



Em busca de novas representações demográficas: O campo de estudos das grades populacionais em tempos de máquinas que aprendem*

Diego Moreira Silva**
Tathiane Mayumi Anazawa***
Silvana Amaral Kampel****
Flávia da Fonseca Feitosa*****
José Irineu Rangel Rigotti*****
Antonio Miguel Vieira Monteiro*****

A distribuição populacional na superfície terrestre revela uma variedade de padrões espaciais que refletem processos sociodemográficos relacionados às dinâmicas histórico-geográficas que os produziram. As grades populacionais têm ganhado destaque como fonte de dados, envolvendo estimativas e distribuição de população em pequenas áreas. Cada grade populacional é composta por células de tamanhos, cobrindo a totalidade do globo ou territórios locais. Este trabalho apresenta uma revisão comentada da literatura no campo destas representações populacionais, especificamente na distribuição e volume da população, bem como na importância das variáveis auxiliares de cunho espacial, denominadas aqui de variáveis sintomáticas. Estas desempenham papel crucial na construção de modelos pautados na realidade, tanto local quanto global, utilizando métodos diversos, que incluem aprendizado de máquina. Destacam-se ainda as principais iniciativas do campo, os produtos globais disponíveis e as bases técnicas das principais metodologias. Além disso, são discutidas as limitações, precauções e as novas oportunidades resultantes da criação dessas grades populacionais.

Palavras-chave: Grades populacionais. Distribuição espacial da população. Aprendizado automático. Modelado espacial. Variáveis sintomáticas. Sensores remotos. Dados censitários.

* O artigo se baseia na dissertação de mestrado de Silva (2023).

** Instituto Nacional de Pesquisas Espaciais (Inpe), São José dos Campos-SP, Brasil (silvamoreiradiego0@gmail.com); <https://orcid.org/0000-0001-7670-4408>).

*** Universidade Estadual de Campinas (Unicamp), Campinas-SP, Brasil (tathiane@unicamp.br); <https://orcid.org/0000-0003-2675-0566>).

**** Instituto Nacional de Pesquisas Espaciais (Inpe), São José dos Campos-SP, Brasil (silvana.amaral@inpe.br); <https://orcid.org/0000-0003-4314-7291>).

***** Universidade Federal do ABC (UFABC), Santo André-SP, Brasil (flavia.feitosa@ufabc.edu.br); <https://orcid.org/0000-0001-7744-5225>).

***** Universidade Federal de Minas Gerais (UFMG), Belo Horizonte-MG, Brasil (rigotti@cedeplar.ufmg.br); <https://orcid.org/0000-0003-1749-5313>).

***** Instituto Nacional de Pesquisas Espaciais (Inpe), São José dos Campos-SP, Brasil (miguel.monteiro@inpe.br); <https://orcid.org/0000-0003-1477-1749>).

Introdução

A distribuição da população na superfície terrestre exhibe uma variedade de padrões espaciais, cujo entendimento é essencial para o desenvolvimento de políticas que visem combater as desigualdades socioterritoriais. No campo da demografia, os padrões espaciais de distribuição da população refletem arranjos e configurações de processos sociais e demográficos associados com as dinâmicas histórico-geográficas que os produziram. Portanto, avançar na produção de representações das distribuições espaciais de variáveis sociodemográficas e nas possibilidades metodológicas para sua estimativa em pequenas áreas é relevante para instrumentalizar os estudos de população e o combate às desigualdades.

Um tipo particular de representação da distribuição espacial de população é conhecido como grade populacional. Cada grade populacional é composta por células de tamanho específico (resolução espacial) que cobrem a totalidade do globo (grades globais) ou um determinado território de interesse (grades nacionais, regionais ou locais). Cada célula de uma grade contém informações relativas à população, como o número de habitantes na célula, densidade populacional e outros dados associados à caracterização demográfica daquela célula (Deichmann; Balk; Yetman, 2001; Tobler, 1979; Tobler *et al.*, 1995, 1997; Yetman; Deichmann; Balk, 2000).

A produção e disseminação de dados de população em grades regulares oferecem novas perspectivas para análises, não exclusivas ao campo dos estudos de população, mas também em estudos integrados abrangendo população, planejamento territorial e estratégias de desenvolvimento com ênfase na justiça social e responsabilidade ambiental (Laituri *et al.*, 2021; ONU-Habitat, 2015, 2017). As grades permitem análises independentes das divisões territoriais político-administrativas e das unidades espaciais definidas para a coleta de informações. Assim, possibilitam a representação e o acesso a dados demográficos em unidades geográficas com tamanhos adequados às condições das informações disponíveis, mantendo sua estabilidade ao longo do tempo. Isso reduz o problema das unidades de área modificáveis (*Modifiable Areal Unit Problem – Maup*) (Openshaw, 1984) nas análises, facilitando tanto estudos temporais quanto comparativos em âmbitos internacional, nacional e em outras escalas. Estas características das grades contribuem para aprimorar a qualidade da localização e caracterização das áreas ocupadas por populações humanas. Entretanto, a construção destas grades traz novos desafios técnicos e analíticos ao campo dos estudos de população. A produção sistemática de grades globais e sua distribuição facilitada pelo acesso aberto, via internet, ampliou muito o uso de dados de população em estudos locais e regionais, para os quais não há dados disponíveis ou períodos intercensitários mais longos.

Apesar dos esforços empreendidos na construção destas grades, trabalhos recentes têm mostrado um conjunto de problemas quando se faz uso direto das grades sem uma avaliação de sua qualidade para contextos específicos (Leyk *et al.*, 2019; Hierink *et al.*, 2022; Kuffer *et al.*, 2024; Thomson *et al.*, 2022). Este trabalho procura, primeiro, revelar

as inovações técnicas presentes na construção de grades populacionais globais, destacando as principais metodologias desenvolvidas nas últimas décadas, sobretudo o uso de métodos baseados em aprendizado por máquina, e, segundo, apresentar um quadro de experiências de produção de grades locais no Brasil. Com isso, podemos discutir as limitações, as oportunidades e os riscos do uso de grades globais sem que haja avaliação de sua adequação para contextos histórico-geográficos específicos. Em particular, à luz da compreensão destas inovações técnicas, destacamos como as bases de informações populacionais e as informações auxiliares produzidas localmente ganham relevância para os ajustes e correções necessários às bases disponíveis.

Nesse contexto, apresentamos um conjunto de iniciativas globais que têm avançado com inovações em métodos e metodologias para estimativas, assim como na (re)distribuição e disseminação de dados de distribuição espacial da população por meio de grades que abrangem todo o planeta. A história da produção e disseminação destas grades em escala global remonta à segunda metade dos anos 1990. A iniciativa começou nos Estados Unidos, focando na formação de cooperações e parcerias acadêmicas e incorporando um arranjo destinado a fortalecer a administração institucional da agenda proposta. Esse desenho institucional foi posteriormente introduzido na Europa nos anos 2000, com o *Consortium for International Earth Science Information Network* (Ciesin) da Columbia University, que desempenha um papel fundamental na expansão da rede. Durante esse período, a rede foi ampliada para incluir o *Joint Research Centre* (JRC) e também envolveu pesquisas em colaboração com o Banco Mundial (Balk *et al.*, 2005, 2006).

Estas iniciativas conseguiram inspirar e mobilizar a agenda para a construção de informação populacional desagregada em grades. Vários grupos em todo o mundo contribuíram para os avanços tecnológicos e metodológicos de novas ferramentas, resultando no refinamento da resolução espacial das grades e na melhoria da qualidade de suas estimativas. Dados de fontes diversas são utilizados, tais como arquivos globais de imagens de satélites de alta resolução espacial, dados censitários e informações geográficas voluntárias.¹ Com os novos dados e bases, surgiram novos métodos de classificação e regressão, utilizando tecnologias de mineração de dados espaciais e algoritmos de aprendizado por máquina (*machine learning*) capazes de lidar com grandes volumes de informação mais apropriadamente (Stevens *et al.*, 2015; Zhou; Wu; Tang, 2002). Em 2020, estas iniciativas, junto com as novas que surgiram ao longo destes anos, estabeleceram uma plataforma colaborativa que mantém grande parte da produção destes dados em escala global, conhecida como *POPGRID Data Collaborative*² (POPGRID, 2020).

As iniciativas globais de produção de grades populacionais desempenham um papel importante na oferta de fontes de dados para países de renda média e baixa que carecem de sistemas de informação estatística e demográfica oficiais adequados (Leyk *et al.*, 2019).

¹ Informações coletadas voluntariamente por cidadãos e cidadãs. Ajudam a construir bases de dados apoiados em colaboração. Um exemplo destas bases é a plataforma aberta *OpenStreetMap* (www.openstreetmap.org).

² O POPGRID pode ser acessado em: <https://popgrid.org/>.

Essas iniciativas globais levam em consideração as informações obtidas nos contextos nacionais dos países, quando essas informações existem.

Vale ressaltar o papel primordial dos censos demográficos nacionais, como adverte o conteúdo presente no *Socioeconomic Data and Applications Center – Sedac/Ciesin*³ e no guia de boas práticas estabelecido no contexto da iniciativa POPGRID (2020). A informação dos censos demográficos é a mais qualificada para trazer os componentes das dinâmicas locais relativas à distribuição de suas populações e do estabelecimento de seus assentamentos. Por isso, torna-se a informação fundamental para a qualificação dos produtos a serem gerados. Esse alerta dos principais produtores de grades globais reforça o que afirma Feitosa (2020, p. 5): “é fundamental, portanto, um debate sobre o papel dos levantamentos tradicionais na era do *big data* e *urban analytics* [...] Não se pode admitir que o entusiasmo com o *big data* reforce discursos e ações que depreciam e degeneram a produção de dados sobre o país”.

