

Avaliação de métodos de desagregação para geração de grades de população

Maria do Carmo Dias Bueno^{1*}
Álvaro de Oliveira D'Antona²

¹ Mestre em Engenharia Civil pela Universidade Federal do Rio de Janeiro (UFRJ), Mestre em Geomática pela UFRJ, Doutoranda em Demografia pelo Instituto de Filosofia e Ciências Humanas (UNICAMPI).

² Mestre em Antropologia pela Universidade Estadual de Campinas (UNICAMPI), Doutor em Ciências Sociais pela UNICAMPI, Doutor em População e Ambiente pela Indiana University, Estados Unidos.

Resumo O objetivo deste trabalho é avaliar métodos de desagregação de dados para gerar grades de população, utilizando como referência uma grade gerada através da agregação de microdados censitários, o que permite obter resultados mais precisos acerca do desempenho desses métodos. Para realizar a avaliação foram selecionadas três metodologias. A primeira utiliza uma matriz contendo os erros e acertos na identificação de áreas povoadas e não povoadas, permitindo avaliar a acurácia da distribuição espacial da população. A segunda utiliza uma regressão linear cujos coeficientes permitem avaliar a qualidade da adequação aos dados de referência. Finalmente, utilizamos uma fórmula para calcular a diferença entre os valores de população, indicando se há uma subestimação ou superestimação. Os resultados sugerem que a escolha da metodologia mais adequada depende dos objetivos a que se destina a grade de população, bem como da disponibilidade e qualidade dos dados auxiliares, além das características da área de estudo.

Palavras-chave: distribuição da população; grades de população; agregação; desagregação.

1. Introdução

Os censos de População e Habitação produzem informações essenciais para a criação e monitoração de políticas públicas nos níveis nacional, estadual e municipal, e também para a tomada de decisão relacionada com investimentos. Os dados obtidos com o censo são utilizados para o planejamento de serviços para a população, como educação, saúde, serviços sociais e serviços públicos (água, esgoto, energia, gás), para o planejamento de serviços de emergência (desastres naturais) e diversos estudos e análises (mapeamento da pobreza, estudos epidemiológicos, análise de mercados). Esses dados também são utilizados para fornecer referências para as projeções populacionais que são utilizadas pelo governo federal na definição das cotas do Fundo de Participação dos Estados e dos Municípios (IBGE, 2010).

Um dos benefícios das geotecnologias aplicadas à área das estatísticas oficiais é a função de integração de dados oferecida pelos Sistemas de Informação Geográfica – SIGs, facilitando o relacionamento de informações oriundas de diferentes áreas. Isso levou, sem sombra de dúvidas, a uma utilização mais abrangente das informações estatísticas e também a uma maior exigência, por parte dos usuários, de dados geograficamente mais desagregados.

Existem dois problemas relacionados com a utilização de dados agregados em unidades geográficas que são bastante conhecidos e citados na literatura. O primeiro deles diz respeito a não coincidência entre as unidades para as quais

os dados censitários são disponibilizados e as unidades para as quais os dados são necessários. Essas últimas podem ser unidades baseadas em características sociais, ambientais ou mesmo unidades artificiais geradas em meio digital, como áreas de influência ou *buffers*. Referências sobre esta questão podem ser vistas, por exemplo, em Clarke e Rhind (1992), Hogan (1992), Deichmann (1996), Ojima e Martine (2012) e Martine e Schensul (2013). O segundo problema está relacionado à estabilidade temporal dessas unidades geográficas, uma vez que tanto as unidades político-administrativas como as unidades operacionais estatísticas sofrem alterações ao longo do tempo, fato este que dificulta a realização de análises temporais (EXETER et al., 2005; NORMAN, REES; BOYLE, 2003).

A utilização de unidades geográficas baseadas em células regulares dispostas num sistema de grade parece ser a solução para esses dois problemas, pois as células têm pequenas dimensões podendo ser consideradas como tijolos que se juntam para formar qualquer recorte espacial desejado além de não se alterarem ao longo do tempo, já que são unidades artificiais e arbitrárias e, uma vez determinadas, não necessitam sofrer alterações (MARTIN, 2000; TAMMILEHTO-LUODE et al., 2000; RUSANEM et al., 2001; TAMMILEHTO-LUODE, 2011).

Existem duas abordagens metodológicas principais para a geração de dados de população em grades: a abordagem *bottom-up* ou de agregação e a *top-down* ou de desagregação. A primeira abordagem utiliza os microdados censitários associados aos seus respectivos atributos de localização para a agregação em células. A geração de

*bueno.mariadocarmo@gmail.com

grades de população utilizando metodologias que tem por base esta abordagem é realizada desde meados da década de 90 em países do norte europeu, como Finlândia e Suécia (TAMMILEHTO-LUODE; BACKER; ROGSTAT, 2010). Isso se tornou possível porque esses países desenvolveram um sistema estatístico com base em pontos codificados que representam as edificações, o que permite o relacionamento entre as diversas bases de registros que são utilizadas para a geração das estatísticas oficiais (NAÇÕES UNIDAS, 2007). Esta metodologia também tem sido utilizada mais recentemente por países que realizam o censo na sua forma tradicional de coleta de dados em campo e que utilizam geotecnologias para a localização espacial dos dados coletados, como, por exemplo, a Estônia (EFGS, 2012).

Nos países onde a abordagem de agregação não é adotada pela agência nacional de estatística, os pesquisadores utilizam métodos espaciais e/ou estatísticos para realocar os dados disponibilizados em unidades irregulares para células regulares com o objetivo de enfrentar os problemas citados anteriormente. Esta abordagem é conhecida como desagregação e envolve diferentes metodologias, algumas com a utilização de dados auxiliares e outras não. Entre os métodos que não utilizam dados auxiliares podemos citar a ponderação zonal, a interpolação picnofilática (TOBLER, 1979) e a estimação *kernel* (BRACKEN; MARTIN, 1989). Entre os métodos que usam dados auxiliares, o mapeamento dasimétrico é o mais conhecido, sendo que diferentes dados podem auxiliar na estimativa de população, como classificação de uso e cobertura das terras derivada de imagens de satélite (EICHER; BREWER, 2001; MENNIS, 2003), malha de sistema viário (XIE, 1995; REIBEL;

BUFALINO, 2005) ou endereços geocodificados (ZANDBERGEN, 2011).

