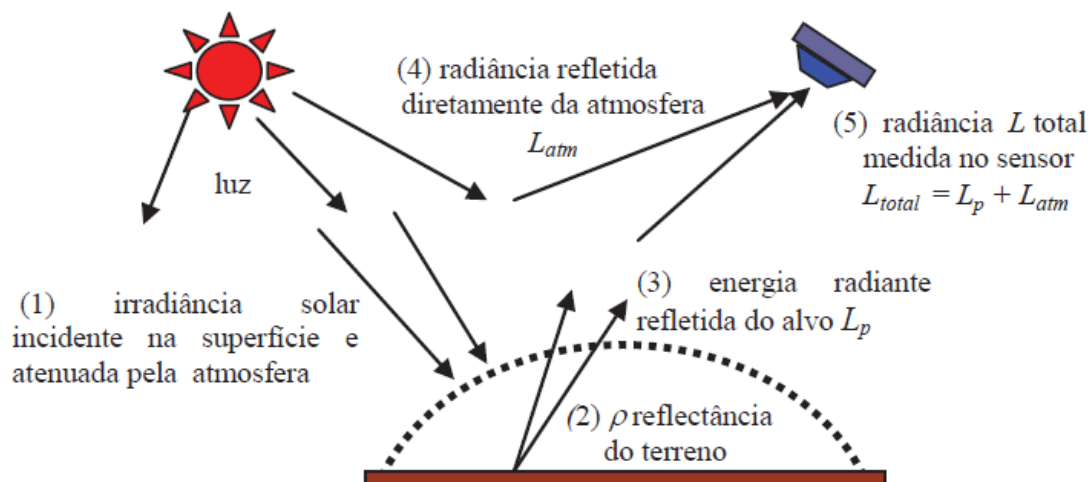
Contudo, os sensores remotos não captam diretamente a reflectância. O fluxo radiante refletido (ou radiância) em diferentes comprimentos de onda e numa determinada direção é que são captados pelos instrumentos imageadores (a bordo dos satélites).

Jensen (2009) define radiância como sendo a “intensidade radiante por unidade de área-fonte projetada numa direção específica”. Em síntese, a radiância que importa para o sensoriamento remoto é aquela projetada em direção ao sensor remoto, pois a mesma é convertida em sinal na forma perceptível ao sistema visual humano (SVH), gerando as imagens (INPE, 2001). Segundo Novo e Ponzoni (2001):

“A composição espectral do fluxo que deixa a superfície sofre alterações que são dependentes das suas características físico-químicas. Assim a radiância medida por um sensor pode ser determinada para um intervalo específico de comprimentos de onda (região ou banda espectral)” (Novo e Ponzoni, 2001).

A quantificação da radiância de uma cena pode ser representada por níveis de cinza (discretos na imagem digital), sendo dado pelo número de bits, por pixel. Imagens com 8 bits, por exemplo, equivale à 256 níveis digitais.

FIGURA 3 – Espalhamento atmosférico e esquema de captação da radiancia por sensores de satélite



Fonte: Meneses e Almeida, 2012.

O ângulo definido pelo detetor (também conhecido por IFOV, *Instantaneous Field of View* ou campo de visada) define a área no solo, representado por um elemento na imagem digital (um pixel), no qual é captada a média da radiancia, expressa por valores correspondentes a níveis de cinza. Essa área no solo é que define a resolução espacial, e, quanto melhor a resolução maior o nível de detalhamento da imagem. Portanto, a resolução espacial está associada à capacidade de discriminação dos elementos da cena, expresso nas diferenças de tamanho dos pixels das imagens (INPE, 2001; Jensen, 2009; Meneses e Almeida, 2012).

A resolução espectral, por sua vez, está relacionada à largura das faixas espectrais de cada banda e o número de bandas do satélite, sendo que “faixas mais largas conferem uma resolução espectral menor ao sensor” (Novo e Ponzoni, 2001). Já a resolução temporal é definida pelo “intervalo de tempo medido entre uma aquisição e outra de imagens” (INPE, 2001; Jensen, 2009), sendo importante na identificação de processos dinâmicos, como mudanças no uso da terra, crescimento de culturas, crescimento urbano, etc.

A resolução radiométrica é definido pelo número de bits que representa um valor digital (ou nível de cinza). Cada pixel de uma imagem possui um nível de cinza, que “representa a intensidade da energia eletromagnética medida pelo sensor, para a

área da superfície da terra correspondente ao tamanho do pixel”, relacionado sempre à “média da intensidade de energia refletida ou emitida pelos diferentes materiais presentes neste pixel” (INPE, 2001). O número de níveis de cinza é dado, conforme equação abaixo.

$$K = 2^m \quad (3.1)$$

Onde: m é o número de bits.

Assim, uma quantização de 8 bits representa 256 níveis de cinza e representa melhor os contornos e os detalhes numa imagem, se comparado com uma representação de 4 bits (com 16 níveis de cinza). Imagens com baixa quantização tendem a apresentar transições bruscas de tons de cinza em regiões com variações suaves (INPE, 2001; Jensen, 2009).

Sobre as correções radiométricas é importante afirmar que, os sensores, responsáveis pela geração das imagens orbitais, em algumas situações apresentam imperfeições que resultam em ruídos, como é o caso das imagens LANDSAT 7 ETM+ (utilizadas neste trabalho), que passaram a apresentar perda de 22% dos dados a partir de maio de 2003 (com ruídos sistemáticos em formato de linhas regulares), em função de um problema no corretor SLC (*Scan Line Corrector*). Nesses casos, são utilizados algoritmos específicos para a remoção dos ruídos, conhecidos como restauração das imagens. A restauração (ou correção radiométrica) corrige as distorções introduzidas pelos sensores, a partir da aplicação de filtros (conhecidos como passa-baixas), que interpolam pixels que apresentam ruptura nos níveis de cinza – por exemplo, em pixels com ausência de informação – resultando na suavização das transições dos níveis de cinza das imagens (JENSEN, 2009, INPE, 2001).

Segundo manual do INPE (2001), o método mais utilizado para a substituição de pixels degradados é através da interpolação dos valores médios dos pixels correspondentes às linhas anterior e posterior. Outra técnica consiste em adotar a média local dessas linhas, e, caso a diferença dessas médias seja menor que um limiar, os pontos com ruído têm seus valores adicionados dessa diferença. Para este trabalho será utilizado o método de interpolação dos valores médios dos pixels

para a correção dos ruídos das imagens LANDSAT 7 ETM+, para os anos de 2010 e 2015.

Já correção geométrica está relacionada à existência de distorções sistemáticas, introduzidas durante a fase de aquisição das imagens. Esse tipo de correção é normalmente utilizado para registrar a posição dos dados da imagem no espaço, para a integração com outras informações, inclusive de períodos distintos, como é o caso da análise de séries temporais. O registro normalmente é realizado a partir de transformações polinomiais de 1º ou de 2º graus, e fazem o vínculo entre as coordenadas da imagem e as coordenadas do sistema de referência, através de pontos de controle (que são feições passíveis de identificação na imagem e no terreno, e cujas coordenadas são conhecidas) [Jensen, 2009; Meneses e Almeida, 2012].

A correção atmosférica, por sua vez, é um dos procedimentos mais importantes na etapa de pré-processamento. Segundo Jensen (2009), Meneses e Almeida (2012) e INPE (2001), a absorção atmosférica reduz os valores de energia do alvo, enquanto o espalhamento atmosférico adiciona valores à resposta original. A considerar que o espalhamento se refere ao efeito de maior proporção, são utilizados diferentes métodos para a correção, como a simulação de propriedades ópticas da atmosfera, ou ainda a utilização de alvos da superfície que tenham a reflectância conhecida, como é o caso de alvos escuros (que, na ausência de espalhamento atmosférico, apresentariam nível digital zero). Nesses casos, a radiância observada nesses alvos é utilizada como fator de subtração no restante da imagem, para a redução da contribuição do efeito atmosférico.

Para este trabalho não foram utilizados métodos de correção atmosférica para a aquisição dos dados de reflectância, pois as informações de reflectância – que constituem nas variáveis independentes dos modelos utilizados para a produção das estimativas – foram obtidas do projeto *Science Processing Architecture* – ESPA, da United States Geological Survey – USGS, que já fornece imagens de satélite com dados de reflectância de superfície.

Para a definição (classificação) das áreas ocupadas, foram selecionadas imagens originais LANDSAT 7 ETM+ da USGS, que já contam com a correção geométrica.

Contudo, ressalta-se que não foram utilizados procedimentos de correção atmosférica. A escolha das imagens LANDSAT 7 ETM+ teve como critério a proximidade com o período de referência dos censos de 2000 e 2010 (1º de agosto), e o baixo percentual de nuvens, com o objetivo de contornar possíveis interferências na identificação das manchas urbanas, o que poderia dificultar o processo de classificação das áreas ocupadas.

3.1.2 Cor, Comportamento Espectral de Alvos Urbanos e Classificação

As variações de cores são muito sensíveis ao olho humano, se comparado às variações de níveis de cinza, sendo que a região do espectro visível varia entre 380 e 710 nanômetros. Segundo Meneses e Almeida (2012), uma tonalidade de cor pode ser representada através de 3 coeficientes de cores, que correspondem a eixos tridimensionais, sendo que cada eixo representa uma cor do sistema RGB.

O sistema RGB trata-se de um sistema de cores primárias, que permite a representação quantitativa de qualquer cor, a partir de um grupo de 3 números (ou coeficientes). Esses 3 coeficientes indicam a proporção das cores primárias – vermelho (R), verde (G) e azul (B) – conjugadas para a produção de qualquer tonalidade de cor. A partir de uma imagem multiespectral, define-se uma composição de 3 bandas, onde cada pixel é representado numericamente por três níveis de cinza (correspondentes a cada banda). Cada ponto na imagem colorida corresponde a uma cor, que é dada pelos níveis de cinza das três bandas selecionadas, e associadas às cores RGB (INPE, 2001).

A composição que representa as cores verdadeiras para as imagens LANDSAT 7 é a R3G2B1 (banda 3 em vermelho, banda 2 em verde e banda 1 em azul), muito embora Jensen (2009) e Meneses e Almeida (2012) afirmem que as composições por falsa cor podem ser muito úteis para a visualização de informações espectrais, como formações florestais, queimadas, água, temperatura, entre outros.

A composição R3G4B5 dos sensores Landsat, utilizada na tese para a interpretação do uso do solo (para as áreas ocupadas) é a melhor opção para a interpretação de áreas urbanas, pois apresentam boa distinção dessa cobertura em

relação às demais, além da maior variação de tonalidades e texturas da área urbana (Jensen, 2009; Souza, 2012). Essa composição ainda mostra claramente os limites do solo e da água, com a vegetação bem discriminada.

Já a composição R4G5B3 também é uma boa alternativa para a representação das áreas ocupadas, com uma banda no visível e as demais no infravermelho, o que permite uma boa diferenciação da vegetação (com tons avermelhados), sendo que o solo exposto e as áreas urbanizadas aparecem em tons de azul-claro (e a água com tom preto).

Contudo, apesar das diversas composições de bandas que contribuem para a melhor visualização do uso do solo, o contraste de brilho das imagens de satélite é normalmente baixo, sendo comum a utilização de técnicas de realce, que aumentam o contraste através da manipulação de histogramas e de filtragens, para a obtenção de imagens coloridas (ou mesmo de tons de cinza) com bom equilíbrio de cores (de brilho) [Meneses e Almeida, 2012]. Contudo, para este trabalho, não foram utilizadas técnicas de realce, com o objetivo de não interferir nos níveis de cinza originais das imagens LANDSAT 7; pois, os níveis de cinza serão utilizados para a conversão em reflectâncias, que, por sua vez, serão fundamentais para a construção dos modelos de regressão que serão elaborados, para o cálculo das estimativas.

Em relação à extração de informações das imagens, realizadas as correções radiométrica (restauração), geométrica e atmosférica, é possível utilizar de um conjunto de procedimentos e técnicas para a extração de informações das imagens. A primeira técnica é a segmentação, que particiona a imagem em regiões que correspondem a uma dada característica ou aplicação. Essas regiões são formadas por um conjunto de pixels contíguos, que apresentam uniformidade em relação a um dado atributo (INPE, 2001). A divisão da imagem em segmentos pode ser realizada através das técnicas, de crescimento das regiões ou de detecção de bordas (INPE, 2001).

A principal limitação da técnica de detecção de bordas é a existência de buracos nas bordas, ou ainda de variações nos níveis de cinza, o que dificulta a geração dos contornos. Já o método de crescimento de regiões tem a vantagem de sempre

gerar contornos fechados, embora também esteja sujeita a erros de delimitação das áreas homogêneas. Essa técnica consiste num processo iterativo onde as regiões adjacentes são agrupadas a partir de um critério de similaridade, e cujo limiar pode ser definido pelo usuário, “a partir de um teste de hipótese, que testa a média entre as regiões” (INPE, 2001).

A classificação, por sua vez, corresponde ao processo de extração de informação em imagens para o reconhecimento de padrões homogêneos de áreas da superfície terrestre. Os pixels são classificados a partir de critérios e o resultado são mapas temáticos de cores, que representam diferentes usos do solo, como áreas urbanas, florestas, tipos de solo, etc. As técnicas podem ser basicamente subdividas em classificação automática e não automática. Nas classificações não automáticas, o usuário utiliza de conhecimento de interpretação para a construção manual das classes, enquanto que, na classificação automática, o computador processa os atributos numéricos de cada pixel na imagem, a partir de parâmetros estabelecidos pelo usuário (Meneses e Almeida, 2012; Jensen, 2009; INPE, 2001).

Na classificação automática (procedimento este que será utilizado na tese para a definição das manchas de ocupação), podem ser utilizadas mais de três bandas espectrais, e “assumindo-se que as bandas espectrais fornecem uma boa discriminação entre as diversas coberturas (...) é razoável supor que os pixels pertencentes a um mesmo alvo irão agrupar-se formando nuvens no espaço de atributos”. Essas nuvens são denominadas de classes (de uso do solo) [Jensen, 2009; Meneses e Almeida, 2012].

As classificações automáticas podem ser divididas entre supervisionada e não supervisionada. Na classificação não supervisionada, cada pixel da imagem é associado a uma classe sem que o usuário informe quantas e quais classes serão criadas, partindo de técnicas que utilizam de algoritmos de agrupamento (similares às técnicas de segmentação).

Nas classificações supervisionadas, são selecionadas um conjunto de amostras de treinamento para cada uma das classes definidas previamente pelo usuário. Segundo o manual do INPE (2001), assume-se que distribuição das classes é feito segundo uma função de densidade de probabilidade, para as quais os “parâmetros

são estimados através de um conjunto de amostras de treinamento, previamente selecionados”. A função de probabilidade será utilizada para a decisão sobre a que classe um pixel pertencerá (INPE, 2001). O método mais utilizado de classificação é o pixel a pixel, que identifica as regiões a partir de informações exclusivamente espectrais, enquanto que classificações por regiões utilizam de etapas de segmentação e do contorno das regiões, para a associação de cada região a uma classe. Ressalta-se que, para o presente trabalho, será utilizado o método de classificação supervisionada por pixel, denominado de Máxima Verossimilhança (MAXVER).

A classificação por Máxima Verossimilhança corresponde ao método por pixel mais utilizado no sensoriamento remoto. A fim de determinar a qual classe um pixel (com valor espectral x) pertence, avalia-se a probabilidade de que a classe w_i seja a correta para um pixel x .

$$p(w_i|\bar{x}), i = 1, \dots, M \quad (3.2)$$

Onde: M é o número total de classes.

$$\bar{x} \in w_i \text{ se } \Rightarrow p(w_i|\bar{x}) > p(w_j|\bar{x}) \text{ para todo } i \neq j \quad (3.3)$$

A partir das amostras de treinamento é possível estimar a distribuição de probabilidade de cada classe, e cada pixel é atribuído à classe de maior probabilidade (Jensen, 2009). Por exemplo, um limite de 95% significa que 5% dos pixels, os de menor probabilidade, serão ignorados. Esse procedimento tem como objetivo compensar a possibilidade de erros na fase de treinamento, que tendem a aumentar a sobreposição entre as distribuições de probabilidades das classes, ou ainda de pixels que estão no limite entre duas classes. Para que toda a imagem seja classificada, é preciso considerar um limiar de 100%. Normalmente utiliza-se uma matriz de classificação, que apresenta a distribuição de porcentagem dos pixels classificados de forma correta e errônea, sendo que os valores da diagonal principal forem próximos de 100% indica a ausência de confusão entre as classes (Meneses e Almeida, 2012).

Em relação à etapa de classificação em estudos de estimativas populacionais baseadas em imagens Landsat, Langford (2013) comenta:

“In addition to the need for specialized skills, further potential limitations are imposed by the relatively coarse 30-m spatial resolution of Landsat imagery. Also, some degree of error is inherent in any multispectral classification, and the precise outcome of the classification process varies somewhat from one analyst to another because it is based on unique decisions regard training site selection and the generation of spectral signatures” (Langford, 2013).

Além do certo grau de arbitrariedade na seleção das amostras de treinamento, Durand (2007) e Souza (2012) comentam sobre algumas dificuldades encontradas quando se trabalha com imagens de média resolução espacial, e que podem resultar em erros de classificação. O mapeamento de outros usos do solo como sendo área urbana é um erro comum nas classificações, especialmente em imagens de média resolução espacial (como no caso das imagens Landsat), onde o solo apresenta comportamento espectral similar às áreas urbanas. Outra dificuldade trata-se do relevo (bem como das sombras geradas), que em muitos casos dificulta a delimitação das áreas urbanizadas. A própria densidade populacional é apontada por Durand (2007), pois as imagens de média resolução espacial não permitem a distinção com precisão das áreas com intensa densidade, marcadas por intenso processo de verticalização.

Ademais, Durand (2007) e Souza (2004) comentam que a variedade de coberturas existentes nas áreas urbanas (como concreto, asfalto, telhados, vegetação dentro das áreas urbanas, dentre outros) torna complexa a distinção por classificação automática, em função das várias respostas espectrais existentes nas áreas ocupadas. Assim, os autores concluem que a resolução espacial é elemento determinante na etapa de classificação de áreas de ocupação humana. Em imagens de resolução espacial média, as manchas urbanas são facilmente identificáveis “por apresentar altos valores de reflectância na faixa do visível” (Souza, 2012), embora ocorram problemas de interpretação com coberturas de solo, como já fora comentado.

Uma das alternativas para melhorar os procedimentos de classificação automática é a exploração visual das imagens, a partir de uma análise qualitativa (Souza, 2012). As técnicas de interpretação permitem incorporar outras informações para além dos dados espectrais, como a forma, brilho, cor, textura, dentre outros elementos (como o próprio conhecimento do pesquisador), que contribuem para

uma melhor distinção entre áreas urbanas e não urbanas (elemento fundamental nos estudos sobre estimativas populacionais e sensoriamento remoto). Diante à importância da análise qualitativa para uma boa classificação, neste trabalho adotou-se o método híbrido, que consiste na classificação automática, seguido pela interpretação visual das imagens, com o objetivo de melhorar a distinção entre áreas ocupadas e não ocupadas, e, conseqüentemente, as estimativas produzidas. Com este procedimento de minimização dos problemas relativos à etapa de classificação, foi possível analisar os resultados das estimativas à luz das vantagens e limitações do modelo proposto, e, principalmente, das diferentes características de ocupação, existente nos municípios selecionados.

3.2 Satélite Landsat 7 ETM+

Inicialmente denominado de ERST-1 (Earth Resources Technology Satellite), o sistema Landsat – assim denominado em janeiro de 1975 – foi desenvolvido pela NASA (National Aeronautics and Space Administration), sendo composto por cinco satélites. Os satélites Landsat 1 e 2 carregaram dois sensores com a mesma resolução espacial, e com capacidade de aquisição de dados multiespectrais: o RBV, com imageamento instantâneo de toda a cena; e o MSS, com imageamento por varredura de linhas (line-scanner) [INPE, 2001; Jensen, 2009; NASA, 2017].

O terceiro satélite teve o RBV modificado, avançando na disponibilização de dados com melhor resolução espacial, em uma única faixa do espectro, além de ter acrescentado uma faixa espectral (na região do infravermelho termal) ao sistema MSS. Os satélites LANDSAT 4 e 5 passaram a contar com o sensor de varredura de linhas multiespectral TM (Thematic Mapper), que operam em oito faixas do espectro, além de diversos aperfeiçoamentos nos componentes ópticos e eletrônicos, que proporcionaram uma resolução espacial mais fina, além da melhor precisão espectral, geométrica e radiométrica, em relação ao sensor MSS (INPE, 2001; NASA, 2017).

O satélite Landsat 7 Enhanced Thematic Mapper Plus (ETM+) foi lançado em 15 de abril de 1999, sendo que a comercialização das imagens ficou sob responsabilidade da *United States Geological Survey* – USGS. A vida útil do

satélite, inicialmente prevista para aproximadamente seis anos, sairá de atividade no início de 2020. Segundo a USGS (2017), em 7 de fevereiro de 2017 ocorreu a 20ª e última manobra de inclinação Delta-I, que mantém a nave espacial na órbita correta para garantir que o satélite mantenha as aquisições de imagens no tempo médio local (10h da manhã). Quando o satélite Landsat 9 for lançado no final de 2020, o Landsat 7 estará em uma órbita inferior, permitindo que o Landsat 9 se mova para a altitude de órbita padrão de 705 quilômetros após o lançamento. Ao atingir o tempo de aquisição de imagens de 9h15 da manhã (horário local), o Landsat 7 não estará mais fornecendo dados científicos válidos e o satélite será desativado, enquanto o Landsat 9 permitirá a continuação de um deslocamento de 8 dias para aquisições de imagens, juntamente com o Landsat 8 (USGS, 2017).

Portanto, o programa de satélites Landsat estabelece uma longa vida aos satélites da série, a partir de uma perspectiva de manutenção da série histórica das informações geradas (sem interrupções). Além disso, os sensores a bordo das plataformas do Programa Landsat (em particular suas bandas espectrais) permitem que novas metodologias sejam aproveitadas para o desenvolvimento de novos sensores, com melhor resolução espacial e mesma resolução radiométrica.

As imagens Landsat 7 ETM+ são compostas por oito bandas espectrais, sendo seis bandas com resolução espacial de 30 metros, uma banda pancromática com resolução de 15 metros, e a banda termal com resolução de 60 metros. O Quadro 1 mostra as principais características do sensor ETM+, com destaque para as faixas do espectro (definida para cada banda):

QUADRO 1 – Principais Características do Sensor ETM+

Características	Parâmetros do Sensor
Bandas Espectrais (μm)	Banda 1 - 0,45 a 0,52 Banda 2 - 0,53 a 0,61 Banda 3 - 0,63 a 0,69 Banda 4 - 0,78 a 0,90 Banda 5 - 1,55 a 1,75 Banda 6 - 10,4 a 12,5 Banda 7 - 2,09 a 2,35 Banda 8 - 0,52 a 0,90
Resolução Espacial	15 metros (banda pancromática) 30 metros (bandas 1 a 5 e 7) 60 metros (banda 6)
Resolução Radiométrica	8 bits (256 níveis de cinza)
Dimensão das Cenas	170 km (norte-sul) / 183 km (leste-oeste)

Fonte: USGS, 2017.

A banda 1 possui grande penetração em corpos d'água, permitindo a análise de turbidez. A banda 2 também possui boa penetração em corpos d'água, tendo grande sensibilidade a sedimentos em suspensão, sendo também útil para o mapeamento da vegetação e de áreas de ocupação humana. A banda 3, por sua vez, é a mais utilizada para a delimitação de manchas urbanas, tendo, também, boa resposta para diferentes tipos de vegetação, além de apresentar bom contraste entre a vegetação e o solo exposto. Já a banda 4 (referente ao infravermelho próximo) é sensível ao relevo, sendo muito utilizada para mapeamentos geológicos, além de apresentar bom contraste entre solo e corpos d'água. A banda 5 é adequada a análise de umidade da vegetação e do solo. A banda 6 detecta a umidade das plantas, além de propriedades termais dos solos, sendo útil para estudos sobre ilhas de calor urbanas. Por fim, a banda 7 possui sensibilidade ao terreno, à umidade do solo e da vegetação.

Em relação à abrangência de captação das imagens do sistema Landsat 7, essas se estendem por toda longitude do globo, além de grande parte da latitude (desde 81° de latitude norte até 81 de latitude sul). A órbita do satélite é descendente (de norte para sul), e cruza a linha do Equador entre 10h e 10h15 (hora local) em cada passagem, sendo que sua órbita é heliosíncrona, ou seja, passa sempre num mesmo local no mesmo horário solar. O satélite realiza uma órbita em aproximadamente 99 minutos, o que permite realizar 14 voltas na Terra por dia, cobrindo a área do planeta em 16 dias (USGS, 2017).

As cenas de Landsat 7 coletadas a partir de 30 de maio de 2003 apresentam lacunas de dados devido à falha do Scan Line Corrector – SLC. Como consequência, há perda de informação em 22% dos pixels de todas as imagens, a partir da data referida. Diante disso, com o objetivo de utilizar um único sensor para o período em análise, no presente trabalho, foi utilizado um procedimento de filtro (de interpolação de pixels) para a correção dos pixels com ausência de informação. Assim como nos demais procedimentos, foi utilizado o software QGIS 2.18.3).

Por fim, vale ressaltar que os números digitais das imagens de satélite correspondem à radiância, para cada banda da imagem. Porém, os modelos de regressão utilizados para o cálculo das estimativas consideram as reflectâncias de superfície. Como já fora comentado, a USGS possui um projeto, denominado ESPA, que disponibiliza imagens de satélite com os dados de reflectância nos pixels, o que dispensou a necessidade de cálculo das reflectâncias, para a construção dos modelos de regressão.

Discutido os principais conceitos de processamento digital de imagens, os próximos itens apresentam os principais métodos de distribuição da população (dentre os quais os métodos dasimétricos), os principais modelos de estimativas populacionais para pequenas áreas (com foco nos métodos que utilizam de imagens orbitais); e, nesse contexto, os modelos que serão utilizados neste trabalho para o cálculo das estimativas via sensoriamento remoto.

3.3 Distribuição Espacial, Estimativas de População e Sensoriamento Remoto

A representação geográfica de dados socioeconômicos em ambiente computacional pode ser feita por meio da agregação dos dados individuais por unidade de área, que incluem as representações temáticas de dados censitários. Entretanto, a utilização de dados censitários em pequenas áreas corresponde a um grande desafio, em função da dificuldade em se comparar dados de censos com outras feições (ou limites) do espaço, o que resulta num grande obstáculo para a integração entre as ciências sociais e naturais. Outro problema é a dificuldade em se realizar análises temporais, dado que, a cada censo, as áreas de enumeração

(e seus limites) mudam em função das mudanças na densidade populacional, como é o caso dos setores censitários que constantemente são desagregados em novos setores. Esse tipo de limitação é conhecido na literatura como “Problema da Unidade de Área Modificável” (ou MAUP, do inglês *Modifiable Areal Unit Problem*).

Nesse contexto, há um conjunto de métodos de interpolação de área que foram elaborados basicamente para solucionar os problemas de transformação das unidades espaciais, e que utilizam, normalmente, dados censitários para a aplicação de técnicas de desagregação para a obtenção de superfícies mais refinadas da distribuição populacional.

Esses métodos de interpolação podem ser subdivididos entre aqueles que não utilizam de dados auxiliares e os que utilizam de informações auxiliares (ou métodos dasimétricos), como é o caso das imagens de satélite. Como o foco deste trabalho insere-se no campo dos métodos dasimétricos, os métodos de interpolação que não utilizam de informações auxiliares serão apenas brevemente introduzidos (no item 3.3.1), tendo como referência os trabalhos de Wu *et al.* (2005) e Kampel (2003). Já o item 3.3.2 aprofunda o tema dos métodos de interpolação que utilizam de informações auxiliares, com foco no sensoriamento remoto.

3.3.1 Breve Introdução aos Métodos de Interpolação de Área que não utilizam de Informações Auxiliares

Um dos primeiros métodos de interpolação de área que não utilizam de dados auxiliares é o de tendência de superfície, que utiliza o “ajuste de um polinômio aos dados pontuais, utilizando-se de uma regressão múltipla dos valores do atributo em função da localização geográfica” (Kampel, 2003). Segundo Kampel (2003), esse método gera superfícies suavizadas, sendo mais adequado para superfícies com tendências mais lineares, sem a existência de interferências naturais. Outros métodos simples também consolidados na literatura são os métodos ponderados pela área e pela distância (aos pontos de informação populacional).

Em relação a métodos estocásticos, a técnica de Krigagem é um dos métodos mais conhecidos na literatura. Segundo Kampel (2003), a técnica de Krigagem “estima a ocorrência de um processo estocástico a partir de observações do mesmo

processo em outros locais”, sendo que a diferença entre os valores entre diferentes localidades é definida pela distância dessas localidades, dada por uma função que explica essa distribuição (Kampel, 2003). Kampel (2003) ressalta que o método de Krigagem preenche toda a área estudada com informação interpolada, sendo, portanto “imprecisa para ser utilizada numa modelagem, porém (...) indicando grandes tendências da distribuição da variável” (Kampel, 2003).

Wu *et al.* (2005) afirma que, quando são utilizados pontos para interpolação, normalmente se utiliza o centroide da área, que, por sua vez, raramente representa a distribuição real da população inserida dentro da unidade espacial. Como consequência, a maioria dos métodos não preserva o volume da população original (Wu *et al.*, 2005).

O método picnofilático desenvolvido por Tobler em 1979 (apud Wu *et al.*, 2005, p. 83), trata-se de uma extensão do método de ponderação por área, e preserva o volume final da superfície estudada. Os valores são ponderados pela distância do centroide, de modo que a superfície de densidade (concêntrica em torno do centroide) é suavizada para o limite das unidades espaciais originais (Kampel, 2003). Assim como nos demais métodos, o método picnofilático indica a presença de população em toda a área estudada (Kampel, 2003).

Ainda segundo Kampel (2003), o método de Kernel corrige essa limitação, ao utilizar interpoladores inteligentes (baseados em funções de decaimento com o aumento da distância) que “geram superfícies de densidade populacional onde não há a obrigatoriedade de se interpolar valores para todas as células da grade” (Kampel, 2003), podendo se observar valores “vazios”, com ausência de população. Wu *et al.* (2005) cita Martin (1996, p.83), que modificou o algoritmo de interpolação de Kernel, para garantir que a população interpolada fosse compatível com a população observada na unidade espacial original.

Um exemplo importante na literatura internacional sobre a aplicação do método Picnofilático trata-se do projeto Gridded Population of the World (GPW), desenvolvido pelo Center for International Earth Science Information Network (CIESIN), da Universidade de Colúmbia, Estados Unidos. Em sua última versão (GPW3) foram disponibilizadas estimativas populacionais para os anos de 1990,

1995 e 2000 para diversos países do globo, a partir da transformação de unidades espaciais administrativas para um zoneamento regular de resolução de 2,5 graus de latitude e longitude (aproximadamente 5 km) [Balk e Yetman, 2004]. Segundo Balk e Yetman (2004) a metodologia não pode ser considerada como dasimétrica, dado que não envolve o procedimento de desagregação com base em informações auxiliares.

Não é objetivo explorar maiores detalhamentos e aplicações dos métodos apresentados acima, pois os métodos que não utilizam de dados auxiliares não são os mais adequados (e utilizados) para a produção de estimativas populacionais em períodos pós-censitários; e, portanto, não fazem parte do escopo deste trabalho. Todavia, a apresentação das principais características dos métodos citados acima serve de base para a compreensão das vantagens da utilização de informações auxiliares para a distribuição da população no espaço.

3.3.2 Métodos Dasimétricos, Estimativas Populacionais e Sensoriamento Remoto

Nas Ciências Sociais, são extensas as possibilidades de utilização de sensoriamento remoto, para as mais diversas finalidades, como destaca Sherbinin (2002):

“In many parts of the world, spatial data on roads or infrastructure, farm sizes, industrial activities, or any number of other variables visible from space are either not available or difficult to obtain. In other instances, an area may simply be inaccessible for reasons of political turmoil or armed conflict. In these cases, remote sensed data, utilized independently or entered into a geographic information system (GIS), may provide an alternative source of data (...) include the following: the world price for a commodity, the farm gate price, the distance to major markets, what other farmers are growing, and the soil type and quality. Remote sensing can provide important information on biophysical parameters such as slope, aspect, soil types, water bodies and vegetation cover, and, in some cases, infrastructure parameters such as roads, pipelines, or power lines, that can impact people’s decision-making or livelihood options” (Sherbinin, 2002, p.18).

No campo das estimativas populacionais, os avanços alcançados nas últimas décadas, em relação aos métodos de interpolação de área via sensoriamento

remoto, têm possibilitado o desenvolvimento de modelos de estimativas populacionais mais precisos.

Para uma ampla visão quanto às potencialidades da utilização de sensoriamento remoto para a produção de estimativas populacionais, faz-se importante compreender as limitações do uso desse tipo de informação. Sobre uma síntese das limitações quanto ao uso do sensoriamento remoto para a produção de estimativas populacionais, Lu *et al.* (2006) comenta:

“The spectral signature-based method for population estimation has a common problem, in that estimations are affected by the external conditions of satellite imaging, such as the complexity of urban landscape, atmospheric conditions, phenological conditions, and sun elevation angles on land cover reflectance. Because population is not directly related to land cover surface reflectance captured by remote sensors, population estimation is still a challenging task based purely on remote sensing spectral signatures” (Lu et al., 2006, p.3554).

Além das limitações externas às imagens comentadas por Lu *et al.* (2006), uma série de outras variáveis interferem na produção de estimativas populacionais, como a resolução espacial das imagens e a escala (ou nível de desagregação) com o qual se pretende estimar a população (no nível municipal ou intramunicipal, por exemplo). Diante da diversidade de aplicações existentes, os próximos itens aprofundam a discussão e as aplicações dos métodos de distribuição espacial da população, com foco na relação entre métodos dasimétricos que utilizam de sensoriamento remoto e modelos matemáticos, utilizados para o cálculo de estimativas populacionais.

Como já comentado, o mapeamento dasimétrico utiliza um conjunto de dados auxiliares (de cobertura da terra) com a finalidade de eliminar áreas de restrição ambiental para as ocupações humanas, possibilitando um maior detalhamento da distribuição da população (e, portanto, de sua densidade) no espaço (Eicher e Brewer, 2001; Mennis, 2003; Holt *et al.*, 2004; Wu *et al.*, 2005; Mennis e Hultgren, 2006).

Segundo Petrov (2008), a popularização dos métodos dasimétricos ocorreu na década de 1960, no contexto do surgimento da geografia computacional, quando os avanços tecnológicos permitiram o desenvolvimento de operações mais

complexas, em relação ao processamento das informações. Petrov (2008) ressalta o interesse crescente pelo mapeamento dasimétrico nos últimos anos, em face à crescente demanda pelo mapeamento da distribuição da população e pela produção de estimativas, para diversas áreas do planejamento urbano e regional; como, por exemplo, para o desenvolvimento de planos de mitigação a desastres, para o planejamento dos serviços de saúde, prevenção contra o crime, modelagem de transportes, entre outras aplicações (Petrov, 2008).

O conceito de mapeamento dasimétrico teve origem no início do Século XX. Em 1911, Benjamin Semenov-Tian-Shansky foi pioneiro ao propor o termo “método dasimétrico” (baseado no grego *dasys* – denso ou densidade, e *metreo* – medida), numa sessão especial de um encontro da Sociedade de Geógrafos da Rússia, no trabalho intitulado “*Dasymetric Map of European Russia*”, que seria publicado anos depois (Petrov, 2008). Segundo Petrov (2008), Semenov-Tian-Shansky definiu mapa dasimétrico como aqueles cuja densidade populacional, independentemente dos limites administrativos, é apresentada em sua distribuição real.

Na década de 1920, Semenov-Tian-Shansky publicou trabalhos que contemplaram 47 mapas dasimétricos da Rússia Européia, baseados no conceito de sobreposição de informações auxiliares, tais como topografia e mapeamento coroplético da população. A metodologia original foi organizada em quatro etapas principais: estimativa da contagem populacional no nível local; identificação de áreas (ou zonas) de concentração populacional; cálculo da densidade populacional; e exclusão das áreas inabitadas (Petrov, 2008). Segundo Petrov (2008), o método desenvolvido por Semenov-Tian-Shansky preserva o princípio do método picnofilático de Tobler, no qual a soma da população na área antiga (uma unidade administrativa, por exemplo) e na área nova (dasimétrica) é a mesma.

Assim como nos trabalhos de Semenov-Tian-Shansky, o trabalho de Wright (1936) é bastante citado na literatura como um dos precursores nos estudos sobre distribuição espacial da população. Partindo da idéia de que os mapas coropléticos não forneciam a distribuição espacial da população de maneira precisa, Wright (1936) propôs a utilização do conhecimento da localidade para o refinamento da distribuição espacial da população. Wright (1936) identificou áreas com diferentes níveis de densidade populacional dentro de setores administrativos, a partir de

partições ou divisões binárias construídas de forma iterativa e preservando o volume populacional (na localidade de Cape Cod, Estados Unidos).

Nas últimas décadas, os avanços nos métodos dasimétricos, especialmente aqueles que utilizam de sensoriamento remoto, estão fortemente relacionados ao desenvolvimento de modelos de estimativas populacionais, e que serão explorados nos próximos subitens deste tópico. A organização e alguns dos principais conceitos gerais que serão apresentados nesses subitens foram baseados em Wu *et al.* (2005), que realizam uma revisão sobre os métodos de interpolação de área, com foco nos métodos que utilizam de imagens de satélite para a produção de mapeamentos dasimétricos e estimativas populacionais. Wu *et al.* (2005) subdividem a abordagem em cinco categorias, baseadas na relação entre população e sensoriamento remoto: correlação com as áreas urbanas; correlação com o uso do solo; contagem das unidades de moradia; e correlação com os pixels das imagens de satélite.

3.3.2.1 Correlação entre População e Áreas Urbanas

Os métodos de correlação com áreas urbanas abordam o relacionamento entre as áreas urbanas e o tamanho populacional. Tobler (1969) foi o precursor nesse tipo de aplicação, ao estudar a relação proporcional entre o raio de abrangência das áreas urbanas com o número de habitantes, a partir da equação:

$$r = aP^b \quad (3.4)$$

Onde: r é o raio da área urbana; P é a população; a e b são parâmetros (ou parâmetros) que definem a relação proporcional entre o raio urbano e a população.

Segundo Wu *et al.* (2005), a crescente disponibilização de imagens de satélite (como as imagens Landsat) e os avanços nas técnicas de processamento de imagens permitiram aos pesquisadores estudar o relacionamento entre população e áreas urbanas.

Lo e Welch (1977) utilizaram imagens Landsat para a definição de áreas urbanas de municípios chineses entre 400 mil e 3 milhões de habitantes, no período entre 1951 e 1956, e utilizaram a equação proposta por Tobler para a obtenção de

estimativas populacionais para o ano de 1970, encontrando coeficientes de correlação de aproximadamente 0,82.

No Brasil, Foresti (1978) utilizou imagens Landsat e fotografias aéreas para a produção de estimativas para um conjunto de municípios do Estado de São Paulo. Num primeiro momento, 35 municípios foram selecionados para a análise comparativa entre as fotografias aéreas e imagens Landsat (para os anos entre 1971 e 1973); enquanto que, num segundo momento, foram selecionados 70 municípios para o estudo dos modelos de estimativa populacional (para o período entre 1975 e 1976). Na análise comparativa entre as fotografias aéreas e as imagens Landsat, foram encontrados coeficientes de correlação da ordem de 0,96, enquanto que, na análise da relação entre população e área urbana a relação encontrada foi de 0,95 e 0,92 (respectivamente às fotografias aéreas e imagens Landsat).

Durand *et al.* (2007) utilizam imagens de média resolução (Landsat) e de alta resolução (IRS) para modelar o crescimento da população urbana e para a inferência populacional, para um conjunto de municípios do estado de São Paulo, no período entre 1985 e 2000. A partir da correlação entre população e área ocupada, estimativas populacionais foram calculadas para os anos de 2005 e 2010. Os autores concluem que o aumento da precisão nos mapeamentos das manchas urbanas (resultante da maior resolução espacial das imagens IRS) aumentam a precisão das estimativas.

Naturalmente, estudos de correlação com áreas urbanas não se restringem ao uso das imagens Landsat. Nas últimas duas décadas tem crescido o número de trabalhos que utilizam, por exemplo, luzes noturnas como indicador do crescimento populacional.

Sutton (1997) utilizou imagens de luzes noturnas do *Programa Defense Meteorological Satellite Program – DMSP/OLS* para a análise da distribuição populacional. O trabalho foi motivado pela crítica de Sutton (1997) à aplicação do modelo original de Tobler (1969) para áreas mais extensas, pois afirma que os parâmetros da relação entre população e áreas urbanas (dado na equação 3.32 por *a* e *b*) variam fortemente entre regiões, em função de determinantes ambientais,

econômicos e culturais. Nesse contexto, Sutton (1997) propõe as imagens DMSP/OLS como uma alternativa para o mapeamento em larga escala (em nível global) da distribuição da população. O autor conclui que a correlação não é tão forte quanto o inicialmente esperado, e indica pela necessidade de avanços na disponibilização de imagens com uma maior resolução (principalmente espacial). Sutton (1997) analisa imagens para os Estados Unidos e demonstra que a correlação entre as imagens e população é maior em áreas saturadas, indicando que 80% da população encontra-se em 10% do território.

No Brasil, merece destaque a tese de Kampel (2003), que utiliza imagens de luzes noturnas DMSP/OLS para a identificação de assentamentos urbanos e a presença humana na Amazônia Legal, para o ano de 1999. Kampel (2003) realizou uma análise sobre as relações entre as áreas com a presença de luzes noturnas, as áreas de manchas urbanas (identificadas através de outras bases de dados, como imagens Landsat e CBERS), dados censitários de população e dados de consumo de energia elétrica. Para os municípios que apresentaram focos de luzes DMSP na Amazônia Legal, foram utilizadas regressões lineares para a identificação da relação entre a população total e urbana e os focos de luzes, do qual foram obtidos coeficientes de determinação de 0,59 para a população total e 0,60 para a população urbana. Ressalta-se ainda que a exclusão dos municípios atípicos melhorou o coeficiente de correlação entre população urbana e luzes da imagem DMSP para 0,79 (Kampel, 2003). Kampel (2003), bem como Kampel *et al.* (2006), mostraram que as luzes noturnas DMSP/OLS detectam a presença da atividade humana, bem como dos níveis de consumo de energia elétrica e da extensão da área urbanizada.

Kampel (2003) e Kampel *et al.* (2012) propõem uma abordagem metodológica de redistribuição da população (obtidas através de dados censitários) em superfícies de densidade, baseados na construção de grids ou células espaciais. A pesquisa foi posteriormente replicada em outros 13 municípios (que se localizam ao longo da rodovia BR 163), para os anos de 2000 e 2007. Para a construção da superfície de ocorrência populacional, foi necessária a incorporação de parâmetros locais da região em estudo, dado que o perfil físico e o processo de ocupação nessa região é bastante particular, por estar localizada na região amazônica.

Primeiramente, a pesquisa constituiu da aplicação do método dasimétrico, para a eliminação das áreas não ocupadas, a partir da classificação do uso e cobertura da terra. Em seguida, foi aplicado um método de interpolação multivariado para gerar a superfície potencial, a partir de inferência fuzzy sobre dados ambientais indicativos de presença humana. Nessa etapa, assim como se observa em metodologias que usam o sensor DMSP/OLS LandScan (Bhaduri *et al.*, 2007), foram definidos critérios tais como: distância de estradas e rios (em metros), distância de distritos, percentual de cobertura florestal, entre outras. Por fim, foi realizada a redistribuição da população em valores para cada célula, de forma proporcional à ocorrência potencial da população, definida na segunda etapa. Para a classificação dos corpos d'água foram utilizadas imagens TM Landsat, de resolução espacial de 30 metros, nas bandas 4 (0,750-0,900 μm , no infravermelho próximo). Para a classificação das florestas foi utilizado o indicador NDVI (normalized vegetation index), com informação espectral no infravermelho próximo na banda 4, e no visível na banda 3 (0,630-0,690 μm).

Kampel *et al.* (2012) concluem que as superfícies de densidade populacional permitiram a interpretação da distribuição da presença humana, em termos de território potencialmente ocupado. Além disso, a representação a partir de células espaciais permite o monitoramento da população através do tempo. Contudo, assim como em Harvey (2002b), Kampel *et al.* (2012) concluíram que a heterogeneidade na cobertura do solo e na forma de ocupação resulta na necessidade de se considerar as particularidades locais para a definição dos métodos de classificação e interpolação a serem empregados.

Mais recentemente, Dória (2015) e Dória *et al.* (2016) produziram estimativas a partir de imagens de luzes noturnas do *Operational Linescan System* (OLS) e do *Visible Infrared Imaging Radiometer Suite* (VIIRS), para o Distrito Florestal Sustentável da BR 163 e para a Região Metropolitana de São Paulo (RMSP). A técnica de regressão linear foi adotada para a análise das estimativas calculadas via OLS para o distrito florestal, para os anos 2000, 2007 e 2010, enquanto que a técnica de regressão geograficamente ponderada foi utilizada para a análise das estimativas baseadas nas imagens VIIRS, para os distritos da RMSP. Dória (2015) e Dória *et al.* (2016) observam que o erro encontrado nas estimativas foi de

aproximadamente 5% para a RMSP, e conclui que as imagens VIIRS contribuem na melhoria das estimativas, dado o potencial de eliminar as luzes não associadas à ocupação, bem como ao potencial de identificar diferentes níveis de densidade populacional.

3.3.2.2 Correlação entre População e Uso do Solo

Com o objetivo de avançar nos estudos que relacionam população com expansão das áreas urbanas, foram desenvolvidos métodos que analisam a correlação entre população e diferentes tipos de uso do solo, a partir da seguinte formulação (Wu *et al.*, 2005):

$$P = \sum_j A_j * D_j \quad (3.5)$$

Onde: P é a população total estimada; A_j é a área do uso do solo j ; D_j é a densidade populacional do uso j (determinado através de análise de regressão).

Segundo Wu *et al.* (2005), a abordagem mais simples é classificar apenas dois tipos de uso, com característica binária (com ou sem a presença de população), e redistribuir a populações alocadas em unidades espaciais censitárias nas áreas definidas como povoadas. Wu *et al.* (2005) e Mennis (2003) comentam que, em 1991, Langford e colegas, em trabalho desenvolvido para Leicestershire (Reino Unido), mapearam superfícies de 1 km de resolução, sendo os primeiros a aplicar técnicas de regressão multivariada para estimar subáreas com diferentes densidades populacionais.

Holt *et al.* (2004) analisa diferentes padrões de densidade populacional através da classificação binária de imagens Landsat 5 para a identificação das áreas de ocupadas de 13 municípios da Região Metropolitana de Atlanta (Georgia) para os anos de 1980 e 1990 e 2000. Métodos dasimétricos foram utilizados para a desagregação da população (obtida por unidades espaciais administrativas) para os pixels classificados como ocupado, para a obtenção da densidade populacional. Holt *et al.* (2004) conclui que, ao utilizar técnica de mapeamento dasimétrico, o simples fato de incorporar à análise áreas inabitadas, já representa um ganho substancial na precisão das medidas de densidade populacional.

