

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

Járvis Campos¹; José Irineu Rangel Rigotti²; Antônio Miguel Vieira Monteiro³; Ilka Afonso Reis⁴; Emerson Augusto Baptista⁵

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 7%, 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: Sensoriamento Remoto, Mapeamento Dasimétrico, Estimativa para Pequenas Áreas.

¹ Professor (DDCA/UFRN)

² Professor (CEDEPLAR/UFMG)

³ Professor (SER/INPE)

⁴ Professora (ICEX/UFMG)

⁵ Professor (Asian Demographic Research Institute)

1. Introdução

Informações sobre o tamanho, a composição e o ritmo de mudança populacional tem sido objeto de interesse crescente do setor público, seja para o planejamento econômico como 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 (5 ou 10 anos) normalmente não é adequado ao ritmo da mudança populacional. Ademais, as informações censitárias são limitadas, devido ao elevado custo e diante à 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).

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 das informações (United Nations, 1952; Shryock e Siegel, 1973; Bryan, 2004; Smith e Morrison, 2005).

Em contrapartida à falta de qualidade dessas fontes de informação, nos últimos anos tem se observado o aumento na demanda por informações para pequenas áreas, por parte dos setores público e privado. Diante disso, um conjunto de métodos que utilizam de variáveis sintomáticas⁶ tem sido desenvolvido para a produção de estimativas de pequenas áreas, e que devem ser compreendidas como aproximações da mudança populacional, fundamentais

⁶ 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).

para o planejamento nos mais diversos segmentos (Shryock e Siegel, 1973; Bryan, 2004; Smith e Morrison, 2005; Jannuzzi, 2007; Cabrera, 2011; Silva e Santos, 2016).

No Brasil, por exemplo, 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 Unidades 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.

Há algum tempo, o uso de dados derivados dos programas espaciais, em particular os dados de imageamento por satélite, chamaram a atenção para a possibilidade de seu uso no auxílio às estimativas populacionais, 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 novos métodos 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; Wu *et al*, 2005; Lu *et al*, 2006; Bhaduri *et al*, 2007; Wu e Murray, 2007; 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) – Contagem, Vespasiano, Sarzedo e Baldim – para os anos de 2000, 2010 e 2015. As estimativas

produzidas foram validadas, e, então, comparadas com dados censitários e com estimativas baseadas em métodos demográficos.

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, dentre aquelas cujo acesso é livre não dispõem 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 à 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

infra-estrutura computacional pública e compartilhada para o tratamento desta informação, cuja manutenção envolve também custos.

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

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 gratuitamente, 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 (compreendem, 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 (Lo e Welch, 1977; Iisaka e Hegedus, 1982; Holt *et al*, 2004; 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), o que justifica a escolha por esse tipo de imagem de média resolução espacial.

2. Metodologia

De maneira geral, para a construção das estimativas via sensoriamento remoto, parte-se primeiramente de anos censitários, no qual a distribuição espacial mais desagregada da população obtida no censo (por exemplo, os setores censitários, por sua vez restritos aos limites administrativos) é transformada para uma distribuição espacial da população ainda mais refinada, a partir da utilização de dados auxiliares. Esse procedimento, denominado de modelos de superfície ou dasimétricos, correspondem, portanto, à redistribuição geográfica de dados populacionais com um maior nível de detalhamento espacial, a partir da utilização de dados auxiliares – neste caso, imagens de satélite –

através do registro das informações populacionais 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). Além disso, esses modelos permitem discriminar estruturas do espaço físico não ocupadas (como vegetação, rios, entre outros), muito embora seja importante observar as incertezas derivadas desse processo (Mennis, 2003; Langford, 2013).

A partir do mapeamento dasimétrico realizado para um dado ano censitário, é construído um modelo econométrico, que por sua vez é aplicado a outro mapa dasimétrico, correspondente a um ano pós-censitário, obtendo-se, assim, a estimativa para o referido ano. O detalhamento desses procedimentos será apresentado no próximo tópico.

