



**INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM SENSORIAMENTO REMOTO**

**A ACURÁCIA DA GRADE POPULACIONAL GLOBAL *WORLDPOP*
PARA AS ÁREAS RURAIS DA AMAZÔNIA PARAENSE:
um exercício com a Região do Baixo Tocantins.**

Gustavo Piva Lopes Salgado

Monografia apresentada como requisito parcial da disciplina de Introdução ao Geoprocessamento, ministrada pela docente Prof^a Dr^a Silvana Amaral Kampel e pelo Prof^o Dr^o Marcos Adami.

RESUMO

A disponibilidade de dados demográficos é fundamental para uma variedade de aplicações, mas muitas áreas do mundo ainda carecem de informações precisas e atualizadas e o Brasil, que possui a maior floresta tropical do mundo, possui deficiências de estimativas populacionais entre os censos, principalmente em áreas rurais. Grades Populacionais Globais (GGP), como o WorldPop Project, fornecem uma fonte de dados para o planejamento e tomada de decisões em áreas onde as informações populacionais são escassas, no entanto, poucos estudos avaliam a qualidade dessas grades. O trabalho propõe um exercício prático de validação da precisão dos dados do WorldPop para a região rural do Baixo-Tocantins, no Pará, comparando-os com dados oficiais do IBGE. Essa análise é relevante devido à lacuna na disponibilidade de dados demográficos atualizados no Brasil. Utilizando-se de métricas estatísticas globais e espaciais o trabalho buscou mensurar a acurácia das estimativas da WorldPop com os dados oficiais do IBGE. Os resultados obtidos, apesar de demonstrarem certa autocorrelação espacial positiva entre as estimativas do WorldPop e as medidas reais do IBGE, evidenciaram uma autocorrelação espacial fraca e insignificante nos erros relativos das estimativas, sugerindo que as diferenças entre as estimativas e as medidas reais não apresentam uma estrutura espacial consistente. Esses resultados ressaltam a importância da avaliação da qualidade e precisão dos dados populacionais, especialmente em áreas rurais com deficiências nas estimativas.

Palavras-Chave: Geoprocessamento, Distribuição de Populações Rurais, Grades Populacionais Globais, Projeto *WorldPop*, Baixo-Tocantins, Amazônia.

ABSTRACT

The availability of demographic data is crucial for a variety of applications, but many areas of the world still lack accurate and up-to-date information, and Brazil, which has the largest tropical forest in the world, faces deficiencies in population estimates between censuses, particularly in rural areas. Global Gridded Population (GGPs), such as the WorldPop Project, provide a data source for planning and decision-making in areas where population information is scarce. However, few studies assess the quality of these grids. This work proposes a practical exercise to validate the accuracy of WorldPop data for the rural region of Baixo-Tocantins in Pará, Brazil, comparing it with official data from Brazilian Institute of Geography and Statistics (IBGE). This analysis is relevant due to the gap in the availability of updated demographic data in Brazil. Using global and spatial statistical metrics, the study aimed to assess the accuracy of WorldPop estimates compared to official data from IBGE. The results obtained, despite showing some positive spatial autocorrelation between WorldPop estimates and IBGE actual measurements, revealed a weak and insignificant spatial autocorrelation in the relative errors of the estimates. This suggests that the differences between the estimates and actual measurements do not exhibit a consistent spatial pattern. These findings highlight the importance of evaluating the quality and accuracy of population data, particularly in rural areas with deficiencies in estimates.

Keywords: Geoprocessing, Rural Population Distribution, Global Gridded Population, *WorldPop* Project, Baixo-Tocantins, Amazon Forest.

LISTA DE FIGURAS

	<u>Pág.</u>
Figura 1 - Fluxograma de Procedimentos	12
Figura 2. Conjuntos de dados de distribuição populacional do IBGE e WorldPop	13
Figura 3. Índice de Moran bivariado local - Estimativas WorldPop x Valores IBGE.....	17
Figura 4. Mapas de agrupamento e de significância para Índice de Moran Bivariado Local - WorldPop x IBGE.....	18
Figura 5. Índice de Moran univariado local - Erro Relativo WorldPop	18
Figura 6. Mapa de agrupamento e mapa de significância estatística para Índice de Moran Univariado Local - Erro Relativo WorldPop.....	19

LISTA DE TABELAS

	<u>Pág.</u>
Tabela 1. Índices estatísticos para avaliação de erros de estimativas	14
Tabela 2. Resultados dos Índices estatísticos entre a grade WorldPop e IBGE.....	16

SUMÁRIO

	Pág.
INTRODUÇÃO	6
WORLDPOP PROJECT	8
GRADE ESTATÍSTICA DO IBGE	8
ÁREA DE ESTUDO	9
MATERIAIS E MÉTODOS	11
RESULTADOS	15
CONSIDERAÇÕES FINAIS	20
REFERÊNCIAS BIBLIOGRÁFICAS	21

INTRODUÇÃO

O mundo passa hoje por desafios que ameaçam o bem estar geral da humanidade e até mesmo sua existência como espécie, sendo a maior parte deles gerados pelas próprias atividades humanas e seus modelos de reprodução social: as já vigentes mudanças climáticas; a perda de biodiversidade ecossistêmicas; escassez de recursos naturais; a pobreza; a desigualdade socioeconômica; violência; guerras; migrações em massa; pandemias, entre tantas outras, são os grandes desafios que o planeta enfrenta hoje em dia e que merecem atenção especial para os formuladores de teorias científicas de qualquer área de estudo e para os tomadores de decisão.

Segundo Archila Bustos et. al. (2020) em estudos de população é fundamental compreender a distribuição da população humana para enfrentar os desafios globais atuais, mas as informações sobre a distribuição populacional são geralmente disponibilizadas de forma agregadas em áreas administrativas, como é o caso dos dados censitários, que pressupõem populações distribuídas de maneira uniforme e podem ser de difícil utilização quando se quer analisar fenômenos irregulares que não seguem os mesmos limites administrativos.

