

**Instituto de Pesquisa Econômica Aplicada**

**Secretaria Nacional de Habitação do  
Ministério do Desenvolvimento Regional**

**Termo de Execução Descentralizada n. 01/2019 SNH/MDR e Ipea**

**Pesquisa de Núcleos Urbanos Informais no Brasil**

**Produto 2 – Tipologia das formas assumidas pelos núcleos  
urbanos informais: desenvolvimento metodológico para sua  
identificação e classificação**

*Elaboração*

Flávia da Fonseca Feitosa  
Gilmara Gonçalves  
Guilherme Frizzi Galdino da Silva  
Miriam Francisca Rodrigues Couto  
Pedro Reis Simões

*Revisão*

Cleandro Henrique Krause

Versão de 11 de setembro de 2020

## **Pesquisa de Núcleos Urbanos Informais no Brasil**

### *Coordenação*

Cleandro Henrique Krause (titular) – Técnico de Planejamento e Pesquisa da  
Diretoria de Estudos e Políticas Regionais, Urbanas e Ambientais (Dirur) do Ipea  
Marco Aurélio Costa (suplente) – Coordenador de Estudos de Desenvolvimento  
Urbano da Dirur-Ipea

### *Equipe da Pesquisa de Núcleos Urbanos Informais no Brasil – bolsistas e colaboradores*

Alexandrina Saldanha Sobreira de Moura – FUNDAJ / Ipea (PNPD)  
Ana Carolina Campos de Melo – Ipea (PNPD)  
André Simionato Castro – Ipea (PNPD)  
Bruno Gallina – UFRGS  
Cátia Wanderley Lubambo – FUNDAJ / Ipea (PNPD)  
David Melo Van Den Brule – Ipea (PNPD)  
Elisa Escosteguy Utzig – Ipea (PNPD)  
Fernanda Balestro – Ipea (PNPD)  
Fernanda Carolina Vieira da Costa – Ipea (PNPD)  
Flávia da Fonseca Feitosa – UFABC / Ipea (PNPD)  
Gilmara Gonçalves – Ipea (PNPD)  
Gabriel Moraes de Outeiro – UNIFESSPA / Ipea (PNPD)  
Giuliana de Freitas – Ipea (PNPD)  
Guilherme Frizzi Galdino da Silva – Ipea (PNPD)  
Heleniza Ávila Campos – UFRGS / Ipea (PNPD)  
Juliana Gomes Petrarolli – Ipea (PNPD)  
Livia Salomão Piccinini – UFRGS  
Manoela Guedes Ferreira Jordão de Vasconcelos – Ipea (PNPD)  
Marcela Rodrigues Santos – Ipea (PNPD)  
Mariana Roberti Bomtempo – Ipea (PNPD)  
Miriam Francisca Rodrigues Couto – Ipea (PNPD)  
Paulo Somlanyi Romeiro – Ipea (PNPD)  
Pedro Reis Simões – Ipea (PNPD)  
Rafael Gonçalves Gumiero – UNIFESSPA  
Raquel de Mattos Viana – FJP / Ipea (PNPD)  
Rosana Denaldi – UFABC / Ipea (PNPD)  
Sergio Moreno Redón – UNIFESSPA / Ipea (PNPD)  
Tatiana Mamede Salum Chaer – Ipea (PNPD)  
Thaís Pires Rubioli – Ipea (PNPD)  
Tiago Gonçalves da Silva – Ipea (PNPD)  
Valéria Barroso da Silveira – Ipea (PNPD)

## Sumário

<b>1. Apresentação</b>	<b>4</b>
<b>2. Levantamentos e análises iniciais para identificação remota de NUI: Índices, indicadores e variáveis</b>	<b>6</b>
2.1 Índice de Vulnerabilidade Social (IVS) e Unidades de Desenvolvimento Humano (UDH)	6
2.1.1 Índice de Vulnerabilidade Social (IVS)	6
2.2 Índice de Moradia Adequada (IMA)	12
2.3 Análise de autocorrelação espacial de área	15
<b>3. Identificação de Núcleos Urbanos Informais: Análise de Metodologias Relacionadas</b>	<b>20</b>
3.1 Metodologia do Centro de Estudos da Metrópole para identificação de assentamentos precários	20
3.1.1 Apresentação das variáveis	21
3.1.2 Filtros aplicados aos dados	24
3.1.3 Construção do modelo de classificação	25
3.1.4 Detalhes de software	26
3.1.5 Apresentação dos resultados	26
3.2 Metodologia para Identificação e Caracterização de Assentamentos Precários em Regiões Metropolitanas Paulistas (MAPPA)	31
3.2.1 Descrição da metodologia MAPPA	31
3.2.2 Aplicação da metodologia MAPPA aos dados da Pesquisa NUI	36
<b>4. Decisões metodológicas e encaminhamentos</b>	<b>45</b>
4.1 Aspectos relevantes para o desenvolvimento da metodologia NUI	45
4.2 Construção e integração de variáveis	48
4.3 Técnicas de Classificação	50
4.4 Mapas interativos	51
<b>5. Considerações Finais</b>	<b>54</b>
<b>6. Referências</b>	<b>56</b>

## 1. Apresentação

O presente relatório refere-se ao Produto 2 previsto no Plano de Trabalho da Pesquisa de Núcleos Urbanos Informais no Brasil, compreendendo o desenvolvimento metodológico necessário para a elaboração de uma tipologia de núcleos urbanos informais (NUI), visando sua identificação e classificação.

A elaboração deste relatório teve como ponto de partida referências presentes no Produto 1 da Pesquisa, apresentado à SNH/MDR em abril de 2020, especialmente em seus itens (i) revisão das formas assumidas pelos núcleos urbanos informais e (ii) análise das técnicas para o registro e a mensuração dos núcleos urbanos informais, avaliando sua acurácia e viabilidade, bem como a aplicabilidade ao caso em tela. Contudo, se no Produto 1 foram revisadas e analisadas referências preexistentes, no presente relatório busca-se desenvolvimento metodológico novo e adequado às necessidades específicas da Pesquisa de Núcleos Urbanos Informais no Brasil.

O Plano de Trabalho, elaborado em setembro de 2019, previu, no escopo do Produto 2, a construção de uma taxonomia das formas assumidas pelos núcleos urbanos informais, que explicitasse sua vinculação a tipologias existentes de assentamentos urbanos e metropolitanos no Brasil, por exemplo, os aglomerados subnormais dos censos demográficos do IBGE. Adicionalmente, esperava-se incluir uma tipologia compreendida no processo de apoio do Ipea à elaboração da Política Nacional de Desenvolvimento Urbano (PNDU), sob responsabilidade da Secretaria Nacional de Desenvolvimento Regional e Urbano do Ministério do Desenvolvimento Regional (SNDRU/MDR). Essa tipologia, portanto, deveria ter sido elaborada previamente, de modo a poder ser incorporada à Pesquisa de Núcleos Urbanos Informais no Brasil. Contudo, a tramitação do TED cujo objeto é o apoio do Ipea à PNDU deu-se em momento posterior ao da presente Pesquisa, inviabilizando, até o momento da apresentação deste Produto 2, a incorporação de tipologias da PNDU, as quais são o objeto de notas técnicas que se encontram em elaboração. Uma vez concluídas as notas técnicas das tipologias da PNDU, seus conteúdos deverão ser levados em consideração e refletirão em produtos subsequentes da presente Pesquisa.

O processo de desenvolvimento metodológico aqui apresentado constitui-se de três etapas: (i) levantamento e análise de índices, indicadores e variáveis potencialmente relevantes para a identificação de NUI; (ii) análise de metodologias de classificação que se relacionam com os objetivos da pesquisa, incluindo sua aplicação (com adaptações) aos dados dos polos de Belo Horizonte, Brasília, Juazeiro do Norte, Marabá, Porto Alegre e Recife<sup>1</sup>; e (iii) estabelecimento de decisões e encaminhamentos para a continuidade da construção da metodologia da pesquisa, a partir dos resultados obtidos nas etapas (i) e (ii).

A primeira etapa, apresentada no Capítulo 2 deste relatório, descreve, aplica e analisa três índices, bem como indicadores e variáveis que os compõem: o Índice de Vulnerabilidade Social

---

<sup>1</sup> Os seis polos citados correspondem aos territórios da pesquisa de campo, conforme estão descritos no Produto 3 da Pesquisa de Núcleos Urbanos Informais no Brasil.

(IVS), o Índice de Moradia Adequada (IMA) e o Índice de Moran. As análises buscaram compreender cada um desses índices, incluindo suas potencialidades e limitações para a presente Pesquisa, e verificar se seus resultados estão significativamente relacionados à presença de NUI de baixa renda.

A segunda etapa, descrita no Capítulo 3, apresenta e explora duas metodologias de identificação de assentamentos precários. A primeira delas, desenvolvida pelo Centro de Estudos da Metrópole (CEM) a pedido do então Ministério das Cidades, consiste na aplicação de técnicas estatísticas para a identificação de setores censitários que apresentam características comparáveis com as dos aglomerados subnormais, mas que não foram originalmente classificados como tal. A metodologia foi aplicada para os seis polos a partir dos dados do Censo Demográfico de 2000 e 2010. Com os dados de 2000, buscou-se validar os procedimentos realizados, verificando se condizem com os originalmente propostos e realizados pelo CEM. Já os dados de 2010 foram utilizados para a obtenção de modelos mais atualizados, cujos resultados de classificação puderam ser avaliados de forma mais detalhada.

A segunda metodologia analisada foi a Metodologia de Identificação e Classificação de Assentamentos Precários em Regiões Metropolitanas Paulistas (MAPPA), desenvolvida pela Universidade Federal do ABC (UFABC) a pedido da Companhia de Desenvolvimento Habitacional e Urbano (CDHU) para o Estado de São Paulo. A metodologia integra variáveis de fontes diversas para a construção de superfícies de probabilidade da presença de distintas tipologias de assentamentos precários. Essa metodologia foi adaptada, considerando dados disponíveis para todo o Brasil, aplicada aos polos da presente Pesquisa e analisada.

Considerando o resultado das análises dos índices e metodologias, o Capítulo 4 apresenta decisões metodológicas e propõe encaminhamentos para a continuidade dessa frente de pesquisa. Esses encaminhamentos incluem estratégias de integração e seleção das variáveis a serem consideradas nos modelos de classificação, definição das técnicas estatísticas mais adequadas para a metodologia e construção de mapas interativos para a comunicação dos resultados da modelagem.

Por fim, as considerações finais apresentam um balanço sobre os resultados já alcançados e os esperados para as próximas etapas do desenvolvimento metodológico.

## **2. Levantamentos e análises iniciais para identificação remota de NUI: Índices, indicadores e variáveis**

### **2.1 Índice de Vulnerabilidade Social (IVS) e Unidades de Desenvolvimento Humano (UDH)**

A vulnerabilidade social é uma das perspectivas consideradas para a análise dos territórios quanto à precariedade de assentamentos e, por conseguinte, suas formas de representação assumem papel relevante na construção de uma metodologia de identificação de NUI. Buscando fomentar o desenvolvimento metodológico da pesquisa, uma análise do Índice de Vulnerabilidade Social (IVS), desenvolvido pelo Ipea (2015), foi conduzida para os polos de Belo Horizonte, Brasília, Juazeiro do Norte, Marabá, Porto Alegre e Recife.

O IVS, além de ampliar o número de indicadores em relação ao Índice de Desenvolvimento Humano Municipal<sup>2</sup> (IDHM), também inova como método de análise por utilizar dados agregados em Unidades de Desenvolvimento Humano – UDH. As UDH são unidades espaciais de área análogas às áreas de ponderação do IBGE, para as quais dados da amostra do censo demográfico podem ser agregados (PNUD; IPEA; FJP, 2016). Entretanto, as UDH são áreas mais homogêneas, constituídas a partir da agregação de setores censitários com semelhanças socioeconômicas. As UDH estão disponíveis para as regiões metropolitanas contempladas pelo Atlas de Desenvolvimento Humano (ADH), que foi desenvolvido em parceria pelo IPEA, PNUD e FJP. Assim, além da análise do IVS, essa seção apresenta também os resultados de uma análise exploratória que busca avaliar o potencial de variáveis disponíveis por UDH para a identificação de NUI.

#### **2.1.1 Índice de Vulnerabilidade Social (IVS)**

O Índice de Vulnerabilidade Social, construído a partir de indicadores do Atlas de Desenvolvimento Social (ADH), objetiva destacar situações de exclusão e vulnerabilidade social no Brasil.

“O IVS é um índice sintético que reúne indicadores do bloco de vulnerabilidade social do ADH, os quais, apresentados por meio de cartogramas e estruturados em diferentes dimensões, servem de suporte para a identificação de porções do território onde há a sobreposição daquelas situações indicativas de exclusão e vulnerabilidade social no território, de modo a orientar gestores públicos municipais, estaduais e federais para o desenho de políticas públicas mais sintonizadas com as carências e necessidades presentes nesses territórios” (IPEA, 2015, p. 12).

O IVS busca produzir de “forma mensurável e ordenada as condições de vida, as situações de pobreza, as carências sociais, a qualidade de vida e outros aspectos que dizem respeito ao bem-estar das populações numa perspectiva multidimensional”, explicam Costa *et al.* (2018, p. 19). O IVS foi pensado para nortear políticas sociais, pesquisadores e acadêmicos que buscam compreender diferentes possibilidades da vulnerabilidade social no Brasil, sendo complementar ao Índice de Desenvolvimento Humano Municipal (IDHM). O IVS utiliza dados do Censo

---

<sup>2</sup> O IDHM brasileiro segue as mesmas três dimensões do IDH global – saúde, educação e renda, mas vai além: adequa a metodologia global ao contexto brasileiro e à disponibilidade de indicadores nacionais (PNUD; IPEA; FJP, 2013).

Demográfico 2010, provenientes dos questionários básico e da amostra, relacionados a situações de vulnerabilidade social. Foi desenvolvido para ser aplicado em diversas escalas e apresentado na forma de mapas que fossem facilmente interpretados.

O IVS se estrutura em três dimensões que correspondem a temas, cujo acesso, ausência e/ou insuficiência indicam alguma(s) vulnerabilidade(s). Para tanto, cada dimensão é representada por um conjunto de indicadores:

i) Infraestrutura urbana: representa as condições de acesso aos serviços de saneamento básico e mobilidade urbana. Compõem esta dimensão a presença de rede de abastecimento de água, de serviço de esgotamento sanitário e coleta de lixo, bem como o indicador do tempo gasto no deslocamento entre a moradia e o local de trabalho pela população ocupada de baixa renda. A dimensão se divide em três indicadores:

- a) Percentual da população em domicílios com abastecimento de água e esgotamento sanitário inadequados;
- b) Percentual da população que vive em domicílios urbanos sem serviço de coleta de lixo;
- c) Percentual de pessoas que vivem em domicílios com renda per capita inferior a meio salário mínimo e que gastam mais de uma hora até o trabalho, considerando o total de pessoas ocupadas de baixa renda que retornam diariamente do trabalho.

ii) Capital humano: apoia-se na inclusão social dos indivíduos, incluindo aspectos como saúde e educação. Neste caso, consideram-se perspectivas de condições atuais e futuras incluindo informações sobre a infância e adolescência. “Adotou-se, para isso, indicadores de mortalidade infantil; da presença, nos domicílios, de crianças e jovens que não frequentam a escola; da presença, nos domicílios, de mães precoces, e de mães chefes de família, com baixa escolaridade e filhos menores; da ocorrência de baixa escolaridade entre os adultos do domicílio; e da presença de jovens que não trabalham e não estudam” (IPEA, 2015, p.14). A dimensão se divide em oito indicadores:

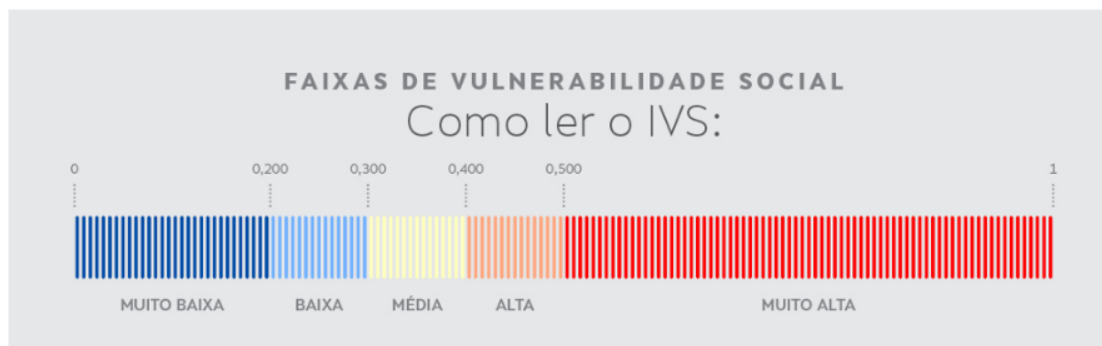
- a) Mortalidade até um ano de idade;
- b) Percentual de crianças de 0 a 5 anos que não frequentam a escola;
- c) Percentual de pessoas de 6 a 14 anos que não frequentam a escola;
- d) Percentual de mulheres de 10 a 17 anos de idade que tiveram filhos;
- e) Percentual de mães chefes de família, sem fundamental completo e com pelo menos um filho menor de 15 anos de idade, no total de mães chefes de família;
- f) Taxa de analfabetismo da população de 15 anos ou mais de idade;
- g) Percentual de crianças que vivem em domicílios em que nenhum dos moradores tem o ensino fundamental completo;
- h) Percentual de pessoas de 15 a 24 anos que não estudam, não trabalham e possuem renda domiciliar per capita igual ou inferior a meio salário mínimo (2010), na população total dessa faixa etária.

iii) Renda e trabalho: agrupa indicadores referentes à insuficiência de renda e fluxo e insegurança de renda, portanto, incluindo renda familiar *per capita*, desocupação de adultos, ocupação informal de adultos pouco escolarizados, dependência a renda de pessoa idosa, e trabalho infantil. Para isso, os indicadores referentes a renda e trabalho são:

- a) Proporção de pessoas com renda domiciliar per capita igual ou inferior a meio salário mínimo (2010);
- b) Taxa de desocupação da população de 18 anos ou mais de idade;
- c) Percentual de pessoas de 18 anos ou mais sem fundamental completo e em ocupação informal;
- d) Percentual de pessoas em domicílios com renda per capita inferior a meio salário mínimo (2010) e dependentes de idosos;
- e) Taxa de atividade das pessoas de 10 a 14 anos de idade.

O IVS resulta da média aritmética dessas três dimensões. Cada indicador tem seu valor normalizado numa escala que varia entre 0 a 1, em que 0 corresponde à situação ideal, ou desejável, e 1 corresponde à pior situação. Cinco faixas de vulnerabilidade social foram geradas a partir dessa escala (Figura 1): Muito baixa (entre 0 e 0,200), Baixa (entre 0,201 e 0,300), Média (entre 0,301 e 0,400), Alta (entre 0,401 e 0,500) e Muito alta (entre 0,501 e 1).

Figura 1 - Escala do Índice de Vulnerabilidade Social – IVS.



Fonte: IPEA, 2020.

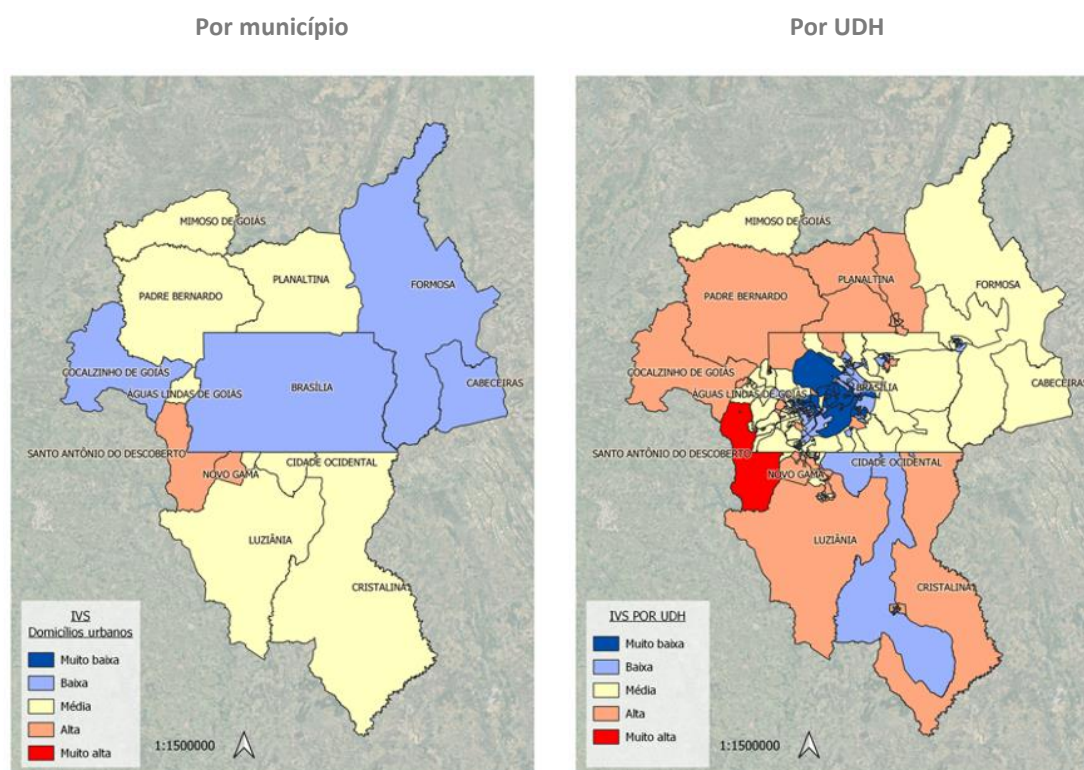
O IVS é aplicável em duas escalas: municipal, contemplando 5.570 municípios brasileiros, e a partir das Unidades de Desenvolvimento Humano (UDH), que foram disponibilizadas para regiões metropolitanas.

