

# ANÁLISE DE DADOS PARA REDISTRIBUIÇÃO DE POPULAÇÃO EM MÚLTIPLAS ESCALAS: ESTUDO DE CASO EM SANTARÉM E BELTERRA (PA)

Mariane Souza Reis\*

Instituto Nacional de Pesquisas Espaciais, Caixa Postal 515 - 12227-010 - São José dos Campos - SP, Brasil

*Trabalho final da disciplina População, Espaço e Ambiente: CST-310-3 e SER-457-3*

21 de setembro de 2017

**Resumo.** Comparar dados do Censo de anos distintos ou integrá-los a análises que utilizam outras variáveis pode ser uma tarefa difícil, se a unidade de análise não corresponder àquela em que os dados estão disponibilizados. Nesses casos, é necessário redistribuir a população contabilizada em uma dada unidade de análise (zona fonte) para outra área (zona alvo). Existem diversas metodologias aplicáveis a essa tarefa, que podem ou não utilizar dados auxiliares. A qualidade dos resultados depende da capacidade do modelo de explicar a distribuição da população, da qualidade dos dados auxiliares se utilizados e da relação entre as zonas fonte e as zonas alvo. De forma a colaborar para o entendimento desses fatores, este trabalho objetivou avaliar dados auxiliares para redistribuir a população contabilizada pelo Censo para zonas alvo de diferentes tamanhos. As análises foram feitas considerando dados de número de moradores em domicílios particulares permanentes ou população residente em domicílios particulares permanentes, contabilizados pelo Censo 2010, nos municípios paraenses de Santarém e Belterra, contidos na Amazônia Legal brasileira. Foram avaliados dados relacionados à distância a hidrografia, distância a rodovias, distância de centros urbanos, distância a localidades, distância a estabelecimentos de saúde, distância a estabelecimentos de ensino, distância a vertentes, declividade média e três abordagens de dados de uso e cobertura da terra. Esses dados foram avaliados de forma individual e conjunta, para redistribuir a população para quatro planos celulares de tamanhos distintos: aquele derivado da Grade Estatística do IBGE e planos celulares de 1x1km, de 2x2km e de 4x4km. Esses resultados foram comparados com duas referências: dados de população total contidos na Grade Estatística e população redistribuída para domicílios obtidos a partir dos dados do Cadastro Nacional de Endereços para Fins Estatísticos, ambos disponibilizados pelo Instituto Brasileiro de Geografia e Estatística. Quanto maior o tamanho da célula utilizada, mais similares foram os dados redistribuídos e as referências. Dados de uso e cobertura da terra retornaram os resultados mais próximos às referências utilizadas. Em especial, atribuir valores distintos de potencial da presença da população para determinados grupos de classes de uso e cobertura retorna resultados mais próximos às referências, mesmo quando comparados a abordagens multivariadas, em que esse e outros dados são utilizados conjuntamente. Esse tipo de dado auxiliar pode ser obtido por meio de imagens de sensoriamento remoto, disponibilizadas para vários anos a partir da década de 70, tratando-se de um dado utilizável para análises que considerem dados censitários anteriores a 2010.

## 1. Introdução

Diversas aplicações tanto nas ciências quanto na política demandam de informações confiáveis acerca da distribuição da população (BRIGGS et al., 2007). No campo ambiental, existe uma longa tradição de se relacionar fatores demográficos e sociais com a pressão em recursos naturais e mudanças ambientais, provavelmente desde a publicação de *An Essay on the Principle of Population* por Thomas Malthus em 1798. Nos últimos anos, no entanto, tem-se considerado a população como um fator que não apenas modifica o meio, mas que interage com outros fatores de origens sociais e ambientais, podendo tanto moldá-los quanto ser moldada por eles (HOGAN, 2001; CÔRTEZ; D'ANTONA, 2014). Nesse sentido, não apenas estudos ambientais prescindem de informações sobre a distribuição e estrutura da população, como a questão do espaço e do ambiente tem sido sistematicamente integradas em estudos demográficos (CÔRTEZ; D'ANTONA, 2014). Relações entre a população e o ambiente são complexas, sendo necessários dados de variadas fontes para entender fatores sociais e biofísicos que fazem parte dessa relação (EVANS; MORAN, 2002).

Dados ambientais espacialmente explícitos, como mapas de uso e cobertura da terra, elevação, hidrografia e outros são disponibilizados em diferentes escalas, que podem ou não ser adequadas para a análise. Dados de sensoriamento remoto têm colaborado para a aquisição de informações em diferentes resoluções espaciais e temporais, sendo que a unidade espacial de análise geralmente corresponde ao pixel (AMARAL et al., 2012).

No Brasil, a principal fonte de dados sobre o número e composição da população é o Censo Demográfico, realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE), de forma decenal. Esses dados são coletados no nível domiciliar e, por motivos operacionais e de manutenção do sigilo das informações, divulgados de forma agregada por unidades territoriais, das quais a menor é o setor censitário (IBGE, 2013). Segundo IBGE (2010), um setor censitário corresponde "*à unidade territorial estabelecida para fins de controle cadastral*" e trata-se de uma "*área contínua, situada em um único quadro urbano ou rural, com dimensão e número de domicílios que permitam o levantamento por um recenseador*". Por um lado, nem todas as informações coletadas pelo Censo estão disponíveis nessa unidade territorial. Por outro, como essa unidade é definida por critérios operacionais, os limites dos setores censitários podem não se manter para diferentes anos de Censo, bem como podem não respeitar limites naturais ou administrativos de interesse, além de possuírem tamanhos muito variados. A título de exemplo, o estado do Pará, localizado na região norte do Brasil e dentro dos limites da Amazônia Legal brasileira, foi dividido em 8919 setores censitários em 2010, sendo que o de menor extensão possui pouco mais de 0,003 km<sup>2</sup>, enquanto o de maior possui mais de 62500 km<sup>2</sup>.

Dessa forma, dados censitários são disponibilizados em unidades territoriais modificáveis e arbitrárias (setores censitários), o que se enquadra na definição do *Problema de Unidade de Área Modificável* (*Modifiable Areal Unit Problem* - MAUP) de Openshaw (1984). Segundo o autor, essas unidades espaciais não possuem nenhum significado geográfico intrínseco, mas é provável que todas as análises derivadas dos mesmos se baseiem na delimitação dessas unidades, o que pode invalidar essas análises. De fato, resultados estatísticos podem variar drasticamente dependendo da escala e padrão das unidades espaciais de análise utilizadas (FLOWERDEW, 2011). Além disso, a simples representação desses dados em mapas coropléticos transmite a impressão de que existem mudanças abruptas na distribuição da população nos limites das unidades espaciais, ao passo em que representa a presença da população de forma contínua no espaço (HOLT et al., 2004).

---

\*mariane.reis@inpe.br

A não conformidade interna entre a unidade espacial de análise de dados censitários, bem como entre dados de diferentes anos e/ou de outras fontes, por vezes implica na necessidade de compatibilizar a unidade de análise, o que envolve redistribuir a população, ou outra variável de interesse, contabilizada em um setor censitário para outra área. A zona, ou unidade de análise, que contém a informação original do dado agregado é definida como zona fonte, enquanto aquela para a qual se deseja redistribuir o dado é chamada de zona alvo (GOODCHILD et al., 1993). A menos que zonas alvo sejam iguais às zonas fonte, ou possam ser adquiridas da agregação destas, necessita-se de métodos de interpolação para redistribuir a população em novas áreas (LAM, 1983; FLOWERDEW; GREEN, 1994; LANGFORD, 2013), o que pode ou não envolver a utilização de dados auxiliares. A comparação entre diferentes dados e métodos de interpolação trata-se de um estudo importante para diversas disciplinas, uma vez que a qualidade dos resultados varia para diferentes áreas de estudo, em que a população se distribui segundo diferentes lógicas (ZANDBERGEN; IGNIZIO, 2010; LANGFORD, 2013).

Para testar os resultados e/ou comparar diferentes métodos, as estimativas (dados de população redistribuídos) comumente são realizadas considerando zonas alvo para a qual a quantidade real da população é conhecida (referência). Segundo Langford (2013), essa necessidade leva vários autores a testarem métodos e/ou dados considerando um nível maior de agregação do Censo como zona fonte e um nível menor de agregação do mesmo dado como zona alvo. No entanto, segundo o autor, essa metodologia implica em diversos questionamentos, do qual se salienta a não certeza de que o modelo que explica a distribuição da população em uma determinada escala é capaz de fazê-lo em escalas consideravelmente mais refinadas. Nesse sentido, em alguns trabalhos as zonas alvo são consideravelmente menores que as zonas fonte, sendo que dados de referência para essas zonas alvo nem sempre estão disponíveis na resolução (espacial ou temporal) desejada (BRIGGS et al., 2007; YANG et al., 2013; DMOWSKA; STEPINSKI, 2017).

Assim, questões sobre a possibilidade de se usar o mesmo modelo para explicar a distribuição da população considerando diferentes resoluções espaciais ainda permanecem, mesmo que se saiba que quanto mais próxima a zona alvo for da zona fonte, mais parecidos sejam também os resultados da estimativa da população com os valores de referência. Na Amazônia brasileira, a relação entre confiabilidade (ou acurácia) dos dados redistribuídos para diferentes escalas espaciais ainda é pouco estudada. Necessita-se avaliar se os modelos existentes explicam a distribuição populacional considerando diferentes tamanhos de zonas fonte. Avanços no conhecimento a esse respeito podem ser úteis para estudos em que a distribuição da população seja analisada conjuntamente com outras variáveis ambientais que, como mencionado anteriormente, podem descrever fenômenos em escalas distintas àsquelas utilizadas pelo IBGE para disponibilização de dados censitários.

De forma a colaborar para o entendimento dessas questões, este trabalho objetivou avaliar dados auxiliares para redistribuir a população contabilizada pelo Censo para zonas alvo de diferentes tamanhos. Para cumprir esse objetivo, a população contabilizada nos setores censitários de dois municípios contidos na Amazônia brasileira foi redistribuída com base em diferentes dados auxiliares, para quatro superfícies regulares, considerando resoluções espaciais distintas para cada uma. Os municípios escolhidos para a análise estão descritos na Seção 2. A justificativa para a escolha dos dados auxiliares, bem como detalhes metodológicos estão apresentada na Seção 3. Os resultados obtidos estão apresentados na Seção 4 e as conclusões desse trabalho estão delineadas na Seção 5.