A avaliação da acurácia da abordagem de desagregação geralmente é realizada utilizando unidades geográficas para onde existem dados censitários publicados, sendo poucos os casos em que é realizada uma comparação com grades de população desenvolvidas com a abordagem de agregação, já que esses dados são regularmente produzidos por poucas agências de estatística. Um exemplo deste tipo de avaliação pode ser visto em Steinnocher et al. (2011). O objetivo deste artigo é avaliar a qualidade de quatro métodos de desagregação em comparação com os resultados obtidos com a agregação de microdados censitários espacializados e agregados em células regulares. Os materiais e métodos utilizados no desenvolvimento do trabalho e a discussão dos resultados serão mostrados a seguir.

2. Materiais e Métodos

A FIG. 1 mostra a área de estudo selecionada para a análise das diferentes metodologias de espacialização da população. A área se localiza no estado de São Paulo, abrangendo totalmente o município de Limeira, tendo cerca de 600 km² e população de 276.022 habitantes, segundo o Censo Demográfico 2010 (IBGE, 2011b). O município de Limeira está inserido na área de interesse do Laboratório de Geografia de Riscos e Resiliência do Centro de Ciências Humanas e Sociais Aplicadas da Faculdade de Ciências Aplicadas da Universidade Estadual de Campinas e, por este motivo, tem sido objeto de diversos estudos e análises.

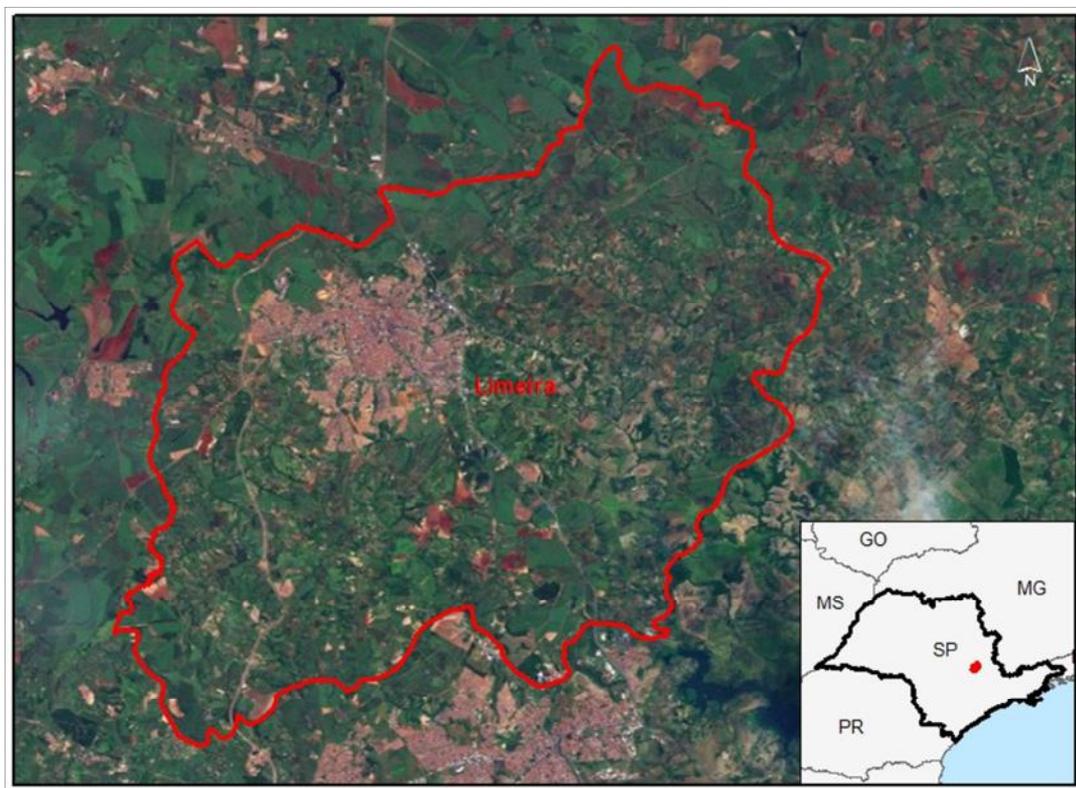


Figura 1: Localização da área de estudo.
Fonte: Malha Municipal Digital 2010, IBGE; ESRI Basemap, s/d.

O compromisso com a confidencialidade dos dados estatísticos, a necessidade de trabalhar com uma resolução espacial melhor do que a dos setores censitários, a adequação simultânea a estudos no nível nacional e a estudos locais e a comparabilidade internacional levaram a optar por utilizar células com dimensões de 1 x 1 km nas áreas rurais e 250 x 250 m nas áreas urbanas.

A abordagem de agregação foi realizada com base nos microdados do Censo Demográfico 2010 e do Cadastro Nacional de Endereços para Fins Estatísticos - CNEFE e é utilizada como referência para a avaliação dos demais métodos. Nas áreas rurais foram utilizadas diretamente as coordenadas das edificações coletadas pela pesquisa, enquanto nas áreas urbanas os endereços das unidades domiciliares visitadas passaram por um processo de geocodificação para serem transformados em pontos. Foi observado uma perda na localização espacial da ordem de 1,5% dos domicílios visitados pelo Censo 2010 no município de Limeira (N = 84.531) devido à falta de informações que permitisse a geocodificação nas áreas urbanas e à falta de coordenadas coletadas nas zonas rurais.