As UDH foram criadas para representar com maior homogeneidade as condições socioeconômicas e ocupação territorial. Produzidas para o Atlas do Desenvolvimento Humano (ADH), as UDH foram geradas a partir do acesso aos dados do questionário da amostra para recortes espaciais que se diferenciam das áreas de ponderação dos Censos do IBGE. Esse acesso foi aprovado por uma comissão técnica, após a submissão de um projeto que respeitou as exigências de confiabilidade estatística. Tal processo buscou gerar áreas mais homogêneas do que as áreas de ponderação do IBGE, já que essas baseiam-se em quesitos técnicos relacionados ao processo de coleta e amostragem. Essa construção e delimitação das UDH contou com apoio e conhecimento técnico de instituições e pesquisadores das regiões metropolitanas. A seguir



(Figura 2) são apresentados mapas do IVS em escala municipal e por UDH ilustrando, assim, o detalhamento espacial das UDH.

Figura 2 - Comparativo da escala do IVS por município e por UDH – Polo Brasília



Fonte: Elaboração própria com base nos dados do IPEA (2020).

Com o intuito de verificar as potencialidades das Unidades de Desenvolvimento Humano, suas variáveis e dimensões na identificação de núcleos urbanos informais, foram realizadas duas análises distintas. A primeira, apresentada no Apêndice A, consiste em uma análise exploratória do IVS e suas dimensões, por município e UDH, para os 6 polos da pesquisa. Já a segunda explora as variáveis disponíveis para as UDH e seu poder explicativo em relação à presença de NUI nos municípios do polo de Brasília.

Na segunda análise testou-se a correlação de Pearson dos indicadores disponíveis por UDH com a presença de Aglomerados Subnormais<sup>3</sup> (Censo Demográfico, 2010) e Áreas de Regularização de Interesse Social (ARIS), disponíveis no Geoportal do Distrito Federal<sup>4</sup>. O experimento se restringiu aos municípios do polo de Brasília devido a disponibilidade e confiabilidade dos dados.

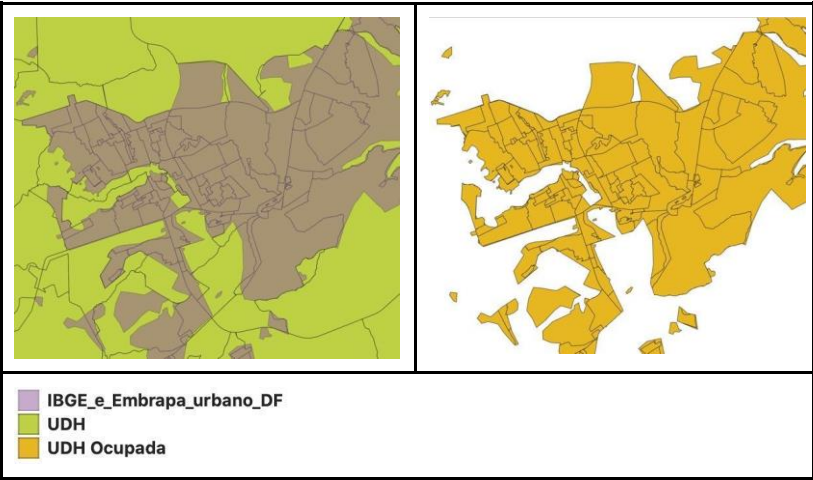
Para a análise de correlação, foram criadas duas variáveis: a) Porcentagem da UDH coberta por Aglomerados Subnormais e; b) Porcentagem da UDH coberta por Áreas de Regularização de Interesse Social. Nesse processo, foram selecionadas apenas as áreas ocupadas das UDH,

<sup>3</sup> Os Aglomerados Subnormais são uma forma de ocupação irregular de terrenos de propriedade alheia para fins de habitação em áreas urbanas e são, em geral, caracterizados por um padrão urbanístico irregular, carência de serviços públicos essenciais e localização em áreas com restrição à ocupação (IBGE, 2010).

<sup>4</sup> <https://www.geoportal.seduh.df.gov.br/mapa/>

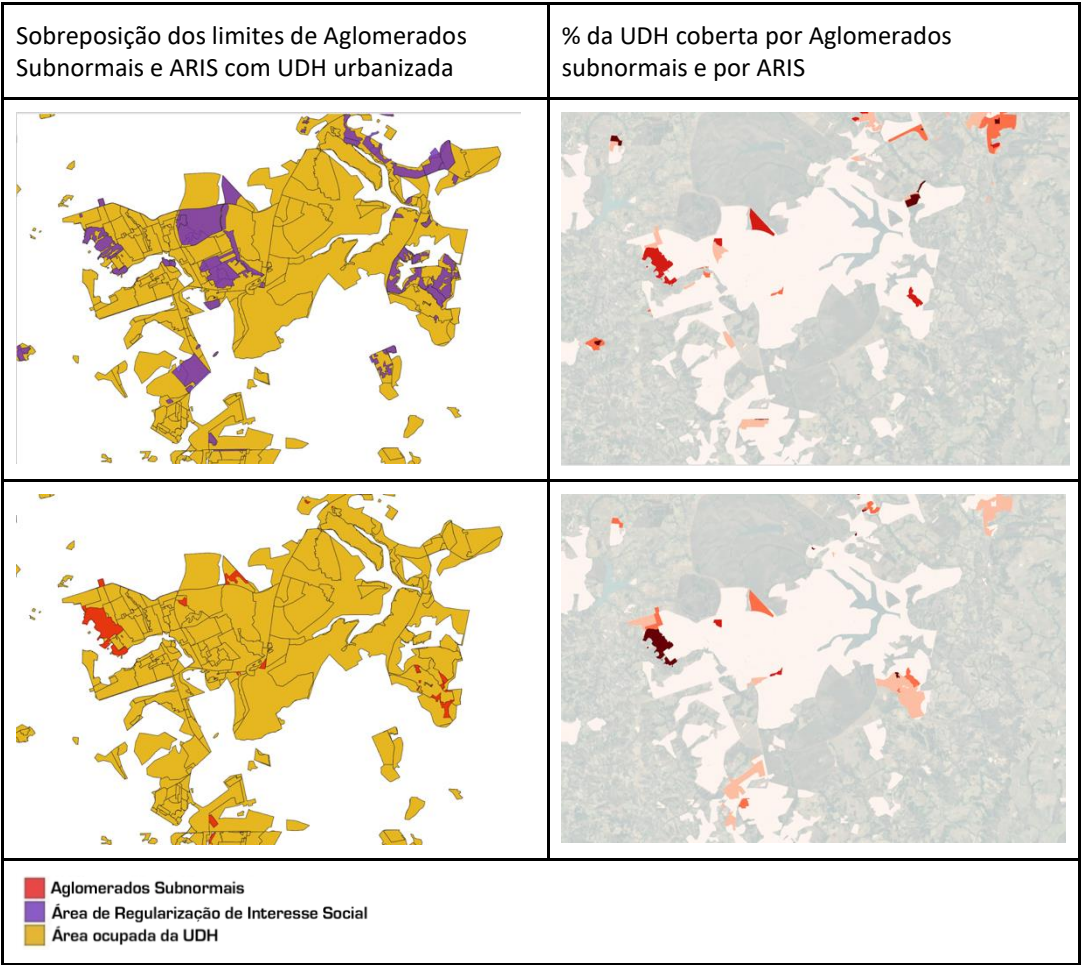
considerando as informações das camadas de áreas urbanizadas IBGE e Embrapa (2015) (Figura 3 e Figura 4).

Figura 3 - Áreas urbanizadas da UDH.



Fonte: Elaboração própria, 2020.

Figura 4 - Porcentagem da UDH coberta por Aglomerados Subnormais/ARIS.



Fonte: Elaboração própria, 2020.

Em seguida, foi gerada uma matriz de correlação de Pearson entre as variáveis disponíveis para as UDH e a porcentagem das UDH coberta por Aglomerados Subnormais/ARIS (Quadro 1). O coeficiente de correlação de Pearson ( $r$ ) assume valores entre -1 e 1, com  $r = 0$  correspondendo à ausência de associação entre variáveis. O coeficiente assume valores positivos em caso de associação positiva, ou seja, à medida que a variável  $x$  aumenta, também aumenta a variável  $y$ . De forma oposta, assume valores negativos em caso de associação negativa, ou seja, à medida que  $x$  aumenta,  $y$  diminui (HIRSCH, 2016).

Quanto maior o valor do módulo de  $r$ , mais forte é a associação. Embora a interpretação da magnitude do valor de  $r$  possa variar dependendo do contexto, um coeficiente de correlação de 0,10 até 0,30 é considerado uma correlação fraca, de 0,40 até 0,60, moderada, e de 0,70 até 1, uma correlação forte (HIRSCH, 2016).

Quadro 1 - Matriz de correlação de Pearson entre a porcentagem de Aglomerados Subnormais (ASmean) e Área de Regularização de Interesse Social (ARISmean) com os indicadores disponíveis para as UDH\*.

Matriz de Correlação			Descrição das variáveis
	ASmean	ARISmean	
ASmean	1	0,283	Porcentagem da UDH coberta por Aglomerados Subnormais
ARISmean	0,283	1	Porcentagem da UDH coberta por Áreas de Regularização de Interesse Social (ARIS)
FEOTOT	0,219	0,244	Taxa de fecundidade total: Número médio de filhos que uma mulher deverá ter ao terminar o período reprodutivo (15 a 49 anos de idade).
MORT1	0,284	0,338	Mortalidade até um ano de idade: Número de crianças que não deverão sobreviver ao primeiro ano de vida em cada 1000 crianças nascidas vivas.
MORT5	0,276	0,324	Mortalidade até cinco anos de idade: Probabilidade de morrer entre o nascimento e a idade exata de 5 anos, por 1000 crianças nascidas vivas.
SOBRE40	-0,223	-0,242	Probabilidade de sobrevivência até 40 anos: Probabilidade de uma criança recém-nascida viver até os 40 anos, se permanecerem constantes ao longo da vida o nível e o padrão de mortalidade por idade prevalentes no ano do Censo.
SOBRE60	-0,197	-0,205	Probabilidade de sobrevivência até 60 anos: Probabilidade de uma criança recém-nascida viver até os 60 anos, se permanecerem constantes ao longo da vida o nível e o padrão de mortalidade por idade prevalentes no ano do Censo.
T_ENV	-0,202	-0,233	Taxa de envelhecimento: Razão entre a população de 65 anos ou mais de idade e a população total multiplicado por 100.
T_AGUA	-0,149	-0,114	Percentual da população que vive em domicílios com água encanada.
T_DENS	0,216	0,259	Percentual da população que vive em domicílios com densidade superior a 2 pessoas por dormitório: Razão entre a população que vive em domicílios particulares permanentes com densidade superior a 2 e a população total residente em domicílios particulares permanentes, multiplicada por 100. A densidade do domicílio é dada pela razão entre o total de moradores do domicílio e o número total de cômodos usados como dormitório.
PAREDE	0,165	0,175	Percentual de pessoas em domicílios com paredes que não sejam de alvenaria ou madeira aparelhada.
T_VULNER_1	0,152	0,2	Percentual de mães chefes de família, sem fundamental completo e com pelo menos um filho menor de 15 anos de idade, no total de mães chefes de família e com filho menor.
Legenda	Correlação positiva e significativa ao nível de 5%		
	Correlação negativa e significativa ao nível de 5%		

\*foram mantidas na tabela apenas as variáveis que apresentaram coeficientes com maior magnitude. Uma lista que inclui essas e outras variáveis disponíveis por UDH que poderão ser consideradas em etapas futuras da pesquisa encontra-se no Apêndice B.

Fonte: Elaboração própria, 2020.

Os resultados apresentados no Quadro 1 revelam correlações com significância estatística ( $p$ -valor < 0,05), embora consideradas fracas. Conclui-se, assim, que as UDH podem fornecer

variáveis e indicadores relacionadas à presença de NUI, aumentando a probabilidade da sua identificação. É o caso das variáveis T\_DENS e PAREDE, que indicam, respectivamente, o percentual da população que vive em domicílios com densidade superior a 2 pessoas por dormitório e percentual de pessoas em domicílios com paredes que não sejam de alvenaria ou madeira aparelhada. Essas duas variáveis são também utilizadas na metodologia para a estimativa do déficit habitacional e inadequação de domicílios no Brasil desenvolvida pela Fundação João Pinheiro (FJP, 2018). A disponibilidade dessas variáveis por UDH, ou seja, unidades de análise mais refinadas e homogêneas que as originais áreas de ponderação, pode auxiliar na construção dos modelos de classificação de NUI. Pretende-se, portanto, incluir as variáveis disponibilizadas por UDH nos modelos estatísticos a serem desenvolvidos nas próximas etapas da pesquisa.

## **2.2 Índice de Moradia Adequada (IMA)**

O Índice de Moradia Adequada (IMA), desenvolvido por Frizzi (2019), também foi analisado com o intuito de avaliar seu potencial para subsidiar a construção da metodologia para a identificação de NUI. A partir da hipótese de que as áreas com um menor Índice de Moradia Adequada têm uma maior probabilidade de serem NUI, foram gerados testes estatísticos entre os resultados dos IMA e a presença de Assentamentos Precários mapeados pelo Diagnóstico Habitacional do Grande ABC (CIGABC/UFABC; DENALDI, 2016). A escolha da Região do Grande ABC como área teste para análise da correlação do IMA com a presença de assentamentos precários se deu em decorrência da preexistência de um mapeamento detalhado desses assentamentos na região.

Segundo Frizzi (2019), os conceitos de moradia adequada e das dimensões que a compõem, foram desenvolvidos a partir do estudo de três documentos chave. O primeiro foi o artigo XI, Item I do Pacto Internacional dos Direitos Econômicos, Sociais e Culturais, de 1966, chamado de *“The Right to Adequate Housing”* [O Direito à Moradia Adequada], aprovado em 1991 (UNITED NATIONS, 1991 apud FRIZZLI, 2019). O segundo foi a publicação *Health principles of housing* [Princípios de saúde da habitação], desenvolvida pela Organização Mundial da Saúde em 1989 (OMS, 1989 apud FRIZZLI, 2019). E por último, os estudos de inadequação habitacional que vêm sendo desenvolvidos pela Fundação João Pinheiro desde 1995 (FJP, 2018), onde são levantados dados sobre o déficit habitacional e sobre a inadequação de domicílios no Brasil com dados do Censo Demográfico e das Pesquisas Nacionais por Amostra de Domicílios (PNAD), ambos elaborados pelo Instituto Brasileiro de Geografia e Estatística (IBGE).

De acordo com o autor, o IMA tem como objetivo instrumentalizar o conceito de moradia adequada em escala intraurbana e, assim, fornecer subsídios para o planejamento e gestão territorial e direcionamento de políticas públicas (FRIZZLI, 2019). A construção metodológica do índice parte de dois princípios básicos que norteiam suas escolhas metodológicas.

O primeiro princípio está relacionado à simplicidade de sua estrutura e garantia de fácil compreensão e aplicação. Para isso, evitou-se cálculos e métodos estatísticos que dificultam e complexificam o seu entendimento e análise. Já o segundo princípio foi pautado na disponibilidade dos dados utilizados para a construção das variáveis e garantia da replicabilidade

do índice em outros territórios. Dessa forma, foram utilizados apenas softwares livres e gratuitos e dados disponíveis para todo o território nacional (FRIZZI, 2019).

Além disso, o IMA utiliza técnicas que permitem análises multiescalares e multidimensionais. Para a garantia desses atributos, assim como para realizar a integração de dados provenientes de diferentes formatos e escalas, o autor integrou os dados do IMA em um plano de informação celular hexagonal com resolução espacial de 100 por 100 metros. As células deste plano, associadas ao banco de dados geográficos, estabelecem uma nova base espacial para a distribuição das variáveis consideradas no processo de construção do IMA. Dessa forma, é possível selecionar diferentes recortes em diferentes escalas territoriais (FRIZZI, 2019). Já para a garantia do atributo multidimensional, o índice é apresentado tanto em sua forma sintética final, ou seja, o acesso à moradia adequada, quanto em suas diferentes dimensões que compõem o conceito de moradia adequada (FRIZZI, 2019).

Com o intuito de sintetizar os entendimentos acerca do conceito da moradia adequada e estabelecer um quadro conceitual único para a elaboração do IMA, o autor apresenta um quadro resumo com as dimensões que compõem o conceito da moradia adequada e suas exigências para o cumprimento da mesma (Quadro 2).

Quadro 2 - Dimensões da moradia adequada e indicadores.

Dimensões	Descrição	Indicadores
<b>Segurança contra remoções</b>	Adequação fundiária urbana e proteção legal contra despejos forçados, assédio e outras ameaças.	% da área em Unidades de Conservação; % da área ocupada por aglomerados subnormais; % da área no limite de obras públicas que envolvem remoções e; % da áreas situação de risco geotécnico (inundação e movimento de massa).
<b>Infraestrutura pública</b>	Acesso à água potável, energia, aquecimento, iluminação, saneamento, armazenamento de alimentos, coleta de lixo e drenagem.	% de domicílios com acesso à água, energia elétrica, esgoto e coleta de lixo. % do entorno dos domicílios com iluminação; bueiro/boca-de-lobo, sem esgoto a céu aberto e sem lixo acumulado no logradouro.
<b>Custo</b>	Os custos associados à moradia não devem comprometer a obtenção e a satisfação de outras necessidades básicas.	% de famílias que ganham até 3 salários mínimos e o aluguel ultrapassa 30% do rendimento mensal.
<b>Habitabilidade</b>	Espaço adequado, não rústico ou improvisado, banheiro exclusivo e sem adensamento excessivo de moradores por dormitório.	% de domicílios rústicos e improvisados; % de domicílios com banheiro exclusivo e; % de domicílios com mais de 3 moradores por dormitório.
<b>Acesso</b>	Grupos desfavorecidos e vulneráveis devem ter acesso assegurado e algum grau de prioridade a recursos habitacionais adequados.	Sem dados disponíveis
<b>Localização</b>	Acesso a opções de emprego, serviços de saúde, escolas, creches e outras instalações sociais.	Proximidade à concentração de equipamentos de saúde e educação; % de pessoas que demoram até 30 min ao trabalho e; proximidade à concentração de empregos.
<b>Adequação cultural</b>	A habitação deve ser adaptada aos modos e costumes da cultura dos seus moradores.	Sem dados disponíveis
<b>Uso consciente</b>	Uso adequado da moradia para atingir o potencial máximo de saúde.	Sem dados disponíveis
<b>Saúde mental</b>	Apoio no desenvolvimento social e psicológico das pessoas e reduzir ao mínimo o estresses psicológicos e sociais ligados ao ambiente habitacional.	Sem dados disponíveis

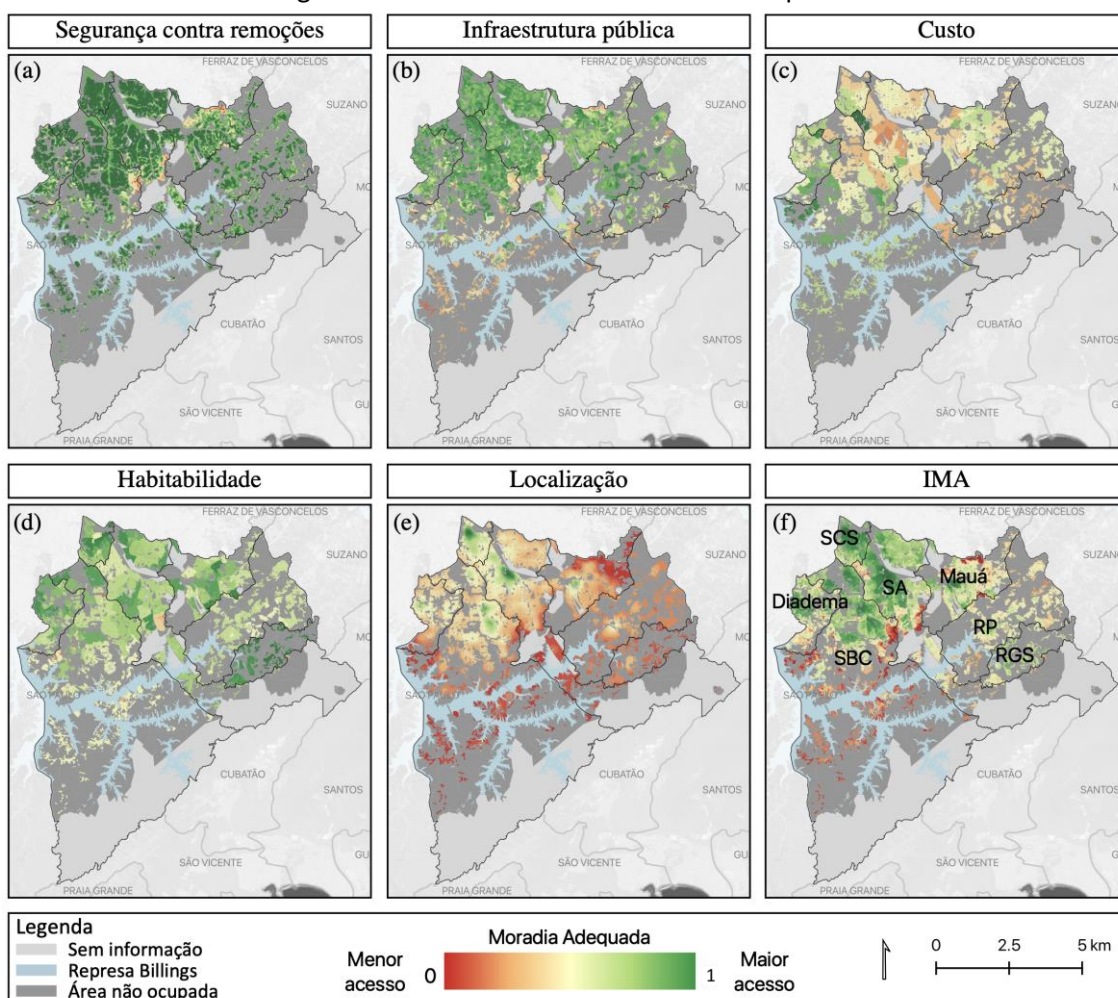
Fonte: adaptado de Frizzi (2019) baseado em United Nations (1991), WHO (2010) e FJP (2018).



