

GRADE ESTATÍSTICA

Presidente da República
Dilma Rousseff

Ministro do Planejamento, Orçamento e Gestão
Valdir Moysés Simão

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE

Presidente
Wasmália Socorro Barata Bivar

Diretor-Executivo
Fernando José de Araújo Abrantes

ÓRGÃOS ESPECÍFICOS SINGULARES

Diretoria de Pesquisas
Roberto Luís Olinto Ramos

Diretoria de Geociências
Wadih João Scandar Neto

Diretoria de Informática
Paulo César Moraes Simões

Centro de Documentação e Disseminação de Informações
David Wu Tai

Escola Nacional de Ciências Estatísticas
Maysa Sacramento de Magalhães

UNIDADE RESPONSÁVEL

Centro de Documentação e Disseminação de Informações

Coordenação de Projetos Especiais
Maria do Carmo Dias Bueno

Ministério do Planejamento, Orçamento e Gestão
Instituto Brasileiro de Geografia e Estatística - IBGE
Centro de Documentação e Disseminação de Informações
Coordenação de Projetos Especiais

Grade Estatística

Rio de Janeiro
2016

Apresentação

As grades estatísticas se constituem em uma forma de disseminação de dados que permite análises detalhadas e independentes das divisões territoriais, visando atender, principalmente, a necessidade de se ter dados em unidades geográficas pequenas e estáveis ao longo do tempo, facilitando sobremaneira a comparação nacional e internacional e fornecendo um aumento significativo do detalhamento, particularmente nas regiões rurais, em comparação com metodologias anteriores.

Com a presente publicação, o Instituto Brasileiro de Geografia e Estatística - IBGE disponibiliza ao público a metodologia empregada no desenvolvimento de sua Grade Estatística para disseminação dos dados do Censo Demográfico 2010. Este relatório contém uma descrição detalhada dos aspectos que foram estudados para tornar possível a aplicação desta forma de disseminação à realidade brasileira, forma esta que já é utilizada em diversos países ao redor do mundo.

Cabe destacar que, no Brasil, a geração da grade estatística se tornou possível graças aos avanços tecnológicos adotados inicialmente nos Censos 2007 e continuados no Censo 2010, como a utilização de dispositivos de coleta eletrônicos equipados com tecnologia que permite a captura de coordenadas geográficas e a criação de um cadastro de endereços conectado com o mapeamento viário.

Com a publicação de dados através de grades estatísticas o IBGE reafirma seu propósito de disponibilizar informações em diferentes formatos e níveis geográficos, ampliando o atendimento aos seus usuários e expandindo o potencial analítico das informações produzidas pelo instituto.

David Wu Tai

Coordenador-Geral do Centro de Documentação e Disseminação de Informações

SUMÁRIO

Introdução	6
Definição e história	6
Vantagens	7
Estabilidade espaço-temporal	7
Adaptação a recortes espaciais	8
Hierarquia e flexibilidade	8
Versatilidade	8
Desafios	9
Abordagens	9
Agregação	10
Desagregação	11
Interpolação	13
Dados	14
Dados estatísticos	14
Dados vetoriais	14
Definição geométrica e espacial da grade	15
Metodologia	16

Introdução

A proposta de criação de um sistema de grades regulares para disseminação de dados estatísticos surge a partir da necessidade de integração de dados de origens diversas e agregados em unidades geográficas incompatíveis, além da necessidade de se ter dados agregados em unidades pequenas e sem variação ao longo do tempo. Todas essas necessidades se tornam mais evidentes a partir da utilização de ferramentas de geoprocessamento nas análises socioeconômicas e demográficas, devido principalmente à facilidade de se criar unidades artificiais de análise nesses ambientes (RUSANEM *et al.*, 2001). Este sistema não é recente, apesar de ainda ser relativamente pouco utilizado e explorado, mas a sua produção tem se tornado cada vez mais comum, tendo em vista as tecnologias espaciais atualmente disponíveis e empregadas na realização dos censos, como utilização de GPS, geoprocessamento e sensoriamento remoto (NAÇÕES UNIDAS, 2000; 2009).

Definição e história

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

Em 1969, o Japão utilizou pela primeira vez este sistema para divulgar estatísticas derivadas de diversas pesquisas para a área metropolitana de Tóquio e, desde então, tanto o governo quanto empresas privadas, realizam a coleta e a manutenção de dados regionais neste sistema, agora abrangendo todo o país.

Na Europa, ao longo da década de 70, outros países, principalmente do norte europeu, também adotaram esse sistema, como a Finlândia, que disponibiliza dados censitários em grade desde 1970, e a Inglaterra, que gerou uma grade com os dados do censo de 1971, mas não manteve esta produção para os censos seguintes.