## 2. Área de estudo

Este estudo foi desenvolvido em uma área de aproximadamente 27300 km<sup>2</sup>, que compreende os municípios de Santarém e Belterra, ambos localizados no sudoeste do estado do Pará, na Amazônia Legal brasileira. A área de estudo está ilustrada na Figura 1.

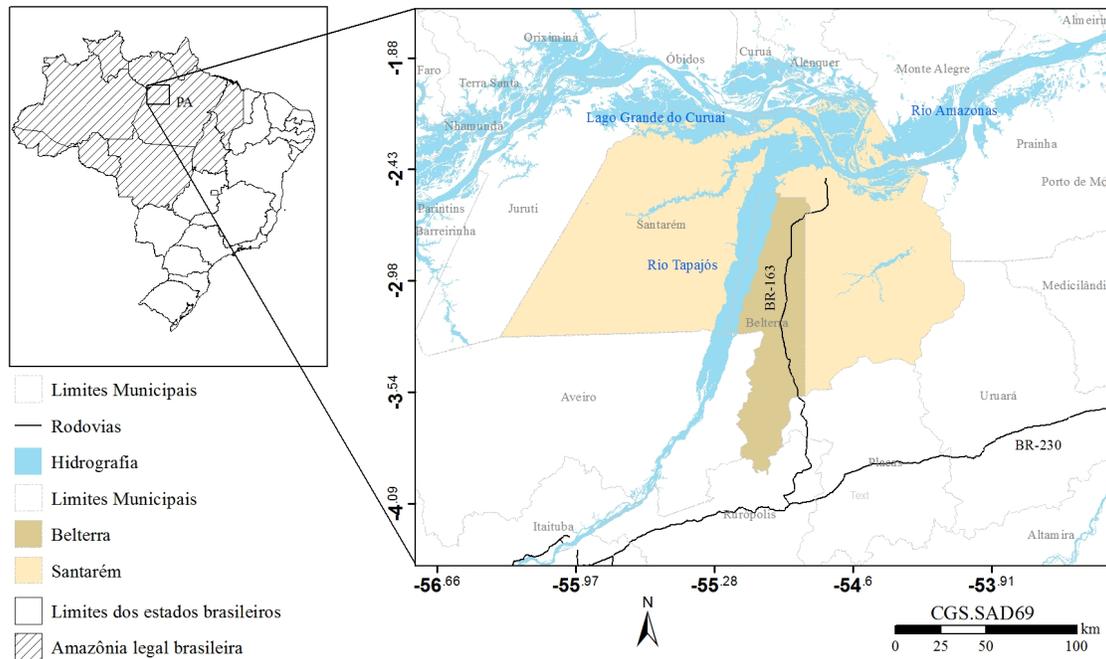


Figura 1: Área de estudo: municípios de Santarém e Belterra, no estado do Pará.

Esses municípios foram escolhidos para a análise por suas características pertinentes para o objetivo proposto. Na região abrangida por esses dois municípios destaca-se a presença de unidades de conservação federais como a Reserva extrativista Tapajós-Arapiuns e a Floresta Nacional Tapajós, bem como a presença do porto da Cargill em Santarém, a proximidade dos municípios das rodovias Cuiabá-Santarém (BR-163) e Transamazônica (BR-230) e dos rios Tapajós e Amazonas, bem como das capitais amazônicas Belém e Manaus. O município de Santarém é classificado como um centro regional de influência C pelo IBGE (IBGE, 2008), sendo então um importante "*centro urbano, financeiro, comercial e cultural do oeste do estado do Pará*" (PREFEITURA DE SANTARÉM, s.d.). Em 2016, possuía unidade territorial de 17898,4 km<sup>2</sup>, sendo a população e a densidade demográfica no ano de 2010 294580 e 12,87 hab/km<sup>2</sup>, respectivamente. Para 2017, a estimativa da população é de 296302 habitantes. Trata-se de uma das cidades mais antigas da região amazônica e a terceira mais populosa do estado do Pará (PREFEITURA DE SANTARÉM, s.d.; IBGE, 2017). O município de Belterra possuía unidade territorial de 4398,4 km<sup>2</sup> em 2016, sendo a população em 2010 contabilizada em 16318 habitantes, com densidade demográfica no mesmo ano de 3,71 hab/km<sup>2</sup>. Estima-se a população para 2017 em 17249 habitantes (IBGE, 2017). São municípios com alta heterogeneidade de uso e cobertura da terra (ESCADA et al., 2009) e diferentes padrões de distribuição populacional, que apresentam setores censitários urbanos e rurais de diferentes tamanhos. São também foco de projetos desenvolvidos no Instituto Nacional de Pesquisas Espaciais (INPE), para as quais dados de diferente datas estão disponíveis (ESCADA et al., 2009; DAL'ASTA et al., 2011; AFFONSO et al., 2016; DAL'ASTA et al., 2017)

### 3. Materiais e métodos

Existem diversos métodos de interpolação que podem ser aplicados para a redistribuição da população. Dentre os mais simples, destaca-se o método ponderado pela área, em que zonas de intersecção entre zonas fonte e alvo são criadas, e cada uma recebe a população que habitaria a zona de intersecção ao se considerar que a população está homogeneamente distribuída na zona fonte (GOODCHILD; LAM, 1980; LANGFORD, 2013). Dados auxiliares não são necessários para esse processo. No entanto, a suposição de que a população se distribui de forma homogênea é, por muitas vezes, falsa, o que pode levar a erros grosseiros na análise. Seus resultados, no entanto, podem ser considerados como um referencial para a comparação de outros métodos e são utilizados dessa forma em trabalhos como Mennis e Hultgren (2006) e Langford (2013). Outros métodos que não utilizam dados auxiliares e que merecem destaque, apesar de não discutidos nesse trabalho, são o picnoflático de Tobler (TOBLER, 1979), análises baseadas em centróides zonais (MARTIN, 1989; BRACKEN; MARTIN, 1989) e krigeagem.

Uma vez identificadas as variáveis que indicam a presença da população em determinada área de estudo, pode-se utilizá-las para auxiliar a redistribuir dados populacionais dentro das zonas fonte. De forma geral, variadas abordagens e dados auxiliares têm sido utilizados para calcular superfícies contínuas usadas como suporte para a distribuição dos dados populacionais: as chamadas superfícies adjacentes. Os métodos existentes na literatura recebem classificações como métodos inteligentes, métodos dasimétricos (WRIGHT, 1936) e métodos multivariados (AMARAL et al., 2012), para citar algumas. No geral, as abordagens podem considerar apenas um dado auxiliar (ZANDBERGEN; IGNIZIO, 2010; LANGFORD, 2013) ou um conjunto de dados de forma integrada (AMARAL et al., 2012; YANG et al., 2013). Há também abordagens binárias (LANGFORD, 2013; JEGA et al., 2017), em que as áreas são classificadas em povoadas e não povoadas e o dado populacional é redistribuído de forma uniforme nas áreas habitadas, ou ainda considerar abordagens multi-classes (SU et al., 2010; DMOWSKA; STEPINSKI, 2017), em que os dados auxiliares são utilizados para extrair informações acerca da variação interna da densidade populacional nas áreas habitadas. Os resultados desses métodos dependem da capacidade do dado auxiliar de explicar a distribuição da variável em análise, ou seja, de o dado refletir as lógicas de ocupação locais, da qualidade e da disponibilidade desses dados (SU et al., 2010; ZANDBERGEN; IGNIZIO, 2010; LANGFORD, 2013). Nesse sentido, é importante conhecer o potencial dos dados auxiliares utilizados, que podem ser avaliados pelos resultados de redistribuição da população obtidos com seu uso.

No presente estudo, identificou-se um conjunto de dados auxiliares relacionados à distribuição da população na região de Santarém e Belterra, tendo como base as análises realizadas no estado do Pará e na região do Distrito Florestal Sustentável (DFS) da BR-163, no qual a área de estudo está parcialmente inserida, descritas em Kempel (2003), Gavlak (2011) e Amaral et al. (2012). De forma individual, superfícies adjacentes foram calculadas para cada dado, considerando um base matricial com pixels de 10x10m. Essas superfícies foram então utilizadas para redistribuir a população também em pixels de 10x10m. A população contida nesses pixels foi então agregada em planos celulares, pela soma da população estimada nos pixels contidos em cada célula. Os resultados foram avaliados pela comparação entre estes e os dados da Grade Estatística do IBGE, bem como por estimativas da população feitas com base nos dados do Cadastro Nacional de Endereços para Fins Estatísticos (CNEFE). Os resultados também foram comparados àqueles obtidos pelo método ponderado pela área. A partir dos resultados obtidos, procedeu-se à análise do uso conjunto de determinados dados auxiliares, para verificação de melhorias na redistribuição da população. Esses processos estão resumidos no fluxograma da Figura 2 e explicadas na sequência. Destaca-se que ambos os dados utilizados como referência não se tratam de uma verdade acerca da distribuição

da população, mas apenas uma base de comparação entre os resultados. Além disso, a escolha da resolução espacial de 10x10m para a desagregação da população possui possíveis impactos para a análise dos resultados obtidos. Esses aspectos também estão discutidos nas próximas seções.