Devido a inexistência de indicadores disponíveis de forma aberta em escala intraurbana e que representassem as dimensões de Acesso, Adequação cultural, Uso consciente e Saúde mental em todo o território nacional, o autor optou por não as apresentar no índice final (FRIZZZI, 2019). Logo, para a composição do Índice de Moradia Adequada proposto no trabalho, foram consideradas as dimensões de Segurança contra remoções, Infraestrutura pública, Custo, Habitabilidade e Localização. Os indicadores que compõem cada uma dessas dimensões também são apresentados no Quadro 2.

Segundo o autor, os indicadores foram integrados em cada uma das suas cinco dimensões. Posteriormente, os resultados das dimensões são combinados para gerar o Índice de Moradia Adequada (FRIZZZI, 2019). Na Figura 5, pode-se observar o resultado do IMA na Região do Grande ABC. Em cada célula hexagonal é possível extrair os valores do acesso a cada uma das dimensões que compõem o IMA (Figuras 5, a, b, c, d, e), assim como o IMA sintético (Figura 5, f). Os acessos a cada uma das dimensões são apresentados em uma escala de 0 a 1. Dessa forma, os valores mais próximos de 1 indicam um maior acesso às dimensões, enquanto os menores valores, mais próximos de 0, representam um menor acesso. A mesma lógica se aplica ao índice sintético IMA.

Figura 5 - Dimensões do IMA em áreas ocupadas.



Fonte: Frizzi (2019).

Como mencionado anteriormente, com o intuito de testar a aplicabilidade do IMA como instrumento para auxiliar na identificação dos Núcleos Urbanos Informais, foram realizadas análises estatísticas de correlação entre a presença de Assentamentos Precários e o IMA. Para isso, foi utilizado o coeficiente de correlação de Pearson. O resultado do cômputo do coeficiente indicou uma correlação negativa de **-0,229** e estatisticamente significativa ( $p\text{-valor} < 0,05$ ).

O resultado obtido para a Região do Grande ABC corrobora com a hipótese de que o Índice de Moradia Adequada está relacionado à presença de assentamentos precários. Áreas de baixo acesso à moradia adequada, segundo o IMA, são áreas estatisticamente mais prováveis de serem assentamentos precários. Apesar disso, o IMA apresentou uma correlação fraca com a presença de assentamentos precários (inferior à 0,3). Dessa forma, podemos afirmar que o IMA, da forma que foi construído, apesar de atender às exigências do estudo, ainda não é um instrumento ideal para indicar presença de núcleos urbanos informais. Entretanto, assim como o IVS e as UDH, apresentadas anteriormente, o IMA dispõe de dimensões e variáveis relevantes para indicar a presença de NUI e que serão consideradas nas próximas etapas da Pesquisa.

### **2.3 Análise de autocorrelação espacial de área**

As técnicas de análise de autocorrelação espacial foram desenvolvidas para contribuir na identificação de regiões onde a distribuição dos valores possa apresentar um padrão específico associado a sua localização geográfica. O resultado que se busca alcançar é o quanto cada área é parecida ou diferente dos seus vizinhos, o que permite identificar padrões de concentração e segregação (CÂMARA et al., 2002).

No âmbito desta Pesquisa, a utilização das técnicas estatísticas espaciais teve como intuito testar a identificação de núcleos urbanos informais a partir da localização de “clusters” ou aglomerados de setores censitários que apresentassem, entre si, valores altos de variáveis consideradas representativas de situações de precariedade.

A estratégia mais utilizada para se medir a dependência espacial é o cômputo do Índice Global de Moran, que pode ser decomposto no Índice Local de Moran para análises mais detalhadas (CÂMARA et al., 2002). De forma geral, o Índice Global de Moran testa a hipótese nula de independência espacial. Valores iguais a 0 indicam correlação nula; valores positivos, entre 0 e 1 indicam correlações positivas; e valores entre 0 e -1 indicam correlações negativas. O índice é calculado a partir da comparação dos atributos de uma feição com a média dos atributos dos seus vizinhos. Os vizinhos são definidos a partir de uma matriz de vizinhança, que comumente se baseia em critérios de contiguidade (ANSELIN, 1995).

Para garantir a validade estatística, geralmente é realizado o teste de pseudo-significância. Para isso são geradas diferentes permutações onde os valores observados são distribuídos aleatoriamente entre as áreas, de tal forma que cada permutação produz uma nova configuração espacial. Em seguida, computa-se o índice de Moran para todas configurações espaciais obtidas por permutação aleatória. A partir dos resultados, constrói-se uma distribuição de frequência dos Índices de Moran em relação à qual o valor do índice computado para a

configuração espacial real é comparado. Se o índice real corresponder a um “extremo” da distribuição de frequência simulada, podemos afirmar que o valor é estatisticamente significativo (CÂMARA et al., 2002).

O Índice Global de Moran fornece um valor único como medida de associação espacial para todo o conjunto de dados, o que é útil para uma análise geral. Para se examinar padrões em escalas mais detalhadas, geralmente utiliza-se o Índice Local de Moran. O indicador local produz um valor específico para cada feição, acarretando na identificação de agrupamentos com valores semelhantes ou “clusters”, valores anômalos e a presença de mais de um padrão espacial.

Em relação à matriz de vizinhança, o experimento realizado utilizou uma matriz do tipo *Queen*, que considera como vizinhos as feições contíguas por arestas e vértices. Para a análise da significância estatística dos resultados, foram realizadas 999 permutações aleatórias e considerou-se como significativos apenas os índices com nível de confiança acima de 95%.

Para a geração dos Índices Locais (LISA), foram selecionadas variáveis do Censo Demográfico de 2010 que fossem representativas da precariedade habitacional. As variáveis selecionadas para a elaboração e análise do indicador podem ser observadas no quadro a seguir (Quadro 3).

Quadro 3 - Variáveis utilizadas para criação e análise do LISA.

Variáveis analisadas		
Nome da variável	Nome da tabela	Descrição da variável
rndpc_d	final_Basico_UF	Rendimento nominal médio per capita por domicílio.
Morad_g	final_Domicilio01_UF	% de moradias atendidas pela rede de abastecimento de água.
Mrd_bnh		% de moradias onde existe banheiro para os moradores.
Mrd_bn		% de moradias onde há banheiros e saneamento adequado.
Str_s_p	final_Entorno01_UF	% do total de domicílios do setor que não existe pavimentação no entorno imediato da moradia.
Str_sgt		% do total de domicílios do setor próximo ao esgoto que não está canalizado.
Setr_lx		% do total de domicílios do setor próximo a logradouros com lixo acumulado.

Fonte: Elaboração própria com base nos dados do Censo IBGE 2010.

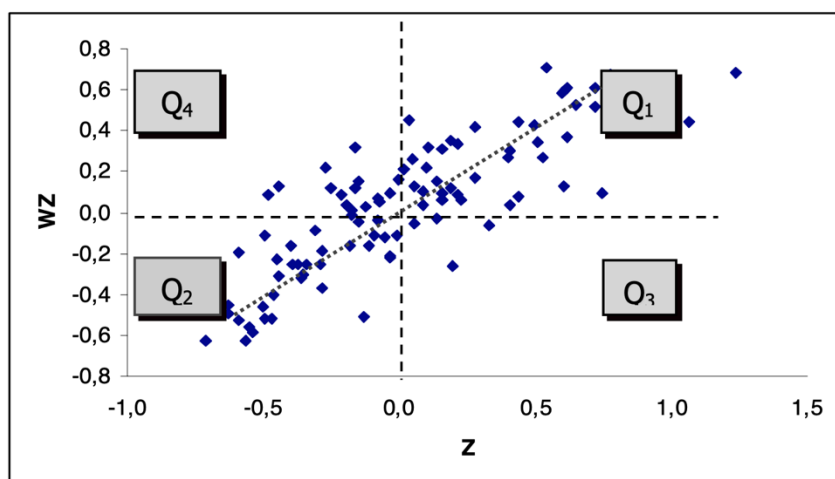
Uma forma adicional de visualizar os resultados dos Índices Locais é pelo Diagrama de Espalhamento de Moran. O diagrama é construído com base nos valores normalizados (valores de atributos subtraídos de sua média e divididos pelo desvio padrão), permitindo assim analisar o comportamento da variabilidade espacial. O diagrama é construído por um gráfico



bidimensional dividido em quatro quadrantes (Figura 6), onde cada quadrante pode ser interpretado como:

- Q1 e Q2: representam áreas que contribuem para autocorrelação positiva e indicam clusters (aglomerações) de valores altos (Alto-Alto) e baixos (Baixo-Baixo).
- Q3 e Q4: representam áreas que contribuem para autocorrelação negativa e indicam áreas de transição, com os valores altos cercados por valores baixos do atributo (Alto-Baixo), e vice-versa (Baixo-Alto).

Figura 6 - Exemplo de Diagrama de Espalhamento de Moran



Fonte: CÂMARA et al. (2002).

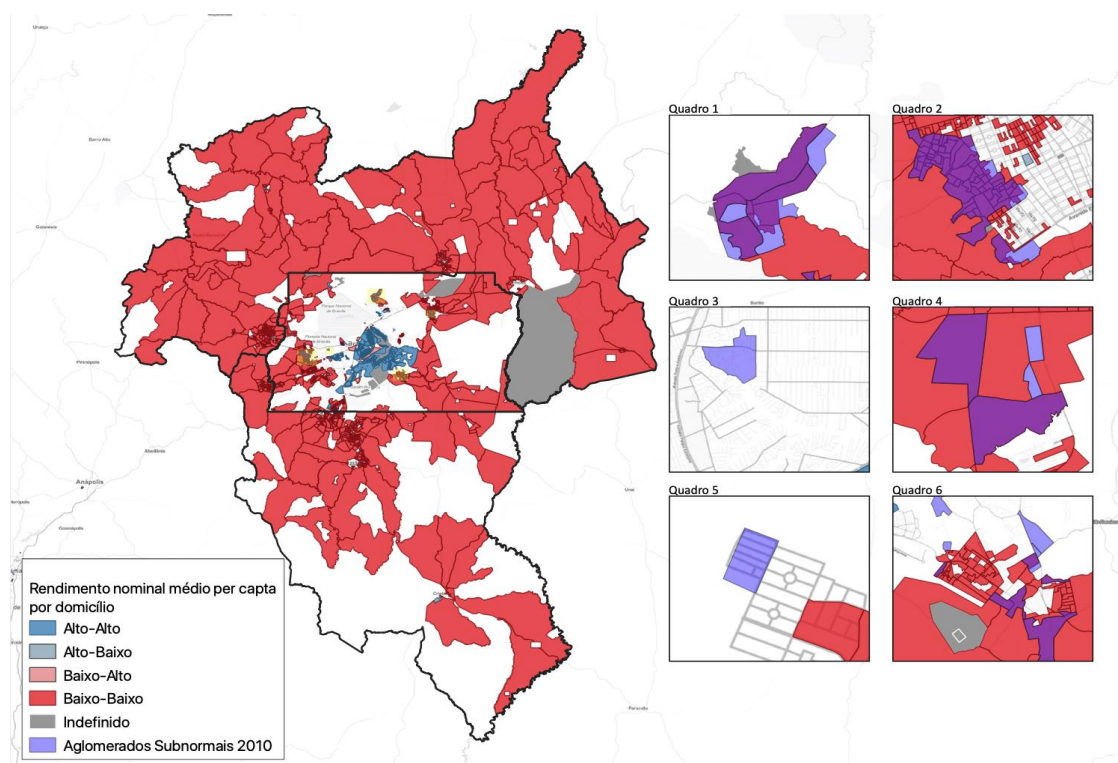
Para cada uma das variáveis censitárias descritas anteriormente no Quadro 3 foram gerados os Diagramas de Espalhamento de Moran. Os índices locais com nível de confiança superior a 95% foram espacializados na forma de mapas temáticos e sobrepostos com os polígonos dos Aglomerados Subnormais (AGSN), oriundos do Censo Demográfico de 2010.

A sobreposição dos Índices Locais de Moran com os polígonos dos AGSN teve como objetivo analisar a relação entre ambos. Parte-se da hipótese de que áreas classificadas no quadrante Q1 (Alto-Alto) de variáveis relacionadas à precariedade, como lixo acumulado, representam clusters de precariedade, com alta probabilidade de serem AGSN e, consequentemente, núcleos urbanos informais. De forma análoga, é esperado também alta probabilidade de ocorrência de NUI em áreas classificadas no quadrante Q2 (Baixo-Baixo), quando se trata de outras variáveis, como o caso da variável de renda. As representações elaboradas para a análise do resultado (Figura 7 e Apêndice C) apresentam em vermelho as áreas (Alto-Alto) ou (Baixo-Baixo), que indicam uma maior concentração de aspectos associados à precariedade e, por conseguinte, também uma maior concentração de AGSN.

No caso de variáveis para as quais o Índice Local de Moran não apresentou relação aparente com a presença de AGSN, foram realizadas também análises baseadas na sobreposição da variável original com imagens orbitais de alta resolução disponíveis na plataforma *Google Earth*. Esse procedimento teve como intuito verificar se o Índice Local de Moran apresentou melhores resultados para indicar a presença de NUI, ou regiões análogas, do que as variáveis originais.

A Figura 7 apresenta espacializações do Índice Local de Moran aplicado à variável **Rendimento nominal médio per capita por domicílio** no polo da Pesquisa que inclui o Distrito Federal e municípios do entorno (polo Brasília). Os resultados e análises para as demais variáveis encontram-se no Apêndice C. Na Figura 7 é possível observar, para o polo de Brasília, os agrupamentos Alto-Alto, Alto-Baixo, Baixo-Alto e Baixo-Baixo do índice local de Moran sobrepostos aos limites dos AGSN. Foram feitas também análises em maior escala para 6 AGSN selecionados e apresentados nos Quadros de 1 a 6 dispostos na Figura 7.

Figura 7 - Rendimento nominal médio per capita por domicílio no DF\*



\*São apresentadas apenas os resultados que obtiveram mais que 95% de confiança

Fonte: Elaboração própria com base nos dados do Censo IBGE 2010.

Os Quadros 1, 2, 4 e 6 da Figura 7, correspondentes a áreas do Distrito Federal, demonstram relação entre agrupamentos de setores censitários (SC) com baixo rendimento médio e a presença de AGSN. Entretanto, alguns SC com baixo rendimento que correspondem a AGSN não possuem vizinhos que também se destacam pelo baixo rendimento e, por conseguinte, não são detectados pelo Índice Local de Moran como clusters significativos (Quadros 3 e 5).

Na Figura 7 é possível observar uma presença generalizada de regiões classificadas como Baixo-Baixo fora do Distrito Federal, revelando que o índice não é capaz de auxiliar na identificação de NUI nessas áreas. É possível que a dificuldade do índice em realizar uma boa distinção de agrupamentos nos municípios vizinhos se deva ao fato de que o mesmo foi gerado para a região do polo como um todo. Logo, a renda do Distrito Federal se apresentou muito mais alta quando comparada com a renda dos municípios do entorno. Assim, todos os outros municípios se classificaram como de baixo rendimento, formando uma grande mancha de Baixo-Baixo.

As análises de autocorrelação espacial realizadas com a variável Rendimento nominal médio, bem como as apresentadas no Apêndice C, revelam que, embora a utilização dos índices locais de Moran possa indicar regiões de precariedade com possível presença de núcleos urbanos informais, apresentam uma série de limitações para essa finalidade. A primeira delas diz respeito ao fato de que NUI não se localizam, necessariamente, em clusters de precariedade. Outra limitação, de maior impacto, refere-se à possibilidade de geração de vastos clusters de precariedade, como demonstrado quanto à variável “Rendimento nominal médio per capita por domicílio” e demais variáveis apresentadas no Apêndice C. De maneira geral, esses clusters de precariedade ocorrem nos municípios vizinhos aos municípios mais ricos e desenvolvidos da região analisada. O padrão, verificado inicialmente no polo de Brasília, se repete nos municípios vizinhos aos municípios-sede dos demais polos da pesquisa (Belo Horizonte, Porto Alegre, Recife, Juazeiro do Norte e Marabá). Além disso, nos locais apontados pelo Índice Local como clusters de precariedade, percebe-se, ao utilizarmos as variáveis em classes e recortes mais específicos, uma heterogeneidade interna que é de extrema relevância para uma indicação mais precisa de áreas de possíveis NUI.

Assim, conclui-se que os indicadores locais de autocorrelação possuem baixa capacidade de identificar áreas e limites mais exatos e pontuais que possam indicar a presença de NUI. Apesar disso, sugere-se sua utilização de forma complementar a outras análises mais precisas para a identificação de núcleos urbanos informais no âmbito desta Pesquisa, podendo ser útil no processo de caracterização dos NUI.

### **3. Identificação de Núcleos Urbanos Informais: Análise de Metodologias Relacionadas**

#### **3.1 Metodologia do Centro de Estudos da Metrópole para identificação de assentamentos precários**

Os setores de aglomerados subnormais, disponibilizados pelo IBGE para todo o território nacional, têm tido sua metodologia de delimitação aprimorada no decorrer dos anos, a qual inclui informações das prefeituras municipais (quando disponíveis), análises de campo, registros administrativos e bases de dados geoespaciais (IBGE, 2020). Entretanto, em que pese sua relevância, diversos autores vêm indicando a subestimação desses assentamentos nos levantamentos realizados até 2010 (DENALDI, 2013; TASHNER, 2008).

Em resposta a essa reconhecida limitação do dado, o Centro de Estudos da Metrópole (CEM) desenvolveu, a pedido do então Ministério das Cidades, uma metodologia pioneira para a identificação de assentamentos precários. A metodologia baseia-se na utilização da técnica de análise discriminante para a identificação de setores censitários com características semelhantes às dos aglomerados subnormais, mas que não haviam sido classificados pelo IBGE como tais (MARQUES, 2007). A metodologia, aplicada inicialmente para os dados censitários de 2000, foi posteriormente reaplicada para os municípios da Macrometrópole Paulista a partir dos dados de 2010 (CEM/FUNDAP, 2013).

A análise discriminante tem como objetivo discriminar categorias através do uso de diferentes variáveis que as descrevem. No contexto da presente Pesquisa, significa utilizar variáveis que descrevem as características de um NUI para tentar identificar áreas não classificadas que possuem as mesmas características, reclassificando-as como NUI também.