Vantagens

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

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

Estabilidade espaço-temporal

Uma característica do sistema de grades é a sua independência de recortes político-administrativos que, como consequência, traz vantagens relacionadas com a estabilidade espaço-temporal dos dados. Esta estabilidade dos limites das unidades ao longo do tempo é vista como uma das principais vantagens deste sistema (MARTIN, 2000; TAMMILEHTO-LUODE *et al.*, 2000; RUSANEM *et al.*, 2001; TAMMILEHTO-LUODE, 2011), coisa que não acontece com unidades geográficas diretamente ligadas com a administração do território (municípios, distritos e outras), nem com unidades operacionais ou funcionais (setores censitários). No primeiro caso, isso acontece porque essas unidades estão sujeitas a alterações nos seus limites físicos, que pode ser devido à criação de novas unidades, que surgem à medida que as regiões vão se tornando mais desenvolvidas e, conseqüentemente, almejam uma maior autonomia no gerenciamento de seu território, ou mesmo pela redefinição dos limites em decorrência de questionamentos legais ou imprecisões conhecidas e que necessitam de correção. No segundo caso, a causa das alterações é inteiramente operacional, uma vez que os limites dos setores censitários é definido de acordo

com o quantitativo de população existente em uma determinada área. Assim, à medida que essas áreas vão crescendo em número de moradores, é necessário que haja uma revisão nos limites para que os parâmetros que regulam a coleta de dados nessas unidades sejam mantidos.

Adaptação a recortes espaciais

Outra vantagem das grades em relação às unidades geográficas tradicionais é a sua forma simples, regular e de pequenas dimensões, o que permite que cada célula funcione como um “tijolo” que pode ser agrupado de diversas formas para se adaptar a diferentes recortes geográficos (TAMMILEHTO-LUODE, 2011). A necessidade de ter dados socioeconômicos e demográficos em unidades espaciais que não sejam as tradicionais unidades político-administrativas é cada vez maior, tendo em vista a crescente necessidade de realização de avaliações e pesquisas envolvendo áreas relacionadas a algum aspecto físico, ambiental e/ou social.

Hierarquia e flexibilidade

Os sistemas de grade com base em células regulares oferecem uma estrutura hierárquica, de forma que a união de um conjunto de células adjacentes leva à formação de uma célula maior, pertencente a um nível hierárquico superior. Esta estrutura hierárquica confere flexibilidade ao sistema, permitindo que diversos níveis hierárquicos possam conviver simultaneamente.

No caso de uma grade estatística, é usual que se tenham células com dimensões diferentes cobrindo áreas urbanas e áreas rurais, com o objetivo de melhor representar a distribuição da população.

Versatilidade

Um sistema de grade pode ser construído utilizando tanto uma estrutura de dados vetorial¹ quanto matricial² e também pode ser facilmente convertido de uma estrutura para outra. A vantagem de se ter a grade em uma estrutura vetorial é que todas as variáveis existentes podem ser alocadas na tabela de atributos e

¹ As estruturas vetoriais representam as entidades geográficas através de três formas básicas: pontos, linhas e áreas (ou polígonos), definidas por suas coordenadas cartesianas (CÂMARA, 2005).

² As estruturas matriciais utilizam uma grade regular sobre a qual se representam os elementos geográficos (CÂMARA, 2005).

relacionadas com a geometria, facilitando as análises conjuntas dessas variáveis (STRAND; BLOCH, 2009). No caso da estrutura matricial, é necessária a geração de um arquivo para cada variável, mas esta estrutura é a geralmente utilizada na execução de modelos de simulação.

Desafios

O sistema de grade apresenta diversas vantagens, mas, também apresenta alguns desafios que necessitam ser enfrentados, tendo em vista a sua crescente produção e utilização.

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

Embora não exista uma recomendação sobre a técnica que deve ser utilizada para a manutenção do sigilo estatístico, qualquer que seja a unidade geográfica utilizada, a supressão de dados é a mais comumente utilizada. Não existe uma norma ou padrão para estabelecer os valores limite para realizar esta supressão, embora a utilização dos valores três e dez indivíduos esteja presente em diversos casos de regras de supressão utilizadas em países europeus para grades estatísticas (KAMINGER, 2007; EFGS, 2012).

Abordagens

A geração de um sistema de grades com objetivos estatísticos consiste em uma operação de transferência de suporte espacial de dados, onde os dados básicos existentes em um suporte (por exemplo, pontos ou polígonos) são transferidos para outro suporte (por exemplo, polígonos ou superfícies). As técnicas para esta transferência de suporte podem ser classificadas de acordo com

três diferentes abordagens (PLUMEJEAUD *et al.*, 2010): agregação, desagregação e interpolação.

Agregação

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

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

A principal característica desta solução é a simplicidade, mas, em contrapartida, ela pressupõe a existência de dados georreferenciados ou, pelo menos, passíveis de serem georreferenciados. Com a utilização crescente de tecnologias nos censos, a adoção desta solução para a divulgação de dados censitários torna-se viável para um número cada vez maior de países. Segundo dados das Nações Unidas (2013), a rodada dos Censos 2010 mostrou um uso sem precedentes de tecnologias em todas as etapas da operação censitária.

Outra característica importante desta abordagem está relacionada com quem pode executá-la. A geração da grade por agregação tem sua origem nos dados coletados pelos censos e, como regra geral, apenas os produtores da pesquisa podem acessar e manipular esses dados. Assim, no caso dos censos,

apenas os institutos nacionais de estatística estão aptos a gerar uma grade estatística.