Quando se deseja uma melhor resolução espacial para obter informações populacionais de pequenas áreas, os modelos de estimativa e redistribuição das grades globais incorporam, implicitamente, elementos de decisão da localização de grupos e, até mesmo, de indivíduos, que são representados nas variáveis (não populacionais) utilizadas nos modelos propostos. Estas variáveis, derivadas de bases de dados globais, nem sempre são as melhores opções para capturar a diversidade intrínseca dos processos que influenciam essas decisões nas diversas regiões do mundo. Isso é compreensível e faz sentido, dada a orientação específica desses produtos.

Entretanto, quando as escalas são relativas às extensões geográfico-administrativas subnacionais, como regiões, estados, microrregiões, municípios e até mesmo distritos, vilas e aglomerados, as grades necessitam de maior acurácia e resoluções cada vez mais detalhadas. Nesse contexto, torna-se vital a seleção criteriosa de variáveis auxiliares que atuem como *proxies*,⁴ também conhecidas por variáveis sintomáticas (Álvarez, 2001; Cabrera, 2011; Campos *et al.*, 2020; Esquivel, 2001), indicativas de evidência de presença humana, que sejam mais adequadas para modelos de distribuição populacional e incorporem conhecimento sobre as diferentes realidades territoriais em que estas populações estão inseridas. Dessa forma, uma maior compreensão dos diferentes aspectos relacionados aos processos histórico-geográficos da ocupação territorial por estas populações em tão diversas geografias passa a ser essencial para a seleção de variáveis sintomáticas.

Os indicadores quantitativos populacionais, como volumes, densidade e taxas, desempenham papel fundamental para o desenho de políticas de enfrentamento das desigualdades socioterritoriais (Nadalin, 1994; Shryock; Siegel; Larmon, 1975). Por isso, estimativas confiáveis e representativas da distribuição espacial em grades populacionais

³ Acesso do Ciesin pode ser realizado em: <https://sedac.ciesin.columbia.edu/>.

⁴ Uma *proxy*, ou variável *proxy*, ou ainda variável substituta no contexto de modelos estatísticos é aquela que pode substituir a variável de interesse, aquela fundamental ao modelo escolhido, porque esta variável pode não estar disponível, por ser muito cara ou muito demorada para se medir. Fazer estimativas sem a variável fundamental ao modelo pode causar um viés de variável omitida. Para evitar isso procura-se por uma variável que possa substituí-la (Bruc; Bruce, 2019).

na escala mais adequada tornam-se instrumentos valiosos para revelar parcelas invisíveis das populações que lutam e resistem para manter seus espaços, respeitando suas particularidades. É nesse contexto que se torna necessário e urgente compreender e experimentar os avanços metodológicos introduzidos nesse campo, adaptando as novas metodologias às capacidades de produção de grades populacionais em escalas locais.

Este trabalho procura contribuir para o conhecimento, uso bem-informado e crítico de bases de dados relativamente pouco exploradas nos estudos demográficos e de planejamento urbano-regional no Brasil, a partir de revisão da literatura que avaliou as produções mais recentes no campo das grades populacionais. A análise abrangeu materiais bibliográficos provenientes de diferentes iniciativas em grades populacionais, incluindo suas descrições e metodologias empregadas, com foco especial em métodos recentemente utilizados, como a aprendizagem por máquina.

Inicialmente, são apresentados os produtos resultantes de iniciativas globais e locais de grades populacionais, abordando diferentes escalas de representação, métodos utilizados, formas de representação dos volumes populacionais e as possibilidades e limitações inerentes a cada abordagem metodológica. Em seguida, são destacados os esforços dedicados à construção de grades populacionais locais no Brasil, assim como os dados frequentemente utilizados para sua produção, incluindo a contribuição dos dados provenientes de sensores remotos.

Grades populacionais: construção de grades globais e locais

A questão da representação de dados de população no contexto de estudos que envolvem a dinâmica populacional e sua integração com outras fontes de dados teve início a partir da agenda estabelecida por Waldo Tobler (1979). O marco inicial foi o *Global Demography Workshop* de 1994, organizado pelo *National Center for Geographic Information and Analysis* (NCGIA) e realizado com o apoio do Ciesin, que solidificou a proposta de construir bases globais para dados de população, utilizando como suporte geométrico uma grade composta de *pixels* regulares de formas variadas (Tobler *et al.*, 1995, 1997), conhecida como grade populacional. Desde então, surgiram diversas iniciativas globais, desenvolvendo métodos e aprimorando a representação da distribuição espacial da população em todo o globo. Uma síntese de algumas destas iniciativas globais é apresentada no Quadro 1.

QUADRO 1
Descrição das principais grades populacionais globais

Grades populacionais	Dados populacionais de entrada	Métodos	Anos disponíveis	Resolução espacial	Características	Limitações	Abrangência
GPW	Rodada de Censos dos anos 2010 (2005-2014)	Ponderação por área	2000, 2005, 2010, 2015 e 2020	30 segundos de arco (1 km)	Baseado na densidade de área; bons resultados para análises mais generalizadas	Densidade e distribuição imprecisas na comparação com as outras grades populacionais	Global
GHS-POP	GPWv4.11	Dasimétrico binário	De 1975 a 2030, a cada 5 anos	3 segundos de arco (100 m) e 30 segundos de arco (1 km)	Dados de entrada podem conter erros associados	A variável de área construída possui imprecisão em sua classificação	Global
GRUMP	Dados dos Censos Demográficos do GPWv3: rodada de Censos dos anos 2000 (1995-2004)	Dasimétrico binário	1990, 1995 e 2000	30 segundos de arco (1 km)	Utiliza dados mais generalizados e com menos precisão	Superestimação e subestimação aumentadas em áreas limítrofes	Global
HRSL	GPWv4	Dasimétrico binário	2015	1 segundo de arco (30 m)	Pode apresentar resultados melhores em áreas menos adensadas	Resulta na superestimação de população em áreas urbanas	Global
Landscan	Censos Demográficos e dados administrativos	Dasimétrico com múltiplas variáveis	Desde 1998, anual	30 segundos de arco (1 km)	Superestimação em áreas menos adensadas	Pode apresentar densidade diferente de zero para grande parte da área de estudo	Global
WorldPop	Dados censitários do GPWv4	Random Forest	2000 a 2020	3 segundos de arco (100 m)	Estimação populacional não possui extrapolação	O seu resultado é diretamente relacionado às variáveis auxiliares utilizadas	Global

Fonte: Adaptada, ampliada e atualizada da Tabela em POPGRID (2020).

Em 1995, foi criada a primeira versão do *Gridded Population of the World – version 1* (GPWv1), coordenada por Waldo Tobler e posteriormente hospedada no Ciesin, na Universidade de Columbia, integrada ao *Socioeconomic Data and Applications Center* (Sedac), da Nasa. Este projeto assumiu a missão de desenvolver uma série de produtos de dados de população, assentamentos humanos e infraestrutura construída em escala global.

Atualmente, o GPW encontra-se em sua quarta versão, especificamente na revisão 11 (GPWv4.11). Os detalhes sobre o conjunto de dados, métodos e técnicas utilizados na geração de informações populacionais em grades de células de [1x1] km estão descritos em Ciesin e Doxsey-Whitfield *et al.* (2015). O GPWv4.11 disponibiliza dados para o período de 2000 a 2020, em intervalos de cinco anos, oferecendo estimativas de características

básicas, como idade e sexo, além de dados sobre população total e densidade populacional. Essas estimativas são obtidas a partir das bases de dados censitários oficiais dos países, sendo ajustadas nacionalmente aos totais reportados pelo *United Nation's World Population Prospects* (UN, 2022).

Uma limitação conhecida e importante do GPWv4.11 reside nas estimativas para populações rurais em áreas de baixa densidade, que tendem a ser superestimadas pelo modelo (Chen *et al.*, 2020). Este problema é especialmente evidente em países com dados de população coletados para unidades espaciais de grande extensão territorial, que geralmente envolvem coletas amostrais sobre grandes áreas.

Em 2011, o *Joint Research Centre* (JRC) lançou o projeto GSHL-*Global Human Settlement Layer*⁵ com o objetivo de produzir informações em formato de grades em escala global. Em 2014, o GSHL foi integrado a uma rede de colaboração mais abrangente, com foco em produtos que abrangem a distribuição espacial de população, áreas construídas (relacionadas à temática de urbanização planetária) e mapeamento de assentamentos humanos (Kemper; Melchiorri; Ehrlich, 2021). Um produto de distribuição de contagem populacional em grades de células de [100x100] m, [250x250] m e de [1x1] km foi produzido, de 1975 a 2030, em intervalos de cinco anos, por meio de regressão linear associada ao uso de técnicas computacionais de mineração de dados, utilizando um conjunto de dados auxiliares. As grades disponíveis para a versão anterior, GSHL Data Package 2019, foram geradas para células de [250x250] m (Florczyk *et al.*, 2019).

Dentre os vários produtos oferecidos, o GSHL disponibiliza o GHS-POP.⁶ A versão mais recente é o GHS-POP R2023A (Schiavina *et al.*, 2023), um conjunto de grades populacionais, em formato matricial, com células de [100x100] m representando a distribuição da população residente, expressa como o número de pessoas por célula. As grades foram produzidas a partir de estimativas populacionais para o período entre 1975 e 2030, com intervalos de cinco anos. A versão R2023A é derivada do produto Sedac-Ciesin GPWv4.11 e consiste em desagregações de unidades censitárias ou administrativas para células da grade.

O processo de geração dessas grades utilizou informações sobre a distribuição, volume e classificação de áreas construídas (*built-up areas*), obtidas por meio de imagens de satélites e técnicas de classificação de imagens não supervisionadas. Essas informações são disponibilizadas nas camadas denominadas produtos GSHL-BUILT, que são de quatro tipos: S (*surface*), H (*height*), V (*volume*) e C (*settlement characteristics*) (European Comission, 2023).