Buscando avaliar a metodologia do CEM perante os objetivos da Pesquisa, buscou-se replicá-la para os dados dos polos. Nesse sentido, foi considerada a descrição da metodologia apresentada na publicação *Assentamentos Precários no Brasil Urbano* (MARQUES, 2007), elaborada pelo Centro de Estudos da Metrópole (CEM). Entretanto, como a publicação não apresenta todos os detalhes dos procedimentos realizados, o primeiro esforço realizado foi no sentido de se apropriar da metodologia, buscando obter os mesmos resultados. Para isso, foram utilizadas exatamente as mesmas unidades espaciais de análise e variáveis apresentadas na publicação, que são provenientes dos resultados do universo do Censo Demográfico 2000.

Os resultados encontrados na réplica foram muito próximos aos originais. Considerando que há um fator aleatório na análise discriminante, condizente com a seleção das amostras para a estimativa da função discriminante, a diferença nos resultados pode ser quase totalmente atribuída a esse fator. Além disso, a pequena diferença pode também ser decorrente da forma como as variáveis são calculadas, uma vez que a metodologia do CEM não a apresenta de forma explícita. Os cálculos efetuados para a construção das variáveis consideradas nas análises realizadas para a presente Pesquisa são apresentados na Tabela 1.

Após a realização da etapa de apropriação da metodologia com os dados do Censo Demográfico 2000, os procedimentos metodológicos foram realizados novamente, dessa vez com variáveis construídas a partir de dados mais recentes, provenientes do Censo Demográfico 2010. Os

procedimentos e resultados descritos a seguir consideram o experimento realizado com dados de 2010.

### **3.1.1 Apresentação das variáveis**

As variáveis escolhidas para a aplicação da técnica foram as mesmas sugeridas pela publicação do CEM e estão descritas na Tabela 1. Todas as variáveis são agregadas por setores censitários.

Entretanto, algumas das variáveis utilizadas pelo CEM, oriundas do Censo Demográfico 2000, não estão presentes nos resultados do universo do Censo 2010. As variáveis sem correspondência em ambos os censos são as seguintes: CEM05, CEM07, CEM12 e CEM13 – duas da dimensão "habitação e infraestrutura" e duas da dimensão "renda e escolaridade do responsável pelo domicílio". Optou-se, assim, por aplicar a técnica de análise discriminante considerando apenas as 15 variáveis restantes como variáveis explicativas.

Tabela 1 - Relação das variáveis explicativas utilizadas para a aplicação da análise discriminante.

Dimensão	ID	Variável
<b>Habitação e infraestrutura</b>	CEM01	% de domicílios sem coleta de lixo
	CEM02	% de domicílios sem ligação à rede de abastecimento de água
	CEM03	% de domicílios sem banheiros ou sanitários
	CEM04	% de domicílios sem ligação à rede de esgoto ou fossa séptica
	CEM05	% de domicílios do tipo cômodo
	CEM06	% de domicílios – outra forma de posse da moradia
	CEM07	% de domicílios – outra forma de posse do terreno
	CEM08	Número de banheiros por habitante
	CEM09	% de responsáveis por domicílio não alfabetizados
	CEM10	% de responsáveis por domicílio com menos de 30 anos não-alfabetizados
<b>Renda e escolaridade do responsável pelo domicílio</b>	CEM11	% de responsáveis por domicílio com renda de até 3 salários mínimos
	CEM12	% de responsáveis por domicílio com menos de 8 anos de estudo
	CEM13	Anos médios de estudo do responsável pelo domicílio
	CEM14	Renda média do responsável pelo domicílio
	CEM15	Número de domicílios particulares permanentes no setor censitário
<b>Aspectos demográficos</b>	CEM16	Número de domicílios improvisados no setor censitário
	CEM17	Número de pessoas residentes no setor censitário
	CEM18	% de responsáveis por domicílio com menos de 30 anos
	CEM19	Número médio de pessoas por domicílio

Fonte: Assentamentos Precários no Brasil Urbano (2007) – CEM/Cebrap.

A Tabela 2 apresenta detalhes do cálculo de cada variável explicativa. Foi atribuído um prefixo para o nome das variáveis do Censo 2010. Isso foi feito para evitar que duas variáveis diferentes recebessem o mesmo nome. Sendo assim, a renomeação seguiu a lógica de colocar o prefixo como um identificador da planilha em que a variável está localizada. Por exemplo, para a variável V002 (Domicílios particulares permanentes) localizada na planilha “Domicílios01\_UF”, foi dado o nome D1V002 - a letra “D” representando a inicial da palavra “Domicílio” e o número “1” indicando em qual das planilhas de domicílio está localizada a variável.

Tabela 2 - Lista das variáveis explicativas utilizadas pelo estudo CEM/Cebrap com as respectivas variáveis utilizadas do censo de 2010 e cálculos realizados para obtê-las.

ID	Variável CEM/Cebrap	Variável(eis) utilizada(s) do Censo 2010	Cálculo
CEM01	% de domicílios sem coleta de lixo	D1V002; D1V035	$1 - (D1V\ 035/D1V\ 002)$
CEM02	% de domicílios sem ligação à rede de abastecimento de água	D1V002; D1V012	$1 - (D1V\ 012/D1V\ 002)$
CEM03	% de domicílios sem banheiros ou sanitários	D1V002; D1V023	$D1V\ 023/D1V\ 002$
CEM04	% de domicílios sem ligação à rede de esgoto ou fossa séptica	D1V002; D1V017; D1V018	$1 - ((D1V\ 017 + D1V\ 018)/D1V\ 002)$
CEM05	% de domicílios do tipo cômodo	NA	NA
CEM06	% de domicílios – outra forma de posse da moradia	D1V002; D1V011	$D1V\ 011/D1V\ 002$
CEM07	% de domicílios – outra forma de posse do terreno	NA	NA
CEM08	Número de banheiros por habitante	D1V025; D1V026; D1V027; D1V028; D1V029; D1V030; D1V031; D1V032; D1V033; D2V001	$(1 \times D1V\ 025 + 2 \times D1V\ 026 + 3 \times D1V\ 027 + 4 \times D1V\ 028 + 5 \times D1V\ 029 + 6 \times D1V\ 030 + 7 \times D1V\ 031 + 8 \times D1V\ 032 + 9 \times D1V\ 033)/D2V\ 001$
CEM09	% de responsáveis por domicílio não alfabetizados	R2V001; R2V093	$R2V\ 093/R2V\ 001$
CEM10	% de responsáveis por domicílio com menos de 30 anos não-alfabetizados	R2V093; R2V094; R2V095; R2V096; R2V097	$1 - ((R2V\ 094 + R2V\ 095 + R2V\ 096 + R2V\ 097)/R2V\ 093)$
CEM11	% de responsáveis por domicílio com renda de até 3 salários mínimos	R1V067; R1V068; R1V069; R1V070; R1V076; R2V001	$(R1V\ 067 + R1V\ 068 + R1V\ 069 + R1V\ 070 + R1V\ 076)/R2V\ 001$
CEM12	% de responsáveis por domicílio com menos de 8 anos de estudo	NA	NA
CEM13	Anos médios de estudo do responsável pelo domicílio	NA	NA
CEM14	Renda média do responsável pelo domicílio	BV005	$BV\ 005$
CEM15	Número de domicílios particulares permanentes no setor censitário	D1V002	$D1V\ 002$
CEM16	Número de domicílios improvisados no setor censitário	DRV001	$DRV\ 001$
CEM17	Número de pessoas residentes no setor censitário	P3V001	$P3V\ 001$
CEM18	% de responsáveis por domicílio com menos de 30 anos	R2V001; R2V093; R2V094; R2V095; R2V096; R2V097	$((CEM10 * R2V\ 093) + R2V\ 094 + R2V\ 095 + R2V\ 096 + R2V\ 097)/R2V\ 001$
CEM19	Número médio de pessoas por domicílio	D1V002; P13V002	$P13V\ 002/D1V\ 002$

Fonte: elaboração própria, 2020.

Para facilitar a localização das variáveis do Censo 2010, segue abaixo a legenda para todos os prefixos utilizados.

- Basico\_UF.xls - B;
- Domicilio01\_UF.xls - D1;
- Domicilio02\_UF.xls - D2;
- DomicílioRenda\_UF.xls - DR;
- Pessoa03\_UF.xls - P3;
- Pessoa13\_UF.xls - P13;
- Responsavel01\_UF.xls - R1;
- Responsavel02\_UF.xls - R2.

Como variável dependente, foi considerada a variável "Tipo\_setor", que classifica cada setor censitário em Aglomerado Subnormal (AGSN) ou Não Especial (NE).

### 3.1.2 Filtros aplicados aos dados

Considerando os dados censitários de 2010, a metodologia de identificação de assentamentos precários proposta pelo CEM foi aplicada aos seis polos de interesse da Pesquisa de Núcleos Urbanos Informais no Brasil: Belo Horizonte, Brasília, Juazeiro do Norte, Marabá, Porto Alegre e Recife. Embora a análise discriminante seja aplicada em todos os setores censitários de interesse, a construção de seu modelo foi realizada a partir de dados filtrados. Foram considerados apenas setores censitários com número de domicílios particulares permanentes maior ou igual a cinquenta ( $CEM15 \geq 50$ ). Além disso, foram considerados apenas setores classificados como urbano ou aglomerado rural com extensão urbana pelo IBGE ("*Situacao\_setor*" = 1,2,3 ou 4).

Entretanto, a aplicação da metodologia resultou na classificação de muitos setores rurais como AGSN. Por conseguinte, a análise discriminante foi realizada novamente com um filtro diferente, que mantém a restrição quanto ao número de domicílios, mas considera apenas setores classificados como urbanos pelo IBGE ("*Situacao\_setor*" = 1,2 ou 3).

A publicação do CEM, diferente do que foi adotado nesse experimento, inclui em seu filtro de dados os setores classificados pelo IBGE como "aglomerados rurais com extensão urbana" ("*Situacao\_setor*" = 4). O filtro original foi utilizado para a recuperação da metodologia, mas para os fins da pesquisa optou-se por alterá-lo, dado que os núcleos urbanos são o foco do estudo.

Com os dados devidamente filtrados, aplicou-se a análise discriminante para todos os setores censitários de cada um dos polos separadamente. A tabela obtida para cada polo possui 16 variáveis: "Tipo\_setor" e as 15 variáveis explicativas apresentadas na Tabela 2.

Importante notar que para alguns setores censitários não é possível obter todas as variáveis explicativas, por conta de valores faltantes ("NA", que significa *Not Available*). Qualquer setor



censitário que possuir uma ou mais variáveis explicativas com valores faltantes deve ser removido da tabela que será usada para a construção do modelo de análise discriminante.

### 3.1.3 Construção do modelo de classificação

Para a construção do modelo de análise discriminante, usa-se uma amostra dos dados e cria-se um subconjunto chamado de “treino”. Na presente análise, o treino corresponde a 75% da quantidade de setores dentro do polo. Para a obtenção deste valor, calcula-se  $0,75 \times n$ , no qual  $n$  é o número de setores censitários dentro do polo (depois da aplicação de todos os filtros anteriores), truncando o resultado. O valor obtido  $X$  representa o tamanho da amostra retirada para treino, amostra essa de natureza aleatória simples e sem reposição. É nesta etapa que ocorre o fator aleatório da técnica e que, por isso, é muito improvável replicá-la obtendo exatamente os mesmos resultados (sem conhecer a amostra selecionada).

Em seguida, aplicou-se a técnica de análise discriminante quadrática para as observações do treino, considerando a variável Tipo\_setor como resposta. Cabe destacar que a publicação do CEM descreve que utilizou a análise discriminante linear nesta etapa, embora os resultados obtidos com a análise discriminante quadrática durante a tentativa de replicar a metodologia tenham gerado valores muito mais próximos dos originais.

Cabe então comparar as premissas entre o uso da análise discriminante linear (ADL) e da análise discriminante quadrática (ADQ). Ambas classificam os dados assumindo que as informações de cada classe (AGSN e NE) foram retiradas de uma população de distribuição Gaussiana. Porém, a ADL assume que as classes possuem a mesma matriz de covariância, enquanto a ADQ assume que as matrizes de covariância são diferentes entre as classes (GARETH et al., 2014).

Na prática, as premissas implicam que a ADL é menos flexível e gera resultados com variância consideravelmente mais baixa – e, portanto, mais precisa. Porém, quando as premissas não são satisfeitas, a ADL gera resultados enviesados. Logo, a ADQ é mais indicada nos casos em que as premissas não podem ser verificadas, correndo menos risco de lidar com problemas de viés.

A ADL é mais indicada para amostras de treino pequenas, enquanto a ADQ, para grandes. Quanto aos resultados, a ADL gera uma função linear e a ADQ, uma função quadrática (GARETH et al., 2014). As funções geradas servem como limites de decisão, ou seja, o valor encontrado na função em determinado setor censitário vai definir a probabilidade dele ser classificado como AGSN e, por complemento, NE. Assim, nos casos em que as variáveis preditoras não são correlacionadas ou são correlacionadas linearmente entre si dentro de cada classe, uma função linear tende a ser mais adequada, enquanto a quadrática é mais adequada para situações nas quais há correlação quadrática entre as variáveis preditoras de cada classe. Essas informações podem ser visualizadas de forma esquemática no Quadro 8.

Cabe salientar ainda que a ADQ é considerada inadequada para casos específicos. Tais casos consistem naqueles caracterizados por possuírem, em pelo menos uma coluna de variáveis, a quantidade de observações (no treino) menor do que a “quantidade de variáveis preditoras +1” (neste caso,  $15+1 = 16$ ). Nesses casos, a ADQ deve ser substituída pela ADL. Assim, optou-se por

utilizar a análise discriminante linear apenas nos casos em que há poucas observações ou grandes quantidades de valores faltantes. Considerando os seis polos de interesse da presente Pesquisa, essa situação ocorreu apenas em um, o de Juazeiro do Norte, que apresenta muitos valores faltantes.

A função estimada por meio da análise discriminante foi aplicada a todos os setores censitários, reclassificando-os como AGSN ou NE. Esses resultados são comparados à classificação original do IBGE, alcançando um dos objetivos da aplicação da metodologia.

A comparação entre a classificação original do IBGE e a classificação resultante da análise discriminante resulta em quatro possibilidades, descritas a seguir:

- AGSN-AGSN: setor classificado como AGSN pelo IBGE e pela análise discriminante;
- AGSN-NE: setor classificado como AGSN pelo IBGE, mas NE pela análise discriminante;
- NE-AGSN: setor classificado como NE pelo IBGE, mas AGSN pela análise discriminante;
- NE-NE: setor classificado como NE pelo IBGE e pela análise discriminante.

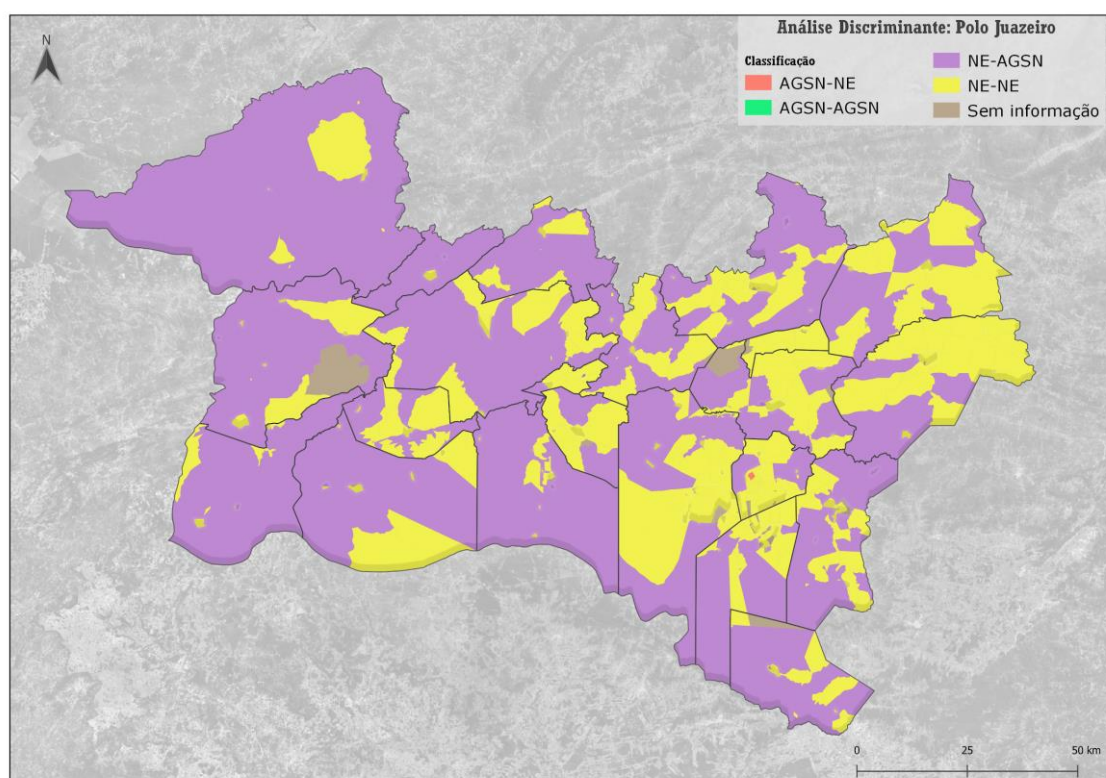
#### **3.1.4 Detalhes de software**

Foi utilizado o software R ver. 3.6.1 para a realização dos cálculos. As funções utilizadas para a obtenção das análises discriminantes foram as seguintes: *qda* (para a análise discriminante quadrática) e a *lda* (para a análise discriminante linear), ambas presentes no pacote MASS versão 7.3-51.5.

#### **3.1.5 Apresentação dos resultados**

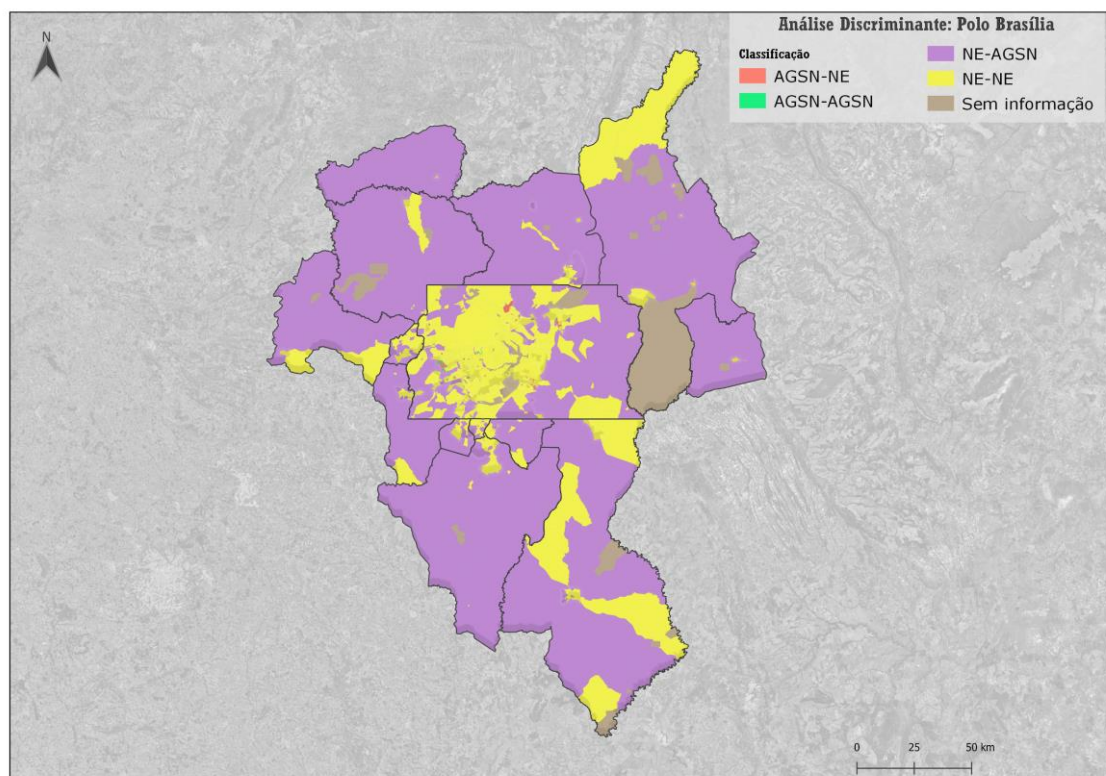
Após a aplicação da técnica nos polos de interesse da Pesquisa, foram obtidos os mapas apresentados a seguir (Figuras 8, 9, 10, 11, 12, 13).

Figura 8 - Resultados da análise discriminante para o polo Juazeiro do Norte.



Fonte: Elaboração própria, 2020.