Todos os métodos de desagregação utilizam os dados do Censo Demográfico 2010 por setor censitário (IBGE, 2011a; 2011b) e o primeiro deles é a ponderação zonal. Este método é o mais simples entre os que se propõem a realocar a população entre duas unidades geográficas diferentes (PLUMEJEAUD et al., 2010). Neste caso, a unidade de origem é o setor censitário e a unidade de destino é a grade de células. O método pressupõe que a distribuição da população é homogênea na unidade geográfica de origem e de destino, de modo que a população da unidade de destino possa ser calculada com base na densidade populacional encontrada na unidade de origem.

O segundo método de desagregação é um método dasimétrico binário (LANGFORD; UNWIN, 1994) com utilização de uma classificação de uso das terras feita com base em imagens do satélite CBERS (pixel de 30 metros) do ano de 2010. A classe “superfície impermeável” foi considerada como sendo povoada enquanto as demais classes, relativas à vegetação e água, foram consideradas não povoadas.

O terceiro método de desagregação é também um método dasimétrico que utiliza o mesmo dado auxiliar do método anterior, sendo que a classe povoada foi reclassificada em alta densidade residencial, baixa densidade residencial e área não urbana habitada. A distribuição de população foi realizada com o auxílio de uma extensão para o programa ArcMap, denominada *Dasymetric Mapping Extension* – DME (SLEETER; GOULD, 2008).

O quarto método de desagregação também é um método dasimétrico, agora utilizando uma malha viária como dado auxiliar. Nas áreas urbanas foi utilizada a malha de ruas do mapeamento censitário do IBGE, com escala aproximada de 1:5.000, enquanto nas áreas rurais foi utilizada a malha viária do IGC, com escala de 1:50.000. Algumas edições foram necessárias nestas camadas vetoriais para eliminar feições que poderiam prejudicar o cálculo da densidade populacional por extensão de via, como canteiros centrais de avenidas e estradas, retornos e outras feições similares.

Foram selecionadas três metodologias para avaliação dos resultados. A primeira delas é uma técnica importada da área de Sensoriamento Remoto que faz uso de uma matriz

com os erros e acertos da classificação e tem por objetivo avaliar a acurácia com que o modelo de realocação de população identifica as regiões povoadas e não povoadas. A técnica é baseada em uma matriz contendo as quantidades de células classificadas de forma correta e incorreta com base em amostras de campo (dados de referência) e a medida final é expressa sob a forma de uma distribuição de percentuais de células classificadas corretamente (SMITS et al, 1999). Também é apresentada outra medida derivada desta mesma matriz – o Índice *Kappa* - que considera no seu cálculo todos os elementos da matriz, o que torna esta medida mais robusta do que a anterior. Utilizou-se uma classificação da qualidade do mapeamento com base neste índice proposta por Landis e Koch (1977).

A segunda metodologia utiliza uma regressão linear ajustada à origem que foi aplicada a todos os métodos de desagregação para avaliar o seu ajuste aos dados de referência. Para realizar a comparação entre os diferentes métodos foram utilizados gráficos de dispersão, coeficiente de correlação, coeficiente de determinação (R^2) e taxa de adequação do modelo (*F ratio*). O gráfico de dispersão nos permite uma avaliação visual entre os valores estimados e os valores de referência, enquanto o coeficiente de correlação indica a força e a direção do relacionamento entre a população estimada e a de referência. O coeficiente de determinação nos dá a proporção da variância da população estimada com relação à população de referência, enquanto a taxa de adequação do modelo permite avaliar o quanto o modelo proporcionou uma melhora na estimativa da população em comparação com os níveis de incerteza do mesmo.

Finalmente, utilizou-se uma medida para avaliar o erro entre a população estimada pela desagregação e a população de referência. A fórmula utilizada para este cálculo é:

$$D = \frac{(POP_{\text{modelo}} - POP_{\text{REF}})}{(POP_{\text{modelo}} + POP_{\text{REF}})}$$

onde D é a diferença normalizada, POP_{modelo} é a população do método de desagregação empregado e POP_{REF} é a população de referência obtida com o método de agregação. Os dados de população foram agrupados em cinco classes (1 a 49, 50 a 199, 200 a 499, 500 a 999 e 1000 e mais) e a média da diferença normalizada para cada classe e para cada método de desagregação é apresentado em forma de gráfico. Esta medida foi selecionada por ser capaz de descrever a direção da diferença entre os valores de população, de modo que valores negativos e positivos ocorrem quando os valores estimados são, respectivamente, menores ou maiores do que os valores de referência, indicando uma subestimação ou uma superestimação. Valores próximos de zero para esta medida indicam que os valores do modelo são praticamente iguais aos valores de referência.

3. Resultados e Discussão

A FIG. 2 mostra sob a forma de mapas os resultados obtidos para a distribuição espacial da população utilizando os métodos avaliados e o método de referência. Os resultados são graficamente similares, com pequenas diferenças visíveis na área rural do município. Entretanto, para avaliar melhor as diferenças entre os métodos é

necessário utilizar medidas mais objetivas, conforme veremos a seguir.

A primeira avaliação está relacionada com a quantidade de células povoadas e não povoadas identificadas corretamente pelos métodos de desagregação em comparação com os dados de agregação que servem de referência para o estudo. Na TAB. 1 podemos observar os valores obtidos para a acurácia do mapeamento de cada classe (povoada e não povoada) e a acurácia geral, além do Índice *Kappa*. De acordo com os dados apresentados, podemos dizer que o método dasimétrico binário e o dasimétrico de 3 classes não apresentam diferenças significativas entre si, mostrando uma acurácia que vai de

razoável à boa. O método de ponderação zonal revelou-se o pior método entre aqueles utilizados neste estudo, obtendo uma classificação de acurácia variando entre péssima e ruim. O método dasimétrico com a utilização de malha viária como dado auxiliar alcançou a melhor classificação, ficando entre razoável e boa. Todos os métodos dasimétricos apresentaram uma capacidade geral de classificação das classes povoada e não povoada superiores a 60%, sendo que este percentual é inferior nas áreas rurais, indicando que os dados auxiliares não conseguem representar de maneira eficiente a distribuição da população em áreas com ocupação dispersa no território.