Embora os problemas com a detecção automática de índices de *built-up areas* tenham sido minimizados em comparação com as grades anteriores dos pacotes de 2019 e 2022 (GSHL Data Package 2019 e GSHL Data Package 2022), as grades mais recentes (GSHL Data Package 2023) continuam apresentando problemas de superestimação em áreas menos adensadas, geralmente associadas a espaços predominantemente rurais (Calka;

⁵ Para mais informações acesse: https://ghsl.jrc.ec.europa.eu/documents/GHSL_Data_Package_2022.pdf?t=1655995832.

⁶ Para mais informações acesse: https://ghsl.jrc.ec.europa.eu/documents/GHSL_Data_Package_2022.pdf?t=1655995832.

Bielecka, 2020; Carneiro-Freire *et al.*, 2016). Além disso, foram identificados problemas em áreas adensadas, predominantemente urbanas, com diversas tipologias de ocupação (Kuffer *et al.*, 2022).

O projeto WorldPop,⁷ iniciado em 2013 e liderado pela *Southampton University*, no Reino Unido, é hoje uma referência global na produção de grades populacionais. Sua origem remonta à necessidade de criar camadas de distribuição espacial de população com resolução espacial fina para estudos de vigilância epidemiológica e de epidemiologia espacial de doenças vetoriais emergentes e reemergentes em países pobres e em desenvolvimento (Hay *et al.*, 2005; Linard; Tatem; Gilbert, 2013).

O WorldPop gera e dissemina, em acesso livre, grades populacionais com células de [100x100] m e [1x1] km. Seus produtos são desenvolvidos por meio de uma abordagem de modelagem dasimétrica semiautomática,⁸ que usa dados censitários e um conjunto de variáveis (covariáveis) associadas à presença humana para identificar pesos dasimétricos e construir um modelo baseado no algoritmo *Random Forest* (Breiman, 2001). *Random Forest* é um método de aprendizado por máquina, baseado na construção de um grande conjunto de árvores de decisão (*decision trees*) durante uma fase de treinamento, denominada *ensemble learning*, que pode ser usado para tarefas de classificação ou para uma regressão (Breiman, 2001). Este modelo é, então, empregado para gerar uma previsão da densidade populacional em uma grade de [100x100] m. Esta previsão, por sua vez, serve como superfície de ponderação para realizar a redistribuição dasimétrica das contagens do censo em nível de país.

O WorldPop utiliza dados provenientes de fontes oficiais, como censo demográfico, estimativas e projeções, além de informações de bases globais de uso e cobertura da terra e novos dados associados ao contexto de informações geográficas produzidas voluntariamente. A metodologia é descrita em Stevens *et al.* (2015), Lloyd, Sorichetta e Tatem (2017) e Tatem (2017). As grades de estimativa e distribuição da população para células de [100x100] m estão disponíveis para o período de 2000 a 2020, em intervalos anuais. Estas grades, juntamente com materiais técnicos relacionados à sua produção (grades *constrained* e grades *unconstrained*⁹), podem ser encontradas nas páginas do *WorldPop gridded population datasets and tools*.¹⁰

⁷ A iniciativa WorldPop produz dados de distribuição populacional através de estimações de idade e sexo para produzir mapas de alta resolução, células de [100x100]m da distribuição populacional para todo o globo. Os dados podem ser acessados em: <https://www.worldpop.org/datacatalog/>.

⁸ A abordagem dasimétrica será explicada na seção “Métodos utilizados para a produção das grades populacionais”.

⁹ O termo *unconstrained* (irrestrito) refere-se à premissa de que todas as áreas de superfície são utilizadas para realizar a alocação, sem restrições para áreas de unidades de conservação, áreas militares, etc. No caso da grade WorldPop são utilizadas variáveis auxiliares provenientes de vias terrestres, dados climatológicos, dados de uso e cobertura da terra, etc. Todos, porém, de bases globais. Para mais informações consultar: https://www.worldpop.org/methods/top_down_constrained_vs_unconstrained/.

¹⁰ Os dados podem ser acessados em: <https://www.worldpop.org/methods/populations/> e <https://hub.worldpop.org/geodata/listing?id=69>.

O que se observa a partir de um conjunto de estudos dedicados à avaliação destas grades populacionais é que tanto as grades GHS-POP, produzidas sem o uso do método de regressão baseado em *Random Forest*, quanto as grades WorldPop apresentam limitações para lidar com áreas de baixa densidade populacional, frequentemente associadas às regiões predominantemente rurais nas classificações nacionais e globais (Archila et al., 2020). Por outro lado, também apresentam limitações em áreas mais densamente povoadas, predominantemente urbanas, quando há tipologias de ocupação muito diferenciadas, especialmente em países de renda *per capita* baixa ou média (Leyk et al., 2019).

Embora as grades GHS-POP e WorldPop sejam amplamente utilizadas em escala global, outras iniciativas têm se destacado na produção de grades para contextos específicos. Um exemplo é o *Global Rural Urban Mapping Project*¹¹ (Grump), uma base de dados publicada em 2007 como resultado de uma parceria estabelecida entre o Ciesin, o *International Food Policy Research Institute* (IFPRI), o Banco Mundial e o *International Center for Tropical Agriculture* (Ciat) (Balk et al., 2006). O Grump foi uma das primeiras grades populacionais cuja metodologia de modelagem foi divulgada de forma transparente e incluiu dados globais de contagem e densidade populacional para 1995, 1990 e 2000, com resolução de aproximadamente [1x1] km. Entretanto, como limitação, tende a superestimar áreas urbanas devido à inclusão de dados de luzes noturnas, que detêm resolução espacial de cerca de 740 m (POPGRID, 2020).

O projeto *High Resolution Settlement Layer*¹² (HRSL) apresenta duas versões de grades, sendo, a primeira, um projeto conjunto entre o Facebook, o Ciesin e o Banco Mundial. Inicialmente, o HRSL foi construído para 30 países e publicado em 2017. A segunda versão foi realizada entre 2018 e 2019, expandindo a disponibilidade da grade para 160 países (POPGRID, 2020). Entre suas características técnicas, destaca-se sua resolução espacial de aproximadamente [30x30] m e a disponibilidade para o ano de 2015. Essas informações foram obtidas a partir de dados censitários e imagens de satélites de alta resolução, como as imagens do *DigitalGlobe* (agora *Maxar*), com resolução de 0,5 m. Além disso, faz uso de dados obtidos das estruturas edificadas a partir de técnicas de visão computacional (Tiecke et al., 2017). O HRSL se diferencia dos demais projetos globais devido à sua resolução espacial detalhada, bem como à disponibilização de dados populacionais segmentados por idade e sexo. Entretanto, apresenta limitações em termos de representatividade geográfica, não abrangendo todo o globo, e em relação à temporalidade, pois fornece dados apenas para 2015 (POPGRID, 2020).

A grade populacional conhecida como *World Population Estimate*¹³ (WPE) consiste em um dos produtos lançados pela *Environmental Systems Research Institute* (Esri), em 2014. Esta grade considera a localização das pessoas em seus locais de residência, também

¹¹ Os dados podem ser acessados em: <https://sedac.ciesin.columbia.edu/data/collection/grump-v1/methods>.

¹² Os dados do HRSL podem ser acessados em: <https://www.ciesin.columbia.edu/data/hrsl/>.

¹³ Os dados podem ser acessados em: <https://www.arcgis.com/apps/View/index.html?appid=9746b14e58e94ef4bc9bd80f55e9d65e>.

chamada de população noturna (*nighttime population*), em contraposição aos lugares onde as pessoas passam o dia. O projeto utiliza imagens do satélite Landsat 8 e variáveis auxiliares como vias, cobertura da terra e corpos hídricos. O WPE fornece estimativas populacionais para 2013, 2015 e 2016, com uma resolução espacial de [250x250] m, para 2013 e 2015, e [150x150] m, para 2016. Entretanto, o acesso a este produto não é livre, pois sua política de distribuição é comercial. Além disso, sua metodologia de estimação varia entre os anos, o que dificulta a comparação entre as distintas versões do produto (Frye, 2017; POPGRID, 2020).

Iniciativas do Brasil para a construção de grades populacionais locais

No Brasil, os investimentos no desenvolvimento e avaliação de abordagens dasimétricas para (re)distribuição espacial da população em grades começaram nos anos 2000, com maior concentração de estudos e projetos no Instituto Nacional de Pesquisas Espaciais (Inpe) e no Núcleo de Estudos de População “Elza Berquó” (Nepo), da Universidade Estadual de Campinas (Unicamp). A motivação principal naqueles anos era o debate sobre as relações entre população e ambiente (Hogan, 1991, 1998; De Sherbinin *et al.*, 2007). No Nepo, os trabalhos foram coordenados pelo Professor Daniel Hogan e, posteriormente, pelo Professor Álvaro D’Antona, consolidando esta agenda no Núcleo (D’Antona; Bueno; Dagnino, 2013; Bueno, 2014; D’Antona; Vanwey; Ludewigs, 2011).

O trabalho de Kampel (2003) inaugurou e consolidou no Inpe uma agenda de pesquisas que teve início no final dos anos 1990 e continua até os dias de hoje, como parte da agenda científica do Laboratório de Investigação em Sistemas Socioambientais (Liss-Inpe) (Amaral *et al.*, 2012; Anazawa; Carmo, 2021). Esta agenda também se estabeleceu no Instituto Brasileiro de Geografia e Estatística (IBGE), com o trabalho de Bueno (2014), cuja tese foi desenvolvida no Nepo, e que coordenaria, mais adiante, a criação da grade estatística do IBGE.¹⁴

A grade estatística do IBGE foi lançada em 2016 como uma representação adicional para a disseminação dos dados estatísticos. Para a sua construção, foram utilizados os microdados do universo do Censo Demográfico 2010, juntamente com um conjunto de variáveis auxiliares de uso e cobertura da terra (Projeto de Monitoramento do Desmatamento dos Biomas Brasileiros por Satélite – PMDBBS e Projeto TerraClass 2010) e vias (Cadastro de Endereços para Fins Estatísticos – CNEFE). A abordagem híbrida empregada associou os dados censitários à grade, combinando técnicas de agregação (para locais onde a totalidade ou grande percentual dos registros apresentava localização geográfica) e desagregação (para locais onde não havia localização geográfica dos domicílios, usando um método dasimétrico). A localização geográfica foi obtida por meio da coleta com GPS (*Global Position System*), nos domicílios visitados pelo Censo Demográfico 2010. Como

¹⁴ Para acesso à grade estatística: https://geoftp.ibge.gov.br/recortes_para_fins_estatisticos/grade_estatistica/censo_2010/grade_estatistica.pdf.

resultado, a grade estatística do IBGE contém informações sobre a população total e por sexo, bem como o número total de domicílios. Ela apresenta células de [1x1] km para áreas rurais e [200x200] m para áreas urbanas (IBGE, 2016). O Quadro 2 mostra uma síntese da grade estatística do IBGE.