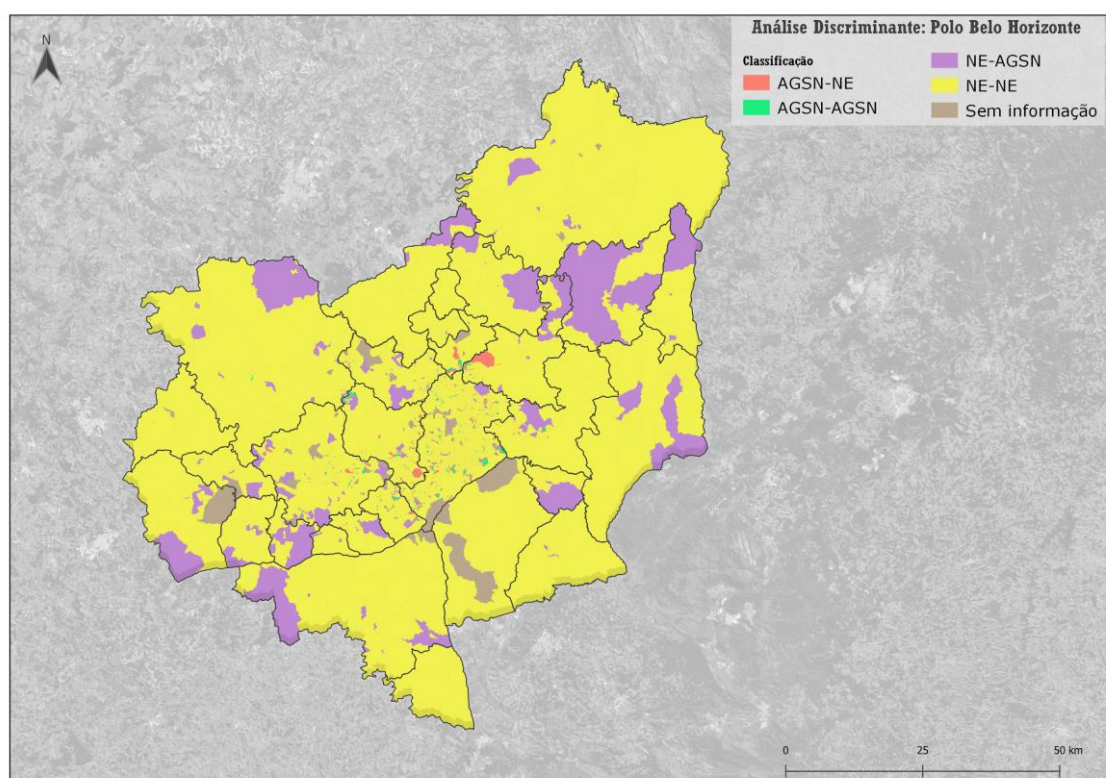
Figura 9 - Resultados da análise discriminante para o polo Brasília.



Fonte: Elaboração própria, 2020.

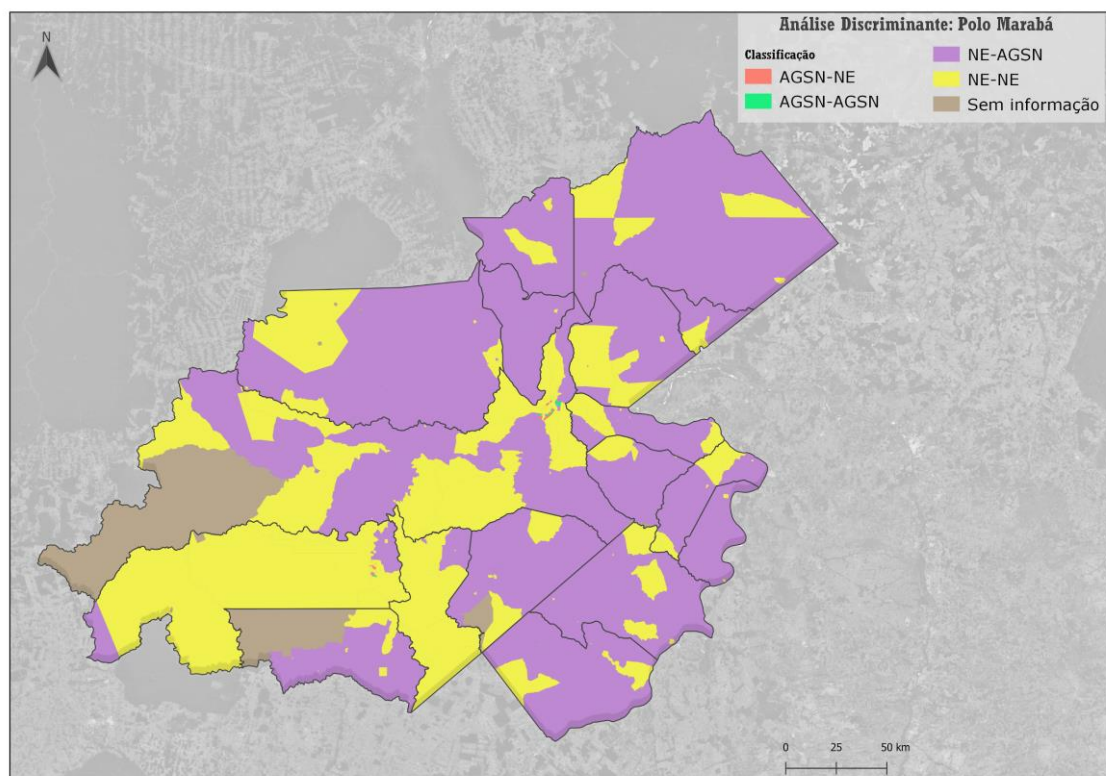


Figura 10 - Resultados da análise discriminante para o polo Belo Horizonte.



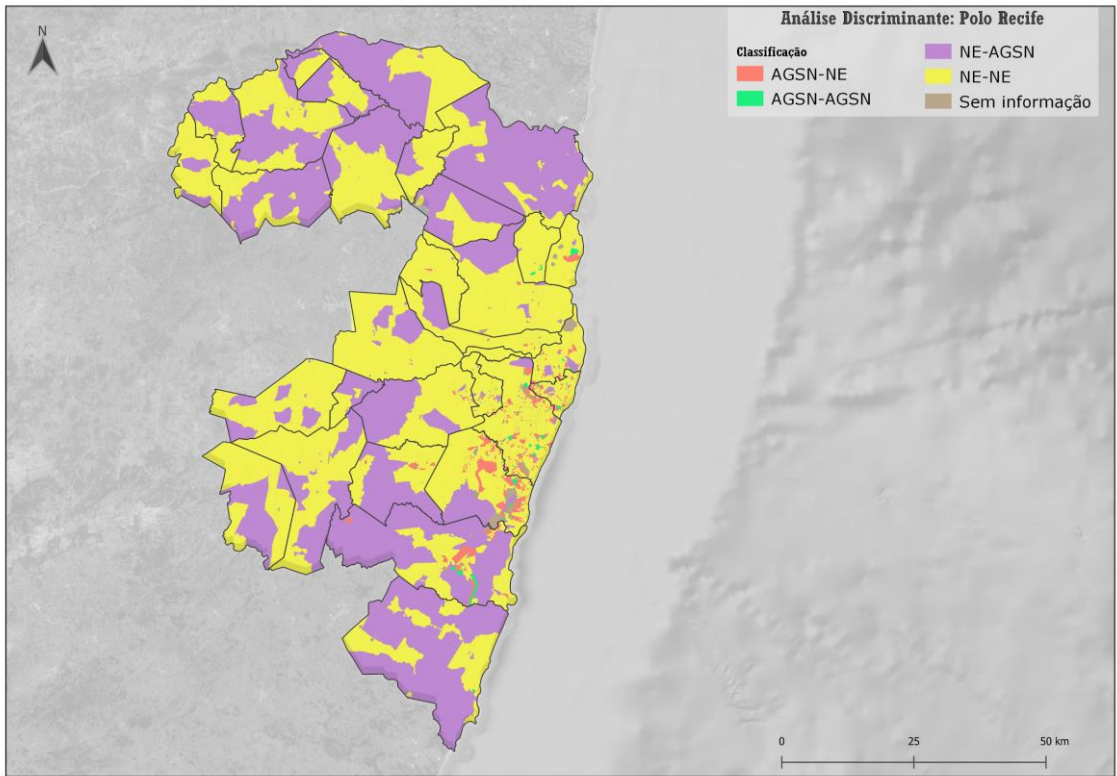
Fonte: Elaboração própria, 2020.

Figura 11 - Resultados da análise discriminante para o polo Marabá.



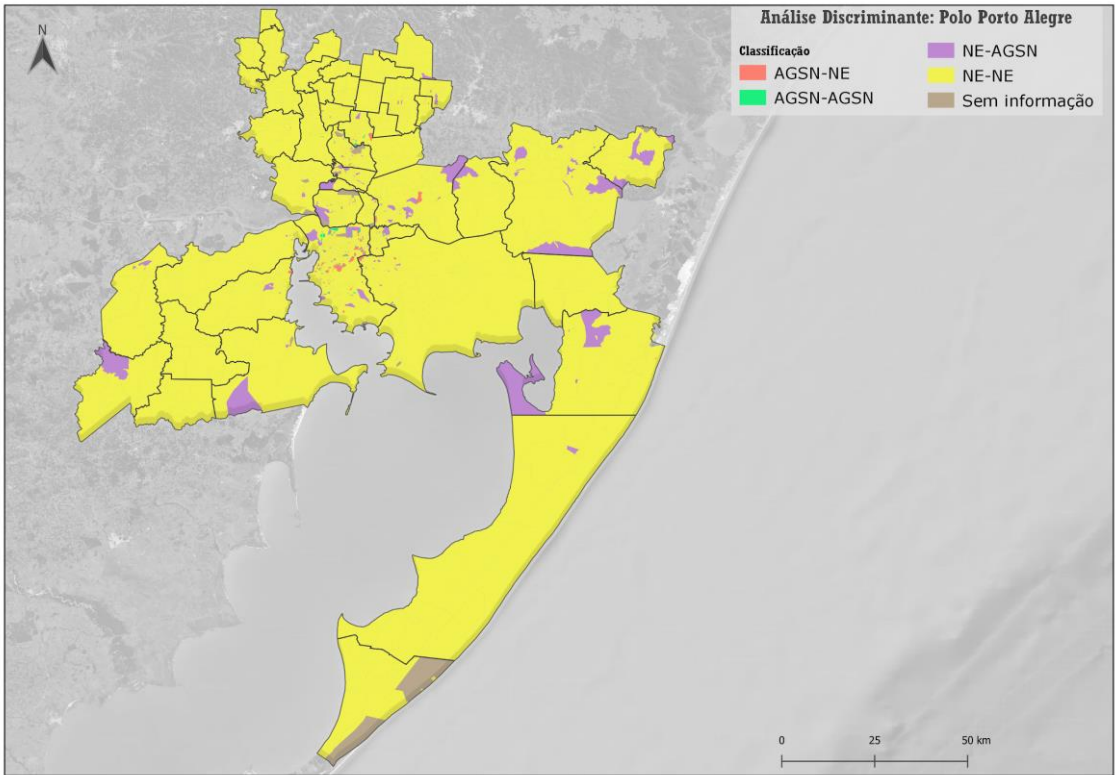
Fonte: Elaboração própria, 2020.

Figura 12 - Resultados da análise discriminante para o polo Recife.



Fonte: Elaboração própria, 2020.

Figura 13 - Resultados da análise discriminante para o polo Porto Alegre.



Fonte: Elaboração própria, 2020.

Para complementar os resultados visuais, as Tabelas 3 e 4 apresentam a quantidade de reclassificações encontradas em cada caso, em números absolutos e em porcentagens.

Tabela 3 - Quantidade (em números absolutos) de reclassificações obtidas através da aplicação da análise discriminante nos seis polos de interesse.

Nome do polo	Reclassificações				NA/ Missing	Total
	AGSN-AS	AGSN-NE	NE-AGSN	NE-NE		
Juazeiro do Norte	0	4	270	1.092	61	1.427
Brasília	38	143	327	5.118	156	5.782
Belo Horizonte	409	387	328	6.347	170	7.641
Marabá	9	36	221	757	22	1.045
Porto Alegre	140	898	297	4.014	104	5.453
Recife	127	220	182	6.283	94	6.906

Fonte: Elaboração própria, 2020.

Tabela 4 - Quantidade (em porcentagens) de reclassificações obtidas através da aplicação da análise discriminante nos seis polos de interesse.

Nome do polo	Reclassificações				NA/ Missing	Total
	AGSN-AGSN	AGSN-NE	NE-AGSN	NE-NE		
Juazeiro do Norte	0,00%	0,28%	18,92%	76,52%	4,27%	100%
Brasília	0,66%	2,47%	5,66%	88,52%	2,70%	100%
Belo Horizonte	5,35%	5,06%	4,29%	83,07%	2,22%	100%
Marabá	0,86%	3,44%	21,15%	72,44%	2,11%	100%
Porto Alegre	2,57%	16,47%	5,45%	73,61%	1,91%	100%
Recife	1,84%	3,19%	2,64%	90,98%	1,36%	100%

Fonte: Elaboração própria, 2020.

Observa-se que em todos os casos o modelo tendeu a ser “conservador”, ou seja, não reclassificou quantitativamente a maioria dos setores NE. Porém, em termos de área reclassificada, os polos Juazeiro do Norte, Brasília e Marabá destacaram-se por conter uma grande mancha que reclassificou setores NE como AGSN. Na maioria dos casos, esses grandes setores são classificados como rurais pelo Censo Demográfico 2010. Em contraste, os polos Belo

Horizonte, Recife e Porto Alegre apresentaram uma mancha maior de setores NE que se mantiveram na mesma classe após a análise discriminante, em convergência com os resultados quantitativos.

Seguindo para análises de resultado quantitativo mais específicas, observa-se que os polos Juazeiro do Norte e Marabá obtiveram os maiores números de reclassificações NE-AGSN, representando 18,92% e 21,15% dos casos, respectivamente; o polo Porto Alegre obteve outro destaque, mas desta vez com reclassificações do tipo AGSN-NE, representando 16,47% dos casos.

Importante dar ênfase aos resultados do polo Juazeiro do Norte, que se mostram pouco confiáveis. Além da grande presença de valores faltantes nos dados censitários, o polo apresenta apenas 4 de seus setores classificados pelo IBGE como AGSN, em 2010. Essas particularidades contribuem para que o modelo de análise discriminante estimado para esse polo apresente uma qualidade inferior.

Conclui-se, por meio desses experimentos, que a metodologia adaptada do CEM apresenta elementos relevantes a serem considerados na construção da metodologia a ser desenvolvida nesta Pesquisa, entre as quais destaca-se a utilização de variáveis disponíveis para todo território brasileiro (variáveis censitárias) e aplicação da análise discriminante como técnica de classificação. Apesar disso, ela ainda não apresenta todos os requisitos necessários para atender às exigências da Pesquisa, conforme será discutido no próximo capítulo deste relatório.

### **3.2 Metodologia para Identificação e Caracterização de Assentamentos Precários em Regiões Metropolitanas Paulistas (MAPPA)**

#### **3.2.1 Descrição da metodologia MAPPA**

Outra metodologia que apresenta elementos relevantes para construção metodológica a ser desenvolvida ao longo desta Pesquisa é a MAPPA - **Metodologia para Identificação e Caracterização de Assentamentos Precários em Regiões Metropolitanas Paulistas**, já apresentada brevemente no Produto 1. A MAPPA, desenvolvida em pesquisa contratada pela CDHU e executada pela UFABC, inovou ao propor procedimentos para a identificação, classificação e caracterização de distintas tipologias de assentamentos precários a partir da utilização intensiva de técnicas de geoprocessamento, estatística e sensoriamento remoto (CDHU e UFABC, 2019a).

O desenvolvimento da MAPPA teve como ponto de partida a definição e caracterização de tipologias de tecidos de assentamentos precários (TECs) a partir da análise de características urbanas e morfológicas de assentamentos precários da Região Metropolitana da Baixada Santista, área caracterizada por grande diversidade de ocupações. Essas análises, realizadas com suporte de interpretação visual de imagens, informações cartográficas e trabalhos de campo, resultaram na definição das seguintes TECs: TEC 1 - Morros; TEC 2 - Palafitas; TEC 3 - Áreas úmidas (APPs de rios e córregos ou aterros de mangues e restingas); TEC 4 - Ocupação

desordenada (sem traçado regulador); TEC 5 - Ocupação ordenada por traçado regulador; e TEC 6 - Ocupação esparsa ou pouco consolidada (CDHU e UFABC, 2019b).

Para a classificação das TECs, a MAPPA adota modelos de regressão logística para geração de superfícies de probabilidade da presença de cada tipologia de tecido de assentamentos precários. A partir das superfícies de probabilidade, constrói-se uma árvore de classificação das tipologias de assentamentos. A regressão logística é uma técnica estatística multivariada utilizada para a classificação de unidades de análise e que se assemelha à análise discriminante, adotada na metodologia proposta pelo CEM (Seção 3.1). Quando as suposições básicas de ambas as técnicas são atendidas, seus resultados classificatórios e preditivos são semelhantes. Entretanto, a regressão logística tem a vantagem de ser menos afetada quando as suposições básicas não são satisfeitas, em específico, a suposição referente à distribuição normal das variáveis (HAIR *et al.*, 2009 apud CDHU e UFABC, 2019a, p. 60).

Outra diferença em relação à metodologia desenvolvida pelo CEM, a qual baseia-se no uso exclusivo de variáveis censitárias, consiste no fato de que os modelos utilizados na MAPPA integram dados de fontes diversas para construção de variáveis representativas de distintos aspectos territoriais: aspectos físico-ambientais e unidades de conservação; malha urbana; população, domicílios e entorno dos domicílios (que incluem variáveis obtidas a partir do Censo Demográfico de 2010) (CDHU e UFABC, 2019a).

Para a integração dos dados provenientes de distintas unidades espaciais e formatos de arquivos, a metodologia MAPPA considerou duas unidades espaciais de análise: Unidades Homogêneas de Uso e Cobertura da Terra (UHCT) e células dispostas na forma de uma grade regular de dimensão de 100x100m. A escolha de utilização de células se deu para que o modelo pudesse ser replicado em áreas com ausência do mapeamento de UHCT, uma vez que as UHCT, até então, foram desenvolvidas apenas para o Estado de São Paulo (CDHU e UFABC, 2019a). Entretanto, variáveis de grau de densidade, consolidação e ordenamento da ocupação, provenientes do mapeamento das UHCT, foram incluídas nos dois modelos.

Para alimentar o modelo como variável dependente, perímetros de assentamentos precários da RMBS foram obtidos a partir da metodologia de mapeamento colaborativo e caracterização dos assentamentos precários desenvolvida pela CDHU (SOUZA; ROSSI; RUDGE, 2018) e sistematizada na plataforma geocolaborativa SIM/Emplasa (SÃO PAULO, s.d.). Esses perímetros foram classificados em tipologias de tecido (TECs) e utilizados na construção das variáveis dependentes dos modelos de identificação e classificação. Os modelos foram desenvolvidos em três etapas: (a) construção e integração de variáveis relevantes para a identificação das TECs; (b) utilização de modelos de regressão logística para a geração de superfícies de probabilidade da presença das distintas TECs; e (c) identificação e classificação dos assentamentos por meio de árvore de decisão.

Na etapa (a) foram construídas e integradas variáveis representativas de distintos aspectos territoriais obtidas a partir de fontes de dados diversas, conforme apresentado no Quadro 4. Os dados utilizados são de livre acesso e disponíveis em plataformas online.



Quadro 4 - Dados utilizados na construção das variáveis.

Dado	Fonte
Modelo Digital de Elevação (MDE), resolução de 30m.	SÃO PAULO (Estado). <i>Modelo Digital de Elevação (MDE) do Estado de São Paulo</i> , 2013.
Hidrografia	SÃO PAULO (Estado). <i>Rede de Drenagem do Estado de São Paulo</i> , 2013.
Unidades de Conservação de Proteção Integral (UC)	BRASIL. Ministério do Meio Ambiente. <i>Cadastro Nacional de Unidades de Conservação</i> , 2017.
Unidades Básicas de Compartimentação do Meio Físico (UBC)	SÃO PAULO (Estado). Instituto Geológico / Coordenadoria de Planejamento Ambiental, 2014.
Mangue	SANTOS WEB ATLAS. <i>Atlas Ambiental e Socioeconômico da Baixada Santista</i> , 2017.
Logradouros	SÃO PAULO (Estado). Secretaria do Meio Ambiente. <i>Base de Dados de Arruamento StreetBase</i> , 2010; <i>OpenStreetMap</i> ®, contribuidores do <i>OpenStreetMap</i> , 2017.
UHCT – Unidades Homogêneas de Uso e Cobertura da Terra	SÃO PAULO (Estado). Instituto Geológico / Coordenadoria de Planejamento Ambiental. <i>Unidades Homogêneas de Uso e Ocupação do Solo Urbano (UHCT) do Estado de São Paulo</i> , 2014.
Censo Demográfico 2010	Instituto Brasileiro de Geografia e Estatística (IBGE). <i>Censo Demográfico</i> , 2010.
Ortofotos	Instituto Geográfico e Cartográfico do Estado de São Paulo (IGC-SP). <i>Ortofotos Leste-SP 2010, 2014; Google Maps / Digital Globe</i> .