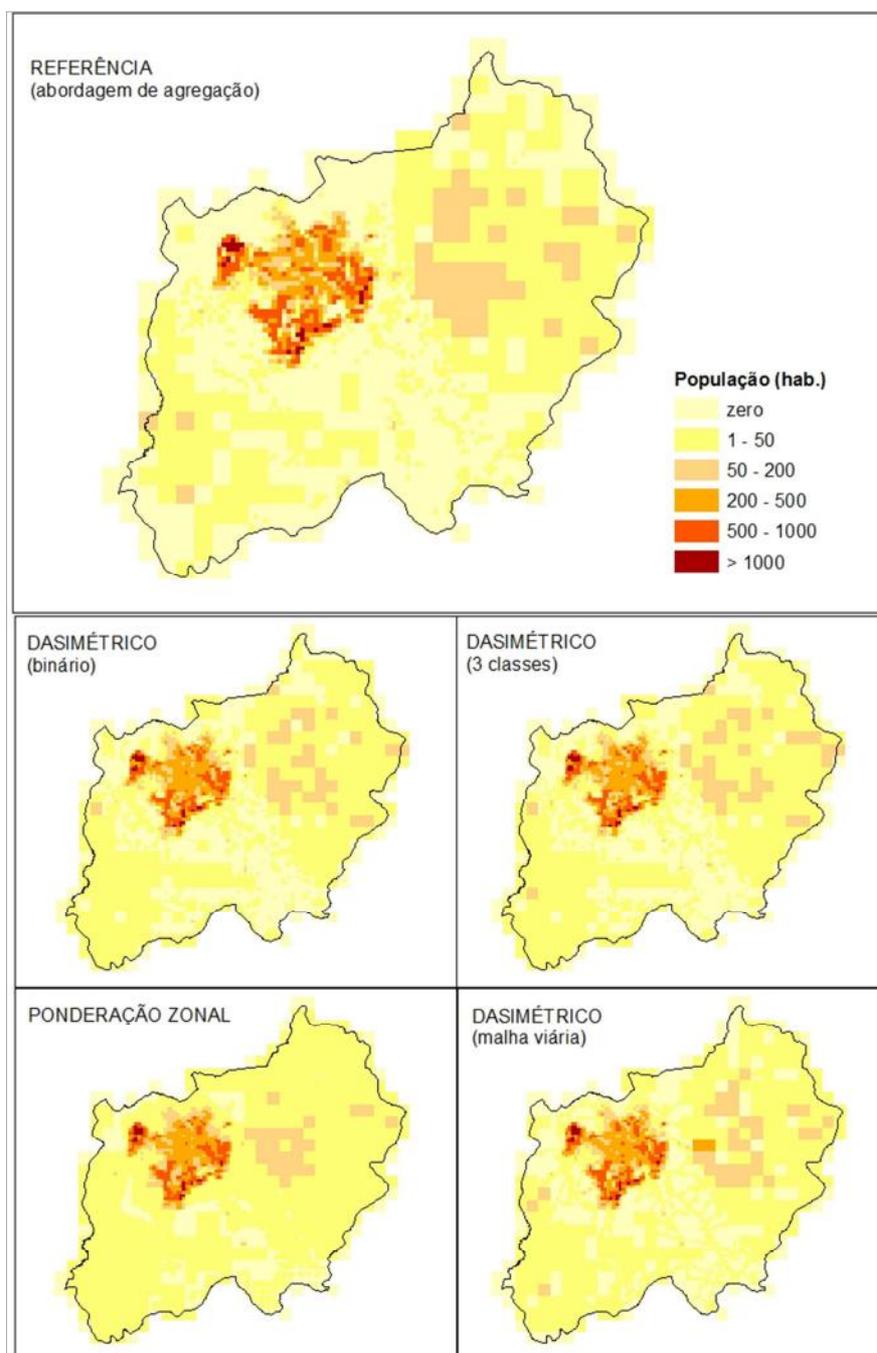


Figura 2: Mapas de distribuição da população de acordo com o método.
Fonte: Censo Demográfico 2010, IBGE.

Localização	Método de desagregação	Classe Povoada		Classe Não povoada		Geral	
		Acurácia	<i>Kappa</i>	Acurácia	<i>Kappa</i>	Acurácia	<i>Kappa</i>
Rural N = 338	Dasim. Binário	65,38	0,625	34,43	0,230	70,71	0,336 (razoável)
	Dasim. 3 classes	65,84	0,636	35,76	0,242	71,30	0,351 (razoável)
	Ponderação zonal	59,94	1,000	0,73	0,004	60,06	0,009 (ruim)
	Dasim. Malha viária	62,85	0,505	31,85	0,200	68,34	0,286 (razoável)
Urbano N = 3122	Dasim. Binário	51,25	0,644	57,83	0,320	70,79	0,423 (boa)
	Dasim. 3 classes	51,26	0,639	58,16	0,323	70,94	0,423 (boa)
	Ponderação zonal	36,85	0,791	0,61	0,002	37,09	0,004 (péssima)
	Dasim. Malha viária	52,81	0,693	58,62	0,332	71,72	0,449 (boa)
Geral N = 3460	Dasim. Binário	53,13	0,650	56,31	0,326	70,78	0,435 (boa)
	Dasim. 3 classes	53,19	0,654	56,70	0,322	70,98	0,432 (boa)
	Ponderação zonal	39,11	0,817	0,62	0,002	39,33	0,004 (péssima)
	Dasim. Malha viária	54,14	0,688	56,79	0,329	71,39	0,445 (boa)

Tabela 1: Acurácia do mapeamento de acordo com a localização e o método.
Fonte: Elaboração Própria.