Ao analisar diferentes padrões de densidade populacional via mapeamento dasimétrico (utilizando imagens Landsat), Holt *et al.* (2004) afirma que *“the traditional manner of computing population density overestimates population density in unpopulated and sparsely populated areas, and it underestimates population density in more-densely populated areas”* (Holt *et al.*, 2004). Este comportamento é similar ao encontrado em outros autores, e trata-se de um elemento importante para a análise deste trabalho, e que justifica em parte a escolha dos municípios selecionados para este estudo.

Uma evolução em relação à classificação binária das imagens de satélite foi o desenvolvimento de novos critérios baseados na maior quantidade de usos do solo, denominados de métodos dasimétricos inteligentes, entre os quais se destacam os trabalhos de Eicher e Brewer (2001), Mennis (2003), Mennis e Hultgren (2006) e Badhuri *et al.* (2007).

Eicher e Brewer (2001) comparou três métodos dasimétricos, para 159 municípios dos estados da Pennsylvania, West Virginia, Maryland, Virginia, além do distrito de Columbia, e avaliou os resultados através da medida de erro raiz média quadrática. O primeiro é o método binário, proposto por Langford, em 1991. O segundo método consiste na utilização de três classes de uso, no qual se define (de forma subjetiva, através da consulta a outros trabalhos) uma distribuição dos usos para cada município. Fora definido que 70% dos dados seria assinalado como uso urbano, 20% como uso agrícola e 10% como florestas. Segundo Eicher e Brewer (2001), a maior limitação do método de três classes *“is that it does not account for the area of each particular land use within each county”*, ou seja, trata-se de uma generalização a distribuição dos usos em face à possibilidade de heterogeneidade nos perfis de uso e ocupação, entre os municípios. O terceiro método, denominado de “variável limitadora”, é também baseado em três classes, e define máximos de densidade para cada tipo de uso, também de forma subjetiva.

Os três métodos ainda foram desagregados pelo critério de mapeamento por área e por grades de células, no qual se constatou que as informações distribuídas por grades apresentam erros menores, além de serem mais funcionais (dado que não são restritas aos limites administrativos, variáveis ao longo do tempo). Na comparação entre os métodos dasimétricos, o método das variáveis limitadoras

apresentou os menores erros, demonstrando o potencial de aplicação dos métodos dasimétricos inteligentes nos estudos de distribuição espacial da população.

Mennis (2003) e Mennis e Hultgren (2006) propõem o aperfeiçoamento do método de três classes (proposto por Eicher e Brewer, 2001), a partir do que os autores denominam como Intelligent Dasymetric Mapping (IDM). Nesse modelo, são realizadas seleções de amostras (baseadas no conhecimento sobre a área), e essas amostras definem a distribuição das densidades de cada classe, e de cada município. A partir dos dados de densidade para as classes de uso, os autores utilizam razões de densidade entre as classes para redistribuir a população na área estudada. Mennis (2003) analisa cinco municípios do sudeste do estado da Pensilvânia (Bucks, Chester, Delaware, Montgomery e Philadelphia), a partir de dados de superfície populacional com 100 metros de resolução, derivados de imagens Landsat TM, de dados de rede rodoviária e de fotografias aéreas, enquanto Mennis e Hultgren (2006) estudam o estado do Colorado, parte de Denver, além dos municípios de Jefferson, Boulder e Adams.

No Brasil, França (2012) aplica o algoritmo do mapeamento dasimétrico inteligente (baseado em Mennis, 2003), a partir da ponderação por logradouros e por endereços, para a Mesorregião Metropolitana de Belém, com o objetivo de avaliar a eficácia diferentes métodos de mapeamento dasimétrico. Dentre os resultados, França (2012) conclui que a ponderação por endereços (oriundos do Cadastro Nacional de Endereços para Fins Estatísticos - CNEFE) produziu as melhores estimativas.

França (2012) afirma que, baseado somente em dados vetoriais, o método de ponderação por endereços é mais acessível e de baixo custo de processamento, embora apresente melhor ajuste nas áreas urbanas, em função da deficiência na cobertura do mapeamento viário em áreas rurais e pequenos municípios. Sobre o método de mapeamento dasimétrico inteligente, embora França (2012) tenha concluído que a série temporal de imagens Landsat permite a construção de mapeamentos para mais de um período censitário, o autor observa que as imagens de média resolução não permitem a distinção adequada dos usos da terra em áreas urbanas. Contudo, é importante ressaltar que França (2012) utilizou como base de dados o TerraClass (INPE e EMBRAPA), que foram construídos a partir de imagens

Landsat do ano de 2008, dois anos antes do Censo de 2010, o que contribuiu para a subestimativa dos resultados encontrados, especialmente nas áreas de maior crescimento populacional. O método de ponderação por endereços (a partir de dados do CNEFE), segundo França (2012), tem a vantagem de possibilitar a localização com maior nível de precisão espacial, especialmente de estruturas residenciais, o que permite a estimação da população em áreas menores.

Outro exemplo de interpolador inteligente é o projeto LandScan. Segundo Badhuri *et al.* (2007), o projeto apresenta uma metodologia baseada em modelagem multi-dimensional, que permite a distribuição espacial da população nos níveis nacional e global, a partir de resolução espacial de 3 arcos por segundo (aproximadamente 90 metros) para os Estados Unidos, e de resolução de 30 arcos por segundo (aproximadamente 1 km) para os demais países.

Badhuri *et al.* (2007) descrevem a aplicação do projeto LandScan para os Estados Unidos, cujo modelo considera um conjunto de variáveis (tendo cada variável um peso associado), tais como proximidade de estradas, uso do solo, parques, escolas, prisões, aeroportos, entre outros, para a definição de um coeficiente de probabilidade para a distribuição da população no espaço (Bhaduri *et al.*, 2007). Para tanto, o projeto utiliza diversas fontes de dados, como censos, imagens de luzes noturnas, entre outras fontes, tendo como inovação a distinção entre estimativas para a população noturna (obtida a partir de estatísticas oficiais, baseadas no local de residência) e para a população diurna (obtida através de dados sobre movimentos pendulares diários). Em teste aplicado para Iowa, foram utilizadas um conjunto de 964 fotografias aéreas, que mostraram precisão de 72,5% na predição de células populacionais e 99% na predição de células sem povoamento (Badhuri *et al.*, 2007).

3.3.2.3 Contagem das Unidades Habitacionais

A população total de uma área pode ser estimada a partir do produto entre o número de unidades habitacionais e o número de pessoas que normalmente vivem nessas unidades, e diversos métodos podem ser utilizados, como categorizar os domicílios

e aplicar diferentes médias de habitantes por domicílio, a partir de cada categoria (Wu *et al.*, 2005).

Segundo Jensen (1999), existem alguns pressupostos que devem ser considerados para a aplicação deste tipo de método, como a resolução das imagens de satélites, que devem ser suficientes para a identificação das estruturas individuais (ou seja, devem ser imagens de alta resolução espacial), além da identificação do uso das edificações, tornando possível a distinção entre estruturas residenciais, comerciais e industriais (Jensen, 1999). É preciso, também, ter disponível informações sobre o número médio de habitantes por domicílio, bem como assumir o pressuposto de que todas as unidades residenciais estão ocupadas (Jensen, 1999).

Um bom exemplo deste tipo de aplicação no Brasil é o trabalho de Souza (2004), que utilizou imagens IKONOS para a estimação da população do município de São José dos Campos. Para tanto, Souza (2004) realizou interpretação de imagens para a identificação do perfil do espaço residencial urbano, e selecionou 3% da área total dos setores amostrais (denominado de áreas homogêneas) para a análise detalhada, baseada na contagem dos domicílios. Dados do Censo 2000 foram utilizados para a definição do número médio de habitantes e da densidade habitacional das áreas homogêneas. O erro foi estimado em aproximadamente 10%, quando comparado aos dados censitários, para aquele ano.

Wu *et al.* (2005) destacam um outro conjunto de métodos de extração de edificações, baseado em técnicas 3D, oriundas de dados LIDAR (*Light Detection And Ranging*). Trata-se de um sensor remoto ativo, que envia sinais à superfície da Terra e registra o sinal refletido, sendo medida a distância a um objeto, através da diferença de tempo entre a emissão de um pulso laser e a detecção do sinal refletido (Jensen, 2009).

Xie *et al.* (2015) utilizaram dados do sensor LIDAR e algoritmos de detecção de estruturas, para a identificação de edificações residenciais, com o objetivo de estimar a população de Indianapolis, Estados Unidos. O método utilizado permitiu a categorização das estruturas em edificações não-residenciais, apartamentos, casas e outras estruturas, além de vegetação e áreas impermeáveis, sendo que a

acurácia na detecção foi de aproximadamente 95%. Num segundo momento, Xie *et al.* (2015) utilizaram modelos de regressão linear a partir de informações como área e unidades domiciliares para analisar a relação entre população e as variáveis extraídas dos dados do sensor LIDAR, sendo que os resultados mostraram uma acurácia de aproximadamente 81,7%,

No Brasil, Tomás (2010) analisa a incorporação de informações sobre elevação para o aprimoramento de métodos de inferência populacional baseados em dados bidimensionais. Para tanto, Tomás (2010) utiliza imagens IKONOS e dados LiDAR para a produção de estimativas da área urbana de Uberlândia, a partir da incorporação de modelos digitais de superfície. Foram calculadas estimativas por setores censitários para o ano de 2004, e os resultados foram comparados com dados de referência, calculados através da aplicação de taxas de crescimento observadas para os bairros, entre 2000 e 2004 (produzido pelo IBGE), de modo que as taxas de crescimento observadas nos bairros foram aplicadas aos setores censitários correspondentes. A comparação foi realizada por meio de regressão linear, sendo que os resultados mostraram uma subestimativa dos valores reais (em torno de 1,35%).

3.3.2.4 Correlação entre População e os Pixels das Imagens de Satélite

Em alguns métodos descritos até aqui, a densidade populacional foi associada às características ou classes de uso do solo. Nos métodos baseados em pixels, a população pode ser correlacionada diretamente à reflectância espectral dos valores das imagens dos pixels (Wu *et al.*, 2005). Iisaka e Hegedus (1982) foram pioneiros ao estimar a população da Região Metropolitana de Tóquio, para os anos de 1972 e 1979. Iisaka e Hegedus (1982) agruparam os pixels em áreas de 500 por 500 metros, e calcularam a média espectral das bandas 4, 6 e 7 do satélite Landsat MSS. Os autores encontraram forte relação linear entre os valores espectrais calculados e a densidade populacional, da ordem de 0,939 para o ano de 1972 e 0,899 para o ano de 1979.

Contudo, os trabalhos elaborados por Harvey (2002a e 2002b) são os mais citados na literatura, considerados clássicos dentre os trabalhos que tratam da correlação

entre população e os pixels das imagens de satélite. Os métodos elaborados por Harvey (2002a e 2002b) foram adotados como referência para as aplicações propostas neste trabalho. Por essas razões, os dois métodos são apresentados de forma mais detalhada a seguir, pois, faz-se necessária uma melhor compreensão (para além das principais características e resultados encontrados), diante da intrínseca relação desses métodos com os resultados e conclusões deste trabalho de tese.

Harvey (2002a) utilizou imagens do sensor TM do satélite Landsat para estimar a população de dois distritos da Austrália: Ballarat Statistical District, localizado no centro da cidade provincial de Ballarat, 110 km a oeste de Melbourne (e que fora utilizado para a construção dos modelos de regressão); e Geelong Statistical District, localizado na cidade portuária de Geelong, a 90 km de Ballarat, sentido sudeste (utilizado para a validação externa dos modelos). Partindo dos pixels pertencentes a uma mesma unidade espacial intraurbana, os valores médios das reflectâncias das bandas 1 a 5 e 7 das imagens Landsat foram utilizados como variáveis explicativas num modelo de regressão linear. Modelos de Mínimos Quadrados Ordinários, com distribuição normal do erro, foram utilizados nesse estudo:

$$p_i = \beta_0 + \sum_{j=1}^n \beta_j r_{ij} + \varepsilon_i \quad (3.6)$$

Onde: p_i é a população ou a densidade populacional da unidade espacial i ; r_{ij} é o indicador j do sensoriamento remoto (por exemplo, uma banda espectral) da unidade espacial i ; n é o número de variáveis explicativas do modelo; e β_i são parâmetros a serem estimados a partir dos dados; e ε é o erro aleatório do modelo de regressão, e que representa a variação da população não explicada pelas variáveis de sensoriamento remoto. Assume-se que o erro é independente, apresenta distribuição normal e variância constante.

Cada variável explicativa normalmente representa uma banda da imagem de satélite, sendo a média da reflectância dos pixels inseridos numa dada unidade espacial (denominada pelo autor de "*individual zones*"). Porém, Harvey (2002a) afirma que outras variáveis podem ser incorporadas, como a razão entre bandas. Da mesma forma, a variável dependente também pode ser transformada, para a melhoria nas estimativas do modelo (como, por exemplo, a transformação para a função logarítmica). Harvey (2002a) ressalta que não utilizou técnicas como a de

autocorrelação espacial para a análise da distribuição dos erros (em função do formato irregular da área de estudo), muito embora este tipo de procedimento possa ser uma alternativa para a avaliação do modelo proposto.

Segundo Harvey (2002a), quando as variáveis explicativas possuem valores mais homogêneos, como as médias das reflectâncias de um dado conjunto de unidades espaciais, a variável dependente deve ser a densidade populacional (como foi o caso deste trabalho de Harvey). Já nos casos onde as variáveis explicativas são dependentes de escala, como a contagem de pixels urbanos (com presença de ocupação humana, por exemplo), a variável dependente deve ser a população (Harvey, 2002a).

Em termos de avaliação dos resultados, o coeficiente de determinação R^2 deve ser interpretado como um indicador da força da relação linear do modelo de regressão. Contudo, foram utilizadas outras medidas de erro para o trabalho, como o Erro Médio Percentual e o Erro Mediano, que correspondem, respectivamente, à média e à mediana da distribuição do erro percentual observado em cada unidade espacial (Harvey, 2002a).

Na primeira etapa dos resultados, os modelos de regressão foram validados internamente, a partir da seleção de amostras de treinamento para o mesmo distrito com o qual o modelo fora criado. As regressões foram analisadas a partir de 132 unidades espaciais, sendo que a variável dependente escolhida foi a média da densidade populacional, enquanto que as variáveis explicativas foram as médias das reflectâncias dos pixels pertencentes a uma mesma unidade espacial, além de um conjunto de variáveis transformadas a partir das bandas espectrais das imagens Landsat TM (como, por exemplo, a razão entre bandas) [Harvey, 2002a]. Além dessas variações, duas variações do modelo básico foram consideradas, o logaritmo e a raiz quadrada da variável dependente.

Segundo Harvey (2002a), a média da reflectância das seis bandas das imagens Landsat TM apresentaram correlação positiva, sendo as maiores correlações observadas nas bandas do visível 1, 2 e 3, enquanto que a maior correlação com a variável dependente foi observada com as bandas 4 e 5 (negativamente, -0,52 e -0,4, respectivamente) e com a banda 1 (positivamente em 0,34). Segundo Harvey

(2002a), esses resultados sugerem que a presença humana é pouco relacionada com a elevada reflectância da banda referente ao azul (a banda 1), talvez devido a superfícies pavimentadas, e pouco relacionada também à relativa baixa reflectância no infravermelho próximo, associado a relativa falta de vegetação.

Na segunda etapa dos resultados, a aplicação dos modelos ao segundo distrito (Geelong) alcançaram R^2 de 0,84 e coeficiente de correlação linear igual a 0,92 entre os valores estimados e reais da densidade populacional.

Dentre os resultados, foram encontrados coeficientes de correlação linear entre densidade e as reflectâncias acima de 0,92 nas amostras de treinamento e de 0,86 na validação externa dos modelos; enquanto que o erro mediano foi de 17,4% na amostra de treinamento e de 18,4% na validação externa. Já a população urbana obteve erros estimados de 1% e -3%, respectivamente à amostra de treinamento e à validação externa (aplicação ao Distrito de Geelong). Harvey (2002a) destaca ainda que os melhores resultados foram encontrados em unidades espaciais de densidade populacional intermediária, sendo que as áreas de elevada densidade populacional foram subestimadas, enquanto que as áreas de menor densidade foram sobre estimadas.

Em contraponto aos modelos que utilizam a agregação dos pixels para o cálculo da média das reflectâncias de uma dada unidade espacial para o cálculo de estimativas, Harvey (2002b) propõe um método de desagregação das populações localizadas em unidades espaciais para o nível dos pixels. O modelo é baseado em procedimento de regressão iterada, a partir do algoritmo “*Expectation-maximization*” (EM), de Dempster (1971, apud Harvey, 2002b), que corresponde a um método estatístico para a estimação de parâmetros em situações de ausência de dados (ou de dados incompletos). O modelo permite a associação entre população e a reflectância de cada pixel da imagem, melhorando a performance das estimativas, principalmente em situações extremas de densidade populacional. Além disso, Harvey (2002b) defende que este tipo de modelo é mais flexível espacialmente, devido à sua independência de unidades administrativas, sendo adequado para a maioria das aplicações demográficas.

A partir das mesmas províncias do estudo anterior (Ballarat e Geelong), Harvey (2002b) utiliza pixels de imagem Landsat TM, inicialmente classificadas como residencial ou não residencial. O algoritmo expectation-maximization (EM) foi usado para a construção e reestimação de uma regressão iterada dos pixels populacionais. A validade preditiva do modelo também foi testada a partir da aplicação da equação ajustada da regressão na segunda imagem (do segundo distrito, como em Harvey, 2002a).

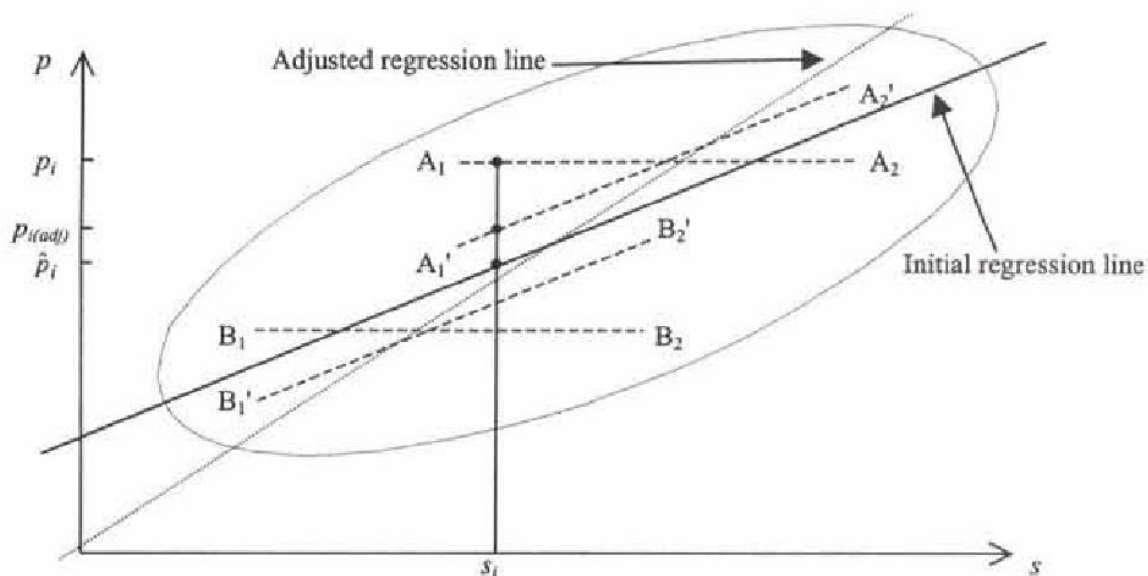
Uma das dificuldades no modelo reside em definir um critério para a distribuição da população no nível dos pixels. Harvey (2002b) sugere a aplicação do pressuposto de densidade constante, que pode ser obtida a partir da razão entre a população de uma dada unidade espacial e o número de pixels contidos nessa unidade, de modo que cada pixel contenha o mesmo número de pessoas.

$$p_{i,j} = \frac{P_j}{n_j} \quad (3.7)$$

Onde: $p_{i,j}$ é a população inicialmente assinada no pixel i , localizado na unidade espacial j ; P_j é a população total da unidade espacial j ; e n_j é o número de pixels na unidade espacial j , classificados como ocupados.

Os pixels residenciais de uma dada unidade espacial são representados na Figura 1. As linhas $A1A2$ e $B1B2$ são exemplos hipotéticos de representação dos pixels residenciais numa dada unidade espacial (sendo que, no exemplo, $B1B2$ apresentam menor densidade populacional). O eixo p representa a densidade populacional do pixel, enquanto o eixo s representa a informação espectral do pixel. Segundo Harvey (2002b), o fato dos pixels terem a mesma população, porém com diferentes valores espectrais, resulta em retas paralelas ao eixo s . Calculado os coeficientes de regressão, a população é estimada para cada pixel, sendo que a regressão estima \hat{p}_i , com densidade inferior à população assinalada por p_i (no exemplo da Figura 4).

FIGURA 4 – Representação Conceitual do Modelo (Harvey, 2002b)



Fonte: Figura Original, extraída de Harvey (2002b).

Partindo do pressuposto de que há uma relação linear entre p e s , é possível melhorar a estimativa da população de cada pixel, como afirma Harvey 2002b), sendo que “CD” corresponde à unidade espacial:

“Assuming that there is in fact a linear relationship between p and s , we can now refine our initial assigned pixel populations, by redistributing population within each CD away from underestimated pixels and towards overestimated pixels while maintaining the known CD total. Intuition suggests, and it can be shown (Harvey, 1999) using the method of Lagrange multipliers, that the optimal such redistribution in a leastsquares sense, which minimizes the sum of squared residuals about the regression line while holding the sum of the p_i values constant” (Harvey, 2002b).

Portanto, Harvey (2002b) propõe a aplicação de um método que otimiza a redistribuição dos mínimos quadrados, tornando os resíduos iguais ao longo da linha de regressão. Com este procedimento, os pixels residenciais de uma dada unidade espacial deixam de apresentar linhas paralelas ao eixo s , passando a apresentar linhas paralelas à linha da regressão inicial (retas $A_1'A_2'$ e $B_1'B_2'$, na Figura 1). A densidade ajustada ($p_{i(ajustado)}$) é obtida a partir das equações abaixo (onde \hat{p}_i é a regressão estimada):

$$P_{i(ajustado)} = \hat{p}_i + \bar{r} \quad (3.8)$$

$$\bar{r} = \frac{\sum_{i=1}^n (p_i - \hat{p}_i)}{n} \quad (3.9)$$

Segundo Reis (2005) “a população ajustada do pixel i é dada pela soma da população estimada do pixel i e da média dos resíduos do setor censitário ao qual pertence o pixel”. Segundo Harvey (2002b), as populações ajustadas para cada unidade espacial, paralelas à linha de regressão, resultam numa elipse dos dados mais estreita e mais alongada, de modo que, quando a linha de regressão é reestimada, a nova linha se ajusta melhor aos pontos ajustados, e o R^2 é aumentado. Segundo Reis (2005):

“Na iteração seguinte, os p_i 's ajustados na iteração anterior substituem as estimativas iniciais de população na variável dependente e a equação de regressão é novamente estimada. Novos p_i 's ajustados são calculados e substituem os p_i 's atuais na iteração seguinte. As iterações continuam e o critério de parada pode ser definido em função de alguma medida de qualidade de ajuste como o coeficiente de determinação (R^2) ou o quadrado médio dos resíduos, por exemplo. As iterações terminariam quando uma destas medidas não tivesse mudanças consideradas relevantes de uma iteração para outra” (Reis, 2005).

Harvey (2002b) afirma que, no transcorrer do processo de iteração, o R^2 continua a crescer, porém com o decréscimo do incremento a cada iteração, até atingir um limite de R^2 menor que 1, servindo, segundo Harvey (2002b), de indicação do limite máximo de precisão do poder de predição que poderia ser atingida, em relação à distribuição da população dos pixels ocupados, de uma dada unidade espacial.

Estimativas negativas de população podem ocorrer devido a não restrição existente em modelos de regressão linear. Diante disso, procedimentos foram adotados no sentido de zerar os pixels que apresentam população negativa, e de ajustar a redistribuição da população nos pixels, para a manutenção da população total da unidade espacial (Harvey, 2002b). Pixels com baixíssima densidade (abaixo de uma pessoa) também foram desconsiderados por Harvey, ou seja, foram considerados como ausência de informação, procedimento que contribuiu para o aumento da eficácia do modelo.

Segundo Harvey (2002b), nos pixels classificados como residenciais, o aumento da densidade populacional está relacionado principalmente com maiores reflectâncias nas bandas 1 e 7, relacionado à predominância de superfícies construídas, e a menores reflectâncias na banda 5, padrão associado à predominância de superfícies naturais, como vegetação, de um modo geral. Já as bandas 2, 3 e 4 apresentam menor correlação com a mudança da densidade populacional (Harvey, 2002b).

Dentre diversos modelos testados, Harvey (2002b) conclui que a combinação linear das bandas das imagens Landsat TM foram as mais efetivas na construção dos modelos, em detrimento a variáveis explicativas mais complexas.

O menor erro relativo mediano encontrado para a população nas unidades espaciais foi de 14%. O modelo baseado em pixels se mostrou mais robusto que o modelo baseado em área, especialmente em áreas de densidade populacional extremas, embora a eficácia do modelo não seja a mesma nessas situações: ou seja, mesmo no modelo baseado em pixels, há tendência de subestimação nas áreas mais adensadas e de sobre estimacão nas áreas menos densas.

A subestimação pode estar associada a não identificação na imagem de estruturas com vários pavimentos (como prédios), algo esperado para imagens de média resolução espacial. Já as sobreestimativas nas áreas pouco adensadas tem provavelmente relação com o erro de classificação nas áreas rurais, como a definição de residências para estruturas não residenciais (como estruturas agrícolas e rodovias). Uma estratégia comentada por Harvey (2002b) para a melhoria da eficácia das estimativas seria a separação da classificação das áreas ocupadas para mais de um estrato, separando as amostras de treinamento para cada tipo de ocupação identificada, muito embora o próprio autor assuma que este tipo de procedimento envolve um grande custo adicional de tempo, além da complexidade deste tipo de procedimento.

Harvey (2002b) conclui ainda que o modelo gerado não deve ser validado para aplicação em qualquer área geográfica, ou para diferentes variações temporais ou culturais, sendo necessário pesquisar e incorporar em novos modelos às diferenças

de cada região, além da necessidade de realização de uma cuidadosa calibração nas classificações, nas fases da construção do modelo.

Wu e Murray (2007) compararam a abordagem por áreas e por pixels, para a estimativa da população da região de Columbus (Estado de Ohio, Estados Unidos), utilizando imagens do sensor Landsat 7 ETM+, para o ano de 2000. Os autores observaram que o erro relativo para a área urbana, calculado pelo método de pixels, foi de 1 a 7% menor quando comparado ao método baseado em área.

No Brasil, Reis (2005) aplicou as técnicas apresentadas em Harvey (2002a e 2002b), na estimação das populações por setores censitários de Belo Horizonte, para o ano de 1996, e com o auxílio do sensor TM Landsat 5 (bandas 1 a 5 e 7). Os resultados baseados na metodologia de Harvey (2002a) tiveram baixa performance, especialmente no nível micro. Ao aplicar o método de Harvey (2002b), constatou-se que os pixels classificados como urbanos tiveram pior desempenho no nível micro, com erro relativo mediano de 30,4% (contra 14% no trabalho de Harvey, 2002b), mas bem superior no nível agregado, com erro relativo total de -0,06% (menor que o encontrado em Harvey, 2002b).

3.3.2.5 Métodos multivariados

A literatura destaca ainda alguns métodos que utilizam de características físicas e socioeconômicas diversas para a produção de mapeamentos dasimétricos e de estimativas populacionais. Segundo Wu *et al.* (2005), áreas urbanas podem ser correlacionadas com a acessibilidade de sistemas de transporte, bases de logradouros (de arruamentos), dentre outras informações, sendo que a robustez desse tipo de abordagem multivariada tende a aumentar na medida em que se acrescentam informações, aumentando a complexidade dos modelos.

O trabalho de Zandbergen *et al.* (2010) faz uma referência à utilização de um conjunto de bases auxiliares para a elaboração de mapeamentos dasimétricos. Foram definidos quatro estados americanos (Connecticut, New Mexico, Oregon e South Carolina) para o estudo, que analisou três tipos de dados auxiliares (impermeabilidade, densidade de rodovias e luzes noturnas) para a comparação com métodos dasimétricos tradicionais, usualmente utilizados na literatura. Os

dados de cobertura da terra foram obtidos da base de dados *2001 National Land Cover Dataset* (NLCD), que consiste em informações de resolução espacial de 30 metros.

Segundo Zandbergen *et al.* (2010), o método de impermeabilidade apresentou muitas células classificadas como ocupadas, mas que constituem de estradas, o que tornou necessária a remoção de algumas características e feições lineares. As redes de estradas, por sua vez, foram obtidas da base de dados *Tiger 2000*, que constitui de um banco de informações vetoriais para o Estados Unidos. Esses arquivos consistem de polilinhas na escala de 1:100.000, com a codificação dos tipos de estradas. Os autores assumiram uma relação linear entre a densidade do arruamento e a densidade populacional, e foram utilizados apenas arruamentos locais, pois alguns tipos de rodovias não apresentaram correspondência com a população local. Já os dados sobre luzes noturnas foram obtidos através do projeto *Operational Linescan System* (OLS/DMSP), do qual foram utilizados os valores médios da luz noturna de intensidade estável, do ano 2000. Para o artigo, os autores aplicaram um relacionamento linear entre a intensidade dos valores e a população assumida.

Zandbergen *et al.* (2010) encontraram uma forte relação entre o método de cobertura da terra convencional e a informação sobre impermeabilidade, em parte por serem métodos baseados na mesma fonte de dados (as imagens Landsat). Contudo, o método de impermeabilidade apresentou informações mais detalhadas quando comparado às quatro classes urbanas do método de cobertura da terra. Os resultados indicam que a performance dos métodos dasimétricos variam fortemente entre as áreas de estudos, e que, portanto, não há uma única técnica que seja consistente para todas as regiões em análise, sendo que o mapeamento dasimétrico é mais difícil de ser trabalhado e calculado em áreas com baixa densidade populacional.

O próximo item apresenta o método multivariado utilizado pelo IBGE para o mapeamento dasimétrico da população do Censo 2010, e que constitui numa das fontes utilizadas na tese para a produção de estimativas via sensoriamento remoto.

3.3.2.6 A Grade Estatística do IBGE

Diante ao problema de incompatibilidade espacial e temporal entre as unidades geográficas para os quais os dados populacionais são disponibilizados, Bueno (2014) propõe uma metodologia híbrida para a produção de células regulares, para a agregação e disseminação de dados censitários, ao qual a autora denominou de “grade estatística”. Essa metodologia, aplicada no trabalho de Bueno (2014) para os estados de São Paulo e Pará, foi utilizada como referência para a disponibilização, por parte do IBGE (2016), da grade estatística para todo o território nacional. Este tópico sintetiza as principais características da grade estatística, pois a mesma será utilizada na elaboração dos resultados deste trabalho. Ressalta-se que o texto deste tópico foi inteiramente elaborado com base no relatório técnico do IBGE (2016), que descreve a metodologia de construção da grade estatística.

O conceito de grade estatística é exatamente o mesmo do mapeamento dasimétrico: a representação de dados espaciais independentes das divisões territoriais/administrativas, possibilitando a análise de dados em unidades geográficas pequenas, especialmente a comparação entre diferentes áreas, ou de uma mesma área, ao longo do tempo.

A utilização da grade estatística no Censo 2010 foi possível, graças à implementação de dispositivos eletrônicos de coleta, que permitiram a captação de coordenadas e a criação de um banco de dados de endereços, conectado ao mapeamento viário (IBGE, 2016). Enquanto nas áreas rurais foram utilizados pontos vetoriais de localização dos domicílios (oriundo do CNEFE), nas áreas urbanas foram utilizadas linhas vetoriais das faces de logradouros e dos limites dos setores censitários (oriundo da base territorial do Censo 2010).

Como informação auxiliar foram utilizadas classificações de uso e cobertura do solo, provenientes do Projeto de Monitoramento do Desmatamento dos Biomas Brasileiros por Satélite (PMDDBS), disponibilizado pelo Ministério do Meio Ambiente em 2007. Disponível para todo o Brasil, com exceção da Amazônia Legal, elas foram feitas a partir de imagens Landsat-5/TM. A segunda fonte provém do projeto TERRACLASS 2010, disponibilizado pela Empresa Brasileira de Pesquisa Agropecuária (Embrapa) e pelo Instituto Nacional de Pesquisas Espaciais (INPE).

Disponível para a Amazônia Legal, foi construído a partir de imagens Landsat-5/TM e do sensor MODIS (Moderate Resolution Imaging Spectroradiometer) [IBGE, 2016]. Nos casos das informações sobre uso e cobertura, as classes foram agrupadas entre “povoada” e “não povoada”, sendo a primeira relacionada a estruturas antrópicas e a segunda, a estruturas naturais (IBGE, 2016).

As informações selecionadas foram organizadas em grades de 1 km x 1 km para as áreas rurais e 200 m x 200 m para as áreas urbanas. Em alguns casos (especialmente em áreas rurais e isoladas), a malha viária estava incompleta ou sem codificação. Da mesma forma, nas áreas rurais algumas edificações não tiveram as coordenadas geográficas captadas, em função de problemas operacionais. Nos setores censitários com ausência de localização inferior a 50% foi adotada a técnica de agregação dos dados censitários, para a construção das grades estatísticas. Nos casos em que se constataram problemas nas malhas viárias ou nos pontos vetoriais (acima de 50% nos setores censitários), foram utilizadas técnicas de desagregação dos dados, o que, todavia, não foi suficiente para a representação do total de domicílios e pessoas levantados no Censo 2010 (foi constatada perda inferior a 5% dos dados censitários) [IBGE, 2016].

Segundo o IBGE (2016), setores censitários pequenos, com menos de 90% de sua área contida numa única célula da grade estatística, são incorporados diretamente nas células correspondentes. Porém, normalmente os setores censitários compreendem mais de uma célula. Nos casos onde há mais de 50% de informação sobre as malhas viárias e pontos de coordenadas (para as estruturas rurais) em relação à localização dos endereços, assumiu-se que a distribuição da população na face do logradouro é homogênea.

Já nas situações em que a ausência de informação sobre as malhas viárias e pontos (em relação ao vínculo com a base de endereços) supera os 50%, dois procedimentos são adotados, para os casos de presença e ausência de malhas viárias. Nos casos em que as malhas viárias estão disponíveis, “considera-se a densidade de domicílios por extensão de vias, por setores censitários, e a densidade é aplicada às células pertencentes aos setores correspondentes” (IBGE, 2016), o que corresponde a um mapeamento dasimétrico, no qual a distribuição

dos domicílios é homogênea no transcorrer das malhas, e cuja densidade varia conforme a distribuição das malhas viárias no espaço.

Nas situações de ausência das malhas viárias, foi adotado um método dasimétrico de classificação do uso e cobertura do solo (baseado no PMDBBS e no TERRACLASS), no qual se calcula a densidade de domicílios por unidade de área, a partir do total de domicílios por setor censitário e da área classificada como ocupada (IBGE, 2016). A partir da distribuição da densidade, são realizados procedimentos de interseção para alocação da densidade nas grades de células, e o volume de população foi obtido através da multiplicação do número de domicílios da célula pelo número médio de residentes do setor censitário correspondente (IBGE, 2015). Nos casos extremos, de ausência de malhas digitais, pontos de coordenadas e de uso e cobertura do solo, foi utilizado o método de desagregação de ponderação zonal simples, que distribui os domicílios de forma homogênea nos setores censitários, para posterior distribuição na grade estatística (IBGE, 2016).

Embora haja perda de informação na grade estatística (inferior à 5%, segundo o IBGE, 2016), essa limitação pode ser compensada pelo maior detalhamento do mapeamento dasimétrico, resultado do conjunto de métodos empregados para melhoria no mapeamento da distribuição dos domicílios, e, conseqüentemente, da população no espaço. Diante disso, a grade estatística foi utilizada neste trabalho para a produção de modelos, que serviram de base para o cálculo de estimativas para o ano de 2015, para os municípios selecionados.

O próximo capítulo apresenta a metodologia do trabalho, onde serão discutidos os métodos e técnicas utilizados em cada etapa da construção dos dados, bem como as estratégias adotadas para a análise da contribuição das imagens de média resolução espacial na produção de estimativas pós-censitárias, no nível municipal.

4 PROPOSTA METODOLÓGICA PARA A CONSTRUÇÃO E AVALIAÇÃO DE ESTIMATIVAS POPULACIONAIS PÓS-CENSITÁRIAS A PARTIR DE IMAGENS ORBITAIS DE MÉDIA RESOLUÇÃO ESPACIAL

O presente capítulo tem como objetivo apresentar os procedimentos utilizados para a produção de estimativas municipais em anos censitários e pós-censitários, via imagens orbitais de média resolução espacial Landsat 7 ETM+, bem como as estratégias adotadas na análise da contribuição desse tipo de informação para o campo das estimativas populacionais para pequenas áreas.

Nos últimos anos, tem se observado o aumento da disponibilização de imagens orbitais de alta resolução espacial, potencialmente mais eficazes se comparado a imagens orbitais de média resolução. Porém, não há, até o presente momento, imagens de satélite de alta resolução com acesso gratuito e disponível para uma série temporal que permita a aplicação dos procedimentos que serão propostos neste capítulo, como, por exemplo, a validação externa das estimativas. Em contraponto, imagens de média resolução espacial – com destaque para as imagens do Programa Landsat – possuem séries históricas que permitem análises temporais, e, diante a essa potencialidade, este trabalho propõe a utilização de imagens do sensor Landsat 7 ETM+, cobrindo um período de 15 anos (entre 2000 e 2015), contemplando, assim, os anos dos últimos dois censos.

Além disso, o tempo de processamento das imagens de alta resolução impede a produção de estimativas para áreas mais extensas e de forma sistemática. A menor área do campo de visada de cada cena das imagens de alta resolução torna inviável, na atualidade, a construção de grandes mosaicos de imagens, em relação ao tempo de processamento dessas informações. Seria necessário um número muito maior de imagens de alta resolução para cobrir uma mesma área, se comparado a imagens de menor resolução espacial, aumentando o tempo de processamento dos dados. Da mesma maneira, a maior quantidade de pixels (decorrente da maior resolução das imagens) acarretaria num tempo maior de

processamento, tornando inviável a produção de estimativas de forma sistemática e para áreas mais extensas (em que pese a possibilidade de produção de estimativas mais precisas) [Jensen, 2009]. Segundo Lu *et al.* (2006):

“Very high spatial resolution data, such as aerial photographs and IKONOS, often prove difficult with regard to data processing and analysis because of their huge data content and geometric distortions in a large area, while low spatial resolution data, such as DMSP OLS, cannot provide sufficiently detailed information for population estimation at the regional and local levels. Hence, in practice, medium spatial resolution data, such as Landsat TM/Enhanced TM (ETM+) imagery, have become the primary source of data for population estimation in recent years” (Lu *et al.*, 2006, p.3554).

Embora avanços tecnológicos possam superar, com o tempo, as limitações apontadas acima – como observado nos 10 anos seguintes ao trabalho de Lu *et al.* (2006) – a demanda pela produção sistemática de estimativas para pequenas áreas, replicáveis para grandes extensões de áreas, e as grandes limitações ainda existentes no processamento de imagens de alta resolução justificam a escolha, neste momento, por imagens de média resolução espacial. Como já fora colocado, as imagens de média resolução são amplamente disponibilizadas (de forma gratuita), possuem série histórica disponível e são consolidadas na literatura em relação à qualidade das estimativas que são produzidas.

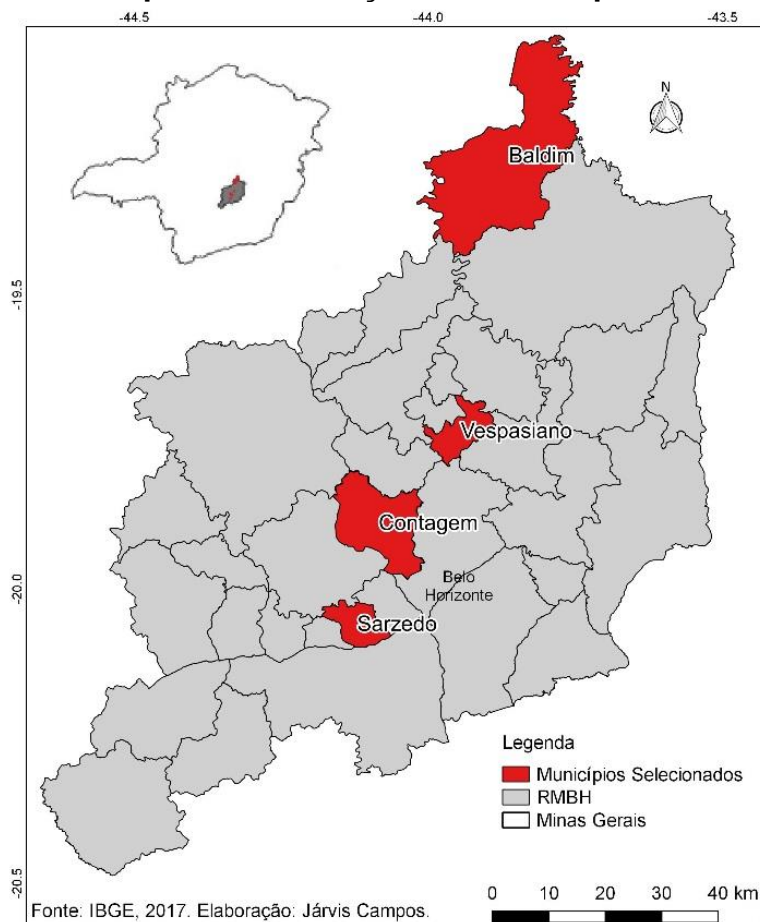
Com base na discussão realizada nos capítulos anteriores, tem-se justificada a utilização das imagens Landsat 7 ETM+, tanto para o mapeamento dasimétrico como para a construção dos modelos. Da mesma forma, os capítulos anteriores mostraram que os modelos selecionados para este trabalho (baseados em Harvey 2002a e 2002b) são referências importantes no campo das estimativas via sensoriamento remoto, apresentam resultados mais robustos em relação aos demais métodos, e há uma carência por novos estudos que utilizam modelos baseados em pixels, especialmente em países em desenvolvimento.

Os tópicos a seguir apresentam os procedimentos (métodos e técnicas) bem como as estratégias de apresentação dos resultados, para o alcance dos objetivos propostos. Definidas, de maneira geral, as fontes de dados que serão utilizadas no trabalho (pelo menos em relação às imagens de satélite Landsat 7 ETM+), bem como os modelos (baseados em Harvey, 2002a e 2002b), o item 4.1 justifica a

seleção dos municípios escolhidos para este trabalho. Já o item 4.2 decorre sobre o método dasimétrico adotado e os modelos utilizados para a produção das estimativas. Por fim, o item 4.3 apresenta a proposta metodológica e a estruturação (esquema/organização) dos resultados; ou seja, os anos selecionados, os dados utilizados para as validações interna e externa (das estimativas via sensoriamento remoto), e as comparações que serão realizadas entre diferentes tipos de estimativas (demográficas e via sensoriamento).

4.1 Critérios para a Seleção dos Municípios

Primeiramente, a escolha por municípios da RMBH deu-se em função do maior conhecimento do autor deste trabalho em relação às tendências demográficas e à dinâmica de uso e ocupação do solo presente nessa área de estudo. Já a seleção dos municípios selecionados teve, como critério, o tamanho e o ritmo de crescimento populacional. Um dos principais desafios nos modelos de estimativas de Harvey (2002a e 2002b) refere-se à subestimativa nas áreas com elevada densidade e a sobrestimativa nas áreas com baixa densidade populacional. Nesse sentido, entende-se que, ao analisar municípios com diferentes tamanhos de população, conseqüentemente estariam sendo selecionados municípios com diferentes padrões de distribuição espacial da população, e, conseqüentemente, diferentes níveis de densidade populacional. Dentre os 34 municípios da RMBH, foram selecionados quatro, a saber: Contagem, Vespasiano, Sarzedo e Baldim, mostrados na Figura 5.

FIGURA 5 – Mapa de Localização dos Municípios Seleccionados

Fonte: Elaboração Própria.

O município de Contagem foi aquele que apresentou, entre 2000 e 2010, o maior ritmo de crescimento populacional, dentre os dois municípios com população acima de 500 mil habitantes (grupo restrito a Contagem e Belo Horizonte). Com população de 603,4 mil habitantes, em 2010, e taxa de crescimento anual de 1,15%, na década de 2000, trata-se de um município metropolitano com características de ocupação bastante peculiares, que justificam sua seleção. Município conurbado com Belo Horizonte, Contagem tem como característica a presença de extensas áreas verticalizadas, com grande densidade populacional, somado à existência de bairros com menor verticalização e densidade, além de um importante complexo industrial (a Cidade Industrial).

Segundo Mennis (2003), um dos maiores problemas na produção de estimativas populacionais em áreas urbanas são os complexos industriais, cujo perfil de ocupação é marcado por residências esparsas, porém normalmente classificadas como de elevado grau de urbanização (ou seja, alta densidade populacional), em

função das densas malhas rodoviárias e das extensas áreas impermeáveis, que resultam em comportamentos espectrais condizentes com áreas de intensa ocupação.

Portanto, a característica heterogênea do espaço ocupado, somada às dificuldades associadas ao perfil industrial de Contagem, justificam a escolha desse município, seja pelo desafio da etapa de classificação (em grande parte devido à presença de indústria), seja pelo tipo de crescimento observado no município, marcado tanto pela expansão da mancha urbana, como pelo processo de verticalização. Além disso, a seleção de um município de grande adensamento representa o perfil de município mais explorado na literatura sobre estimativas por sensoriamento remoto.

O município de Vespasiano, por sua vez, foi o escolhido por ser o município entre 100 mil e 500 mil habitantes (104,5 mil habitantes em 2010) que apresentou a maior taxa de crescimento populacional no Estado, entre 2000 e 2010 (3,13% ao ano). A considerar que os modelos de Harvey são bem ajustados para municípios com acelerado ritmo de crescimento populacional, há uma maior possibilidade das diferenças nos resultados das estimativas (em relação a Contagem, por exemplo) estarem muito mais relacionadas às diferenças de densidade entre os municípios do que ao ritmo de crescimento de Vespasiano (dado que ambos apresentaram taxa de crescimento positiva superiores a 1% ao ano).

Ao norte e limítrofe com Belo Horizonte, Vespasiano pode ser considerado um município dormitório, com baixa densidade populacional, e que convive com bairros de baixo poder aquisitivo, além de chácaras (de maior poder aquisitivo), além de ocupações mais esparsadas. O recente e acelerado crescimento populacional desse município está associado ao desenvolvimento do “vetor norte” de Belo Horizonte, motivado pela duplicação da Linha Verde, que liga a capital e demais municípios ao Aeroporto Tancredo Neves (em Confins), e, principalmente, pela construção da Cidade Administrativa (em área de Belo Horizonte, vizinha do município de Vespasiano) que concentra os órgãos da administração do Estado de Minas Gerais.