Desagregação

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

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

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

(REIBEL; AGRAWAL, 2007), embora diversas outras informações também sejam utilizadas.

O método dasimétrico mais simples, dentre os que utilizam dados de sensoriamento remoto, é o binário (LANGFORD *et al.*, 1991), que utiliza apenas duas classes de uso das terras: habitada e inabitada. O total de população na unidade de origem (geralmente o setor censitário) é, então, redistribuído apenas na classe habitada, através da densidade populacional dasimetricamente derivada. A vantagem desta modalidade é a sua simplicidade, tendo em vista que a classificação de uso das terras envolve apenas duas classes, muito embora esta divisão seja subjetiva e dependa de conhecimentos da região mapeada (EICHER; BREWER, 2001); a desvantagem é não considerar a existência de variações na densidade populacional nas áreas povoadas (MAANTAY; MAROKO; GERRMANN, 2007; FRANÇA, 2012).

Outros tipos de dados podem ser utilizados na execução do método dasimétrico. Informações também derivadas de imagens de Sensoriamento Remoto, como luzes noturnas e superfícies impermeáveis do solo, também são utilizadas como apoio ao método dasimétrico. As imagens de luzes noturnas são indicativas da presença humana e são valiosas para projetos de mapeamento da distribuição da população em escala global (SUTTON, 1997; TURNER; OPENSHAW, 2001). As superfícies impermeáveis têm uma grande correlação com a densidade populacional e têm a vantagem de não necessitarem de uma calibração complexa para a sua classificação a partir de imagens de sensores remotos. A utilização desta informação tem como base a hipótese de que a densidade de edificações ou de população é proporcional ao grau de impermeabilidade. No entanto, como isso não é sempre verdadeiro, será necessário efetuar algumas edições na camada de superfícies impermeáveis para excluir algumas áreas que não são residenciais, como áreas comerciais e industriais, além da rede viária e ferroviária (STEINNOCHER *et al.*, 2010).

Outros tipos de informação de origens as mais variadas possíveis também podem ser utilizados para auxiliar a redistribuição da população através de métodos dasimétricos, como, por exemplo, redes de transporte, endereços e mapeamento das edificações, entre outros.

Uma avaliação geral da acurácia do método dasimétrico, principalmente com a utilização de uso das terras como informação auxiliar, foi feita por Zandbergen e Ignizio (2010), que apresentaram três fatores que influenciam os erros envolvidos na metodologia. O primeiro fator é o tamanho das unidades de origem e de destino, uma vez que os erros são menores quando o método é aplicado em unidades de origem de pequena extensão e unidades de destino de grande extensão. O segundo fator é o tipo de organização espacial existente nas unidades de origem e destino: quanto mais essas unidades tiverem uma organização semelhante, menor será o erro. E, finalmente, a qualidade dos dados auxiliares, estando aqui incluída a resolução espacial, que deve ser maior do que a dos dados populacionais (TATEM *et al.*, 2007) e a classificação do uso das terras, que deve ser o mais detalhado possível.

Interpolação

Podemos classificar os métodos de interpolação entre aqueles que são baseados em pontos e aqueles que são baseados em áreas. Na primeira classe, podemos citar os métodos de interpolação por polinômios, os que utilizam distância, os geoestatísticos (*kriging*) e os modelos de tendência utilizando diferentes funções (WU *et al.*, 2005). Todos esses métodos assumem que os centroides das áreas de enumeração são representativos da distribuição da população naquela área e, portanto, podem ser utilizados para a redistribuição da população. Os problemas relacionados com estes métodos dizem respeito à escolha do centroide para representar a unidade areal, já que caso esta área seja irregular e não simétrica, o mesmo pode estar localizado em uma posição que afeta os resultados, como por exemplo, fora do polígono (LAM, 1983; LIU, 2003). Outros problemas com estes métodos apontados por Martin (1996) estão relacionados com a não preservação do valor total da população em cada zona de origem e com a não consideração de regiões desabitadas.

Na segunda classe de métodos de interpolação o método mais conhecido é o picnofilático proposto por Tobler (1979), que utiliza a população em unidades de área como base para a criação de uma superfície suavizada em grade. Este método restringe a movimentação de população entre as unidades areais, garantindo que a população permaneça constante. O método picnofilático é mais apropriado para os casos em que não se tem nenhuma informação auxiliar disponível e quando as

unidades espaciais de origem são relativamente homogêneas (DEICHMANN, 1996). Também podem ser incluídos nesta classe os métodos propostos por Kyriakidis, nos quais são utilizados métodos geoestatísticos na realização de interpolações do tipo área-ponto e área-área (KYRIAKIDIS; YOO, 2003; KYRIAKIDIS, 2004; YOO; KYRIAKIDIS, 2006; KYRIAKIDIS; SCHENEIDER; GOODCHILD, 2005; KYRIAKIDIS, 2011).

Dados

Dados estatísticos

Os dados numéricos utilizados para a geração da Grade Estatística são os microdados do universo do Censo Demográfico 2010.