QUADRO 2
Descrição das principais características da grade estatística do IBGE

Grades populacionais	Dados populacionais de entrada	Métodos	Anos disponíveis	Resolução espacial	Características	Limitações	Abrangência
Grade Estatística (IBGE)	Microdados Censo Demográfico 2010 + dados auxiliares	Abordagem híbrida	2010	1 km e 200 m	Áreas urbanas mais representativas	Pode apresentar níveis de generalização maiores nas áreas rurais	Local

Fonte: IBGE (2016).

No campo dos métodos de estimativas de população em pequenas áreas, baseados em técnicas de regressão e fazendo uso de imagens de satélite de observação da Terra e dados auxiliares derivados destas imagens, um conjunto de trabalhos de autoria de Ilka Reis, professora da Universidade Federal de Minas Gerais (UFMG), foi desenvolvido no Inpe (Silva; Reis; Reis, 2009). Mais recentemente, destacam-se o trabalho de Campos et al. (2020), evidenciando um campo de estudos em franco desenvolvimento no Brasil, e o estudo de Silva (2023), que produziu e avaliou uma nova grade populacional, na escala local, denominada VALIN-POP.¹⁵ Utilizando bases de dados e conhecimentos locais para a construção de novas variáveis indicadoras da presença e ocupação humana, VALIN-POP foi construída a partir da abordagem de aprendizado por máquina, especificamente a aplicação da regressão baseada em *Random Forest*, utilizando dados oficiais do Censo 2010 e variáveis sintomáticas provenientes de bases locais e globais.

A produção destas grades brasileiras está comprometida com agendas locais, diferentemente das grades globais que possuem interesse em representar a população de modo mais geral, podendo omitir volumes populacionais, tornando parcelas do globo menos representativas.

Variáveis sintomáticas: seus contextos e utilizações para a produção de grades populacionais

Dados espaciais, em geral, apresentam três grandes suportes, do ponto de vista de suas geometrias: (i) como elementos vetoriais, na forma de pontos, linhas e polígonos; (ii) como elementos matriciais, na forma de uma grade, que tem *pixels* de geometria regular e valor único, formando um reticulado (*lattice*), como suportes matriciais; e (iii) como uma representação híbrida, que se estabelece como uma grade de polígonos regulares,

¹⁵ A grade populacional VALIN-POP pode ser acessada em: <https://zenodo.org/records/10140007>.

células, com múltiplos valores, que se organizam em uma topologia de matrizes, como suportes celulares.

Normalmente, os dados populacionais que são coletados, medidos ou observados em uma unidade espacial de referência são agregados em unidades de áreas poligonais para fins de disseminação. Para a produção de grades populacionais, geralmente é necessário aplicar uma série de transformações sobre estes dados agregados, a fim de incorporá-los a um novo suporte geométrico: seja ele matricial ou celular. Essas transformações, no contexto das grades populacionais, estão associadas às abordagens dasimétricas, que têm uma longa história na análise da distribuição espacial de população (Dmowska, 2019; Semenov-Tian-Shansky, 1928; Wright, 1936). No entanto, os novos procedimentos dasimétricos, como os propostos em Mennis e Hultgren (2006), Nagle *et al.* (2014), Dória, Amaral e Monteiro (2016) e Hu *et al.* (2019), também não estão isentos de erros e/ou incertezas.

O método dasimétrico realiza a subdivisão de áreas e alocação de informação a partir de um conjunto de variáveis sintomáticas. Estas variáveis auxiliam na estimativa de diferentes pesos para a desagregação da informação que antes era apresentada de forma agregada.

No campo dos métodos e tecnologias para geração de grades, houve avanços significativos e rápidos, especialmente nas tecnologias computacionais, em relação à capacidade de estoque de informação e velocidade de processamento. Essas tecnologias agora possuem capacidades aprimoradas para captura, armazenamento, processamento e disseminação de dados e informações com referência de localização. Destacam-se os avanços nos métodos estatísticos, impulsionados pelos inovadores ambientes de computação estatística e pela evolução das técnicas de aprendizado por máquinas e reconhecimento de padrões no campo da inteligência artificial (Calian; Zuppardo; Hardarson, 2023).

Para além do aprimoramento de tecnologias e metodologias, avanços significativos também foram observados no que diz respeito à produção de dados provenientes tanto de setores públicos como privados e parcerias: dados de censos demográficos e *surveys* das agências nacionais podem ser refinados espacialmente com o auxílio de informações provenientes de sensoriamento remoto orbital por satélites de observação da Terra, assim como de um conjunto de novos dados com referência espacial e temporal. Dentre estes últimos, destacam-se os dados de grafos em associação com as novas tecnologias de informação e comunicação (TICs): informação geográfica voluntária (*Volunteered Geographic Information – VGI*), mídias sociais e colaboração coletiva de dados e informações (*Social Media and Crowdsourcing*) e ciência cidadã (*Citizen Science*), principalmente (See *et al.*, 2016). Quando devidamente tratadas, essas fontes de dados apresentam potencial para uso no contexto dos estudos populacionais, em particular na análise da distribuição espacial da população.

Na demografia formal, mais especificamente no campo estrito dos métodos e técnicas para estimativas e projeções de população em pequenas áreas administrativas, há sempre um desafio: a disponibilidade e produção de dados primários que possam auxiliar o

refinamento dos modelos utilizados para estimativas e projeções, visto que são originados de dados oficiais (Alves; Cavenaghi; 2012; Jannuzzi, 2007). Para contornar esta dificuldade, a solução foi incorporar aos recentes modelos produzidos, até então, um conjunto de variáveis auxiliares relacionadas ao crescimento da população e à sua mobilidade, que são coletadas regularmente (Reed *et al.*, 2018; Stevens *et al.*, 2015).

As variáveis auxiliares são essenciais para capturar as alterações da dinâmica populacional em pequenas áreas não amostradas, e são conhecidas como variáveis sintomáticas no *stricto sensu* da demografia formal. Elas se relacionam com a informação coletada regularmente por empresas e instituições públicas ou privadas para fins administrativos. Incluem matrículas escolares, declarações fiscais, o número de assinantes de serviços básicos (eletricidade, água, etc.), o número de eleitores inscritos nos cadastros eleitorais, entre outras possibilidades. Embora a qualidade dos registros administrativos, principalmente em países de baixa e média renda (LMICs – *Low-and Middle- Income Countries*), precise melhorar significativamente (Mesa, 2019), o uso destas variáveis sintomáticas tem sido amplamente explorado na literatura acadêmica nesse campo de estudos (Alvarez, 2001; Brito; Cavenaghi; Jannuzzi, 2010; Cabrera, 2011; Esquivel, 2001).

Neste trabalho, o termo *variável sintomática* é usado com certa liberdade, não associado ao seu uso mais restrito, que aparece no contexto da demografia formal. Aqui, uma variável sintomática é considerada uma variável auxiliar, que pode ser construída a partir de bases de dados não populacionais, que são sistematicamente produzidas e não estão associadas a registros civis, porém acompanham a dinâmica populacional existente em determinado território. Estas bases de dados devem conter informações associadas à possibilidade de uso de certos territórios para fins habitacionais, servindo como potenciais indicadores de presença/ocupação humana.

Nesse contexto, as variáveis sintomáticas auxiliam o processo de alocação espacial nas células das grades de população estimada, sendo produzidas a partir de imagens de satélites de observação da Terra, como as bases de dados de uso e cobertura da terra e dados construídos por grandes bases de informação geográfica voluntariamente sistematizadas, com alguma curadoria, como o *OpenStreetMap*.¹⁶ Estas bases são relativamente novas no contexto dos usos na demografia e permitem acrescentar às grades populacionais locais informações e características específicas de cada população e sua distribuição, permitindo assim um detalhamento das condições de distribuição espacial da população que as grades globais tratam de forma homogênea.

¹⁶ Acesso do OpenStreetMap em: <https://www.openstreetmap.org/about>.

Grades populacionais e dados de sensoriamento remoto

As iniciativas para o uso de dados de sensoriamento remoto orbital em aplicações envolvendo indivíduos, grupos e populações remontam aos primórdios do Programa de Satélites de Observação da Terra, da Nasa, inicialmente denominado *Earth Resources Technology Satellite* (Erts) e posteriormente conhecido como programa Landsat (Landsat, 2017). Em 1972, Charles E. Fuller, na época chefe do grupo de Ecologia em Saúde Pública do Centro de Naves Espaciais Tripuladas da Nasa, circulou um documento interno intitulado *Public health applications of remote sensing* (Fuller, 1972), estabelecendo a agenda de pesquisas do grupo em associação com as missões do programa emergente de satélites de observação da Terra. Nesse momento, o estudo das relações entre indivíduos, grupos e populações e o desenvolvimento de certos processos de saúde e doença começaram a desempenhar um papel instrumental na produção de informações auxiliares, especialmente em contextos de estudos de saúde em que os impactos sobre as populações eram o foco da atenção.

No entanto, somente em 1998, 26 anos depois, o *National Research Council* (NRC) dos EUA lançou um livro que organizou as diversas pesquisas e aplicações no campo das ciências sociais que buscavam introduzir o uso de imagens de satélites de observação da Terra em seus métodos e metodologias, abrangendo também o campo de estudos de população (Liverman *et al.*, 1998).