Fonte: Adaptado de CDHU e UFABC (2019b).

Os dados apresentados no Quadro 4 foram então integrados em uma mesma base espaço temporal, uma vez que todas as variáveis devem estar disponíveis em unidades espaciais de análise comuns. Como mencionado, para essa integração, foram consideradas duas unidades espaciais de análise: Unidades Homogêneas de Uso e Cobertura da Terra (UHCTs) (Figura 14, a) e células dispostas na forma de uma grade regular de 100mx100m (Figura 14, b).

Figura 14 - Unidades espaciais de análise: UHCT e células.

(a) UHCTs



(b) células



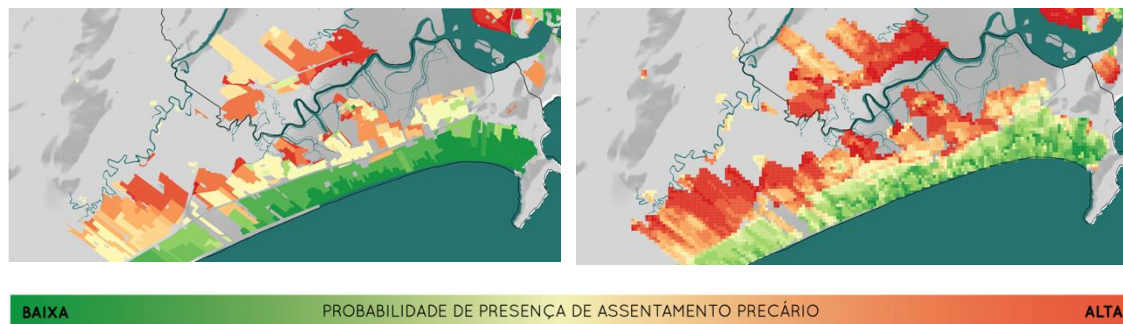
Fonte: Adaptado de CDHU e UFABC (2019b).

Após a integração, foram geradas superfícies de probabilidade da presença de cada uma das TECs a partir da construção de modelos de regressão logística, tanto para as UHCTs como para as células, conforme Figura 15(a) e Figura 15(b), respectivamente. As superfícies de probabilidade foram construídas a partir de modelos correspondentes à presença ou ausência das TECs.

Figura 15 - superfícies de probabilidade da presença de assentamento precário: UHCT e células.

(a) UHCTs

(b) células



Fonte: Adaptado de CDHU e UFABC (2019b).

O Quadro 5 apresenta uma síntese das variáveis incorporadas nos modelos de regressão logística estimados a partir de dados da RMBS. O símbolo “↑” foi utilizado quando a variável aumenta a probabilidade de presença da TEC em questão (correlação positiva), enquanto o símbolo “↓” representa o oposto. Foram construídos modelos específicos para a RMBS e modelos generalizáveis para outras regiões paulistas. Nestes últimos, foram desconsideradas variáveis muito particulares, tais como cobertura de mangue.

Dada a geometria diferenciada das unidades espaciais de análise (UHCT e células), as variáveis consideradas em cada modelo também podem divergir. É importante destacar que foram incluídas nos modelos apenas as variáveis que estavam de acordo com o conceito da TEC em questão, como, por exemplo, a probabilidade de ser uma “TEC 1 - Morros” aumenta em áreas de alta declividade e alto índice de escorregamento.

Para cada TEC, foram estimados quatro modelos:

- específico para a RMBS, com UHCT como unidade espacial (↓ ou ↑);
- Específico para a RMBS, com células (↓ ou ↑);
- Generalizável para outras regiões, com UHCT (↓ ou ↑);
- Generalizável para outras regiões, com células (↓ ou ↑).

Quadro 5 - Variáveis consideradas nos modelos MAPPA.

Variável	Modelos					
	NTEC	TEC 1	TEC 2/3	TEC 4	TEC 5	TEC 6
Declividade de 0% a 15%		↓ ↓ ↓ ↓ ↓		↑ ↑	↑ ↑	
Declividade média	↓		↓ ↓			
Distância de 15 m ao curso d'água			↑ ↑ ↑ ↑			
Distância de 30 m ao curso d'água	↑ ↑ ↑					
Unidades de conservação de proteção integral	↑ ↑ ↑					↑ ↑ ↑ ↑
Média do índice de perigo de inundação						↑ ↑
Perigo de escorregamento alto ou muito alto	↓	↑ ↑				↑ ↑
Área de mangue			↑ ↑			
Curvatura média do terreno		↓ ↓				
Distância de até 50 metros de vias carroçáveis					↑ ↑ ↑ ↑	
Média do índice de regularidade de forma das quadras ou bolsões de ocupação	↓ ↓ ↓ ↓ ↓	↑ ↑ ↑	↑ ↑ ↑ ↑			↑ ↑ ↑
Densidade de ocupação (tamanho dos lotes)	↓ ↓ ↓ ↓ ↓	↑ ↑ ↑ ↑	↑ ↑ ↑ ↑	↑ ↑ ↑ ↑	↑ ↑ ↑ ↑	
Estágio de ocupação		↑		↑ ↑	↑ ↑	↓ ↓ ↓ ↓
Grau de ordenamento urbano	↑ ↑ ↑ ↑	↓ ↓		↓ ↓ ↓ ↓		↓ ↓
Altura média em relação ao curso d'água mais próximo			↓ ↓			
Densidade de moradores em domicílios	↓ ↓ ↓	↑		↑ ↑		
Média do número de moradores em domicílios	↓ ↓ ↓ ↓ ↓	↑ ↑ ↑	↑ ↑ ↑ ↑	↑ ↑	↑ ↑ ↑ ↑	
Densidade de domicílios particulares permanentes					↑ ↑ ↑ ↑	
% de domicílios do tipo casa				↑ ↑ ↑ ↑	↑ ↑ ↑ ↑	
% de domicílios do tipo apartamento	↑ ↑ ↑ ↑					
% de domicílios sem meio-fio		↑				
% de domicílios com esgoto a céu aberto	↓					↑ ↑ ↑
% de domicílios com lixo acumulado nos logradouros		↑ ↑				
% de domicílios sem energia elétrica	↓ ↓					↑
% de domicílios sem medidor de uso exclusivo			↑ ↑			↑ ↑
% de domicílios sem coleta de lixo	↓ ↓ ↓					
% de domicílios sem coleta de lixo direta	↓	↑				
Média de número banheiros por domicílio	↑ ↑					↓ ↓ ↓ ↓
% de domicílios - moradia adequada	↑			↓ ↓		
% de domicílios com esgotamento sanitário via vala		↑ ↑ ↑				
% de domicílios com esgotamento sanitário via fossa rudimentar, vala ou outro escoadouro						↑ ↑ ↑ ↑
% de domicílios com esgotamento sanitário via rio, lago ou mar			↑ ↑ ↑ ↑			
% de pessoas responsáveis com rendimento nominal mensal de até 3 salários mínimos		↑		↑ ↑		
% de domicílios com rendimento nominal mensal domiciliar per capita de até 1/2 salários mínimos	↓ ↓ ↓	↑ ↑	↑ ↑ ↑ ↑	↑ ↑		↑ ↑

Fonte: CDHU e UFABC (2019b).

A metodologia desenvolvida na pesquisa MAPPA se configura como um dos instrumentos mais completos, dentre os estudados e apresentados, para a identificação de assentamentos precários ou, no caso, núcleos urbanos informais. Ao mesmo passo que a MAPPA permite a identificação da probabilidade de a área ser um NUI, ou não, sem se ater necessariamente a uma classificação categórica, a metodologia ainda se propõe a identificar diferentes tipologias de tecido, o que pode indicar diferentes estratégias de regularização e reurbanização dessas áreas.

Para que ocorra a identificação e classificação dessas áreas em assentamentos precários e suas tipologias, de forma categórica, as superfícies de probabilidade devem ser combinadas em árvores de decisão. A MAPPA propõe que os limiares de classificação dessas árvores sejam estabelecidos por analistas, com o auxílio de imagens orbitais e outros dados disponíveis. Dessa

forma, a MAPPA está sujeita a decisões variadas para a escolha desses limiares, dado que não estão definidos *a priori* na metodologia e podem variar dependendo da região.

Uma limitação da MAPPA diz respeito à utilização das UHCT (Unidades Homogêneas de Uso e Cobertura da Terra). As variáveis de grau de densidade, consolidação e ordenamento da ocupação, que apresentaram destacado poder explicativo nos modelos de regressão logística estimados, estão disponíveis apenas para áreas onde ocorreu o mapeamento das UHCT, ou seja, para o Estado de São Paulo. Ademais, as UHCT são áreas definidas e atribuídas de forma manual a partir da interpretação visual de imagens aéreas de alta resolução espacial (SMA, 2010), ou seja, um produto oneroso e de difícil reprodução.

Além disso, a MAPPA foi estabelecida como uma metodologia que tinha como intuito a aplicação apenas no Estado de São Paulo. Dessa forma, os modelos foram ajustados tendo como base os assentamentos precários de São Paulo, não possuindo ainda nenhuma análise sobre sua adequabilidade e aplicabilidade para outras regiões brasileiras. Para que a MAPPA seja replicada em território nacional seria necessária uma base unificada e classificada de assentamentos precários de outras regiões brasileiras, que serviria como variável dependente para o ajuste dos modelos. Trata-se de um trabalho extensivo e ainda não realizado. Em contrapartida, existem os Aglomerados Subnormais do Censo IBGE, que apesar de não possuírem atributos classificatórios de tecidos urbanos, podem proporcionar uma variável dependente genérica.

Um estudo aplicado e simplificado da metodologia MAPPA nos municípios polos da Pesquisa, tendo como variáveis dos modelos apenas dados censitários, e com a utilização dos Aglomerados Subnormais do Censo Demográfico 2010 como variável dependente para o ajuste dos modelos, é apresentado na próxima seção. Nesse estudo, foram geradas superfícies de probabilidade da presença dos núcleos urbanos informais nos seis polos.

### **3.2.2 Aplicação da metodologia MAPPA aos dados da Pesquisa NUI**

Com o intuito de testar a aplicação da metodologia MAPPA e suas superfícies de probabilidade da presença de assentamentos precários para a identificação dos NUI, foi elaborada uma versão preliminar simplificada da metodologia desenvolvida pelo Projeto MAPPA. Dessa forma, assim como as superfícies de probabilidade da MAPPA indicam a probabilidade de as unidades de análise serem determinadas tipologias de assentamentos precários, o modelo de regressão logística desenvolvido nesse experimento estima as probabilidades das unidades de análise serem ou não NUI.

Uma vez que não existe o mapeamento das UHCTs para todo o território nacional e a geração da malha celular demandaria um alto tempo de processamento, o teste realizado considerou setores censitários como unidades espaciais de análise.

Tendo em vista também o caráter de avaliação do modelo, para a elaboração das variáveis preditoras foram utilizados, em sua maioria, dados censitários, apenas incluindo no campo das variáveis físico-ambientais a variável de declividade do terreno (Quadro 6). Todas as variáveis listadas no Quadro 6 foram incluídas nos modelos simplificados das superfícies de probabilidade

de cada um dos seis polos a fim de testar o comportamento das variáveis. Para os próximos experimentos, pretende-se aprimorar a seleção das variáveis, considerando o poder explicativo e a significância estatística de cada uma delas.

Quadro 6 - Lista das variáveis preditoras.

Nome	VARIÁVEL	Fonte
<b>%Ate1/2SM</b>	Porcentagem de domicílios com rendimento nominal mensal domiciliar per capita de até ½ salários mínimos	Censo IBGE 2010
<b>MédiaMoradores</b>	Média do número de moradores em domicílios particulares permanentes	Censo IBGE 2010
<b>%Apartamento</b>	Porcentagem de domicílios particulares permanentes do tipo apartamento	Censo IBGE 2010
<b>%Energia</b>	Porcentagem de domicílios sem energia elétrica	Censo IBGE 2010
<b>%Lixo</b>	Porcentagem de domicílios sem coleta de lixo	Censo IBGE 2010
<b>%LixoDireta</b>	Porcentagem de domicílios sem coleta de lixo direta	Censo IBGE 2010
<b>MédiaBanheiro</b>	Média de número de banheiros por domicílio	Censo IBGE 2010
<b>%Adequada</b>	Porcentagem de domicílios - moradia adequada (Moradia adequada - Domicílio particular permanente com rede geral de abastecimento de água, rede geral de esgoto, coleta de lixo.)	Censo IBGE 2010
<b>%Ilumina</b>	Porcentagem de domicílios do setor com iluminação no entorno	Censo IBGE 2010
<b>%Esgoto</b>	% de domicílios com esgoto a céu aberto	Censo IBGE 2010
<b>Declividade</b>	Declividade média do terreno dentro do Setor Censitário	SRTM

Fonte: Elaboração própria, 2020.

Como variável dependente, optou-se pela utilização dos limites de Aglomerados Subnormais, extraídos do Censo Demográfico de 2010, dado que a presente Pesquisa ainda não dispõe de uma ampla base com limites de núcleos urbanos informais para todos os municípios dos polos, como no caso da metodologia MAPPA. Além disso, tendo em vista o avanço na coleta dos limites de núcleos urbanos informais apenas para os municípios dos polos Brasília, estes foram utilizados no processo de validação do modelo desenvolvido a partir dos AGSN.

Na Tabela 5, a seguir, são apresentados os coeficientes das variáveis e o Nagelkerke  $R^2$  para cada um dos modelos, desenvolvidos individualmente para os respectivos polos.

Tabela 5 - Coeficientes e R quadrado dos modelos preliminares.

	<b>Brasília</b>	<b>Marabá</b>	<b>Recife</b>	<b>Porto Alegre</b>	<b>Belo Horizonte</b>	<b>Juazeiro do Norte</b>
<b>Declividade</b>	0,101*	-0,013	-0,136*	0,137*	0,056*	14,071
<b>Ate1/2SM</b>	-0,009*	-0,010	0,025*	0,053*	0,069*	-8,146
<b>MédiaMoradores</b>	1,074*	-0,058	-0,891*	1,481*	0,640*	230,405
<b>%Apartamento</b>	-0,124	0,020	-0,014*	0,012	-0,030*	0,159
<b>%Energia</b>	-0,097*	-0,247*	-0,236*	-0,386*	-0,441*	13,398
<b>%Lixo</b>	-0,022*	-0,012	-0,037*	-0,096*	-0,078*	33,388
<b>%LixoDireta</b>	0,021	-	0,001	-0,009	0,270*	-33,406
<b>%MédiaBanheiro</b>	-2,139*	-1,156	-0,931*	-4,599*	-4,638*	-614,811
<b>%Adequada</b>	-0,006*	-0,017	0,002	0,019*	0,190*	-0,695
<b>%Ilumina</b>	-	-0,028*	-0,037*	-0,031*	-0,035*	0,030
<b>%Esgoto</b>	0,043*	0,019*	0,010*	0,015*	0,004	13,998
<b>Constante</b>	-4,691*	0,209	4,375*	-2,768*	1,045	-1295,99
<b>R quadrado Nagelkerke</b>	<b>0,493</b>	<b>0,193</b>	<b>0,482</b>	<b>0,478</b>	<b>0,467</b>	<b>1</b>

\*coeficientes significativos ao nível de 5%, p-valor < 0,05.

Fonte: Elaboração própria, 2020.

Os coeficientes das variáveis independentes representam o impacto, ou o peso, que elas possuem na predição da variável dependente. Coeficientes positivos representam correlações positivas com a variável dependente, enquanto coeficientes negativos representam uma correlação negativa. Nesse sentido, esperar-se-ia que variáveis representantes da precariedade habitacional apresentassem correlações positivas com a variável dependente, os Aglomerados Subnormais. Ou seja, que em setores censitários classificados como subnormais a probabilidade de terem, por exemplo, uma renda mais baixa ou uma porcentagem mais alta de domicílios sem coleta de lixo, fosse mais alta do que em setores não precários. Entretanto, essa expectativa não se concretizou em alguns casos, os quais estão grifados em vermelho na tabela.

Para as situações nas quais a variável apresentou um coeficiente significativo com sinal contrário ao esperado, são levantadas hipóteses distintas. A primeira decorre da necessidade de se reestimar os modelos, descartando variáveis explicativas correlacionadas entre si (multicolinearidade do modelo) ou que não contribuem significativamente para melhorar a qualidade dos resultados. A segunda hipótese decorre da qualidade da delimitação dos aglomerados subnormais, que apresenta reconhecidas limitações relacionadas ao subdimensionamento, conforme já exposto no Produto 1 desta Pesquisa. Partindo dessa perspectiva, grande parte das áreas precárias podem não ter sido classificadas como AGSN, o

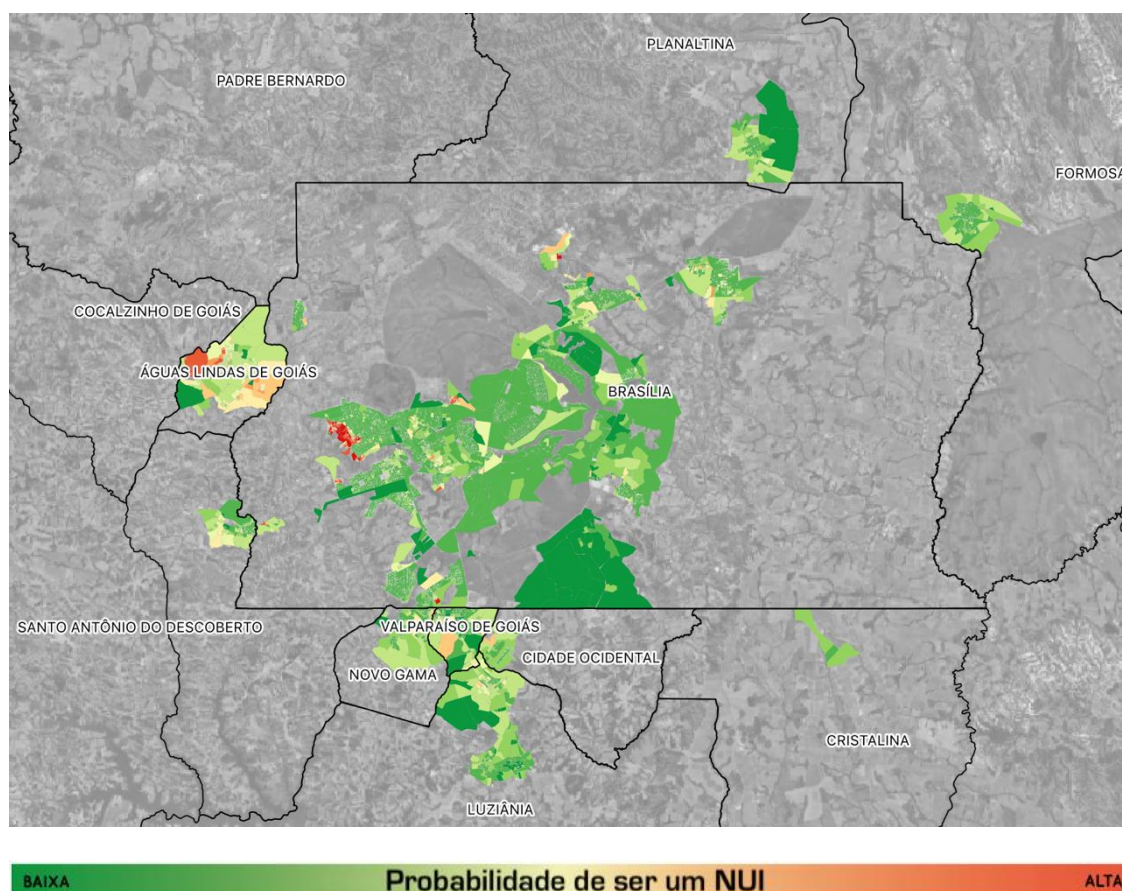
que explicaria níveis de precariedade relativa mais altos em setores não subnormais. Por fim, devemos considerar a heterogeneidade dos municípios que compõem cada um dos polos. Nesse sentido, parte dos aglomerados subnormais de um município sede do polo, como é o caso de Brasília/DF, podem apresentar alguns índices de precariedade mais baixos do que setores censitários não classificados como subnormais em municípios vizinhos mais pobres. Por último, variáveis que apresentam coeficientes com sinais contrários ao esperado devem ser analisadas caso a caso, considerando a realidade local. Caso o problema persista em versões futuras do modelo, a variável poderá ser removida ou substituída por outra similar.