Dados vetoriais

Os dados vetoriais diretamente utilizados correspondem aos pontos de localização dos domicílios nas áreas rurais, às linhas vetoriais das faces de logradouros e às divisões de setores censitários. Os primeiros são oriundos do CNEFE e os demais fazem parte da Base Territorial.

Foram utilizadas como informação auxiliar classificações de uso e cobertura das terras proveniente de duas fontes, de acordo com a disponibilidade de dados:

- Projeto de Monitoramento do Desmatamento dos Biomas Brasileiros por Satélite – PMDBBS, disponível para todo o Brasil, exceto Amazônia Legal, desenvolvido a partir da classificação de imagens Landsat-5/TM e disponibilizado pelo MMA em 2007;
- Projeto TERRACCLASS 2010, disponível para a Amazônia Legal, desenvolvido a partir da classificação de imagens Landsat-5/TM e com apoio de imagens do sensor MODIS (Moderate Resolution Imaging Spectroradiometer) e executado pela Empresa Brasileira de Pesquisa Agropecuária (Embrapa) e Instituto Nacional de Pesquisas Espaciais (INPE).

Nos dois casos, as diversas classes de uso/cobertura foram agrupadas, sendo classificadas como “povoada” as classes relacionadas com características

antrópicas e como “não povoada” as classes relacionadas com características naturais.

Definição geométrica e espacial da grade

A Diretoria de Geociências – DGC do IBGE através da Coordenação de Cartografia - CCAR, elaborou um estudo (FRANÇA e MARANHÃO, 2015) de geração de uma grade estatística para o Brasil buscando minimizar as distorções decorrentes da grande extensão territorial do país. Neste contexto, foi feita a opção pela utilização da Projeção Equivalente de Albers, que apresenta como característica principal a equivalência em área. Essa projeção é formada por um cone ao redor do globo, sob o qual as feições da superfície terrestre são projetadas. Este cone intercepta o globo em dois paralelos padrão, com os meridianos formando linhas retas com a origem em um ponto central na geratriz do cone, enquanto os paralelos formam círculos concêntricos em torno deste ponto. O *datum* horizontal adotado é o SIRGAS2000.

Os parâmetros da projeção adotada são:

Meridiano Central	-54°
Latitude de Origem	-12°
1° Paralelo Padrão	-2°
2° Paralelo Padrão	-22°
Origem E:	5.000.000
Origem N:	10.000.000
Área de Abrangência	canto inferior esquerdo (E,N): 2.800.000, 7.350.000 canto superior direito (E,N): 8.210.000, 12.200.000

Para o propósito de disseminação de dados estatísticos, foram adotadas células com dimensões de 1 km x 1 km nas áreas rurais e 200 m x 200 m nas áreas urbanas. A base da classificação urbano-rural utilizada para determinar esta diferença no tamanho das células foi extraída da malha de setores censitários para o Censo Demográfico de 2010. Também com o propósito de disseminação, o produto final foi convertido para Projeção Geográfica.

Metodologia

Para enfrentar a complexidade da operação censitária, o IBGE busca aprimorar cada vez mais os métodos e técnicas de coleta, apuração e disseminação dos resultados. Dois desses avanços adotados no Censo Demográfico 2010 merecem ser destacados devido a sua relação com a geração da grade estatística.

O primeiro avanço foi o emprego de equipamentos de coleta eletrônicos com dispositivos GPS acoplados, que permitiu a captura de informações de localização dos domicílios visitados pelo censo. Nas áreas rurais, com algumas exceções, não existe um traçado regular de vias que permita identificação das edificações através de endereços tradicionais, então, nessas áreas foram capturados as coordenadas geográficas de cada edificação, os quais foram incluídos no Cadastro de Endereços.

O segundo avanço está relacionado com a associação do Cadastro de Endereços para Fins Estatísticos – CNEFE ao mapeamento censitário, fato que permitiu que a aplicação eletrônica desenvolvida para a coleta de dados pudesse colocar os mapas e a lista de endereços juntos numa mesma tela, aperfeiçoando e facilitando a operação censitária. Este relacionamento entre as duas bases de dados somente foi possível de ser realizado nas áreas urbanas, onde a associação do cadastro de endereços com as linhas que representam as faces de quarteirão no mapeamento foi feita através de códigos. O código de cada quadra/face acoplado ao código do setor censitário a que pertence cria uma identificação única, que permite a associação do cadastro de endereços ao mapeamento censitário e, conseqüentemente, a localização espacial dos endereços ali existentes.