2.1. Métodos

Ao longo das décadas observou-se um processo de evolução nos métodos utilizados para a produção de estimativas via Sensoriamento Remoto. Os primeiros métodos consideravam a relação direta entre crescimento das áreas urbanas e o tamanho populacional (Tobler, 1969; Lo e Welch, 1977; Foresti, 1978; Sutton, 1997). 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 (Holt *et al*, 2004; Eicher e Brewer, 2001; Mennis, 2003; Mennis e Hultgren, 2006; Badhuri *et al*, 2007). Há ainda um conjunto de métodos que estimam a população através 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 (Souza, 2004; Xie *et al*, 2015; Tomás, 2010).

Nos métodos descritos até aqui, a densidade populacional foi associada às características ou classes de uso do solo. Mais recentemente, surgiram um conjunto de métodos baseados em pixels, no qual a população pode ser correlacionada diretamente à reflectância espectral dos valores das imagens dos pixels (Wu *et al*, 2005). Os trabalhos elaborados por Harvey (2002a e 2002b) são os mais citados na literatura internacional, são considerados clássicos dentre os

trabalhos que tratam da correlação entre população e os pixels das imagens de satélite; e, portanto, foram escolhidos para a realização deste trabalho.

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 (MQO).

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*"). Já a variável dependente pode ser transformada, para a melhoria nas estimativas do modelo (como, por exemplo, a transformação da população para a função logaritmica).

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].

Dentre os resultados, ressalta-se que 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, 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).

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).

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 (sendo que define-se um critério de parada, quando o aumento do R^2 não for significativo).

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.

No Brasil, Reis (2005) aplicou as técnicas apresentadas em Harvey (2002a e 2002b), na estimacão 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 mostraram um erro relativo mediano de 30,4% (contra 14% no trabalho de Harvey, 2002b), mas bem superior no nível agregado (ou seja, no nível municipal), com erro relativo total de apenas -0,06%.

2.2. Estruturação dos Dados

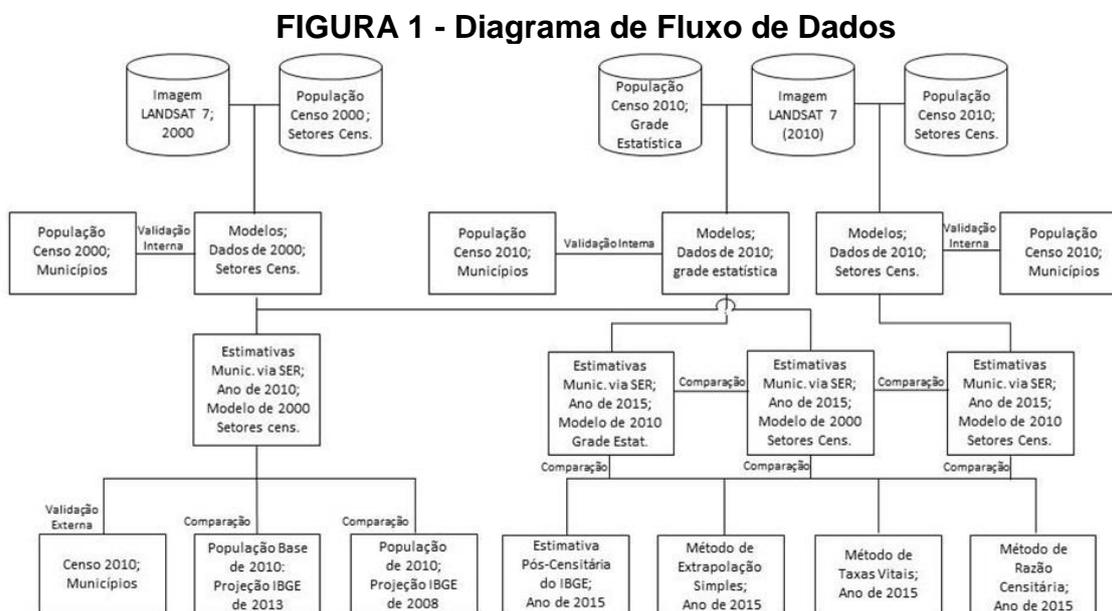
Para a construção dos modelos e das estimativas, a primeira etapa consistiu na classificação de imagens Landsat 7 ETM+ (registradas, sem correção atmosférica e com baixo percentual de nuvens), no qual foram mapeadas as áreas ocupadas para cada um dos quatro municípios, e para os três anos em análise (2000, 2010 e 2015). 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 as áreas e anos estudados, o que representou 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+. A partir da distribuição da população dos municípios na escala dos pixels – pelo critério de distribuição homogênea, para cada conjunto de pixels localizados num dado setor censitário ou grade estatística – 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), por meio da utilização de grades de células vetoriais. Na etapa seguinte, essas informações serviram de base para a construção dos modelos.