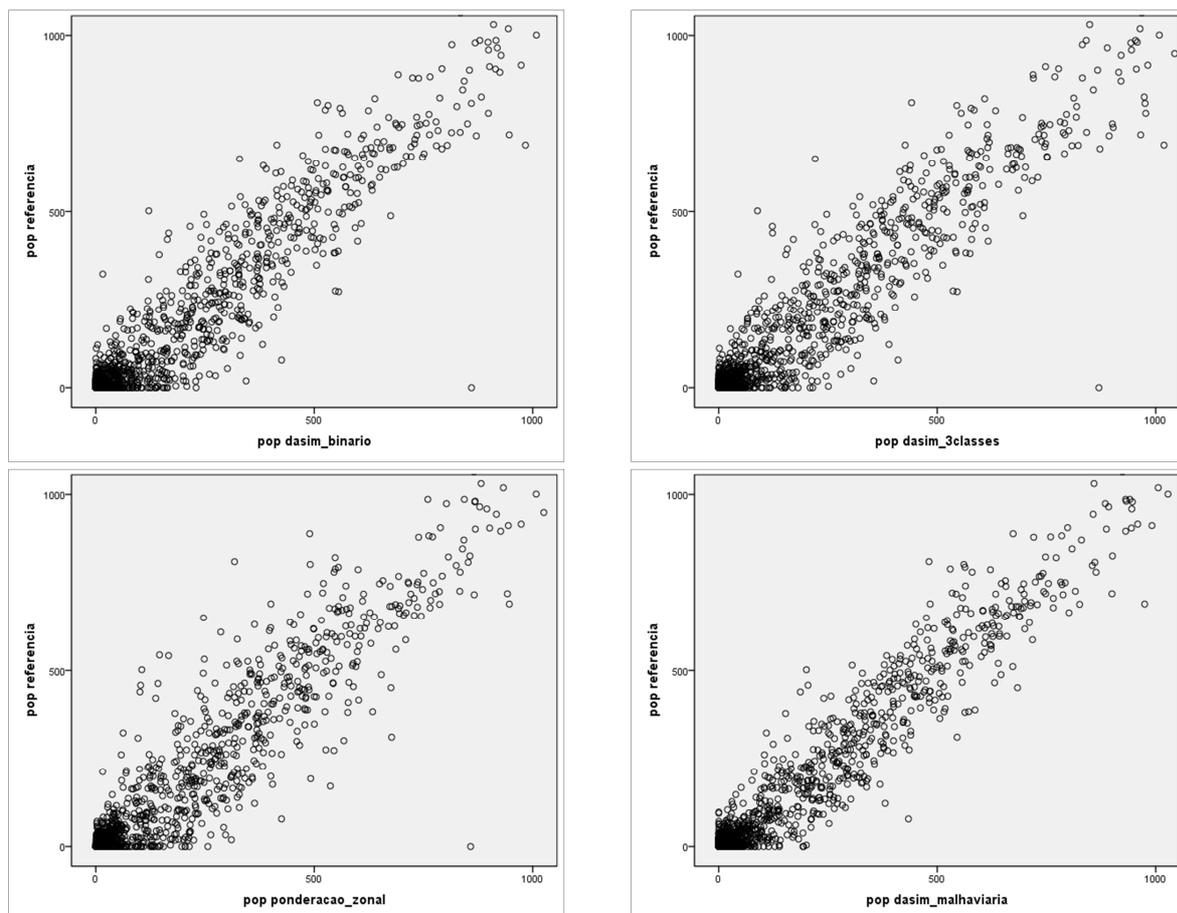


Figura 3: Gráficos de dispersão de acordo com o método.
 Fonte: Censo Demográfico 2010, IBGE.

O método de ponderação zonal não tem habilidade para identificar áreas não povoadas, uma vez que ele pressupõe uma distribuição homogênea da população na região considerada e os valores baixos, mas não nulos, para a acurácia da classe “Não Povoada” são explicados pela quantidade de casas decimais utilizadas para representar os valores de população (apenas duas). Especula-se que o valor relativamente alto da acurácia geral para áreas rurais deste método (60,06%) é devido ao número pequeno de células consideradas na análise e também às características locais, pois em estudos realizados para outras áreas isso não se confirma (BUENO; MARTIN; D’ANTONA, 2013).

A FIG. 3 mostra os gráficos de dispersão construídos com os dados obtidos da regressão linear para toda a área de estudo. Visualmente o melhor resultado parece ser o do método dasimétrico com a utilização de vias como dado auxiliar, pois os pontos se apresentam menos dispersos em relação a uma linha de tendência. Além disso, o gráfico desse modelo dasimétrico exibe uma menor quantidade de pontos ao longo do eixo X, que representam valores estimados de população nulos para valores reais de população não nulos. Com relação à quantidade de pontos

ao longo do eixo Y, que representam valores estimados de população não nulos para valores reais nulos, aparentemente a quantidade é semelhante em todos os quatro gráficos.

A TAB. 2 apresenta os principais coeficientes obtidos com uma regressão linear ajustada à origem aplicada aos quatro modelos de desagregação da população. O coeficiente de correlação indica a existência de uma correlação forte entre os dados analisados para a área urbana e média para a área rural, mas prevalecendo uma correlação forte quando analisamos o município integralmente. Esta conclusão se repete para o coeficiente de determinação, onde para a área urbana os modelos aplicados conseguem explicar mais de 90% da variância dos valores de população; para a área rural temos valores mais baixos, variando entre 46 e 71%. No caso da análise para toda a área de estudo, os valores permanecem altos, acima de 90%. Com relação ao *F ratio*, o método dasimétrico com malha viária se destaca em relação aos demais, indicando um melhor ajuste do modelo na área urbana e na área total. De uma maneira geral, podemos dizer que todos os métodos apresentam resultados satisfatórios para as áreas urbanas, mas não tão bons para as áreas rurais.