Trata-se, portanto, de um município importante para o estudo, devido ao perfil recorrente do tipo de uso e ocupação de Vespasiano, comum a muitas regiões

metropolitanas (com a presença de ocupações do tipo chácaras, e de menor densidade populacional, se comparado a grandes centros urbanos). Esse tipo de ocupação representa um desafio para a etapa de classificação das áreas ocupadas, e conseqüentemente, ao mapeamento dasimétrico e produção das estimativas, sendo, portanto, um importante município a ser investigado.

Sarzedo representa o município abaixo de 100 mil habitantes com maior ritmo de crescimento populacional entre 2000 e 2010 (4,02% ao ano). Com uma população de 25,8 mil habitantes, Sarzedo trata-se de um município com densidade bem inferior, se comparado a Contagem e Vespasiano, além de ser um município de baixo contingente populacional, pouco estudado na literatura sobre estimativas por imagens orbitais. O rápido ritmo de crescimento observado nos últimos anos, somado ao tipo de expansão das áreas ocupadas (marcado pela expansão horizontal da mancha urbana), tendem a ser elementos que contribuam para a qualidade das estimativas.

Ademais, a investigação das estimativas para o município de Sarzedo pode lançar luzes sobre a eficácia das estimativas produzidas via sensoriamento remoto para municípios com baixo contingente populacional, especialmente se for considerado que, em relação ao crescimento da população, o comportamento observado em Sarzedo é bastante adequado ao modelo proposto. Como as estimativas populacionais obtidas por sensoriamento remoto costumam ser boas em municípios com crescimento positivo, a análise dos resultados das estimativas de Sarzedo poderá ser mais fortemente associada à baixa densidade populacional do que a problemas referentes à ausência de crescimento ou mesmo decréscimo. Sendo assim, Sarzedo é um município relevante para a análise da eficácia das estimativas produzidas através de imagens orbitais.

Por fim, Baldim foi o único município da RMBH que apresentou ritmo de crescimento negativo entre 2000 e 2010 (-0,3% ao ano). Com baixíssima densidade e pequeno contingente populacional, Baldim observou a redução da população, de 8.155 habitantes, em 2000, para 7.913 habitantes, em 2010. Os modelos propostos neste trabalho tendem, por hipótese, a não serem bem ajustados aos dados neste tipo de situação, por uma razão simples: o fato de a população decrescer num município, num dado período, não pode ser captado por imagens de sensores

remotos, pois a saída de pessoas não representa, por hipótese, uma mudança no uso do solo. Assim, uma estrutura de ocupação observada no início do período (por exemplo, uma casa) não tende, em princípio, a sofrer mudança para um uso representativo de ausência de ocupação (como, por exemplo, solo exposto ou vegetação).

O decréscimo de população pode ser eventualmente resultante de um fluxo emigratório (como normalmente ocorre em áreas de perda populacional), que tenha causado, num dado período, um saldo líquido migratório negativo. No entanto, pode ter havido, embora em menor número, a chegada de imigrantes, que, por sua vez, pode resultar na expansão da área ocupada de um município. A chegada de imigrantes, ou quaisquer outros fatores associados à dinâmica imobiliária ou econômica do município, podem resultar na expansão da mancha urbana (ou de ocupação), mesmo num contexto de decréscimo populacional, o que resultaria, por hipótese, em estimativas positivas de crescimento populacional, mesmo a tendência sendo a de redução da população residente, num dado período.

Contudo, mesmo ponderando todas essas questões, entende-se que o teste é válido, pois, à parte da mancha urbana que tende, no curto prazo, a aumentar de tamanho (ou se manter estática), mas não à sua redução, o cálculo das estimativas dos modelos de Harvey é baseado nas reflectâncias dos pixels, que, podem, eventualmente, apresentar mudanças de comportamento decorrentes da perda populacional.

Portanto, ainda que a hipótese seja a de ausência de assertividade (e de refutação dos resultados), até mesmo por uma questão de pressuposto em relação ao comportamento da variável sintomática (das imagens), cuja mancha de ocupação restringe-se ao crescimento, ao longo do tempo, algumas ponderações devem ser consideradas. Este procedimento ainda não foi testado, de modo que a análise deste tipo de município é um desafio para o modelo proposto, e pode trazer algum indício de associação entre decréscimo populacional e os modelos propostos, contribuindo para a avaliação da eficácia das estimativas calculadas via sensoriamento remoto.

Portanto, os municípios selecionados representam não são apenas diferentes tipos de tamanho e densidade populacionais, como também diferentes ritmos de crescimento populacional. Além disso, os municípios selecionados possuem tipos de uso e ocupação do solo bastante diferenciados, assim como a dinâmica de expansão das áreas ocupadas; elementos que, somados, contribuem para a análise da eficácia das estimativas produzidas por imagens Landsat 7 ETM+. O próximo item analisa o método dasimétrico, a estruturação dos resultados e os modelos adotados para o cálculo das estimativas.

4.2 Bases de Dados, Mapeamento Dasimétrico, Estruturação dos Resultados e os Modelos Adotados para o Cálculo das Estimativas

4.2.1 Mapeamento Dasimétrico e Construção das Bases de Dados

As bases de dados utilizadas para as etapas de mapeamento dasimétrico foram: shapes (ou camadas vetoriais) correspondentes aos setores censitários dos censos de 2000 e 2010; shapes da grade estatística do IBGE de 2010; população total por setores censitários (obtido do Universo dos Censos de 2000 e 2010); população total por grades estatísticas (publicação do IBGE, 2015); e duas cenas do satélite Landsat 7 ETM+, para os anos de 2000, 2010 e 2015.

Com o objetivo de reduzir o tempo de processamento, as imagens foram cortadas aos limites dos municípios, gerando 12 novas imagens correspondentes aos quatro municípios em estudo e três anos de análise (2000, 2010 e 2015). A considerar que a literatura indica que os modelos de regressão devem representar cada município individualmente, o corte das imagens ao limite dos municípios atende a este requisito, além de diminuir o tempo de processamento no SIG. Ressalta-se que as

imagens de satélite foram obtidas no site *Earth Explorer*, da *United States Geological Survey* (USGS)¹¹.

A USGS possui um projeto, denominado *Science Processing Architecture* (ESPA), que disponibiliza imagens de satélite com os dados de reflectância de superfície, o que dispensou a necessidade de cálculo das reflectâncias para a construção dos modelos de regressão. Além disso, as imagens do projeto ESPA possuem correção geométrica; ou seja, são registradas e disponibilizadas em formato “GeoTiff”, no sistema WGS84.

Porém, as imagens do projeto ESPA foram utilizadas apenas para a obtenção das reflectâncias de superfície, necessárias para a construção das variáveis independentes dos modelos de regressão. Para a etapa anterior de classificação (que corresponde à definição das áreas ocupadas) foram utilizadas as imagens originais disponibilizadas pela USGS, que, por sua vez não dispõem de correção atmosférica nas imagens; limitação esta que pode, em alguns casos, interferir na etapa de classificação. Como se optou pela não realização do procedimento de correção atmosférica, foram selecionadas imagens com baixo percentual de nuvens, abaixo de 10%, com o objetivo de contornar possíveis interferências na identificação das manchas de ocupação humana.

O detalhamento dessas informações, tais como o número de setores, o número de grades estatísticas e o número de pixels classificados como ocupados por município, bem como a data das imagens Landsat 7 ETM+, serão apresentados no próximo capítulo, referente aos resultados do trabalho.

Para a construção dos modelos e das estimativas, a primeira etapa consistiu na classificação das imagens Landsat 7 ETM+, no qual foram identificadas as classes “áreas ocupadas”, “água”, “vegetação” e “solo exposto”. Foram mapeadas as áreas ocupadas para cada um dos quatro municípios, e para os três anos em análise.

¹¹ Para tanto, foram utilizados os softwares de Sistemas de Informação Geográfica QGIS 2.16.3 e Grass 7.0.4, para todas as etapas de processamento digital de imagens e de mapeamento dasimétrico. Já o software estatístico R 3.4.0 foi utilizado para as etapas de construção dos modelos de regressão.

Para tanto, foi aplicado o método de classificação supervisionada Maxver, para 4 classes (áreas ocupadas, vegetação, água e solo), para as 12 imagens selecionadas (4 municípios e 3 períodos, 2000, 2010 e 2015), a partir do plugin SCP, do software QGIS. Em seguida, foi realizada uma segunda etapa de classificação, que consistiu na interpretação manual de imagens de satélite de alta resolução espacial (da Google), para a totalidade das áreas estudadas e para os anos analisados, o que permitiu um importante ganho de qualidade na delimitação das áreas de ocupação humana.

De posse dessas informações, a etapa seguinte consistiu nos mapeamentos dasimétricos da população, que se deu a partir da distribuição da população na mesma escala espacial das imagens Landsat ETM+, para as áreas classificadas como de ocupação humana (ou residencial). A partir da distribuição da população dos municípios na mesma escala em relação às imagens (ou seja, no nível dos pixels), foram construídas bases de dados georreferenciadas e compatibilizadas, entre as informações de população e as reflectâncias de cada banda espectral das imagens Landsat (bandas 1 a 5 e 7). Na etapa seguinte, essas informações serviram de base para a construção dos modelos, que foram construídos conforme a literatura apresentada no capítulo anterior (item 3.3.2.4, tendo como referência as metodologias desenvolvidas por Harvey 2002a e 2002b).

Para a realização dessa etapa, foram criadas grades de células (no software QGIS 2.18.3), em formato vetorial, georreferenciadas em relação aos pixels das imagens Landsat 7 ETM+ – ou seja, cada célula da grade sobrepõe com precisão cada pixel da imagem – e limitadas às manchas de ocupação urbana e rural. Esse procedimento foi realizado para as 12 imagens classificadas, referentes aos quatro municípios selecionados e três períodos em análise (2000, 2010 e 2015).

Definidas as grades de células, os setores censitários de 2000 e 2010 e as grades estatísticas (referentes a 2010) foram sobrepostos às malhas vetoriais das grades de células, e foram utilizadas ferramentas de consulta espacial (no QGIS 2.18.3) para a compatibilização e alocação das células vetoriais aos respectivos setores. A população informada em cada setor foi, então, redistribuída de forma homogênea no conjunto de células classificadas como ocupadas e pertencentes a cada setor,

a partir de procedimentos de consulta espacial (assim como proposto em Harvey, 2002a e 2002b).

O critério adotado para a correspondência entre as grades células e os setores foi a utilização do centróide das grades de células, de modo que cada célula fosse relacionada somente a um setor censitário. Embora esse procedimento não seja tão preciso quanto redistribuir proporcionalmente a área das células pertencentes a cada setor, entende-se que o método utilizado atende aos propósitos do trabalho, sob o pressuposto de compensação entre as células localizadas nas áreas limítrofes aos setores (ora captadas pelo setor censitário de referência, ora pelos setores vizinhos), em função da localização dos centroides.

Assim como no caso da população, a partir da sobreposição das grades vetoriais com as imagens de satélite, procedimentos de consulta espacial permitiram a alocação dos dados de reflectância nas grades de células, o que possibilitou, ao final dessa etapa, a obtenção de dados de população e das reflectâncias das bandas espectrais no nível das grades de células (georreferenciadas e com a mesma resolução dos pixels das imagens Landsat ETM+). Resume-se que procedimentos de consulta espacial similares foram utilizados para a desagregação das informações de situação (urbano/rural) e aglomerados subnormais no nível dos pixels das imagens de satélite. Ao final, os bancos de dados criados foram utilizados para a construção dos modelos de regressão e das estimativas populacionais.

4.2.2 Estruturação dos Resultados

Num primeiro momento, os bancos de dados mencionados no item anterior foram criados para os quatro municípios selecionados, para a construção de modelos baseados em dados de 2000 e 2010 e para o cálculo de estimativas para os anos de 2010 e 2015 (como será detalhado mais adiante).

Como já fora comentado, para os mapeamentos dasimétricos baseados nos dados de 2000, foram utilizados os setores censitários do Censo 2000, que corresponde à menor unidade espacial disponibilizada pelo IBGE, para divulgação dos resultados do Universo, do Censo 2000. Assim, para cada município foram criados

dois modelos a partir de dados de 2000: no nível dos setores censitários (como em Harvey, 2002a) e no nível dos pixels (como em Harvey, 2002b), e que serviram para a produção de estimativas para os anos de 2010 e 2015. Porém, não foram construídos os modelos no nível dos setores (como em Harvey 2002a) para os municípios de Sarzedo e Baldim, devido ao baixo número de setores censitários, o que impossibilitou a construção dos modelos de regressão. O baixo número de setores para esses dois municípios é agravado por problemas nas malhas digitais do IBGE, e que será tema de discussão no capítulo de resultados.

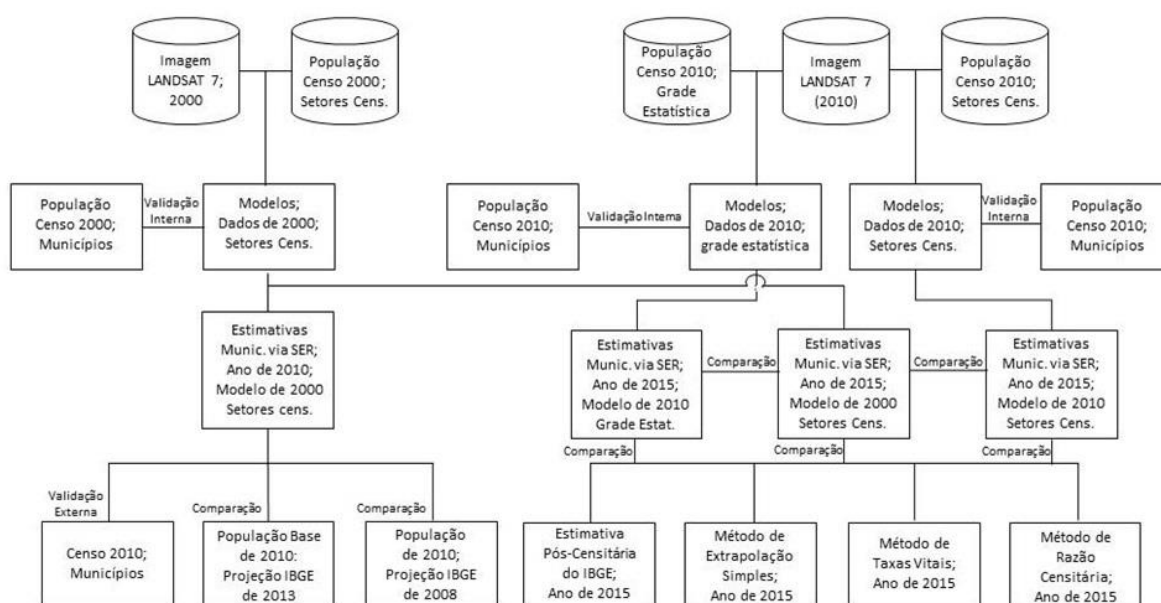
Para o ano de 2010, o mapeamento dasimétrico foi elaborado primeiramente a partir dos setores censitários do Censo 2010 e, num segundo momento, a partir das grades estatísticas do IBGE. Assim, foram construídos três modelos baseados em dados de 2010 (utilizados para o cálculo de estimativas para o ano de 2015), a saber: no nível dos setores censitários (como em Harvey, 2002a), no nível dos pixels, a partir da redistribuição dos dados dos setores censitários (como em Harvey, 2002b), e no nível dos pixels, porém a partir da redistribuição dos dados da Grade Estatística do IBGE (segundo a mesma lógica de Harvey, 2002b). Da mesma forma, em função do baixo número de setores censitários, os modelos no nível dos setores (como em Harvey, 2002a) não serão construídos para os municípios de Sarzedo e Baldim.

Portanto, as estimativas construídas para os anos de 2010 e 2015 foram baseadas nos dois métodos de Harvey. No método proposto em Harvey (2002a) – que utiliza a média das reflectâncias dos pixels pertencentes a uma dada unidade espacial (neste trabalho, a unidade espacial refere-se aos setores censitários e à grade estatística) – foram calculadas para os municípios de Contagem e Vespasiano. Como fora comentado, os municípios de Sarzedo e Baldim possuem poucos setores censitários, o que impossibilitou o cálculo das estimativas através do método de setores de Harvey (2002a). Já o método por pixels de Harvey (2002b) foi utilizado para o cálculo de estimativas dos quatro municípios em estudo.

A Figura 6 apresenta o diagrama de fluxo de dados do trabalho. Ressalta-se que no diagrama não foram apresentados: os shapes (camadas vetoriais) utilizados no trabalho; os detalhes referentes às imagens de satélite; a descrição dos modelos utilizados (se Harvey 2002a e 2002b); bem como os municípios selecionados. Este

diagrama tem, contudo, o objetivo de apresentar a estrutura geral do fluxo de dados, com foco principal: no ano a que se referem as fontes de dados de população (2000 ou 2010); no ano com o qual os modelos foram construídos; no caso dos dados de 2010, se a unidade espacial se trata do setor censitário ou da grade estatística; no caso das estimativas, o ano em que a população foi estimada e o ano do modelo (se calculado através de modelo produzido a partir de dados de 2000 ou de 2010); e as fontes de dados utilizadas para a validação externa e comparação dos resultados produzidos via sensoriamento remoto.

FIGURA 6 - Diagrama de Fluxo de Dados



Fonte: Elaboração Própria.

A primeira linha do diagrama corresponde às fontes de dados, enquanto na segunda linha são apresentados os modelos baseados em 2000 e 2010. Construídos os modelos, é realizada uma etapa de validação interna, a partir da seleção de amostras aleatórias dos próprios dados utilizados para os ajustes dos modelos. A terceira linha mostra as estimativas para o ano de 2010 (baseado em modelos construídos com dados de 2000), e as estimativas para o ano de 2015 (a partir de modelos construídos com dados de 2000 e 2010); sendo que, no caso dos modelos de 2010, ainda distintos entre aqueles produzidos por setores censitários e por grade estatística do IBGE. Ressalta-se, a exceção dos modelos e estimativas produzidos via grade estatística (que são restritos ao método de Harvey 2002b), os demais modelos e estimativas, baseados em setores censitários, foram

produzidos nos dois formatos: no nível dos setores censitários (como em Harvey 2002a) e no nível dos pixels (como em Harvey 2002b).

Porém, um aspecto ainda não comentado sobre os resultados, e presente na quarta linha do diagrama, diz respeito às estratégias adotadas para a análise da contribuição do sensoriamento remoto na produção de estimativas populacionais. Esse trabalho propõe comparar todas as estimativas produzidas via sensoriamento remoto com um conjunto de informações, baseadas em censos, projeções e estimativas oficiais, além de estimativas produzidas a partir de variáveis sintomáticas.

Os resultados das estimativas calculadas para o ano de 2010 foram comparados, num primeiro momento, com a população do Censo 2010, com o objetivo de se realizar uma “validação externa temporal” (ou seja, entre fontes de anos distintos), procedimento este ainda não realizado na literatura, dentre os trabalhos que utilizam dos métodos de Harvey (2002a e 2002b). Basta lembrar que a validação externa das estimativas calculadas por Harvey foi realizada para o mesmo ano, enquanto que a etapa de validação do trabalho de Reis (2005) foi restrita à validação interna.

Porém, essa etapa de validação externa apresenta limitações, pois, na literatura sobre estimativas via sensoriamento remoto, devem-se considerar sempre as especificidades de cada área e região: ou seja, modelos devem ser contruídos para cada área de estudo, em função das características de uso e cobertura de cada localidade, levando-se em consideração que o comportamento dos sensores remotos em resposta aos diferentes usos pode se alterar ao longo do tempo. Em que pese todos esses elementos, será utilizado a terminologia “validação externa” para a comparação das estimativas via sensoriamento remoto de 2010 (construídas a partir de modelos baseados em dados de 2000) com os dados de população do Censo 2010, porém, ressaltando as limitações aqui comentadas, seja de subenumeração, seja das limitações relacionadas ao uso de imagens de satélite de diferentes anos.

Num segundo momento, as estimativas por sensoriamento remoto, calculadas para o ano de 2010 (a partir de modelos construídos com dados de 2000), foram

comparadas com estimativas produzidas através da desagregação das projeções por Unidades da Federação do IBGE (revisões 2008 e 2013) para o nível dos municípios, através do método AiBi, para o ano de 2010. Como já fora comentado, não é objetivo, a partir dessa comparação, definir qual seria o “melhor modelo”, pois, em nenhum momento desse trabalho pretende-se cometer o erro de associar a produção de estimativas por variáveis sintomáticas com o conceito de previsão populacional. Contudo, a comparação das estimativas produzidas por sensoriamento remoto com estimativas baseadas em projeções anteriores ao Censo 2010 (no caso da revisão de 2008) contribui para a análise da eficácia das imagens orbitais de média resolução na produção de estimativas. Um bom ajuste das estimativas via sensoriamento remoto, por exemplo, pode ser um indicativo da potencialidade da utilização das imagens de satélite como informação auxiliar, na análise qualitativa e preliminar para a construção de modelos de projeção, especialmente em situações em que os dados censitários não estão disponíveis, bem como uma ferramenta para a produção de estimativas em períodos inter e pós-censitários.

Já as estimativas produzidas para o ano de 2015 (tanto aquelas oriundas de modelos baseados em dados de 2000 como de 2010), foram comparadas a quatro fontes de dados, a saber: estimativas pós-censitárias do IBGE; método de Extrapolação Simples; método das Taxas Vitais; e método da Razão Censitária.

Embora as fontes de dados de 2015 sejam restritas a estimativas, dada a ausência de pesquisas censitárias – o que, por esse motivo, não permite chamar de validação externa as comparações entre estimativas de 2015 – a comparação entre diferentes métodos pode contribuir para a análise da eficácia das estimativas produzidas via imagens orbitais. Resultados muito similares, por exemplo, podem ser analisados à luz dos custos relativos ao tempo necessário para a aplicação de cada tipo de método, especialmente se os resultados dos métodos propostos neste trabalho forem muito similares às estimativas pós-censitárias do IBGE. Ao mesmo tempo, resultados muito díspares podem ser um indicativo de problemas no método, para um dado município, por exemplo.

O método de Extrapolação Simples foi escolhido para a comparação com as estimativas via sensoriamento remoto produzidas para 2015, em função do

baixíssimo custo de tempo necessário para o cálculo, e sob a hipótese de que este método é adequado à proposta do trabalho, qual seja a de produzir estimativas para intervalos curtos de tempo (em períodos pós-censitários). O método de Taxas Vitais, por sua vez, foi escolhido para comparação devido aos avanços alcançados nos últimos anos, em relação à diminuição da subenumeração dos registros de nascimentos e óbitos. Por fim, o método da Razão Censitária foi escolhido para a comparação com as estimativas via sensoriamento remoto produzidas para 2015 por se tratar de um método que utiliza apenas uma variável sintomática. Para este método foi escolhida informações sobre o número de matrículas do Ensino Fundamental, obtidas no Censo Escolar, motivado pelo fato do Brasil já ter alcançado há alguns anos a universalização do Ensino Fundamental.

Portanto, foram selecionadas variáveis sintomáticas já consolidadas na literatura como sendo as que apresentam os melhores resultados (como fora observado no Capítulo 2), e a partir de métodos de fácil manipulação. É importante ressaltar que, embora o método da Correlação de Razões seja o que apresenta os melhores resultados na literatura (dentro dos métodos que utilizam variáveis auxiliares), ele demanda um maior número de variáveis sintomáticas. Para a análise deste trabalho, optou-se por métodos que utilizassem apenas uma variável auxiliar, para que os resultados pudessem ser analisados à luz das limitações da variável sintomática proposta.

4.2.3 Modelos Adotados para o Cálculo das Estimativas via Sensoriamento Remoto

O modelo de Harvey (2002b) foi escolhido para este estudo primeiramente devido a sua flexibilidade em relação às unidades espaciais, pois estimativas produzidas pelo método dos pixels podem ser comparadas ao longo do tempo. Em segundo, e igualmente importante, o modelo de Harvey (2002b) supera a limitação da distribuição homogênea da população dentro das unidades espaciais. Como já fora discutido, este método redistribui a população nos pixels (num processo iterativo, que cessa quando o nível de ajuste do modelo é estabilizado), mantendo o total da população observado para cada unidade espacial. Já o modelo proposto por Harvey (2002a) foi considerado apenas para os municípios de Contagem e

Vespasiano, dado que Sarzedo e Baldim possuem poucos setores censitários, o que impede a construção de modelos de regressão.

O Quadro 2 mostra o conjunto de variáveis utilizadas para a construção dos modelos deste trabalho, para todos os anos, considerando o modelo de Harvey (2002a). Já o Quadro 3 apresenta as variáveis do modelo de Harvey (2002b), e cujas diferença principal diz respeito à unidade de análise, neste caso no nível do pixel. Observa-se que, no modelo de Harvey (2002a) a variável dependente é a densidade populacional (ou ainda o logaritmo da densidade), pois apresenta melhor ajuste com as médias das reflectâncias, se comparado à população. Já no modelo de Harvey (2002b), a variável dependente é a população.

Ressalta-se que além das bandas das imagens Landsat 7, foram ainda consideradas nos modelos as variáveis “percentual de área ocupada” (para os modelos oriundos de Harvey 2002a), situação (urbano/rural, para os modelos de Harvey 2002b) e “aglomerado subnormal” (para os modelos oriundos de Harvey 2002a e 2002b); sendo variáveis que representam medidas de ocupação que contribuem para um melhor ajuste dos modelos, como será discutido nos resultados.

QUADRO 2 – Descrição das Variáveis utilizadas para a construção de modelos baseados em Harvey (2002a)

Siglas (nos modelos)	Variáveis	Fonte
Dens00	Densidade do Setor Censitário em 2000	IBGE
Dens10	Densidade do Setor Censitário em 2010	IBGE
Log(Dens)00	Logaritmo da Densidade do Setor Censitário em 2000	IBGE
Log(Dens)10	Logaritmo da Densidade do Setor Censitário em 2010	IBGE
TM1	Média da Reflectância na banda 1 (Landsat 7 ETM+)	USGS
TM2	Média da Reflectância na banda 2 (Landsat 7 ETM+)	USGS
TM3	Média da Reflectância na banda 3 (Landsat 7 ETM+)	USGS
TM4	Média da Reflectância na banda 4 (Landsat 7 ETM+)	USGS
TM5	Média da Reflectância na banda 5 (Landsat 7 ETM+)	USGS
TM7	Média da Reflectância na banda 7 (Landsat 7 ETM+)	USGS
AreaOcup00	Percentual de Área Ocupada em 2000	USGS/IBGE
AreaOcup10	Percentual de Área Ocupada em 2010	USGS/IBGE
AgSub00	Aglomerado Subnormal em 2000	IBGE
AgSub10	Aglomerado Subnormal em 2010	IBGE

Fonte: Elaboração Própria.

QUADRO 3 – Descrição das Variáveis utilizadas para a construção de modelos baseados em Harvey (2002b)

Siglas (nos modelos)	Variáveis	Fonte
Pop00pixel(setor)	População do pixel em 2000	IBGE
Pop10pixel(setor ou grade)	População do pixel em 2010	IBGE
TM1	Reflectância do pixel na banda 1 (Landsat 7 ETM+)	USGS
TM2	Reflectância do pixel na banda 2 (Landsat 7 ETM+)	USGS
TM3	Reflectância do pixel na banda 3 (Landsat 7 ETM+)	USGS
TM4	Reflectância do pixel na banda 4 (Landsat 7 ETM+)	USGS
TM5	Reflectância do pixel na banda 5 (Landsat 7 ETM+)	USGS
TM7	Reflectância do pixel na banda 7 (Landsat 7 ETM+)	USGS
Urbano00	Situação (Urbano/Rural) em 2000	IBGE
Urbano10	Situação (Urbano/Rural) em 2010	IBGE
AgSub00	Aglomerado Subnormal em 2000	IBGE
AgSub10	Aglomerado Subnormal em 2010	IBGE

Fonte: Elaboração Própria.

Para cada município e ano foi testado 7 modelos (no caso dos modelos baseados em Harvey 2002a) e 6 modelos (no caso daqueles baseados em Harvey 2002b). Segue abaixo a relação dos modelos baseados em Harvey (2002a):

- Modelo 1: Var. Dependente: Densidade; Var. Explicativas: bandas 1 a 5 e 7;
- Modelo 2: Var. Dependente: Logaritmo da Densidade; Var. Explicativas: bandas 1 a 5 e 7;
- Modelo 3: Var. Dependente: Logaritmo da Densidade; Var. Explicativas: bandas 1 a 5 e 7 e percentual de área ocupada;
- Modelo 4: Var. Dependente: Logaritmo da Densidade; Var. Explicativas: bandas 1 a 5 e 7, percentual de área ocupada e aglomerado subnormal;
- Modelo 5: Var. Dependente: Logaritmo da Densidade; Var. Explicativas: bandas 1, 4, 5 e 7, percentual de área ocupada e aglomerado subnormal;
- Modelo 6: Var. Dependente: Logaritmo da Densidade; Var. Explicativas: bandas 1, 4 e 7, percentual de área ocupada e aglomerado subnormal;
- Modelo 7: Var. Dependente: Logaritmo da Densidade; Var. Explicativas: bandas 1, 4 e 7, percentual de área ocupada, aglomerado subnormal e três variáveis de interação entre as bandas 1, 4 e 7 e aglomerado subnormal;

Portanto, primeiramente foram testados modelos com a incorporação de todas as bandas, tal como observado em Harvey (2002a). Porém, os melhores ajustes foram verificados em modelos cuja variável dependente é o logaritmo da densidade. Assim, os modelos 2 a 7 incorporam essa alteração, sendo que no modelo 3 é

incorporada a variável de percentual de população ocupada (que, em muitos casos, melhorou o poder de predição dos modelos, como será discutido no capítulo de resultados). No modelo 4 é incorporada a variável de aglomerado subnormal, no modelo 5 são retiradas as bandas 2 e 3, no modelo 6 é retirada a banda 5 e, por fim, no modelo 7 são incorporadas variáveis de interação entre as bandas 1, 4 e 7 com aglomerado subnormal, com o objetivo de verificar a influência da favela no efeito que cada banda tem sobre a densidade.

A correlação linear é alta entre as reflectâncias das bandas 1, 2 e 3, bem como entre as bandas 5 e 7, enquanto que a banda 4 foi a que apresentou as menores correlações com as outras bandas, como será comentado nos resultados. Como afirma Reis (2005) “estes relacionamentos entre as variáveis independentes podem causar problemas de multicolinearidade, afetando a estimação dos coeficientes do modelo”. De um modo geral, a retirada das bandas 2, 3 e 5 resultaram em modelos mais bem ajustados, em Reis (2005).

Segue abaixo a relação dos modelos baseados em Harvey (2002b):

- Modelo 1: Var. Dependente: População; Var. Explicativas: bandas 1 a 5 e 7;
- Modelo 2: Var. Dependente: População; Var. Explicativas: bandas 1 a 5 e 7 e aglomerado subnormal;
- Modelo 3: Var. Dependente: População; Var. Explicativas: bandas 1 a 5 e 7, aglomerado subnormal e situação (urbano/rural);
- Modelo 4: Var. Dependente: População; Var. Explicativas: bandas 1, 4, 5 e 7, aglomerado subnormal e situação (urbano/rural);
- Modelo 5: Var. Dependente: População; Var. Explicativas: bandas 1, 4 e 7, aglomerado subnormal e situação (urbano/rural);
- Modelo 6: Var. Dependente: População; Var. Explicativas: bandas 1, 4 e 7, aglomerado subnormal, situação (urbano/rural) e três variáveis de interação entre as bandas 1, 4 e 7 e aglomerado subnormal;

De forma similar, os modelos baseados em Harvey (2002b) partem do modelo 1, que apresenta todas as bandas como variáveis explicativas (tal como na proposta original de Harvey, 2002b). O modelo 2 é a incorporação da variável referente a aglomerado subnormal, o que geralmente melhor o poder de predição dos modelos,

enquanto o modelo 3 incorpora a variável urbano/rural. O modelo 4 corresponde à exclusão das bandas 2 e 3, enquanto o modelo 5 refere-se à exclusão da banda 5. Por fim, o modelo 6 incorpora três variáveis de interação entre as bandas 1, 4 e 7 e aglomerado subnormal.

No próximo capítulo serão apresentados quais modelos se ajustaram melhor a cada situação (além dos coeficientes dos modelos com melhor ajuste). A equação 4.1 apresenta o modelo teórico (baseado no ano de 2000). Ressalta-se que a equação 4.1 serve tanto para o modelo baseado em Harvey (2002a) como em Harvey (2002b). Em Harvey (2002a), as variáveis de sensoriamento remoto são as médias das reflectâncias de cada banda e a variável dependente é a densidade do setor censitário. Em Harvey (2002b), as variáveis de sensoriamento remoto são as reflectâncias de cada pixel e a variável dependente é a população. No caso dos modelos baseados em Harvey (2002b), deve-se retirar a variável relativa ao percentual de área ocupada e incluir a variável de situação (urbano/rural).

$$Pop00_i = \beta_0 + \beta_1 TM1_i + \beta_2 TM2_i + \dots + \beta_7 TM7_i + ArOcup00_i + AgSub00_i + \varepsilon \quad (4.1)$$

Onde: Utilizando como referência Harvey (2002a), $Pop00_i$ é a densidade populacional do setor censitário i ; β são parâmetros a serem estimados a partir dos dados; $TM1_i$ ao $TM7_i$ são as médias das reflectâncias das bandas do sensor Landsat 7, no setor censitário i ; $ArOcup00_i$ é o percentual de área ocupada no setor censitário i ; $AgSub00_i$ discrimina se o setor censitário i é aglomerado subnormal; e ε é o erro aleatório do modelo. Assume-se que o erro é independente, apresenta distribuição normal e variância constante.

As medidas de erro utilizadas para a análise da validação interna foram o Erro Relativo Mediano, o R^2 (back) e o Erro Total. O Erro Relativo Mediano corresponde à mediana dos valores absolutos dos erros relativos, observado para cada setor censitário (no caso de Harvey 2002a), e observado para cada pixel (no caso de Harvey 2002b). A segunda medida foi o R^2 (back), que corresponde ao quadrado do coeficiente de correlação linear entre as estimativas de população para os setores (ou pixels) e os valores populacionais reais desses setores (ou pixels), sendo que, quanto mais próximo de um (1), maior o ajuste do modelo. Por ser calculado no nível dos setores, o R^2 (back) está mais ligado ao Erro Relativo Mediano do que ao Erro Total, e, portanto, tende a apresentar valores baixos (como verificado na literatura). A terceira medida trata-se do Erro Relativo Total (citado de forma recorrente no Capítulo 6 como “erro total”), que representa a variação do total

estimado em relação ao total observado para o conjunto de setores (ou pixels); ou seja, para o município como um todo. Essa medida corresponde ao que Harvey (2002a, 2002b) denomina de “Erro Macro”.

Para a validação externa, bem como para as comparações com estimativas oficiais e demográficas que foram produzidas, foi adotada a diferença percentual (que nada mais é do que o Erro Relativo Total, ou “erro total”), tendo como referência a diferença das estimativas via sensoriamento remoto em relação às populações do Censo 2010 (no caso da validação externa, em 2010); e a diferença das estimativas via sensoriamento remoto em relação às estimativas pós-censitárias oficiais do IBGE (para o ano de 2015).

Apresentada a metodologia do trabalho, o próximo capítulo discute os resultados, onde serão analisados os mapeamentos dasimétricos, os modelos, as estimativas calculadas a partir dos modelos, e as demais fontes de dados, utilizadas para a validação e comparação com as estimativas produzidas via sensoriamento remoto.

5 ESTIMATIVAS POPULACIONAIS A PARTIR DE IMAGENS ORBITAIS DE MÉDIA RESOLUÇÃO ESPACIAL: RESULTADOS PARA OS MUNICÍPIOS SELECIONADOS

Este capítulo apresenta os resultados conforme a proposta metodológica descrita no capítulo anterior, sendo organizado em quatro tópicos, referentes a cada município (selecionado para o estudo): Contagem, Vespasiano, Sarzedo e Baldim.

A organização dos resultados por municípios e não por etapas de construção dos dados ou ainda por métodos adotados para o cálculo das estimativas se justifica, pois, os modelos criados e as estimativas calculadas apresentam intrínseca relação com a configuração de cada município (em relação ao tamanho e densidade populacional, e às etapas de construção dos dados). Assim, será possível observar a forte associação entre as fontes de dados, os mapeamentos dasimétricos e as estimativas calculadas para cada município em particular.

Os tópicos referentes aos municípios apresentam, num primeiro momento, toda a etapa de mapeamento dasimétrico e construção dos bancos de dados, seguido pelos resultados dos modelos no nível dos setores censitários, e, por fim, dos modelos no nível dos pixels. O próximo tópico, referente ao município de Contagem, é o que descreve de forma mais detalhada as diversas etapas de elaboração dos dados, dos modelos e das estimativas, bem como dos indicadores utilizados. O maior detalhamento das informações apresentadas em Contagem tem como objetivo contribuir para uma melhor compreensão de toda a estrutura dos resultados do trabalho, especialmente em relação aos demais municípios, cuja estrutura (e formas de representação dos resultados) segue a mesma lógica.

5.1 Contagem

Entre 1991 e 2010, o município de Contagem experimentou um incremento de pouco mais de um terço da população (34,2%, de 449,5 mil para 603,4 mil pessoas), porém, em paralelo, houve uma redução no ritmo de crescimento

populacional, passando de 2% ao ano na década de 1990, para 1,15% na década seguinte. A tendência observada encontra-se em consonância com o estágio atual da transição demográfica no Brasil, marcado rapidíssima redução das taxas de fecundidade.

Enquanto a mortalidade infantil caiu de 28,5 para 15,4 por mil habitantes entre 2000 e 2010, a probabilidade de sobrevivência até os 60 anos subiu de 79,8% para 82,4%. Já a Taxa de Fecundidade Total caiu de 2,4, em 1991, para 2,0 em 2000 e 1,77 em 2010 (dados da PNUD, 2017). Já as migrações apresentaram tendência à maior oscilação, dado que apresentou tendência de redução nas décadas de 1990 e 2000. Embora o saldo migratório quinquenal não seja uma medida de estoque das migrações ocorridas no decênio, o saldo pode ser útil para uma análise geral das mudanças no comportamento migratório, e foi possível observar a redução dos saldos positivos entre os quinquênios 1995-2000 (26.701) e 2005-2010 (22.333), o que representou uma redução de 4,96 para 3,70 na Taxa Líquida de Migração (respectivamente aos dois quinquênios).

O comportamento das componentes demográficas e a tendência recente de redução do ritmo de crescimento populacional, somado ao tamanho populacional do município de Contagem, permite inferir que Contagem continuará crescendo no período 2010-2015, porém com a continuidade da redução do ritmo de crescimento, diante à ausência de fatores econômicos locais e regionais relevantes, inclusive no nível nacional, a considerar que a crise econômica observada nos últimos anos teve seu agravamento a partir de 2016.

5.1.1 Mapeamento Dasimétrico

Com o objetivo de produzir as estimativas para Contagem, a primeira etapa consistiu na aquisição de imagens Landsat 7 ETM+ para os anos de 2000, 2010 e 2015. Foram selecionadas imagens com percentual de nuvens abaixo de 10%, e com datas as mais próximas da data de referência dos Censos de 2000 e 2010 (1º de agosto), dado que a base para a produção dos modelos bem como a principal

fonte de comparação das estimativas produzidas foram as informações censitárias¹².

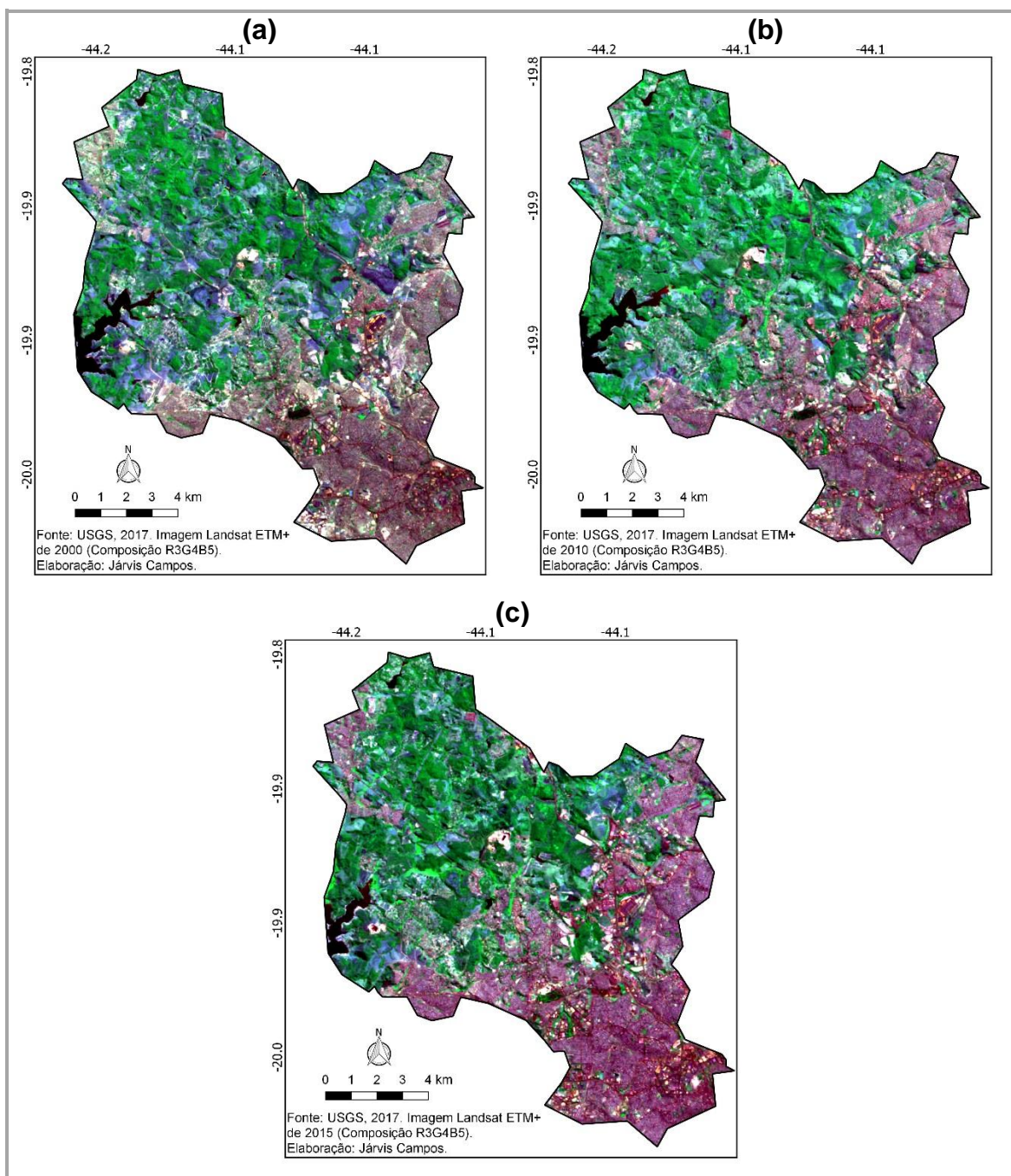
A Figura 7 apresenta a composição R3G4B5 das imagens Landsat 7 ETM+ para Contagem nos anos 2000, 2010 e 2015, onde se observa a concentração urbana nas porções sudeste e leste do município (onde predomina a cor roxa), justamente nas áreas que são conurbadas com o município de Belo Horizonte. Essas áreas conturbadas são, portanto, marcadas por um maior adensamento urbano, com destaque para a porção sudeste do município, que conta com a presença da “Cidade Industrial” de Contagem.

Nas porções central, sudoeste e norte do município, há a predominância de áreas de baixa densidade populacional, marcadas por áreas rurais e de ausência de ocupação (onde predominam as cores verdes associadas à vegetação, bem como a cor azul correspondente ao solo exposto); sendo que, na porção sudoeste, está localizada a represa Vargem das Flores, marcada pela presença de chácaras e de ocupações de baixa densidade. Vale destacar ainda a porção noroeste, marcada por uma mancha de ocupação urbana isolada da sede municipal, denominada “Nova Contagem”, e cuja expansão urbana está associada principalmente à presença da penitenciária Nelson Hungria.

Contagem apresenta, portanto, uma configuração de ocupação bastante heterogênea, com a presença de áreas industriais, áreas verticalizadas de grande adensamento, em paralelo a áreas rurais e bairros de baixo adensamento, o que torna um desafio a produção de estimativas para essa localidade.

¹² Para os quatro municípios em análise as datas foram as mesmas (13 de agosto de 2000, 08 de julho de 2010 e 07 de agosto de 2015), sendo que os municípios de Contagem, Vespasiano e Sarzedo pertencem à mesma cena (código 218074), enquanto o município de Baldim pertence à cena de código 218073.

FIGURA 7 – Composições R3G4B5 de Imagens Landsat 7 ETM+ para os anos de 2000 (a), 2010 (b) e 2015 (c) – Município de Contagem

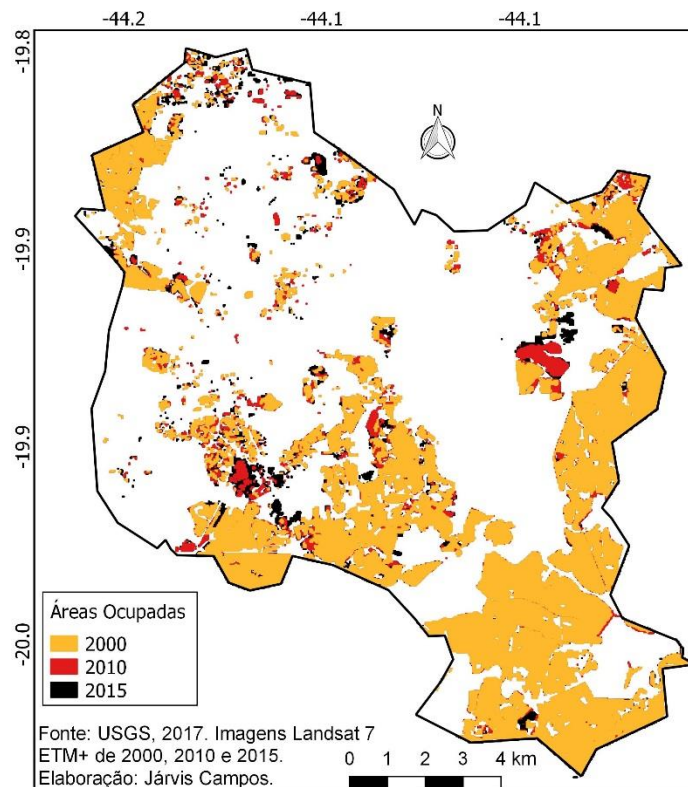


Fonte: USGS, 2017. Elaboração Própria.

A composição R3G4B5 foi utilizada para a coleta das amostras de treinamento, para a classificação das áreas ocupadas. Já a Figura 8 mostra a evolução das manchas de ocupação para os anos 2000, 2010 e 2015, resultado não apenas da etapa de classificação automática por pixels (Maxver), mas, também, da etapa de interpretação manual de imagens de satélite de alta resolução (Google), para toda

a área estudada e os 3 anos analisados, esforço este fundamental para a validação e aprimoramento das classificações automáticas.

FIGURA 8 – Evolução das Manchas de Ocupação (2000, 2010 e 2015) a partir da classificação de Imagens Landsat 7 ETM+ – Município de Contagem



Fonte: USGS, 2017. Elaboração Própria.