Para os mapeamentos dasimétricos baseados nos dados de 2000, foram utilizados os setores censitários 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.

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 (2015), para o ano de 2010. 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). A Figura 1 apresenta o diagrama de fluxo dos bancos de dados.



Fonte: Elaboração Própria.

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 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.

A Figura 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 a Figura 3 apresenta as variáveis do modelo de Harvey (2002b).

FIGURA 1 – 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.

FIGURA 2 – 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), que por sua vez diferem em relação à utilização das variáveis explicativas. Por uma questão tamanho do artigo, optou-se por não apresentar os

modelos, de modo que os resultados que serão apresentados correspondem aos modelos com melhor ajuste, para cada ano e município.

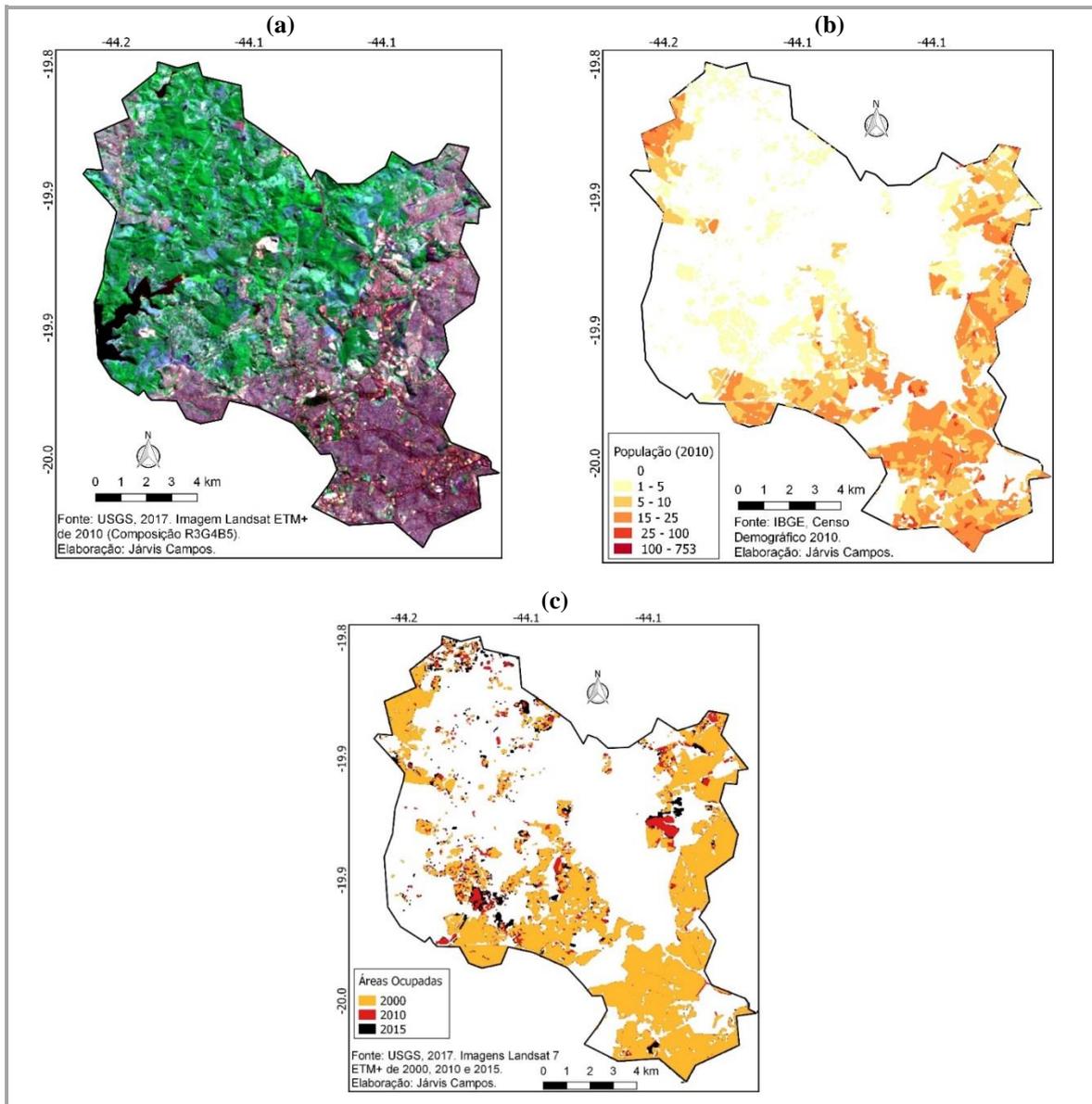
Foram definidos três municípios que apresentaram taxas de crescimento positiva e com diferentes tamanhos populacionais (Contagem, Vespasiano e Sarzedo), o que contribui para a análise da eficácia das estimativas. Isso ocorre, pois, partindo da taxa de crescimento positiva (que representa o comportamento ideal para os modelos, que se ajustam bem aos municípios que ganham população), os erros das estimativas passam a ser resultado muito mais do tamanho e do nível de densidade dos municípios. Contudo, foi selecionado ainda o município de Baldim, que, por sua vez, foi o único a apresentar taxa de crescimento negativa no período 2000-2010.