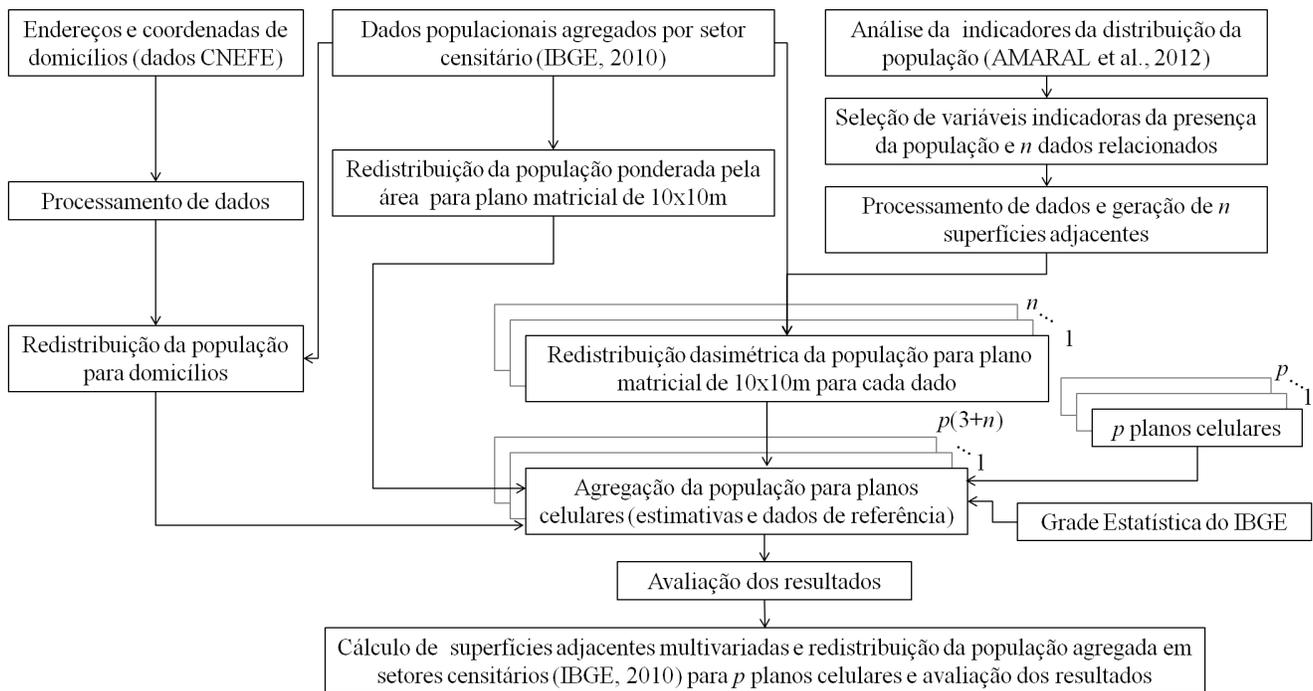


Figura 2: Metodologia adotada para análise de dados auxiliares para redistribuição da população.

### 3.1. Variáveis indicadoras da presença da população e dados utilizados

A ocupação da área de estudo se deu em duas etapas principais. Inicialmente, a ocupação de seu nas proximidades dos rios, primeiramente por indígenas e então por jesuítas, no século XVII, com o povoamento da região do Rio Tapajós se consolidando pelas relações comerciais influenciadas pelas atividades portuárias (CARDOSO, 2010). Na década de 1970 foi lançado o Projeto de Integração Nacional (PIN), que se tratou do principal estímulo para criação de obras de infraestrutura na região, decretando a construção imediata das rodovias Transamazônica (BR-230) e Cuiabá-Santarém (BR-163), com áreas ao redor dessas rodovias destinadas para colonização e reforma agrária (BRASIL, 1970; BECKER, 1991). Dessa forma, supõe-se que a população se distribua preferencialmente nas proximidades de rios e rodovias. Segundo Gavlak (2011), 90% das comunidades do DFS estão a menos de 20km de rios, com distância média de 7,6km. Similarmente, 90% das comunidades da região também estão situadas a menos de 20km de estradas, com valor médio de distância entre estas de 9,7km.

Além do processo de ocupação, acessibilidade, disponibilidade de recursos naturais, presença de infraestrutura e características biofísicas também se relacionam com a presença da população em uma região específica (AMARAL et al., 2012). No que se relaciona à presença de infraestrutura, considerou-se nesse trabalho que a população se concentra, preferencialmente, próxima a centros urbanos, vilas ou comunidades. Adicionalmente, a construção de estabelecimentos de ensino e de saúde segue uma lógica de planejamento relacionada à distribuição da população. Portanto, a presença desses estabelecimentos também trata-se de um indicador da concentração populacional.

Segundo Gavlak (2011), 90% das localidades do DFS estão localizadas a mais de 1km de vertentes e nenhuma ocorre em até 200m de distância. Kampel (2003) também observou que no estado do Pará, os distritos se localizam em áreas de declividades médias entre 0 e 7,3%, com 50% dos distritos com declividade média inferior a 0,27%. Assim, a distância a vertentes e a declividade do terreno também foram considerados como condicionantes da presença da população neste trabalho.

Adicionalmente, considerou-se que a população habita preferencialmente regiões com determinados tipos de uso e cobertura da terra, que decorrem dessa ocupação, como áreas urbanas e áreas rurais com indícios de cultivo familiar. Em contrapartida, não habita ou habita com menor intensidade áreas com cobertura florestal densa ou rios.

Em resumo, as variáveis e os dados relacionados considerados são:

- distância a hidrografia, calculada a partir dos dados de hidrografia fornecidos em escala de 1:250000 para a Amazônia Legal brasileira pelo IBGE (IBGE, 2009);
- distância a rodovias, calculada a partir dos dados de rodovias do mesmo conjunto de dados (IBGE, 2009);
- distância a centros urbanos, sendo centros urbanos os polígonos classificados como *Área Urbana* nos dados de 2010 do Mapeamento do Uso e Cobertura da Terra nas Áreas Desflorestadas da Amazônia Legal (TerraClass) (ALMEIDA et al., 2016) ;
- distância a localidades, com localidades pertencentes ao conjunto de dados disponibilizados por IBGE (2009);
- distância a estabelecimentos de saúde, com a localização dos estabelecimento de saúde extraídos dos dados de 2010 do CNEFE, após etapa de pré-processamento descrito na Seção 3.3;
- distância a estabelecimentos de ensino, sendo que a localização dos estabelecimentos de ensino também definidos pelos dados de 2010 do CNEFE;
- distância a vertentes. Regiões de vertentes foram estimadas a partir de dados do *Shuttle Radar Topography Mission* (SRTM), com resolução espacial de 3 arcos de segundo (aproximadamente 90m) e metodologia definida por Pinheiro et al. (2016) e Rennó et al. (2008);
- declividade média, que corresponde à média de janelas de 5x5 pixels dos valores de declividade calculados a partir dos dados SRTM com resolução espacial de 3 arcos de segundo;
- uso e cobertura da terra em 2010, proveniente dos dados TerraClass.

Para a variável de uso e cobertura da terra, três abordagens foram consideradas. Na primeira delas, considerou-se que pessoas não habitam áreas de florestas e rios, sendo que a distribuição da população nas demais áreas não segue padrão conhecido; supondo-se então distribuição homogênea. Na segunda abordagem, diferentes pesos foram atribuídos para áreas não-habitadas (florestas e rios) e habitadas (demais classes), considerando se estas áreas possuem ou não em sua vizinhança áreas habitadas e não-habitadas, respectivamente. Nessa abordagem, uma região habitada localizada entre regiões habitadas possui maior potencial de presença de população que uma região habitada localizada nas proximidades de uma região não-habitada, uma área não-habitada próxima de uma área habitada possui um pequeno potencial de presença de população e uma área não-habitada

cercada de áreas não-habitadas possui potencial nulo de presença de população. Na terceira abordagem foi suposto que algumas classes de uso e cobertura da terra possuem maior adensamento populacional que outras, então estas receberam diferentes pesos para redistribuição da população.

A classificação de áreas de vertentes, a média da declividade e os planos de uso e cobertura da terra foram gerados considerando a resolução original dos dados e reamostrados para 10m após o cálculo das superfícies adjacentes (Seção 3.2). Todas as medidas de distância, incluindo a distância a vertentes, foram calculadas considerando uma matriz vetorial com pixels de 10x10m e distância Euclidiana entre o centro do pixel e a feição mais próxima, não importando a resolução dos dados de entrada (dados TerraClass possuem resolução de 30m, enquanto dados SRTM foram utilizados com aproximadamente 90m). Esse tamanho de pixel foi adotado por ser a menor resolução possível de ser utilizada com a capacidade de processamento disponível. Acredita-se que, apesar da imensa geração de dados redundantes, os dados calculados nesse tamanho de pixel também conservem informação o suficiente para permitir a reamostragem dos dados, superfícies adjacentes ou mesmo população redistribuída em diferentes escalas, com problemas de pixelização ou perda de limites de bordas minimizados. A média da declividade, para cada pixel de aproximadamente 90m, foi calculada considerando uma janela de 5x5pixels, sendo o pixel estimado aquele localizado no centro da janela. Analogamente, na abordagem do uso de e cobertura da terra que considera a porcentagem de áreas habitadas em uma vizinhança, cada pixel de 30m recebe o número de pixels não-habitados em uma janela de 5x5 pixels, sendo o pixel estimado considerado na janela e localizado no centro da mesma.

A respeito dos dados de uso e cobertura da terra, nas duas primeiras abordagens as classes *Floresta* e *Hidrografia* foram consideradas como não-habitadas, sendo as demais classes consideradas habitadas. Para a terceira abordagem, as classes foram reclassificadas em cinco grupos com diferentes potenciais de ocupação, sendo o peso atribuído para cada grupo definido por AHP (*Analytic Hierarchy Process*)(SAATY, 1980), livremente baseado na análise AHP realizada por Dal’Asta (2016). São cinco grupos, ordenados pelo potencial de adensamento populacional atribuído (do menor para o maior):

- Grupo 1: classes *Floresta*, *Hidrografia* e *Reflorestamento*. Reflorestamento não havia sido considerada anteriormente porque informações de *Floresta* e *Hidrografia* podem ser adquiridos para a Amazônia Legal brasileira de forma anual desde 2000, a partir de dados do Programa de Cálculo de Desflorestamento da Amazônia (PRODES), enquanto áreas de *Reflorestamento* estão contidas sob a máscara utilizada pelo projeto. Dessa forma, ao se excluir a classe *Reflorestamento* das demais abordagens, permite-se que essas sejam reproduzidas para outras áreas e anos com uso de dados PRODES. No entanto, áreas de reflorestamento podem possuir densa cobertura florestal, não condizentes com altas densidades populacionais, de forma análoga às florestas;
- Grupo 2: classes *Mineração*, *Área não observada*, *Desflorestamento*, *Outros* e *Vegetação Secundária*. Tratam-se de classes de baixo potencial de presença de população ou classes que recentemente foram mapeadas como *Florestas* e não se pode prever o atual potencial de adensamento populacional, por não apresentarem uso definido;
- Grupo 3: classes *Agricultura Anual*, *Não floresta*, *Pasto limpo*, *Pasto sujo*, *Regeneração com pasto* e *Pasto com solo exposto*.
- Grupo 4: classe *Mosaico de ocupações*. Essa classe denota a presença de pequenos produtores e comunidades;

- Grupo 5: classe *Área urbana*.