Sem dúvida, a interpretação de imagens – realizada para os quatro municípios – aumentou sobremaneira o tempo de processamento das informações, a um nível que comprometeria, por exemplo, a replicação dessa metodologia para áreas mais extensas; o que é, sem dúvida, um aspecto importante quando se considera a produção de estimativas via sensoriamento remoto para o campo da Demografia. Embora a escolha, neste trabalho, por imagens gratuitas e de média resolução tenha forte relação com a idéia de testar estimativas baseadas em métodos que utilizam de sensoriamento remoto e que sejam replicáveis para áreas mais extensas – com o potencial de contribuir, numa escala regional, para o campo das estimativas e das projeções demográficas – para este trabalho optou-se por um processamento dispendioso em tempo, na classificação das áreas ocupadas.

O objetivo foi o de excluir os erros das estimativas a problemas/limitações associados à etapa de classificação (das manchas de ocupação), cuja parcela de

erro não seria passível de mensuração. Portanto, o maior processamento e melhor captação das áreas ocupadas permitiu associar os resultados das estimativas às limitações relacionadas à resolução espacial das imagens Landsat 7 ETM+, bem como ao perfil dos municípios quanto ao tamanho populacional, densidade e perfil de ocupação, ou ainda à mudança da relação entre população e reflectância ao longo do tempo; sem que essas análises fossem prejudicadas pelas limitações associadas à etapa de classificação.

Contudo, como já fora discutido no capítulo 3, ressalta-se que a literatura internacional indica que os erros da classificação automática por pixels não prejudicam sobremaneira as estimativas no nível municipal, dado que os erros por excesso (em especial a classificação de solo exposto como sendo urbano) são em alguma medida compensados pelas limitações por falta, relacionadas à incapacidade das imagens Landsat ETM+ de selecionar de forma diferenciada os tipos de ocupação de elevada densidade, como as áreas verticalizadas (de intensa ocupação).

Em relação à expansão da mancha de ocupação de Contagem, a Figura 8 permite observar a expansão nas porções leste (marcada por áreas mais verticalizadas), e principalmente nas regiões sudoeste e norte (marcadas por áreas de menor adensamento). Ressalta-se que na porção sudeste não fora considerada uma área aparentemente definida como ocupada (na Figura 7, quando se observa a composição R3G4B5), pois, essa porção do território corresponde à área industrial do município de Contagem.

Em 2000, 2010 e 2015 foram classificados, respectivamente, 57.174, 63.857 e 69.565 pixels como áreas de ocupação humana, o que representou um crescimento de 11,7% no período 2000-2010 e 8,9% no período 2010-2015. A considerar que 2010-2015 representa a metade do período anterior, constata-se que o ritmo de crescimento horizontal da mancha de ocupação neste quinquênio foi superior ao observado na década de 2000. Como consequência, em 2000, a mediana do percentual de áreas ocupadas por setores censitários foi de 83,1%, passando para 85,3% em 2010 e 86% em 2015.

Todavia, essas informações não permitem afirmar que o ritmo de crescimento populacional em 2010-2015 foi superior ao período 2000-2010, pois os pixels ocupados não refletem o processo de verticalização do município, que, eventualmente, pode ter sido mais intenso na década de 2000, principalmente a considerar que a expansão dos pixels no período 2010-2015 ocorreu principalmente em áreas de baixo adensamento populacional.

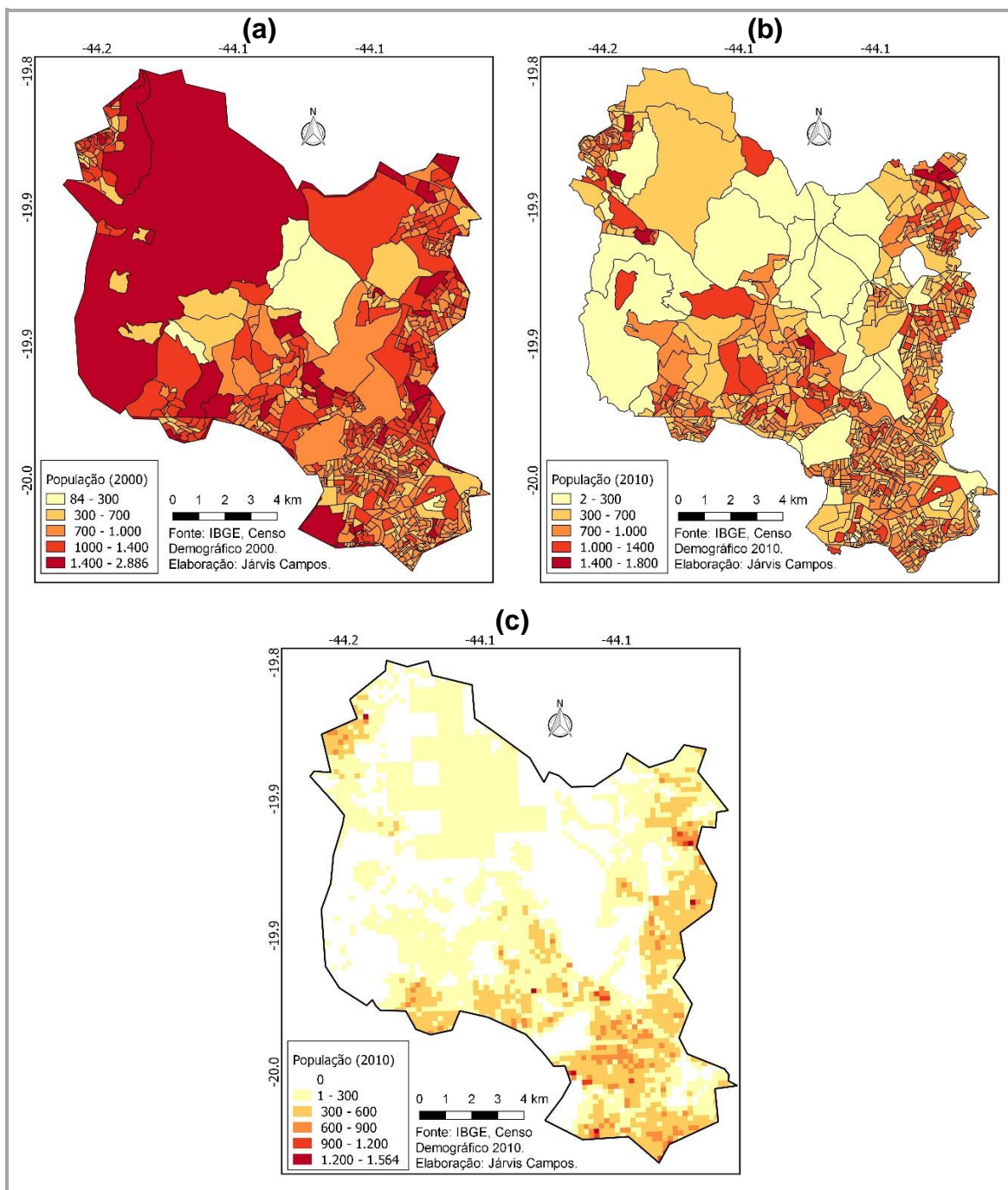
Dentre o total de pixels classificados como ocupados no período, 95,5% estão localizados em setores censitários urbanos no ano 2000, 96,5% em 2010 e 95,2% em 2015¹³. Ressalta-se que, diferentemente do trabalho de Reis (2005), os setores rurais não foram excluídos dos bancos de dados para a produção das estimativas, o que permitiu a incorporação da variável “situação (urbano/rural)” no estudo¹⁴. Já em relação aos aglomerados subnormais, em 2000, apenas 5% dos pixels estão localizados em favelas, percentual que aumentou para 5,6% em 2010 e 5,4% em 2015.

A Figura 9 mostra os setores censitários de 2000 e 2010, além da grade estatística de 2010, e a distribuição das populações nessas unidades espaciais. Como a população de cada setor censitário tende a variar pouco, o aglomerado de setores censitários pequenos em área representam, naturalmente, as áreas de maior adensamento.

¹³ A classificação dos pixels como urbano/rural e aglomerado subnormal em 2015 foi definida a partir da relação entre a localização dos pixels e os setores censitários de 2010.

¹⁴ A variável “situação (urbano/rural)” foi testada para os modelos no nível dos pixels (baseados em Harvey, 2002b), pois, nos modelos no nível dos setores censitários (baseados em Harvey, 2002a), além da variável “aglomerado subnormal”, fora testada a variável “percentual de área ocupada por setor” (esta última própria do modelo baseado em setores).

FIGURA 9 – Distribuição da População por Setores Censitários de 2000 (a), Setores Censitários de 2010 (b) e Grades Estatísticas de 2010 (c), a partir de Dados do Censo 2010 – Município de Contagem



Fonte: IBGE, Censos Demográficos 2000 e 2010. IBGE, 2015. Elaboração Própria.

O principal aspecto a ser analisado – que é elemento fundamental na qualidade dos modelos e das estimativas produzidas – trata da quantidade de setores censitários: nos modelos baseados em Harvey (2002a), a média das reflectâncias será relacionada à densidade populacional em cada setor censitário, e quanto maior o número de setores censitários (e consequentemente menor a área desses

setores), melhor (mais precisa) será a relação entre as reflectâncias e a população. Isso ocorre porque a média das reflectâncias dos pixels nos setores menores se torna um melhor representante do que acontece em termos de ocupação humana do setor, pois a variabilidade tende a diminuir com menos pixels no setor censitário. Portanto, quanto maior o número de setores e menor a área desses setores, melhor será a relação entre o nível da reflectância das bandas e o tamanho populacional, e, por consequência, maior será o poder de predição dos modelos.

Em Contagem é possível observar uma grande quantidade de setores censitários, tanto em 2000, com 594 setores (dos quais foram utilizados um total de 583), como em 2010, com 884 setores (tendo sido utilizados 863 setores). Foram considerados apenas os setores comuns e os aglomerados subnormais, tendo sido excluídos os setores definidos pelo IBGE como quartéis e bases militares, alojamentos/acampamentos, embarcações, aldeias indígenas, penitenciárias, além de asilos, orfanatos, hospitais, conventos, dentre outros (ou seja, estruturas de adensamento populacional que tendem a prejudicar a relação entre reflectância e população).

A malha de setores censitários do Censo 2000, disponibilizada gratuitamente no site do IBGE, não contempla toda a área do município, ficando excluída grande parte das áreas rurais. Na Figura 9a, a grande área a noroeste do município de Contagem na realidade corresponde a um polígono referente à área rural, porém não disponibilizada pela malha digital do IBGE, o que dificultou a compatibilização do banco de dados; pois, nessa grande área, foram classificadas áreas ocupadas e que necessitavam de associação com população (ou densidade) para a construção dos modelos. Diante disso, foi calculada a diferença entre o total da população do município (em 2000) e a soma da população referente aos setores censitários inseridos na malha digital de setores de 2000. A diferença foi alocada nessa área inicialmente vazia (não coberta pela malha de setores censitários), e finalmente inserida no banco de dados como um novo setor. Portanto, parte-se do pressuposto que a melhor qualidade da malha de setores censitários de 2010 possibilita a construção de modelos e de estimativas mais precisas em relação a 2000, seja por fornecer todos os setores existentes nos municípios, seja por apresentar um número maior de setores.

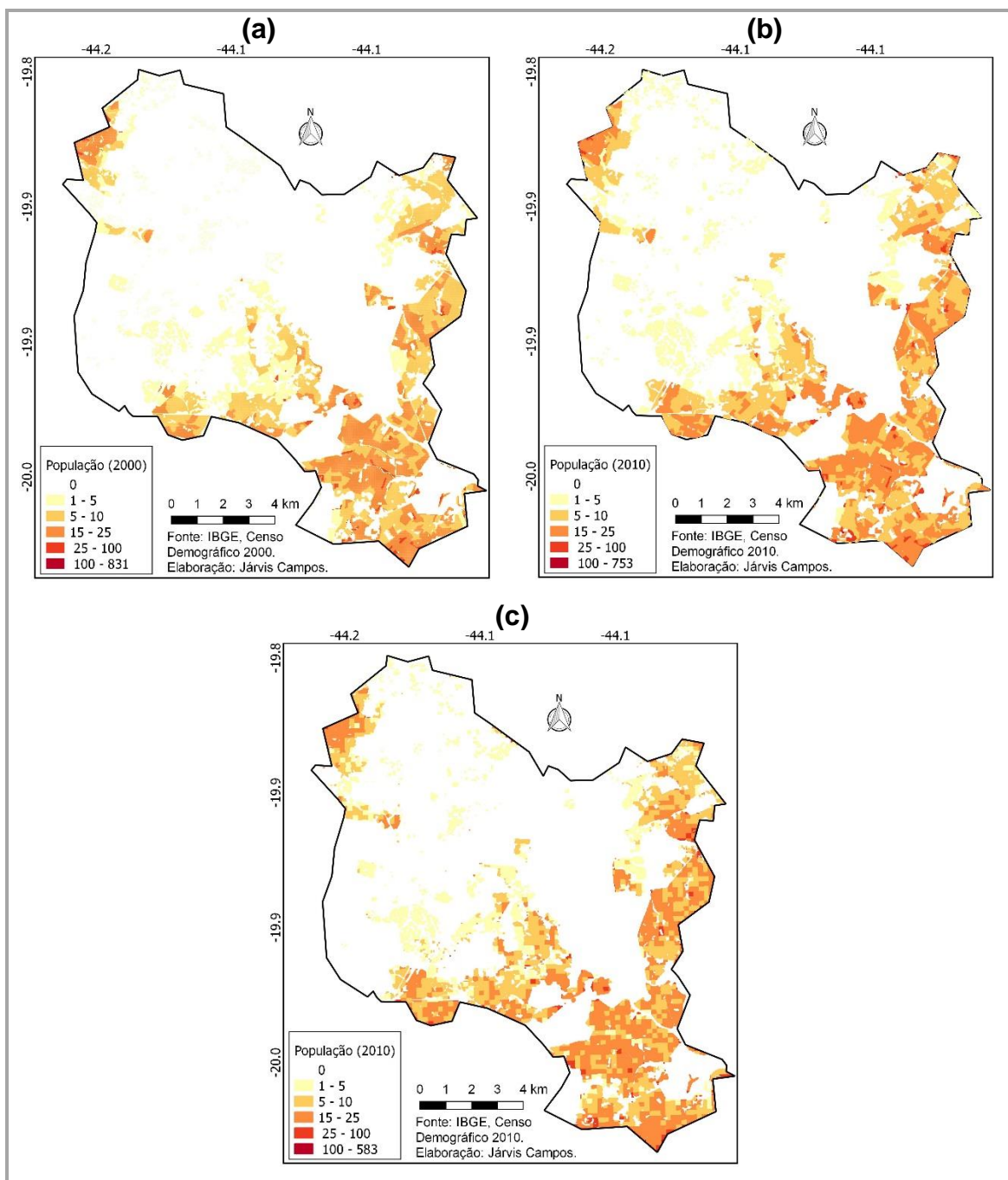
Já em relação às grades estatísticas, essas somaram 4.462 áreas no município de Contagem, número cinco vezes superior à quantidade de setores censitários identificados em Contagem em 2010. Por hipótese, mesmo se considerar a distribuição espacial mais precisa da população, oriunda das grades estatísticas, como essa fonte de dados será utilizada para a construção no nível dos pixels – que, por sua vez, contempla o processo iterativo de redistribuição da população (inicialmente distribuída de forma homogênea a partir da área maior) – espera-se que os resultados entre as estimativas de 2015, produzidas por setores e pela grade estatística, não apresentem resultados muito díspares. Infelizmente, por conta da disponibilização das grades estatísticas somente para o ano de 2010, a análise dos resultados será restrita à validação interna e à análise comparativa com outras estimativas demográficas, para o ano de 2015¹⁵.

No que diz respeito à distribuição da população nos pixels (ou mapeamento dasimétrico), a mediana da população por pixel foi de 8,6 pessoas em 2000, ao passo que em 2010 o número aumentou para 8,8, considerando a distribuição a partir de setores censitários, e 9, na distribuição através da grade estatística. Os resultados mostram um leve aumento na densidade dos pixels entre 2000 e 2010, fenômeno que pode (eventualmente) estar associado ao processo de verticalização.

A Figura 10 apresenta os mapeamentos dasimétricos oriundos dos setores censitários de 2000 e 2010, e da Grade Estatística de 2010, no qual é possível observar uma grande similaridade (ao menos visual) nas distribuições de população, considerando o mesmo intervalo de classe (a mesma legenda) para as três fontes de dados.

¹⁵ Como comentado na metodologia, a validação externa é realizada somente para os modelos construídos com dados de 2000, pois as estimativas produzidas para 2010 (a partir dos modelos criados com dados de 2000) são comparadas à população do Censo 2010. O Censo 2010, por sua vez, não se trata de estimativa e sim de pesquisa censitária (a partir de coleta de dados em campo), o que permite a denominação de “validação externa” as diferenças percentuais entre as estimativas de 2010 (que partem de modelos criados em 2000) via Sensoriamento remoto e a população do Censo 2010. Já as estimativas produzidas para o ano de 2015 (tanto por modelos criados com dados de 2000 como de 2010) serão comparadas com outras estimativas demográficas (cada qual com seus pressupostos e limitações), não sendo, portanto, considerado “validação externa”, mas tão somente uma “análise comparativa” entre as estimativas.

FIGURA 10 – Mapeamento Dasimétrico por meio dos Pixels de Imagens Landsat 7 ETM+ e de Dados dos Setores Censitários do Censo 2000 (a), Setores Censitários do Censo 2010 (b), e da Grade Estatística de 2010 (c) – Município de Contagem



Fonte: IBGE, Censos Demográficos 2000 e 2010. IBGE, 2015. USGS, 2017. Elaboração Própria.

5.1.2 Modelos e Estimativas no nível dos Setores Censitários

Para a definição das variáveis explicativas e dos modelos no nível dos setores censitários, primeiramente foi verificada a correlação entre as bandas das imagens

Landsat ETM+ selecionadas. A Tabela 1 apresenta a matriz de correlação entre as médias das reflectâncias das bandas da imagem recortada para o município de Contagem, para o ano de 2000, e a densidade populacional por setores censitários. Verifica-se a baixa correlação entre densidade e todas as bandas (com as maiores correlações observadas entre densidade e as bandas 1 e 4), a alta correlação entre as bandas vizinhas (como era esperado, em função da sobreposição existente entre as faixas do espectro), com destaque para a alta correlação entre as bandas 1, 2 e 3 (próximas a 1). Ressalta-se que em 2010 o comportamento das correlações entre densidade e bandas foi bastante similar.

TABELA 1 – Matriz de Correlação entre Densidade e as Bandas 1 a 5 e 7: Modelo Baseado em Setores Censitários de 2000 –Contagem

	Densidade	banda 1	banda 2	banda 3	banda 4	banda 5	banda 7
Densidade	1,000	0,101	0,038	-0,003	0,122	-0,064	0,031
banda 1	0,101	1,000	0,923	0,820	0,478	0,548	0,689
banda 2	0,038	0,923	1,000	0,958	0,731	0,769	0,824
banda 3	-0,003	0,820	0,958	1,000	0,808	0,857	0,872
banda 4	-0,122	0,478	0,731	0,808	1,000	0,908	0,750
banda 5	-0,064	0,548	0,769	0,857	0,908	1,000	0,928
banda 7	0,031	0,689	0,824	0,872	0,750	0,928	1,000

Fonte: Elaboração Própria.

O modelo 1 considera todas as bandas (1 a 5 e 7), assim como no artigo original de Harvey (2002a). Já o modelo 2 considera a transformação logarítmica da variável dependente (densidade), o que aumenta o poder de predição dos modelos, como observado em Harvey (2002a) e Reis (2005). Já o modelo 3 considera a inserção da variável “percentual de área ocupada”, enquanto o modelo 4 incorpora a variável “aglomerado subnormal”, com o objetivo de testar a eficiência dessas variáveis na produção de estimativas. O modelo 5 testa a exclusão das bandas 2 e 3 (altamente correlacionadas), enquanto o modelo 6, já partindo da exclusão dessas duas bandas, testa a exclusão da banda 5 (altamente correlacionada com a banda 4). Por fim, o modelo 7, partindo do modelo 5 – que se restringe às bandas 1, 4 e 7, mais percentual de área ocupada e aglomerado subnormal – testa a incorporação de termos de interação entre as três bandas (1, 4 e 7) e aglomerado subnormal, com o objetivo de verificar a influência dos aglomerados subnormais (favelas) no efeito que a reflectância em cada banda tem sobre a densidade populacional. Ressalta-se que esses modelos propostos foram utilizados não

apenas em Contagem, mas também para o município de Vespasiano, mantendo-se a mesma estrutura¹⁶.

Para a construção dos modelos, 2/3 dos bancos de dados foram definidos de forma aleatória para o ajuste de cada modelo, enquanto os outros 1/3 foram utilizados para o teste referente à validação interna. Ressalta-se que este critério foi adotado para todos os municípios, e todos os modelos (baseados em Harvey 2002a e 2002b). Assim, no caso dos modelos no nível dos setores e construídos a partir de dados dos setores censitários de 2000, foram considerados 389 setores para o ajuste dos modelos e 194 setores para a etapa de validação interna; enquanto que, para os modelos no nível dos setores e construídos a partir dos setores censitários de 2010, foram utilizados 575 setores para o ajuste dos modelos e 288 para a validação interna.

5.1.2.1 Validação Interna

A Tabela 2 apresenta os resultados para os modelos no nível dos setores, construídos a partir de dados de 2000 e 2010 (Censos 2000 e 2010 e imagens para esses anos).

¹⁶ Para os municípios de Sarzedo e Baldim não foram trabalhados os modelos no nível dos setores censitários (baseados em Harvey, 2002a), como será discutido mais adiante.

TABELA 2 – Validação Interna dos Modelos no nível dos Setores Censitários (2000 e 2010) - Contagem

Ano (base de dados)	Indicadores	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Modelo 7
	Var. Dependente = Var. Explicativas	Dens. = Bandas 1 a 5 e 7	Log. (Dens.) = Bandas 1 a 5 e 7	Log. (Dens.) = Bandas 1 a 5 e 7, (%) de Área Ocup.	Log. (Dens.) = Bandas 1 a 5 e 7, (%) de Área Ocup. e Agl. Subnormal	Log. (Dens.) = Bandas 1, 4, 5 e 7, Agl. Subnormal e (%) de Área Ocup.	Log. (Dens.) = Bandas 1, 4 e 7, Agl. Subnormal e (%) de Área Ocup.	Log. (Dens.) = Bandas 1, 4 e 7, Agl. Subnormal e (%) de Área Ocup. e Interação (Produto) de Agl. Subnormal com as Bandas 1 (v1), 4 (v2) e 7 (v3)
2000 (setor)	R ² (back)	0,061	0,032	0,232	0,323	0,322	0,341*	0,341
	Erro relat. mediano	0,361	0,399	0,346	0,349	0,330	0,340*	0,321
	Erro total (%)	115,70	44,40	10,52	6,56	6,81	6,29*	6,18
2010 (setor)	R ² (back)	0,262	0,179	0,281	0,322	0,335	0,341*	0,341
	Erro relat. mediano	0,316	0,386	0,305	0,308	0,292	0,300*	0,298
	Erro total (%)	25,82	17,42	-0,26	-1,82	-1,76	-1,5*	-1,52

* Modelo adotado para o cálculo das estimativas.

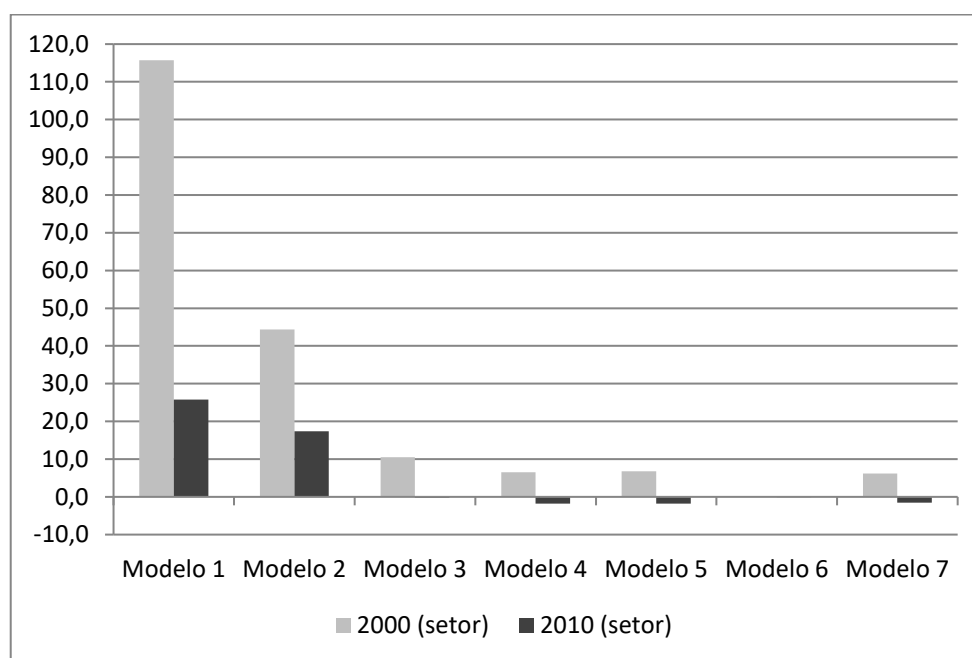
Fonte: Elaboração Própria.

De maneira geral, o R^2 (back) foi baixo para todos os modelos, tanto em 2000 como em 2010, sendo que os melhores valores foram alcançados pelo modelo 6 (0,341 para 2000 e 2010). É importante lembrar que, por ser calculado no nível dos setores, o R^2 (back) está mais ligado ao erro relativo mediano dos Setores do que ao Erro Total. No Erro total (nível macro), setores em que a população foi superestimada acabam sendo compensados por setores em que a população foi subestimada. Como o erro total trabalha com a soma das populações estimadas, os erros individuais dos setores acabam sendo diluídos. O R^2 (back) não tem essa chance de "compensação de erros", pois o cálculo "enxerga" os valores individuais dos setores.

Já em relação ao erro relativo mediano dos setores censitários (ERM), foi possível observar pouca diferença entre os modelos, tanto em 2000 (entre 32,1% e 39,9%) como em 2010 (entre 29,2% e 38,6%), com os melhores resultados encontrados no modelo 2, em 2000 e 2010.

Em relação à validação interna, o Gráfico 1 apresenta a comparação do erro total entre os modelos 1 a 7 (no nível dos setores) e entre os modelos construídos a partir de 2000 e 2010.

GRÁFICO 1 – Erro Total dos Modelos baseados em Setores Censitários (2000 e 2010) - Contagem



Fonte: Elaboração Própria.

Na validação interna, os modelos que obtiveram os melhores resultados foram o modelo 7 (erro total de 6,2%, baseado em dados de 2000) e o modelo 3 (erro total de -0,3%, a partir de dados de 2010), o que pode ser considerado um resultado satisfatório, especialmente no caso de 2010.

É possível notar que a utilização do logaritmo na variável dependente reduziu drasticamente o erro total (de 115,7% para 44,4% em 2000 e de 25,8% para 17,4% em 2010). Da mesma maneira, a incorporação da variável “percentual de área ocupada” (modelo 3) melhorou sobremaneira os resultados, para níveis de aproximadamente 10,5% em 2000 e -0,26% em 2010. A incorporação da variável “aglomerado subnormal” reduziu de 10,5% para 6,6% o erro total em 2000, enquanto em 2010 aumentou o erro de -0,26% para -1,82%, o que, em todo caso, manteve o baixo erro total observado. Em intensidade bem menor, a exclusão das bandas 2, 3 e 5 e a exclusão dos termos de interação (do modelo 7) não melhoraram significativamente o poder de predição do modelo.

No caso de 2000, embora o menor erro total tenha sido observado no modelo 7 (6,2%), o modelo 6 apresentou um erro total muito similar (6,3%), e sem a necessidade de incorporação das variáveis de interação entre as bandas e aglomerado subnormal. Diante disso, o modelo 6 foi utilizado para o cálculo das estimativas, para a validação externa (ano de 2010) e para a comparação com outras estimativas demográficas (ano de 2015). A equação 5.1 apresenta o modelo escolhido.

$$\begin{aligned} \text{Log}(\text{Dens})00\text{setor}_i = & 6,66 + 13,18\text{TM}1_i + (-5,46)\text{TM}4_i + 1,06\text{TM}7_i + \\ & 2,75\text{AreaOcup}00_i + 0,73\text{AgSub}00_i + \varepsilon \end{aligned} \quad (5.1)$$

Já em relação a 2010, embora o modelo 3 tenha apresentado o menor erro total (-0,26%), o modelo 6 apresentou erro total muito pequeno (-1,5%) e um erro relativo mediano (ERM) menor do que o observado no modelo 3 (30% contra 30,5%), tendo sido, portanto, adotado como referência para o cálculo das estimativas para o ano de 2010 (para a validação externa) e para 2015 (para a comparação com outras estimativas demográficas). A equação 5.2 apresenta o modelo escolhido.

$$\begin{aligned} \text{Log}(\text{Dens})10\text{setor}_i = & 70,73 + 261,17\text{TM}1_i + (-170,29)\text{TM}4_i + 52,85\text{TM}7_i + \\ & 25,50\text{Areacup}10_i + 0,57\text{AgSub}10_i + \varepsilon \end{aligned} \quad (5.2)$$

Portanto, os melhores modelos identificados no nível dos setores (para Contagem) foram aqueles que não consideram as bandas 2, 3 e 5, e que incorporam as variáveis “aglomerado subnormal” e “percentual de área ocupada”.

5.1.2.2 Validação Externa e Análise Comparativa

A Tabela 3 apresenta os resultados da validação externa, através das diferenças percentuais entre a estimativa via sensoriamento remoto de 2010 (construída através do modelo criado em 2000, no nível dos setores) e a população do Censo 2010, para o município de Contagem. A tabela mostra ainda as diferenças percentuais da população de Contagem, obtida através das projeções do IBGE (revisões 2008 e 2013) com a população do Censo 2010.

TABELA 3 – Validação Externa: Diferenças Percentuais da Estimativa via Sensoriamento Remoto para o ano de 2010 (baseado em modelo criado em 2000, no nível dos setores) e das Estimativas Oficiais de 2010, em relação à População do Censo 2010 - Contagem

Municípios	Abs / Dif. Perc.	Censo 2010	Pop. de 2010 (projeção IBGE 2013)	Pop. de 2010 (projeção 2008)	Modelo de 2000 - Setores
Contagem	Abs.	603.442	630.352	633.361	670.287
	Dif. (%)	0,0	4,5	5,0	11,1

Fonte: Elaboração Própria.

O modelo de 2000, construído no nível dos setores censitários (equação 6.1), quando aplicado aos dados de 2010, resultou em erro relativo mediano (dos setores) ou ERM de 27%, enquanto que as estimativas de 2010, produzidas a partir desse modelo de 2000 (o que corresponde à validação externa), apresentou erro total de 11,1% (670.287), quando comparado à população de Contagem (603.442 pessoas, obtido no Censo 2010)¹⁷. Esse erro pode ser considerado aceitável, conforme Smith *et al.* (2002) e Wilson e Rowe (2011), como já fora discutido no

¹⁷ Como fora comentado na metodologia, para as validações externas, optou-se como método de avaliação dos erros das estimativas a diferença percentual (ou erro total), por ser indicada, segundo Smith *et al.* (2002), para a avaliação comparativa entre diferentes regiões. Ademais, a grande maioria dos métodos de avaliação de erros descritos no Capítulo 3 (tópico 3.7) é aplicável quando se dispõem de um conjunto de resultados, como, por exemplo, um conjunto de municípios.

Capítulo 3 (tópico 3.7)¹⁸. Já a desagregação das projeções do IBGE (revisões 2008 e 2013) para o nível do município de Contagem (via aplicação do método AiBi) apresentaram erro total de 5,0% (633.361) e 4,5% (630.352), respectivamente; ou seja, erros inferiores à estimativa via sensoriamento remoto (no nível dos setores censitários e partindo de modelos construídos com dados de 2000).

É importante ressaltar que os bons resultados encontrados na etapa de validação interna estão em consonância com as referências internacionais e nacionais sobre o tema. Por outro lado, a validação externa deste estudo trata de uma inovação, pois, a validação externa realizada por Harvey (2002a) se restringiu à aplicação de modelos para o mesmo ano (e outra área), enquanto as validações externas propostas neste trabalho foram testadas para outro ano censitário (e, por consequência, a partir de outra imagem), o que aumenta o desafio em relação à assertividade das estimativas.

A utilização de outra imagem (referente a outro ano) para a validação externa tem forte relação com a análise da contribuição das estimativas via sensoriamento remoto para o campo da Demografia, justamente pela possibilidade de produção de estimativas para períodos intercensitários e pós-censitários. Contudo, esse tipo de “validação externa temporal” tem como implicação a possibilidade de mudanças na relação entre as reflectâncias e a população, seja por diferenças no nível de iluminação solar, bem como em função de interferências atmosféricas (como diferenças no nível de cobertura de nuvens, ainda que tenham sido utilizadas imagens com baixo percentual de nuvens). Diante do exposto, o erro de 11,1%, embora maior se comparado às estimativas oriundas das projeções do IBGE, pode ser considerado um bom resultado, principalmente se levar em consideração a existência de subcobertura no Censo 2010.

Os modelos de 2000 e 2010 (no nível dos setores) foram utilizados para a construção de estimativas para o ano de 2015, sendo que, para a análise

¹⁸ Segundo Smith *et al.* (2002), erros entre 8% e 14% são aceitáveis para um horizonte de 10 anos, enquanto Wilson e Rowe (2011) consideram 5% de erros nas estimativas como de baixa magnitude, 10% como erros aceitáveis, e acima de 20% como graves problemas.

comparativa, será utilizado como parâmetro a estimativa pós-censitária do IBGE para o município de Contagem (que, no ano de 2015, foi de 648.766 pessoas). Assim, a Tabela 4 mostra as diferenças percentuais entre as estimativas calculadas através de modelos no nível dos setores (baseados em dados de 2000 e 2010) e a estimativa pós-censitária do IBGE. A tabela mostra ainda a comparação entre estimativas demográficas – pelos métodos de extrapolação simples, taxas vitais e razão censitária – com a estimativa pós-censitária do IBGE.

TABELA 4 – Análise Comparativa: Diferenças Percentuais das Estimativas via Sensoriamento Remoto (baseados em modelos criados em 2000 e 2010, no nível dos setores) e de estimativas demográficas, para o ano de 2015, em relação às Estimativas Pós-Censitárias do IBG

Municípios	Abs / Diferença Percentual	Estim. Pós-Censitária IBGE (ano de 2015)	Extrap. Simples	Taxas Vitais	Razão Censitária	Modelo de 2000 - Setores	Modelo de 2010 - Setores
Contagem	Abs.	648.766	636.155	666.439	622.170	701.389	691.204
	Dif. (%)	0,0	-1,9	2,7	-4,1	8,1	6,5

Fonte: Elaboração Própria.

O modelo no nível dos setores baseado em dados de 2000 apresentou uma estimativa de 701.389 pessoas (diferença de 8,1% em relação à estimativa do IBGE), enquanto que o modelo no nível dos setores baseado em dados de 2010 estimou a população em 691.204 (diferença de 6,5% em relação à estimativa do IBGE). Já as estimativas demográficas por Extrapolação Simples, Taxas Vitais e Razão Censitária apresentaram diferenças percentuais de, respectivamente, -1,9%, 2,7% e -4,1%, portanto, erros menores se comparados às estimativas por sensoriamento remoto.

Se o erro de 4,5%, entre a população estimada para 2010 – da projeção do IBGE (revisão 2013) desagregada via AiBi – em relação à população do Censo 2010, fosse mantida para o ano de 2015, a estimativa com melhor ajuste para 2015 seria a do método por razão censitária. Contudo, não é uma boa alternativa a busca pela assertividade em relação à população de 2015, em primeiro porque o próprio Censo 2010 – utilizado para o cálculo do erro da projeção (revisão 2013) em 2010 – apresenta problema de subcobertura, cujo valor ainda não foi divulgado e que sem dúvida difere para cada região e município. Por hipótese, o erro total de 4,5% da projeção (revisão 2013) em 2010 deve ser diminuído se corrigida a subcobertura

do Censo 2010, o que, por consequência, modificaria a análise de erro dentre as estimativas calculadas neste trabalho. Por exemplo, se o erro em 2010 da projeção do IBGE (revisão 2013) passasse – por hipótese – de 4,5% para 2%, após a correção da cobertura do Censo 2010, o método mais ajustado, dentre aqueles apresentados na Tabela 4, seria o de Extrapolação Simples.

Ademais, deve-se considerar que a população do IBGE para 2015 trata-se apenas de uma estimativa pós-censitária, calculada através da desagregação da projeção por componentes do IBGE (revisão 2013) para o nível dos municípios através do método AiBi, e que apresenta limitações (como qualquer método de extrapolação). O método AiBi projeta a população da área pequena a partir de sua contribuição no crescimento absoluto da população esperada na área maior, e pressupõe a relação linear entre o crescimento da área maior e da área menor, o que pode não ser verificado, especialmente em áreas menores, como é o caso dos municípios. Da mesma forma, as demais estimativas demográficas e via sensoriamento remoto apresentam um conjunto de pressupostos e limitações.

A validação externa em 2010 (Tabela 3), cujo erro total foi de 11,1% (para o modelo no nível dos setores e construído por setores de 2000) não foi verificado no ano de 2015, pois, o erro encontrado na comparação com a estimativa pós-censitária do IBGE (8,1%) foi inferior à diferença da validação externa (11,1%). Ressalta-se que nessa comparação não foi considerado o problema de subcobertura no Censo 2010 (que poderiam, por sua vez, aproximar as diferenças percentuais observadas).

Em todo caso, não se pode afirmar que a validação externa não seja aplicável a 2015, pois a população estimada para 2015 pela projeção do IBGE (revisão 2013) é desagregada via método AiBi (como já fora comentado) e pode não refletir a população real (não conhecida) para este ano, dada à possibilidade de mudança na dinâmica populacional do município, num curto intervalo de tempo. Assim, a comparação entre as estimativas de 2015 servem apenas como um “parâmetro”, bem como para avaliar o nível de similaridade existente entre os diferentes métodos utilizados.

Mesmo diante à ausência de pesquisa censitária para a realização de comparação que permita a validação externa para o ano de 2015, as estimativas produzidas via

sensoriamento remoto (a partir de modelos no nível dos setores) apresentaram pequenas diferenças percentuais em relação às estimativas pós-censitárias do IBGE – 8,1% para o modelo contruído a partir de dados de 2000, e 6,5%, a partir de dados de 2010 – e diferença abaixo de 12% se comparado a todas as estimativas demográficas testadas.

Contudo, é possível inferir que a população real de 2015 de Contagem esteja mais próxima das estimativas demográficas, pois, as estimativas via sensoriamento remoto sugerem o aumento no ritmo de crescimento no quinquênio 2010-2015 para níveis acima de 2,7%, fenômeno pouco provável no contexto atual da transição demográfica (e pouco provável de ser explicada por um surto migratório, para um município do tamanho populacional de Contagem). Entre 2000 e 2010, a taxa de crescimento anual do País como um todo caiu de 2% para 1,15%, a TFT caiu de 2 para 1,77 (PNUD, 2017) e a TLM reduziu de 4,96% para 3,70%, entre os quinquênios 1995-2000 e 2005-2010 (IBGE, 2000, 2010), enquanto que a estimativa pós-censitária do IBGE sugere um crescimento anual de 1,45%, pouco maior ao observado na década de 2000.

Todavia, os pequenos erros totais podem ser considerados aceitáveis, o que mostra o potencial das estimativas por imagens orbitais de média resolução, ao menos para municípios de elevada densidade populacional, como é o caso de Contagem.

5.1.3 Modelos e Estimativas no nível dos Pixels

Assim como nos modelos baseados em setores censitários, a construção dos modelos no nível dos pixels partiu da análise da correlação entre as bandas das imagens recortadas para o município de Contagem. As Tabelas 5a e 5b apresentam as matrizes de correlação entre população e as bandas dos modelos no nível dos pixels (respectivamente, baseado nos setores de 2000 e na grade estatística de 2010), onde se observa a maior correlação entre as bandas vizinhas (como já era esperado), além da alta correlação entre as bandas 1, 2 e 3, bem como entre as bandas 5 e 7, o que justifica o teste de modelos com a exclusão das

bandas 2, 3 e 5, assim como fora proposto nos modelos no nível dos setores censitários.

TABELA 5 – Matriz de Correlação entre População e as Bandas 1 a 5 e 7: Modelo Baseado em Pixels a partir de Setores Censitários de 2000 (a) e a partir da Grade Estatística de 2010 (b) –Contagem

(a)

	População	banda 1	banda 2	banda 3	banda 4	banda 5	banda 7
População	1,000	0,131	0,040	-0,021	-0,186	-0,158	-0,026
banda 1	0,131	1,000	0,920	0,805	0,365	0,460	0,683
banda 2	0,040	0,920	1,000	0,949	0,627	0,695	0,810
banda 3	-0,021	0,805	0,949	1,000	0,719	0,815	0,870
banda 4	-0,186	0,365	0,627	0,719	1,000	0,859	0,636
banda 5	-0,158	0,460	0,695	0,815	0,859	1,000	0,870
banda 7	-0,026	0,683	0,810	0,870	0,636	0,870	1,000

Fonte: Elaboração Própria.

(b)

	População	banda 1	banda 2	banda 3	banda 4	banda 5	banda 7
População	1,000	0,172	0,102	0,080	-0,243	-0,073	0,093
banda 1	0,172	1,000	0,924	0,853	0,184	0,523	0,712
banda 2	0,102	0,924	1,000	0,949	0,407	0,697	0,795
banda 3	0,080	0,853	0,949	1,000	0,441	0,779	0,869
banda 4	-0,243	0,184	0,407	0,441	1,000	0,741	0,373
banda 5	-0,073	0,523	0,697	0,779	0,741	1,000	0,856
banda 7	0,093	0,712	0,795	0,869	0,373	0,856	1,000

Fonte: Elaboração Própria.

A partir da análise de correlação e considerando as referências sobre o tema (em especial Harvey, 2002b e Reis, 2005), o primeiro modelo contempla todas as bandas (1 a 5 e 7), tendo a população como variável dependente (assim como para os demais modelos). O modelo 2 mantém todas as bandas e incorpora a variável de “aglomerado subnormal”, enquanto o modelo 3 mantém as variáveis do modelo 2, além de incorporar “situação (urbano/rural)”. Já o modelo 4 testa a exclusão das bandas 2 e 3, enquanto o modelo 5 testa a exclusão da banda 5 (em ambos os casos mantendo-se as variáveis de “aglomerado subnormal” e “situação urbano/rural”). Por fim, o modelo 6 considera as bandas 1, 4 e 7, “aglomerado subnormal”, “situação (urbano/rural)”, assim como no modelo 5, e incorpora termos de interação entre as três bandas (1, 4 e 7), a partir do produto com a variável “aglomerado subnormal”, com o objetivo de testar a influência da favela no efeito que cada banda tem sobre a densidade. Ressalta-se que esses modelos propostos foram utilizados não apenas em Contagem, mas também para os municípios de Vespasiano, Sarzedo e Baldim, mantendo-se a mesma estrutura.

Assim como nos modelos baseados em setores, para a construção dos modelos no nível dos pixels foram considerados 2/3 dos bancos de dados, a partir de seleção aleatória. Os modelos no nível dos pixels, construídos a partir de dados de 2000, tiveram 38.116 pixels alocados para o ajuste e 19.058 alocados para a amostra de teste (para a validação interna). Já os modelos no nível dos pixels e baseados em dados dos setores censitários de 2010 contaram com 42.571 pixels para o ajuste e 21.286 para a validação interna.

5.1.3.1 Validação Interna

A Tabela 6 mostra os resultados para os modelos no nível dos pixels, construídos a partir de dados de 2000 (por setores censitários) e 2010 (por setores censitários e grades estatísticas).

TABELA 6 – Validação Interna dos Modelos baseados em Pixels (2000 e 2010) - Contagem

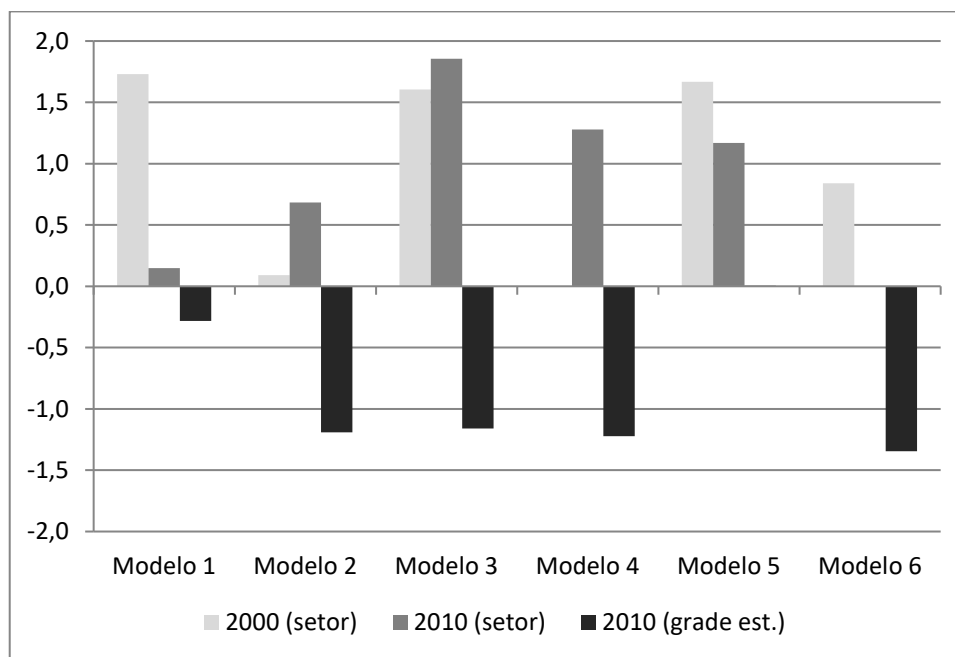
Ano (base de dados)	Indicadores	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6
	Variáveis explicativas	Pop, = Bandas 1 a 5 e 7	Pop, = Bandas 1 a 5 e 7 e Agl, Subnormal	Pop, = Bandas 1 a 5 e 7, Urb./Rur, e Agl, Subnormal	Pop, = Bandas 1, 4, 5 e 7, Urb./Rur, e Agl, Subnormal	Pop, = Bandas 1, 4 e 7, Urb./Rur, e Agl, Subnormal	Pop, = Bandas 1, 4 e 7, Urb./Rur, e Agl, Subnormal, e Interação (Produto) de Agl, Subnormal com as Bandas 1 (v1), 4 (v2) e 7 (v3)
2000 (setor)	R ² (back)	0,077	0,184	0,110	0,126*	0,110	0,146
	Erro relat. mediano	0,353	0,315	0,317	0,314*	0,316	0,318
	Erro total (%)	1,73	0,09	1,61	0,02*	1,67	0,84
2010 (setor)	R ² (back)	0,144	0,186	0,154	0,171	0,170	0,269*
	Erro relat. mediano	0,312	0,293	0,302	0,282	0,281	0,280*
	Erro total (%)	0,15	0,68	1,85	1,28	1,17	-0,40*
2010 (grade estatística)	R ² (back)	0,099	0,127	0,136	0,110	0,170*	0,141
	Erro relat. mediano	0,330	0,323	0,328	0,315	0,318*	0,319
	Erro total (%)	-0,28	-1,19	-1,16	-1,22	0,40*	-1,34

* Modelo adotado para o cálculo das estimativas,

Fonte: Elaboração Própria.