3. Resultados e Discussão

Conforme foi descrito no item anterior, para cada município foram elaborados mapas referentes às imagens de satélite Landsat 7 (ETM+), denominados de composição R3G4B5, bem como mapas dasimétricos, para todos os municípios e anos em estudo. Contudo, por uma questão de limitação de espaço no artigo, para efeito de ilustração será apresentado apenas o mapa de composição R3G4B5 e o mapa dasimétrico, ambos referentes a 2010, e o mapa de evolução da mancha de ocupação, para o município de Contagem (Figura 4).

Para os quatro municípios em análise, os baixíssimos erros encontrados na etapa de validação interna (não apresentados em tabela) – tanto para os modelos no nível dos setores censitários como para todos os modelos no nível dos pixels, encontram-se em consonância com a literatura e comprova o grande potencial desses métodos (especialmente o baseado em Harvey 2002b) em distribuir a população no espaço contínuo, bem como um grande potencial em se estimar populações para o mesmo ano e mesma cena; o que pode ser muito útil em áreas cuja cobertura censitária não tenha sido satisfatória.

FIGURA 3 – Composição R3G4B5 de Imagens Landsat 7 ETM+ para o ano de 2010 (a), mapeamento dasimétrico para o ano de 2010 (b) e evolução da mancha de ocupação entre 2000 e 2015 (c) – Município de Contagem



Fonte: USGS, 2017. Elaboração Própria.

Para o município de Contagem, 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%, enquanto que, em 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 (Tabela 1).

TABELA 1 – 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.

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, também para os modelos no nível dos setores, apresentaram erros de 8,1% (pelo modelo de 2000) e 6,5% (pelo modelo de 2010) [Tabela 2].