Para uso da AHP, os grupos de classes foram comparados de forma pareada, sendo atribuídos valores relacionados ao quão mais relevante é um grupo de classes em relação a outro para a variável analisada (presença da população). Essa comparação está ilustrada na Tabela 1. Os pesos atribuídos para cada grupo de classes foram então calculados, sendo que critérios mínimos para a aceitação dos pesos foram atingidos <sup>1</sup>. Foram atribuídos os pesos: 0,04 (Grupo 1), 0,06 (Grupo 2), 0,11 (Grupo 3), 0,22 (Grupo 4) e 0,57 (Grupo 5).

Tabela 1: Comparação pareada entre potencial de presença de população para os grupos de classes de uso e cobertura da terra. Grau de importância dado do grupo da linha para o grupo da coluna.

	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5
Grupo 1	1	1/2	1/3	1/5	1/9
Grupo 2	2	1	1/3	1/4	1/8
Grupo 3	3	3	1	1/4	1/7
Grupo 4	5	4	4	1	1/5
Grupo 5	9	8	7	5	1

Em que: 1=igual importância; 2=um pouco melhor; 3=algo melhor; 4=moderadamente melhor; 5=melhor; 6=bem melhor; 7=muito melhor; 8=criticamente melhor e 9=absolutamente melhor.

Os dados auxiliares, bem como a delimitação dos setores censitários e os dados CNEFE e da Grade Estatística do IBGE, foram convertidos para UTM WGS84, fuso 21S e geometricamente corrigidos, quando houve deslocamento observável. Dados TerraClass 2010 e aqueles baseados em SRTM foram utilizados em formato matriz e resolução original (30 e 90m, respectivamente). Os setores censitários também foram utilizados em formato matricial, com pixels de 10x10m e contendo a informação de moradores em domicílios particulares permanentes ou população residente em domicílios particulares permanentes. do CENSO 2010. Os demais dados foram utilizados em formato vetorial.

### 3.2. Geração de superfícies adjacentes e redistribuição da população

Para a geração das superfícies adjacentes baseadas nos dados de uso e cobertura da terra e abordagem binária (áreas dividida em habitadas ou não-habitadas), os pixels rotulados como *Floresta* ou *Hidrografia* receberam potencial de presença da população 0 e os demais 1. Para a abordagem ponderada (diferentes pesos para grupos de classes), os potenciais adotados são aqueles calculados por AHP. As superfícies adjacentes baseadas nos demais dados auxiliares foram calculadas utilizando funções de pertinência *fuzzy* (ZADEH, 1988), em que cada pixel recebe um valor entre 0 e 1, que denotam pertinências mais baixas ou mais altas a determinado conjunto, ou ainda como é no caso, menor ou maior potencial de presença da população. Tendo como base trabalhos anteriormente desenvolvidos na área de estudo (KAMPEL, 2003; GAVLAK, 2011; AMARAL et al., 2012), a função quadrática para todos os dados. Nos trabalhos anteriores, a função quadrática foi desenvolvida adotando-se um limite mínimo (abaixo do qual a função recebe valor 0) e um máximo (acima do qual a função retorna 1). No trabalho presente, descartou-se a utilização de um dos limites, para evitar transições abruptas do potencial de presença da população. A função quadrática adotada é descrita como:

<sup>1</sup>Saaty (1980) estabelece a necessidade de se atingir valores menores que 0,1 para a medida de Razão de Consistência. A presente análise possui Razão de Consistência igual a 0,07.

$$f(x) = \begin{cases} 1 & \text{se } x \geq x_1 \\ 1/(1 + \alpha(x - \beta)^2) & \end{cases} \quad (1)$$

em que  $x$  é o valor da variável indicadora da presença da população e  $\alpha$  e  $\beta$  podem ser estimados pelo conhecimento dos valores de  $x_1$  e  $x_2$ , respectivamente os valores de  $x$  que retornam  $f(x_1) = 1.0$  e  $f(x_2) = 0.5$ .<sup>2</sup> Exclusivamente para o cálculo da superfície adjacente dos dados de distância a vertentes, têm-se que  $f(x) = 1$  se  $x \leq x_1$ . Os valores calculados para cada variável estão apresentados na Tabela 2, sendo que os valores de  $x_1$  e  $x_2$  adotados correspondem ou são baseados naqueles apresentados em Amaral et al. (2012).

Tabela 2: Valores determinados para inferência *fuzzy*.

Variável	$x_1$	$x_2$	$\alpha$	$\beta$
Distância a hidrografia	900	7700	2,16E-08	900
Distância a rodovias	900	9700	1,29-08	900
Distância de centros urbanos	2000	30000	1,28E-09	2000
Distância a localidades	2000	30000	1,28E-09	2000
Distância a estabelecimentos de saúde	2000	30000	1,28E-09	2000
Distância a estabelecimentos de ensino	2000	30000	1,28E-09	2000
Distância a vertentes	1000	500	4,00E-06	1000
Declividade média (%)	0,27	0,58	10,41	0,27
Uso e cobertura da terra (% de vizinhos não-habitados)	0,05	0,30	16,00	0,05

A operação proposta retorna superfícies adjacentes com valores de potencial da presença da população máximo para áreas de hidrografia tanto considerando as variáveis distância a hidrografia e declividade média. Para evitar que grande parte da população seja alocadas nessas áreas, o que não é esperado para a área de estudo, atribuiu-se o valor 0 para pixels que correspondem a corpos hídricos segundos os dados de hidrografia previamente utilizados.

Após o cálculo das superfícies adjacentes, aquelas calculadas considerando outros tamanhos de pixels foram reamostradas para 10x10m. A estimativa da população para cada superfície adjacente é dada por:

$$\widehat{P}_i = P_{ki} \times \frac{F_i}{\sum_{l=1} F_{lk}} \quad (2)$$

em que  $\widehat{P}_i$  corresponde à população estimada para o pixel  $i$ .  $P_{ki}$  é a população contabilizada no setor  $k$  a que pertence o pixel  $i$ ,  $F_i$  é o valor da superfície adjacente em  $i$  e  $F_{lk}$  é o valor de todos os pixels  $l$  contidos no setor censitário  $k$ .

### 3.3. Cadastro Nacional de Endereços para Fins Estatísticos: processamento dos dados e redistribuição da população para domicílios

O Cadastro Nacional de Endereços para Fins Estatísticos (CNEFE) trata-se de um cadastro de endereços de apoio às pesquisas domiciliares do IBGE (IBGE, 2013). O dado corresponde a uma lista de estabelecimentos identificados por setor censitário, sendo que parte dos estabelecimentos localizados em setores rurais possui informação de localização em coordenadas geográficas (latitude

<sup>2</sup>O valor atribuído a  $\beta$  corresponde ao  $x_1$ , e  $\alpha$  é dado por  $\alpha = 1/(x_2 - \beta)^2$ .

e longitude) e endereço da face do logradouro, enquanto outros estabelecimentos nesses setores e aqueles localizados em setores urbanos possuem apenas a informação do endereço da face do logradouro. Adicionalmente, os estabelecimentos estão identificados em *domicílio particular*, *domicílio coletivo*, *estabelecimento agropecuário*, *estabelecimento de saúde*, *estabelecimento de outras finalidades* e *edificação em construção*.

O CNEFE dispõe de 106299 endereços para os municípios de Santarém e Belterra, dos quais apenas 19463 possuem coordenadas geográficas. Para espacializar os demais endereços, utilizou-se a ferramenta de geocodificação disponibilizada em <https://www.doogal.co.uk/BatchGeocoding.php>, baseada no Google API. Dos 86836 endereços geocodificados, 10688 falharam (não foram alocados) e 35056 foram alocados em um município ou setor incorreto, sendo 3762 pertencentes a setores rurais e 31294 pertencentes a setores urbanos. Dos endereços geocodificados válidos (alocados no setor censitário correto), 10 pertencem a setores rurais e 41099 a setores urbanos. Juntamente com os endereços cuja coordenada foi fornecida pelo IBGE, tem-se que 56,98% dos endereços disponíveis são considerados como pontos válidos.

A informação de localização dos estabelecimentos de saúde e dos estabelecimentos de ensino foi utilizada para a geração dos dados auxiliares correspondentes. Pontos rotulados como *domicílio particular* foram utilizados para redistribuir a população agregada em setores censitários, considerando que a distribuição é uniforme por domicílios contidos em um mesmo setor e que os domicílios não localizados provavelmente se distribuem nas proximidades dos demais, ou seja, a alocação da população em um ponto válido não possui grandes interferências na análise. Em função da impossibilidade de se utilizar grande parte da informação do CNEFE, não há pontos válidos alocados em alguns setores censitários com população contabilizada maior que zero. Nesses casos, incluiu-se um único ponto no centróide do setor, no qual toda a população daquele setor foi alocada. Foram incluídos 4 pontos desse tipo.

### 3.4. Agregação da população para planos celulares

A população redistribuída para planos matriciais de 10x10m foi então agregada para quatro planos celulares:

- a Grade Estatística do IBGE, em que setores urbanos estão divididos em células de 200x200m e setores rurais em células de 1x1km;
- plano celular regular 1x1km;
- plano celular regular 2x2km;
- plano celular regular 4x4km;

pela soma dos valores contidos dentro dos polígonos de cada plano celular. Os valores de população contidos na Grade Estatística do IBGE também foram agregados para os planos celulares regulares de 1km, 2km e 4km, pela soma ponderada pela área, e utilizados como referência para comparar os demais resultados obtidos. Ressalta-se que os valores da Grade Estatística do IBGE também são estimados e a agregação destes não tende a retornar a população contabilizada no setor censitário ou no município. Assim, esse dado não se trata de uma verdade sobre a distribuição da população e sim de uma estimativa realizada pelo próprio IBGE. No entanto, trata-se de uma referência aceitável para análises realizadas com dados de 2010.