Em todos os modelos baseados em pixels (a partir de dados de setores de 2000 e 2010, e das grades estatísticas de 2010), o R^2 back apresentou valores baixos, que remetem à baixa relação entre população e as variáveis explicativas no nível dos pixels, e que é corroborado pelos ERM's elevados no nível dos pixels, de aproximadamente 31%. Porém, quando se analisa a validação interna dos modelos, as estimativas apresentam erros totais muito baixos, inferiores a 2% em todos os modelos, nas três bases de dados em análise (modelos por setores de 2000, por setores de 2010 e por grade estatística de 2010), o que pode ser considerado como um excelente resultado, mesmo considerando a subcobertura (ainda não conhecida) do Censo 2010. O Gráfico 2 mostra o erro total para cada base de dados e modelo utilizado. Importante ressaltar que a escala do gráfico apresenta baixa amplitude (entre -2% e 2%), de modo que as diferenças nos erros não são significativas.

GRÁFICO 2 – Erro Total dos Modelos baseados em Pixels, a partir de Setores Censitários (de 2000 e 2010) e Grade Estatística (de 2010) - Contagem



Fonte: Elaboração Própria.

Os modelos de pixels que partem de dados de setores censitários de 2000 tiveram o maior R^2 observado no modelo 6 (0,146) seguido pelo modelo 4 (0,126), e os menores ERM's no modelo 4 (0,314). O menor erro total foi observado no modelo 4, com 0,02% de erro total, abaixo inclusive dos resultados encontrados por Reis

(2005), de -0,06%, muito embora as diferenças entre as bases de dados utilizadas em cada trabalho não permitam a comparação direta entre os erros. Em todo caso, pode-se afirmar que os resultados encontrados na validação interna são satisfatórios. Ressalta-se ainda que a incorporação da variável “aglomerado subnormal” (modelo 2) melhorou sobremaneira o erro total estimado, enquanto que a incorporação da variável “situação urbano/rural” (modelo 3) não melhorou a estimativa calculada.

A equação 5.3 apresenta o modelo 4 no nível dos pixels e com dados de 2000, construído a partir de 10 iterações, tendo alcançado um R^2 de 0,20.

$$Pop00pixel(setor)_i = 7,05 + 170,30TM1_i + (-50,17)TM4_i + (-42,53)TM5_i + 43,12TM7_i + 0,68Urbano00_i + 9,11AgSub00_i + \varepsilon \quad (5.3)$$

Enquanto o modelo 4 foi escolhido para o cálculo das estimativas de 2010 e 2015, dentre aqueles construídos a partir de dados de setores censitários de 2000, o modelo 6 foi o que obteve os melhores resultados dentre aqueles que partem de dados dos setores censitários de 2010. Embora com R^2 back de 0,269 (contra 0,144 do modelo 1), o modelo 6 apresentou o menor erro relativo mediano (ERM de 0,28%) e o menor erro total (-0,4%). Ressalta-se que a incorporação das variáveis “aglomerado subnormal” (modelo 2) e “situação urbano/rural” (modelo 3) não melhorou o erro total, quando comparado ao modelo que utiliza somente as reflectâncias das bandas 1 a 5 e 7 (modelo 1). A equação 5.4 mostra o modelo 6, construído a através de 12 iterações, o que resultou num R^2 de 0,41.

$$Pop10pixel(setor)_i = 13,93 + 96,61TM1_i + (-92,32)TM4_i + 32,65TM7_i + (-0,90)Urbano10_i + 10,54AgSub10_i + 108,69TM1_i * AgSub10_i + (-56,60)TM4_i * AgSub10_i + (-71,41)TM7_i * AgSub10_i + \varepsilon \quad (5.4)$$

Já os modelos oriundos da grade estatística obtiveram o melhor ajuste observado no modelo 5, com 0,170 de R^2 back, 31,8% de ERM e 0,4% de erro total (tendo sido escolhido para o cálculo da estimativa). Ressalta-se que a incorporação das variáveis “aglomerado subnormal” (modelo 2) e “situação urbano/rural” (modelo 3) não melhorou o erro total, quando comparado ao modelo que utiliza somente as reflectâncias das bandas 1 a 5 e 7 (modelo 1). A equação 5.5 apresenta o modelo 5, construído após 6 iterações (com R^2 de 0,53).

$$Pop10_{pixel}(grade)_i = 14,02 + 122,38_1 TM1_i + (-92,62)TM4_i + 29,45TM7_i + (-1,89)Urban0_i + 4,87AgSub10_i + \varepsilon \quad (5.5)$$

Os resultados evidenciam a superioridade dos modelos baseados em pixels em comparação aos modelos no nível dos setores censitários, especialmente em relação à validação interna; com erros totais inferiores à 0,5% nos modelos no nível dos pixels, contra -6,3% e 1,5% nos modelos no nível dos setores, respectivamente às bases de dados de 2000 e 2010.

Ademais, os indicadores mostram que a base de dados por grade estatística, diferentemente da hipótese inicial, não apresentou os melhores resultados, muito embora a diferença seja muito pequena. Os modelos baseados em grade estatística apresentaram resultados similares aos modelos construídos por setores censitários de 2010. Destaca-se ainda que, nas três bases de dados em análise, os modelos com as melhores validações internas (4, 6 e 5) são aqueles que não consideram as bandas 2 e 3, bem como aqueles que utilizam variáveis de termos de iteração.

5.1.3.2 Validação Externa e Análise Comparativa

O modelo no nível dos pixels, baseado em dados dos setores censitários de 2000 (equação 5.3), foi utilizado para a produção de estimativa para o ano de 2010, com o objetivo de se realizar a validação externa. Já os modelos no nível dos pixels, baseados em dados dos setores censitários de 2010 (equação 5.4) e baseados na grade estatística de 2010 (equação 5.5) foram utilizados para a produção de estimativas para o ano de 2015, com o objetivo de se realizar uma análise comparativa dessas estimativas com estimativas demográficas, tendo como referências as estimativas pós-censitárias do IBGE (para o ano de 2015).

A Tabela 7 apresenta a validação externa, através da diferença percentual entre a estimativa via sensoriamento remoto, para o ano de 2010 (baseado em modelo no nível dos pixels, criado em 2000, através de dados dos setores censitários) e a população enumerada pelo Censo 2010. A tabela mostra ainda as diferenças percentuais entre as projeções do IBGE (revisão 2008 e 2013), desagregadas via

método AiBi para o município de Contagem, em relação à população do Censo 2010.

TABELA 7 – Validação Externa: Diferenças Percentuais da Estimativa via Sensoriamento Remoto para o ano de 2010 (baseado em modelo criado em 2000, no nível dos pixels) e das Estimativas Oficiais de 2010, em relação à População do Censo 2010 - Contagem

Municípios	Abs / Dif. Perc.	Censo 2010	Pop. de 2010 (projeção IBGE 2013)	Pop. de 2010 (projeção 2008)	Modelo de 2000 - Pixels (a partir de Setores)
Contagem	Abs.	603.442	630.352	633.361	599.371
	Dif. (%)	0,0	4,5	5,0	-0,7

Fonte: Elaboração Própria.

Em relação à validação externa, a aplicação do modelo de pixel baseado em dados de 2000 para a produção de estimativa em 2010 resultou no ERM dos setores de 28,9%, e população estimada de 599.371; o que representa uma diferença percentual de -0,7% em relação à população de Contagem, obtida pelo Censo 2010 (603.442). Mesmo se considerar a subcobertura do Censo 2010, o erro pode ser considerado muito pequeno, e muito superior em qualidade se comparado ao erro total encontrado a partir do modelo no nível dos setores censitários, que partem de dados de 2000 (11,1%).

As projeções do IBGE (revisão 2008 e 2013), desagregadas pelo método AiBi para o nível municipal, apresentaram erro total de, respectivamente, 5,0% (633.361) e 4,5% (630.352), o que mostra, num primeiro momento, o melhor ajuste das estimativas calculadas no nível dos pixels, via imagens Landsat 7 e dados dos setores do Censo 2000. Todavia, os erros de cobertura do Censo 2010 (ainda não conhecido) podem equilibrar as diferenças percentuais (o erro total) entre as estimativas por sensoriamento remoto e as estimativas por projeções populacionais. Em todo caso, o erro obtido através do modelo construído com dados de 2000 deve ser considerado um excelente resultado, com bom ajuste do modelo à validação externa (ou seja, à aplicação do modelo a outra imagem, com 10 anos de temporalidade).

No que diz respeito à análise comparativa para o ano de 2015, a Tabela 8 apresenta as diferenças percentuais das estimativas via sensoriamento remoto baseados em modelos criados em 2000 e 2010, no nível dos pixels, e de estimativas

demográficas, em relação às estimativas pós-censitárias do IBGE, em todos os casos para o ano de 2015.

TABELA 8 – Análise Comparativa: Diferenças Percentuais das Estimativas via Sensoriamento Remoto (baseados em modelos criados em 2000 e 2010, no nível dos pixels) e de estimativas demográficas, para o ano de 2015, em relação às Estimativas Pós-Censitárias do IBGE

Municípios	Abs / Dif. (%)	Estim. Pós-Censit. IBGE (2015)	Extrap. Simples	Taxas Vitais	Razão Censit.	Modelo de 2000 - Pixels (a partir de Setores)	Modelo de 2010 - Pixels (a partir de Setores)	Modelo de 2010 - Pixels (a partir de Grades Estat.)
Contagem	Abs.	648.766	636.155	666.439	622.170	701.199	664.205	683.835
	Dif. (%)	0,0	-1,9	2,7	-4,1	8,1	2,4	5,4

Fonte: Elaboração Própria.

As estimativas por imagens orbitais Landsat 7 para o ano de 2015 resultaram em populações estimadas e erro total de: 701.199 e 8,1% (modelo de 2000, a partir de dados de setores), 664.205 e 2,4% (modelo de 2010, a partir de dados de setores) e 683.835 e 5,4% (modelo de 2010, através de dados da grade estatística); frente à população estimada em 648.766, pelo IBGE e tomada como referência para o cálculo dos erros totais (macro).

Porém, ao analisar as diferenças percentuais das estimativas demográficas – extrapolação simples (-1,9%), taxas vitais (2,7%) e razão censitária (-4,1%) – os erros totais negativos (abaixo da estimativa pós-censitária do IBGE) encontrados no método de extrapolação simples e de razão censitária, parecem ser mais ajustados numa primeira análise. Entretanto, como já fora ressaltado, embora as estimativas pós-censitárias do IBGE tenham sido tomadas como referência, essas também estão sujeitas aos erros implícitos na projeção (revisão 2013) do IBGE e na aplicação do método AiBi, não podendo tomar os resultados dessa estimativa como a “população real”.

É pouco provável que tenha havido aumento no ritmo de crescimento populacional no quinquênio 2010-2015, em relação ao período 2000-2010 em Contagem – para níveis próximos de 2%a.a., como o observado pela estimativa via sensoriamento remoto com menor erro total (2,4%), calculada pelo modelo no nível dos pixels, a partir de dados de 2010 – o que sugere sobreestimativa das populações calculadas

pelos modelos via sensoriamento remoto no nível dos pixels. Vale ressaltar que a estimativa pós-censitária do IBGE sugere uma taxa de crescimento anual de 1,45% entre 2010-2015 (contra 1,15% observado na década anterior). De toda forma, a proximidade das estimativas calculadas por imagens Landsat com as demais estimativas demográficas (incluindo a estimativa do IBGE) pode ser considerado como um ótimo indicativo.

Vale ressaltar ainda que a validação externa obtida através de modelo no nível dos pixels, a partir de dados de 2000 (-0,7%), não apresenta um padrão de similaridade com o erro total do mesmo modelo calculado para o ano de 2015 (8,1%), o que pode estar associado tanto ao erro das estimativas pós-censitárias do IBGE para 2015, ou, o mais provável, a dificuldade de se obter um padrão de erro entre imagens de anos diferentes, aplicadas a um mesmo modelo. Isso pode ter ocorrido em função das diferentes respostas que a reflectância pode fornecer para diferentes imagens, em diferentes períodos no tempo; ou ainda os dois fatores associados.

As estimativas via sensoriamento remoto produzidas para o município de Contagem mostram resultados robustos, tanto nos testes de validação interna como nos de validação externa, reforçando o potencial de aplicação dos métodos utilizados para municípios com alta densidade populacional, em consonância com a literatura nacional e internacional sobre o tema. Ademais, foi possível constatar a superioridade dos modelos baseados em pixels em relação aos modelos que utilizam da média das reflectâncias. Em relação aos modelos no nível dos pixels, verificou-se que os modelos construídos através da grade estatística apresentaram resultados relativamente próximos, em comparação aos modelos no nível dos pixels, elaborados a partir dos setores censitários.

5.2 Vespasiano

Entre 1991 e 2010, a população do município de Vespasiano praticamente dobrou, passando de 54.868 em 1991 para 76.422 em 2000 e 104.527 em 2010. Assim, mesmo com a importante redução dos níveis de fecundidade observados no período – a TFT de 2,62 em 1991 caiu para 2,52 em 2000 e 1,78 em 2010 (segundo

dados da PNUD, 2017) – a taxa de crescimento anual manteve-se elevada, com uma pequena redução, de 3,68% ao ano na década de 1990 para 3,13% na década de 2000. A mortalidade, por sua vez, seguiu a tendência esperada de redução dos níveis de mortalidade infantil, de 32,8 para 15,3 por mil entre 2000 e 2010, em paralelo ao aumento da probabilidade de sobrevivência aos 60 anos, de 77,3 para 79,7 (dados da PNUD, 2017).

A manutenção do elevado ritmo de crescimento populacional nas décadas de 1990 e 2000 tiveram relação com a fecundidade elevada, mas, também, com o intenso volume migratório observado no período. De fato, como já mencionado no capítulo anterior, o crescimento de Vespasiano está ligado ao desenvolvimento do “vetor norte” de Belo Horizonte, com a construção da Cidade Administrativa do Estado de Minas Gerais, além da duplicação da linha verde, o que resultou na valorização e expansão imobiliária do município. Como consequência, o saldo migratório no quinquênio 1995-2000 foi de 10.853 (o que representou uma TLM de 14,2%), enquanto o quinquênio 2005-2010 teve um saldo de 9.222 (com TLM de 8,8%).

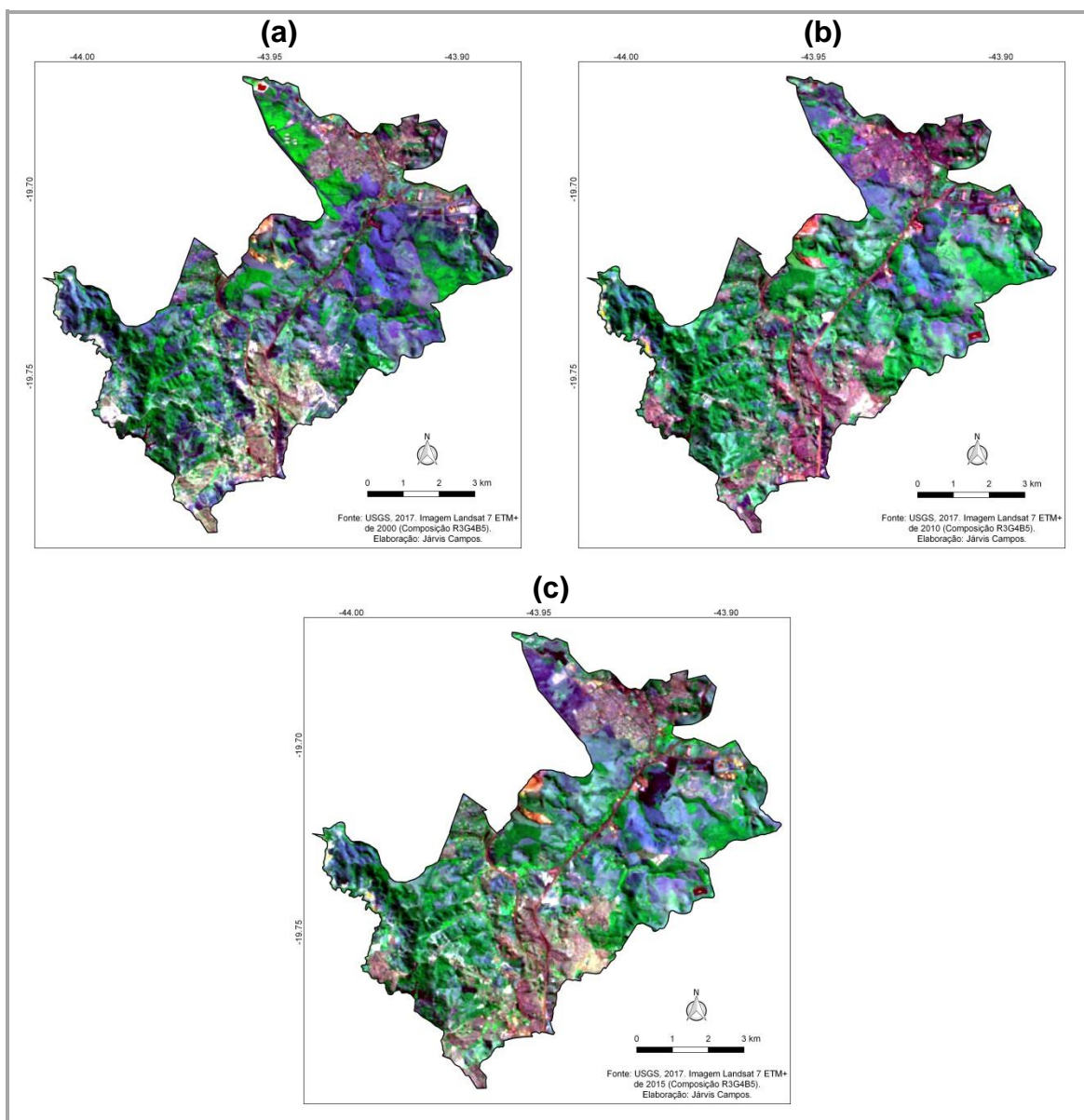
De toda forma, a considerar a baixa TFT alcançada em 2010 (1,78), é de se esperar, na ausência de mudanças abruptas na componente migratória, a redução gradativa do ritmo de crescimento populacional em Vespasiano, a partir do ano de 2011.

5.2.1 Mapeamento Dasimétrico

A Figura 11 apresenta a composição R3G4B5 para os anos de 2000, 2010 e 2015. Ressalta-se que as imagens de Vespasiano pertencem à mesma cena de Contagem¹⁹.

¹⁹ Vale reiterar que as datas foram selecionadas tendo como critérios o baixo percentual de nuvens e a proximidade com a data de referência dos Censos de 2000 e 2010 (13 de agosto de 2000, 08 de julho de 2010 e 07 de agosto de 2015).

FIGURA 11 – Composições R3G4B5 de Imagens Landsat 7 ETM+ para os anos de 2000 (a), 2010 (b) e 2015 (c) – Município de Vespasiano

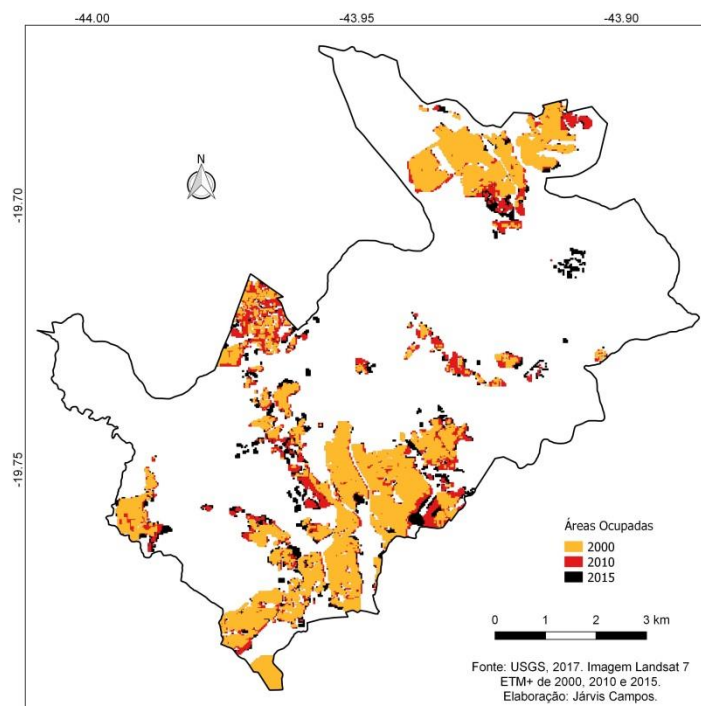


Fonte: USGS, 2017. Elaboração Própria.

Em que pese o ritmo acelerado de crescimento populacional, Vespasiano é marcado por ocupações de baixa densidade populacional, com a presença de duas manchas urbanas bem demarcadas, a primeira ao sul do município, na região conurbada com o município de Belo Horizonte, e no extremo norte, na divisa com o município de Lagoa Santa, e ambas ligadas pela linha verde, que corta o município no sentido norte-sul. Indústrias são observadas no município, especialmente na porção sudeste, enquanto as demais são marcadas por loteamentos de baixa densidade, chácaras e bairros isolados.

A Figura 12 mostra a evolução das manchas de ocupação no período analisado, onde se verifica que a expansão ocorre de forma generalizada em todo o território.

FIGURA 12 – Evolução das Manchas de Ocupação (2000, 2010 e 2015) a partir da classificação de Imagens Landsat 7 ETM+ – Município de Vespasiano



Fonte: USGS, 2017. Elaboração Própria.

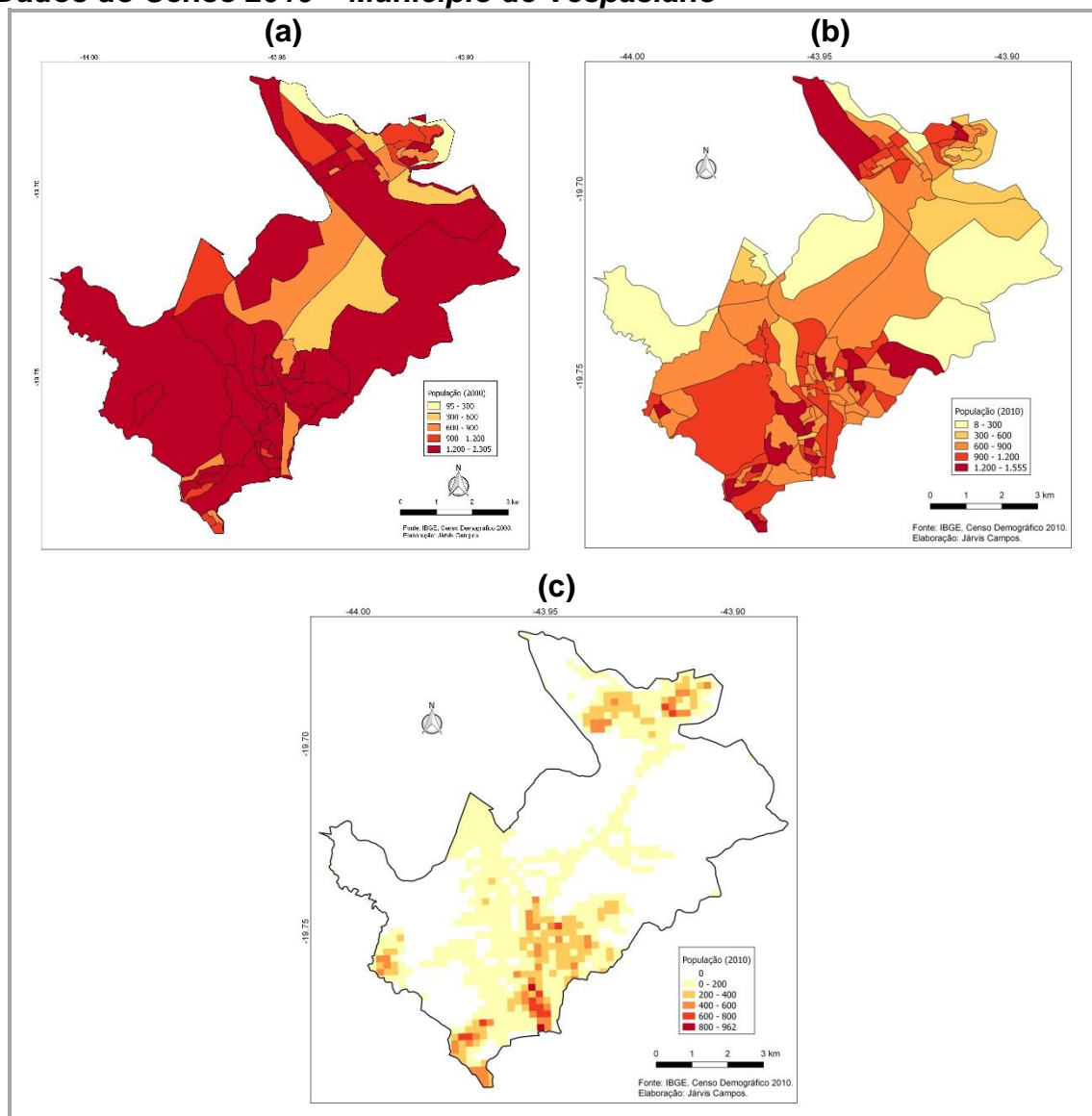
Entre 2000 e 2010, o número de pixels classificados como de ocupação humana cresceu de 11.036 para 14.706 (33,3%), enquanto que, entre 2010 e 2015 o número cresceu 15,7%, passando para 17.014 pixels. O acelerado processo de expansão horizontal da mancha de ocupação também pode ser verificado no crescimento da proporção de área ocupada por setores censitários, cuja mediana aumentou de 60% para 65,5%, entre 2000 e 2010.

Vale ressaltar que, enquanto a proporção do número de pixels considerados como aglomerado subnormal diminuiu no período (14,8% para 13,9%), os pixels localizados em setores censitários urbanos cresceram de 88,3% em 2000 para 100% em 2010, dado que todos os setores censitários no último censo demográfico foram considerados como urbano, o que comprometeu a utilização da variável “situação urbano/rural” nos modelos baseados em dados de 2010.

Para a construção dos mapeamentos dasimétricos, foi adotada a mesma metodologia em relação ao município de Contagem, qual seja a distribuição da população dos setores censitários de 2000 e 2010 e da grade estatística de 2010 para o nível dos pixels. A Figura 13 apresenta as bases de dados oficiais do IBGE, com a distribuição da população.

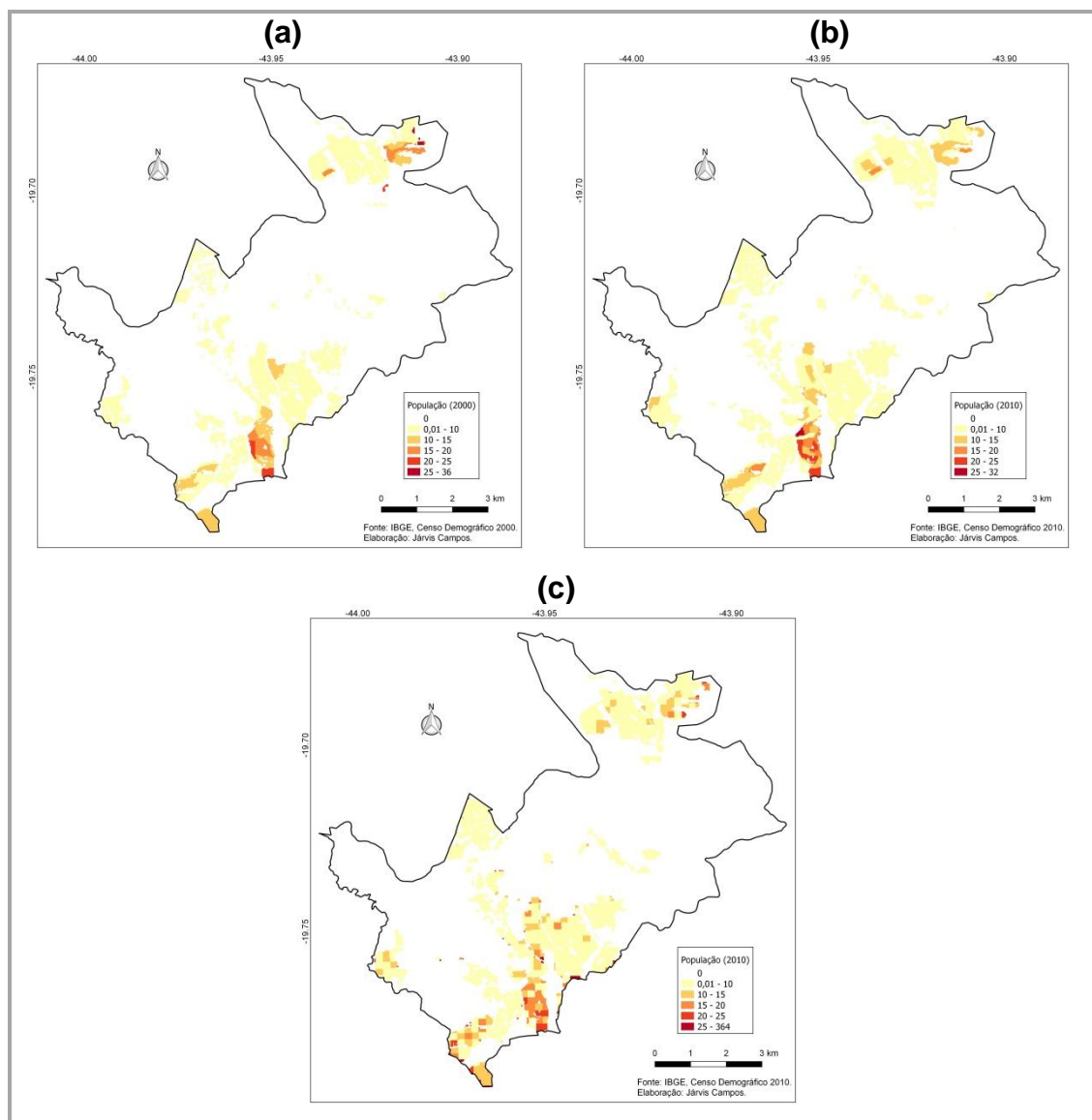
É importante destacar que a malha digital do ano de 2000 contava com apenas 61 setores censitários, sendo todos urbanos, enquanto que a grande área em vermelho apresentada na Figura 13a (nas porções sudoeste e nordeste) corresponde à área rural de 2000, sendo marcada pela ausência de informação, ou seja, a malha de setores de 2000 do IBGE não contempla as áreas rurais, representadas nas porções sudoeste e nordeste do município. Assim, a diferença entre a população enumerada pelo Censo 2000 e o somatório das populações dos setores censitários urbanos foi alocada nessa grande área, para a distribuição da população nos pixels. Já em 2010, o número de setores censitários foi de 121, enquanto que a grade estatística apresentou um total de 778 polígonos.

FIGURA 13 – Distribuição da População por Setores Censitários de 2000 (a), Setores Censitários de 2010 (b) e Grades Estatísticas de 2010 (c), a partir de Dados do Censo 2010 – Município de Vespasiano



A Figura 14 apresenta o mapeamento dasimétrico para os três anos em estudo, onde se verifica uma grande similaridade na distribuição da população entre as bases de dados utilizadas.

FIGURA 14 – Mapeamento Dasimétrico por meio dos Pixels de Imagens Landsat 7 ETM+ e de Dados dos Setores Censitários do Censo 2000 (a), Setores Censitários do Censo 2010 (b), e da Grade Estatística de 2010 (c) – Município de Vespasiano



Fonte: IBGE, Censos Demográficos 2000 e 2010. IBGE, 2015. USGS, 2017. Elaboração Própria.

5.2.2 Modelos e Estimativas no nível dos Setores Censitários

A Tabela 9 apresenta a matriz de correlação entre as médias das reflectâncias das bandas da imagem recortada para o município de Vespasiano, para o ano de 2000, e a densidade populacional por setores censitários.. Para além da alta correlação entre as bandas 2 e 3 e entre as bandas 5 e 7, a Tabela 9 mostra uma maior relação da banda 1 e da banda 2 com a densidade, o que justifica em parte a escolha por

modelos que consideram todas as bandas da imagem (1 a 5 e 7), como será verificado mais adiante.

**TABELA 9 – Matriz de Correlação entre Densidade e as Bandas 1 a 5 e 7:
Modelo Baseado em Setores Censitários de 2000 –Vespasiano**

	Densidade	banda 1	banda 2	banda 3	banda 4	banda 5	banda 7
Densidade	1,000	0,511	0,331	0,181	-0,166	-0,314	-0,048
banda 1	0,511	1,000	0,929	0,803	0,424	0,324	0,629
banda 2	0,331	0,929	1,000	0,954	0,678	0,581	0,760
banda 3	0,181	0,803	0,954	1,000	0,773	0,750	0,834
banda 4	-0,166	0,424	0,678	0,773	1,000	0,868	0,710
banda 5	-0,314	0,324	0,581	0,750	0,868	1,000	0,869
banda 7	-0,048	0,629	0,760	0,834	0,710	0,869	1,000

Fonte: Elaboração Própria.

5.2.2.1 Validação Interna

A Tabela 10 apresenta os resultados para os 7 modelos no nível dos setores censitários, tanto aqueles relativos aos dados de setores de 2000 como de 2010. Para a construção dos modelos foram utilizados 2/3 dos bancos de dados, selecionados de forma aleatória, enquanto que, para a validação interna, foram considerados os outros 1/3 das informações.

TABELA 10 – Validação Interna dos Modelos baseados em Setores Censitários (2000 e 2010) - Vespasiano

Ano (base de dados)	Indicadores	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Modelo 7
	Var. Dependente = Var. Explicativas	Dens, = Bandas 1 a 5 e 7	Log, (Dens,) = Bandas 1 a 5 e 7	Log, (Dens,) = Bandas 1 a 5 e 7, (%) de Área Ocup,	Log, (Dens,) = Bandas 1 a 5 e 7, (%) de Área Ocup, e Agl, Subnormal	Log, (Dens,) = Bandas 1, 4, 5 e 7, Agl, Subnormal e (%) de Área Ocup,	Log, (Dens,) = Bandas 1, 4 e 7, Agl, Subnormal e (%) de Área Ocup,	Log, (Dens,) = Bandas 1, 4 e 7, Agl, Subnormal e (%) de Área Ocup, e Interação (Produto) de Agl, Subnormal com as Bandas 1 (v1), 4 (v2) e 7 (v3)
2000 (setor)	R ² (back)	0,514	0,572	0,797	0,858*	0,896	0,903	0,903
	Erro relat. mediano	0,448	0,511	0,492	0,518*	0,457	0,457	0,456
	Erro total (%)	-150,92	-18,36	-2,79	1,50*	-13,01	-13,09	-11,15
2010 (setor)	R ² (back)	0,583	0,475	0,632	0,665*	0,700	0,704	0,704
	Erro relat. mediano	0,505	0,499	0,395	0,343*	0,393	0,413	0,389
	Erro total (%)	4,54	64,69	-6,19	2,00*	7,14	4,64	11,58

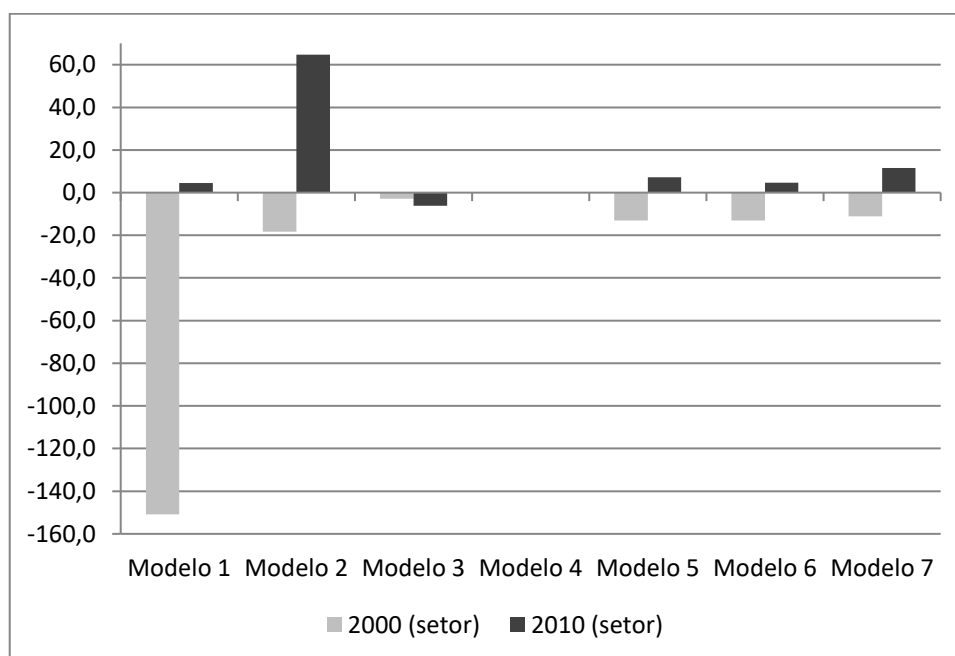
* Modelo adotado para o cálculo das estimativas.

Fonte: Elaboração Própria.

O R^2 (back) apresentou valores elevados, tanto nos modelos que partem de dados de 2000 como de 2010, alcançando 0,903 em 2000 e 0,704 em 2010, em ambos os casos através do modelo 6 (que incorpora os termos de interação das bandas 1, 4 e 7 com a variável “aglomerado subnormal”). Por outro lado, o ERM dos setores foi elevado para todos os modelos, com os menores valores observados no modelo 1 em 2000 (44,8%) e no modelo 4 em 2010 (34,3%).

Em relação à validação interna, o Gráfico 3 apresenta a comparação do erro total entre os modelos 1 a 7 (no nível dos setores) e entre os modelos construídos a partir de 2000 e 2010.

GRÁFICO 3 – Erro Total dos Modelos baseados em Setores Censitários (2000 e 2010) - Município de Vespasiano



Fonte: Elaboração Própria.

Nos modelos baseados em dados de 2000, a incorporação do logaritmo na variável dependente (modelo 2) reduziu o erro total de -150,9% para -18,4%, enquanto que, nos modelos construídos a partir de dados de 2010, o logaritmo na variável dependente aumentou o erro total, de 4,5% para 64,7%. Em que pese a falta de um comportamento padrão em relação ao logaritmo da densidade, para os modelos baseados em 2000 e 2010, a incorporação da variável “percentual de área ocupada” (modelo 3) reduziu significativamente os erros totais, para -2,8% nos modelos de 2000 e -6,2% nos modelos de 2010. De forma menos acentuada, a

incorporação da variável “aglomerado subnormal” (modelo 4) reduziu o erro total para 1,5% no modelo de dados de 2000 e para 2% no modelo de dados de 2010. Estes valores do modelo 4 representam os melhores resultados encontrados, tanto entre os modelos que utilizam dados de 2000 como de 2010, sendo que esses erros podem ser considerados satisfatórios. Ressalta-se ainda que a exclusão das bandas 2, 3 e 5, bem como a incorporação dos termos de interação pioraram as estimativas, para erros totais em torno de 10%.

As equações 5.6 e 5.7 apresentam, respectivamente, os modelos baseados em dados de 2000 e 2010, no nível dos setores, que serão utilizados para a produção de estimativas populacionais, para os anos de 2010 e 2015 (no caso do modelo que parte de dados de 2000) e para o ano de 2015 (no caso do modelo que parte de dados de 2010).

$$\text{Log}(Dens)00setor_i = 7,14 + (-70,10)TM1_i + 4,51TM2_i + 70,74TM3 + 19,69TM4 + (-72,80)TM5_i + 41,21TM7 + 33,16ArOcup00_i + 0,62AgSub00_i + \varepsilon \quad (5.6)$$

$$\text{Log}(Dens)10setor_i = 5,83 + (-94,77)TM1_i + 186,90TM2_i + (-57,86)TM3_i + (-34,16)TM4_i + 16,53TM5_i + (-8,90)TM7_i + 3,18ArOcup10_i + 0,75AgSub10_i + \varepsilon \quad (5.7)$$

5.2.2.2 Validação Externa e Análise Comparativa

A Tabela 11 mostra a validação externa, através da diferença percentual entre a estimativa para 2010, calculada através do modelo construído em 2000, e a população enumerada no Censo 2010²⁰. Para efeito de comparação, a tabela ainda apresenta as diferenças percentuais entre as estimativas oriundas das projeções do IBGE (revisão 2008 e 2013) e a população do Censo 2010.

²⁰ Na Tabela 11 são apresentados os resultados para Vespasiano e também para Contagem. Para todas as tabelas referentes à validação externa, serão apresentados novamente os resultados dos municípios já analisados (portanto, de forma cumulativa), com o objetivo de se realizar uma análise integrada dos erros totais, correspondentes às validações externas.

TABELA 11 – Validação Externa: Diferenças Percentuais da Estimativa via Sensoriamento Remoto para o ano de 2010 (baseado em modelo criado em 2000, no nível dos setores) e das Estimativas Oficiais de 2010, em relação à População do Censo 2010-Vespasiano e Contagem

Municípios	Abs / Dif. Perc.	Censo 2010	Pop. de 2010 (projeção IBGE 2013)	Pop. de 2010 (projeção 2008)	Modelo de 2000 - Setores
Contagem	Abs.	603.442	630.352	633.361	670.287
	Dif. (%)	0,0	4,5	5,0	11,1
Vespasiano	Abs.	104.527	98.928	99.662	106.467
	Dif. (%)	0,0	-5,4	-4,7	1,9

Fonte: Elaboração Própria.

O modelo 4 no nível dos setores e construído a partir de dados de 2000 foi aplicado aos dados de 2010, tendo apresentado ERM de 44,2% e erro total (correspondente à validação externa) de 1,9%, erro menor se comparado às projeções do IBGE, cujas diferenças relativas foram de -4,7% (revisão de 2008) e -5,4% (revisão de 2013). O baixo erro total na estimativa calculada pelo modelo no nível dos setores (1,9%) chama a atenção, principalmente se comparado ao mesmo resultado do município de Contagem (validação externa de 11,1%). Este resultado deixa claro a eficácia do modelo no nível dos setores (como em Harvey 2002a), tanto em situações de alta densidade (como em Contagem) como em contextos de baixa densidade (como é o caso de Vespasiano).

A Tabela 12 apresenta a análise comparativa para o ano de 2015, entre as estimativas calculadas por modelos no nível dos setores (a partir de fonte de dados de 2000 e 2010) e as estimativas demográficas, em relação às estimativas pós-censitárias do IBGE.

TABELA 12 – Análise Comparativa: Dif. Perc. das Estimativas via Sens. Remoto (baseados em modelos criados em 2000 e 2010, no nível dos setores) e de estimativas demog., para o ano de 2015, em relação às Estimativas Pós-Cens. do IBGE de 2015-Vespasiano e Contagem

Municípios	Abs / Diferença Percentual	Estim. Pós-Censitária IBGE (ano de 2015)	Extrap. Simples	Taxas Vitais	Razão Censitária	Modelo de 2000 - Setores	Modelo de 2010 - Setores
Contagem	Abs.	648.766	636.155	666.439	622.170	701.389	691.204
	Dif. (%)	0,0	-1,9	2,7	-4,1	8,1	6,5
Vespasiano	Abs.	118.557	118.580	128.846	117.544	128.142	133.478
	Dif. (%)	0,0	0,02	8,7	-0,9	8,1	12,6

Fonte: Elaboração Própria.

Os modelos no nível dos setores apresentaram ERM de 44,2% e 43,6%, e erro total de 8,1% e 12,6%, respectivamente às fontes de dados de 2000 e 2010. Embora os resultados estejam numa margem de erro aceitável, enquanto as estimativas foram similares entre Vespasiano e Contagem no modelo no nível dos setores e baseados em fonte de dados de 2000 (8,1% para ambos), o erro total do modelo com fonte de dados em 2010 foi o dobro para Vespasiano (12,6%) em relação à Contagem (6,5%). Já as estimativas demográficas apresentaram resultados mais próximos das estimativas pós-censitárias do IBGE, especialmente os métodos de Extrapolação Simples (0,02%) e de Razão Censitária (-0,9%).

A considerar que o município de Vespasiano apresentou uma forte redução da TFT no período 2000-2010 (de 2,52 para 1,78, dados da PNUD, 2017), em paralelo à redução da TLM positiva (14,2% para 8,8% entre 2000 e 2010) [IBGE, 2000, 2010], e com a conseqüente redução do ritmo de crescimento populacional (3,7% para 3,1% entre 2000 e 2010), espera-se que a população real de 2015 esteja próximo às estimativas pós-censitárias do IBGE. A população estimada pelo IBGE para 2015 foi de 118.557, o que resulta numa taxa anual de crescimento de 2,52% ao ano (entre 2010-2015) – contra 3,13% observado na década de 2000 – coerente, portanto, ao pressuposto de redução do ritmo de crescimento no período 2010-2015. De toda forma, os erros totais na casa dos 8,1% e 12,6% (em que pese resultarem num crescimento anual superior a 4% entre 2010-2015) podem ser considerados aceitáveis, especialmente se considerar que trata-se de um município de pequeno porte populacional (mais sujeito à oscilação da população, decorrente de uma eventual mudança no padrão migratório). Portanto, esses resultados mostram o potencial das estimativas via imagens de satélite calculadas a partir de modelos no nível dos setores (como em Harvey, 2002a).

Chama a atenção os diferenciais entre os resultados da validação externa de 2010 (1,9%) e a análise comparativa de 2015 (8,1%), em relação ao modelo no nível dos setores construído a partir de dados de 2000, o que sugere a dificuldade de manutenção da relação entre reflectância e densidade populacional ao longo do tempo (entre diferentes imagens de satélite).

5.2.3 Modelos e Estimativas no nível dos Pixels

A Tabela 13 apresenta as matrizes de correlação entre população e as bandas, para os bancos de dados no nível dos pixels e construídos a partir dos setores censitários de 2000 e da grade estatística de 2010. Assim como nas demais matrizes de correlação verificadas até este ponto, observa-se a alta correlação entre as bandas 2 e 3 e as bandas 5 e 7. Como já fora comentado, os seis modelos são os mesmos descritos e utilizados para o município de Contagem.

TABELA 13 – Matriz de Correlação entre População e as Bandas 1 a 5 e 7: Modelo Baseado em Pixels a partir de Setores Censitários de 2000 (a) e a partir da Grade Estatística de 2010 (b) –Vespasiano

(a)

	População	banda 1	banda 2	banda 3	banda 4	banda 5	banda 7
População	1,000	0,136	0,026	-0,062	-0,177	-0,264	-0,122
banda 1	0,136	1,000	0,938	0,848	0,579	0,558	0,731
banda 2	0,026	0,938	1,000	0,959	0,740	0,723	0,833
banda 3	-0,062	0,848	0,959	1,000	0,780	0,825	0,881
banda 4	-0,177	0,579	0,740	0,780	1,000	0,782	0,690
banda 5	-0,264	0,558	0,723	0,825	0,782	1,000	0,883
banda 7	-0,122	0,731	0,833	0,881	0,690	0,883	1,000

Fonte: Elaboração Própria.