Já no município de Vespasiano, a análise da validação externa – 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 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 (Tabela 1).

Na análise comparativa para o ano 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%).

TABELA 1 – 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-Baldirim, 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
Baldirim	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 município de Sarzedo (que não contou com a construção de modelos no nível dos setores censitários) apresentou um erro de validação externa na ordem de 20,1%, 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 conseqüentemente outra imagem) não resulta em boas estimativas, em função da mudança da relação entre população e as reflectâncias das bandas.

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.

O município de Baldim, o único a apresentar taxa de crescimento negativa nas últimas duas décadas, mostrou 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 negativo.

Diante ao 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, se mostrou 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.

4. Conclusões

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, podem-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.

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).

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 comprometer os resultados.

Foi realizada uma correção por interpolação (via aplicação de filtro nas imagens), mas que, por sua vez, pode mascarar particularidades no uso das ocupações – e consequentemente 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.

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 comparado 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 comparada 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 podem 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.

Sobre os municípios de baixa densidade populacional, em que pese 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 pelo desenvolvimento de outros estudos que corroborem os resultados encontrados neste trabalho – pode-se considerar o resultado de validação externa de Baldim como 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).

5. Referências Bibliográficas

- 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.
- 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.
- CABRERA, M. **Estimación de Población em áreas menores con métodos que utilizan variables sintomáticas**. Uruguay: Comisión Sectorial de Población, 2011, 61p.
- 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.
- 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. Dasymeric Estimation of Population Density and Areal Interpolation of Census Data. **Cartography and Geographic Information Science**. v.31, n.2, p.103-121, 2004.
- IASAKA, J., HEGEDUS, E. Population Estimation from Landsat Imagery. **Remote Sensing of Environment**. v.12, p.259-272, 1982.
- 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.
- 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.
- LANGFORD, M. An Evaluation of Small Area Population Estimation Techniques Using Open Access Ancillary Data. **Geographical Analysis**. v.45, p.324-344, 2013.
- 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.
- MENNIS, J. Generating Surface Models of Population Using Dasymeric Mapping. **The Professional Geographer**. v.55, n.1, p.31-42, 2003.

MENNIS, J., HULTGREN, T. Intelligent Dasyetric Mapping and Its Application to Areal Interpolation. **Cartography and Geographic Information Science**. v.33, n.3, p.179-194, 2006.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION – NASA. **Landsat 7 Science Data Users Handbook**. United States of America, 2017. 186p.

REIS, I. A. Estimaco da populao dos setores censitrios de Belo Horizonte usando imagens de satlite. In: SIMPSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 12, 2005. Goinia. **Anais**. Goinia: SBSR, 2005, v.1, p.2741-2748.

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 utilizao de Variveis Sintomticas para Estimativas de Populaes Municipais do Estado do Paran. In: ENCONTRO DA ASSOCIAO BRASILEIRA DE ESTUDOS POPULACIONAIS, 20, 2016. Foz do Iguau. **Anais**. Foz do Iguau: 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.

SOUZA, I. M. **Anlise do Espao Intra-urbano para Estimativa Populacional Intercensitria Utilizando Dados Orbitais de Alta Resoluo Espacial**. 2004. 92f. (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais (INPE), So Jos 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.

TOMS, L. R. **Inferncia populacional urbana baseada no volume de edificaes residenciais usando imagens IKONOS-II e dados LIDAR**. 2010. 108f. Tese (Doutorado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais (INPE), So Jos dos Campos, 2010.

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 STATES GEOLOGICAL SURVEY (USGS). **Landsat Enhanced Thematic Mapper Plus (ETM+)**. Estados Unidos, 2017. Disponvel em: < <https://lta.cr.usgs.gov/LETMP>>. Acesso em: 01 jan. 2017.

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 Dasyetric Mapping Techniques for Small-Area Population Estimates. **Cartography and Geographic Information Science**. v.37, n.3, p.199-214, 2010.