Para atenuar o problema acima, nas últimas décadas foram desenvolvidos projetos de modelagem da distribuição populacional em nível de grades, permitindo uma escala e formato mais úteis, que são chamados de Grades Populacionais Globais, ou GGP (*Global Gridded Population*) . Eles ajudaram a reduzir problemas estatísticos, como o da unidade de área modificável (MAUP - *Modifiable Areal Unit Problem*)¹ e aumentaram a disponibilidade de dados para estudos que geralmente requerem dados desagregados como entrada. (ARCHILA BUSTOS et. al., 2020)

Com projetos disponibilizados desde o início da década de 1990, o uso das grades populacionais globais aumentou significativamente na década de 2010, por causa da popularização do uso de cinco conjuntos de dados importantes por pesquisadores e em tomada de decisões de órgãos executivos, provavelmente porque estão disponíveis gratuitamente *online* (ARCHILA BUSTOS et. al., 2020). São eles:

¹ As diferentes formas de desagregação de dados podem levar a resultados diferentes, influenciando as conclusões e interpretações de um estudo. Isso ocorre porque a escolha das unidades de agregação pode afetar a variabilidade dos dados e a distribuição espacial das observações.

Gridded Population of the World (GPW); Global Rural Urban Mapping Project (GRUMP); Global Human Settlement Layer-Population (GHS-POP); LandScan Global Population Database e WorldPop Project.

As aplicações de uso das grades populacionais globais podem ser encontradas em diversos campos, desde o gerenciamento de desastres naturais; acompanhamento de riscos e vulnerabilidades; estudos sobre mudanças climáticas e saúde pública, entre outros. Entretanto, poucos estudos demográficos *stricto sensu* fazem uso desse tipo de dado, porque, normalmente, não se pode avaliar com confiabilidade quão precisos são os resultados das distribuições populacionais. (HALL, STROH & PAYA, 2015)

Realizados usualmente de 10 em 10 anos no Brasil, o intervalo entre os censos pode ser longo demais para capturar mudanças demográficas e sociais em curso, especialmente em regiões dinâmicas e populações em rápida expansão, como na região amazônica. Além desse problema inerente aos estudos da demografia brasileira, nos últimos anos observou-se uma lacuna significativa na disponibilidade de dados demográficos atualizados devido ao atraso na execução do censo brasileiro por causa da pandemia de covid19 e do negacionismo científico do então governo federal.

A falta de informações atualizadas e de qualidade pode ser problemática para governos, pesquisadores e planejadores que dependem desses dados para tomar decisões e desenvolver políticas. Diante dessa situação, é fundamental explorar novas alternativas para preencher essa lacuna e obter estimativas demográficas mais precisas e abrangentes. Pensando nisso, este trabalho buscou verificar como o uso das Grades Estatísticas Populacionais pode refletir a distribuição populacional de seres humanos em áreas rurais da Amazônia paraense, com base em uma grade populacional oficial de referência. Para isso, será realizado um exercício prático para as porções rurais na Região do Baixo-Tocantins, Estado do Pará, a partir de uma validação de acurácia dos dados da *WorldPop* com os dados de referência oficiais do IBGE. Essa grade foi escolhida porque seu ano base de projeção é o mesmo da grade oficial do governo brasileiro, que é 2010, o que dispensa qualquer tipo de interpolação de dados e possíveis acúmulos de erros. Seus dados foram comparados com a base de grades estatísticas do IBGE para o Censo de 2010 a fim de avaliar a qualidade e precisão do conjunto de dados.

WORLDPOP PROJECT

O projeto *WorldPop* é desenvolvido na Escola de Geografia e Ciências Ambientais da Universidade de Southampton, no Reino Unido. Criado a partir de uma combinação dos projetos de grades populacionais AfriPop, AmeriPop e AsiaPop em 2013, o WorldPop participa de projetos demográficos geoespaciais com governos e instituições em países de baixa e média renda, além de colaborações com organizações parceiras, provedores de dados comerciais e outras organizações internacionais de desenvolvimento e o oferecimento de treinamentos em modelagem populacional para ministérios da saúde e escritórios nacionais de estatística em países de baixa e média renda, para apoiar pesquisas em saúde e demografia. (WORLDPOP, 2023)

Os conjuntos de dados da WorldPop são criados usando redistribuição dasimétrica de dados censitários baseada em *Random-Forest*, que é um modelo computacional de aprendizado de máquinas que gera previsões de densidade populacional com base em dados auxiliares, também chamados de covariáveis, como: uso do solo; dados de altimetria e estimativas de declive; luzes noturnas; clima; infraestruturas de estradas; vias navegáveis; povoados; áreas de reservas e instalações públicas como escolas, hospitais e clínicas de saúde e podem variar por país com base na disponibilidade dos dados e na importância relativa para a estimativa populacional em cada localização. (ARCHILA BUSTOS et al. 2020)

Segundo Archila Bustos et. al. (2020) a metodologia *Random-Forest* da WorldPop é considerada superior a outras metodologias porque incorpora muitos conjuntos de dados auxiliares com pouco ajuste ou supervisão, mas que possuem a limitação dos dados à faixa de densidades populacionais das entradas. Isso pode ter um efeito em predições de áreas administrativas maiores, onde a população mensurada estará concentrada em um determinado local adjacente às estradas.