Segundo D'Antona et al. (2013) "*a agregação, em qualquer unidade de área, de dados individuais obtidos em registros administrativos ou em pesquisas domiciliares, com critérios estabelecidos*

*de qualidade e regularidade, apresenta resultados superiores a qualquer método estatístico ou geoestatístico utilizado para estimação de dados*". Os autores afirmam que a metodologia de agregar pontos individuais não se trata de uma técnica de estimação, não estando sujeita aos mesmos erros dos métodos estatísticos ou geoestatísticos (D'ANTONA et al., 2013). Dessa forma, os dados pontuais da distribuição da população com base no CNEFE foram posteriormente agregados em planos celulares e utilizados como referência para avaliar os resultados obtidos com apoio dos demais dados auxiliares. No entanto, como a população pode não se distribuir de forma uniforme, menos de 57% dos endereços foram considerados válidos e os centróides dos setores para os quais não houve domicílios alocados foram definidos de forma arbitrária, ressalta-se que esse dado também não se trata de uma referência sobre a real distribuição da população, mas de outra estimativa.

### 3.5. Avaliação dos resultados e análise com superfícies multivariadas

A população agregada em cada célula de cada plano celular foi avaliada pelo valor do Coeficiente de Covariância (CoV) calculado entre a população estimada em cada célula e a célula correspondente nos dados de referência. Trata-se de um valor derivado do Erro Médio Quadrático (RMSE - *Root Mean Square Error*) utilizado em diferentes artigos relacionados (LANGFORD, 2013; JEGA et al., 2017) por se tratar de uma medida que permite a comparação entre resultados de diferentes resoluções. O CoV pode ser calculado como:

$$CoV = \frac{1}{\bar{r}} \sqrt{\frac{\sum_{c=1}^m (y_c - r_c)^2}{m}} \quad (3)$$

em que  $m$  é o número de células em um plano celular,  $y_c$  e  $r_c$  são os valores de população contidos na célula  $c$  para os dados sendo avaliados e os de referência, respectivamente, e  $\bar{r}$  é a média da população por célula no dado de referência. Por se tratar de uma medida de erro, quando menor o CoV, mais parecidos são os dados comparados.

A partir da análise dos resultados obtidos, novas superfícies adjacentes foram calculadas, utilizando dois ou mais dados auxiliares. Essas superfícies foram então utilizadas para redistribuir a população primeiro para planos matriciais de 10x10m e, na sequência, para os planos celulares propostos. Essas novas estimativas também foram avaliadas pelo cálculo da CoV entre os resultados obtidos e os dois dados de referência.

## 4. Resultados e discussões

Para avaliação inicial dos resultados, considerou-se que determinada superfície adjacente, para ser considerada efetiva na redistribuição da população na escala analisada, deve retornar Coeficientes de Covariância menores que o método ponderado pela área. Essa consideração se deve ao fato de que o método ponderado pela área não faz uso de dados auxiliares, enquanto o uso de superfícies adjacentes acaba por redistribuir a população de forma similar à ponderada pela área se não houver diferenças significativas nos valores de potencial de presença da população em um determinado setor censitário. Logo, uma superfície adjacente que retorna resultados menos acurados que o método ponderado pela área provavelmente não representa a distribuição da população esperada pelo registro dos dados de referência. Nesse sentido, ressalta-se que os dados de referência utilizados não se tratam de uma verdade sobre a distribuição da população e sim do que foi considerado neste estudo como a melhor aproximação disponível. Para complementar a avaliação, optou-se por também calcular o Coeficiente de Covariância entre os dados da Grade Estatística e os dados agregados a partir da redistribuição por domicílios dos dados do CNEFE. Os resultados de cada

superfície adjacente foram avaliados em função do plano celular utilizado, primeiro considerando apenas uma superfície por dado auxiliar utilizado e depois considerando superfícies adjacentes multivariadas, calculadas a partir de um ou mais dados auxiliares.

#### 4.1. Análise de superfícies adjacentes baseadas em um dado auxiliar

Os resultados estão apresentados do plano celular de maior para o de menor resolução espacial. Os valores de Coeficiente de Covariação calculados para a população agregada ao plano celular correspondente à Grade Estatística do IBGE estão apresentados na Tabela 3.

Tabela 3: Valores do Coeficiente de Covariância entre população redistribuída e dados de referência, agregados para o plano celular correspondente à Grade Estatística do IBGE, e comparação adicional entre dados de referência distribuídos para o mesmo plano celular.

Superfície adjacente utilizada	Coeficiente de Covariância	
	Grade Estatística	domicílios (CNEFE)
Distância a hidrografia	<b>2,54</b>	<b>3,58</b>
Distância a rodovias	2,68	<b>3,58</b>
Distância de centros urbanos	2,68	3,67
Distância a localidades	<b>2,67</b>	3,67
Distância a estabelecimentos de saúde	2,68	3,67
Distância a estabelecimentos de ensino	2,68	3,67
Distância a vertentes	2,76	3,76
Declividade média (%)	3,00	3,95
Uso e cobertura da terra (binário)	<b>2,55</b>	<b>3,59</b>
Uso e cobertura da terra (% de vizinhos não-habitados)	<b>2,53</b>	<b>3,58</b>
Uso e cobertura da terra (ponderado)	<b>2,47</b>	<b>3,54</b>
Ponderado pela área	2,68	3,67
Domicílios (CNEFE)	2,91	-

Nota: valores de Coeficiente de Covariação menores que o apresentado pelo método ponderado pela área estão destacados em negrito.

Para esse plano celular, as superfícies calculadas a partir da distância a hidrografia, distância a localidades e dados de uso e cobertura da terra apresentaram Coeficientes de Covariação menores que o método ponderado pela área ao se considerar a Grade Estatística como referência. Ao se considerar a agregação dos dados redistribuídos por domicílios como referência, as superfícies que apresentam os melhores resultados foram calculadas a partir de distância a hidrografia, distância a rodovias e dados de uso e cobertura da terra. Em ambos os casos, o menor Coeficiente de Varição foi obtidos ao utilizar os dados de uso e cobertura da terra de forma ponderada. Também é interessante observar que para dados agregados para esse plano celular, somente a superfície relacionada a declividade média retornou um Coeficiente de Covariação maior que a redistribuição por domicílios ao se adotar a Grade Estatística como referência, denotando que todos os demais dados auxiliares retornam resultados mais parecidos com a Grade Estatística que o uso dos dados do CNEFE, no qual a própria Grade Estatística se baseia. Isso provavelmente ocorre pelo fato de não se utilizar 43% dos dados originais do CNEFE, em função de problemas de geocodificação. Nesse nível de agregação, as suposições de que os domicílios não localizados devem se distribuir próximos aos localizados, que a população se concentra em um ponto no centróide do setor censitário quando este não apresenta domicílios, ou ainda que a população se distribua de forma uniforme por

domicílios em um dado setor censitário não devem ser válidas. Dados os resultados semelhantes entre as superfícies relacionadas a distâncias a centros urbanos, a localidades, a estabelecimentos de saúde e a estabelecimentos de ensino, acredita-se que a impossibilidade em atribuir coordenadas aos pontos do CNEFE não possui impactos significativos no cálculo das superfícies de densidade (distância a estabelecimentos de saúde e distância a estabelecimentos de ensino) se pelo menos um estabelecimento da espécie analisada tenha sido alocado nas proximidades da vila ou centro urbano correspondente.

Os valores de Coeficiente de Covariação calculados para a população agregada ao plano celular de 1x1km estão ilustrados na Tabela 4. Para esta análise, ressalta-se que apesar das células possuírem o mesmo tamanho de parte das células contidas na Grade Estatística do IBGE (1km), os limites entre as células não são coincidentes. Por esse motivo, a população agregada na Grade Estatística foi redistribuída para esse plano celular, pela soma ponderada pela área. Tendo em vista que mesmo na resolução espacial de 1km não se pode afirmar a homogeneidade da distribuição espacial da população por unidade de análise da Grade Estatística, essa alteração do dado de referência pode afastá-lo ainda mais de uma possível verdade de campo. Adicionalmente, a agregação da população alocada nos domicílios nos setores urbanos (200x200m no plano celular usado anteriormente) para células de 1x1km altera consideravelmente a configuração do dado de referência baseado na distribuição da população por domicílios, uma vez que todos esses domicílios tiveram sua localização determinada por geocodificação, do qual grande parte da informação não pode ser utilizada. Os maiores valores de Coeficiente de Variação apresentados no plano celular de 1x1km que os obtidos com uso do plano celular referente à Grade Estatística do IBGE endossam essas suposições, bem como problemas próprios de análises em áreas modificáveis.

Tabela 4: Valores do Coeficiente de Covariância entre população redistribuída e dados de referência, agregados para plano celular de 1x1km, e comparação adicional entre dados de referência distribuídos para o mesmo plano celular.

Superfície adjacente utilizada	Coeficiente de Covariância	
	Grade Estatística	domicílios (CNEFE)
Distância a hidrografia	<b>2,54</b>	<b>3,66</b>
Distância a rodovias	3,20	4,25
Distância de centros urbanos	3,20	4,25
Distância a localidades	3,20	4,25
Distância a estabelecimentos de saúde	3,20	<b>4,24</b>
Distância a estabelecimentos de ensino	3,20	4,25
Distância a vertentes	3,39	4,43
Declividade média (%)	<b>2,94</b>	<b>3,93</b>
Uso e cobertura da terra (binário)	<b>2,59</b>	<b>3,68</b>
Uso e cobertura da terra (% de vizinhos não-habitados)	<b>2,54</b>	<b>3,65</b>
Uso e cobertura da terra (ponderado)	<b>2,37</b>	<b>3,50</b>
Ponderado pela área	3,20	4,25
Domicílios (CNEFE)	2,57	-

Nota: valores de Coeficiente de Covariação menores que o apresentado pelo método ponderado pela área estão destacados em negrito.