(b)

	População	banda 1	banda 2	banda 3	banda 4	banda 5	banda 7
População	1,000	0,178	0,126	0,105	-0,122	0,010	0,110
banda 1	0,178	1,000	0,948	0,902	0,314	0,670	0,798
banda 2	0,126	0,948	1,000	0,964	0,478	0,788	0,856
banda 3	0,105	0,902	0,964	1,000	0,499	0,836	0,891
banda 4	-0,122	0,314	0,478	0,499	1,000	0,716	0,448
banda 5	0,010	0,670	0,788	0,836	0,716	1,000	0,906
banda 7	0,110	0,798	0,856	0,891	0,448	0,906	1,000

Fonte: Elaboração Própria.

A construção dos modelos considerou 2/3 dos bancos de dados, enquanto os 1/3 restantes foram utilizados para o cálculo da validação interna. Assim, para os modelos no nível dos pixels e baseados em setores de 2000, foram utilizados 7.357 pixels para ajuste do modelo e 3.679 para o teste (validação interna). Já para os modelos no nível dos pixels construídos a partir de setores e grade estatística de 2010, foram selecionados 9.804 pixels para o ajuste e 4.902 pixels para o teste dos modelos.

5.2.3.1 Validação Interna

A Tabela 14 apresenta os resultados para os modelos no nível dos pixels, construídos a partir de dados de 2000 (por setores censitários) e 2010 (por setores censitários e grades estatísticas).

TABELA 14 – Validação Interna dos Modelos baseados em Pixels (2000 e 2010) - Vespasiano

Ano (base de dados)	Indicadores	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6
	Var. Dependente = Var. Explicativas	Pop, = Bandas 1 a 5 e 7	Pop, = Bandas 1 a 5 e 7 e Agl, Subnormal	Pop, = Bandas 1 a 5 e 7, Urb./Rur, e Agl, Subnormal	Pop, = Bandas 1, 4, 5 e 7, Urb./Rur, e Agl, Subnormal	Pop, = Bandas 1, 4 e 7, Urb./Rur, e Agl, Subnormal	Pop, = Bandas 1, 4 e 7, Urb./Rur, e Agl, Subnormal, e Interação (Produto) de Agl, Subnormal com as Bandas 1 (v1), 4 (v2) e 7 (v3)
2000 (setor)	R ² (back)	0,163	0,204	0,208	0,212*	0,217	0,218
	Erro relat. mediano	0,481	0,496	0,476	0,459*	0,475	0,462
	Erro total (%)	7,06	7,11	7,93	5,20*	5,50	7,36
2010 (setor)	R ² (back)	0,188	0,241	-----	0,242*	0,241	0,232
	Erro relat. mediano	0,421	0,404	-----	0,376*	0,384	0,395
	Erro total (%)	-0,65	0,20	-----	0,20*	2,87	0,71
2010 (grade estatística)	R ² (back)	0,110	0,077	-----	0,067	0,066	0,088*
	Erro relat. mediano	0,461	0,461	-----	0,454	0,463	0,440*
	Erro total (%)	-2,29	2,48	-----	3,03	4,50	1,60*

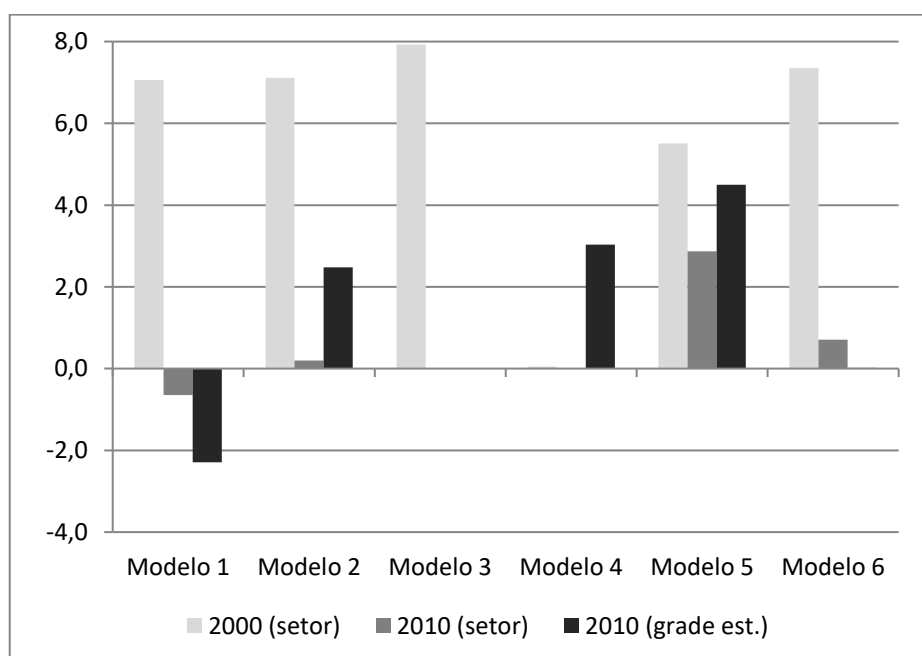
* Modelo adotado para o cálculo das estimativas,

Fonte: Elaboração Própria.

Assim como foi possível constatar para o município de Contagem, o R^2 (back) apresentou valores inferiores no nível dos pixels, se comparado ao mesmo indicador no nível dos setores censitários. Nos modelos no nível dos pixels, os melhores R^2 (back) foram encontrados no modelo 4 (0,212 e 0,242, respectivamente aos modelos de 2000 e 2010). Já os modelos no nível dos pixels, construídos por meio de grade estatística, apresentaram R^2 (back) de 0,09 (no modelo 7). Já o ERM dos setores teve o patamar mais baixo (de 46%) observado nos modelos por pixels, que partem de dados de setores de 2000 (modelo 4). Já o ERM dos modelos no nível dos pixels que utilizam dados dos setores de 2010 foi de 37,6% (modelo 4), enquanto que os modelos no nível dos pixels e construídos através da grade estatística foi de 44% (no modelo 7).

Já em relação ao erro total no nível macro (validação interna), o Gráfico 4 mostra os resultados dos três tipos de modelos baseados em pixels (setores de 2000, de 2010 e grade estatística de 2010) e para cada um dos seis modelos adotados.

GRÁFICO 4 – Erro Total dos Modelos baseados em Pixels, a partir de Setores Censitários (de 2000 e 2010) e Grade Estatística (de 2010) - Vespasiano



Fonte: Elaboração Própria.

Os maiores erros encontrados pertencem aos modelos de pixel construídos através de setores de 2000 (com o menor erro estimado em 5,2% no modelo 4), enquanto os modelos de pixel construídos através de setores de 2010 tiveram o menor erro

total estimado no modelo 4 (0,2%). Já os modelos de pixels elaborados por meio de grade estatística de 2010 tiveram o menor erro total observado no modelo 6 (1,6%). Ainda que, de forma geral, os erros dessa validação interna tenham sido superiores aos erros observados em Contagem (com erro total sempre abaixo de 0,5%), os resultados dessa validação interna podem ser considerados ótimos, especialmente os modelos baseados em dados de 2010.

Contudo, é curioso notar para Vespasiano e Contagem que, diferentemente dos modelos construídos no nível dos setores (como em Harvey 2002a), não há um modelo único no nível dos pixels que tenha apresentado os menores erros totais, de modo que o melhor ajuste (ou a melhor composição das variáveis explicativas) depende da fonte de dados em relação à população e às imagens.

As equações 5.8 a 5.10 apresentam os parâmetros dos modelos adotados para cada base de dados (para o cálculo das estimativas para 2010 e 2015). O modelo de pixel que parte dos setores de 2000 (equação 5.8, referente ao modelo 4) foi construído após 10 iterações, alcançando um R^2 de 0,56.

$$\begin{aligned} Pop00pixel(setor)_i = & 12,10 + 296,52TM1_i + (-46,24)TM4_i + (-9,46)TM5_i + \\ & (-63,25)TM7_i + 0,48Urbano00_i + 27,69gSub00_i + \varepsilon \end{aligned} \quad (5.8)$$

Já o modelo 4 foi escolhido para ajuste dentre os modelos de pixel que partem dos setores de 2010 (equação 5.9), e processou 8 iterações, com um R^2 de 0,55.

$$\begin{aligned} Pop10pixel(setor)_i = & 8,41 + 122,40TM1_i + 1,59TM4_i + (-137,30)TM5_i + \\ & 124,37TM7_i + 2,57AgSub10_i + \varepsilon \end{aligned} \quad (5.9)$$

O modelo 6 foi escolhido para ajuste dentre os modelos de pixel que partem de grade estatística de 2010 (equação 5.10), tendo processado 75 iterações e alcançando um R^2 de 0,49.

$$\begin{aligned} Pop10pixel(grade)_i = & 11,26 + 195,18TM1_i + (-72,26)TM4_i + (-12,59)TM7_i + \\ & 3,03AgSub10_i + (-389,77)TM1_i * AgSub10_i + 13,85TM4_i * AgSub10_i + \\ & 131,92TM7_i * AgSub10_i + \varepsilon \end{aligned} \quad (5.10)$$

5.2.3.2 Validação Externa e Análise Comparativa

A Tabela 15 apresenta as diferenças percentuais da estimativa via sensoriamento remoto para o ano de 2010 (baseado em modelo no nível dos pixels, criado com dados de 2000) e das estimativas oficiais (projeções de 2008 e 2013 do IBGE, desagregadas para o nível municipal através do método AiBi), em relação à população enumerada pelo Censo 2010, para os municípios de Vespasiano e Contagem (este último incluso para efeito de comparação).

TABELA 15 – Validação Externa: Diferenças Percentuais da Estimativa via Sensoriamento Remoto para o ano de 2010 (baseado em modelo criado em 2000, no nível dos pixels) e das Estimativas Oficiais de 2010, em relação à População do Censo 2010-Vespasiano e Contagem

Municípios	Abs / Dif. Perc.	Censo 2010	Pop. de 2010 (projeção IBGE 2013)	Pop. de 2010 (projeção 2008)	Modelo de 2000 - Pixels (a partir de Setores)
Contagem	Abs.	603.442	630.352	633.361	599.371
	Dif. (%)	0,0	4,5	5,0	-0,7
Vespasiano	Abs.	104.527	98.928	99.662	137.546
	Dif. (%)	0,0	-5,4	-4,7	31,6

Fonte: Elaboração Própria.

O ERM dos setores para o município de Vespasiano foi de 49,4%, enquanto o erro total foi de 31,6%, o que é considerado como um resultado ruim, segundo Wilson e Rowe (2011). O resultado parece condizente com a validação interna, que apresentou erro total (5,2%) bem superior aos demais erros de validação interna observados, normalmente inferiores a 2%. Após extensa conferência das etapas de classificação, de construção dos bancos de dados de setores censitários e de pixels, e da construção dos modelos, concluiu-se que o erro não está associado às etapas de produção dos dados, mas provavelmente às limitações do método para aplicações em municípios de baixa densidade populacional, no que diz respeito à aplicação para anos diferentes (em relação ao ano de construção do modelo e ao ano de produção da estimativa).

É importante ressaltar que os resultados da validação interna foram satisfatórios, e que a validação externa – testada entre imagens de anos diferentes – implica no pressuposto de manutenção da relação entre as reflectâncias das bandas e a população, o que parece não se confirmar no município de Vespasiano,

caracterizado como de baixa densidade populacional. Este resultado para modelos no nível dos pixels é curioso, se observar que a validação externa do modelo no nível dos setores apresentou pequeno erro total (1,9%), para Vespasiano. Esses resultados permitem inferir que, para municípios de baixa densidade populacional, a relação entre reflectância e população sofre alteração ao longo dos anos, enquanto que a relação entre a média das reflectâncias e o logaritmo da densidade se mantém mais estável (ao longo do tempo).

Em relação ao erro total das projeções, observa-se a manutenção da diferença percentual em torno de 5%, bem superior em qualidade se comparado ao erro total da estimativa calculada através do modelo por pixels, construído a partir de setores de 2000 (31,6%).

A Tabela 16 apresenta a análise comparativa das estimativas via sensoriamento remoto e das estimativas demográficas em relação às estimativas pós-censitárias oficiais do IBGE, para o ano de 2015.

TABELA 16 – Análise Comparativa: Dif. Perc. das Estimativas via Sens. Remoto (baseados em modelos criados em 2000 e 2010, no nível dos pixels) e de estimativas demog., para o ano de 2015, em relação às Estimativas Pós-Cens. do IBGE de 2015-Vespasiano e Contagem

Municípios	Abs / Dif. (%)	Estim. Pós-Censit. IBGE (2015)	Extrap. Simples	Taxas Vitais	Razão Censit.	Modelo de 2000 Pixels (a partir de Setores)	Modelo de 2010 - Pixels (a partir de Setores)	Modelo de 2010 - Pixels (a partir de Grades Estat.)
Contagem	Abs.	648.766	636.155	666.439	622.170	701.199	664.205	683.835
	Dif. (%)	0,0	-1,9	2,7	-4,1	8,1	2,4	5,4
Vespasiano	Abs.	118.557	118.580	128.846	117.544	175.513	145.634	139.602
	Dif. (%)	0,0	0,02	8,7	-0,9	48,0	22,8	17,8

Fonte: Elaboração Própria.

O erro total entre o modelo de pixel construído a partir de setores de 2000 e a estimativa do IBGE, para o município de Vespasiano, foi de 48%, mantendo o mesmo padrão em relação ao erro total elevado, observado na validação externa. Interessante notar que, embora os erros totais dos modelos de pixels construídos a partir de dados de 2010 apresentem, também, erros bastante elevados – 22,8% e 17,8%, respectivamente aos modelos construídos por setores de 2010 e grade estatística de 2010 – esses ainda são melhores se comparados ao modelo que

parte de dados de setores de 2000 (48%). Esses resultados podem estar associados à menor distância temporal (5 anos) dos modelos criados com dados de 2010, em relação às estimativas produzidas para o ano de 2015, se comparado ao modelo que parte de dados de 2000. A hipótese é de que, quanto maior o intervalo temporal, maior a diferença na relação entre população e reflectâncias, entre imagens e bases de dados de anos diferentes.

Os elevados erros totais observados por diferentes modelos, com diferentes bases de dados, confirmam a baixa eficácia do método de estimativas populacionais via sensoriamento remoto baseado no nível dos pixels, para a aplicação em outro conjunto de imagens, referentes a anos distintos, para municípios de baixa densidade populacional.

Em relação à comparação das estimativas demográficas com as estimativas pós-censitárias do IBGE, vale apenas ressaltar que foram observadas pequenas diferenças percentuais, especialmente os métodos de extrapolação simples (0,02%) e razão censitária (-0,9%), o que corrobora os grandes erros oriundos das estimativas por sensoriamento remoto.

5.3 Sarzedo

No período entre 1991 e 2010, o município de Sarzedo experimentou um incremento populacional da ordem de 171,6%, passando de 9.504 para 25.815 habitantes no período, sendo que nas décadas de 1990 e 2000, as taxas de crescimento anual foram de 6,6% e 4%, respectivamente. Tal crescimento está associado tanto à elevada TFT – que em 1991 era de 2,96 filhos por mulher, caindo em 2000 para 2,69 – como à migração, cujos saldos e TLM's nos quinquênios 1995-2000 e 2005-2010 foram de, respectivamente, 4.335 (25,1%) e 3.246 (12,6%). Já a mortalidade infantil caiu de 23,6 para 12 por mil entre 2000 e 2010, enquanto a probabilidade de sobrevivência aos 60 anos subiu de 82,7% para 85,8%, no mesmo período.

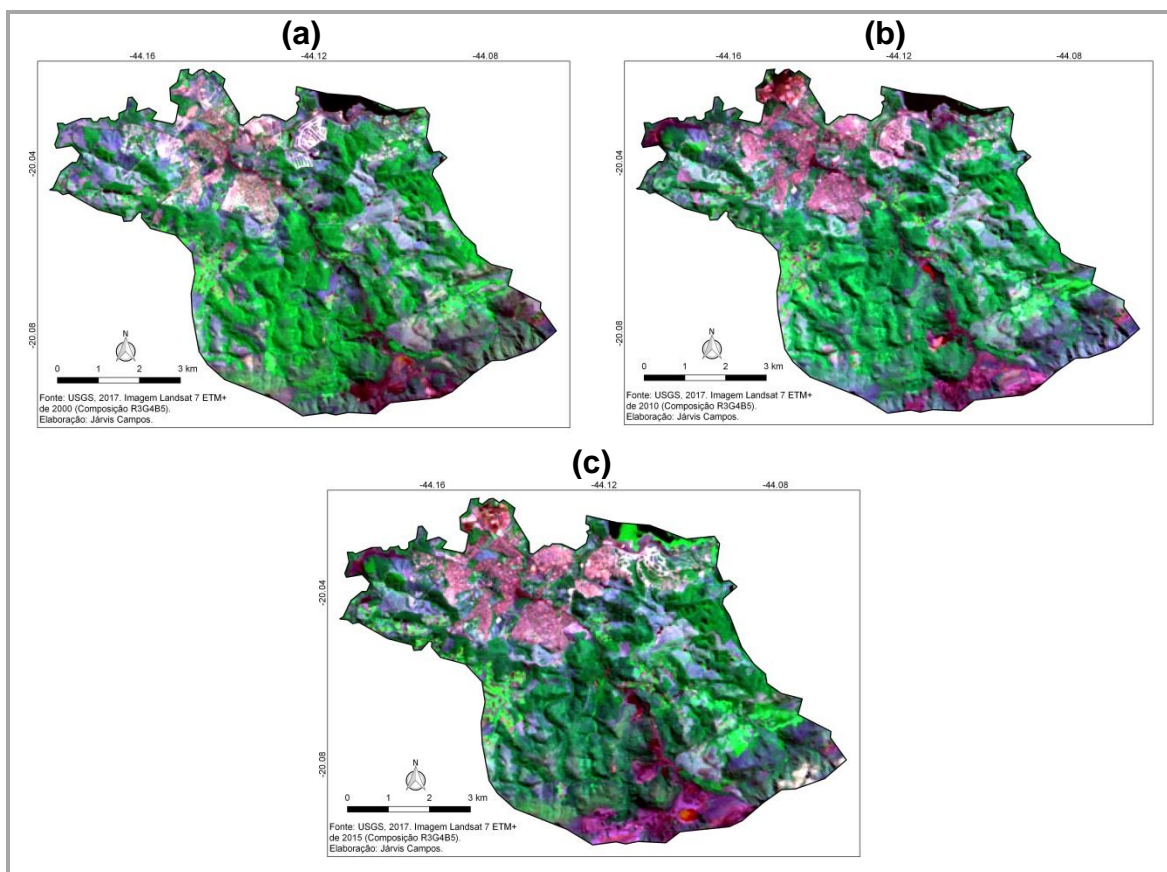
A redução da fecundidade para níveis abaixo da reposição populacional (com TFT de 2,07 em 2010), em paralelo à redução gradativa dos saldos migratórios positivos, observado nas últimas duas décadas, permite inferir que a população do

município de Sarzedo continuará crescendo no quinquênio 2010-2015, porém em ritmo menor, se comparado às taxas de crescimento observadas nas duas décadas anteriores.

5.3.1 Mapeamento Dasimétrico

A Figura 15 apresenta a composição R3G4B5 das imagens Landsat ETM+ para os anos de 2000, 2010 e 2015, para o município de Sarzedo.

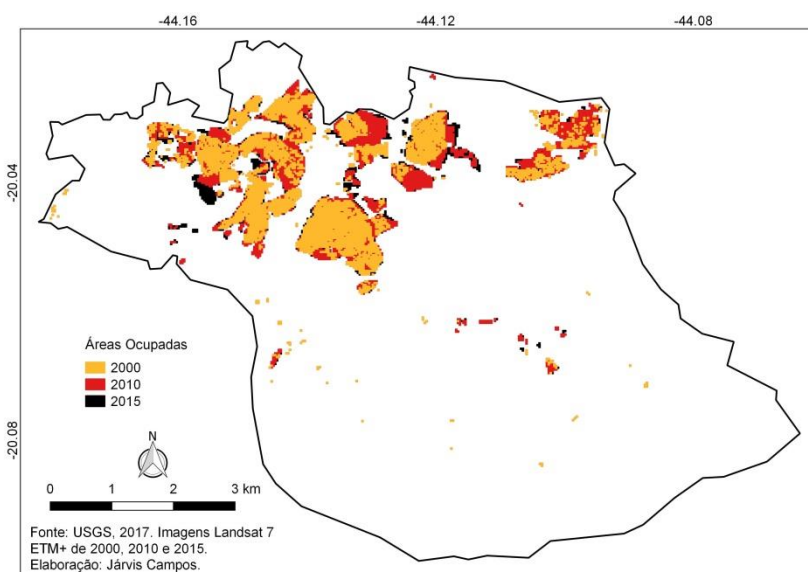
FIGURA 15 – Composições R3G4B5 de Imagens Landsat 7 ETM+ para os anos de 2000 (a), 2010 (b) e 2015 (c) – Município de Sarzedo



Fonte: USGS, 2017. Elaboração Própria.

Já a Figura 16 mostra a evolução das manchas de ocupação entre 2000 e 2015.

FIGURA 16 – Evolução das Manchas de Ocupação (2000, 2010 e 2015) a partir da classificação de Imagens Landsat 7 ETM+ – Município de Sarzedo



Fonte: USGS, 2017. Elaboração Própria.

As áreas ocupadas estão concentradas na porção noroeste, próximo à divisa com o município de Ibitité, e que corresponde à área urbana de Sarzedo, sendo marcada por ocupações de baixa densidade. Já as demais áreas do município são caracterizadas por ocupações rurais isoladas.

Em 2000 foram classificados 4.628 pixels como de áreas de ocupação humana, número que cresceu para 7.245 em 2010 (o que representou um crescimento de 56,5%), e 7.982 pixels em 2015 (ou 10,2% de crescimento entre 2010-2015). Se considerar a ausência de um processo acelerado de verticalização no período, pode-se inferir que o ritmo de crescimento populacional entre 2010-2015 foi inferior ao observado na década de 2000.

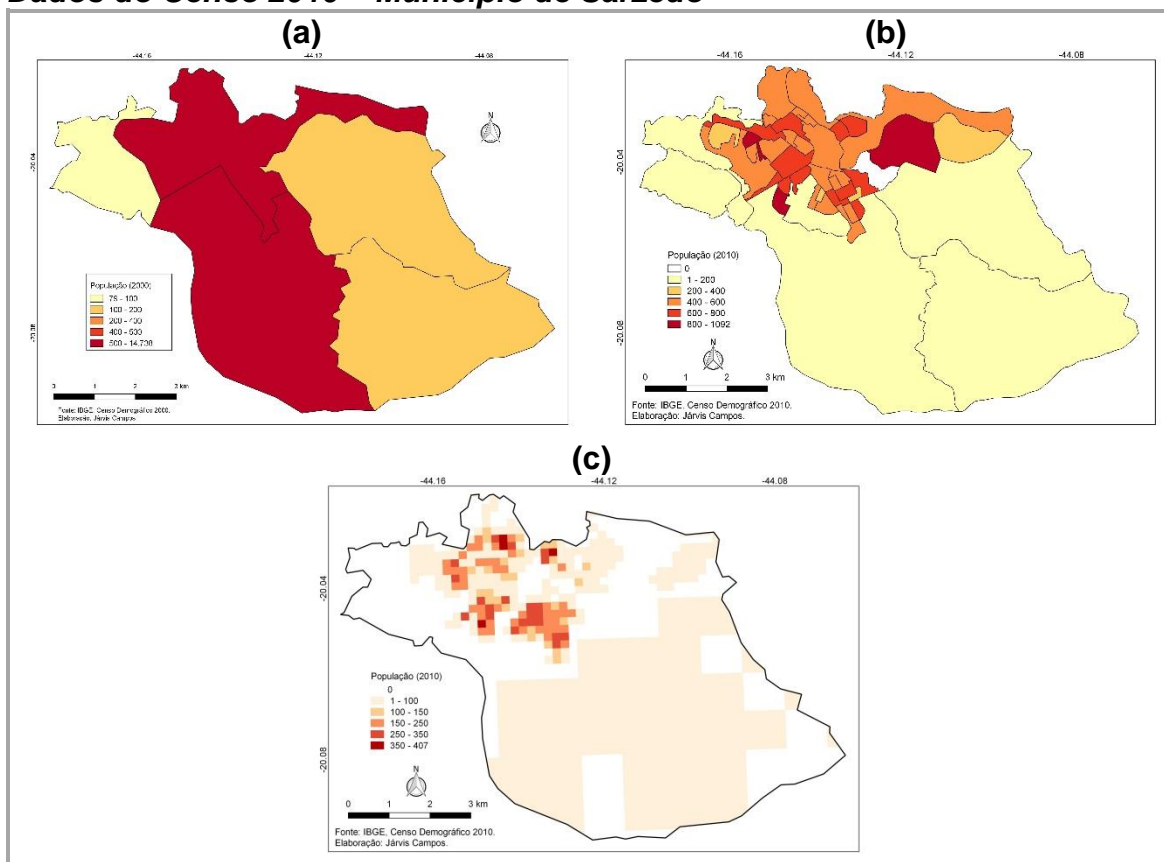
Em relação à situação urbano/rural, no ano 2000 apenas 3.154 pixels (68,2%) foram classificados como urbanos, número que cresceu para 6.990 (96,5%) em 2010 e 7.674 (96,1%) em 2015. Portanto, diferentemente de Vespasiano, que passou a contar com 100% da população urbana em 2010, o que resultou na não utilização da variável “situação urbano/rural” nos modelos de regressão, os dados de Sarzedo permitiram a utilização dessa variável. Porém, tanto em 2000 como em 2010 não foram identificados setores caracterizados como aglomerado subnormal, o que impossibilitou a construção do modelo 2 (que incorpora a variável “aglomerado subnormal” juntamente com as bandas 1 a 5 e 7) e do modelo 6 (que

testa os termos de interação entre as bandas 1, 4 e 7 com a variável “aglomerado subnormal”).

A Figura 17 apresenta a distribuição da população por setores censitários de 2000 e 2010 e pela grade estatística de 2010. O baixo número de setores em 2000 (apenas cinco) impossibilitou a construção das estimativas no nível dos setores censitários (através do método proposto por Harvey, 2002a). Em 2010, os 44 setores censitários permitiriam a construção dos modelos no nível dos setores, mas, devido à ausência de validação externa – em função da impossibilidade de construção de modelos baseados em dados de 2000 – optou-se pela análise somente de modelos baseados em pixels (como em Harvey, 2002b). Portanto, problemas nas bases de dados espaciais são limitadores importantes que devem ser considerados entre os desafios na produção de estimativas via sensoriamento remoto.

Ressalta-se que fora adotada a mesma estratégia para o município de Baldim, em função dos mesmos problemas com a malha de setores censitários de 2000, e diante à ausência de aglomerados subnormais. Sobre a grade estatística, vale ainda destacar que foram utilizados 777 polígonos para o ano de 2010, no município de Sarzedo.

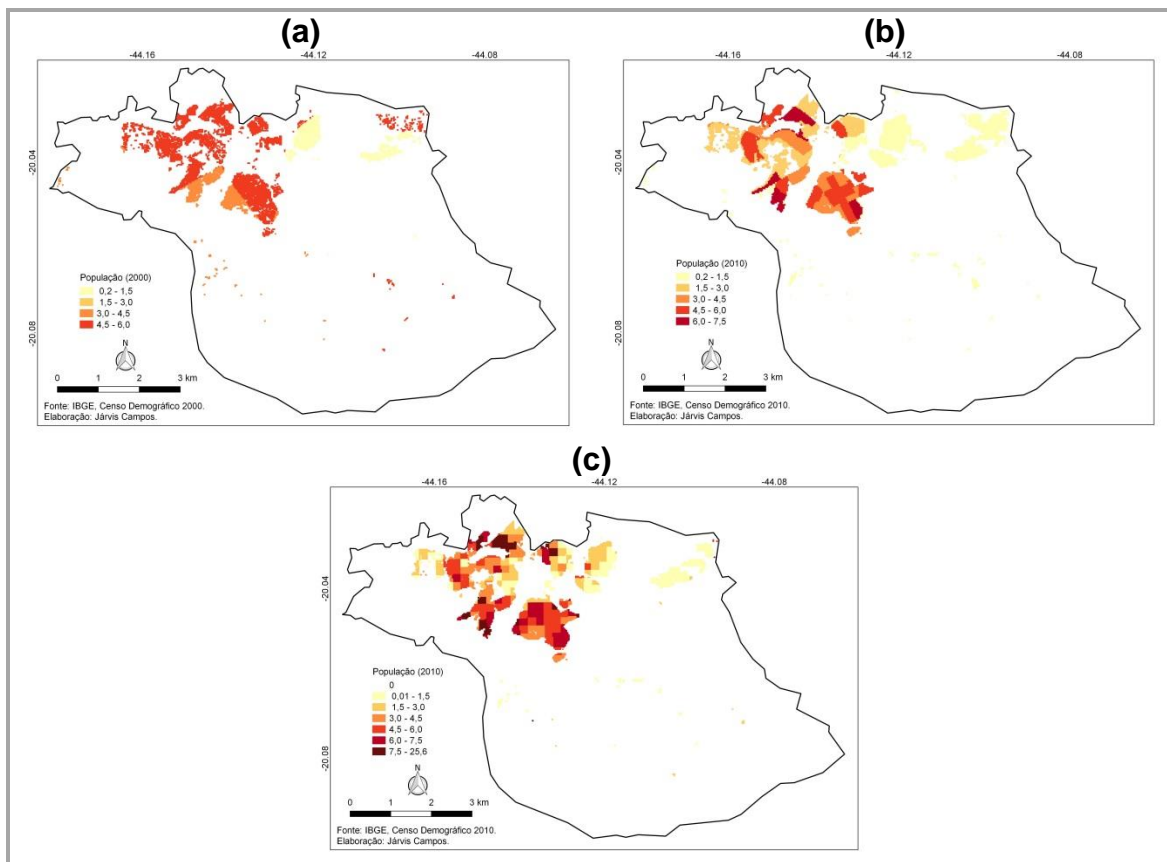
FIGURA 17 – Distribuição da População por Setores Censitários de 2000 (a), Setores Censitários de 2010 (b) e Grades Estatísticas de 2010 (c), a partir de Dados do Censo 2010 – Município de Sarzedo



Fonte: IBGE, Censos Demográficos 2000 e 2010. IBGE, 2015. Elaboração Própria.

A Figura 18 apresenta o mapeamento dasimétrico no nível dos pixels das imagens Landsat 7, a partir dos dados de setores censitários do Censo 2000 e do Censo 2010, além dos dados da grade estatística de 2010, para o município de Sarzedo. É possível notar uma diferença na distribuição da população entre os mapas de cada base de dados, sendo que as medianas da população por pixels foi de 4,7 (para o mapa de pixels baseado nos setores de 2000), 3,3 (baseado nos setores de 2010) e 2,7 (baseado na grade estatística de 2010).

FIGURA 18 – Mapeamento Dasimétrico por meio dos Pixels de Imagens Landsat 7 ETM+ e de Dados dos Setores Censitários do Censo 2000 (a), Setores Censitários do Censo 2010 (b), e da Grade Estatística de 2010 (c) – Município de Sarzedo



Fonte: IBGE, Censos Demográficos 2000 e 2010. IBGE, 2015. USGS, 2017. Elaboração Própria.

5.3.2 Modelos e Estimativas no nível dos Pixels

A Tabela 17 apresenta a matriz de correlação entre população e as bandas 1 a 5 e 7, para as bases de dados de pixels, a partir dos setores censitários de 2000 e da grade estatística de 2010. Assim como nas demais matrizes de correlação, dos demais municípios, é possível notar a alta correlação entre as bandas 1, 2 e 3 e entre as bandas 5 e 7.

TABELA 17 – Matriz de Correlação entre População e as Bandas 1 a 5 e 7: Modelo Baseado em Pixels, a partir de Setores Censitários de 2000 (a) e Modelo Baseado em Pixels, a partir da Grade Estatística de 2010 – Sarzedo

(a)

	População	banda 1	banda 2	banda 3	banda 4	banda 5	banda 7
População	1,000	-0,129	-0,150	-0,174	-0,277	-0,301	-0,061
banda 1	-0,129	1,000	0,952	0,870	0,673	0,713	0,771
banda 2	-0,150	0,952	1,000	0,961	0,777	0,813	0,853
banda 3	-0,174	0,870	0,961	1,000	0,774	0,857	0,887
banda 4	-0,277	0,673	0,777	0,774	1,000	0,851	0,739
banda 5	-0,301	0,713	0,813	0,857	0,851	1,000	0,897
banda 7	-0,061	0,771	0,853	0,887	0,739	0,897	1,000

Fonte: Elaboração Própria.

(b)

	População	banda 1	banda 2	banda 3	banda 4	banda 5	banda 7
População	1,000	0,108	0,051	0,056	-0,376	-0,115	0,125
banda 1	0,108	1,000	0,912	0,880	0,455	0,728	0,802
banda 2	0,051	0,912	1,000	0,969	0,591	0,845	0,877
banda 3	0,056	0,880	0,969	1,000	0,565	0,868	0,911
banda 4	-0,376	0,455	0,591	0,565	1,000	0,781	0,525
banda 5	-0,115	0,728	0,845	0,868	0,781	1,000	0,902
banda 7	0,125	0,802	0,877	0,911	0,525	0,902	1,000

Fonte: Elaboração Própria.

Os modelos foram construídos a partir de 2/3 dos bancos de dados, de modo que, para o ajuste dos modelos de pixel baseados nos setores de 2000 foram considerados 3.085 pixels, enquanto que 1.543 foram utilizados para a validação interna. Já o ajuste dos modelos de pixel baseados nos setores e na grade estatística de 2010 contaram com 4.831 pixels (sendo 2.416 pixels para o teste de validação interna).

5.3.2.1 Validação Interna

A Tabela 18 mostra os resultados para os modelos no nível dos pixels, construídos a partir de dados de 2000 (por setores censitários) e 2010 (por setores censitários e grades estatísticas).

TABELA 18 – Validação Interna dos Modelos baseados em Pixels (2000 e 2010) - Sarzedo

Ano (base de dados)	Indicadores	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6
	Var. Dependente = Var. Explicativas	Pop, = Bandas 1 a 5 e 7	Pop, = Bandas 1 a 5 e 7 e Agl, Subnormal	Pop, = Bandas 1 a 5 e 7, Urb./Rur, e Agl, Subnormal	Pop, = Bandas 1, 4, 5 e 7, Urb./Rur, e Agl, Subnormal	Pop, = Bandas 1, 4 e 7, Urb./Rur, e Agl, Subnormal	Pop, = Bandas 1, 4 e 7, Urb./Rur, e Agl, Subnormal, e Interação (Produto) de Agl, Subnormal com as Bandas 1 (v_1), 4 (v_2) e 7 (v_3)
2000 (setor)	R ² (back)	0,730*	-----	0,730	0,670	0,681	-----
	Erro relat. mediano	0,279*	-----	0,138	0,141	0,123	-----
	Erro total (%)	0,09*	-----	0,72	-0,10	0,40	-----
2010 (setor)	R ² (back)	0,190	-----	0,207*	0,314	0,331	-----
	Erro relat. mediano	0,503	-----	0,465*	0,440	0,460	-----
	Erro total (%)	9,07	-----	-0,10*	-0,20	1,58	-----
2010 (grade estatística)	R ² (back)	0,290	-----	0,353	0,349	0,331*	-----
	Erro relat. mediano	0,417	-----	0,440	0,387	0,420*	-----
	Erro total (%)	0,70	-----	0,92	-1,80	-0,40*	-----

* Modelo adotado para o cálculo das estimativas,

Fonte: Elaboração Própria.

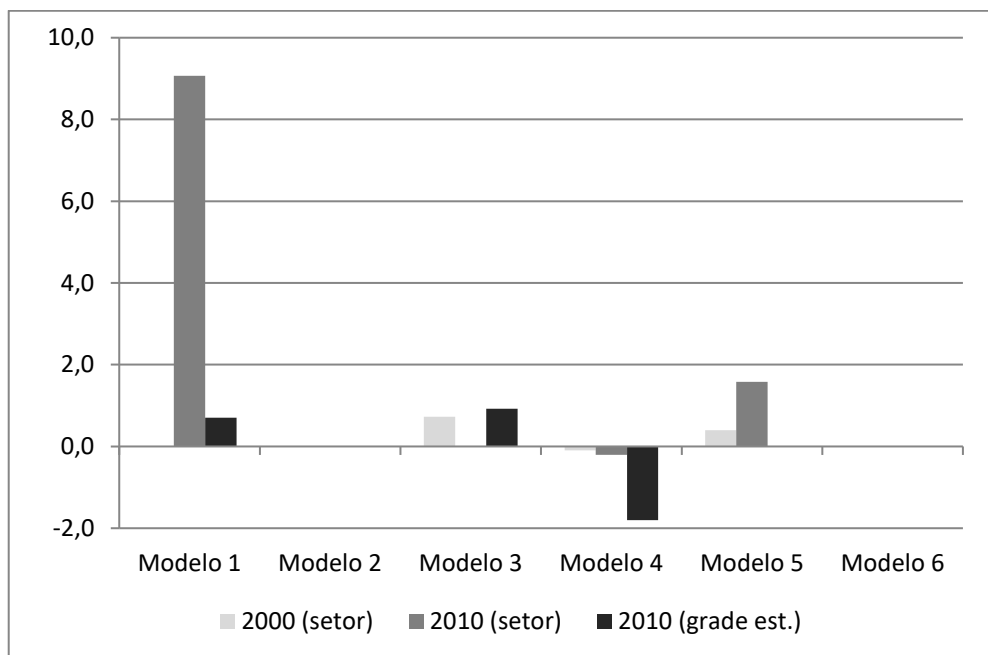
O R^2 (back) apresentou valores maiores nos modelos de pixel que partem dos setores censitários de 2000, alcançando 0,73 no modelo 1, contra o valor máximo de 0,331 no modelo 5, construído a partir dos setores de 2010, e 0,353 no modelo 3, construído através de dados da grade estatística de 2010.

Em relação à definição dos modelos para o cálculo das estimativas, o modelo 1 foi escolhido para o cálculo das estimativas por pixel, oriunda dos dados dos setores de 2000, pois, embora apresente o maior ERM (27,9% contra 12,3% do modelo 5), o erro total foi o menor observado dentre os modelos (0,09%). Dentre os modelos por pixel baseado nos setores de 2010, foi definido o modelo 3 para o cálculo da estimativa de 2015, pois apresenta o menor erro total (-0,10%), e ERM de 46,5%, pouco superior ao menor ERM observado (de 44% no modelo 4). Da mesma forma, o modelo escolhido para o cálculo da estimativa de 2015 dentre os modelos por pixel produzidos através da grade estatística de 2010, teve como critério o menor erro total. O modelo 5 obteve erro total de -0,4% e ERM de 42%, sendo que o menor ERM foi de 38,7%, observado no modelo 4.

A validação interna com erros totais abaixo de 0,4% pode ser considerada muito satisfatória. Entretanto, os modelos que apresentaram os menores erros totais foram diferentes, conforme a base de dados utilizada (ou seja, conforme a fonte de dados utilizada para a distribuição da população e conforme a imagem utilizada). Essa análise mostra que não há um modelo único, com um único conjunto de variáveis explicativas que possa ser utilizado de forma generalizada para a produção de estimativas via sensoriamento remoto.

O Gráfico 5 apresenta os erros totais para as três bases de dados e os 4 tipos de modelos considerados (os modelos 2 e 6 não foram calculados).

GRÁFICO 5 – Erro Total dos Modelos baseados em Pixels, a partir de Setores Censitários (de 2000 e 2010) e Grade Estatística (de 2010) - Sarzedo



Fonte: Elaboração Própria.

É importante ressaltar que, assim como em Vespasiano e diferentemente do observado em Contagem, não foi observado uma grande diminuição dos erros totais (da validação interna) na medida em que se incorporam e que se excluem variáveis. Ou seja, nos municípios de baixa densidade (Vespasiano e Sarzedo), em quase todos os casos (ou seja, para quase todas as bases de dados), a incorporação da variável “situação urbano/rural”, bem como a exclusão das bandas 2, 3 e 5, não resultou em ganho no poder de predição (ou seja, em redução do erro total), entre os modelos 1, 3, 4 e 5 (dado que os modelos 2 e 6 não foram calculados para Sarzedo).

A equação 5.11 corresponde ao modelo 1 por pixel e foi construído a partir dos setores censitários de 2000. O modelo foi calculado após 1.324 iterações, o que pode ter relação com o baixo número de setores censitários, o que demandou por um longo processo de redistribuição da população nos pixels (dado que o R^2 foi sendo continuamente melhorado a cada iteração, alcançando 0,99).

$$Pop00pixel(setor)_i = 5,38 + (-23,42)TM1_i + 145,54TM2_i + (-68,93)TM3_i + (-51,15)TM4_i + 46,83TM5_i + (-35,42)TM7_i + \varepsilon \quad (5.11)$$

Já a equação 5.12 representa o modelo 3 por pixel, baseado nos setores de 2010, que contou com apenas duas iterações, alcançando um R^2 de 0,33.

$$Pop10pixel(setor)_i = 5,75 + (-0,61)TM1_i + 35,89TM2_i + (-22,98)TM3_i + (-24,70)TM4_i + (-34,63)TM5_i + 53,49TM7_i + 0,65Urbano10_i + \varepsilon \quad (5.12)$$

A equação 5.13 corresponde ao modelo 5 por pixel, construído através da grade estatística e foi consolidado após 4 iterações, alcançando um R^2 de 0,94.

$$Pop10pixel(grade)_i = 5,25 + 6,68TM1_i + (-31,98)TM4_i + 16,70TM7_i + 0,77Urbano10_i + \varepsilon \quad (5.13)$$

5.3.2.2 Validação Externa e Análise Comparativa

A Tabela 19 apresenta a validação externa, através da diferença percentual da estimativa via sensoriamento remoto, por pixel e baseada nos setores de 2000, em relação à população enumerada pelo Censo 2010, para Sarzedo, bem como para os municípios de Contagem e Vespasiano (para efeito de comparação). A tabela mostra ainda as diferenças percentuais entre as projeções do IBGE (revisão 2008 e 2013) e a população do Censo 2010, para esses municípios.

TABELA 19 – Validação Externa: Dif. Perc. da Estimativa via Sens. Remoto para o ano de 2010 (baseado em modelo criado em 2000, no nível dos pixels) e das Estimativas Oficiais de 2010, em relação à População do Censo 2010- Sarzedo, Contagem e Vespasiano

Municípios	Abs / Dif. Perc.	Censo 2010	Pop. de 2010 (projeção IBGE 2013)	Pop. de 2010 (projeção 2008)	Modelo de 2000 - Pixels (a partir de Setores)
Contagem	Abs.	603.442	630.352	633.361	599.371
	Dif. (%)	0,0	4,5	5,0	-0,7
Vespasiano	Abs.	104.527	98.928	99.662	137.546
	Dif. (%)	0,0	-5,4	-4,7	31,6
Sarzedo	Abs.	25.814	25.387	25.652	31.010
	Dif. (%)	0,0	-1,7	-0,6	20,1

Fonte: Elaboração Própria.

A estimativa do modelo no nível do pixel, baseado nos setores de 2000, quando aplicado aos dados de 2010 (para Sarzedo), apresentou ERM dos setores de 48,5% e erro total (correspondente à validação externa) de 20,1%, o que pode ser considerado como uma estimativa ruim, segundo Wilson e Rowe (2011). Chama a

atenção a diferença nos resultados, quando comparado à validação interna (cujo erro para o mesmo modelo foi de 0,09%). Assim como fora observado para o município de Vespasiano, a utilização de modelos construídos num dado ano para outro ano (e outra imagem) não fornece bons resultados – talvez em função da mudança na relação entre população e as reflectâncias das bandas – quando se trata de municípios de baixa densidade populacional. Já as projeções do IBGE, desagregadas via método AiBi apresentaram pequenas diferenças percentuais.

A Tabela 20 apresenta a análise comparativa entre as estimativas calculadas para o ano de 2015 e as estimativas pós-censitárias do IBGE, para o mesmo ano.

TABELA 20 – Análise Comparativa: Dif. Perc. das Estimativas via Sens. Remoto (baseados em modelos criados em 2000 e 2010) e de estimativas demog., para o ano de 2015, em relação às Estimativas Pós-Cens. do IBGE de 2015-Sarzedo, Contagem e Vespasiano

Municípios	Abs / Dif. (%)	Estim. Pós-Censit. IBGE (2015)	Extrap. Simples	Taxas Vitais	Razão Censit.	Modelo de 2000 - Pixels (a partir de Setores)	Modelo de 2010 - Pixels (a partir de Setores)	Modelo de 2010 - Pixels (a partir de Grades Estat.)
Contagem	Abs.	648.766	636.155	666.439	622.170	701.199	664.205	683.835
	Dif. (%)	0,0	-1,9	2,7	-4,1	8,1	2,4	5,4
Vespasiano	Abs.	118.557	118.580	128.846	117.544	175.513	145.634	139.602
	Dif. (%)	0,0	0,02	8,7	-0,9	48,0	22,8	17,8
Sarzedo	Abs.	29.889	30.084	36.342	32.090	28.293	34.698	23.902
	Dif. (%)	0,0	0,7	21,6	7,4	-5,3	16,1	-20,0

Fonte: Elaboração Própria.

A diferença percentual da estimativa calculada pelo modelo por pixel baseado nos setores de 2000 foi de -5,3%, o que pode ser considerada aceitável, pois considera o crescimento positivo de Sarzedo entre 2010-2015, sem apresentar uma grande discrepância em relação à estimativa pós-censitária do IBGE. Porém, este resultado mostra que a validação externa calculada pelo mesmo modelo, para o ano de 2010 (20,1%), não serve como referência para a análise do erro de estimativas calculadas para outro ano. Partindo do pressuposto de que a estimativa pós-censitária do IBGE para 2015 seja pequena em relação à “população real”, a diferença dos erros encontrados entre 2010 e 2015 mostram que a relação entre população e as reflectâncias de fato não se mantém constante ao longo do tempo

(entre imagens de diferentes anos), para os municípios de baixa densidade populacional (já verificado em Vespasiano e agora em Sarzedo).

Os erros totais dos modelos por pixels baseados em dados de 2010 foram de 16,1% e -20% (respectivamente às bases de dados dos setores de 2010 e da grade estatística de 2010). Além dos erros totais serem elevados, os resultados são muito díspares, pois o erro da estimativa calculada via setores de 2000 foi positivo (em 16,1%), o que representaria uma aceleração no ritmo de crescimento do município, enquanto o erro da estimativa via grade estatística foi negativo (-20%), o que sugere a inversão no ritmo de crescimento, de positivo para negativo, fenômeno pouco provável para Sarzedo, que apresentou taxa de crescimento anual de 4% ao ano na década de 2000. Considera-se ainda a grande diferença desses dois resultados em relação ao erro total do modelo construído através de dados de 2000 (-5,3%). Esses resultados permitem afirmar que a utilização de dados de 2010 não contribuiu para a redução do erro total (e para a obtenção de melhores estimativas), bem como não há um padrão de erro que permita conclusões seguras quanto à melhor base de dados, ou o melhor modelo, que seria mais ajustado à realidade de Sarzedo.