GRADES ESTATÍSTICAS DO IBGE

A Grade Estatística do Censo Demográfico de 2010, elaborada pelo Instituto Brasileiro de Geografia e Estatística (IBGE), é uma ferramenta utilizada para a organização e apresentação dos dados coletados durante o Censo 2010. Composta

por células, que são áreas de coleta de dados geográficos e estatísticos delimitadas por linhas de latitude e longitude, cada uma **dela** possui uma área aproximada de 0,25 km², e são identificadas por um código numérico único. (IBGE, 2023)

A Grade Estatística permite a produção de informações estatísticas de diferentes tipos, tais como dados sobre a população, habitação, educação, trabalho e rendimento, entre outros. Além disso, ela é utilizada como referência espacial para outras pesquisas e estudos realizados pelo IBGE e por outros órgãos públicos e privados. (IBGE, 2023)

A Grade Estatística do Censo Demográfico de 2010 é uma ferramenta importante para a produção e análise de informações estatísticas sobre a população humana do território brasileiro. Com ela, é possível obter dados precisos e atualizados que auxiliam na formulação e implementação de políticas públicas e no planejamento de atividades econômicas e sociais em todo o país, apresentando-se assim como o parâmetro ótimo de análise de acurácia para qualquer grade estatística que proponha-se a estimar valores demográficos do território nacional.

ÁREA DE ESTUDO

O Brasil é o país que concentra a maior parte da maior floresta tropical do mundo: a floresta amazônica. **Uma importante parcela** dos povos que habitam essa floresta estão na microrregião de Cametá, no Pará, também conhecida como Baixo-Tocantins, por **ser a região do local da jusante do rio** Tocantins. (ALMEIDA, 2010)

Localizada entre a Amazônia Central e a Amazônia Oriental, a microrregião do Baixo-Tocantins está no nordeste da Amazônia Paraense e é formada pelos seguintes municípios: Abaetetuba, Acará, Baião, Barcarena, Cametá, Igarapé Miri, Moju, Tailândia, Oeiras do Pará, Limoeiro do Ajuru e Mocajuba.

Com uma cobertura vegetal classificada como floresta equatorial densa, a região do Baixo-Tocantins possui atividades econômicas ligadas à riqueza natural local, como a pesca, agricultura familiar e extração do açaí, mas também, devido ao grande potencial hídrico da Bacia do Tocantins-Araguaia, constituiu-se como uma área de interesse estratégico nacional, com enfoque em transporte e geração de energia. Um exemplo desse potencial já em aproveitamento pelo planejamento governamental e por interesses econômicos é a Usina Hidrelétrica de Tucuruí, construída em 1984 e que se

mantém ativa até hoje. A usina trouxe certo desenvolvimento econômico e integração nacional à região, mas também gerou impactos provocados por sua barragem, como a inundação de vasta área de floresta; migração forçada de populações indígenas; não indenização de famílias deslocadas; redução do volume do pescado e poluição; erosão do leito e das margens do rio e elevado índice de malária. (ALMEIDA, 2010)

Segundo Almeida (2010) a exploração do cacau e da seringueira configuraram a cena econômica até meados da década de 1970, seguida da exploração madeireira, que antecipou a monocultura da pimenta-do-reino, duas atividades **danosas ao equilíbrio ambiental**. Nos dias atuais, as atividades de agricultura e do extrativismo regem a economia local. A **renda agrícola advinda da agricultura e do extrativismo responde por mais de 60% da economia dos municípios da região do Baixo-Tocantins**.

Com a agricultura e o extrativismo encabeçando a atividade econômica local, a presença da população **rural ainda é muito forte na região**, o Baixo-Tocantins apresenta **duas dinâmicas distintas para a distribuição da população rural: terra firme e a região das ilhas**. Na primeira **as áreas de terra firme desflorestadas** são ocupadas por populações que trabalham na agricultura tradicional de corte e queima, onde basicamente se cultiva mandioca, o principal produto dessa zona. Na segunda, a população tem a palmeira do açaí como a principal atividade produtiva, mas também trabalham com uma grande incidência de buritizais. (ALMEIDA, 2010)

Tendo em vista as condições de carência estrutural e de serviços públicos, de desequilíbrios sócio-econômicos e ecológicos, bem como outros impactos sociais e ambientais que se prenunciam para as populações locais de agricultores, pescadores e extrativistas, esta microrregião foi escolhida pela importância de se entender a dinâmica da população rural amazonense frente **às diversas que as ameaçam** e a necessidade de se dispor de opções de estimativas demográficas minimamente confiáveis entre censos para desenvolver estudos e políticas, o que será verificado para a grade populacional WorldPop, através de métodos de validação de acurácia.

MATERIAIS E MÉTODOS

Para a execução do exercício prático foram utilizados como dados de entrada um dado de avaliação e um dado de validação, ou padrão “ouro”. O parâmetro oficial e padrão “ouro” de dados demográficos de áreas rurais da Amazônia paraense foram obtidos do censo geral IBGE 2010, no formato vetorial de Grades Estatísticas - número id75 que é a referência para a região do **baixo-tocantins**. **Ele** estava no sistema de referência EPSG 4674, que é o SIRGAS 2000, e em coordenadas geográficas. Também do IBGE foi utilizado o arquivo vetorial de municípios do estado do Pará, que foi posteriormente filtrado para apenas os municípios da região do baixo-tocantins, também em SIRGAS 2000 e em coordenadas geográficas.

O dado de avaliação escolhido para a região em questão foi o do projeto WorldPop e foi obtido pelo próprio site do projeto em arquivo raster e denominado “bra_2010_1km_ASCII_XYZ”, que é o arquivo referência para o Brasil no ano de 2010. Ele estava em sistema de referência EPSG 4326, que é o WGS84 e também em coordenadas geográficas. Como as diferenças entre SIRGAS2000 e WGS84 são negligenciáveis em muitas aplicações práticas que não exigem alta precisão geodésica, então foi considerado os dois como equivalentes para esse trabalho.