De forma semelhante ao observado anteriormente, superfícies adjacentes baseadas em distâncias a centros urbanos, a localidades, a estabelecimentos de saúde e a estabelecimentos de ensino apresentaram valores similares de Coeficiente de Covariação, tanto na comparação com a agrega-

ção de dados da Grade Estatística quando da população distribuída por domicílios. Os valores apresentados por distância a rodovias também são similares a esses. As superfícies baseadas em distância a hidrografia, declividade média e dados de uso e cobertura da terra apresentaram resultados melhores que o ponderado pela área considerando os dois dados de referência, sendo os menores valores de Coeficiente de Covariância apresentados pelo uso de dados de uso e cobertura da terra de forma ponderada. Para essa análise, apenas as superfícies baseadas em distância a hidrografia e dados de uso e cobertura da terra nas abordagens % de vizinhos não-habitados e ponderada apresentaram Coeficientes de Variação menores que a agregação da população distribuídas por domicílios, ao se considerar a redistribuição da Grade Estatística como referência.

As células do planos celulares de 2x2km e 4x4km podem ser obtidas pela agregação das células de 1x1km. Os valores de Coeficientes de Covariância dos planos celulares de 2x2km e 4x4km estão apresentados nas Tabelas 5 e 6, respectivamente.

Tabela 5: Valores do Coeficiente de Covariância entre população redistribuída e dados de referência, agregados para plano celular de 2x2km, e comparação adicional entre dados de referência distribuídos para o mesmo plano celular.

Superfície adjacente utilizada	Coeficiente de Covariância	
	Grade Estatística	domicílios (CNEFE)
Distância a hidrografia	<b>1,40</b>	<b>1,87</b>
Distância a rodovias	1,68	2,13
Distância de centros urbanos	1,67	2,13
Distância a localidades	1,67	<b>2,12</b>
Distância a estabelecimentos de saúde	<b>1,66</b>	<b>2,12</b>
Distância a estabelecimentos de ensino	1,67	2,13
Distância a vertentes	1,80	2,24
Declividade média (%)	<b>1,63</b>	<b>2,01</b>
Uso e cobertura da terra (binário)	<b>1,39</b>	<b>1,82</b>
Uso e cobertura da terra (% de vizinhos não-habitados)	<b>1,37</b>	<b>1,81</b>
Uso e cobertura da terra (ponderado)	<b>1,23</b>	<b>1,67</b>
Ponderado pela área	1,67	2,13
Domicílios (CNEFE)	1,11	-

Nota: valores de Coeficiente de Covariância menores que o apresentado pelo método ponderado pela área estão destacados em negrito.

Para ambos os planos celulares, verificou-se que a similaridade entre os Coeficientes de Covariância se mantém para superfícies baseadas nas distâncias a rodovias, a centros urbanos, a localidades, a estabelecimentos de saúde e a estabelecimentos de ensino. Para o plano celular de 2x2km, as superfícies adjacentes baseadas em distância a hidrografia, distância a estabelecimentos de saúde, declividade média e dados de uso e cobertura da terra apresentaram Coeficientes de Covariância menores que ponderado pela área, para os dois dados de referência. Para o plano celular de 4x4km, todas as superfícies adjacentes retornaram valores menores ou similares de Coeficiente de Covariância que o ponderado pela área. Adicionalmente, para esses dois planos celulares observa-se uma melhor correspondência entre a agregação da população distribuída por domicílios e os dados agregados da Grade Estatística do que o uso dos demais dados auxiliares.

Tabela 6: Valores do Coeficiente de Covariância entre população redistribuída e dados de referência, agregados para plano celular de 4x4km, e comparação adicional entre dados de referência distribuídos para o mesmo plano celular.

Superfície adjacente utilizada	Coeficiente de Covariância	
	Grade Estatística	domicílios (CNEFE)
Distância a hidrografia	<b>0,69</b>	<b>0,87</b>
Distância a rodovias	0,81	1,01
Distância de centros urbanos	<b>0,80</b>	<b>1,00</b>
Distância a localidades	<b>0,79</b>	<b>1,00</b>
Distância a estabelecimentos de saúde	<b>0,80</b>	<b>1,00</b>
Distância a estabelecimentos de ensino	<b>0,80</b>	<b>1,00</b>
Distância a vertentes	<b>0,80</b>	<b>0,98</b>
Declividade média (%)	<b>0,76</b>	<b>0,88</b>
Uso e cobertura da terra (binário)	<b>0,66</b>	<b>0,84</b>
Uso e cobertura da terra (% de vizinhos não-habitados)	<b>0,65</b>	<b>0,84</b>
Uso e cobertura da terra (ponderado)	<b>0,60</b>	<b>0,79</b>
Ponderado pela área	0,81	1,01
Domicílios (CNEFE)	0,54	-

Nota: valores de Coeficiente de Covariância menores que o apresentado pelo método ponderado pela área estão destacados em negrito.

Para as superfícies adjacentes analisadas, verificou-se que a baseada em dados de uso e cobertura da terra retornam os resultados mais parecidos com as referências utilizadas. Dados similares têm se sobressaído para a redistribuição da população, com bons resultados obtidos em estudos como Zandbergen e Ignizio (2010), Langford (2013) e Jega et al. (2017). Na Figura 3 ilustra-se um recorte dos dados da população agregada em setores censitários e da redistribuída para plano matricial de 10x10m com uso de dados de uso e cobertura da terra na abordagem ponderada, avaliados como os mais próximos dos dados de referência para todos os planos celulares considerados. Para efeito de análise, esses dados estão apresentados em termos de densidade populacional (habitantes/km<sup>2</sup>).

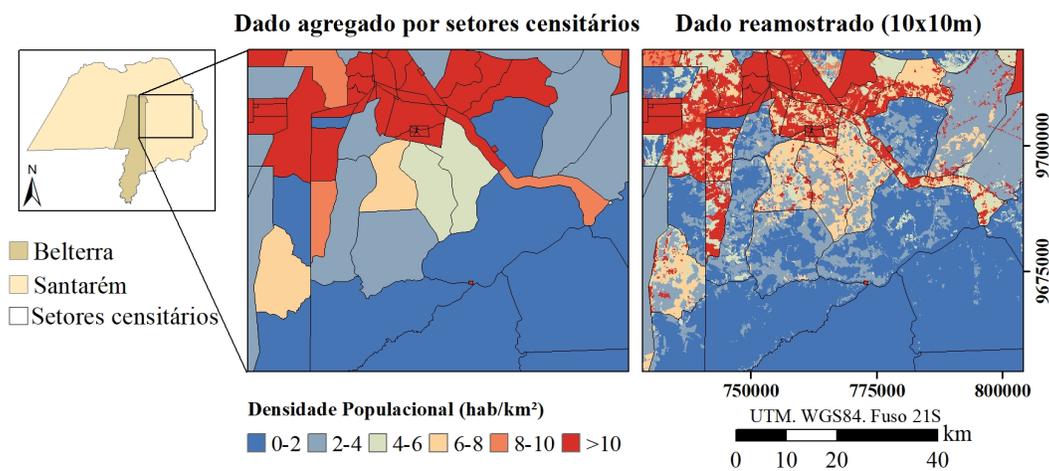


Figura 3: Moradores em domicílios particulares permanentes ou população residente em domicílios particulares permanentes. Informação agregada em setores censitários e redistribuída para plano matricial de 10x10m por meio da superfície adjacente calculada sobre dados de uso e cobertura da terra (ponderado).

Duas características são notáveis ao se comparar os dados apresentados na Figura 3: a primeira é que variações abruptas na densidade populacional ocorrem nos limites dos setores censitários, tanto nos dados agregados por setor, como era o esperado, quanto nos dados redistribuídos para o plano matricial de 10x10m. A segunda característica observada é que é possível, com o uso do dado auxiliar proposto, estimar parte da heterogeneidade da distribuição da população no interior dos setores, sendo que em alguns casos as variações abruptas entre setores são minimizadas.

Os dados de referência agregados pelo plano celular da Grade Estatística do IBGE estão ilustrados na Figura 4. Nesta figura também são apresentados os planos celulares com a informação da população redistribuída com dados de uso e cobertura da terra na abordagem ponderada, para comparação.