Outro ponto a ser destacado é a impossibilidade de ajuste do modelo de regressão logística a partir dos dados do polo de Juazeiro do Norte, dado que a região possui apenas 4 AGSN, conforme o Censo Demográfico de 2010. Considerando o número reduzido de AGSN, é importante que se analisem alternativas para a construção de modelos para este polo em experimentos futuros. Tais alternativas podem abranger a inclusão de outras fontes de informação para a construção da variável dependente, tais como os dados dos Aglomerados Subnormais de 2019 (IBGE, 2020), ou a aplicação de um modelo construído a partir de dados de outro polo.

A seguir são apresentados os resultados do modelo desenvolvido para o polo de Brasília, representados na forma de mapas de superfície de probabilidade (Figura 16). Com o intuito de avaliar o resultado e a acurácia do modelo, as superfícies de probabilidade foram sobrepostas com a base inicial dos limites de NUI mapeados pelos pesquisadores de campo do polo Brasília.



Figura 16 - Visão geral da superfície de probabilidade do polo Brasília em áreas urbanas.



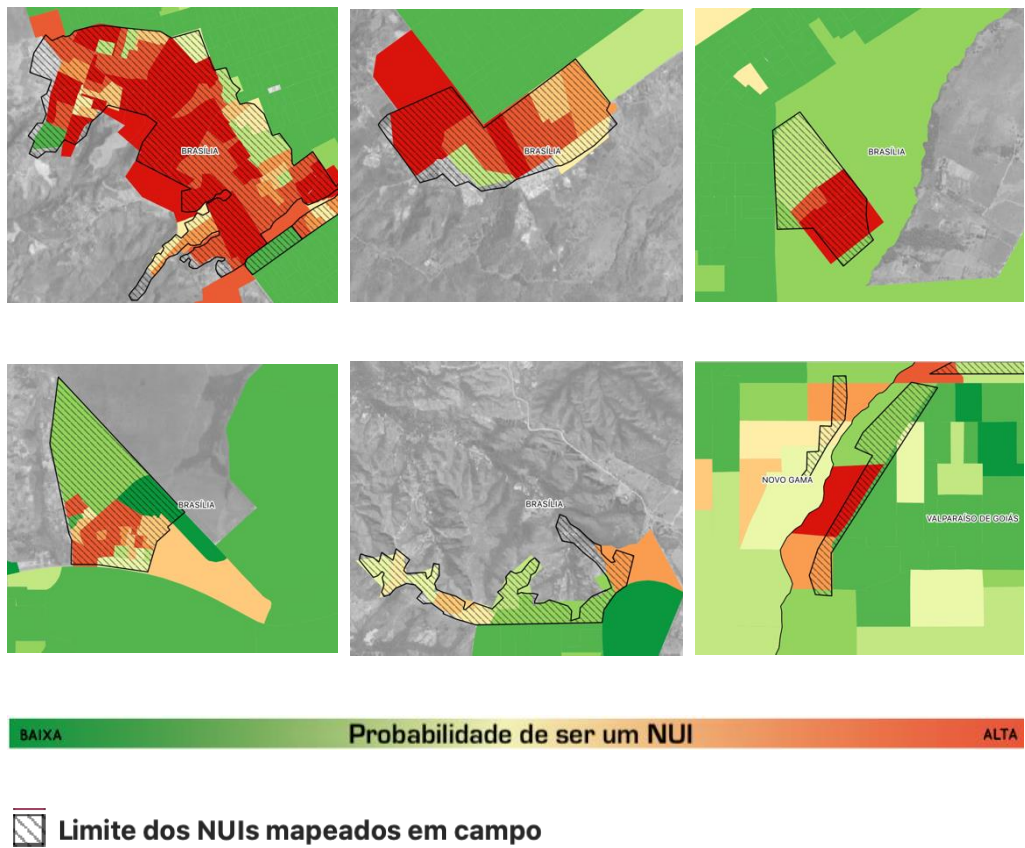
Fonte: Elaboração própria, 2020.

Até a conclusão deste relatório, dentre os 13 municípios e Distrito Federal, que compõem o polo Brasília, apenas os municípios Cidade Ocidental, Novo Gama e Valparaíso de Goiás, assim como o Distrito Federal, possuíam dados de NUI mapeados em campo. A seguir, são apresentadas as análises das superfícies de probabilidade referentes a estes casos.

A partir do cruzamento da superfície de probabilidade com os limites de NUI mapeados em campo temos três situações possíveis: a) áreas mapeadas como NUI onde o modelo apresenta alta probabilidade de ser NUI (Figura 17); b) áreas mapeadas como NUI onde o modelo apresenta baixa probabilidade de ser NUI (Figura 18) e; c) áreas com alta probabilidade de ser NUI que ainda não foram mapeadas como NUI (Figura 19). A seguir apresentaremos imagens onde ocorrem essas situações.

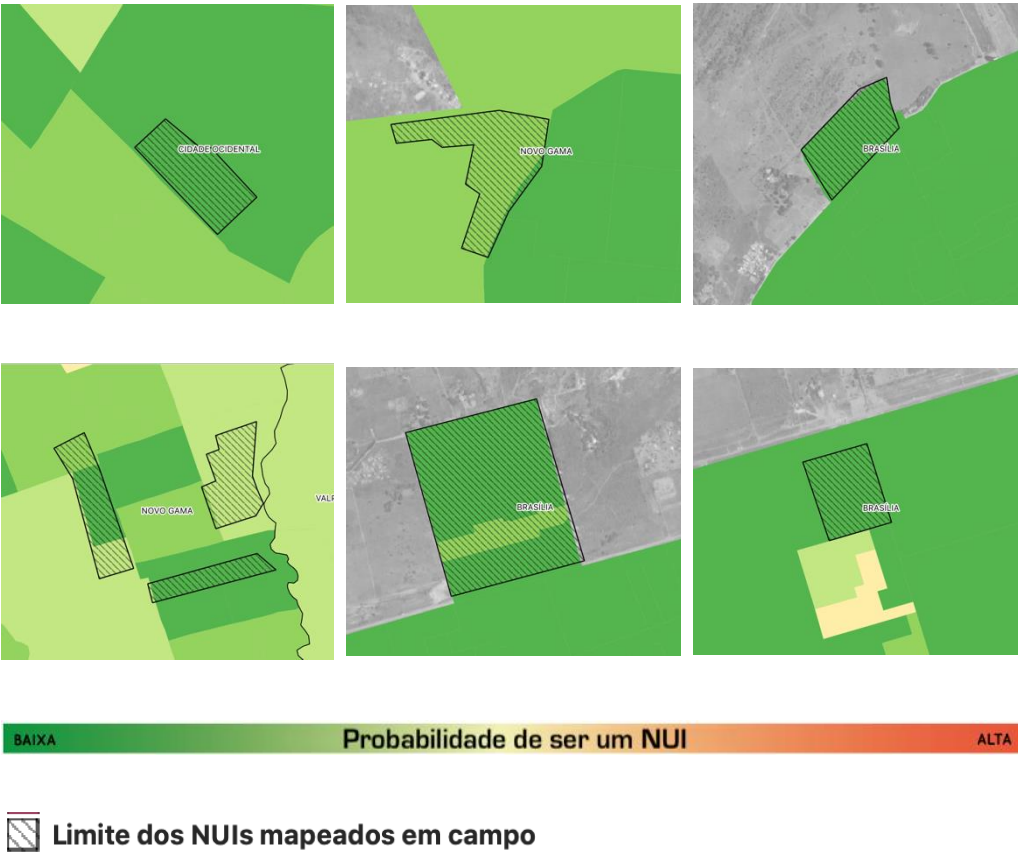


Figura 17 - Situação A: áreas mapeadas como NUI onde o modelo apresenta alta probabilidade de ser NUI.



Fonte: Elaboração própria, 2020.

Figura 18 - Situação B: áreas mapeadas como NUI onde o modelo apresenta baixa probabilidade de ser NUI.



Fonte: Elaboração própria, 2020.

Figura 19 - Situação C: áreas com alta probabilidade de ser NUI que ainda não foram mapeadas como NUI.



 **Limite dos NUIs mapeados em campo**

Fonte: Elaboração própria, 2020.

Com base na interpretação das Figuras 17, 18 e 19, podemos observar que o modelo teste de superfície de probabilidades, utilizando como suporte a metodologia MAPPA, apresentou aspectos promissores, embora ainda apresente resultados limitados, que deverão ser avaliados após os esforços de desenvolvimento metodológico planejados para as próximas etapas da pesquisa (Capítulo 4). Cabe ressaltar, entretanto, que resultados divergentes como os apresentados nas Figuras 18 e 19 são esperados em modelos de classificação dessa natureza.

Por conseguinte, é fundamental que a metodologia a ser desenvolvida e os modelos a partir dela estimados não sejam somente avaliados a partir de uma perspectiva determinística, focada na identificação de áreas classificadas corretamente ou erroneamente.

As superfícies de probabilidade, um dos produtos resultantes do processo de modelagem, deverão ser compreendidas como um plano de informação adicional sobre os NUI. Esse plano de informação, longe de representar uma verdade absoluta, serve como contraponto às informações oficiais sobre NUI, tais como os AGSN do IBGE ou informações disponibilizadas por órgãos municipais. Situações de convergência entre distintos planos de informação revelam condições de menor incerteza, ao passo que situações de divergência deverão ser investigadas com maior profundidade, que podem incluir estratégias de verificação em campo ou remota (por meio de imagens orbitais de alta resolução).

Para além das superfícies de probabilidade, cabe valorizar ainda o processo de construção dos modelos de classificação, incluindo uma análise minuciosa das variáveis identificadas como relevantes em cada um deles. Observar o papel de cada variável em cada modelo estimado é um aspecto fundamental para ampliar a compreensão das condições relacionadas à presença de NUI e como elas se diferenciam em diferentes regiões do país.

## 4. Decisões metodológicas e encaminhamentos

### 4.1 Aspectos relevantes para o desenvolvimento da metodologia NUI

A partir da apropriação e análise das metodologias desenvolvidas pelo CEM e pela UFABC/CDHU para a classificação de assentamentos precários, foi possível realizar uma comparação de diferentes aspectos inerentes a esses estudos e um balanço de suas limitações e potencialidades em relação aos objetivos da presente Pesquisa. A partir dessa comparação e buscando conciliar as contribuições das duas metodologias, foram estabelecidas características desejáveis ou viáveis a serem consideradas na construção da metodologia a ser desenvolvida nesta Pesquisa, a qual será referenciada neste relatório como *metodologia NUI*. Essas características deverão servir como referência para a continuidade do trabalho.

O primeiro aspecto analisado diz respeito ao objetivo das metodologias. A metodologia CEM busca identificar assentamentos precários, ao passo que a MAPPA busca não apenas identificá-los como também classificá-los em distintas tipologias de tecido urbano. A metodologia NUI buscará identificar e caracterizar os núcleos urbanos informais de baixa renda. Para tanto, será avaliada uma abordagem constituída por duas etapas. A primeira delas consiste na construção de modelos de classificação para a identificação de núcleos urbanos informais. Essa etapa resultará em uma função, a ser estimada para cada polo, que representará o que há de comum nas características dos NUI desses polos. A segunda etapa consistirá na caracterização dos NUI identificados a partir de indicadores construídos com variáveis diversas, que representam distintas dimensões das necessidades habitacionais, mesmo que essas variáveis não tenham apresentado significância estatística para sua inclusão no modelo de classificação. Ao contrário da primeira etapa, que representa o que há de comum, essa segunda etapa deverá dar visibilidade à heterogeneidade dos NUI.

Apesar do estabelecimento de duas etapas para a construção da metodologia NUI, as demais considerações desta seção dizem respeito apenas à primeira delas, a construção de modelos de classificação, visto que é a etapa mais complexa e que apresenta correspondência com as metodologias CEM e MAPPA.

Em relação à abrangência geográfica, a metodologia CEM foi desenvolvida para todo o território nacional, ao passo que a MAPPA foi construída para o Estado de São Paulo e considera tipologias de assentamentos precários que, embora possam servir de referência para outras regiões, foram desenvolvidas a partir da realidade do estado. A metodologia NUI, apesar de construída a partir dos seis polos da pesquisa, deverá ser passível de generalização para todo o país.

A abrangência geográfica das metodologias tem relação com os dados considerados no processo de classificação. Enquanto a metodologia CEM utiliza dados disponíveis para todo o Brasil, a MAPPA foca na utilização de dados disponíveis para o Estado de São Paulo, que possui uma infraestrutura de dados espaciais mais completa do que a maioria dos estados brasileiros, o que permite a construção de uma diversificada gama de variáveis. A metodologia NUI deverá considerar somente variáveis que possam ser obtidas a partir de dados disponíveis para todo território nacional.

Quanto à natureza dos dados, a metodologia CEM utiliza apenas dados censitários, enquanto a MAPPA integra dados de distintas fontes, incluindo tanto dados provenientes do censo demográfico como também dados relativos à malha urbana, físico-ambientais e de unidades de conservação. A metodologia NUI, assim como a MAPPA, também deverá considerar dados de fontes e naturezas diversas.

A natureza dos dados está relacionada a outro aspecto metodológico relevante: a unidade espacial de análise. A metodologia CEM adota o setor censitário como unidade espacial de análise. A escolha pelo setor censitário é coerente, visto que a metodologia utiliza apenas variáveis provenientes do questionário básico do censo demográfico, já agregadas por setor censitário. Entretanto, cabe salientar que a delimitação do setor censitário segue princípios operacionais, que não dialogam necessariamente com a forma urbana e, por conseguinte, pode apresentar geometrias muito divergentes das poligonais dos assentamentos precários. Outra limitação do uso do setor censitário como unidade espacial de análise é que ele dificulta a integração de dados de fontes distintas, um aspecto que não era relevante no caso da metodologia CEM.

A MAPPA, por utilizar dados diversos, envolve a problemática da integração de dados e demanda maior atenção na escolha da unidade espacial de análise. A metodologia explora duas unidades espaciais de análise distintas: as Unidades Homogêneas de Uso e Cobertura da Terra (UHCT) e células. A vantagem do uso da UHCT é que ela, por ter sido delimitada a partir de ortofotos, seguindo critérios morfológicos, dialoga diretamente com a forma urbana e frequentemente coincide com os limites de assentamentos e loteamentos, incluindo os precários. No entanto, assim como o setor censitário, a UHCT não representa uma unidade espacial adequada para a integração de dados diversos. Outra limitação da UHCT diz respeito ao fato de ela estar disponível apenas para o Estado de São Paulo. Já as células podem ser geradas facilmente para qualquer área de estudo e são unidades mais adequadas para a integração de dados. Caso não tenham dimensões muito grandes, as células constituem unidades espaciais flexíveis, "pixels" que podem indicar a presença de NUI e que, aglomerados, representam também sua abrangência em termos de área.

Dado que a metodologia NUI pretende utilizar dados diversos, que demandam integração, a célula apresenta-se como alternativa mais adequada e deve ser explorada de forma mais detalhada nas próximas etapas da pesquisa, conforme será apresentado na próxima seção (4.2).

Outro aspecto metodológico importante diz respeito à forma de apresentação dos resultados da classificação. Na metodologia CEM, os resultados da classificação são apresentados de forma binária, ou seja, os setores censitários são classificados como assentamentos precários ou não. Na MAPPA, os resultados são apresentados de duas formas distintas: contínua, como superfícies que indicam a probabilidade da presença de determinada tipologia de assentamento precário; e categórica, na qual as UHCTs ou células são classificadas como não precárias ou como uma das tipologias de assentamento precário estabelecidas (no total, 6 classes). Na metodologia NUI pretende-se priorizar a representação contínua, por meio de superfícies de probabilidade da

presença de NUI. Caso considere importante, o analista poderá escolher um limiar de probabilidade para obter uma classificação binária (presença/ausência de NUI).

Relacionada à forma de apresentação dos resultados, está a escolha da técnica de classificação adotada na metodologia. Na metodologia CEM optou-se pela análise discriminante, ao passo que a MAPPA adotou a análise de regressão logística. Embora envolvam procedimentos estatísticos distintos, as duas técnicas são análogas quanto ao tipo de resultado produzido e ambas permitem a estimativa de probabilidades. Por conseguinte, ambas são promissoras para a metodologia NUI e deverão ser comparadas e avaliadas nas próximas etapas da pesquisa, desta vez a partir de um mesmo conjunto de dados (ver Seção 4.3).

O Quadro 7 sintetiza a comparação entre as metodologias CEM e MAPPA, indicando diretrizes e encaminhamentos para o desenvolvimento da metodologia NUI.

Quadro 7 - Aspectos comparativos entre as metodologias CEM, MAPPA e NUI.

<b>Aspecto Analisado</b>	<b>Metodologia CEM (MARQUES, 2007)</b>	<b>MAPPA (UFABC/CDHU, 2019)</b>	<b>Metodologia NUI <i>Decisões e Encaminhamentos</i></b>
Objetivo	Identificar assentamentos precários	Identificar assentamentos precários e classificá-los em distintas tipologias de tecido urbano (TECs)	Identificar núcleos urbanos informais e caracterizá-los
Abrangência geográfica	Território nacional	Território do Estado de São Paulo	Território nacional
Disponibilidade dos dados	Dados abertos para o território nacional	Dados abertos para o Estado de São Paulo	Dados abertos para o território nacional.
Tipo de dados	Uso exclusivo de dados censitários	Dados provenientes de fontes e naturezas diversas	Dados provenientes de fontes e naturezas diversas
Unidade espacial de análise	Setor censitário: caráter operacional, que não necessariamente dialoga com a forma urbana; não é adequado para integração de dados	UHCT e célula: tem aderência à forma urbana (UHCT) ou flexibilidade para tal (célula); facilita integração de dados (célula)	Célula: adequada para integração de dados e tem flexibilidade para a representação de NUI
Apresentação dos resultados	Classificação binária (precário ou não)	Resultados apresentados de forma contínua (superfícies de probabilidade) e de forma categórica (6 classes, uma para cada tipologia).	Resultados apresentados de forma contínua (superfícies de probabilidade da presença de NUI), que podem ser categorizados após a escolha de um limiar de probabilidade
Técnica de classificação	Análise discriminante	Análise de regressão logística	Análise discriminante ou regressão logística (a definir)

Fonte: Elaboração própria (2020) com base em Marques (2007) e UFABC/CDHU (2019a).

## 4.2 Construção e integração de variáveis

Partindo das diretrizes estabelecidas no Quadro 7, os dados a serem considerados na metodologia NUI deverão ser abertos e disponíveis para todo território nacional, provenientes de fontes diversas e integrados em unidades espaciais de análise do tipo célula. A presente seção apresenta o processo, em andamento, de levantamento desses dados e de construção e integração de variáveis consideradas potencialmente relevantes para o desenvolvimento de modelos de classificação de NUI para os 6 polos da pesquisa.

A partir dos levantamentos e testes apresentados nos capítulos 2 e 3 foram selecionados os seguintes dados para a pesquisa:

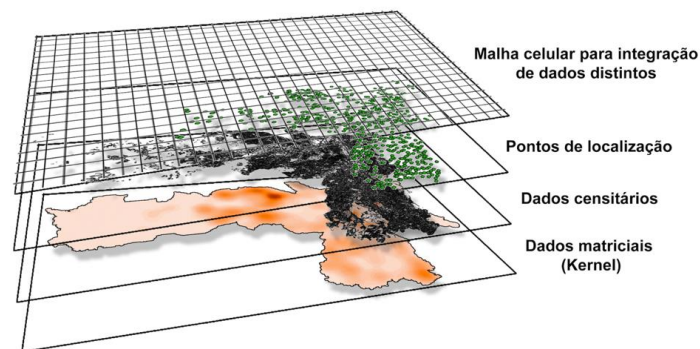
- Censo Demográfico 2010, que abrange informações relacionadas à população, domicílios e entorno dos domicílios. Serão considerados dados disponibilizados em diferentes escalas, com ênfase para os resultados do universo agregados em setores censitários. Dados provenientes da amostra também serão considerados, porém agregados por Unidades de Desenvolvimento Humano (UDH), conforme apresentado na Seção 2.1. Considerando a maior homogeneidade dessas unidades em relação às áreas de ponderação, espera-se que as variáveis provenientes desses dados, mesmo que originalmente agregadas em unidades espaciais abrangentes, possam servir como *proxies* da situação em unidades espaciais menores (células). Os dados disponibilizados por meio da grade estatística (número de habitantes e domicílios) também serão considerados (IBGE, 2010);
- Aglomerados Subnormais 2019 (IBGE, 2020), que serão testados na composição da variável dependente dos modelos (presença de NUI);
- Modelos Digitais de Terreno provenientes da *Shuttle Radar Topography Mission* (SRTM), a partir dos quais foram processadas informações de declividade e curvatura do terreno (NASA, 2010);
- Logradouros extraídos da plataforma *OpenStreetMap*, a partir dos quais foram computadas variáveis relacionadas à regularidade da forma de quadras e bolsões de ocupação e distância de vias carroçáveis (OPENSTREETMAPS, 2020);
- Dados de hidrografia da Fundação Brasileira para o Desenvolvimento Sustentável (FBDS), a partir dos quais foram gerados dados sobre a proximidade de cursos d'água (FBDS, 2015);
- Unidades de conservação de proteção integral, disponibilizados pelo Ministério do Meio Ambiente (MMA, 2020);
- Faixas de servidão de linhas de alta tensão, disponibilizadas pela Agência Nacional de Energia Elétrica (ANEEL, 2020);
- Faixas de servidão de dutos, disponibilizadas pela Agência Nacional do Petróleo, Gás Nacional e Combustível (ANP, 2020).