Complementarmente, em 2004, um livro editado pelo *Center for Spatially Integrated Social Science* (CSISS), que surgiu no contexto do consórcio *National Center for Geographic Information and Analysis* (NCGIA), ambas iniciativas da *National Science Foundation* (NSF) dos EUA, organizou estudos aplicados no campo das ciências sociais, com técnicas, métodos, metodologias e tecnologias associadas à análise de dados geográficos (Goodchild; Janelle, 2004).

Em um artigo mais recente, Kugler *et al.* (2019) demonstram a evolução conjunta dos métodos e dados, em particular nos estudos envolvendo interações entre população e ambiente. Avanços no campo da geração, disseminação e políticas de acesso aos dados de observação da Terra por satélites ocorreram com velocidade sem precedentes. Os programas eram de natureza nacional e internacional, envolvendo principalmente entidades públicas, mas também privadas. Em consonância com isso, ocorreram avanços disruptivos nos métodos, técnicas e na infraestrutura computacional de processamento de grandes volumes de dados de sensoriamento remoto e sua integração com outras bases de dados espaciais.

No Brasil, o projeto BDC – *Brazil Data Cube*¹⁷ (Ferreira *et al.*, 2020) é um excelente exemplo. Desde os anos 2000, as condições supracitadas possibilitaram uma evolução consistente, com melhorias que se refletem na oferta de grades populacionais globais, com algumas limitações, mas também oportunidades.

¹⁷ Acesso ao BDC em: <http://www.brazildatacube.org/en/home-page-2/>.

Decisões sobre a localização de indivíduos, famílias e grupos populacionais são influenciadas por suas relações históricas, abrangendo aspectos sociais, econômicos e culturais, com seus territórios de vida e vivência. O espaço escolhido para viver, ou seja, os locais habitados pelos seres humanos, é intrinsecamente uma construção social. Cada sociedade deixa marcas distintas na paisagem, que refletem seus modos de habitar e transformar (Segaud, 2016). Portanto, aspectos relacionados às paisagens em constante transformação, como os padrões de uso e cobertura da terra, oferecem informações indiretas sobre a estrutura que guia a decisão de ocupar determinados lugares. Isso nos permite identificar estruturas criadas pelo ser humano para habitação, produção e circulação, fornecendo pistas da possível presença humana em recortes territoriais específicos.

Informações que sinalizam a possibilidade de presença de assentamentos e infraestruturas e, portanto, de presença humana podem ser extraídas de várias fontes de dados provenientes de sensoriamento remoto orbital. Isso inclui bases de dados de uso e cobertura da terra derivadas de sensores orbitais de observação da Terra, bases de dados de satélites de observação da Terra com sensores para capturar luzes noturnas, que indicam atividades humanas, e bases de dados que identificam potenciais áreas edificadas (*built-up areas*) a partir de imagens de sensoriamento remoto.

Portanto, os dados de observação da Terra gerados por plataformas orbitais desempenham um papel fundamental na análise das mudanças ocorridas nessas paisagens, que são moldadas pelas populações e suas dinâmicas socioterritoriais. Utilizando estas bases de dados, é possível criar um conjunto de variáveis auxiliares, ou sintomáticas, que indicam a provável presença humana em unidades espaciais de referência. Estas variáveis incluem informações sobre o uso e cobertura da terra, índices de áreas edificadas, indicadores derivados de dados de luzes noturnas e redes de caminhos, vias, estradas e rodovias que informam sobre modos de circulação das pessoas.

São variáveis desta natureza que os modelos de geração de grades populacionais utilizam principalmente durante a fase de distribuição espacial da população estimada. As grades globais empregam bases de dados globais que definem classes de uso e cobertura da terra em suas legendas. O aumento na disponibilidade desses mapeamentos e bases de dados levou a uma preocupação com a compatibilização das legendas associadas às diversas fontes de informações sobre uso e cobertura da terra em todo o mundo.

Essa preocupação motivou as Nações Unidas a definir um programa que estabelece um sistema de classificação de cobertura da terra (um *Land Cover Classification System*),¹⁸ em uma tentativa de estabelecer métodos para harmonizar as distintas legendas que estavam surgindo.

¹⁸ Para mais informações acesse: <https://www.fao.org/4/x0596e/x0596e00.htm>.

Métodos utilizados para a produção das grades populacionais

A produção de grades populacionais baseia-se em métodos de desagregação e agregação de dados demográficos. O procedimento de desagregação, também conhecido como abordagem “*top-down*”, é a forma como contagens populacionais são subdivididas para unidades menores, a partir de atribuição de características específicas a cada subunidade com apoio de dados geoespaciais (Stevens *et al.*, 2015). Este método envolve a realocação e fragmentação de dados quantitativos populacionais originados de resolução espacial menos detalhada (como municípios, setores censitários e distritos) para um alvo de resolução mais refinada, como unidades celulares com resolução de [100x100] m. Por sua vez, o método “*bottom-up*” estabelece uma agregação a partir de unidades individuais, sendo mais indicado quando há ausência de informações censitárias atuais. Nestes casos, em substituição aos dados censitários que possuem intervalos de captura de informação da população, como nos anos censitários, são utilizados dados de registro civil, de nascimentos, mortalidade, etc., a exemplo da Fundação Seade, que realiza a projeção da população através do método dos componentes demográficos (Seade, 2023).

Abordagens “*top-down*” e “*bottom-up*”, no entanto, podem possibilitar vantagens e desvantagens quando empregadas. A utilização de dados de baixa qualidade ou dados censitários desatualizados pode comprometer severamente as abordagens “*top-down*”, sobretudo para áreas urbanas, cuja população tende a ser subestimada. Por outro lado, abordagens “*bottom-up*”, apesar de serem vantajosas para os casos em que há ausência parcial ou dados não confiáveis, podem necessitar de variáveis sintomáticas que possuam relação com os quantitativos populacionais (Schug *et al.*, 2021).

De modo geral, existem duas maneiras de se obter a desagregação: uma baseada em informações auxiliares e outra que não as utilizam. A seguir, são descritas as principais abordagens e métodos.

O método de ponderação por área (*areal interpolation*) envolve a realocação de dados censitários de uma camada fonte para uma camada alvo, geralmente com resolução espacial mais fina, sem a necessidade de introduzir informações auxiliares durante o processo (Goodchild; Lam, 1980, Mennis; Hultgren, 2006). A interpolação da população pressupõe homeogeneidade de distribuição populacional na área de origem, sendo que o valor a ser atribuído à camada alvo é proporcional à sua área na camada fonte (Goodchild; Lam, 1980). Com suas variações, este é o método mais utilizado nas abordagens que não fazem uso de informações auxiliares na alocação espacial da população estimada em aplicações de grades populacionais.

O mapeamento dasimétrico é o método mais utilizado no contexto das metodologias que utilizam dados auxiliares para melhorar a representação da distribuição espacial da população. Como técnica para desagregação de um conjunto de dados de entrada, o método dasimétrico utiliza um suporte baseado em unidades espaciais de referência maiores para redistribuição em unidades espaciais mais detalhadas, como as células de uma grade populacional. Os dados auxiliares, aqui referidos como variáveis sintomáticas, fornecem

informação para subdivisões dos quantitativos populacionais. Estas estão relacionadas ao uso e cobertura da terra, à infraestrutura de circulação, entre outros, contribuindo para distribuir os dados populacionais de forma mais adequada.

Na abordagem dasimétrica, o método dasimétrico binário e o método dasimétrico inteligente (Leyk *et al.*, 2019; Wright, 1936) são os mais utilizados. No método dasimétrico binário, é definida uma estratégia de ponderação binária a partir de uma variável sintomática, utilizando pesos binários para apontar a presença ou a ausência de população na unidade mais detalhada para onde a realocação deve ocorrer (Eicher; Brewer, 2001, Mennis *et al.*, 2006).

O método dasimétrico inteligente ou não binário também utiliza variáveis auxiliares. No entanto, o conjunto de variáveis sintomáticas pode ter diferentes pesos e contribuições, a depender da relevância da variável para indicar a presença de população. Geralmente, a integração de múltiplas camadas de dados auxiliares promove uma melhoria na qualidade das estimativas (Amaral *et al.*, 2005; De Mattos; Bertolotto, 2020; França, 2012; Langford, 2013).

Alguns artigos discutem a contribuição das variáveis sintomáticas na modelagem da distribuição de população. Zhou, Wu e Tang (2002), por exemplo, estimaram a distribuição populacional para áreas de 30 m de resolução espacial em uma região no sul da China, utilizando uma abordagem de regressão do algoritmo *Random Forest* e dados de fontes variadas, como agências governamentais, organizações internacionais e instituições de pesquisa. Dentre os dados utilizados, pontos de interesse (POIs), caracterizados por pontos de localização diretamente relacionados com a presença populacional, se mostraram mais efetivos para modelar a distribuição populacional do que dados de luzes noturnas e dados topográficos, principalmente em áreas urbanas.

Por sua vez, Wan *et al.* (2021) empregaram dados de uso e cobertura da terra para assentamentos denominados *Built-Up Property Records* (BUPR), provenientes do *Historical Settlement Data Compilation for the United States* (HISDAC-US), para desagregar dados populacionais dos censos para células. O método baseado em BUPR superou outros conjuntos de dados auxiliares utilizados para o refinamento de dados populacionais, exceto para aquelas áreas em que esta informação estava incompleta. Neste caso, os autores sugeriram, como alternativa, complementar tais dados por outros como rede de vias e informações sobre cobertura.