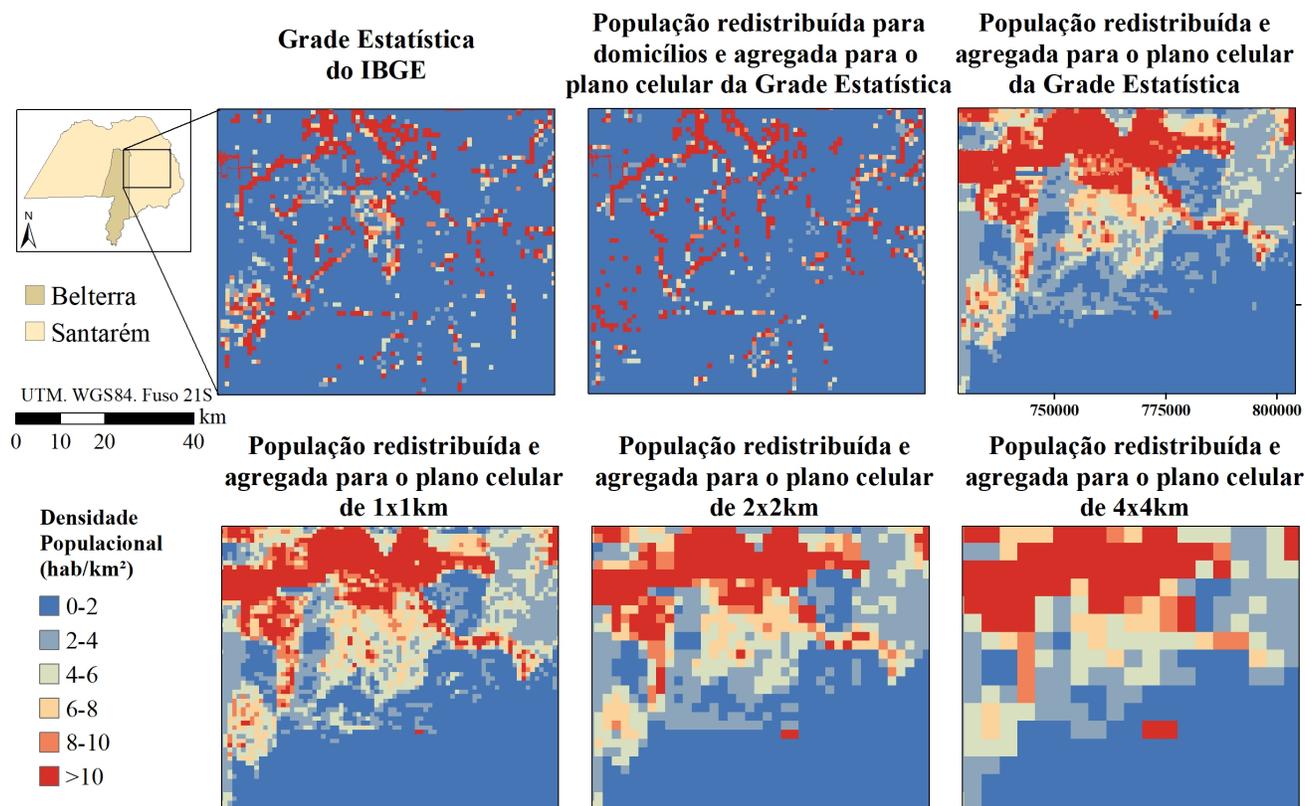


Figura 4: Moradores em domicílios particulares permanentes ou população residente em domicílios particulares permanentes. Dados de referência (Grade Estatística do IBGE e população redistribuída por domicílios) e dados redistribuídos por superfície adjacente calculada sobre dados de uso e cobertura da terra (ponderado) e agregados para diferentes planos celulares.

É perceptível que a impossibilidade de se utilizar cerca de 43% dos dados disponíveis no CNEFE para redistribuir a população causa diferenças nos dados de população agregados no plano celular da Grade Estatística, embora parte da heterogeneidade da distribuição da população seja recuperável com o uso desses dados. Nesse sentido, como se supõe homogeneidade na distribuição da população de um mesmo setor censitário em uma mesma classe de uso e cobertura da terra, não há variação interna da distribuição da população nessas áreas ao se considerar os dados redistribuídos com auxílio de dados de uso e cobertura da terra e todos os planos celulares, o que se observa em função de manchas contínuas com a mesma densidade populacional. Esses resultados corroboram, de forma visual, que uma fraqueza desses métodos é ignorar diferenças internas em uma classe

específica (SU et al., 2010). Uma maneira de sobrepujar essa fraqueza seria ponderar diferenças internas com uso de imagens de sensoriamento remoto de melhor resolução espacial ou outros dados auxiliares capazes de indicar a distribuição da população interna aos setores censitários. Para verificar se com as variáveis propostas neste estudo é possível melhorar a distribuição interna da população por classe de uso e cobertura da terra e setor censitário, superfícies adjacentes multivariadas (que utilizam mais de um dado auxiliar) foram calculadas, conforme apresentadas na próxima subseção.

#### 4.2. Análise de superfícies adjacentes baseadas em múltiplos dados auxiliares

Considerando os resultados apresentados na subseção anterior, os dados de população agregados em planos celulares mais similares às referências são aqueles baseados em dados de uso e cobertura da terra, ordenados em abordagens ponderada, % de vizinhos habitados e binária, e distância a hidrografia. Para os planos celulares que não o baseado na Grade Estatística do IBGE, também se observa bons resultados para uso de declividade média. Para analisar se o uso conjunto dos dados proporciona resultados da distribuição da população mais parecidos com as referências, superfícies adjacentes representando cada variável foram utilizadas para calcular uma nova superfície. Como declividade média e distância a vertentes se relacionam ao relevo, dados de uso e cobertura são instâncias do mesmo dado e distâncias a localidades, a centros urbanos, a estabelecimentos de ensino/saúde se relacionam à distância a locais de provável adensamento humano, provendo resultados similares a distância a rodovias, os seguintes conjuntos de dados foram utilizados:

- C1: distância a hidrografia, declividade média e uso e cobertura da terra (ponderada);
- C2: distância a hidrografia, declividade média, distância a localidades e uso e cobertura da terra (% de vizinhos não habitados).

O conjunto C1 representa dados que consistentemente apresentaram bons resultados, sendo que do conjunto de dados de uso e cobertura da terra, selecionou-se o que retornou os resultados mais próximos da referência. Os dados do conjunto C2 foram selecionados por, além de também representarem diferentes variáveis relacionadas à distribuição da população, possuírem processamento de menor custo. Nesse sentido, ressalta-se que, para a Amazônia Legal brasileira, os dados originais de hidrografia e localidades são disponibilizados pelo IBGE, declividade pode ser calculada diretamente a partir de dados SRTM livremente distribuídos e a informação de áreas não habitadas pode ser derivada de dados PRODES.

Seguindo a abordagem inicial utilizada por Amaral et al. (2012), as superfícies adjacentes multivariadas foram calculadas por meio de média simples. Um segundo teste foi realizado considerando o produto entre as variáveis. Os resultados dessas superfícies adjacentes foram avaliados seguindo a mesma metodologia dos resultados de cada dado usado de forma individual. Os valores de Coeficiente de Covariância entre esses resultados e os dados de referência estão apresentados na Tabela 7.

Observa-se que tanto a média simples quanto o produto das superfícies adjacentes dos conjuntos selecionados não retornaram melhores resultados que a superfície adjacente obtida a partir dos dados de uso e cobertura da terra (com abordagem ponderada). Adicionalmente, as superfícies adjacentes obtidas pelo produto de C1 e C2 retornaram valores de Coeficiente de Covariância maiores que o método ponderado pela área ao se agregar a população no plano celular relativo à Grade do IBGE (200x200m nas áreas urbanas e 1x1km nas áreas rurais). Uma possível explicação para o ocorrido reside em o uso e cobertura da terra se tratar de uma variável que não apenas

aumenta o potencial de presença de população, mas também de um registro deixado no terreno por atividades antrópicas, sejam os agentes dessas atividades a população residente ou não. No entanto, salienta-se novamente que os resultados aqui apresentados foram avaliados em comparação a outros dados estimados, sendo possível que as superfícies multivariadas resultem em uma população distribuída em um padrão semelhante ao real, cuja avaliação não foi possível.

Tabela 7: Valores do Coeficiente de Covariância entre população redistribuída usando superfícies adjacentes multivariadas e dados de referência.

Plano celular	Superfície adjacente	Coeficiente de Covariância	
		Grade Estatística	domicílios (CNEFE)
Grade Estatística do IBGE	C1 (média)	2,51	3,57
	C1 (produto)	2,84	3,83
	C2 (média)	2,57	3,60
	C2 (produto)	2,92	3,90
1x1km	C1 (média)	2,26	3,60
	C1 (produto)	2,54	3,57
	C2 (média)	2,73	3,84
	C2 (produto)	2,66	3,64
2x2km	C1 (média)	1,36	1,82
	C1 (produto)	1,35	1,71
	C2 (média)	1,49	1,94
	C2 (produto)	1,43	1,80
4x4km	C1 (média)	0,67	0,85
	C1 (produto)	0,68	0,80
	C2 (média)	0,71	0,90
	C2 (produto)	0,68	0,80

## 5. Conclusões

Este trabalho objetivou avaliar dados auxiliares relacionados a diferentes variáveis indicadoras da presença da população quanto ao potencial de redistribuição dos dados de número de moradores em domicílios particulares permanentes ou população residente em domicílios particulares permanentes dos municípios de Santarém e Belterra para o Censo de 2010, agregados em setores censitários, para zonas alvo de diferentes tamanhos. Foram utilizados 11 dados auxiliares (incluindo o mesmo dado com mais de uma abordagem), avaliados de forma individual e conjunta, para redistribuir a população em 4 planos celulares. Esses são: aquele derivado da Grade Estatística do IBGE (200x200m em setores urbanos e 1x1km em setores rurais), plano celular de 1x1km, plano celular de 2x2km e plano celular de 4x4km. Os resultados foram analisados pela comparação destes com a Grade Estatística do IBGE (informação de população total) e por dados redistribuídos por domicílios, contabilizados a partir do CNEFE, ambos agregados nos respectivos planos celulares para comparação.

Para todos os tamanhos de células, dados de uso e cobertura da terra retornaram os resultados mais próximos às referências. Das três abordagens utilizadas, a que retorna os resultados mais similares é aquela em que se atribui valores distintos de potencial da presença da população para grupos de classes de uso e cobertura, para as quais se supõe homogeneidade interna da distribuição da população contida também em um mesmo setor censitário. Esse dado retornou resultados mais

próximos das referências, inclusive, que abordagens multivariadas, em que esse e outros dados são utilizados conjuntamente.

Observou-se que com o aumento do tamanho das células, os resultados e as referências (agregadas no mesmo plano celular para análise) diferem menos. Especialmente para células de 4x4km, são observados valores de Coeficiente de Covariação próximos aos encontrados por Langford (2013) para Cardiff no Reino Unido (entre 0,410 e 0,731) e por Jega et al. (2017) para Leicester, também no Reino Unido (entre 0,460 e 0,635), que se tratam de regiões onde, teoricamente, a estimativa da população por meio de dados auxiliares é mais simples que na Amazônia brasileira. Nesse sentido, cabe ao analista definir quais são a resolução e a exatidão necessárias para análise dos dados, e qual o melhor compromisso entre esses fatores.

No entanto, o fato de as próprias referências tratarem-se de estimativas limita a análise dos resultados obtidos, uma vez que não há informações disponíveis sobre a distribuição real da população de 2010 dos municípios analisados. No entanto, as estimativas utilizadas como referências são aceitas como indicadores da distribuição da população em escala mais refinada que a dos setores censitários. Nesse sentido, destaca-se que os dados utilizados como referência (Grade Estatística e CNEFE) não estão disponíveis para dados do Censo anteriores a 2010. Ou seja, para análises que consideram outros anos de Censo e que prescindam da informação da população redistribuída, é importante saber quais dados auxiliares utilizar. Em particular, dados de uso e cobertura da terra podem ser obtidos a partir de imagens de sensoriamento remoto, disponibilizadas para vários anos a partir da década de 70.