Portanto, os resultados observados para os municípios de Vespasiano e Sarzedo permitem afirmar que o método baseado em pixel (Harvey, 2002b) tem grande potencial e aplicabilidade, sendo indicado para a redistribuição da população e para a produção de estimativas em áreas de uma mesma cena (de um mesmo ano). Porém, o método baseado em pixel é adequado para a produção de estimativas para períodos pós-censitários (para outro ano, a partir de outra imagem) apenas para municípios de alta densidade populacional (como Contagem), não sendo indicada a produção de modelos em um ano (e imagem) e produção de estimativas para anos diferentes, para municípios de baixa densidade populacional (como observado em Vespasiano e Sarzedo). É importante ressaltar que, embora o número de municípios testados tenha sido pequeno, em função do alto custo de processamento e de construção das bases de dados, foram testados uma grande quantidade de modelos e bases de dados diferentes – e produzidos de forma independente – para cada município, o que permite concluir ao menos no nível da inferência quanto aos resultados aqui apresentados. Contudo, naturalmente, há a

necessidade de continuidade na produção de estimativas via sensoriamento remoto, para a consolidação dos resultados e indicativos apresentados neste trabalho.

5.4 Baldim

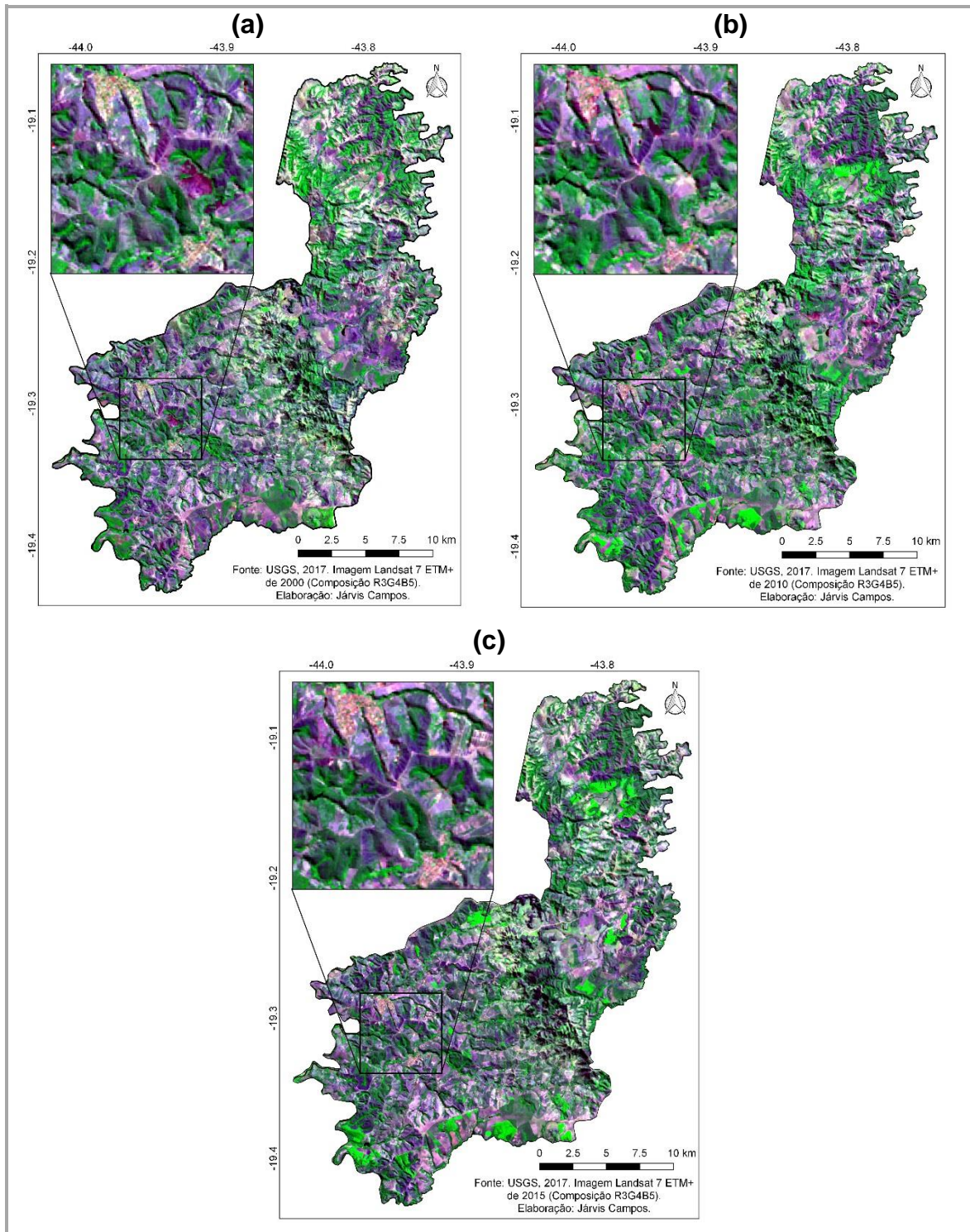
O município de Baldim foi o único da RMBH a apresentar taxa de crescimento anual negativa entre 1991 e 2000 (-0,31%), e entre 2000 e 2010 (-0,3%), caindo de 8.383 habitantes em 1991, para 8.155 em 2000 e 7.913 em 2010. Embora tenha sido observado decréscimo populacional no município, a fecundidade, que apresentou redução no período (de 2,82 em 1991 para 2,31 em 2010), ainda se encontra acima do nível de reposição populacional. Se considerar a inversão do saldo migratório negativo entre os quinquênios 1995-2000 (-99) e 2005-2010 (129), somado ao nível da fecundidade em 2010, é realmente difícil prever a direção do crescimento no quinquênio 2010-2015. Sobre a componente mortalidade, ressalta-se a redução (ainda que pequena) da mortalidade infantil entre 2000 (25,8 por mil) e 2010 (20 por mil) e o leve aumento da probabilidade de sobrevivência até os 60 anos, de 77,4 em 2000 para 78,3 em 2010.

Como já fora comentado no tópico 5.1, tem-se como hipótese que os modelos propostos neste trabalho (baseados em Harvey 2002a e 2002b) não são bem ajustados ao decréscimo populacional, pois pressupõe-se que a perda de população num dado período não pode ser captada pelas imagens de satélite; pois a saída de pessoas não representaria, por hipótese, numa mudança do uso do solo. Deve-se ainda considerar que, apesar do saldo migratório negativo, há também a chegada de imigrantes, que, por sua vez, refere-se a um evento que pode resultar na expansão da área ocupada de um município; sem contar em fatores associados à dinâmica imobiliária do município, que podem resultar na expansão da mancha de ocupação humana. Contudo, o teste é válido, pois, à parte da expansão das manchas de ocupação, deve-se considerar que os modelos baseados em Harvey partem das reflectâncias, que podem eventualmente apresentar mudanças no comportamento decorrentes da perda de população.

5.4.1 Mapeamento Dasimétrico

A Figura 19 apresenta a composição R3G4B5 das imagens Landsat ETM+ para os anos de 2000, 2010 e 2015, para o município de Baldim. A ocupação parte de uma pequena mancha urbana (referente à sede) na porção oeste de Baldim, além de ocupações rurais isoladas, distribuídas nas demais áreas do município. Marcado pela baixa densidade populacional, no ano de 2000 foram identificados 3.428 pixels de ocupação humana em Baldim, dos quais somente 527 (15,4%) são urbanos. Em 2010, o número de pixels subiu para 4.566 (um crescimento de 33,2%), mesmo com a perda de população observada no período. Já a proporção de pixels urbanos cresceu (2.069 ou 45,3% do total). Em 2015, o número de pixels cresceu, porém em ritmo menor (4.566 pixels e 4,9% de crescimento entre 2010-2015), dos quais 2.102 pixels (43,9%) são urbanos.

FIGURA 19 – Composições R3G4B5 de Imagens Landsat 7 ETM+ para os anos de 2000 (a), 2010 (b) e 2015 (c) – Município de Baldim

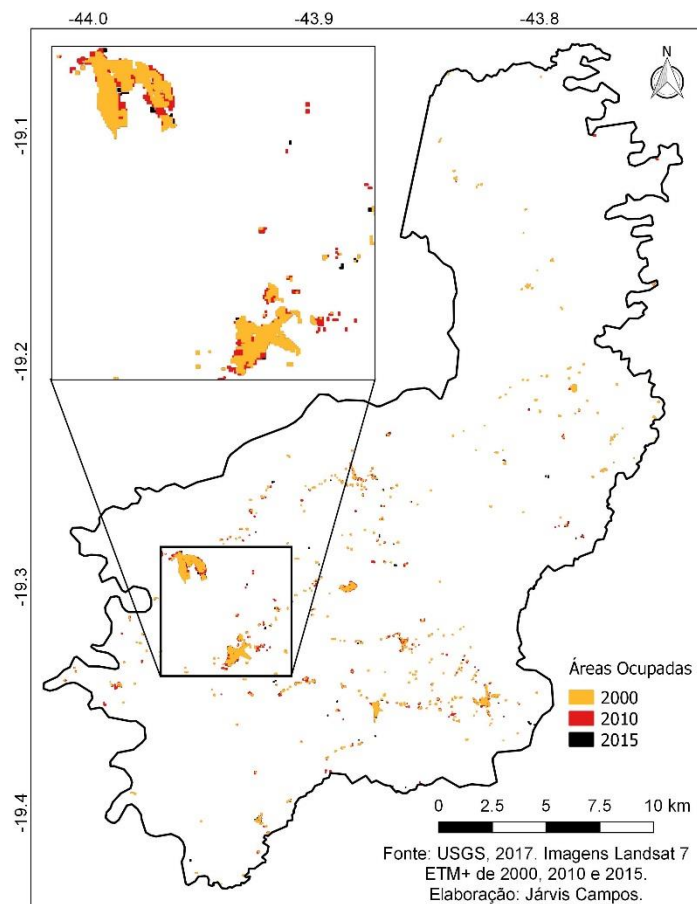


Fonte: USGS, 2017. Elaboração Própria.

A Figura 20 mostra a evolução das manchas de ocupação entre 2000, 2010 e 2015, resultado das classificações das imagens Landsat 7, onde se observa que o

crescimento horizontal das áreas ocupadas se deu de forma espreada, em todo o município.

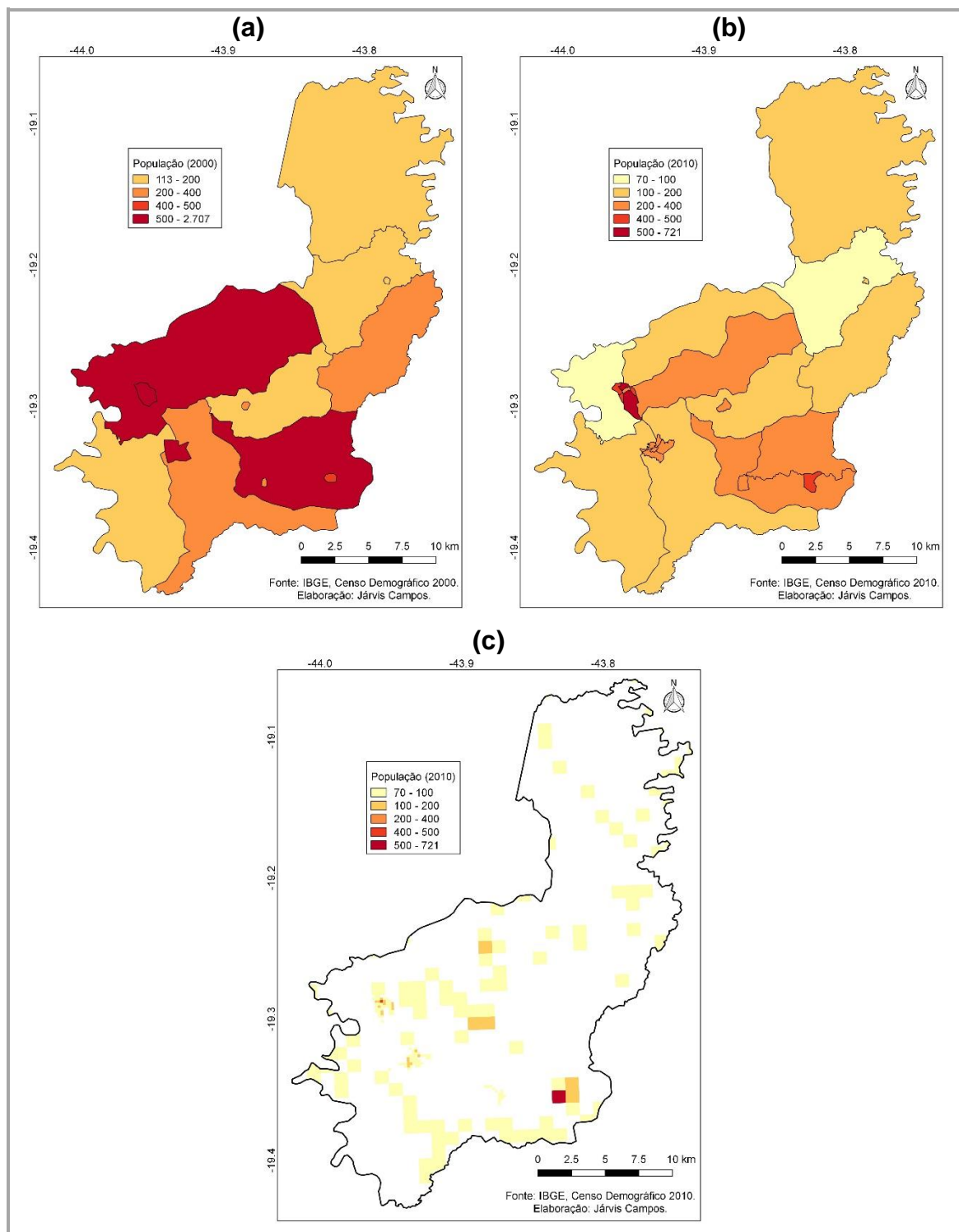
FIGURA 20 – Evolução das Manchas de Ocupação (2000, 2010 e 2015) a partir da classificação de Imagens Landsat 7 ETM+ – Município de Baldim



Fonte: USGS, 2017. Elaboração Própria.

Já a Figura 21 apresenta a distribuição da população por setores censitários de 2000 e de 2010 e por grade estatística de 2010.

FIGURA 21 – Distribuição da População por Setores Censitários de 2000 (a), Setores Censitários de 2010 (b) e Grades Estatísticas de 2010 (c), a partir de Dados do Censo 2010 – Município de Baldim

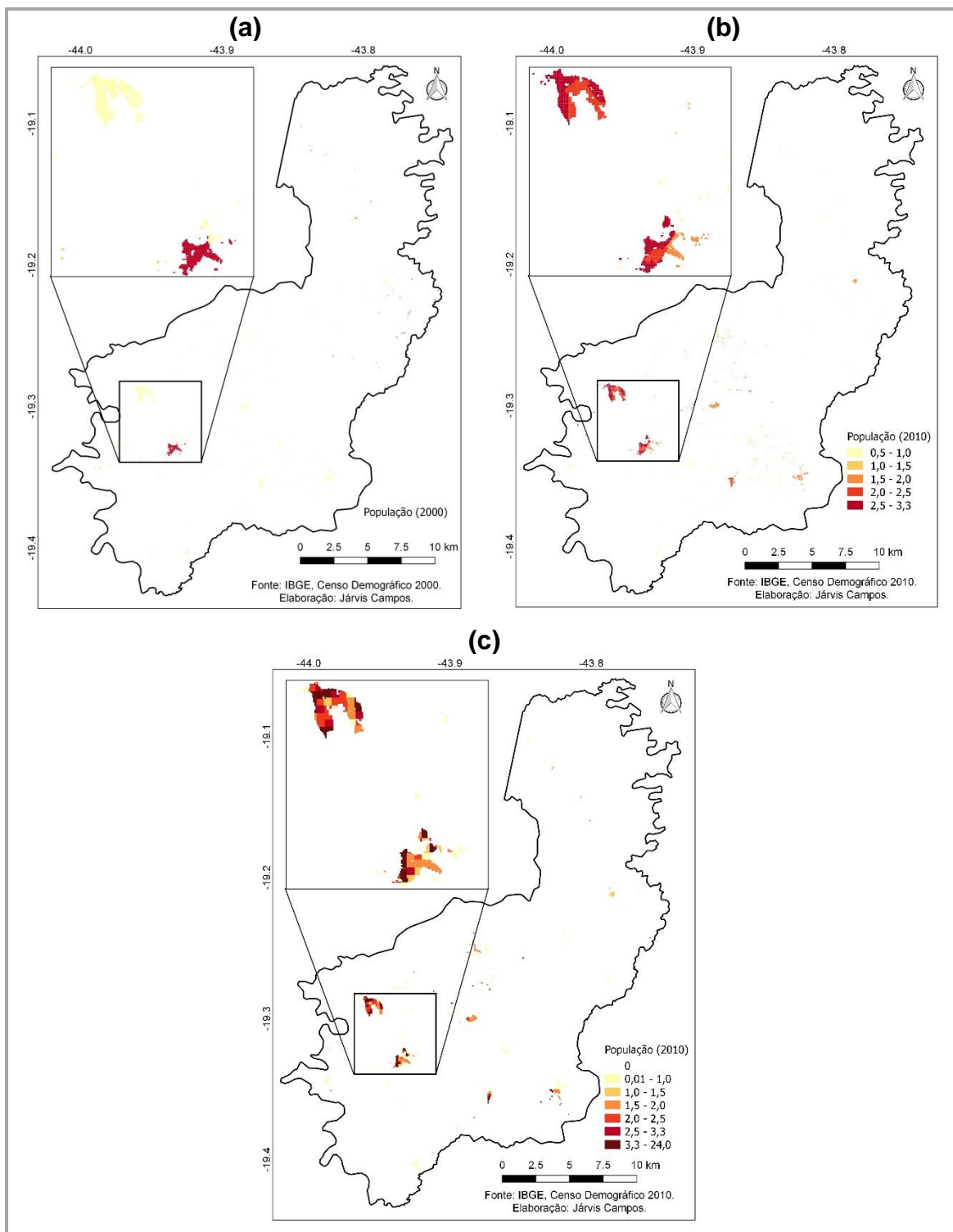


Fonte: IBGE, Censos Demográficos 2000 e 2010. Elaboração Própria.

O pequeno número de setores em 2000 (14) não permitiu a construção dos modelos no nível dos setores (como em Harvey, 2002a), e, diante à impossibilidade de se realizar a validação externa, optou-se por desenvolver somente os modelos

no nível dos pixels. Ressalta-se que as malhas digitais de 2010 contemplam 26 setores e 212 polígonos, referente à grade estatística.

FIGURA 22 – Mapeamento Dasimétrico por meio dos Pixels de Imagens Landsat 7 ETM+ e de Dados dos Setores Censitários do Censo 2000 (a), Setores Censitários do Censo 2010 (b), e da Grade Estatística de 2010 (c) – Município de Baldim



Fonte: IBGE, Censos Demográficos 2000 e 2010. IBGE, 2015. USGS, 2017. Elaboração Própria.

A diferença do número de unidades espaciais entre as três bases de dados apresentadas na Figura 21 resultou numa distribuição da população bastante diferenciada, como pode ser observado na Figura 22. Como consequência, a mediana da população por pixel foi de 0,21 para o mapeamento dasimétrico por pixel a partir dos setores de 2000, 1,2 para o mapeamento via setores de 2010 e 1,8 para o mapeamento via grade estatística de 2010.

5.4.2 Modelos e Estimativas no nível dos Pixels

A Tabela 21 mostra a matriz de correlação entre população e as bandas 1 a 5 e 7, para as bases de dados de pixels, a partir dos setores censitários de 2000 e da grade estatística de 2010, para o município de Baldim. Observa-se a alta correlação entre as bandas 1, 2 e 3 e entre as bandas 5 e 7, além da baixíssima correlação entre população e todas as bandas, em níveis inferiores ao observado para os demais municípios.

TABELA 21 – Matriz de Correlação entre População e as Bandas 1 a 5 e 7: Modelo Baseado em Pixels, a partir de Setores Censitários de 2000 (a) e Modelo Baseado em Pixels, a partir da Grade Estatística de 2010 – Baldim

(a)

	População	banda 1	banda 2	banda 3	banda 4	banda 5	banda 7
População	1,000	0,059	0,031	-0,013	-0,136	-0,039	0,032
banda 1	0,059	1,000	0,937	0,835	0,446	0,560	0,715
banda 2	0,031	0,937	1,000	0,945	0,603	0,665	0,799
banda 3	-0,013	0,835	0,945	1,000	0,627	0,779	0,875
banda 4	-0,136	0,446	0,603	0,627	1,000	0,562	0,523
banda 5	-0,039	0,560	0,665	0,779	0,562	1,000	0,932
banda 7	0,032	0,715	0,799	0,875	0,523	0,932	1,000

Fonte: Elaboração Própria.

(b)

	População	banda 1	banda 2	banda 3	banda 4	banda 5	banda 7
População	1,000	0,193	0,161	0,102	-0,060	-0,037	0,097
banda 1	0,193	1,000	0,928	0,824	0,246	0,596	0,743
banda 2	0,161	0,928	1,000	0,922	0,401	0,714	0,821
banda 3	0,102	0,824	0,922	1,000	0,405	0,784	0,860
banda 4	-0,060	0,246	0,401	0,405	1,000	0,534	0,362
banda 5	-0,037	0,596	0,714	0,784	0,534	1,000	0,895
banda 7	0,097	0,743	0,821	0,860	0,362	0,895	1,000

Fonte: Elaboração Própria.

Os modelos foram construídos através da seleção aleatória de 2/3 dos bancos de dados, enquanto os 1/3 restantes foram utilizados para o teste (validação interna)

dos modelos. Para o ajuste dos modelos de pixel baseados nos setores de 2000 foram considerados 2.285 pixels, enquanto 1.143 foram utilizados para a validação interna. Por sua vez, o ajuste dos modelos de pixel baseados nos setores e na grade estatística de 2010 contaram com 3.044 pixels (sendo 1.522 pixels para o teste de validação interna).

5.4.2.1 Validação Interna

A Tabela 22 apresenta os resultados para os modelos no nível dos pixels, construídos a partir de dados de 2000 (por setores censitários) e 2010 (por setores censitários e grades estatísticas).

Em relação ao R^2 (back), assim como fora observado em Sarzedo, os modelos por pixels construídos através dos setores censitários de 2000 forneceram os melhores resultados. O modelo 4 apresentou um R^2 (back) de 0,886, contra 0,401 do modelo 3 baseado nos setores de 2010 e 0,167 do modelo 4, construído através da grade estatística de 2010.

Os modelos que apresentaram os menores erros relativos, e que foram utilizados para o cálculo das estimativas para 2010 e 2015 variaram conforme a base de dados. O modelo 3 – com R^2 (back) de 0,805, ERM dos setores de 23,7% e erro total de 0,30% – foi o escolhido dentre os modelos por pixel, construídos a partir dos setores de 2000. Dentre os modelos por pixels que partem dos setores de 2010, foi escolhido o modelo 1, com R^2 (back) de 0,275, ERM de 81,5% e erro total de -0,30% (sendo que os demais modelos apresentaram ERM elevado, de aproximadamente 75%). Por fim, o modelo 4 – com R^2 (back) de 0,167, ERM dos setores de 27,6% e erro total de -0,06% – foi o escolhido dentre os modelos por pixel, construídos através da grade estatística de 2010.

TABELA 22 –Validação Interna dos Modelos baseados em Pixels (2000 e 2010) - Baldim

Ano (base de dados)	Indicadores	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6
	Var. Dependente = Var. Explicativas	Pop, = Bandas 1 a 5 e 7	Pop, = Bandas 1 a 5 e 7 e Agl, Subnormal	Pop, = Bandas 1 a 5 e 7, Urb./Rur, e Agl, Subnormal	Pop, = Bandas 1, 4, 5 e 7, Urb./Rur, e Agl, Subnormal	Pop, = Bandas 1, 4 e 7, Urb./Rur, e Agl, Subnormal	Pop, = Bandas 1, 4 e 7, Urb./Rur, e Agl, Subnormal, e Interação (Produto) de Agl, Subnormal com as Bandas 1 (v1), 4 (v2) e 7 (v3)
2000 (setor)	R ² (back)	0,062	-----	0,805*	0,886	0,835	-----
	Erro relat. mediano	0,639	-----	0,237*	0,148	0,238	-----
	Erro total	-1,92	-----	0,30*	1,13	0,90	-----
2010 (setor)	R ² (back)	0,275*	-----	0,401	0,391	0,391	-----
	Erro relat. mediano	0,815*	-----	0,728	0,773	0,716	-----
	Erro total	-0,30*	-----	3,97	1,70	-1,21	-----
2010 (grade estatística)	R ² (back)	0,096	-----	0,152	0,167*	0,121	-----
	Erro relat. mediano	0,354	-----	0,275	0,276*	0,292	-----
	Erro total	1,60	-----	-1,72	-0,06*	1,16	-----

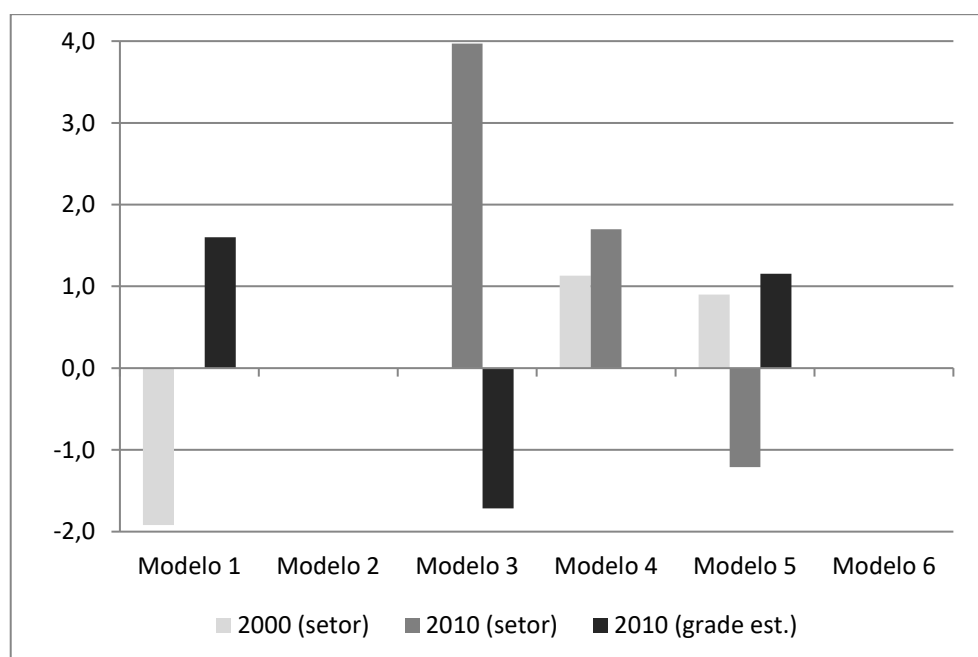
* Modelo adotado para o cálculo das estimativas,

Fonte: Elaboração Própria.

Chama à atenção os baixos erros totais, abaixo de 0,4%, o que pode ser considerado como excelente resultado de validação interna. Por outro lado, nota-se uma falta de padrão em relação aos modelos que apresentaram os menores erros totais, o que ratifica os resultados apresentados nos outros municípios. O conjunto de variáveis explicativas mais ajustadas à variável dependente (população) varia conforme a base de dados; ou seja, varia conforme a fonte de dados utilizada para a distribuição da população, bem como às reflectâncias das bandas, ao longo do tempo (cuja relação com a população pode variar para cada imagem).

O Gráfico 6 mostra os erros totais para as três bases de dados e para os quatro modelos considerados.

GRÁFICO 6 – Erro Total dos Modelos baseados em Pixels, a partir de Setores Censitários (de 2000 e 2010) e Grade Estatística (de 2010) - Baldim



Fonte: Elaboração Própria.

A equação 5.14 corresponde ao modelo 3 no nível dos pixels, construído através dos setores do Censo 2000, e contou com 140 iterações para a definição dos parâmetros (o que resultou no R^2 de 0,99).

$$Pop00_{pixel(setor)_i} = 0,83 + 70,34TM1_i + (-76,10)TM2_i + 19,76TM3_i + 2,39TM4_i + (-5,01)TM5_i + 7,06TM7_i + 2,44Urbano00_i + \varepsilon \quad (5.14)$$

Já a equação 5.15 refere-se ao modelo 1 por pixel, calculado através dos setores de 2010, sendo resultado de 18 iterações (alcançando um R^2 de 0,82).

$$Pop10\ pixel(setor)_i = 3,36 + 40,14TM1_i + 55,86TM2_i + (-16,82)TM3_i + (-10,10)TM4_i + (-28,81)TM5_i + 20,89TM7_i + \varepsilon \quad (5.15)$$

Por fim, a equação 5.16 remete ao modelo 4 por pixel, baseado na grade estatística de 2010, e calculado após 128 iterações, que alcançaram um R^2 de 0,89.

$$Pop10\ pixel(grade)_i = 3,52 + 50,26TM1_i + (-2,63)TM4_i + (-21,14)TM5_i + 10,60TM7_i + 0,49Urbanol0_i + \varepsilon \quad (5.16)$$

5.4.2.2 Validação Externa e Análise Comparativa

A Tabela 23 mostra a diferença percentual entre a estimativa para 2010, produzida através de imagem Landsat 7 e bases de setores censitários de 2000, e a população enumerada pelo Censo 2010, para o município de Baldim (bem como para Contagem, Vespasiano e Sarzedo). A tabela ainda mostra a diferença percentual entre as projeções do IBGE (revisões 2008 e 2013), em relação à população do Censo 2010.

TABELA 23 – Validação Externa: Dif. Perc. das Estimativas via Sens. Remoto para o ano de 2010 (baseados em modelos criados em 2000) e das Estimativas Oficiais de 2010, em relação à População do Censo 2010- Baldim, Contagem, Vespasiano e Sarzedo

Municípios	Abs / Dif. Perc.	Censo 2010	Pop. de 2010 (projeção IBGE 2013)	Pop. de 2010 (projeção 2008)	Modelo de 2000 - Setores	Modelo de 2000 - Pixels (a partir de Setores)
Contagem	Abs.	603.442	630.352	633.361	670.287	599.371
	Dif. (%)	0,0	4,5	5,0	11,1	-0,7
Vespasiano	Abs.	104.527	98.928	99.662	106.467	137.546
	Dif. (%)	0,0	-5,4	-4,7	1,9	31,6
Sarzedo	Abs.	25.814	25.387	25.652	-----	31.010
	Dif. (%)	0,0	-1,7	-0,6	-----	20,1
Baldim	Abs.	7.913	7.917	7.909	-----	7.768
	Dif. (%)	0,0	0,05	-0,05	-----	-1,8

Fonte: Elaboração Própria.

O ERM entre o modelo por pixel baseado nos setores de 2000 e a população do Censo 2010 foi de 44,2%, enquanto que o erro total (no nível macro, que corresponde à validação externa) foi de apenas -1,8%. Interessante observar que

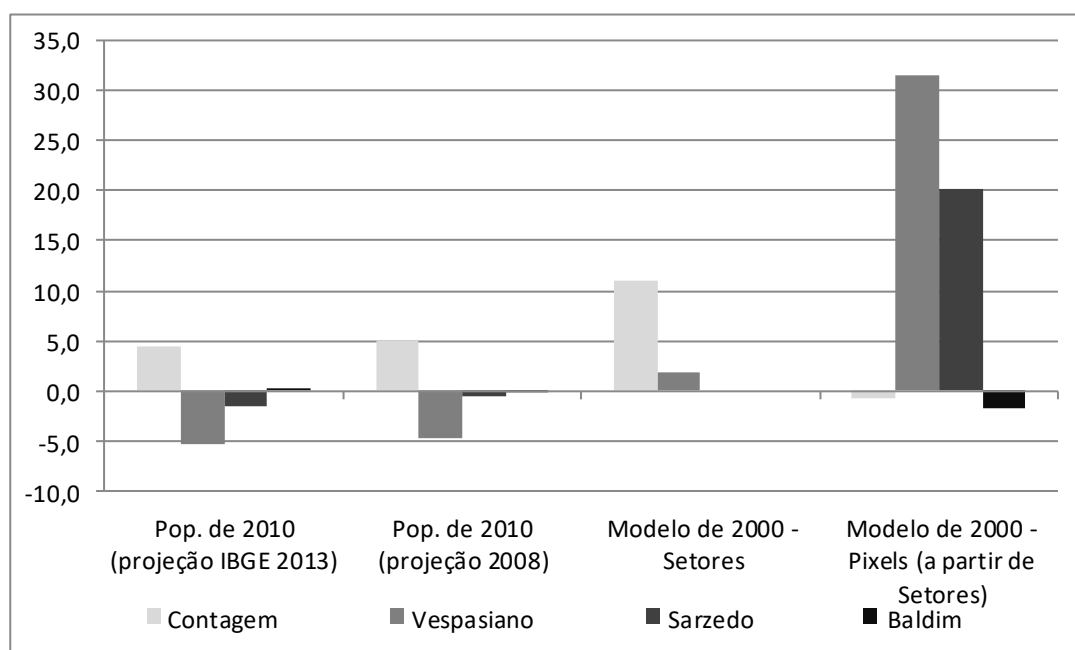
o modelo calculou, além de uma diferença pequena, uma população menor do que a observada em 2000 e um pouco inferior à população de 2010, para o município de Baldim. Este é um resultado muito interessante e um bom indicativo de aplicabilidade do modelo baseado em pixels na produção de estimativas para municípios que apresentam decréscimo populacional.

Mesmo se considerar que as projeções do IBGE para Baldim apresentaram erros relativos menores que o observado para outros municípios (0,05% para a revisão 2013 e -0,05% para a revisão 2008), e, ainda que a subcobertura do Censo 2010 indique que a diferença de -1,8% entre a estimativa por sensoriamento remoto e a população do Censo 2010 tende a ser um pouco maior, deve-se ressaltar que o erro total foi pequeno e muito satisfatório. Contudo, o mais importante para efeito de análise da contribuição das estimativas via sensoriamento remoto para a Demografia, é que a aplicação do modelo por pixel e baseado em dados de 2000 para o ano de 2010 estimou uma população menor que a observada em 2000, o que mostra a potencialidade do sensoriamento remoto na produção de estimativas, a partir de diferentes comportamentos demográficos.

Ainda assim, deve-se considerar que os resultados são incipientes, somado ao fato de que a ausência de um comportamento padrão nas validações externas de municípios com menor contingente populacional – visto que os erros totais para os municípios de Vespasiano e Sarzedo foram elevados (31,6% e 20,1%) – não permitem concluir sobre a eficácia dos métodos baseados em pixels para estes municípios, marcados pela baixa densidade populacional.

O Gráfico 7 mostra os erros totais das projeções do IBGE, das estimativas no nível dos setores e das estimativas no nível dos pixels, em relação à população do Censo 2010, para os quatro municípios em análise.

GRÁFICO 7 – Diferenças Percentuais das Estimativas via Sensoriamento Remoto para o ano de 2010 (baseados em modelos criados em 2000) e das Estimativas Oficiais de 2010, em relação à População do Censo 2010 - Baldim, Contagem, Vespasiano e Sarzedo



Fonte: Elaboração Própria.

Em que pese a validação interna ter apresentado erros totais sempre menores para os modelos no nível dos pixels (quase sempre abaixo de 1%) em relação aos modelos no nível dos setores (com erros um pouco superiores, entre 1% e 5% em média), os resultados da validação externa apresentados no Gráfico 7 mostram um melhor ajuste dos modelos por setores em municípios com baixa densidade populacional. Embora este teste tenha sido realizado somente para Vespasiano – por problemas de bases de dados do IBGE em relação às malhas dos setores de 2000, cujo pequeno número de setores impediu a produção de estimativas no nível dos setores para Sarzedo e Baldim – foram testados vários modelos e para dois anos distintos, através do cálculo de estimativas para o ano de 2010, bem como para o ano de 2015.

Portanto, os modelos no nível dos pixels apresentaram erros elevados para os municípios com baixa densidade populacional (no caso, Vespasiano e Sarzedo). Já os modelos baseados em pixels foram superiores nos testes de validação externa para o município de Contagem, o que mostra o melhor ajuste dos modelos

baseados em Harvey (2002b) para municípios com elevada densidade populacional, quando comparado aos modelos no nível dos setores.

Em relação às projeções populacionais (desagregadas pelo método AiBi), os erros totais para Baldim foram muito pequenos (entre -0,05% e 0,05%), embora essas diferenças tenham uma tendência ao aumento, na hipótese de correção da subcobertura do Censo 2010.

A Tabela 24 mostra a análise comparativa das estimativas via sensoriamento remoto (produzidas a partir de dados de 2000 e 2010) e das estimativas demográficas, em relação às estimativas pós-censitárias do IBGE, para o ano de 2015.

Na análise comparativa entre as estimativas para o ano de 2015 tem-se considerado que as estimativas pós-censitárias do IBGE apresentam um bom ajuste (ou pequeno erro total) em relação à “população real”, que não é conhecida. Essa hipótese pode ser justificada pelo pequeno intervalo de projeção entre 2010 e 2015, bem como pelos pequenos erros totais apresentados pelas projeções demográficas para o ano de 2010 (que oscilam para algo em torno de 5%). Se o erro total da estimativa pós-censitária do IBGE para o ano de 2015 for pequena, os métodos de Extrapolação Simples apresentaram erros pequenos, de apenas -3,5% e 5,7%, respectivamente, em contraponto ao método de Razão Censitária, que apresentou um erro da ordem de 17,5%. Mesmo diante às oscilações das estimativas demográficas para o município de Baldim espera-se, de fato, que a estimativa do IBGE para 2015 esteja próxima da realidade, pois a mesma sugere um ritmo de crescimento de 0,4% entre 2010-2015, algo relativamente próximo dos -0,3% ao ano, observado nas duas décadas anteriores.

TABELA 24 – Análise Comparativa: Dif. Perc. das Estimativas via Sens. Remoto (baseadas em modelos criados em 2000 e 2010) e de estim. demog., para o ano de 2015, em relação às Estimativas Pós-Cens. do IBGE de 2015-Baldir, Contagem, Vespasiano e Sarzedo

Municípios	Abs / Diferença Percentual	Estim. Pós- Censitária IBGE (ano de 2015)	Extrap. Simples	Taxas Vitais	Razão Censitária	Modelo de 2000 Setores	Modelo de 2010 Setores	Modelo de 2000 - Pixels (a partir de Setores)	Modelo de 2010 - Pixels (a partir de Setores)	Modelo de 2010 - Pixels (a partir de Grades Estat.)
Contagem	Abs.	648.766	636.155	666.439	622.170	701.389	691.204	701.199	664.205	683.835
	Dif. (%)	0,0	-1,9	2,7	-4,1	8,1	6,5	8,1	2,4	5,4
Vespasiano	Abs.	118.557	118.580	128.846	117.544	128.142	133.478	175.513	145.634	139.602
	Dif. (%)	0,0	0,02	8,7	-0,9	8,1	12,6	48,0	22,8	17,8
Sarzedo	Abs.	29.889	30.084	36.342	32.090	-----	-----	28.293	34.698	23.902
	Dif. (%)	0,0	0,7	21,6	7,4	-----	-----	-5,3	16,1	-20,0
Baldir	Abs.	8.071	7.792	8.529	9.483	-----	-----	8.615	10.530	9.217
	Dif. (%)	0,0	-3,5	5,7	17,5	-----	-----	6,7	30,5	14,2

Fonte: Elaboração Própria.

O modelo por pixel (baseado em dados de 2000) apresentou ERM dos setores de 44,2% e erro total (macro) de 6,7% para o município de Baldim. Este resultado, além de ser considerado aceitável, é mais coerente com a validação externa (para o ano de 2010, cujo erro total foi de -1,8%), se comparado com Sarzedo, que apresentou erro total de 20,1% em 2010 e erro total de -5,3% em 2015.

A ausência de um padrão nas estimativas calculadas por modelos no nível dos pixels e para municípios de baixa densidade se confirma novamente, pois, os erros totais observados pelos modelos construídos a partir de dados de 2010 foram bem superiores ao erro total de 6,7%, estimado pelo modelo que parte de dados de 2000: ERM dos setores de 82,1% e erro total de 30,5% pelo modelo de pixel que utiliza os setores de 2010, e ERM por setores de 28,2% e erro total de 14,2% pelo modelo de pixels que utiliza a grade estatística de 2010. Esperava-se que o menor intervalo temporal entre as imagens de 2010 e 2015, bem como a menor mudança populacional observada no quinquênio (se comparado ao período 2000-2015) fosse contribuir para a produção de estimativas mais precisas, o que não ocorreu, a partir dos dados de Baldim (bem como para Sarzedo).

A qualidade das estimativas no nível dos pixels encontra-se condicionada à manutenção entre população e reflectâncias (das bandas espectrais) ao longo do tempo, entre diferentes imagens, o que parece não ocorrer nos municípios de baixa densidade populacional. Ao mesmo tempo, o padrão de mudança dessa relação parece não ser bem estabelecido, talvez devido a fatores associados à imagem, como diferentes níveis de iluminação (ou de cobertura de nuvens), ou ainda à perda de 22% de informação nas imagens Landsat 7 a partir de 2003 (referente a um problema de ruído), que, em alguma medida, pode comprometer os resultados.

A correção por interpolação (via aplicação de filtro nas imagens) pode mascarar particularidades no uso das ocupações (e conseqüentemente no nível das reflectâncias), que refletiriam, por exemplo, na relação das variáveis explicativas (as reflectâncias) com a variável dependente (população), comprometendo o ajuste dos modelos. Uma hipótese é que em áreas de baixa densidade populacional, a interpolação referente à perda de informação em 22% dos pixels resulte no maior comprometimento das estimativas, pois, nesses municípios, há uma maior alteração do uso do solo ao longo da imagem.

Já a justificativa para os baixos erros encontrados nas estimativas baseadas em modelos no nível dos pixels para Contagem seria, talvez, a do menor impacto que a perda de informação das imagens (e a correção por interpolação) geraria para municípios com maior densidade populacional, marcados por uma mancha densa e praticamente única, com menos mudanças no uso ao longo do município. Já os municípios com menor densidade apresentam uma maior alteração no uso do solo ao longo da imagem (dentro do município), com manchas de ocupação mais esparsadas, o que comprometeria a qualidade na definição das áreas ocupadas, após a correção das imagens, e, conseqüentemente, a qualidade das estimativas.

Porém, deve-se ressaltar que a classificação das áreas de ocupação humana não foi restrita ao procedimento automático, realizado através das imagens Landsat 7. Fora realizada uma segunda etapa de interpretação de imagens de satélite de alta resolução, que minimizam os problemas de definição de áreas ocupadas mencionado anteriormente. Em todo o caso, perde-se em qualidade no que diz respeito à informação da reflectância propriamente dita, que por sua vez é resultado de interpolação; o que, sem dúvida, deve ser considerado como um problema, porém de difícil mensuração.

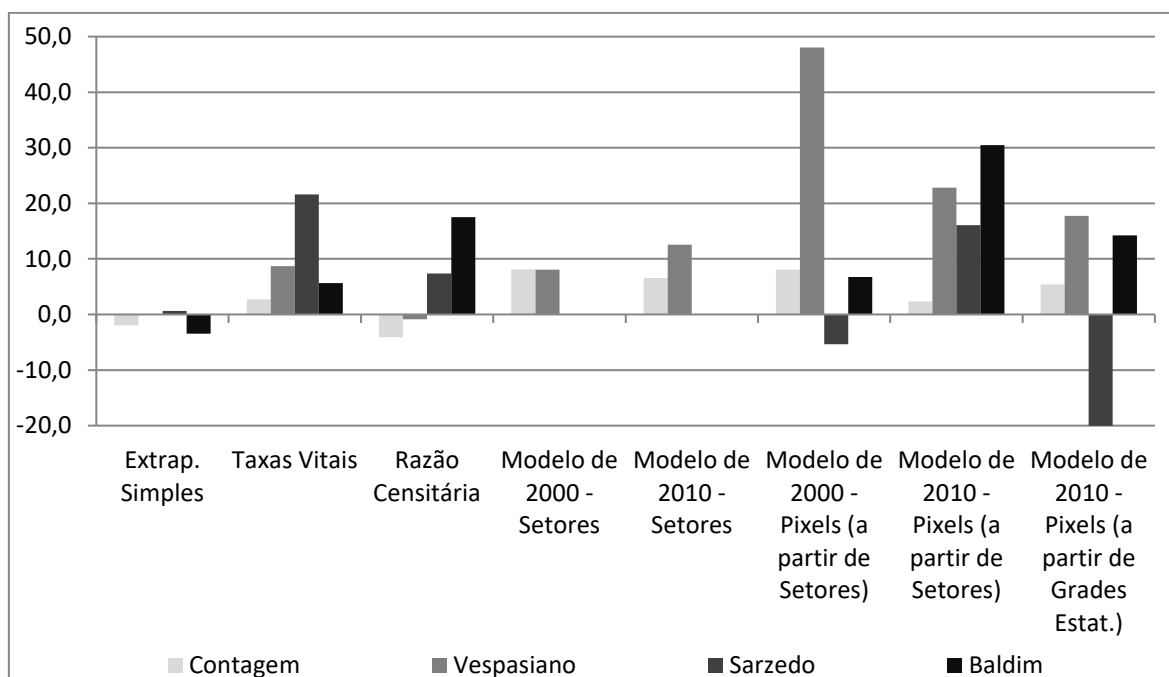
A Gráfico 8 mostra as diferenças percentuais de todas as estimativas (demográficas e via sensoriamento remoto) em relação às estimativas pós-censitárias do IBGE, para o ano de 2015 e para os quatro municípios em estudo.

O Gráfico 8 mostra a dificuldade de se verificar um padrão bem estabelecido de erro para 2015, muito embora Contagem (que expressa ocupação de alta densidade) tenha apresentado bom ajuste aos dois métodos, principalmente pelo método dos pixels, enquanto Vespasiano mostrou erro menor nas estimativas calculadas pelo método dos setores censitários (município que expressa a ocupação de baixa densidade). Baldim e Sarzedo, cujo pequeno número de setores censitários não permitiu a construção de modelos no nível dos setores e a partir da média das reflectâncias, apresentaram os piores resultados, nos modelos baseados em pixel.

Por outro lado, as estimativas demográficas mostraram, de forma geral, menores oscilações e menores erros totais, se comparados às estimativas via sensoriamento

remoto, especialmente o método por Extrapolação Simples, no qual se ressalta a facilidade e a rapidez de processamento.

GRÁFICO 8 – Diferenças Percentuais das Estimativas via Sensoriamento Remoto (baseadas em modelos criados em 2000 e 2010) e de estimativas demográficas, para o ano de 2015, em relação às Estimativas Pós-Censitárias do IBGE de 2015



O próximo capítulo trata da conclusão deste trabalho, onde serão analisadas as contribuições dos resultados encontrados para o campo das estimativas via sensoriamento remoto. No contexto das principais vantagens e limitações dos métodos empregados, pretende-se discutir sobre a eficácia das estimativas calculadas por imagens orbitais de média resolução espacial Landsat 7 ETM+ no campo das estimativas populacionais para pequenas áreas, e, de maneira geral, para a Demografia. Pretende-se ainda propor agendas futuras para esse campo de estudo.

6 CONCLUSÃO

Nas últimas décadas, tem se observado o aumento pela busca por informações mais detalhadas sobre a dinâmica populacional, diante à crescente demanda dos setores público e privado para o planejamento nas mais diferentes áreas – como educação, saúde, meio ambiente, economia, pesquisas de mercado, dentre outras – e, como consequência, o intervalo temporal das pesquisas censitárias (entre 5 e 10 anos) tem sido insuficiente para atender ao aumento dessa demanda. Ademais, o aumento da complexidade na busca por informações populacionais não se restringe à questão temporal: nos últimos anos, tem crescido a demanda por informações para pequenas áreas, seja para o setor privado, como também para a alocação de fundos governamentais, bem como para o delineamento e monitoramento de políticas públicas.