Com relação à abordagem para a associação dos dados censitários à grade, constatou-se a existência de uma quantidade significativa de registros sem dados de localização, fazendo com que a abordagem de agregação não fosse suficiente para tratar a totalidade dos dados. Nas áreas urbanas existem duas fontes potenciais para a ausência de dados de localização. A primeira é a malha viária, que pode estar incompleta, e a segunda é a ausência de codificação desta malha viária, o que impede o seu relacionamento com os dados estatísticos. Estas duas

situações se apresentam geralmente em áreas não urbanizadas, áreas urbanas isoladas e aglomerados rurais. No caso das áreas estritamente rurais, nem todas as edificações tiveram as suas coordenadas geográficas registradas devido a razões operacionais e/ou técnicas. Diante dessas impossibilidades, constatou-se não ser possível empregar apenas a abordagem de agregação para poder representar todos os dados do Censo 2010 em uma grade estatística, sendo necessária a utilização de uma abordagem híbrida, combinando agregação e desagregação. Nos locais onde a totalidade ou um grande percentual dos registros apresenta dados locais, foi utilizada a abordagem de agregação; naquelas regiões onde isso não ocorreu, a abordagem de desagregação foi utilizada. Mesmo com a adoção desta estratégia ainda há um percentual de dados que não está representado através da grade estatística, ou seja, o total de população e de domicílios registrado no Censo 2010 é maior do que o valor obtido com a grade estatística. No entanto, esta diferença pode ser considerada desprezível.

Diante da decisão de utilização de uma abordagem híbrida, fez-se necessário determinar o limite para utilização de cada abordagem e o método de desagregação mais adequado. Para a determinação do limite de utilização de cada uma das abordagens foi realizada uma avaliação quantitativa da espacialização dos dados, ou seja, avaliou-se a quantidade de registros que tiveram a sua localização geográfica possível de ser realizada. Para isso, utilizou-se o setor censitário como unidade espacial e a quantidade de domicílios permanentes ocupados como variável de avaliação. A quantidade de domicílios foi calculada de duas maneiras distintas para cada setor censitário: somando o número de registros no banco de microdados e somando a quantidade de registros localizados espacialmente com utilização de relacionamentos entre a base geográfica e as bases de endereços e de dados coletados. Com esses dois resultados em mãos, calculou-se a diferença entre eles, a qual foi denominada “ausência de localização”. O mesmo cálculo foi feito para cada setor censitário, sendo registrada a ausência de localização relativa por setor. A ausência de localização relativa foi classificada em faixas e foi obtida a quantidade de setores e de domicílios pertencentes a cada uma dessas faixas. Como a quantidade de dados sem localização varia ao longo do território, optou-se por utilizar um valor médio como limite para a adoção de cada uma das abordagens. Desta maneira, nos setores censitários onde a ausência de localização é inferior a 50%, foi adotada a agregação de dados; a abordagem de desagregação

foi adotada nos setores censitários onde a ausência de localização é superior a 50%. Com esta opção, mesmo havendo perda de dados, esta perda será inferior a 5%, que é um valor aceitável diante das circunstâncias apresentadas.

Para auxiliar a escolha do método de desagregação mais adequado, foram utilizados os resultados de estudos empíricos (BUENO, 2014), cujos resultados ajudaram a esclarecer algumas questões relacionadas com as diferentes técnicas de desagregação. Avaliando os resultados obtidos com a desagregação com aqueles obtidos com a agregação, que foi considerada como referência, e considerando tanto a acurácia do mapeamento quanto o valor estimado da população, foi constatado que o método dasimétrico com dados auxiliares de vias apresenta o melhor desempenho; a segunda opção recai sobre o método dasimétrico com dados auxiliares de classificação de uso/cobertura das terras; e como última opção está o método de ponderação zonal, utilizado quando não há disponibilidade de dados auxiliares.