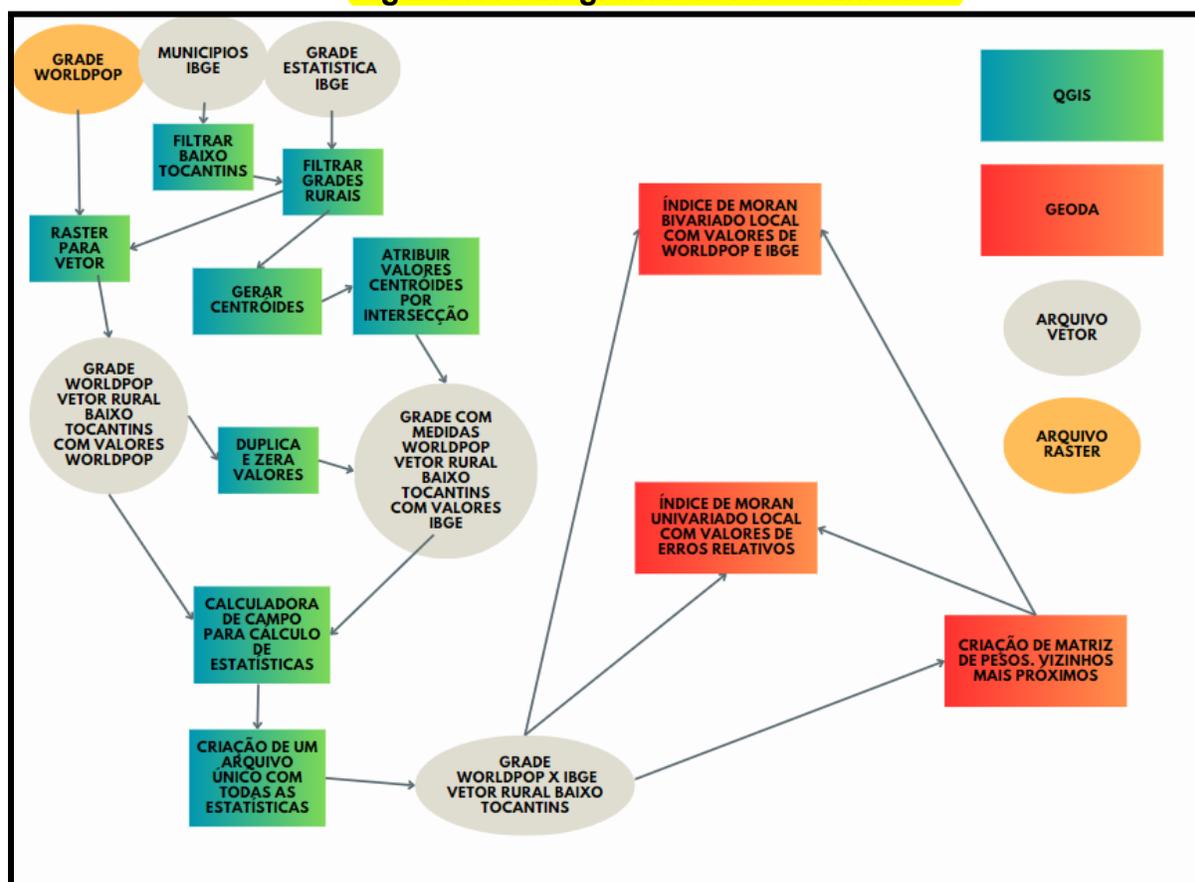
Os sistemas de informações geográficas utilizados foram o Quantum GIS (QGIS) e o GeoDa. O primeiro para manipulações, transformações e estatísticas básicas e o segundo para análise de dados espaciais.

Na etapa de pré-processamento dos dados, o dado **da** WorldPop que estava em arquivo raster foi transformado em arquivo vetorial shapefile pelo software QGIS a fim de reduzir seu tamanho e facilitar as etapas de geoprocessamento. Feito isso, tanto o arquivo vetorial da WorldPop quanto o arquivo vetorial do IBGE foram recortados pela **“máscara” do arquivo IBGE de municípios da região.**

O arquivo do IBGE foi então filtrado pelo tamanho de suas grades: **grades cujos lados representam 250m** foram retiradas porque abrangem aglomerados urbanos. Então só restaram **as grades de 1km de lado**, que abrangem as áreas rurais que o trabalho buscou avaliar. Após isso **o a** grade da WorldPop também foi recortada pela grade estatística do IBGE a fim de retirar as áreas urbanas.

O valor total da população rural para a região, antes de continuar o processamento dos dados, foi 301.537 para a grade estatística do IBGE e 377.484 para WorldPop. Uma diferença de 25,19% de superestimação.