Os diferentes métodos empregados para a produção de grades populacionais têm contribuído para uma melhor compreensão de seus resultados. Archila *et al.* (2020) compararam diferentes produtos de grades: GHS-POP, GPW, Grump, LandScan e WorldPop, descritos anteriormente no Quadro 1. As grades populacionais globais GHS-POP, LandScan e WorldPop obtiveram melhores resultados. No entanto, WorldPop, GPWv4 e Grump identificaram a densidade populacional das células povoadas e não povoadas com maiores percentuais de acerto, diferentemente dos dados do GHS-POP e LandScan, que identificaram corretamente cerca de 80% cada. Os erros estiveram mais associados a células não povoadas para as grades do WorldPop, GPWv4 e Grump.

Recentemente, alguns trabalhos se debruçaram na construção de grades populacionais baseadas na utilização de aprendizagem profunda, a partir de imagens de sensoriamento remoto. Huang *et al.* (2021) compararam diferentes arquiteturas (VGG, ResNet, Xception e DenseNet) de aprendizado profundo mais recentemente utilizadas para a construção de grades populacionais, a partir de padrões de imagens de satélite do sensor Sentinel-2 e a grade de quantitativos populacionais do *LandScan population grid*, para gerar dados de volume e distribuição da população de duas regiões metropolitanas dos Estados Unidos. Dentre as arquiteturas empregadas, os resultados mostraram que o DenseNet obteve melhores resultados das métricas de avaliação, considerando todos os tamanhos de vizinhança selecionados e desempenhando melhor com a heterogeneidade da distribuição populacional, relacionada com os padrões da imagem de satélite. No entanto, padrões de vizinhança presentes na imagem diminuíram o desempenho de estimação da população.

Robinson, Hohman e Dilkina (2017) fizeram uso da rede de neurônios para realizar a predição da população em uma grade populacional refinada para os Estados Unidos a partir de imagens do Landsat e dados populacionais do Ciesin. Os resultados mostraram superestimação de população para algumas áreas mais extensas.

Considerações finais

Este artigo evidencia algumas contribuições na evolução da produção de grades populacionais, com mais e melhores técnicas e com o suporte de dados variados, incluindo os não populacionais. Ressalta-se, contudo, que, apesar de as grades populacionais contribuírem para a compreensão de certas dinâmicas demográficas, é necessário considerar a forma como são construídas e acessar sua qualidade. A avaliação deve ser feita com muita cautela, principalmente em relação aos novos produtos e seus métodos utilizados para produção de dados populacionais detalhados gerados para a escala global. A aderência destes produtos globais ao contexto de escalas locais, como escalas metropolitanas e regionais, pode não ser adequada para a representação da distribuição da população naquela região geográfica específica. Isso contribui para a invisibilidade de parcelas da população e a omissão das políticas públicas voltadas à população naquelas áreas, geralmente as mais carentes.

As iniciativas apresentadas aqui fornecem contribuições para aprimorar a produção de grades, com qualidade e acurácia. A seleção e escolha de grades populacionais existentes para os diferentes usos devem se pautar na consideração das heterogeneidades e particularidades presentes nestes territórios, além de considerar as mudanças temporais da distribuição populacional e seu acompanhamento. As possibilidades de integração de dados de diferentes fontes, como dados de sensores remotos, registros administrativos e dados de informação geográfica voluntária, vêm contribuindo para a qualidade da representação de volumes e distribuição em grades populacionais.

De fato, os erros das bases de dados populacionais não são homogêneos ou aleatórios no espaço, assim como a intensidade da correlação das variáveis sintomáticas com

a densidade demográfica não é a mesma para todas as escalas de análise. Portanto, estimativas a partir das metodologias conhecidas como “*top-down*” inevitavelmente irão reproduzir os problemas das informações de base e das escolhas das variáveis que servem de “*inputs*” aos modelos. Assim, a própria escolha das variáveis e dos modelos deve ser precedida de estudos substantivos aprofundados sobre a distribuição espacial da população e seu contexto. Aliados à disponibilidade cada vez maior de informações para todo o território nacional, a exemplo das coleções de dados prontos para análise (*Analysis-Ready Data* – ARD) do projeto BDC/Inpe, abre-se um novo leque de oportunidades e desafios para os estudos populacionais baseados em evidências empíricas de alta qualidade e expertise nacionais.

Como a ênfase deste trabalho foram as grades globais, os métodos “*bottom-up*” não foram abordados. Porém, para áreas de baixa cobertura censitária, ou longo interstício temporal antes do último recenseamento ou contagem populacional, a abordagem “*bottom-up*” pode ser valiosa. Em que pese a necessidade de dados de alta resolução espacial, as inovações tecnológicas atuais apontam para uma década profícua para a integração de novos dados de entrada e metodologias de construção de grades populacionais.

Por fim, apesar de a produção de dados geoespaciais para todo o globo ter se desenvolvido ao longo do tempo, a seleção destas informações, que apoiem a distribuição local da população em grades, ainda carece de mais compreensão do contexto histórico de ocupação singular em diferentes circunstâncias e escalas. Esta compreensão configura uma etapa inicial para a construção de modelos mais robustos somados aos distintos métodos, evidenciando a necessidade de abordagens interdisciplinares e multiescalares. Em outras palavras, antes mesmo de selecionar métodos robustos que modelam a população em células, devemos propor criteriosamente os dados de entrada para estes modelos.

Referências

ALVAREZ, G. O. **Estimación de población en áreas menores mediante variables sintomáticas: una aplicación para los departamentos de la República Argentina (1991 y 1996)**. Celade – División de Población, Cepal, 2001.

ALVES, J. E. D.; CAVENAGHI, S. Tendências demográficas, dos domicílios e das famílias no Brasil. **Aparte: Inclusão Social em Debate**, v. 24, p. 1-33, 2012.

AMARAL, S. *et al.* Estimating population and energy consumption in Brazilian Amazonia using DMSP night-time satellite data. **Computers, Environment and Urban Systems**, v. 29, n. 2, p. 179-195, 2005.

AMARAL, S. *et al.* Using remote sensing and census tract data to improve representation of population spatial distribution: case studies in the Brazilian Amazon. **Population and Environment**, v. 34, p. 142-170, 2012.

ANAZAWA, T. M.; CARMO, R. L. O risco construído: reflexões sobre o desastre ocorrido em Mariana, estado de Minas Gerais, em 2015. **Revista Vértices**, v. 23, n. 1, p. 234-255, 2021.

ARCHILA, B. M. F. *et al.* A pixel level evaluation of five multitemporal global gridded population datasets: a case study in Sweden, 1990-2015. **Population and Environment**, v. 42, p. 255-277, 2020.

BALK, D. L. *et al.* Determining global population distribution: methods, applications and data. **Advances in Parasitology**, v. 62, p. 119-156, 2006.

BALK, D. *et al.* The distribution of people and the dimension of place: methodologies to improve the global estimation of urban extents. *In*: INTERNATIONAL SOCIETY FOR PHOTOGRAMMETRY AND REMOTE SENSING. **Proceedings [...]**. Tempe, Arizona, 2005. p. 14-16.

BREIMAN, L. Random forests. **Machine Learning**, v. 45, p. 5-32, 2001.

BRITO, L. P. G.; CAVENAGHI, S.; JANNUZZI, P. Estimativas e projeções populacionais para pequenos domínios: uma avaliação da precisão para municípios do Rio de Janeiro em 2000 e 2007. **Revista Brasileira de Estudos de População**, v. 27, n. 1, p. 35-57, 2010.

BRUCE, A.; BRUCE, P. **Estatística prática para cientistas de dados**. Alta Books, 2019.

BUENO, M. C. D. **Grade estatística: uma abordagem para ampliar o potencial analítico de dados censitários**. Tese (Doutorado em Demografia) – Universidade Estadual de Campinas (Unicamp), Campinas, 2014.

CABRERA, M. **Estimación de población en áreas menores con métodos que utilizan variables sintomáticas**. Uruguay: Comisión Sectorial de Población, 2011.

CALIAN, V.; ZUPPARDO, M.; HARDARSON, O. Machine learning estimation of the resident population. **Statistical Journal of the IAOS**, v. 39, n. 4, p. 947-960, 2023.

CALKA, B.; BIELECKA, E. GHS-POP accuracy assessment: Poland and Portugal case study. **Remote Sensing**, v. 12, n. 7, p. 1105, 2020.

CAMPOS, J. *et al.* Population estimates from orbital data of medium spatial resolution: applications for a Brazilian municipality. **Sustainability**, v. 12, n. 9, p. 3565, 2020.

CARNEIRO-FREIRE, S. *et al.* Development of new open and free multi-temporal global population grids at 250 m resolution. *In*: ASSOCIATION OF GEOGRAPHIC INFORMATION LABORATORIES IN EUROPE (AGILE). **Proceedings [...]**. Helsinki, Finland: AGILE, 2016.

CHEN, R. *et al.* Multiple global population datasets: differences and spatial distribution characteristics. **ISPRS International Journal of Geo-Information**, v. 9, n. 11, p. 637, 2020.

D'ANTONA, Á. 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 de População**, v. 30, n. 2, p. 401-428, 2013.

D'ANTONA, A.; VANWEY, L.; LUDEWIGS, T. Polarização da estrutura fundiária e mudanças no uso e na cobertura da terra na Amazônia. **Acta Amazonica**, v. 41, p. 223-232, 2011.

DE SHERBININ, A. *et al.* Population and environment. **Annual Review of Environment and Resources**, v. 32, p. 345-373, 2007.

DE MATTOS, A. C. H.; BERTOLOTTI, G. M. M. Assessing the quality of gridded population data for quantifying the population living in deprived communities. *In*: CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS (NEURIPS 2020), 34. **Proceedings [...]**. Vancouver, Canada: NeurIPS, 2020.

DEICHMANN, U.; BALK, D.; YETMAN, G. **Transforming population data for interdisciplinary usages: from census to grid**. Washington, DC: Center for International Earth Science Information Network, 2001.

DMOWSKA, A. Dasymetric modelling of population distribution—large data approach. **Quaestiones Geographicae**, v. 38, n. 1, p. 15-27, 2019.