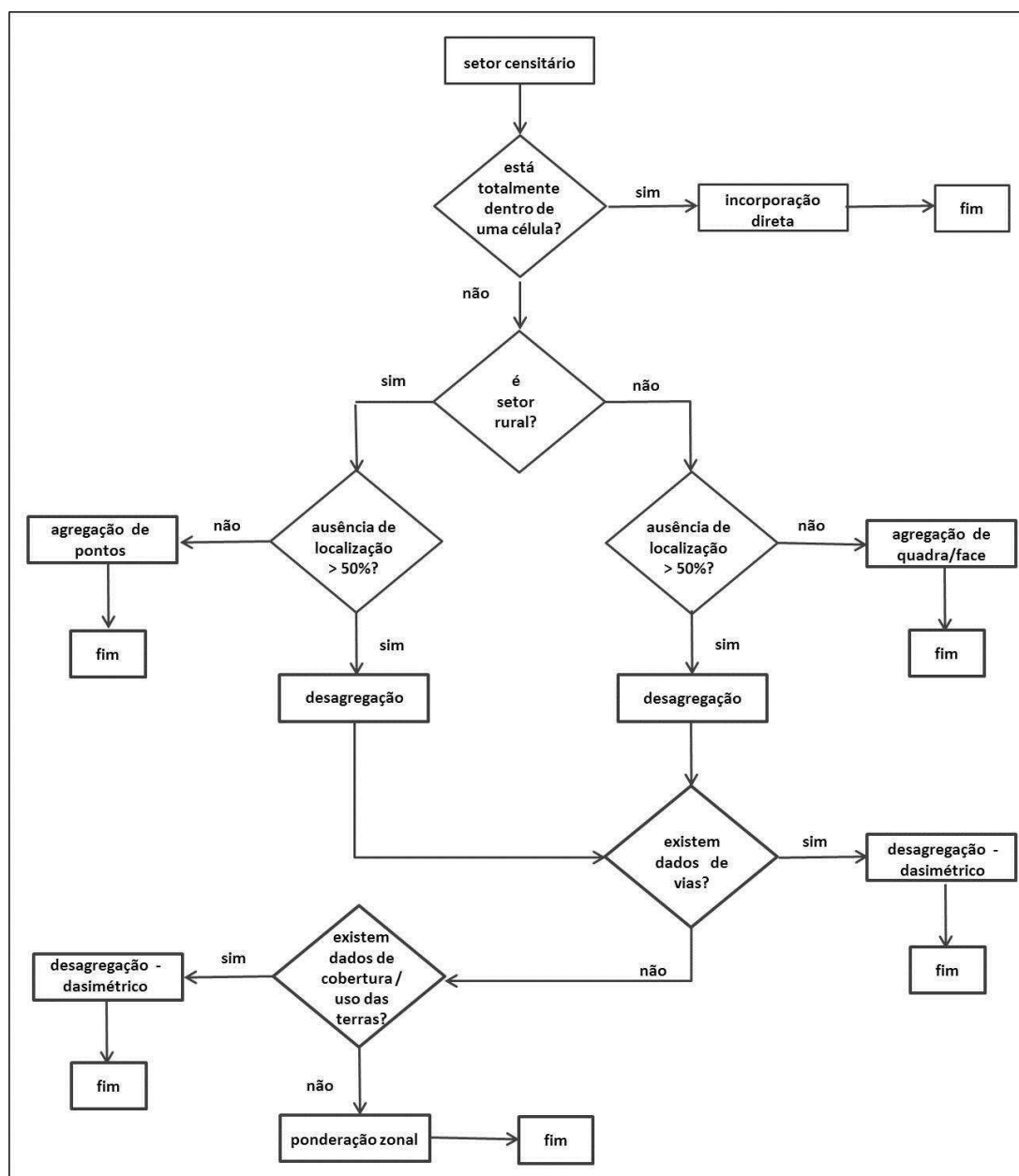
De posse das opções apresentadas para a escolha da abordagem e da técnica de desagregação (item 4) e das considerações feitas acima, foi criada uma árvore de decisão que sintetiza as opções de processamento para popular as células da grade estatística. Esta árvore de decisão, como pode ser visto na Figura 1, tem por unidade o setor censitário e considera a área do setor com relação à célula da grade, a situação territorial (rural ou urbana), o valor da ausência de localização espacial e a existência de dados auxiliares.

O procedimento para a agregação de pontos foi a espacialização das coordenadas das unidades visitadas de maneira direta seguida de uma interseção espacial entre esses pontos e os polígonos da grade.

O procedimento para agregação de quadra/face utilizou a codificação única desses elementos dentro de cada setor censitário, permitindo que os domicílios pudessem ser associados a um trecho de logradouro. Utilizando estes códigos presentes tanto no cadastro de endereços como no mapeamento censitário foi possível localizar espacialmente os domicílios das áreas urbanas que foram visitados pelo censo. Nos casos em que a face está totalmente inserida em uma única célula, o procedimento é simplesmente de soma dos dados. Nos casos em que a face não estiver totalmente inserida dentro de uma única célula, assumiu-se que a face de logradouro é homogênea, ou seja, que a distribuição dos domicílios

pode ser considerada uniforme em toda a extensão da face. A operacionalização destes procedimentos foi executada através de uma operação de interseção espacial entre a face de logradouro e a grade estatística, de modo a alocar em cada célula da grade um percentual de dados equivalentes à extensão da face localizada no interior da célula.

Figura 1: Árvore de decisão para escolha da abordagem e da técnica de desagregação.



Fonte: Adaptado de BUENO, 2014.

O procedimento adotado para o tratamento de setores censitários com dimensões pequenas quando comparado com as células da grade estatística foi a incorporação direta dos dados. Nos casos em que o setor não estiver totalmente inserido em uma célula, adotou-se um critério de tolerância: se 90% ou mais da superfície do setor estiver dentro da célula, o setor foi considerado totalmente inserido na célula.

A metodologia de desagregação foi realizada tendo como unidade de origem os setores censitários e como unidade de destino as células da grade estatística.

Um dos métodos dasimétricos empregados utiliza a malha viária como dado auxiliar, ou seja, como aproximação para a existência de residências e, conseqüentemente, de população. O procedimento operacional se inicia com a totalização do comprimento das vias no interior do setor censitário. De posse desse comprimento e da quantidade total de domicílios, calcula-se a densidade de domicílios por extensão de via. Em cada célula inserida dentro de um único setor censitário, multiplicando-se a extensão de vias dentro do perímetro celular pela densidade de domicílios por extensão de via, obtém-se a quantidade total de domicílios. Para cada setor censitário foi calculada a quantidade média de moradores por domicílio a partir da divisão da população residente total pela quantidade total de domicílios existentes naquele setor. Esta quantidade é utilizada para calcular a população nas células a partir da quantidade de domicílios obtida anteriormente através da densidade de domicílios por extensão de via.