## Referências

AFFONSO, A. G.; ESCADA, M. I. S.; AMARAL, S.; SOUZA, A. R.; SIQUEIRA, J. M.; TORRES, N. C.; CAMILOTTI, V. L.; DAL'ASTA, A. P.; COSTA, L. C. O.; SOARES, F. R. **As comunidades ribeirinhas do baixo Tapajós(PA): infraestrutura, mobilidade, serviços sócio ambientais e conectividade**: Relatório técnico de atividade de campo - estudo das Trajetórias de Padrões e Processos na caracterização das Dinâmicas do Desmatamento na Amazônia. subprojeto 5 do Monitoramento Ambiental por Satélites no Bioma Amazônia. São José dos Campos: INPE, 2016. 153p.

ALMEIDA, C.; COUTINHO, A.; ESQUERDO, J.; ADAMI, M.; VENTURIERI, A.; DINIZ, C.; DESSAY, N.; DURIEUX, L.; GOMES, A. High spatial resolution land use and land cover mapping of the Brazilian Legal Amazon in 2008 using Landsat-5/TM and MODIS data. **Acta Amazonica**, v. 46, n. 3, p. 291–302, 2016.

AMARAL, S.; GAVLAK, A. A.; ESCADA, M. I. S.; MONTEIRO, A. M. V. Using remote sensing and census tract data to improve representation of population spatial distribution: case studies in the Brazilian Amazon. **Population and Environment**, v. 34, n. 1, p. 142–170, 2012.

BECKER, B. **Amazônia**. 2. ed. São Paulo: Ática, 1991. 112 p.

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

BRASIL. Decreto-lei no 1.106, de 16.06.1970. Cria o Programa de Integração Nacional, altera a legislação do imposto de renda das pessoas jurídicas na parte referente a incentivos fiscais e dá outras providências. **Jus Brasil, Legislação**, 1970.

- BRIGGS, D. J.; GULLIVER, J.; FECHT, D.; VIENNEAU, D. M. Dasymetric modelling of small-area population distribution using land cover and light emissions data. **Remote Sensing of Environment**, v. 108, n. 4, p. 451 – 466, 2007. ISSN 0034-4257.
- CARDOSO, A. C. D. **O rural e o urbano na Amazônia: diferentes olhares em perspectivas**. Belém: EDUFPA, 2010. 4807–4816 p. 215p.
- CÔRTEZ, J. C.; D'ANTONA, Á. O. Dinâmicas no uso e cobertura da terra: perspectivas e desafios da demografia. **Revista Brasileira de Estudos de População**, v. 31, n. 1, p. 191–210, 2014.
- DAL'ASTA, A.; GAVLAK, A.; ESCADA, M.; BRIGATTI, N.; AMARAL, S. **Núcleos de ocupação humana e usos da terra entre Santarém e novo Progresso, ao longo da BR-163 (PA): Relatório técnico de atividade de campo**. São José dos Campos: INPE, 2011. 64p.
- DAL'ASTA, A. P. **Representações do fenômeno urbano na Amazônia contemporânea: observações no sudoeste do Pará**. 181 p. Dissertação (Mestrado) — Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2016.
- DAL'ASTA, A. P.; SOUZA, A. R.; PINHO, C. M. D.; SOARES, F. R.; REGO, G. F. J.; SIQUEIRA, J. M.; ESCADA, M. I. S.; BRIGATTI, N.; AMARAL, S.; CAMILOTTI, V. L.; DÓRIA, V. E. M.; COSTA, L. C. **As comunidades de terra firme do sudoeste do Pará: população, infraestrutura, serviços, uso da terra e conectividades. Expedição de campo 2013, versão atualizada: Relatório técnico de atividade de campo - projeto URBISAMAZÔNIA/ FUNDAÇÃO VALE e projeto LUA-IAM/FAPESP**. São José dos Campos: INPE, 2017. 211p.
- D'ANTONA, A. O.; BUENO, M. C. D.; DAGNINO, R. S. Estimativa da população em unidades de conservação na amazônia legal brasileira – uma aplicação de grades regulares a partir da contagem 2007. **Revista brasileira de Estudos da População**, v. 30, n. 2, p. 401–428, 2013.
- DMOWSKA, A.; STEPINSKI, T. F. A high resolution population grid for the conterminous united states: the 2010 edition. **Computers, Environment and Urban Systems**, v. 61, p. 13–23, 2017.
- ESCADA, M.; AMARAL, S.; RENNÓ, C.; PINHEIRO, T. **Levantamento do Uso e Cobertura da Terra e da Rede de Infraestrutura no Distrito Florestal da BR-163**. São José dos Campos: INPE, 2009. 52p.
- EVANS, T. P.; MORAN, E. F. Spatial integration of social and biophysical factors related to landcover change. **Population and Development Review**, v. 28, p. 165–186, 2002.
- FLOWERDEW, R. How serious is the modifiable areal unit problem for analysis of english census data? **Population trends**, v. 145, n. 1, p. 106–118, 2011.
- FLOWERDEW, R.; GREEN, M. Areal interpolation and types of data. **Spatial analysis and GIS**, London: Talyor and Francis, v. 121, p. 145, 1994.
- GAVLAK, A. A. **Padrões de mudança de cobertura da terra e dinâmica populacional no Distrito Florestal Sustentável da BR-163: população, espaço e ambiente**. 155 p. Dissertação (Mestrado) — Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2011.

GOODCHILD, M.; LAM, N. Areal interpolation: a variant of the traditional spatial problem. **Geo-Processing**, v. 1, p. 297–312, 1980.

GOODCHILD, M. F.; ANSELIN, L.; DEICHMANN, U. A framework for the areal interpolation of socioeconomic data. **Environment and planning A**, v. 25, n. 3, p. 383–397, 1993.

HOGAN, D. J. Demographic dynamics and environmental change in Brazil. **Ambiente & Sociedade**, n. 9, p. 43–73, 2001.

HOLT, J.; LO, C.; HODLER, T. Dasymetric estimation of population density and areal interpolation of census data. **Cartography and Geographic Information Science**, v. 31, n. 2, p. 103–121, 2004.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Regiões de Influência das Cidades**, IBGE, Rio de Janeiro, 2008.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Dados base 1:250000 para Amazônia Legal brasileira**, 2009. Disponível em: <<ftp://geoftp.ibge.gov.br/>>. Acesso em: 09 set. 2014.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Censo 2010**, 2010. Disponível em: <<http://censo2010.ibge.gov.br/>>. Acesso em: 09 set. 2014.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Metodologia do Censo Demográfico 2010**, Rio de Janeiro, 2013. 712p.

IBGE. Instituto Brasileiro de Geografia e Estatística. **IBGE cidades**, 2017. Disponível em: <<http://http://cidades.ibge.gov.br/xtras/home.php>>. Acesso em: 09 set. 2014.

JEGA, I.; COMBER, A.; TATE, N. A comparison of methods for spatial interpolation across different spatial scales. **SSRG International Journal of Geo informatics and Geological Science (SSRG-IJGGS)**, v. 4, n. 3, p. 12–22, 2017.

KAMPEL, S. A. **Geoinformação para estudos demográficos: representação espacial de dados de população na Amazônia brasileira**. 166 p. Tese (Doutorado) — Escola Politécnica da Universidade de São Paulo, São Paulo, 2003.

LAM, N. S. Spatial interpolation methods: a review. **The American Cartographer**, Taylor & Francis, v. 10, n. 2, p. 129–150, 1983.

LANGFORD, M. An evaluation of small area population estimation techniques using open access ancillary data. **Geographical Analysis**, Wiley Online Library, v. 45, n. 3, p. 324–344, 2013.

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

MENNIS, J.; HULTGREN, T. Intelligent dasymetric mapping and its application to areal interpolation. **Cartography and Geographic Information Science**, v. 33, n. 3, p. 179–194, 2006.

OPENSHAW, S. **The Modifiable Areal Unit Problem**. **CATMOG n. 38**. Norwich: Geo Books, 1984. 41 p.

PINHEIRO, T.; RENNÓ, C. D.; ESCADA, M. I. S. Mapeamento de ambientes de terra firme no Distrito Florestal Sustentável da BR-163 utilizando o descritor de terreno HAND. **Relatório de Pesquisa**, Belém:INPE, 2016. 12p.

PREFEITURA DE SANTARÉM. **Santarém**. [s.n.], s.d. Disponível em: <<http://www.santarem.pa.gov.br/>>. Acesso em: 09 set. 2014.

RENNÓ, C. D.; NOBRE, A. D.; CUARTAS, L. A.; SOARES, J. V.; HODNETT, M. G.; TOMASELLA, J.; WATERLOO, M. J. **Remote Sensing of Environment**, v. 112, p. 3469–3481, 2008.

SAATY, T. L. **The analytic hierarchy process**. New York: McGraw-Hill, 1980.

SU, M.; LIN, M.; HSIEH, H.; TSAI, B.; LIN, C. Multi-layer multi-class dasymetric mapping to estimate population distribution. **Science of the total environment**, v. 208, p. 4807–4816, 2010.

TOBLER, W. Smooth pycnophylactic interpolation for geographical regions. **Journal of the American Statistical Association**, v. 74, n. 367, p. 519–530, 1979.

WRIGHT, J. A method of mapping densities of population with cape cod as an example. **Geographical Review**, v. 26, p. 103–111, 1936.

YANG, X.; YUE, W.; GAO, D. Spatial improvement of human population distribution based on multi-sensor remote-sensing data: an input for exposure assessment. **International journal of remote sensing**, v. 34, n. 15, p. 5569–5583, 2013.

ZADEH, L. A. Fuzzy logic. **Computer**, v. 21, p. 83–92, 1988.

ZANDBERGEN, P. A.; IGNIZIO, D. A. Comparison of dasymetric mapping techniques for small-area population estimates. **Cartography and Geographic Information Science**, v. 37, n. 3, p. 199–214, 2010.