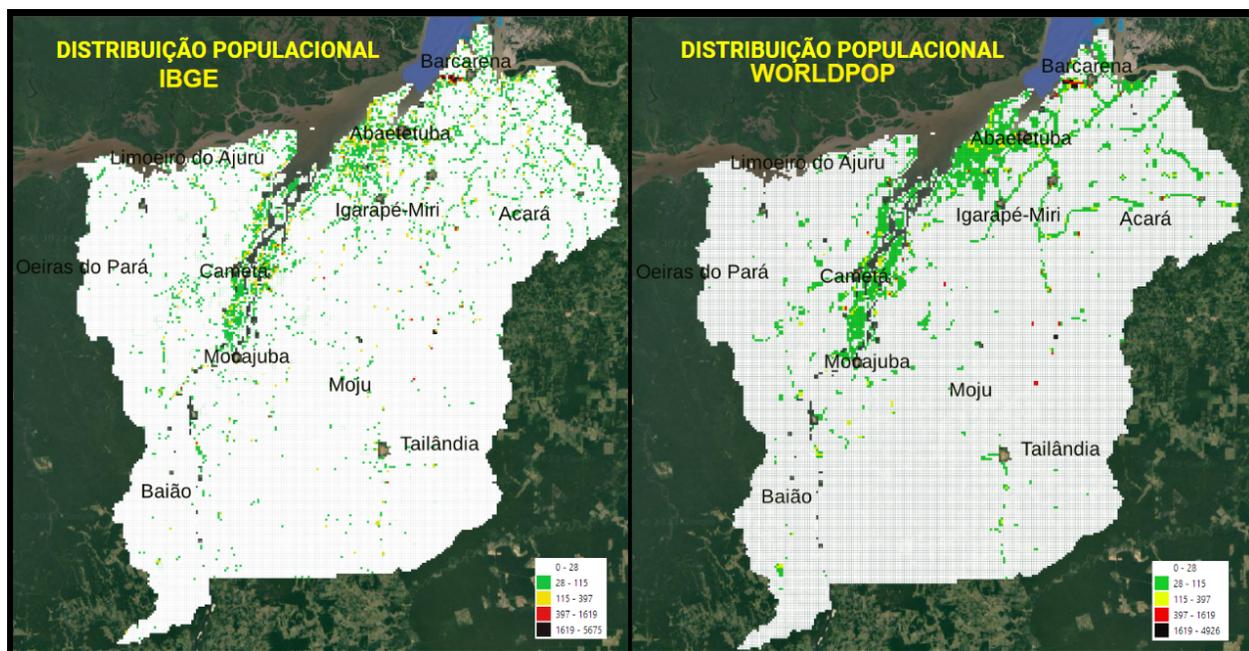
Figura 1. Fluxograma de Procedimentos



FONTE: Autoria própria.

A grade da WorldPop foi então duplicada e sua cópia teve os valores de população zerados. Ela receberia a partir daí os valores referência do IBGE e isso foi necessário porque ambas não estavam gradeadas com as mesmas medidas, o que inviabilizaria a análise comparativa. O próximo passo então foi gerar os centróides dos polígonos de cada grade do arquivo referência das grades do IBGE e atribuir os valores dos centróides ao arquivo cópia do WorldPop por meio de análise por intersecção. A população total do IBGE dentro de cada célula da cópia da grade WorldPop foi então calculada como a soma dos centróides da grade do IBGE que intersectam cada célula da cópia da grade WorldPop (Figura 2.).

Figura 2. Conjuntos de dados de distribuição populacional para o Baixo-Tocantins - IBGE e Grade Populacional Global WorldPop.



FONTE: Autoria própria com dados de WorldPop e IBGE.

Com os dois arquivos igualmente gradeados segundo a grade original da WorldPop, mas com valores do IBGE e da WorldPop em cada uma, foi então possível iniciar uma etapa de análise comparativa de grades populacionais conforme realizado por Archila Bustos et al. (2020). Os autores realizaram a avaliação da acurácia espacial das estimativas populacionais de 5 conjuntos de grades populacionais globais (WorldPop incluso) no território sueco e com um padrão ouro de validação dado pelo Instituto de Estatística da Suécia (Statistiska centralbyrån, SCB).

Reproduziu-se a ideia e os índices utilizados por Archila Bustos et. al. (2020), mas adaptou-se segundo as condições e objetivos deste trabalho, ao utilizar apenas a grade WorldPop e o padrão ouro de dados para avaliação tinham que ser os dados oficiais brasileiros. Deste modo, utilizou-se medidas estatísticas (Tabela 1.) para avaliar a precisão dos conjuntos de dados da WorldPop em relação à grade de referência do IBGE. Isso foi feito a partir do uso da calculadora de campo do QGIS.

Tabela 1. Índices estatísticos para avaliação de erros de estimativas

NOME	SIGLA	FÓRMULA	OBJETIVO
Erro Absoluto	EA	$EA = X - \bar{X} $	avalia a precisão da estimativa em termos de magnitude
Erro Relativo	ER	$ER = (X - Y / Y) \times 100$	compara os erros em relação ao tamanho dos valores de referência
Erro Percentual Médio Absoluto	EPMA	$EPMA = (1/n) \times \sum((X - Y) / Y) \times 100$	avalia o desempenho médio da estimativa em relação aos valores de referência
Erro Quadrático Médio	EQM	$EQM = (1/n) \times \sum((X - Y)^2)$	mede a dispersão dos erros e penaliza erros maiores de forma mais significativa
Raiz do Erro Quadrático Médio	RMSE	$RMSE = \sqrt{EQM}$	estima o erro médio em unidades originais
Coefficiente de Determinação	R ²	$R^2 = 1 - (\sum(X - Y)^2 / \sum(Y - \bar{Y})^2)$	avalia o grau de ajuste de um modelo de regressão aos dados e a qualidade da estimativa
Coefficiente de Correlação de Pearson	r	$r = \sum((X - \bar{X})(Y - \bar{Y})) / \sqrt{(\sum(X - \bar{X})^2 * \sum(Y - \bar{Y})^2)}$	verifica a correlação entre os valores estimados e os valores de referência

FONTE: Autoria própria.

Os resultados dos índices ~~acima calculados no software QGIS~~ foram então compilados juntamente com os dados de entrada da WorldPop e do IBGE em apenas **um arquivo vetorial denominado "worldpop_x_ibge"** e este foi então utilizado para a etapa de análise de correlação espacial feita no software GeoDa.

No software GeoDa foi realizada o cálculo do Índice de Moran local, que é um índice que avalia a autocorrelação espacial entre conjuntos de dados, onde inúmeras permutações aleatórias dos valores dos atributos de cada área calcularam o Índice de Moran para cada uma destas permutações aleatórias e **associaram** isso a um mapa de peso estatístico (mapa de pseudo significância / p-valor), que foi dado pela própria distribuição de polígonos das grades e cujo nível de confiança foi de 95% de probabilidade de um intervalo pseudo normal conter o valor encontrado para cada célula (não há autocorrelação espacial / aleatoriedade) e 5% de probabilidade de o valor encontrado em cada célula não ser contido dentro da normalidade (há

autocorrelação espacial / dependência espacial). Não foi necessário executar nenhum tipo de fórmula matemática porque o software GeoDa executa tudo isso de forma automática.