DÓRIA, V. E. M.; AMARAL, S.; MONTEIRO, A. M. V. Estimativa e distribuição espacial de população urbana com imagens de satélites de luzes noturnas: um estudo para a região metropolitana de São Paulo, Brasil, com o sensor visible/infrared imaging radiometer suite (VIIRS). **Geografia**, v. 41, n. 3, p. 527-548, 2016.

DOXSEY-WHITFIELD, E. *et al.* Taking advantage of the improved availability of census data: a first look at the gridded population of the world, version 4. **Papers in Applied Geography**, v. 1, n. 3, p. 226-234, 2015.

EICHER, C. L.; BREWER, C. A. Dasymetric mapping and areal interpolation: implementation and evaluation. **Cartography and Geographic Information Science**, v. 28, n. 2, p. 125-138, 2001.

ESQUIVEL, E. A. C. Variables sintomáticas en las estimaciones poblacionales a nivel cantonal em Costa Rica. **Notas de Población**, n. 71, p. 51-72, 2001.

EUROPEAN COMMISSION. **GHSL data package 2023**. 2023. Disponível em: https://ghsl.jrc.ec.europa.eu/documents/GHSL_Data_Package_2023.pdf?t=1683540422. Acesso em: 02 set. 2022.

FEITOSA, F. F. Big data e urban analytics à brasileira: questões inerentes a um país profundamente desigual. **Revista de Morfologia Urbana**, v. 8, n. 1, artigo e00141, 2020.

FERREIRA, K. R. *et al.* Earth observation data cubes for Brazil: requirements, methodology and products. **Remote Sensing**, v. 12, n. 24, p. 4033, 2020.

FLORCZYK, A. J. *et al.* **GHSL data package 2019**. Luxembourg: EU, 2019.

FRANÇA, V. O. **Avaliação de métodos dasimétricos para estimativa populacional em pequenas áreas**. Dissertação (Mestrado em Estudos Populacionais e Pesquisas Sociais) – Escola Nacional de Ciências Estatísticas (Ence), Rio de Janeiro, 2012.

FRYE, C. **Esri's 2016 world population estimate methodology**. Redlands, CA: Esri, 2017. (Esri® White Paper).

FULLER, C. E. Public health applications of remote sensing. **4th Annual Earth Resources Program Review**. Volume 1 – National Aeronautics and Space Programs. Houston, TX: Nasa, 1972.

GOODCHILD, M. F.; JANELLE, D. G. (Ed.). **Spatially integrated social science**. Oxford University Press, 2004.

GOODCHILD, M. F.; LAM, N. S. Areal interpolation: a variant of the traditional spatial problem. **Geo-processing**, v. 1, n. 3, p. 297-312, 1980.

HAY, S. I. *et al.* Urbanization, malaria transmission and disease burden in Africa. **Nature Reviews Microbiology**, v. 3, n. 1, p. 81-90, 2005.

HIERINK, F. *et al.* Differences between gridded population data impact measures of geographic access to healthcare in sub-Saharan Africa. **Communications Medicine**, v. 2, n. 1, p. 117, 2022.

HOGAN, D. J. Crescimento demográfico e meio ambiente. **Revista Brasileira de Estudos de População**, v. 8, n. 1/2, p. 61-71, 1991.

HOGAN, D. J. Mobilidade populacional e meio ambiente. **Revista Brasileira de Estudos de População**, v. 15, n. 2, p. 83-92, 1998.

HU, W. *et al.* Mapping missing population in rural India: a deep learning approach with satellite imagery. In: AAAI/ACM CONFERENCE ON AI, ETHICS, AND SOCIETY. **Proceedings [...]**. New York: Association for Computing Machinery, 2019.

HUANG, X. *et al.* Sensing population distribution from satellite imagery via deep learning: model selection, neighboring effects, and systematic biases. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 14, p. 5137-5151, 2021.

IBGE – Instituto Brasileiro de Geografia e Estatística. **Grade Estatística**. 2016. Disponível em: ftp://geoftp.ibge.gov.br/malhas_digitais/censo_2010/grade_estatistica/Grade_Estatistica.pdf. Acesso em: 04 ago. 2021.

JANNUZZI, P. M. Cenários futuros e projeções populacionais para pequenas áreas: método e aplicação para distritos paulistanos 2000-2010. **Revista Brasileira de Estudos de População**, v. 24, n. 1, p. 109-137, 2007.

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

KEMPER, T.; MELCHIORRI, M.; EHRlich, D. **Global human settlement layer**. Luxembourg: Publications Office of the European Union, 2021.

KUFFER, M. *et al.* The missing millions in maps: exploring causes of uncertainties in global gridded population datasets. **ISPRS International Journal of Geo-Information**, v. 11, n. 7, p. 403, 2022. Disponível em: <https://doi.org/10.3390/ijgi11070403>. Acesso em: maio 2024.

KUGLER, T. A. *et al.* People and pixels 20 years later: the current data landscape and research trends blending population and environmental data. **Population and Environment**, v. 41, p. 209-234, 2019.

LAITURI, M. *et al.* SDG Indicator 11.3. 1 and secondary cities: an analysis and assessment. **ISPRS International Journal of Geo-Information**, v. 10, n. 11, p. 713, 2021.

LANDSAT. **Landsat's enduring legacy: pioneering global land observation from space**. Landsat Legacy Team. American Society for Photogrammetry and Remote Sensing. Digital Free, 2017.

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

LEYK, S. *et al.* The spatial allocation of population: a review of large-scale gridded population data products and their fitness for use. **Earth System Science Data**, v. 11, n. 3, p. 1385-1409, 2019.

LINARD, C.; TATEM, A. J.; GILBERT, M. Modelling spatial patterns of urban growth in Africa. **Applied Geography**, v. 44, p. 23-32, 2013.

LIVERMAN, D. *et al.* **People and pixels: linking remote sensing and social science**. Washington, DC: National Academies Press, 1998.

LLOYD, C. T.; SORICHETTA, A. I.; TATEM, A. J. High resolution global gridded data for use in population studies. **Scientific Data**, v. 4, n. 1, p. 1-17, 2017.

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.

MESA, S. **Silêncio administrativo**. La pobreza en el laberinto burocratico. Barcelona: Anagrama, 2019.

NADALIN, S. O. **A demografia numa perspectiva histórica**. Abep, 1994.

NAGLE, N. N. *et al.* Dasymetric modeling and uncertainty. **Annals of the Association of American Geographers**, v. 104, n. 1, p. 80-95, 2014.

NEVES, A. K. *et al.* Assessment of TerraClass and MapBiomias data on legend and map agreement for the Brazilian Amazon biome. **Acta Amazonica**, v. 50, p. 170-182, 2020.

ONU-HABITAT. **Diretrizes internacionais para planejamento urbano e territorial**. Nairóbi: ONU-Habitat, 2015.

ONU-HABITAT. **Nova agenda urbana**. Declaração de Quito sobre Cidades e Assentamentos Humanos Sustentáveis. ONU-Habitat, 2017.

OPENSHAW, S. Ecological fallacies and the analysis of areal census data. **Environment and Planning A**, v. 16, n. 1, p. 17-31, 1984.

POPGRID. **Leaving no one off the map: a guide for gridded population data for sustainable development**. POPGRID Data Collaborative, 2020. Disponível em: https://www.popgrid.org/sites/default/files/documents/Leaving_no_one_off_the_map.pdf. Acesso em: 10 jul. 2023.

REED, F. J. *et al.* Gridded population maps informed by different built settlement products. **Data**, v. 3, n. 3, p. 33, 2018.

ROBINSON, C.; HOHMAN, F.; DILKINA, B. A deep learning approach for population estimation from satellite imagery. *In: ACM SIGSPATIAL WORKSHOP ON GEOSPATIAL HUMANITIES*, 1. **Proceedings [...]**. New York: Association for Computing Machinery, 2017.

SCHIAVINA, M. *et al.* **GHS-POP R2023A – GHS population grid multitemporal (1975-2030)**. European Commission, Joint Research Centre (JRC), 2023. Disponível em: <http://data.europa.eu/89h/2ff68a52-5b5b-4a22-8f40-c41da8332cfe>. Acesso em: 20 jul. 2023.

SCHUG, F. *et al.* Gridded population mapping for Germany based on building density, height and type from Earth observation data using census disaggregation and bottom-up estimates. **PloS One**, v. 16, n. 3, article e0249044, 2021.

SEADE – Fundação Sistema Estadual de Análise de Dados. **Banco de Dados de Informações dos Municípios Paulistas**. São Paulo, 2023.

SEGAUD, M. **Antropologia do espaço: habitar, fundar, distribuir, transformar**. São Paulo: Sesc, 2016.

SEE, L. *et al.* Crowdsourcing, citizen science or volunteered geographic information? The current state of crowdsourced geographic information. **ISPRS International Journal of Geo-Information**, v. 5, n. 55, 2016.

SEMENOV-TIAN-SHANSKY, B. Russia: territory and population. A perspective on the 1926 Census. **Geographical Review**, v. 18, p. 616-640, 1928.

SHRYOCK, H. S.; SIEGEL, J. S.; LARMON, E. A. **The methods and materials of demography**. US Bureau of the Census, 1975.

SILVA, D. M. **VALiN-POP – uma grade populacional adaptada para a Região Metropolitana do Vale do Paraíba e Litoral Norte, SP**. Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais (Inpe), São José dos Campos, 2023.

SILVA, V. L.; REIS I. A.; REIS, E. A. **Modelos de regressão para estimação de populações via imagens de satélite**. Relatório Técnico. Belo Horizonte: Universidade Federal de Minas Gerais (UFMG), Departamento de Estatística, 2009. (Série Pesquisa RTP 03/2009). Disponível em: https://www.est.ufmg.br/portal/arquivos/rts/Relatorio_Tecnico_EST_publicado.p df. Acesso em: 03 mar. 2023.

STEVENS, F. R. *et al.* Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. **PloS One**, v. 10, n. 2, article e0107042, 2015.