No caso de utilização de método dasimétrico com dados de classificação de uso/cobertura das terras como uma aproximação para a localização dos domicílios, optou-se por um método binário, que implica a transformação das classes existentes em apenas duas: uma que contém população, e outra que não contém população. Este método é o mais simples descrito na literatura (LANGFORD et al., 1991) e foi selecionado pela sua simplicidade e facilidade de execução, tendo em vista a necessidade de combinar classificações de origens diversas e que apresentam classes diferentes. As classes relacionadas com corpos d'água e cobertura vegetal nativa foram consideradas como não povoadas e aquelas relacionadas com áreas urbanas ou com agricultura/agropecuária foram consideradas povoadas. O fluxo de operações necessárias para o processamento desta técnica é iniciado com a agregação das classes de uso/cobertura em classe

povoada e não povoada e com a determinação da área da classe povoada para o setor censitário considerado. Utilizando esta área e o total de domicílios no setor, calcula-se a densidade de domicílios por unidade de área. Após a execução de uma interseção espacial entre o vetor da grade e os polígonos da classe povoada, foi calculada a área de cada um desses novos polígonos contidos nas células. Com estes dados foi calculada a quantidade de domicílios através da multiplicação da área do polígono povoado inserida na célula pela respectiva densidade de domicílios calculada para o setor censitário. O volume de população foi obtido multiplicando-se a quantidade de domicílios da célula pelo número de moradores por domicílio obtido para o setor censitário através da divisão da população residente total pelo número total de domicílios existentes no setor.

Nos casos em que não há possibilidade de agregação de dados e também não existem dados auxiliares que possam auxiliar na metodologia de desagregação foi empregada a ponderação zonal simples para transformar os dados agregados em setores censitários para dados agregados em células da grade estatística. O parâmetro utilizado para realizar esta operação foi a área do setor censitário e o pressuposto de homogeneidade interna mais uma vez foi considerado verdadeiro. A quantidade de domicílios existentes no setor censitário foi redistribuída para as células de acordo com o percentual de área do setor que estava inserido em cada uma das células e o volume de população foi calculado a partir da quantidade de moradores por domicílio existente no setor censitário.

Com relação às variáveis relacionadas com características da população ou do domicílio utilizou-se a proporção destas variáveis existente no setor censitário e foi feita a replicação desta proporção para as células da grade, ou seja, calculou-se a variável por habitante ou por domicílio no setor censitário e multiplicou-se este valor pela quantidade de população ou de domicílios existente na célula.

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

A regra de proporcionalidade gera nos resultados números não inteiros para as variáveis e, como regra geral, adotou-se que os valores numéricos não serão arredondados em nenhuma etapa intermediária, sendo esta operação realizada apenas no final do processamento. Com relação aos dados espúrios originados pela aplicação da metodologia, os mesmos foram suprimidos da base de dados para evitar incongruências.

Referências

- BUENO, Maria do Carmo D. **Grade estatística: uma abordagem para ampliar o potencial analítico de dados censitários**. Tese (Doutorado) – Universidade Estadual de Campinas, Instituto de Filosofia e Ciências Humanas, Campinas, 2014.
- CÂMARA, Gilberto. Representações computacionais do espaço geográfico. In CASANOVA, Marco; CÂMARA, Gilberto; DAVIS, Clodoveu; QUEIROZ, Gilberto R. (Ed.). **Bancos de Dados Geográficos**. Curitiba: MundoGEO, 2005. Disponível em: <<http://www.dpi.inpe.br/livros/bdados/>>. Acesso em: 10 out. 2013.
- DEICHMANN, Uwe. **A Review of Spatial Population Database Design and Modelling**. Technical Report TR-96-3. National Center for Geographic Information and Analysis (NCGIA), University of California, Santa Barbara, USA, 1996.
- EFGS – EUROPEAN FORUM FOR GEOSTATISTICS. **GEOSTAT 1A – Representing Census data in a European population grid - Final Report**. 2012.
- EICHER, Cory L; BREWER, Cynthia A. Dasymetric mapping and areal interpolation: implementation and evaluation. **Cartography and Geographic Information Science**, vol. 28, n. 2, p. 125-138, 2001.
- FRANÇA, Vitor de Oliveira. **Avaliação de Métodos Dasimétricos para Estimativa Populacional em Pequenas Áreas**. Dissertação (Mestrado) - Escola Nacional de Ciências Estatísticas, Rio de Janeiro, 2012.
- _____; DINIZ, Viviane Barboza; BATISTA, Ivone Lopes; MARANHÃO, Marcelo Rodrigues de Albuquerque. **Uma Proposta para o Estabelecimento de Grades de Referência Estatística para o Brasil**. Relatório Técnico IBGE/DGC/CCAR. 2015.
- GRASLAND, Claude; MADELIN, Malika. **The Modifiable Areas Unit Problem – Final Report**. ESPON - European Spatial Planning Observation Network, 2006.
- GUTMANN, Myron P.; WITKOWSKI, Kristine; COLYER, Corey; O ´ROURKE, JoAnne M.; MCNALLY, James. Providing Spatial Data for Secondary Analysis: Issues and Current Practices Relating to Confidentiality. **Population Research and Policy Review**, vol. 27, n. 6, p. 639-665, 2008.