Foram realizados dois testes para o índice de Moran local: um bivariado entre os valores de população estimados pelo WorldPop com os valores de referência da grade estatística do IBGE para entender a associação dos dados da WorldPop em termos de magnitude com os dados do IBGE; outro univariado apenas com o índice ER, a fim de entender um pouco melhor se existe algum tipo de associação espacial para os erros de estimativas da WorldPop.

RESULTADOS

A população rural total das grades estatísticas do IBGE para a região do Baixo Tocantins em 2010 foi calculada como sendo de 345.049 pessoas, enquanto que a população estimada para a região foi de 401.828,45 pessoas para o mesmo ano, com uma diferença de 16,46% de superestimação por parte da WorldPop. Devido ao uso de valores decimais pelo projeto WorldPop, nenhuma célula obteve valores estimados iguais aos valores de referência do IBGE, portanto a porcentagem de células identificadas corretamente foi zero.

Tabela 2. Resultados dos Índices estatísticos entre a grade WorldPop e IBGE

NOME	SIGLA	RESULTADO
Erro Absoluto	EA	56779,45 pessoas
Erro Absoluto Médio	EA \bar{X}	1,86 pessoas
Erro Relativo	ER	16,46%
Erro Relativo Médio	ER \bar{X}	38,32%
Erro Percentual Médio Absoluto	EPMA	136,91%
Erro Quadrático Médio	EQM	6702,82 pessoas
Raiz do Erro Quadrático Médio	RMSE	81,87 pessoas
Coefficiente de Determinação	R ²	0,3655
Coefficiente de Correlação de Pearson	r	0,605

FONTE: Autoria própria.

Ao analisar-se os indicadores de erro, constatou-se que o Erro Absoluto médio foi de 1,86 pessoas, com um desvio padrão de 81,85 pessoas. Esses valores sugerem uma certa variabilidade nos erros das estimativas da WorldPop em relação aos números oficiais do IBGE. Além disso, o coeficiente de variação de cerca de 43,86 pessoas indica uma dispersão considerável nos erros absolutos.

O Erro Percentual Médio Absoluto (EPMA) foi de 136,91%, o que demonstra que, em média, as estimativas da WorldPop têm um desempenho muito aquém do desejado, com uma diferença substancial em relação aos valores oficiais do IBGE.

O Erro Quadrático Médio (EQM) de 6.702,82 pessoas revela uma grande dispersão entre as estimativas da WorldPop e os valores do IBGE. Esse resultado indica que as diferenças entre os dois conjuntos de dados são significativas.

A Raiz do Erro Quadrático Médio (RMSE) de 81,87 pessoas representa a magnitude média dessas diferenças em termos absolutos. Essa medida é útil para compreender o tamanho médio dos erros nas estimativas da WorldPop em relação aos dados do IBGE.

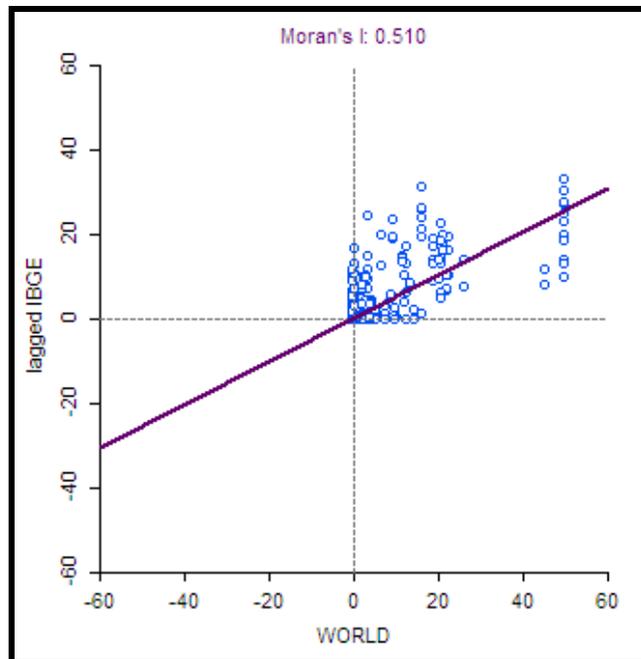
A análise do Coeficiente de Correlação de Pearson revelou um valor de 0,605, o que significa que as duas variáveis estão positivamente relacionadas. No entanto, essa relação não é muito forte, sugerindo que as estimativas da WorldPop não refletem completamente os valores oficiais fornecidos pelo IBGE.

O resultado do Coeficiente de Determinação (R^2) constatou que o modelo de correlação linear explica apenas 36,55% da variância dos valores dos erros da WorldPop em relação aos dados oficiais do IBGE, ou seja, cerca de 63,45% da variabilidade das estimativas da WorldPop ainda permanece sem explicações pelo modelo. Esse valor indica que a correlação entre as estimativas das duas fontes existe, mas ainda é moderada.

Por fim, para verificar a dependência espacial dos erros de estimativas da WorldPop em relação a grade estatística do IBGE, foi utilizado o Índice de Moran Univariado Local, que avalia uma determinada variável em cada unidade espacial com os valores dessa mesma variável nas unidades espaciais vizinhas. O Índice de Moran é baseado na ideia de autocorrelação espacial, que indica se há uma tendência de valores semelhantes se agruparem em certas áreas geográficas, formando clusters, ou

se há uma distribuição aleatória desses valores. O método utilizado para ponderar o peso espacial das variáveis foi o k-nearest neighbours (vizinho mais próximo), que considerou os 8 polígonos mais próximos do polígono em questão, independentemente da direção e da distância. O resultado pode ser visualizado nas figuras 3 e 4.