Diante a essa crescente demanda, fontes de dados alternativos vêm sendo utilizados, muito embora essas informações – como registros civis, censos escolares e cadastramentos governamentais – não sejam disponíveis (ou são restritas) para a maioria dos países, além do problema de sub registro e de falta de qualidade das informações.

A disseminação dos Sistemas de Informações Geográficas, observado nas últimas décadas, somado à crescente disponibilização de imagens de sensoriamento remoto, tem contribuído para o desenvolvimento de metodologias no campo das estimativas populacionais para pequenas áreas, com destaque para um conjunto de métodos que utilizam de imagens orbitais. Nesse contexto, o presente trabalho teve como objetivo contribuir na avaliação da capacidade das imagens orbitais de média resolução espacial Landsat ETM+ na produção de estimativas populacionais no nível municipal.

Para a avaliação da eficácia das estimativas produzidas via imagens orbitais de média resolução espacial, buscou-se a seleção de municípios que apresentassem diferentes comportamentos em relação à taxa de crescimento, mas, principalmente, municípios que apresentassem diferentes tamanhos e níveis de densidade

populacionais. Assim, estimativas via sensoriamento remoto foram calculadas para quatro municípios localizados na RMBH: Contagem (alta densidade e taxa de crescimento positiva), Vespasiano (baixa densidade e taxa de crescimento positiva e acelerada), Sarzedo (baixa densidade e taxa de crescimento positiva e acelerada) e Baldim (baixíssima densidade e taxa de crescimento negativa).

Foram construídos modelos baseados em dados de 2000 e 2010, primeiramente através do método desenvolvido por Harvey (2002a), que utiliza a média das reflectâncias das bandas da imagem Landsat 7 ETM+, observada no nível dos setores censitários (como variáveis explicativas), e a densidade populacional dos setores (como variável dependente), aplicado somente nos municípios de Contagem e Vespasiano, em função de problemas nas bases de setores censitários dos municípios de Sarzedo e Baldim. Num segundo momento, modelos foram construídos através do método desenvolvido por Harvey (2002b), que utiliza as reflectâncias e a população distribuída para o nível dos pixels, através de um modelo denominado de “regressão iterada”, etapa que fora aplicada para os quatro municípios em estudo.

Para todos os modelos foi realizada uma etapa de validação, através da seleção de uma amostra (de 2/3) dos bancos de dados para a construção dos modelos e o restante dos dados para a produção de estimativas, a partir da mesma base de dados (ou seja, a partir da mesma imagem de satélite). Essa etapa constitui a validação interna, na medida em que testa o ajuste dos modelos à mesma imagem utilizada para a construção dos parâmetros dos modelos, sendo este o teste usualmente verificado na literatura na análise das estimativas via sensoriamento remoto.

A partir dos modelos baseados em dados de 2000, foram calculadas estimativas para os anos de 2010 e 2015. As estimativas para 2010 foram comparadas com o Censo 2010, etapa denominada como de validação externa, na medida em que utiliza outro banco de dados (a partir de outra imagem) para o teste dos modelos; o que representa uma inovação nos estudos sobre estimativas que partem dos métodos empregados neste trabalho. As estimativas de 2010 foram ainda comparadas com a população de 2010, oriunda das projeções de 2008 e 2013 do IBGE (desagregadas para o nível municipal, através do método AiBi). Neste caso,

teve-se como objetivo analisar os quão similares são as estimativas produzidas a partir de diferentes fontes de dados, o que contribui para a discussão quanto à eficácia das estimativas produzidas por imagens orbitais Landsat 7 ETM+.

Num segundo momento, modelos construídos a partir de dados de 2000 e de 2010 foram utilizados para a elaboração de estimativas para o ano de 2015. No caso dos modelos baseados em dados de 2000 foi utilizada a população dos setores censitários, enquanto que, nos modelos baseados em dados de 2010, foram construídos modelos a partir dos setores censitários de 2010 e através da grade estatística de 2010 (do IBGE). Essas estimativas via sensoriamento remoto para 2015 foram então comparadas com estimativas pós-censitárias do IBGE e com outros três tipos de estimativas demográficas (pelos métodos de Extrapolação Simples, Taxas Vitais e Razão Censitária).

As diferenças encontradas entre as estimativas permitem identificar se há diferença significativa entre as estimativas baseadas em dados de 2000 (mais distantes temporalmente em relação a 2015) e as estimativas que utilizam dados mais recentes, de 2010, bem como permitem a análise da diferença de erro entre as estimativas via sensoriamento remoto e as estimativas demográficas, o que contribui na análise da contribuição das estimativas baseadas em imagens orbitais de média resolução para o campo das estimativas para pequenas áreas e, de forma mais abrangente, para a Demografia.

Para o município de Contagem, marcado pela elevada densidade populacional, os resultados mostraram o potencial dos métodos empregados para diversas finalidades no campo das estimativas para pequenas áreas, bem como para a Demografia. Em primeiro lugar, os baixíssimos erros encontrados na etapa de validação interna – tanto para os modelos no nível dos setores censitários (6,3% pelo modelo de 2000 e -1,5% pelo modelo de 2010), como para todos os modelos no nível dos pixels (0,02% pelo modelo de 2000, -0,4% pelo modelo de setores de 2010 e 0,4% pelo modelo de grade estatística de 2010) – encontram-se em consonância com a literatura (0,47% encontrado por Reis, 2005, para o município de Belo Horizonte, também marcado por elevada densidade populacional) e comprova o grande potencial desses métodos (especialmente o baseado em Harvey 2002b) em distribuir a população no espaço contínuo. Os bons resultados

da validação interna para Contagem também comprovam o grande potencial em se estimar áreas de elevada densidade, para o mesmo ano e mesma cena; o que pode ser muito útil em áreas cuja cobertura censitária não tenha sido satisfatória.

Já os bons resultados encontrados nas etapas de validação externa e de análise comparativa, especialmente dos modelos no nível dos pixels, mostram o potencial desse método para a produção de estimativas pós-censitárias confiáveis, para municípios com elevada densidade populacional. O modelo no nível dos pixels apresentou validação externa de -0,7% (sem considerar o problema de cobertura censitária), ficando abaixo dos erros das projeções do IBGE (em torno de 5%). Já na análise comparativa com a estimativa pós-censitária do IBGE para o ano de 2015, os erros totais ficaram entre 2,4% (modelo de 2010 por setores) e 8,1% (modelo de 2000 por setores), podendo ser considerados aceitáveis, embora estejam em níveis superiores às estimativas demográficas, que ficaram entre 2% e -4%. Já o erro de 11,1% na validação externa do modelo no nível dos setores pode também ser considerado aceitável, sendo que os resultados da análise comparativa para o ano de 2015 apresentaram erros similares aos modelos no nível dos pixels (8,1% pelo modelo de 2000 e 6,5% pelo modelo de 2010).

Em que pese os bons resultados, foram significativas as diferenças entre as validações externas e as análises comparativas para o ano de 2015 (tendo como referência os mesmos modelos), o que evidencia a dificuldade de manutenção da relação entre reflectância e densidade populacional no transcorrer do tempo (entre diferentes imagens de satélite).

Seja do ponto de vista das estimativas propriamente ditas, seja como ferramenta para o planejamento urbano ou ainda como ferramenta auxiliar na análise de dados, para a definição de tendências em estudos de projeção populacional, pode-se considerar os resultados para o município de Contagem como muito promissores, o que resulta numa gama de possibilidades de aplicação para áreas de elevada densidade populacional, com grande potencial de aplicação para o campo das estimativas para pequenas áreas e para a Demografia.

Já o município de Vespasiano é marcado pela baixa densidade populacional e pelo crescimento populacional positivo (na década de 2000). Os erros de validação

interna dos modelos baseados em setores censitários foram baixíssimos (1,5% pelo modelo que parte de dados de 2000 e 2000 e 2% pelo modelo de 2010); padrão seguido pelos modelos no nível dos pixels, cujos erros – de 5,2% pelo modelo de 2000, 0,2% pelo modelo de 2010 por setores censitários e 1,6% pelo modelo de 2010 por grade estatística – são considerados aceitáveis. Esses resultados mostram que a eficácia das estimativas via imagens orbitais de média resolução espacial não está restrita aos municípios de elevada densidade populacional, e indicam aplicar-se, também, aos de baixa densidade, quando a análise do erro é realizada por meio de dados do mesmo ano (a partir da mesma imagem), denominada neste trabalho de validação interna.

Porém, a análise da validação externa para Vespasiano – 1,9% pelo modelo no nível dos setores e 31,6% pelo modelo no nível dos pixels (ambos a partir de dados de 2000) – mostra que, em municípios de baixa densidade, a relação entre reflectância e população aparentemente sofre alteração ao longo dos anos, enquanto que a relação entre a média das reflectâncias e o logaritmo da densidade se mantém num nível bem mais estável, ao longo do tempo.

Na análise comparativa, a partir de dados de 2015, o modelo no nível dos setores apresentou erro total de 8,1% e 12,6%, respectivamente às fontes de dados de 2000 e 2010, que, embora aceitáveis, apresentam patamares superiores ao observado na validação externa (1,9%), o que representa mais uma incompatibilidade temporal na relação população e reflectância. No nível dos pixels, os modelos apresentaram erros totais elevados – 48%, 22,8% e 17,8%, respectivamente aos modelos por setores de 2000, de 2010 e por grade estatística de 2010 – corroborando o resultado ruim da validação externa (da ordem de 31,6%).

Da mesma forma, o município de Sarzedo (que não contou com a construção de modelos no nível dos setores censitários) apresentou validação interna 0,09% (para o modelo de 2000), -0,10% (para o modelo de setores de 2010) e -0,40% (para o modelo de grades estatísticas de 2010). Ao contrário do bom ajuste à produção de estimativas baseados na mesma fonte de dados, o teste de validação externa foi de 20,1% de erro total, corroborando o resultado observado para o município de Vespasiano, e a conclusão de que, em municípios de baixa densidade

populacional, modelos construídos num dado ano para um outro ano (e consequentemente outra imagem) não resulta em boas estimativas, em função da mudança da relação entre população e as reflectâncias das bandas,

Em contrapartida, as estimativas demográficas mostraram resultados mais robustos (mais próximos das estimativas pós-censitárias do IBGE), com destaque para o método de Extrapolação Simples (0,02%). Em relação à análise comparativa para o ano de 2015, em que pese o pequeno erro total do modelo no nível dos pixels e por setores de 2000 (-5,3%), a discrepância em relação ao erro total do mesmo modelo para o ano de 2010 (20,1% na validação externa) mostra novamente a mudança da relação entre população e reflectância entre imagens de satélite de anos diferentes. Ademais, os erros totais do modelo baseado em setores de 2010 (16,1%) e do modelo baseado em grade estatística de 2010 (-20,0%), além de serem elevados, apresentaram resultados inferiores à estimativa calculada através do modelo que parte de dados de 2000, o que ratifica a ausência de um padrão temporal na relação entre população e reflectância, para municípios com baixa densidade populacional.

Os resultados para Contagem, Vespasiano e Sarzedo mostram a limitação do modelo no nível dos pixels (tal como em Harvey, 2002b) para a estimativa de municípios marcados por baixa densidade populacional, em períodos pós-censitários. Por outro lado, as estimativas via modelos baseados em pixels mostram-se bem ajustados para a produção de estimativas pós-censitárias em municípios com alta densidade populacional, bem como para a distribuição espacial da população, e para a construção de estimativas para um mesmo ano. Já os modelos no nível dos setores se mostraram bem ajustados a municípios com elevada densidade (embora com erros totais superiores aos modelos no nível dos pixels), e, principalmente, bem ajustados a municípios com baixa densidade populacional.

O município de Baldim, o único a apresentar taxa de crescimento negativa nas últimas duas décadas, mostrou bom ajuste nos testes de validação interna (entre -0,3% e 0,3%) e erro pequeno na validação externa (-1,8%). Por um lado, este último resultado contraria a tendência observada em Vespasiano e Sarzedo, quanto à ineficácia dos municípios de baixa densidade populacional em relação à validação

externa. Por outro, o erro total na análise comparativa para 2015 – 6,7% para o modelo de 2000, 30,5% para o modelo de setores de 2010 e 14,2% para o modelo de grade estatística de 2010 – mostram que o bom resultado da validação externa para Baldim pode estar relacionado à manutenção da relação entre população e reflectância entre as imagens de 2000 e 2015 (o que não ocorre necessariamente entre outras imagens).

Contudo, o destaque para Baldim está na validação externa, no qual fora estimada uma população inferior àquela observada nos Censos de 2000 e 2010, o que representa um bom indicativo de aplicabilidade do modelo por pixels na produção de estimativas para municípios que apresentam taxa de crescimento negativa. Diante do indicativo de bom ajuste dos modelos no nível dos setores para a produção de estimativas pós-censitárias, em municípios com baixa densidade populacional (observado em Vespasiano), o cálculo de estimativas no nível dos setores para 2015, para Sarzedo e Baldim, poderia corroborar a eficácia dos modelos de Harvey (2002a) para municípios de baixa densidade, bem como o potencial desses métodos para municípios que perdem população. Contudo optou-se pela não produção das estimativas no nível dos setores para Sarzedo e Baldim, para 2015, em função da ausência de validação externa; o que, após análise dos resultados, mostrou-se pouco eficiente, na medida em que a validação externa não serviu de referência (de padrão em relação ao percentual de erro) para as estimativas pós-censitárias, para municípios de baixa densidade.

A qualidade das estimativas no nível dos pixels encontra-se condicionada à manutenção da relação entre população e reflectâncias (das bandas espectrais) ao longo do tempo, entre diferentes imagens, o que parece não ocorrer nos municípios de baixa densidade populacional. Ao mesmo tempo, o padrão de mudança dessa relação parece não ser bem estabelecido, talvez devido a fatores associados à imagem, como diferentes níveis de iluminação (ou de cobertura de nuvens), ou ainda à perda de 22% de informação nas imagens Landsat 7 a partir de 2003 (referente a um problema de ruído), que, em alguma medida, pode ter comprometido os resultados.

A correção por interpolação (via aplicação de filtro nas imagens) pode mascarar particularidades no uso das ocupações – e conseqüentemente no nível das

reflectâncias – especialmente nos municípios com baixa densidade populacional (no qual apresentam uma maior alteração no uso do solo, ao longo da imagem) comprometendo o ajuste dos modelos.

No ano de 2010, as estimativas calculadas através das projeções do IBGE apresentaram erro pequeno, inferior à 6% para todos os municípios, sendo que a diferença entre as duas projeções (Revisão 2008 e 2013) foi muito pequena para todos os municípios analisados (sempre próximo a 1%). A pequena diferença entre as projeções é interessante, a considerar que a Revisão 2008 parte da população de 1980 e se trata de uma projeção no nível do País, desagregada por AiBi para as Unidades da Federação, para que, em seguida, fosse desagregada para os municípios deste trabalho. Contudo, o mais importante a considerar refere-se aos pequenos erros observados, o que, somado à facilidade de construção das informações, pode-se afirmar que representam boas alternativas para a produção de estimativas no nível municipal.

No ano de 2015, os resultados das estimativas demográficas oscilaram mais (entre -4,1% e 21,6%), se comparados às projeções do IBGE, muito embora deva ser considerado que a própria informação de referência (as estimativas pós-censitárias do IBGE) é uma projeção, que pode, eventualmente, não representar a realidade dos municípios no período pós-censo. Em todo o caso, o método de Extrapolação Simples apresentou resultados coerentes com as estimativas pós-censitárias do IBGE. Somado ao baixo custo de construção e de interpretação do indicador, pode-se afirmar que este método deve ser considerado como referência para a produção de estimativas em pequenos intervalos temporais. Ainda assim, deve-se considerar sempre que, em pequenas áreas, as oscilações no tamanho dos municípios podem, eventualmente, não reproduzir num período pós-censitário – mesmo considerando um curto intervalo de tempo – as tendências observadas no passado recente.

De um modo geral, o maior equilíbrio dos erros percentuais (por volta de 5%) das projeções do IBGE desagregadas para os municípios em 2010 (em paralelo à oscilação das validações externas, conforme município e densidade), e os menores erros das estimativas demográficas em 2015, quando comparadas às estimativas pós-censitárias via sensoriamento remoto, sugerem, num primeiro olhar (talvez descuidado), que as diferentes estimativas demográficas são mais eficazes –

inclusive se considerar o baixo custo para aplicação dos métodos – se comparadas às estimativas por imagens orbitais. Porém, ao analisar os melhores resultados encontrados pelas estimativas via sensoriamento remoto, é possível identificar importantes contribuições das imagens orbitais de média resolução espacial Landsat 7 ETM+ para o campo das estimativas.

Em primeiro lugar, para todos os municípios, bases de dados e modelos, a validação interna foi abaixo de 6% (e em quase todos os casos abaixo de 1%), o que mostra a grande eficácia das estimativas via sensoriamento remoto nas situações em que o modelo construído é aplicado à mesma cena da imagem e ao mesmo ano. Esse tipo de aplicação pode ser muito útil na construção de estimativas para municípios que apresentam problemas em estimativas demográficas, ou em situações em que se constata problema de cobertura censitária, bem como para o planejamento regional de uma dada região (ou conjunto de regiões).

Estimativas calculadas a partir de dados de um mesmo ano também pode ser úteis para a distribuição da população no espaço contínuo – por exemplo, através do método de Harvey (2002b) – que utiliza o método de iteração para a redistribuição da população no nível dos pixels, o que representa um tipo de informação estratégica para os setores público e privado, para uma diversidade de estudos, tais como: alocação de equipamentos públicos (como escolas, hospitais, serviços de transporte público), para a alocação de programas sociais, bem como para pesquisa de mercado, dentre outras finalidades.

Em relação à validação externa, o método no nível dos pixels (inspirado em Harvey, 2002b) apresentou resultados consistentes para município com alta densidade populacional – com validação externa de -0,7% de erro total, a partir do modelo por pixel e setores de 2000 e erro total de 2,4%, por meio de modelo por pixel e setores de 2010 – o que permite concluir sobre a eficácia dos métodos via imagens orbitais de média resolução espacial, para a produção de estimativas pós-censitárias, em municípios de alta densidade. Aliás, este tipo de município, normalmente de grande porte populacional, é o que demanda por métodos automáticos de produção de estimativas, dada a dificuldade em se utilizar métodos baseado em contagem domiciliar, por exemplo. Essa é uma importante contribuição do trabalho, pois, como já fora comentado, a literatura sobre o tema é restrita à testes de validação

interna, o que limita a análise do potencial e da contribuição das estimativas por imagens orbitais.

O potencial de se produzir estimativas para anos pós-censitários, com baixo percentual de erro, pode ser utilizado como informação sintomática da mudança populacional, e que pode ser ainda utilizado como informação auxiliar para a elaboração de projeções populacionais; como, por exemplo, para a elaboração de cenários de migração. Há uma grande dificuldade em se criar cenários para a componente migratória, especialmente em pequenas áreas, e, sob o pressuposto de que os cenários de fecundidade e mortalidade são confiáveis num curto intervalo temporal, a diferença entre a população esperada (na ausência de migração), mantida as taxas projetadas para fecundidade e mortalidade, e a população estimada via sensoriamento remoto, pode, por exemplo, ser considerado o saldo migratório ocorrido no período.

O mesmo conjunto de aplicações citados acima pode ser utilizado para municípios de baixa densidade populacional, diante aos bons resultados de validação externa encontrados a partir do método no nível dos setores (inspirado em Harvey, 2002a), para o município de Vespasiano. O erro total de 1,9% em 2010 (referente à validação externa), e de 8,1% em 2015, a partir da mesma base de dados (de setores de 2000) são considerados aceitáveis, o que mostra o potencial do sensoriamento remoto para a produção de estimativas para municípios com baixa densidade populacional.

Ainda sobre este tipo de município, em que pesem as estimativas ruins calculadas através do método no nível dos pixels para municípios de baixa densidade, em Baldim – para além do bom ajuste do modelo no nível dos pixels na validação externa (-1,8%), resultado não corroborado pelas estimativas de 2015 – verificou-se a produção de estimativas via sensoriamento remoto que indicam o decréscimo populacional. Ainda que este resultado seja incipiente – pois sem dúvida demanda o desenvolvimento de outros estudos que corroborem os resultados encontrados neste trabalho – pode-se considerar o resultado de validação externa de Baldim como muito promissor para o campo das estimativas via imagens orbitais de média resolução espacial. Promissor na medida em que amplia as possibilidades de utilização dos métodos via sensoriamento remoto para a produção de estimativas

de municípios com os mais variados tipos de comportamento demográfico, tanto em relação ao tamanho como em relação à direção da mudança populacional (se crescimento ou decréscimo).

Esse resultado sugere, portanto, que os modelos captam a perda de população, ao contrário da hipótese inicial, de que não há mudança no uso do solo em decorrência de perda populacional (causada, por exemplo, por um surto emigratório, que não transformaria as estruturas ocupadas em outro uso). De fato, deve-se aprofundar os estudos nesse sentido, pois, há de se ressaltar que o resultado encontrado foi calculado através do método por pixels, que, por sua vez, apresentou resultados ruins para municípios com baixa densidade populacional, como é o caso de Baldim.

Contudo, são grandes as limitações das estimativas produzidas a partir de imagens de média resolução espacial, especialmente no contexto da avaliação quanto à contribuição e eficácia desse tipo de aplicação para o campo das estimativas para pequenas áreas e para a Demografia. De um modo geral, nos capítulos teóricos ficou evidenciado, através de muitos exemplos, a importância do conceito de replicabilidade, ou seja, do desenvolvimento de métodos que possam ser utilizados para áreas mais extensas, de modo a atender as demandas de estudos voltados para o planejamento regional (como, por exemplo, estudos para um conjunto de municípios, tal como nas projeções demográficas convencionais).

Sobre este aspecto, em primeiro lugar, ainda que se observe o aumento da disponibilização de imagens gratuitas e de melhor resolução nos últimos anos, e mesmo considerando os avanços tecnológicos, a construção dos bancos de dados, as classificações das imagens e as etapas de modelagem ainda representam um elevado custo operacional, o que dificulta a produção de estimativas para áreas mais extensas (como foi possível constatar na etapa de produção dos dados deste trabalho).

Em segundo, foi possível constatar que não há um modelo padrão que possa ser replicado para todas as unidades espaciais (ex.: municípios), como afirma Harvey (2002b). Cada município apresenta suas particularidades em relação ao uso e ocupação, e, conseqüentemente, demandam não apenas por diferentes variáveis explicativas, mas também por bases de dados diferentes e por métodos diferentes.

Foi possível constatar ao longo de todo o trabalho questões como a adequação de métodos específicos para municípios com alta e baixa densidade populacional, ou a alternância em relação à base de dados com melhor ajuste (ora os setores de 2000 ou de 2010, ora as grades estatísticas). Em relação às variáveis, deve-se destacar que, o logaritmo da densidade e o percentual de área ocupada melhoraram sobremaneira os resultados dos modelos no nível dos setores, enquanto que aglomerado subnormal melhoraram os modelos, porém em menor proporção (como observado em Reis, 2005). Porém, nos modelos construídos no nível dos pixels, não foi possível identificar um padrão (um conjunto de variáveis) bem definido, com melhores resultados.

Deve-se considerar ainda as próprias características da fonte de dados, que se traduzem em limitações para a produção das estimativas. As imagens de satélite sofrem um conjunto de interferências, como a data da imagem, a iluminação solar, o percentual de nuvens, além de limitações dos próprios sensores – como o problema de perda de 22% das informações nas imagens Landsat7 ETM+ desde 2003, aspecto este que inclusive fora comentado neste trabalho como hipótese para os problemas encontrados nas estimativas, por modelos no nível dos pixels, para municípios de baixa densidade populacional. Portanto, por todos os elementos discutidos, o autor deste trabalho concorda com Wu *et al.* (2005), quando esses afirmam que “*more studies are needed before remote sensing can be applied to population estimation on an operational basis*” (Wu *et al.*, 2005).

Outra limitação diz respeito à escala com o qual as estimativas são produzidas. Se, de um lado, o mapeamento dasimétrico permite o cálculo de estimativas para quaisquer unidades administrativas (inclusive intramunicipais), os ERM's dos setores – com resultados de aproximadamente 35%, próximos ao encontrado por Reis (2005), de 30,4%, e acima do encontrado em Harvey (2002b), de 14% – impedem que sejam produzidas estimativas para unidades muito desagregadas.

Algumas limitações dos resultados encontrados podem estar associadas às escolhas definidas neste trabalho. A ineficiência da base de dados por grade estatística na produção de estimativas de melhor qualidade, quando comparado às bases de dados por setores censitários, serve como agenda para estudos futuros. O caminho é concentrar os esforços de processamento num número maior de

municípios, em detrimento à utilização de diferentes bases de dados populacionais. Vale lembrar que o método baseado em pixels utiliza a regressão iterada, que redistribui a população até que o ajuste do modelo seja estabilizado, o que diminui a influência do número de unidades espaciais da fonte de dado populacional (sem contar que a grade estatística possui aproximadamente 5% de subenumeração, como fora comentado no tópico 4.3.2.6).

Outro aspecto diz respeito à variável “situação urbano/rural”. A considerar que essa variável não melhorou a qualidade das estimativas, talvez um caminho para estudos futuros seja a exclusão de pixels de baixíssima densidade (normalmente localizado em áreas rurais), como adotado por Harvey (2000b) e Reis (2005) e cujos resultados, segundo estes autores, melhoram as estimativas. Por outro lado, como afirma Reis (2005) “o modelo baseado em pixels não leva em conta a densidade populacional”, de modo que “a incorporação de uma variável auxiliar como indicação de densidade populacional permitiria que os modelos fossem ajustados separadamente para setores muito densos e pouco densos” (Reis, 2005). Contudo, ressalta-se a dificuldade em se obter este tipo de informação, principalmente em períodos pós-censitários. Uma alternativa apontada por Harvey (2002b) seria a separação da classificação das áreas ocupadas para mais de um estrato, separando as amostras de treinamento para cada tipo de ocupação identificada, muito embora o próprio autor assuma que este tipo de procedimento envolve um grande custo adicional de tempo.

O autor deste trabalho concorda com Reis (2005), quando afirma que as estimativas por sensoriamento remoto “seria uma tecnologia auxiliar na obtenção e informações sobre o número de habitantes de uma região, não substituindo a realização das contagens populacionais nem tampouco dos censos demográficos”. Contudo, a gama de possibilidades de aplicação, atualmente – considerando as imagens orbitais de média resolução, objeto deste trabalho – são inúmeras (como amplamente discutido neste trabalho), o que tende a ampliar cada vez mais a agenda de pesquisa sobre estimativas populacionais via sensoriamento remoto, inclusive através do aumento de estudos interdisciplinares no campo da Demografia.

No futuro, o avanço na disponibilização de novas fontes de informação, através de novas imagens, com melhores resoluções espacial, espectral e radiométrica, em paralelo aos avanços computacionais e tecnológicos, ampliará, ainda mais, as oportunidades no campo das estimativas para pequenas áreas. Na visão do autor deste trabalho, a direção é a da utilização de imagens de alta resolução espacial para a produção de estimativas populacionais em áreas intra-urbanas, e a partir de métodos mais sofisticados de detecção e mensuração da densidade populacional (como é o caso das informações por radar, que captam a elevação das edificações), como já tem sido verificado em alguns esforços pioneiros.

Porém, em termos de replicabilidade para a escala regional – que é, na visão do autor deste trabalho, o elemento mais importante na análise da eficácia e da contribuição das estimativas via sensoriamento remoto para o campo da Demografia, o que inclusive motivou a utilização de imagens gratuitas e de resolução espacial que possibilite a replicação das estimativas para áreas mais extensas - essa perspectiva de imagens de alta resolução disponíveis de forma massificada, com série temporal disponível e com viabilidade em termos de processamento, é sem dúvida uma perspectiva de muito longo prazo.

Num cenário de curto e médio prazos, a ampliação das séries históricas de imagens gratuitas, como as imagens dos sensores Landsat, em especial o sensor Landsat 8 (que conta com avanços importantes), além do surgimento e disponibilização de novas imagens (com melhor resolução espacial), permitirá o aprofundamento desse campo do conhecimento, trazendo consigo o aprofundamento de uma agenda de pesquisa com novas perspectivas na direção da construção de estimativas pós-censitárias municipais ainda mais confiáveis. Aliás, o aumento da precisão das estimativas via imagens orbitais – resultado dos avanços tecnológicos e da crescente disponibilização de dados – deverá caminhar juntamente com a maior precisão quanto aos métodos, bases de dados e variáveis a serem adotadas, como resultado inclusive do aumento do número de pesquisas e estudos que incorporam essa agenda de pesquisa.

Esses avanços esperados, somados ao aumento do potencial de replicabilidade das estimativas para áreas mais extensas (a partir dos avanços computacionais), muito provavelmente ampliarão, pelo menos no médio prazo, a eficiência e a

contribuição das estimativas via sensoriamento remoto para o campo das estimativas para pequenas áreas, e, especialmente, para a Demografia.

REFERÊNCIAS BIBLIOGRÁFICAS

ALHO, J. M., SPENCER, B. D. **Statistical Demography and Forecasting**. New York: Springer, 2005, 410p. Cap. 10. Errors in Census Numbers, p.296-326.

ALMEIDA, C. M. **Modelagem da dinâmica espacial como uma ferramenta auxiliar ao planejamento: simulação de mudanças de uso da terra em áreas urbanas para as cidades de Bauru e Piracicaba (SP), Brasil**. 2004. 321f. Tese (Doutorado em sensoriamento remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2004.

ÁLVAREZ, G. **Estimación de población en áreas menores mediante variables sintomáticas: una aplicación para los departamentos de la República Argentina (1991 y 1996)**. Santiago, Chile: Centro Latinoamericano y Caribeño de Demografía (CELADE), jan., 2001. 36p.

ASSUNÇÃO, R. M. Método Bayesiano de Relação de Coortes para Projeções de Pequenas Áreas. In: ENCONTRO DA ASSOCIAÇÃO BRASILEIRA DE ESTUDOS POPULACIONAIS, 8, 2002. Ouro Preto. **Anais**. Ouro Preto: ABEP, 2002.

BAKER, W. L. A Review of Models of Landscape Change. **Landscape Ecology**. v.2, n.2, p.111-133, 1989.

BALK, D., YETMAN, G. **The Global Distribution of Population: Evaluating the gains in resolution refinement**. New York: Center for International Earth Science Information Network (CIESIN). Columbia University. 2004, 15p.

BATTY, M. COUCLELIS, H., EICHEN, M. Urban Systems as Cellular Automata. **Environment and Planning B: Planning and Design**. v.24, p.159-164, 1997.

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, p.103-117, 2007.

BRITO, L. P. G., CAVENAGHI, S., JANNUZZI, P. M. Estimativas e projeções populacionais para pequenos domínios: uma avaliação da precisão para municípios do Rio de Janeiro em 2000 e 2007. **Revista Brasileira de Estudos Populacionais**. v.27, n.1, p.35-57, jan./jun., 2010

BRYAN, T. Population Estimates. In: SIEGEL, J. S., SWANSON, D. A. (Orgs.) **The methods and materials of demography**. San Diego, California: Elsevier Academic Press, 2004. p.523-560.

BUENO, M. C. D. **Grade estatística: uma abordagem para ampliar o potencial analítico de dados censitários**. 2014. 263f. Tese (Doutorado em Demografia) – Instituto de Filosofia e Ciências Humanas: Universidade Estadual de Campinas, Campinas, 2014.

CABRERA, M. **Estimación de Población em áreas menores con métodos que utilizan variables sintomáticas**. Uruguay: Comisión Sectorial de Población, 2011, 61p.

CEDEPLAR. Estimativas de População para o Brasil: Total do País, Unidades Federativas e Municípios, 2010-2030. **Relatório de Pesquisa**. Belo Horizonte, set, 2014.

DÓRIA, V. E. M. **Sensoriamento Remoto de Luzes Noturnas para Estimativas Populacionais em Escalas Regional e Local: os Casos do Distrito Florestal Sustentável da BR 163 (PA) e da Região Metropolitana de São Paulo**. 2015. 129f. Dissertação (Mestrado em sensoriamento remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2015.

DÓRIA, V. E. M., KAMPEL, S. A., MONTEIRO, A. M. V. Estimativa e distribuição espacial da população urbana com imagens de satélite de luzes noturnas: um estudo para a Região Metropolitana de São Paulo, Brasil, com o sensor Visible / Infrared Imaging Radiometer Suite (VIIRS). **Geografia (Rio Claro. Online)**. v.41, n.1, 2016.

DUCHESNE, L. Proyecciones de poblacion por sexo y edad para áreas intermedias y menores: método “relación de cohortes”. In: CENTRO LATINO AMERICANO DE DEMOGRAFIA – CELADE (Org.) **Métodos para Proyecciones Subnacionales de Población**. Santiago: CELADE e IDRC, 1989. p.71-126.

DURAND, C., PEREIRA, M. N., MOREIRA, J. C., FREITAS, C. C. Análise da Correlação entre População e Área Urbana (km²) visando a Inferência Populacional por Meio do Uso de Imagens Orbitais. **Geografia**. v.16, n.2, p.113-142, jul./dez. 2007.

EICHER, C. L., BREWER C. A. Dasymeric Mapping and Areal Interpolation: Implementation and Evaluation. **Cartography and Geographic Information Science**. v.28, n.2, p.125-138, 2001.

ESQUIVEL, E. A. C. Variables sintomáticas em las estimaciones poblacionales a nivel cantonal em Costa Rica. **Notas de Población**. n.71, p.51-72, 2001.

FORESTI, C. **Estimativas Populacionais e de Crescimento de Áreas Urbanas no Estado de São Paulo, com Utilização de Imagens LANDSAT**. 1978. 124f. Tese (Dissertação em sensoriamento remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 1978.

FRANÇA, V. O. **Avaliação de métodos basimétricos para estimativa populacional em pequenas áreas**. 2012. 103f. Dissertação (Mestrado em Estudos Populacionais e Pesquisa Social) – Escola Nacional de Ciências Estatísticas, Rio de Janeiro, 2012.

FREIRE, F. H. M. A. Projeção Populacional com Estimadores Bayesianos Espaciais. In: ENCONTRA DA ASSOCIAÇÃO BRASILEIRA DE ESTUDOS POPULACIONAIS, 13,2002. Ouro Preto. **Anais**. Ouro Preto: ABEP, 2002.

GEORGE, M. V., SMITH, S. K., SWANSON, D. A., TAYMAN, J. Population projections. SIEGEL, J. S., SWANSON, D. A. (Orgs.) **The methods and materials of demography**. San Diego, California: Elsevier Academic Press, p.561-602, 2004.

GONZÁLEZ, L. M., TORRES, E. Estimaciones de población en áreas menores en América Latina: revisión de métodos utilizados. In: CAVENAGHI, S. **Estimaciones y proyecciones de población em América Latina: Desafios de uma agenda pendiente**. Rio de Janeiro, ALAP, 2012.

GONZÁLEZ, R. C., WOODS, R. E. **Processamento de Imagens Digitais**. São Paulo, 2000.

HARVEY, J. T. Estimating census district populations from satellite imagery: some approaches and limitations. **International Journal of Remote Sensing**. v. 23, n. 10, p. 2071-2095, 2002a.

_____. Population estimation models based on individual TM pixels. **Photogrammetric Engineering & Remote Sensing**. v.68, n.11, p. 1181-1192, 2002b.

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

IMHOFF, E. V., POST, W. Microsimulation methods for population projection. **Population: An English Selection**. v.10, n.1, p.97-138, 1998.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Estimativas da População Residente para os Municípios e para as Unidades da Federação Brasileiros com Data de Referência em 1º de Julho de 2015**. Rio de Janeiro, ago., 2015.

_____. **Grade Estatística**. Rio de Janeiro, 2016.

_____. **Projeções da População: Brasil e Unidades da Federação**. Rio de Janeiro, 2013.

_____. **Censo Demográfico Brasileiro**. Rio de Janeiro: Série Relatórios Metodológicos, 2010.

_____. **Projeção da População do Brasil por Sexo e Idade – 1980-2050: Revisão 2008**. Rio de Janeiro: Estudos & Pesquisas, Informação Demográfica e Socioeconômica, 2008.

_____. **Metodologia adotada nas estimativas populacionais municipais**. Rio de Janeiro, 2002.

_____. **Censo Demográfico Brasileiro**. Rio de Janeiro, 2000.

INSTITUTO NACIONAL DE PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP). **Censo Escolar**. Brasília, 2017. Disponível em: <<http://portal.inep.gov.br/web/guest>>. Acesso em 01 jan. 2017.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE). **Processamento Digital de Imagens**. São José dos Campos, 2001.

JANNUZZI, P. M. Cenários futuros e projeções populacionais para pequenas áreas: método e aplicação para distritos paulistanos 2000-2010. **Revista Brasileira de Estudos de População**. v.24, n.1, p.109-136, 2007.

JARDIM, M. L. T. **Metodologias de estimativas e projeções populacionais para o Rio Grande do Sul e seus municípios**. Porto Alegre: Secretaria da Coordenação e Planejamento – Fundação de Economia e Estatística Siegfried Emanuel Heuser, 2003. 68p.

JARDIM, M. L. T. Uso de variables sintomáticas para estimar la distribución espacial de población: aplicación a los municipios de Rio Grande do Sul, Brasil. **Notas de Población**. n.71, p.21-50, 2001.

JENSEN, J. R. **Sensoriamento remoto do Ambiente: Uma Perspectiva em Recursos Terrestres**. Tradução: EPIPHANIO, J. C. N. *et al.*. São José dos Campos, 2009. 598p..

JENSEN, J. R., COWEN, D. C. Remote Sensing of Urban/Suburban Infrastructure and Socio-Economic Attributes. **Photogrammetric Engineering & Remote Sensing**. v.65, n.5, p.611-622, maio. 1999.

KAMPEL, S. A. **Geoinformação para estudos demográficos: representação espacial de dados de população na Amazônia Brasileira**. 2003. 166f. Tese (Doutorado em Engenharia) – Universidade de São Paulo, Escola Politécnica, São Paulo, 2003.

KAMPEL, S. A., MONTEIRO, A. M. V., CÂMARA, G., QUINTANILHA, J. A. DMSP/OLS night-time light imagery for urban population estimates in the Brazilian Amazon. **International Journal of Remote Sensing**. v.27, n.5, p.855-870, mar. 2006.

KAMPEL, S. A., 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 and Environment**. v.34, p.142-170, 2012.

LANGFORD, M. An Evaluation of Small Area Population Estimation Techniques Using Open Access Ancillary Data. **Geographical Analysis**. v.45, p.324-344, 2013.

LEDENT, J., ZENG, Y. Multistate Demography. In: ZENG, Y. (Org.) **Demography: Encyclopedia of life support systems**. Oxford, UK: Eolss Publishers Co. Ltd., 2010. p.139-163.

IIASAKA, J., HEGEDUS, E. Population Estimation from Landsat Imagery. **Remote Sensing of Environment**. v.12, p.259-272, 1982.

LIU, X., CLARKE, K. **Estimation of Residential Population using High Resolution Satellite Imagery**. UCSB: Santa Bárbara, Department of Geography, University of California, 2002.

LO, C. P., WELCH, R. Chinese urban population estimates. **ANNALS OF THE ASSOCIATION OF AMERICAN GEOGRAPHERS**. v.67, n.2, p.246-253, 1977.

LU, D., WENG, Q., LI, G. Residential population estimation using a remote sensing derived impervious surface approach. **International Journal of Remote Sensing**. v.27, n.16, p.3553-2570, ago. 2006.

MACHADO, C. M. **Projeções multirregionais de população: o caso brasileiro (1980-2020)**. 1993. 147f. Tese (Doutorado em Demografia) – Centro de Desenvolvimento e Planejamento Regional, Universidade Federal de Minas Gerais, Belo Horizonte, 1993.

MADEIRA, J. L., SIMÕES, C. C. S. Estimativas preliminares da população urbana e rural segundo as Unidades da Federação, de 1960/1980 por uma nova metodologia. **Revista Brasileira de Estatística**. v.33, n.129, p.3-11, 1972.

MENESES, P. R., ALMEIDA, T. **Introdução ao Processamento de Imagens de Sensoriamento Remoto**. Brasília, 2012.

MENNIS, J. Generating Surface Models of Population Using Dasymetric Mapping. **The Professional Geographer**. v.55, n.1, p.31–42, 2003.

MENNIS, J., HULTGREN, T. Intelligent Dasymetric Mapping and Its Application to Areal Interpolation. **Cartography and Geographic Information Science**. v.33, n.3, p.179-194, 2006.

MUNIZ, J. O. **Análise e pRevisão demográfica através de matrizes de crescimento e distribuição populacional intermunicipal**. Belo Horizonte: Centro de Pesquisas Quantitativas em Ciências Sociais – CEPQS/UFMG, 2013. 29p.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION – NASA. **Landsat 7 Science Data Users Handbook**. United States of America, 2017. 186p.

O'NEILL, B. C., BALK, D., BRICKMAN, M., EZRA, M. **A Guide to Global Population Projections**. **Demographic Research**. v.4, n.8, p.203-288, jun.2001.

OLIVEIRA, L. C. S., FREITAS, M. P. S., DIAS, M. R. M. L. (*et al.*). **Censo Demográfico 2000: resultados da pesquisa de avaliação da cobertura da cobertura**. Rio de Janeiro: IBGE, Coordenação Técnica do Censo Demográfico, 2003. 143p.

PETROV, A. One Hundred Years of Dasymetric Mapping: Back to the Origin. **The Cartographic Journal**. v.49, n.3, p.256-264, ago., 2012.

REIS, I. A. Estimação da população dos setores censitários de Belo Horizonte usando imagens de satélite. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 12, 2005. Goiânia. **Anais**. Goiânia: SBSR, 2005, v.1, p.2741-2748.

SANTÉ, I., GARCÍA, A. M., MIRANDA, D., CRECENTE, R. Cellular automata models for the simulation of real-world urban processes: a review and analysis. **Landscape and Urban Planning**. v.96, p.108-122, 2010.

SANTOS, R. O., BARBIERI, A. F. Projeções populacionais em pequenas áreas: uma avaliação comparativa de técnicas de extrapolação matemática. **Revista Brasileira de Estudos Populacionais**. Rio de Janeiro, v.32, n.1, p.139-163, jan./abr. 2015.

SHERBININ, A., BALK, D., YAGER, (*et al.*). **A CIESIN Thematic Guide to Social Science Applications of Remote Sensing**. New York: Center for International Earth Science Information Network (CIESIN). Columbia University. out., 2002, 68p.

SHRYOCK, H. S., SIEGEL, J. S. **The Methods and Materials of Demography**. Washington, 2.ed. D.C.: U.S. Government Printing Office, 1973. 888p. Cap. 23. Population Estimates, p.725-809.

SILVA, L. G. C, SANTOS, R. O. A utilização de Variáveis Sintomáticas para Estimativas de Populações Municipais do Estado do Paraná. In: ENCONTRO DA ASSOCIAÇÃO BRASILEIRA DE ESTUDOS POPULACIONAIS, 20, 2016. Foz do Iguaçu. **Anais**. Foz do Iguaçu: ABEP, 2016.

SMITH, S. K., MORRISON, P. A. Small-Area and Business Demography. In: POSTON, D., MICKLIN, M. (Orgs.) **Handbook of Population**. New York: Springer Publishers, 2005.

SMITH, S. K., TAYMAN, J., SWANSON, D. A. **State and Local Population Projections: Methodology and Analysis**. New York, Boston, Dordrecht, London, Moscow: Kluwer Academic Publishers, 2002, 426p.

SOARES-FILHO, B. S., CERQUEIRA, G. C., PENNACHIN, C. L. DINAMICA – a stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier. **Ecological Modelling**. v.154, p.217-235, 2002.

SOUZA, I. M. **Sensoriamento Remoto Orbital Aplicado a Estudos Urbanos**. São José dos Campos: Instituto Nacional de Pesquisas Espaciais – INPE. 2012. 39p.

SOUZA, I. M. **Análise do Espaço Intra-urbano para Estimativa Populacional Intercensitária Utilizando Dados Orbitais de Alta Resolução Espacial**. 2004. 92f. (Mestrado em sensoriamento remoto) – Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2004.

SUTTON, P., ROBERTS, D., ELVIDGE, C., MEIJ, H. A Comparison of Nighttime Satellite Imagery and Population Density for the Continental United States. **Photogrammetric Engineering & Remote Sensing**. v.63, n.11, p.1303-1313, nov. 1997.

TOBLER, W. R. Satellite Confirmation of Settlement Size Coefficients. **Area**. v.1, n.3, p.30-34, 1969.

TOMÁS, L. R. **Inferência populacional urbana baseada no volume de edificações residenciais usando imagens IKONOS-II e dados LIDAR**. 2010. 108f. Tese (Doutorado em sensoriamento remoto) – Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2010.

UMBELINO, G. J. M. **Simulações de distribuição espacial domiciliar e projeções demográficas intraurbanas com apoio de geotecnologias**. 2012. 194f. Tese (Doutorado em Demografia) – Centro de Desenvolvimento e Planejamento Regional (CEDEPLAR/UFMG), Belo Horizonte, 2012.

UNESCO. The effect of varying population estimates on the calculation of enrolment rates and out-of-school rates. UNESCO: Institute of Statistics. **Information Paper**. n.36, out., 2017.

UNITED NATIONS. Manuals on methods of estimating population: manual I - methods of estimating total population for current dates. **Population Studies**. New York, n.10, 1952.

UNITED NATIONS. Manuals on methods of estimating population: manual III - methods for population projections by sex and age. **Population Studies**. New York, n.25, 1956.

UNITED STATES GEOLOGICAL SURVEY (USGS). **Landsat Enhanced Thematic Mapper Plus (ETM+)**. Estados Unidos, 2017. Disponível em: <<https://lta.cr.usgs.gov/LETMP>>. Acesso em: 01 jan. 2017.

UNITED STATES GEOLOGICAL SURVEY (USGS). **Science Processing Architecture – ESPA**. Estados Unidos, 2017. Disponível em: <<https://espa.cr.usgs.gov/index>>. Acesso em: 01 jan. 2017.

WALDVOGEL, B. **Técnicas de projeção populacional para o planejamento regional**. Belo Horizonte: UFMG/CEDEPLAR, 1998.

WHITE, R., ENGELEN, G. Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns. **Environment and Planning A**. v.25, p.1175-1199, 1993.

WILLEKENS, F. **Description of the multistate projection model (multistate model for biographic analysis and projection)**. Netherlands: Netherlands Interdisciplinary Demographic Institute (NIDI), 2006. 122p.

WILSON, T., ROWE, F. The Forecast Accuracy of Local Government Area Population Projections: A Case Study of Queensland. **Australasian Journal of Regional Studies**. v.17, n.2, 2011.

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, C., MURRAY, A. Population Estimation Using Landsat Enhanced Thematic Mapper Imagery. **Geographical Analysis**. v.39, p.26-43, 2007.

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.

XIE, Y., WENG, A., WENG, Q. Population Estimation of Urban Residential Communities Using Remotely Sensed Morphologic Data. **IEEE Geoscience and Remote Sensing Letters**. v.12, n.5, p.1111-1115, maio. 2015.

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.