TATEM, A. J. WorldPop, open data for spatial demography. **Scientific Data**, v. 4, n. 1, p. 1-4, 2017.

THOMSON, D. R. *et al.* How accurate are WorldPop-Global-Unconstrained gridded population data at the cell-level? A simulation analysis in urban Namibia. **PLoS One**, v. 17, n. 7, article e0271504, 2022.

TIECKE, T. G. *et al.* Mapping the world population one building at a time. **arXiv preprint** 1712.05839, 2017.

TOBLER, W. R. Cellular geography. In: GALE, S.; OLSSON, G. (Ed.). **Philosophy in geography**. Dordrecht: Riedel, 1979. p. 379-386.

TOBLER, W. *et al.* World population in a grid of spherical quadrilaterals. **International Journal of Population Geography**, v. 3, n. 3, p. 203-225, 1997.

TOBLER, W. *et al.* **The Global Demography Project**. Santa Barbara, CA: National Center for Geographic Information and Analysis. Department of Geography, University of California, 1995. (Technical Report TR-95-6).

UNITED NATIONS. **World population prospects 2022**. United Nations, Department of Economic and Social Affairs, Population Division, 2022.

WAN, H. *et al.* Population downscaling using high-resolution, temporally-rich US property data. **Cartography and Geographic Information Science**, v. 49, n. 1, p. 18-31, 2022.

WRIGHT, J. K. A method of mapping densities of population: with Cape Cod as an example. **Geographical Review**, v. 26, n. 1, p. 103-110, 1936.

YETMAN, G.; DEICHMANN, U.; BALK, D. Creating a global grid of human population. In: ANNUAL ESRI INTERNATIONAL USER CONFERENCE, 20. **Proceedings** [...]. San Diego, CA, 2000. p. 26-30.

ZHOU, Z.; WU, J.; TANG, W. Ensembling neural networks: many could be better than all. **Artificial Intelligence**, v. 137, n. 1-2, p. 239-263, 2002.

Sobre os autores

Diego Moreira Silva é mestre em Sensoriamento Remoto pelo Instituto Nacional de Pesquisas Espaciais (Inpe) e graduado em Geografia pela Universidade Federal Fluminense, polo de Campos dos Goytacazes – RJ.

Tathiane Mayumi Anazawa possui pós-doutorado pelo Instituto Nacional de Pesquisas Espaciais (bolsista Fapesp), com estágio pós-doutoral no Centro de Estudios Demográficos (CED), Universidad Autónoma de Barcelona. É doutora em Demografia pela Universidade Estadual de Campinas (Unicamp), mestre em Sensoriamento Remoto pelo Instituto Nacional de Pesquisas Espaciais (Inpe) e graduada em Ciências Biológicas pela Universidade Estadual de Ponta Grossa. Pesquisadora junto ao LiSS – Laboratório de investigação em Sistemas Socioambientais, do Inpe.

Silvana Amaral Kampel é doutora em Engenharia/Informação Espacial pela Escola Politécnica da Universidade de São Paulo (USP), mestre em Sensoriamento Remoto pelo Instituto Nacional de Pesquisas Espaciais (Inpe) e ecóloga pela Universidade Estadual Paulista (Unesp) de Rio Claro. Funcionária e pesquisadora do Inpe, na Divisão de Processamento de Imagens, Coordenação de Observação da Terra (OBT).

Flávia da Fonseca Feitosa é doutora em Geografia pela Rheinisch Friedrich-Wilhelms-Universität Bonn (Alemanha), mestre em Sensoriamento Remoto pelo Instituto Nacional de Pesquisas Espaciais (Inpe) e graduada em Arquitetura e Urbanismo pela Universidade Federal de Santa Catarina (UFSC). Professora do Bacharelado em Planejamento Territorial e do Programa de Pós-Graduação em Planejamento e Gestão do Território da Universidade Federal do ABC (UFABC).

José Irineu Rangel Rigotti é doutor e mestre em Demografia pelo Centro de Desenvolvimento e Planejamento Regional (Cedeplar) da Universidade Federal de Minas Gerais (UFMG) e graduado em Geografia pela UFMG. Professor associado II, do Departamento de Demografia, da Universidade Federal de Minas Gerais e pesquisador do Centro de Desenvolvimento e Planejamento Regional (Cedeplar).

Antonio Miguel Vieira Monteiro é doutor em Engenharia Eletrônica e Controle/Ciência da Computação pelo Centro de Ciências Espaciais da Escola de Engenharia e Ciências Aplicadas da Universidade de Sussex, em Brighton, Reino Unido, mestre em Computação Aplicada pelo Instituto Nacional de Pesquisas Espaciais (Inpe) e graduado em Engenharia Elétrica pela Universidade Federal do Espírito Santo (Ufes). Funcionário na Divisão de Processamento de Imagens (DPI) do Inpe. É orientador nos programas de Computação Aplicada, de Sensoriamento Remoto e de Ciência do Sistema Terrestre do Inpe. Coordenador do Programa Institucional Espaço e Sociedade do Inpe e do LiSS – Laboratório de Investigação em Sistemas Socioambientais na Área de Observação da Terra no Inpe.

Endereço para correspondência

Diego Moreira Silva

Instituto Nacional de Pesquisas Espaciais (Inpe)
Av. dos Astronautas, 1758, Jardim da Granja
12227-010 – São José dos Campos-SP, Brasil

Tathiane Mayumi Anazawa

Universidade Estadual de Campinas (Unicamp)
Rua da Reitoria nº 121, Cidade Universitária Zeferino Vaz, Barão Geraldo
13083-970 – Campinas-SP, Brasil

Silvana Amaral Kampel

Instituto Nacional de Pesquisas Espaciais (Inpe)
Av. dos Astronautas, 1758, Jardim da Granja
12227-010 – São José dos Campos-SP, Brasil

Flávia da Fonseca Feitosa

Universidade Federal do ABC (UFABC)
Av. dos Estados, Bangú, Santo André
09280-560, Santo André-SP, Brasil

José Irineu Rangel Rigotti

Universidade Federal de Minas Gerais (UFMG)
Av. Pres. Antônio Carlos, 6627, Pampulha
31270-901 – Belo Horizonte-MG, Brasil

Antonio Miguel Vieira Monteiro

Instituto Nacional de Pesquisas Espaciais (Inpe)
Av. dos Astronautas, 1758, Jardim da Granja
12227-010 – São José dos Campos-SP, Brasil

CRedit

Reconhecimentos: Não aplicável.

Financiamento: Este estudo foi financiado pela agência brasileira de Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (Capes), código financeiro 001.

Conflitos de interesse: Os autores atestam que não possuem qualquer interesse pessoal, comercial, acadêmico, político ou financeiro que possa gerar um conflito de interesses em relação ao manuscrito.

Aprovação ética: Os autores certificam que o trabalho não inclui seres humanos ou animais.

Disponibilidade de dados e material: Não aplicável.

Contribuições dos autores:

Diego Moreira Silva: investigação; metodologia; *software*; curadoria de dados; análise formal; validação; escrita – rascunho original; escrita – revisão e edição.

Tathiane Mayumi Anazawa: investigação; metodologia; supervisão; análise formal; validação; escrita – revisão e edição.

Silvana Amaral Kampel: validação; escrita – revisão e edição.

Flávia da Fonseca Feitosa: validação; escrita – revisão e edição.

José Irineu Rangel Rigotti: validação; escrita – revisão e edição.

Antonio Miguel Vieira Monteiro: metodologia; supervisão; análise formal; administração do projeto; aquisição de financiamento; escrita – rascunho original; escrita – revisão e edição.

Editores: Cassio Turra e Igor Cavallini Johansen

Abstract

In search of new demographic representations: the field of population grids in the era of machine learning

The population distribution on the Earth's surface reveals a variety of spatial patterns that reflect sociodemographic processes related to the historical-geographical dynamics that produced them. Population grids have gained prominence as a source of population data, involving estimates and distribution in small areas. Each population grid consists of cells of specific sizes, covering the entire globe or specific local territories. This work presents a commented literature review in the field of these population representations, specifically in the distribution and volume of the population, and the importance of spatial auxiliary variables, referred to here as symptomatic variables. These play a crucial role in building reality-based models, both locally and globally, using various methods, including Machine Learning. The main initiatives in the field, available global products, and the technical foundations of the main methodologies are also highlighted.

Additionally, the paper discusses limitations, precautions, and new opportunities resulting from the creation of these population grids.

Keywords: Population grids. Spatial distribution of the population. Machine learning. Spatial modeling. Symptomatic variables. Remote sensing. Census data.

Resumen

En busca de nuevas representaciones demográficas: El campo de estudios de las rejillas poblacionales en la era del aprendizaje automático

La distribución poblacional en la superficie terrestre revela una variedad de patrones espaciales que reflejan procesos sociodemográficos relacionados con las dinámicas histórico-geográficas que los produjeron. Las grillas poblacionales han cobrado protagonismo como fuente de datos, involucrando estimaciones y distribución en áreas pequeñas. Cada una de ellas está compuesta por celdas de tamaños específicos y abarcan la totalidad del globo o territorios locales específicos. Este trabajo presenta una revisión comentada de la literatura en el campo de estas representaciones poblacionales, específicamente en la distribución y el volumen de la población, y la importancia de las variables auxiliares espaciales, denominadas aquí *variables sintomáticas*. Estas desempeñan un papel crucial en la construcción de modelos basados en la realidad tanto local como global, utilizando diversos métodos, incluido el aprendizaje automático. También se destacan las principales iniciativas en el campo, los productos globales disponibles y las bases técnicas de las principales metodologías. Además, se discuten las limitaciones, precauciones y las nuevas oportunidades resultantes de la creación de estas grillas poblacionales.

Palabras clave: Grillas poblacionales. Distribución espacial de la población. Aprendizaje automático. Modelado espacial. Variables sintomáticas. Sensores remotos. Datos censales.

Recebido para publicação em 21/12/2023

Aceito para publicação em 19/08/2024