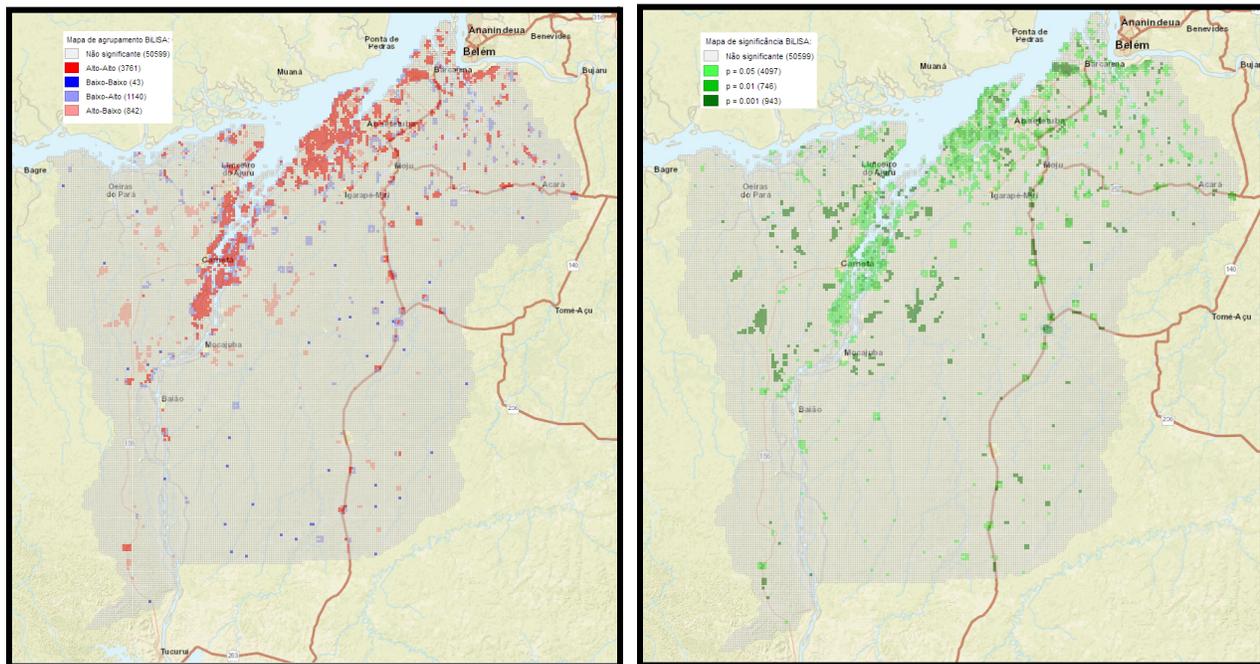
Figura 3. Índice de Moran bivariado local - Estimativas WorldPop x Valores IBGE



FONTE: Software GeoDa

Um índice de Moran bivariado local de 0,510 entre o conjunto de estimativas WorldPop e o conjunto de medidas do IBGE indica a presença de autocorrelação espacial positiva entre esses dois conjuntos de dados. Isso significa que as estimativas e as medidas reais apresentam uma tendência de se agrupar espacialmente, ou seja, áreas com estimativas altas tendem a estar próximas a áreas com medidas reais altas, e áreas com estimativas baixas tendem a estar próximas a áreas com medidas reais baixas.

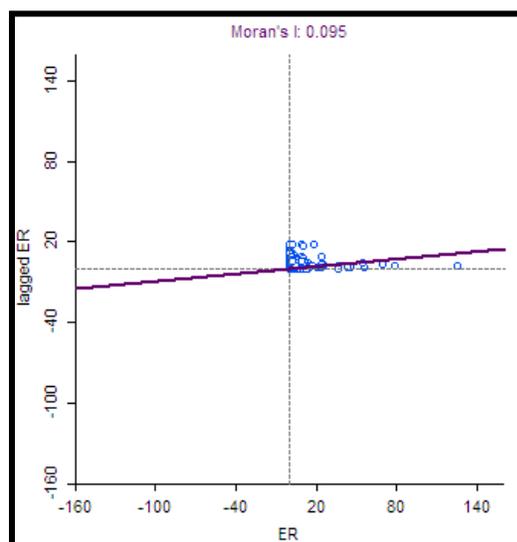
Figura 4. Mapa de agrupamento e mapa de significância estatística para Índice de Moran Bivariado Local - WorldPop x IBGE