Localização	Modelo	Coef. correlação	Coef. determinação	Fratio **
Rural N = 338	Dasim. binário	0,539	0,463	290,553
	Dasim. 3 classes	0,546	0,468	296,370
	Ponderação zonal	0,767	0,712	833,010
	Dasim. Malha viária	0,656	0,579	463,869
Urbano N = 3 122	Dasim. binário	0,952	0,918	35 024,880
	Dasim. 3 classes	0,950	0,915	33 589,578
	Ponderação zonal	0,945	0,906	30 238,245
	Dasim. Malha viária	0,966	0,941	49 803,529
Geral N = 3 460	Dasim. binário	0,952	0,917	38 157,222
	Dasim. 3 classes	0,950	0,914	36 638,838
	Ponderação zonal	0,945	0,906	33 334,743
	Dasim. Malha viária	0,965	0,940	54 154,504

Tabela 2: Sumário da regressão linear* de acordo com a localização e o método.
* Regressão linear ajustada à origem. ** Valores estatisticamente significantes (p value = 0).

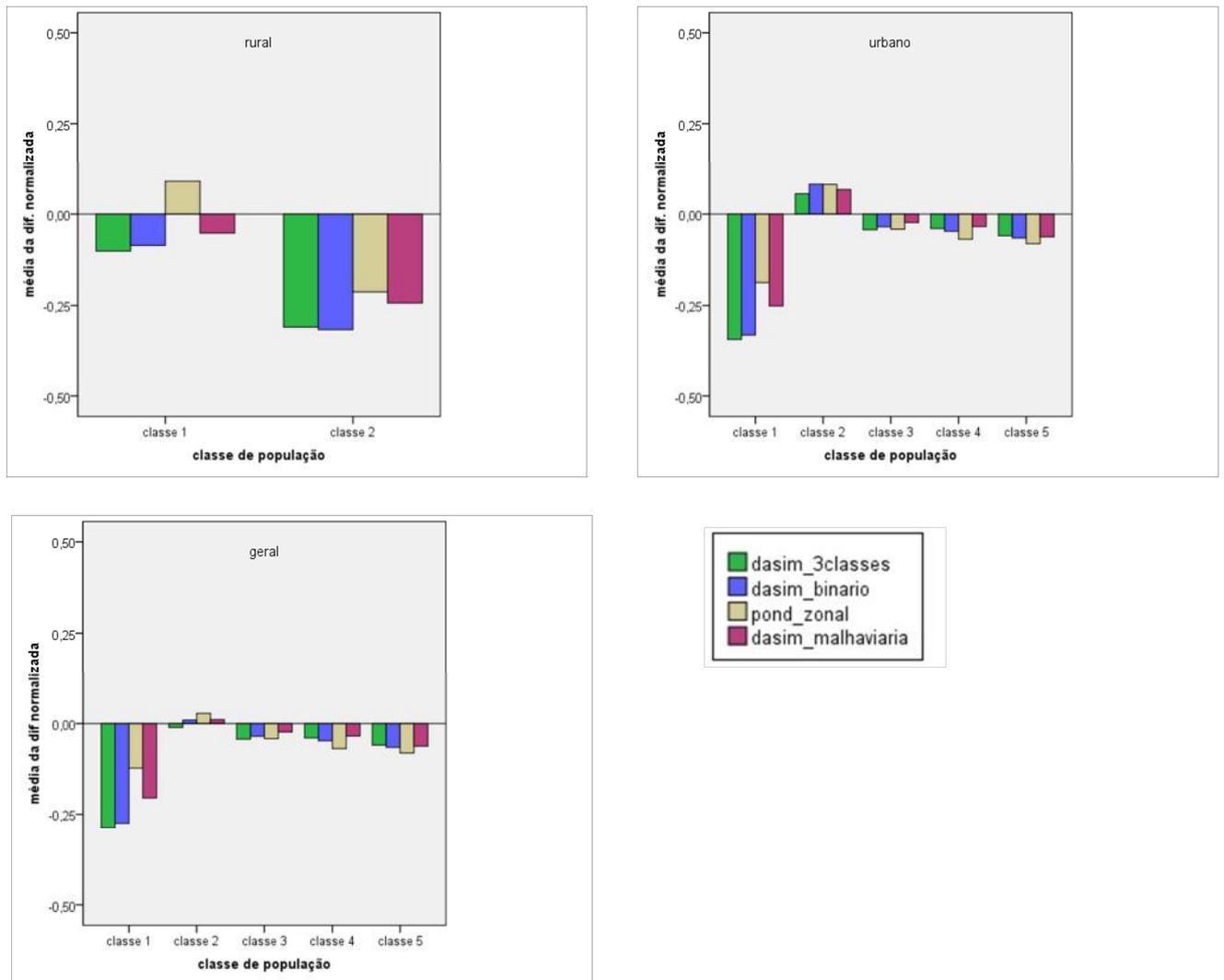


Figura 4: Média da Diferença normalizada por classe de população de acordo com a localização e o método.

Pode-se observar que foram obtidos valores comparativamente altos para todos os coeficientes com o método de ponderação zonal em área rural. Como já mencionado anteriormente, acredita-se que isso se deva a fatores locais, pois este método não costuma apresentar resultados muito bons em áreas rurais, apesar de apresentar resultados satisfatórios para áreas urbanas, principalmente as mais densas.

Observando a FIG. 4 pode-se dizer que, de uma maneira geral, os modelos de desagregação subestimam os valores de população. Tem-se como exceção a classe 2 de população (50 – 200) na área urbana. As maiores diferenças médias surgem para os menores valores de população, ou seja, para aqueles que pertencem à classe 1 (1 – 50 habitantes). A classe 2, com valores entre 200 e 500 habitantes, apresenta diferenças negativas na área rural e positivas na área urbana, o que nos leva a uma diferença média geral próxima de zero para todos os modelos. Para as classes com maior quantidade de população o método de ponderação zonal apresenta as maiores diferenças; em seguida, estão os modelos dasimétricos com utilização de imagens classificadas com valores bem próximos entre si; por último, está o modelo dasimétrico com vias.

Comparando os resultados obtidos para as diferenças médias por classe de população com aqueles obtidos para

outras áreas no Brasil (BUENO; MARTIN; D'ANTONA, 2013), pode-se observar que não há um padrão visível de subestimação ou superestimação de acordo com as classes de população, no entanto, são necessários estudos mais aprofundados para avaliar mais profundamente esta questão.