- GUZMÁN, José Miguel; SCHENSUL, Daniel; ZHANG, Sainan. Understanding Vulnerability and Adaptation Using Census Data. In: MARTINE, George;
- SCHENSUL, Daniel (Ed.). **The Demography of Adaptation to Climate Change**. New York, London and Mexico City: UNFPA, IIED e El Colegio de México, 2013.
- KAMINGER, Ingrid. **Assuring confidentiality with grid data**. In: GIS FOR STATISTICS, Luxemburgo, 2007.
- KARR, Alan F.; KOHNEN, Christine N.; OGANIAN, A.; REITER, J.P.; SANIL, A.P..A Framework for Evaluating the Utility of Data Altered to Protect Confidentiality. **The American Statistician**, vol. 60, n. 3, p. 224-232, 2006.
- LAM, Nina Siu-Ngan. Spatial Interpolation Methods: A Review. **The American Cartographer**, vol. 10, n. 2, p.129-150, 1983.
- LANGFORD, Mitchel; MAGUIRE, D. J; UNWIN, David J. The areal interpolation problem: estimating population using remote sensing in a GIS framework. In: MASSER, Ian; BLAKEMORE, Michael (Ed.). **Handling geographical information: Methodology and potential applications**. London: Longman, 1991.
- LIU, XiaoHang. **Estimation of the Spatial Distribution of Urban Population Using High Spatial Resolution Satellite Imagery**. Tese (Doutorado) – University of California, Santa Barbara, 2003.
- KYRIAKIDIS, Phaedon C.; YOO, Eun-Hye. Geostatistical prediction/simulation of point values from areal data. In: MARTIN, David (Ed.). **Proceedings of the 7th International Conference on Geocomputation**, Southampton, UK, 2003.
- _____. A geostatistical framework for area-to-point spatial interpolation. **Geographical Analysis**, vol. 36, n. 3, p. 259-289, 2004.
- _____; SCHNEIDER, P.; GOODCHILD, Michael F. Fast geostatistical areal interpolation. In: XIE, Yichun; BROWN, Daniel G. (Ed.). **Proceedings of the 8th International Conference on Geocomputation**, Ann Arbor, Michigan, 2005.

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

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

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

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

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

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

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

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

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

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

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

REIBEL, Michael; AGRAWAL, Aditya. Areal Interpolation of Population Counts Using Preclassified Land Cover Data. **Population Research and Policy Review**, vol. 26, p. 619–633, 2007.

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

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

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

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

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

- TAMMILEHTO-LUODE, Marja. **Opportunities and challenges of grid-based statistics**. In: WORLD STATISTICS CONGRESS OF THE INTERNATIONAL STATISTICAL INSTITUTE, Irlanda, 2011.
- TATEM, Andrew J.; NOOR, Abdisalam M.; VON HAGEN, Craig; DI GREGORIO, Antonio; HAY, Simon I. High Resolution Population Maps for Low Income Nations: Combining Land Cover and Census in East Africa. **PLoS ONE**, vol. 2, n. 12, e1298, 2007.
- TOBLER, Waldo R. Smooth Pycnophylactic Interpolation for Geographical Regions. **Journal of the American Statistical Association**, vol. 74, n. 367, p. 519-530, 1979.
- TURNER, Andy; OPENSHAW, Stan. **Disaggregative Spatial Interpolation**. In: GISRUK, Glamorgan, País de Gales, 2001.
- WRIGHT, John K. A method of mapping densities of population: With Cape Cod as an example. **Geographical Review**, vol. 26, n. 1, p. 103-110, 1936.
- WU, Shuo-sheng; QIU, Xiaomin; WANG, Le. Population Estimation Methods in GIS and Remote Sensing: A Review. **GIScience and Remote Sensing**, vol. 42, n. 1, p. 58-74, 2005.
- YOO, Eun-Hye; KYRIAKIDIS, Phaedon C. Area-to-point kriging with inequality type data. **Journal of Geographical Systems**, vol. 8, n. 4, p. 357-390, 2006.
- ZANDBERGEN, Paul A.; IGNIZIO, Drew A. Comparison of Dasymetric Mapping Techniques for Small-Area Population Estimates. **Cartography and Geographic Information Science**, vol. 37, n. 3, p. 199-214, 2010.

Equipe técnica

Centro de Documentação e Disseminação de Informações

Coordenação de Projetos Especiais

Maria do Carmo Dias Bueno

Equipe Técnica

Herlan Cassio de Alcântara Pacheco

Maria do Carmo Dias Bueno

Milene Santos Couto

Ricardo Neves de Souza Lima

Tiago de Almeida Silva

Se o assunto é **Brasil**,
procure o **IBGE**.

www.ibge.gov.br
www.twitter.com/ibgecomunica
www.facebook.com/ibgeoficial

Atendimento
0800 721 8181