FONTE: Software GeoDa

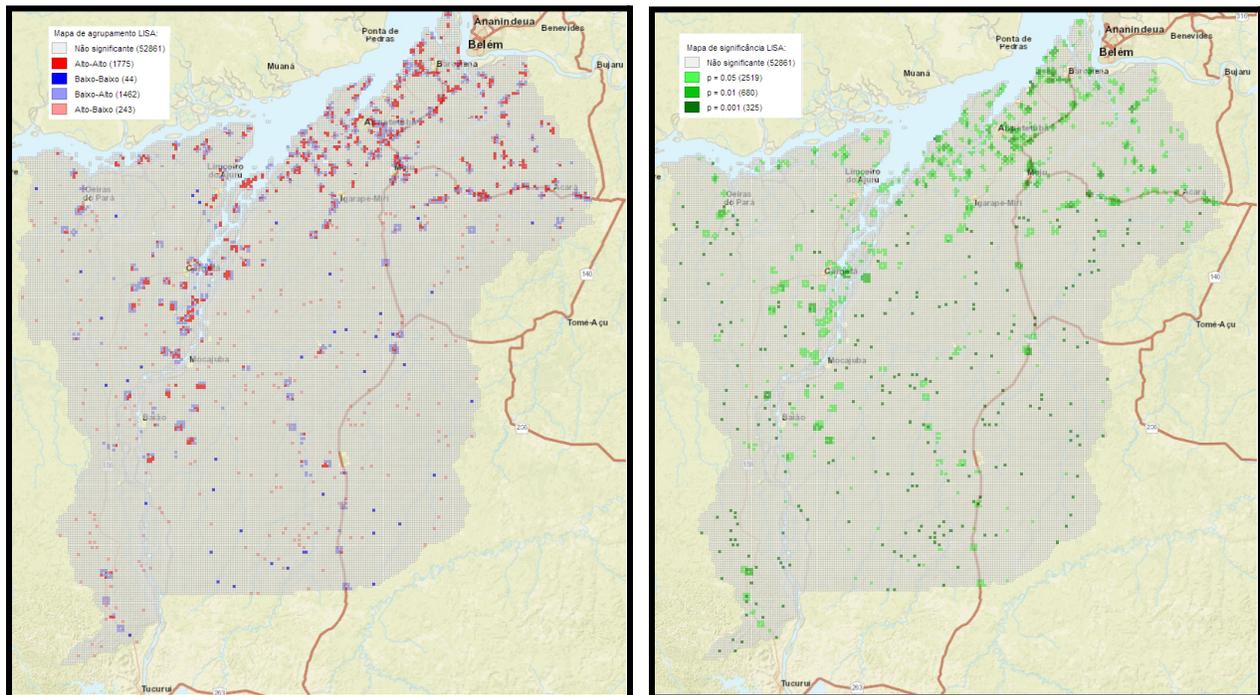
Para avaliar a dependência espacial dos erros de estimativas populacionais referente às próprias magnitudes de valores de referência oficiais do IBGE, foi utilizado o Índice de Moran Univariado local com a coluna de valores que indicavam o erro relativo em cada célula da grade. O resultado pode ser visualizado nas figuras 5 e 6.

Figura 5. Índice de Moran univariado local - Erro Relativo WorldPop



FONTE: Software GeoDa

Figura 6. Mapa de agrupamento e mapa de significância estatística para Índice de Moran Univariado Local - Erro Relativo WorldPop



FONTE: Software GeoDa

Um índice de Moran Local obtido foi de 0,095, que indica que a tendência de valores semelhantes de erros se agruparem no espaço é fraca ou insignificante. Isso significa que não há uma tendência clara de agrupamento espacial dos erros de estimativas e autocorrelação espacial fraca sugere os erros não apresentam uma estrutura espacial consistente, ou seja, áreas com altos valores de erro relativo não estão necessariamente próximas a outras áreas com altos valores de erro relativo, e o mesmo se aplica a baixos valores de erro relativo. Isso pode ser resultado de fatores aleatórios ou de uma variação espacialmente heterogênea nas diferenças entre as estimativas e as medidas reais.

CONSIDERAÇÕES FINAIS

A grande concentração de erros relativos na região das Ilhas e em assentamentos próximos a áreas urbanas ao longo de estradas e vias navegáveis indicam que a precisão das estimativas da WorldPop é menor nessas áreas, provavelmente devido a desafios na identificação e caracterização de populações rurais em locais com maior complexidade espacial.

Os resultados obtidos dos indicadores estatísticos na maior parte do Baixo-Tocantins indicaram que as estimativas da WorldPop apresentaram baixa acurácia na mensuração da população rural nessa região, sugerindo que os valores estimados divergem consideravelmente dos valores oficiais do IBGE, o que torna a utilização desse tipo de dado ainda uma opção duvidosa, principalmente para pesquisadores e tomadores de decisão que almejam uma precisão maior no embasamento de seus dados.

Uma amostragem local para o treinamento do modelo de *machine learning* da WorldPop, baseada em *Random Forest*, poderia trazer resultados mais precisos, considerando as características e peculiaridades da região.

A mensuração da acurácia de outros projetos de grades populacionais poderia proporcionar um resultado mais assertivo e/ou indicar particularidades de erros em cada modelo, pois diferentes fontes de dados e metodologias podem gerar estimativas populacionais com diferentes níveis de precisão. Seria importante avaliar e comparar as diferentes grades populacionais disponíveis para determinar qual delas oferece a melhor correspondência com os valores oficiais do IBGE.

REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA, R. Amazônia, Pará e o mundo das águas do Baixo Tocantins. **Estudos Avançados**, São Paulo, v. 24, n. 68, p. 269-288, 2010. DOI: 10.1590/S0103-40142010000100020. Disponível em: <<https://www.scielo.br/j/ea/a/SjrQ9BqjDpRtD4ndNnNfxcM/?lang=pt>>. Acesso em: 14 mai. 2023.

ARCHILA BUSTOS, M. F. et al. A pixel level evaluation of five multitemporal global gridded population datasets: a case study in Sweden, 1990–2015. **Popul Environ**, v. 42, p. 255-277, 2020. Disponível em: <<https://doi.org/10.1007/s11111-020-00360-8>>. Acesso em: 14 abr. 2023.

HALL, Ola; STROH, Emilie; PAYA, Fredy. From Census to Grids: Comparing Gridded Population of the World with Swedish Census Records. **The Open Geography Journal**, 2012, 5, 1-5, 2015. Disponível em: <<https://benthamopen.com/contents/pdf/TOGEOGJ/TOGEOGJ-5-1.pdf>>. Acesso em: 12 mai. 2023.

IBGE - Instituto Brasileiro de Geografia e Estatística. **Grade Estatística do Censo 2010**. Disponível em: <https://geoftp.ibge.gov.br/recortes_para_fins_estatisticos/grade_estatistica/censo_2010/grade_estatistica.pdf>. Acesso em: 14 abr. 2023.

IBGE - Instituto Brasileiro de Geografia e Estatística. **Malhas Territoriais**. [Recurso eletrônico]. Data de acesso: 13/05/2023. Disponível em: <<https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/15774-malhas.html?=&t=downloads>>.

WORLDPOP. WorldPop Population Database. **University of Southampton and Flowminder Foundation**, 2021. Disponível em: <<https://www.worldpop.org/>>. Acesso em: 14 abr. 2023.