4. Conclusões

A avaliação entre os diferentes métodos de desagregação realizada neste estudo sugere que não existe um método melhor do que o outro, pois esta decisão depende do objetivo da análise, da disponibilidade de informações para auxiliar a distribuição da população e das características de ocupação da área de estudo. Assim, nos casos em que o objetivo principal da grade é determinar a presença ou ausência de população em detrimento da quantidade, os três métodos dasimétricos apresentados levam a um resultado satisfatório. Neste caso, o método da ponderação zonal seria descartado por não apresentar bons resultados. Quando além da presença ou ausência de população, a determinação da quantidade de população for também importante, os resultados apontam o método dasimétrico com malha viária como informação auxiliar como sendo o mais adequado, no entanto, os demais métodos também apresentaram bons resultados.

Analisando sob o ponto de vista da disponibilidade de informações auxiliares como suporte aos modelos de distribuição espacial da população, a primeira opção recai sobre dados detalhados relativos ao sistema viário, tanto na área urbana como na área rural. No caso da não existência desses dados, imagens de satélite classificadas quanto ao uso das terras são uma boa opção, não havendo aparentemente diferenças na utilização de uma classificação binária – classe povoada e classe não povoada – ou de uma classificação mais detalhada, considerando a densidade de população – três classes povoadas e uma classe não povoada. Na impossibilidade de aquisição de qualquer dado auxiliar, o método da ponderação zonal pode ser utilizado, mas com conhecimento das suas limitações e das incertezas envolvidas.

A localização da área de estudo também pode orientar na escolha do método para desagregação da população. De uma maneira geral, todos os métodos apresentam resultados melhores para áreas urbanas do que para áreas rurais, indicando que é necessário o desenvolvimento de metodologias específicas para estas áreas ou a utilização de dados auxiliares mais representativos da ocupação rural.

Como conclusão geral, pode-se dizer que os métodos de desagregação representam uma solução viável quando não há a disponibilidade de dados obtidos por meio de agregação direta dos dados coletados, devendo-se, no entanto, ser consideradas nas análises as incertezas proporcionadas pela metodologia adotada.

Agradecimentos

Este estudo foi apoiado pelo Instituto Brasileiro de Geografia e Estatística - IBGE e pela Coordenação de Aperfeiçoamento de Nível Superior – CAPES (processo 17235-12-0). Os microdados e informações cartográficas detalhadas do Censo Demográfico 2010 foram cedidos pelo IBGE em caráter excepcional.

REFERÊNCIAS

- [1] BRACKEN, I.; MARTIN, . The Generation of Spatial Population Distributions from Census Centroid Data Source. **Environment and Planning A**, v. 21, n. 4, p. 537-543, 1989.
- [2] BUENO, M. C. D.; MARTIN, D.; D'ANTONA, A. O. The Brazilian Population Grid: a hybrid approach. In: **European Forum for Geostatistics – SOFIA CONFERENCE**, 2013, Bulgária. Disponível em <http://www.nsi.bg/efgs2013/data/uploads/presentations/DAY2_WS1_5_Paper_BUENO_ok.pdf>. Acesso em outubro de 2013.
- [3] CLARKE, J. I.; RHIND, D. W. **Human Dimensions of Global Environmental Change**. International Social Science Council e UNESCO: 1992.
- [4] DEICHMANN, U. **A Review of Spatial Population Database Design and Modelling**. National Center for Geographic Information and Analysis (NCGIA). 1996.
- [5] EICHER, C. L.; BREWER, C. A. Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation. **Cartography and Geographic Information Science**, v. 28, n. 2, p.125-138, 2001.
- [6] EFGS – European Forum for Geostatistics. **GEOSTAT 1A - Representing Census data in a European population grid**, Final Report: 2012.
- [7] EXETER, D. J.; BOYLE, P.; FENG, Z.; FLOWERDEW, R.; SCHIERLOH, N.. The creation of Consistent Areas Through Time (CATTs) in Scotland, 1981-2001. **Population Trends**, v. 119, n.1, 2005.
- [8] HOGAN, D. J. Crescimento populacional, padrões de assentamento e o ambiente físico. In: ANPOCS (ed.). **Ciências sociais hoje**. Rio de Janeiro: ANPOCS, 1992.
- [9] IBGE – Instituto Brasileiro de Geografia e Estatística. **Censo 2010 – Síntese das Etapas da Pesquisa**, 2010. Disponível em <http://censo2010.ibge.gov.br/images/pdf/censo2010/sintese/sintese_censo2010_portugues.pdf> . Acesso em março de 2011.
- [10] _____. **Censo Demográfico 2010 - Malha de Setores Censitários**, 2011a. Disponível em <ftp://geoftp.ibge.gov.br/malhas_digitais/censo_2010/>. Acesso em outubro de 2011.
- [11] _____. **Base de informações do Censo Demográfico 2010: Resultados do Universo por Setor Censitário**, 2011b. Disponível em <ftp://ftp.ibge.gov.br/Censos/Censo_Demografico_2010/Resultados_do_Universo/Agregados_por_Setores_Censitarios/>. Acesso em maio de 2012.
- [12] LANDIS, J.R.; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, v. 33, n. 1, p.159–174, 1977.
- [13] LANGFORD, M.; UNWIN, D. J. Generating and mapping population density surfaces within a geographical information system. **The Cartographic Journal**, v. 31, n. 1, p.21-26,1994.
- [14] MARTIN, D. Census 2001: making the best of zonal geographies. In: **The census of population: 2000 AND BEYOND**, Reino Unido: 2000.
- [15] MARTINE, G.; SCHENSUL, D. (eds.). **The Demography of Adaptation to Climate Change**. New York, London and Mexico City: UNFPA, IIED e El Colegio de México, 2013.
- [16] MENNIS, J. Generating Surface Models of Population Using Dasymetric Mapping. **Professional Geographer**, v. 55, n. 1, p. 31-42, 2003.
- [17] NAÇÕES UNIDAS, United Nations Economic Commission for Europe. **Register-based statistics in the Nordic countries - Review of best practices with focus on population and social statistics**. New York and Geneva: United Nations, 2007. Disponível em <<http://www.unece.org/fileadmin/DAM/stats/pu>>

blications/Register_based_statistics_in_Nordic_countries.pdf>. Acesso em julho de 2013.

- [18] NORMAN, P.; REES, P.; BOYLE, P. Achieving Data Compatibility over Space and Time: Creating Consistent Geographical Zones. **International Journal of Population Geography**, v. 9, n.1,p. 365-386, 2003.
- [19] OJIMA, R.; MARTINE, G. Resgates sobre População e Ambiente: breve análise da Dinâmica Demográfica e a Urbanização nos Biomas Brasileiros. **Ideias**, n. 5, p. 55-70, 2012.
- [20] PLUMEJEAUD, C.; PRUD'HOMME, J.; DAVOINE, P.-A.; GENSEL, J. Transferring Indicators into Different Partitions of Geographic Space. In: TANIAR, D.; GERVASI, O.; MURGANTE, B.; PARDEDE,E.; APDUHAN, B. O. (eds.) **ICCSA'10 Proceedings of the 2010 international conference on Computational Science and Its Applications - Volume Part I**. Berlin, Heidelberg: Springer-Verlag, 2010.
- [21] REIBEL, M.; BUFALINO, M. E. A test of street weighted areal interpolation using geographic information systems. **Environment and Planning A**, v. 37, n.1,p. 127–139, 2005.
- [22] RUSANEM, J.; MUILU, T.; COLPAERT, A.; NAUKKARINEN, A. Finnish socio-economic grid data, GIS and the hidden geography of unemployment. **Tijdschrift voor Economische en Sociale Geographie**, v. 92, n. 2, p. 139-147, 2001.
- [23] SMITS, P. C.; DELLEPIANE, S. G.; SCHOWENGERT, R. A. Quality assessment of image classification algorithms for land-cover mapping: a review and a proposal for a cost based approach. **International Journal of Remote Sensing**, v. 20, n.8, p. 1461-1486, 1999.
- [24] SLEETER, R.; GOULD, M. **Geographic Information System Software to Remodel Population Data Using Dasymetric Mapping Methods**. Techniques and Methods 11–C2. U.S. Department of the Interior e U.S. Geological Survey, 2008.
- [25] STEINNOCHER K., KAMINGER I., KOSTL M., WEICHSELBAUM J.. Gridded Population – new data sets for an improved disaggregation approach. In: **European Forum for Geostatistics Workshop**, 2010, Estonia. Disponível em <<http://www.efgs.info/workshops/efgs-2010-tallinn-estonia>> . Acesso em janeiro de 2011.
- [26] TAMMILEHTO-LUODE, M., BACKER, L.; ROGSTAT, L. Grid data and area delimitation by definition. Towards a better European territorial statistical system. In: **Conference of European Statisticians**, Suíça, 2000.
- [27] TAMMILEHTO-LUODE, M. Opportunities and challenges of grid-based statistics. In: **World Statistics Congress of the International Statistical Institute**, 2011, Irlanda.
- [28] TOBLER, W. R. Smooth pycnophylactic interpolation for geographical regions. **Journal of the American Statistical Association**, v. 74,n.1, p. 519-530, 1979.
- [29] XIE, Y. The overlaid network algorithms for areal interpolation problem. **Computer, Environment and Urban Systems**, v. 19, n. 4, p. 287-306, 1995.
- [30] ZANDBERGEN, P. A. Dasymetric Mapping Using High Resolution Address Point Datasets. **Transactions in GIS**, v. 15 (s1), p. 5–27, 2011.

Evaluation of disaggregation methods to generate population grids

Maria do Carmo Dias Bueno¹
Álvaro de Oliveira D'Antona²

¹ Civil Engineer, Master in Geomatics UERJ, Rio de Janeiro (RJ), PhD candidate in Demography at the Institute of Philosophy and Human Sciences - UNICAMP, Campinas (SP), technologist of the Brazilian Institute of Geography and Statistics.

² Bachelor of Economics, Master in Anthropology, Doctor of Social Sciences (Population Studies), a research collaborator at the Center for Population Studies (NEPO-Unicamp), professor in the School of Applied Sciences (FCA-Unicamp) and the Program postgraduate Demography (IFCH-Unicamp).

Abstract The purpose of this work is to evaluate disaggregation methods used to generate a population grid, using as reference a population grid built with the aggregation of census micro-data, which allows more accurate results about the performance of these methods. Three methods were selected to conduct the evaluation. The first uses an array of errors and successes in identifying populated and non-populated areas, allowing an evaluation of the accuracy of the spatial distribution. The second method uses coefficients reported by a linear regression to evaluate the data fit. Finally, it is used a formula to calculate the difference between the population values, showing whether there is underestimation or overestimation. The results suggest that the choice of the most appropriate method depends on the purpose of the study, the quality and availability of the ancillary data as well as the features of the interest area.

Keyword: population distribution; population grid; disaggregation; Aggregation.

Informações sobre os autores

Maria do Carmo Dias Bueno (IBGE)

Endereço para correspondência: Centro de Documentação e Disseminação de Informações, Rua General Canabarro, 706 – sala 206 – Maracanã – CEP 20271-205 – Rio de Janeiro, RJ.

E-mail: bueno.mariadocarmo@gmail.com.

Link para o currículo lattes: <http://lattes.cnpq.br/7979542180039260>.

Álvaro de Oliveira D'Antona (FCA/Unicamp)

Endereço para correspondência: Rua Pedro Zaccharia, 1300, Cidade Universitária – CEP 13484-350 - Limeira, SP .

E-mail: alvaro.dantona@fca.unicamp.br.

Link para o currículo lattes: <http://lattes.cnpq.br/1771971577733548>.

Artigo Recebido em: 14-03-2014

Artigo Aprovado em: 25-05-2014