Considerando as variáveis disponíveis para esses dados ou passíveis de processamento a partir deles, foram sistematizadas aquelas potencialmente relevantes para a identificação de NUI), as quais serão testadas como variáveis explicativas durante o processo de construção de modelos de classificação. As variáveis levantadas são apresentadas no Apêndice B.



A heterogeneidade dos dados e variáveis a serem considerados na construção dos modelos NUI impõe dificuldades para a integração dos mesmos, pois envolve unidades espaciais de análise de distintas fontes e formatos. Por exemplo, variáveis censitárias são disponibilizadas por setores censitários, unidades de conservação são apresentadas na forma de polígonos condizentes com sua localização e dimensão e dados sobre declividade ou curvatura do terreno são gerados para células. Para que análises possam ser conduzidas a partir desses dados, é necessário integrá-los em uma mesma base espaço-temporal, visto que o processo de modelagem a ser conduzido requer que todas as variáveis compartilhem da mesma unidade espacial de análise. Conforme discutido na seção anterior, estabeleceu-se que a integração dos dados e variáveis para a metodologia NUI considerará a construção de grades celulares, compostas por células regulares (Figura 20).

Figura 20 - Imagem esquemática da integração de dados distintos em uma grade celular.



Fonte: Frizzi e Pinho (2017).

A grade celular é um dado vetorial formado por um conjunto de células, cada uma representada por um polígono associado a vários atributos. A resolução das células pode ser alterada e é definida considerando os dados disponíveis, a escala definida para análise e a capacidade de processamento dos recursos computacionais disponíveis (CDHU e UFABC, 2019a). No caso da metodologia NUI, será estudada a possibilidade de se estabelecer grades cujas células sejam subdivisões da grade estatística do IBGE. A grade estatística do Censo Demográfico de 2010, publicada em 2016, foi gerada com o intuito de disseminar dados estatísticos a partir dos microdados do universo do Censo (IBGE, 2016). Em áreas urbanas, as células da grade estatística apresentam dimensões de 200m X 200m. Para a metodologia NUI, pretende-se testar células de 100m X 100m. Assim, 4 células da metodologia NUI estarão contidas em 1 célula urbana da grade estatística. Dessa forma, variáveis obtidas a partir da grade estatística, tais como densidade populacional ou de domicílios, poderão ser transferidas para a célula NUI de forma direta.

Embora análises preliminares tenham indicado alguns problemas em relação aos dados disponíveis para a grade estatística, particularmente em áreas sem um sistema viário bem definido, essa decisão metodológica visa estabelecer uma base celular permanente, que atenda estudos temporais, incluindo aqueles que usem dados da grade estatística do próximo censo demográfico.

### 4.3 Técnicas de Classificação

Uma vez finalizado o processo de construção e integração de variáveis em uma base celular comum, as duas técnicas estatísticas de classificação exploradas no decorrer da Pesquisa - análise discriminante e regressão logística - deverão ser analisadas e comparadas sob as mesmas condições, ou seja, considerando o mesmo conjunto de variáveis resposta e explicativas e as mesmas áreas (6 polos da pesquisa).

A partir dos resultados da análise comparativa das superfícies de probabilidade geradas pelas duas técnicas, uma delas deverá ser escolhida para compor a metodologia NUI.

Conforme apresentado no Capítulo 3, tanto a análise discriminante quando a regressão logística são técnicas que permitem classificar (e, portanto, identificar) espacialidades como AGSN ou NUI. Entretanto, embora sejam análogas, cada uma das técnicas possui características próprias, vantagens e desvantagens.

A análise discriminante assume, para estimar os coeficientes da função de classificação resultante, que os dados de cada uma das classes que se pretende classificar são provenientes de uma população de distribuição aproximadamente Gaussiana. Já a regressão logística não necessita dessa premissa, usando o estimador de máxima verossimilhança para as estimativas (GARETH et al., 2014).

O Quadro 8 apresenta uma comparação entre a análise discriminante em duas de suas variações - linear e quadrática - e a regressão logística, destacando em quais situações cada uma delas tende a ser mais adequada (↑) ou menos adequada (↓).

Quadro 8 - Comparação de performance (↑ - tende a ser adequada; ↓ - tende a não ser adequada) entre os métodos de classificação: análise discriminante linear (ADL), análise discriminante quadrática (ADQ) e regressão Logística (RL).

Condição	ADL	ADQ	RL
População das classes seguem distribuição Gaussiana	↑	↑	↑
População das classes não seguem distribuição Gaussiana	↓	↓	↑
Há correlação linear entre as variáveis preditoras de cada classe	↑	↓	↑
Há correlação quadrática entre as variáveis preditoras de cada classe	↓	↑	↓
Não há correlação entre as variáveis preditoras de cada classe	↑	↓	↑
Amostra de treino muito grande (condição válida apenas para AD)	↓	↑	-
Amostra de treino muito pequena (condição válida apenas para AD)	↑	↓	-

Fonte: Elaboração própria, 2020.

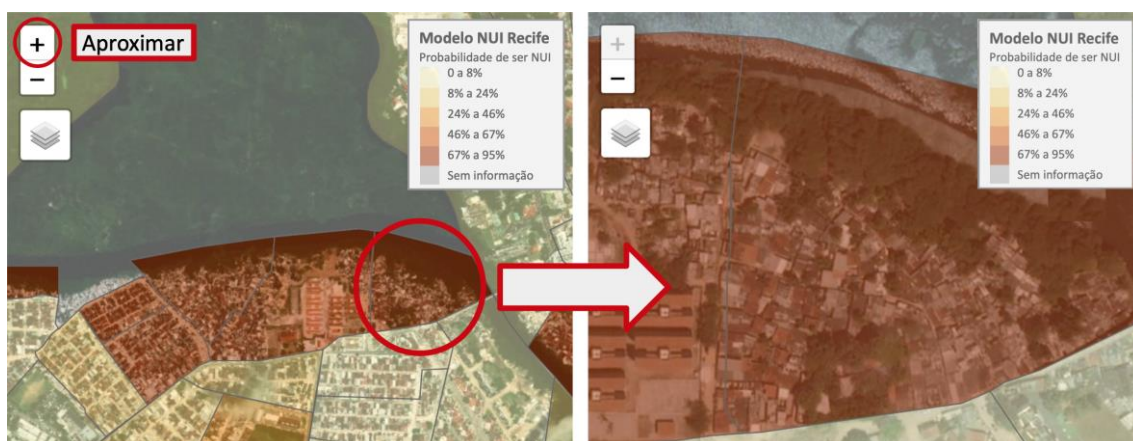
#### 4.4 Mapas interativos

Um importante e desafiador aspecto da modelagem diz respeito à comunicação de seus resultados. A modelagem espacial, apesar de ser uma representação generalizada da realidade, pode envolver enormes conjuntos de dados combinados e sintetizados por meio de uma função matemática (por exemplo, função discriminante ou função logística). A apresentação de seus resultados não é uma tarefa trivial e demanda especial atenção. Às unidades espaciais de análise consideradas do modelo, sejam elas células ou perímetros operacionais, estão associadas todas as variáveis que compõem o modelo e os resultados do cômputo da função que o representa. Os resultados, assim, envolvem extensas tabelas compostas por diversas colunas, que representam cada uma das variáveis, e com tantas linhas quanto o número de unidades espaciais de análise.

Entretanto, os modelos costumam ser apresentados apenas na forma de funções, com seus respectivos coeficientes estimados e testes estatísticos, e mapas estáticos que espacializam o resultado da aplicação do modelo. É comum, assim, o surgimento de dúvidas em relação aos resultados da modelagem. No caso particular da modelagem para identificação de núcleos urbanos informais, são frequentes as dúvidas sobre quais as condições (variáveis) que levaram uma determinada região a apresentar alta probabilidade de ser um NUI. O atendimento de dúvidas como essas demanda formas alternativas e mais interativas de comunicação dos resultados.

A produção de mapas, ou a cartografia, é uma arte milenar que envolve intuição, comunicação e um elemento de criatividade. Os mapas estáticos têm uma comunicação direta e não permitem interações com o seu usuário. Os mapas interativos, por sua vez, elevam a comunicação dos resultados a outro nível. A interatividade desses mapas pode assumir muitas formas, a mais comum é a capacidade de mover e ampliar qualquer parte de um conjunto de dados geográficos sobreposto em uma imagem de satélite para mostrar o contexto urbano. Assim, um usuário pode visualizar, em distintas escalas, os padrões de tecido urbano nas áreas apresentadas pelo modelo como um provável NUI, o que auxilia no processo de avaliação dos resultados da modelagem. A Figura 21 ilustra alguns testes realizados no âmbito desta Pesquisa que avançam nesse sentido, permitem a visualização de imagens orbitais de alta resolução em diferentes níveis de aproximação, o que facilita a identificação do padrão construtivo predominante em NUI de diferentes extensões territoriais.

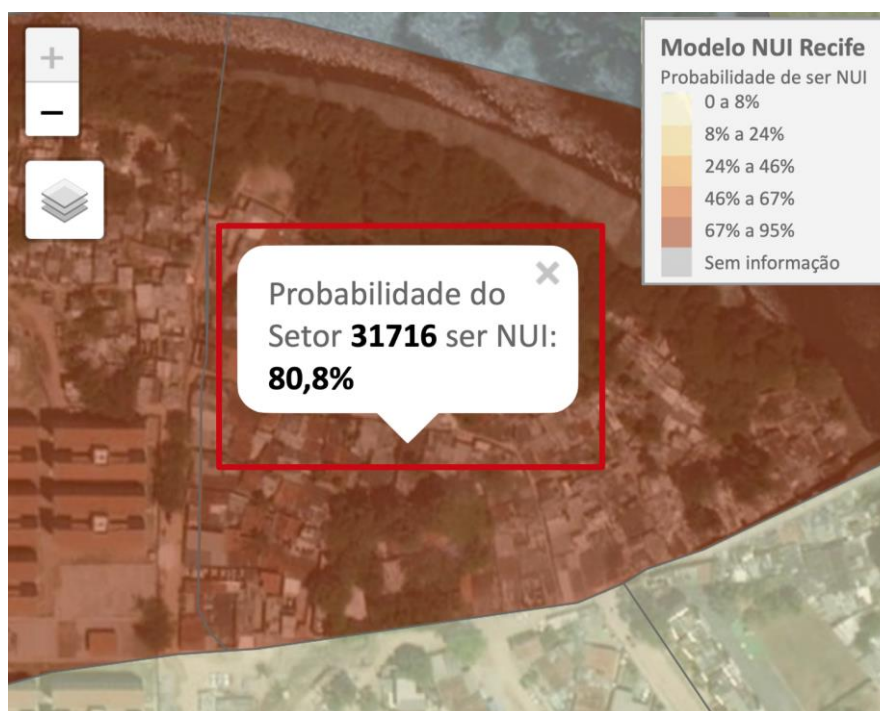
Figura 21 - Movendo e ampliando uma imagem orbital em mapa interativo.



Fonte: Elaboração própria, 2020.

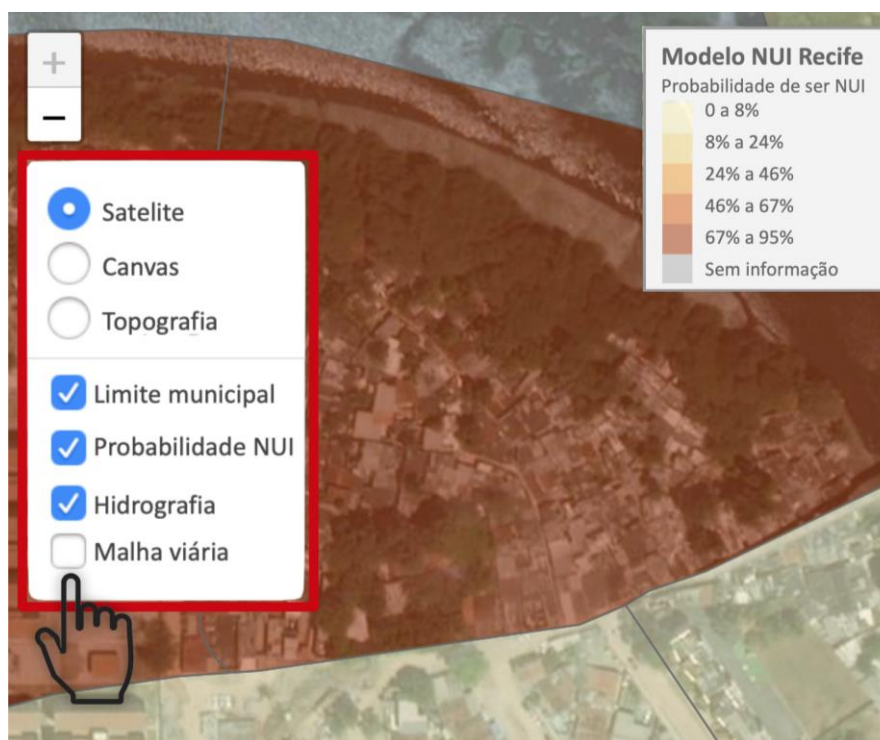
Outras funcionalidades dos mapas interativos incluem janelas que aparecem quando o usuário clica em diferentes feições ou recursos, uma espécie de rótulo dinâmico (Figura 22). Destaca-se ainda a possibilidade de “ligar e desligar” camadas de informações adicionais que podem esclarecer, por exemplo, dúvidas em relação às dimensões da precariedade que se encontram em situação mais crítica em uma área identificada como NUI (Figura 23). Assim, um usuário com dúvidas pode facilmente acessar as variáveis consideradas na modelagem e melhor compreender seus resultados. A facilidade de manipular distintas camadas de informação também permite ao usuário sobrepor bases como malha viária, sedes de municípios, equipamentos públicos ou outros pontos de referência que permitam uma melhor interpretação dos resultados.

Figura 22 - Informações das feições em mapas interativos.



Fonte: Elaboração própria, 2020.

Figura 23 - Ligar ou desligar camadas adicionais de informações em mapas interativos.



Fonte: Elaboração própria, 2020.

Os modelos têm muito a ganhar associando-se aos mapas interativos. A partir de qualquer navegador web, técnicos e tomadores de decisão podem examinar os resultados dos modelos, as variáveis neles considerados e dados adicionais sob múltiplas perspectivas. Esse processo facilita tanto a divulgação e acesso aos resultados, como também sua análise e interpretação.

Os mapas interativos dos resultados dos modelos NUI, além de incluir todas as vantagens já citadas, estão sendo desenvolvidos nas diversas etapas do desenvolvimento metodológico a fim de auxiliar na interpretação e validação dos resultados. Considerando que os modelos NUI não devem ser validados a partir de uma perspectiva binária e determinística, que busque estabelecer se determinada área está classificada de forma correta ou não, os mapas interativos oferecem uma poderosa ferramenta de análise para auxiliar atores locais, técnicos e especialistas na interpretação e análise de variáveis e estimativas de probabilidade, calibração de limiares para classificação de NUI, e apreensão dos diferentes níveis de incerteza atribuídos aos resultados da modelagem.

## 5. Considerações Finais

O presente relatório apresenta o processo, em andamento, de desenvolvimento da metodologia NUI, que objetiva a construção de modelos que auxiliem na identificação de núcleos urbanos informais.

A primeira etapa de desenvolvimento da metodologia envolveu a análise de índices já existentes, com o objetivo de investigar se esses poderiam atender aos objetivos da Pesquisa. Observou-se que, embora seja possível observar uma correlação entre os índices e a presença de NUI, os resultados não se revelaram suficientes para que se adotasse uma metodologia baseada apenas no cômputo desses índices.

Apesar disso, as análises dos índices mostraram-se promissoras em aspectos relevantes para o desenvolvimento da metodologia NUI. O primeiro deles diz respeito aos dados e variáveis utilizados nesses índices, cujo levantamento e sistematização foi importante e será considerado nas etapas seguintes da Pesquisa. Os índices também demonstraram ser promissores para a caracterização dos aspectos socioeconômicos e físico-territoriais dos NUI que serão identificados pela Pesquisa, revelando a heterogeneidade a eles inerente.

Considerando os resultados obtidos nesta etapa, optou-se por realizar uma análise aprofundada de metodologias que aplicam técnicas estatísticas para a classificação de assentamentos precários. As metodologias CEM e MAPPA foram adaptadas e aplicadas aos polos da Pesquisa. A partir do estudo e comparação das metodologias foi possível estabelecer aspectos desejáveis para a metodologia NUI e delinear decisões metodológicas e encaminhamentos para a Pesquisa.

Entre os aspectos considerados importantes para a construção da metodologia NUI estão a capacidade de ser generalizável para todo o território nacional, o que demanda a utilização de dados abertos para todo o Brasil e de técnicas aplicáveis a diferentes realidades, tais como a estimativa de modelos que possam combinar diferentes variáveis dependendo da área de estudo.

Considerou-se desejável ainda que a metodologia NUI não envolvesse apenas variáveis censitárias, mas que fosse também capaz de compreender outras dimensões territoriais por meio da integração de dados de fontes e naturezas diversas. Dada a necessidade de integração, optou-se pela adoção de células como unidades espaciais de análise. Mais do que isso, a malha celular que servirá de base para a construção dos modelos NUI deverá ser compatível com a grade estatística do IBGE, o que tende a ampliar ainda mais sua capacidade de integração com outros tipos de dados e resultados de outros estudos.

Outras decisões importantes foram tomadas em relação ao tipo de resultado esperado da metodologia NUI. Ao invés de classificações binárias, que apontam se determinada área é ou não um NUI, optou-se pelo desenvolvimento de uma metodologia cujos resultados sejam apresentados de forma contínua, na forma de superfícies de probabilidade da presença de NUI, o que explicita tanto os diferentes níveis de precariedade quanto as incertezas inerentes aos

resultados da classificação. Caso o analista prefira uma categorização binária, poderá fazê-lo a partir do estabelecimento de um limiar de probabilidade que considere adequado.

Para a geração das superfícies de probabilidade, duas técnicas estatísticas serão consideradas, a análise discriminante (em suas versões linear e quadrática) e a análise de regressão logística. Ambas serão amplamente testadas a partir dos dados obtidos em pesquisa de campo nos polos, devidamente integrados na grade celular que será estabelecida como base para o processo de modelagem. Caso uma das técnicas se mostre mais eficaz, será adotada como referência para a metodologia NUI.

Por fim, uma questão importante a ser considerada diz respeito à forma de uso e comunicação dos resultados da modelagem. As superfícies de probabilidade deverão servir com um plano de informação complementar às informações existentes sobre NUI (por exemplo, levantamentos municipais e perímetros dos aglomerados subnormais do IBGE). A combinação das superfícies de probabilidade a esses dados deverá dar origem a planos de informação adicionais, que indicam as áreas onde é possível ter maior/menor incerteza sobre a presença/ausência de NUI. Esses produtos deverão configurar uma importante referência para o planejamento de verificações em campo ou remota (por meio de análise de imagens orbitais).

Para facilitar esse processo, os resultados da modelagem e produtos associados deverão ser disponibilizados na forma de mapas interativos, de fácil acesso e manipulação. Esses mapas interativos deverão abranger ainda índices, indicadores e variáveis que permitam uma caracterização detalhada dos NUI identificados.

Espera-se, por meio dessas estratégias, que os resultados da metodologia NUI promovam a produção e disseminação de informações que subsidiem, de forma efetiva, a elaboração e aprimoramento de políticas e programas habitacionais.

## 6. Referências

- Agência Nacional de Energia Elétrica (ANEEL). Shapefile das linhas de alta tensão em território nacional, 2020.
- Agência Nacional do Petróleo, Gás Natural e Combustível (ANP). Shapefile das dutovias em território nacional, 2020.
- ANSELIN, L. (1995) *Local indicators of spatial association – LISA. Geographical analysis*, 27(2), p. 93-115. Disponível em: <[http://dces.wisc.edu/wp-content/uploads/sites/30/2013/08/W4\\_An\\_selin1995.pdf](http://dces.wisc.edu/wp-content/uploads/sites/30/2013/08/W4_An_selin1995.pdf)>. Acesso em: 4 jul. 2020.
- BRASIL. Ministério do Meio Ambiente. *Cadastro Nacional de Unidades de Conservação*, 2020.
- CÂMARA, G.; CARVALHO, M. S.; CRUZ, O. G.; CORREA, V. (2002) Análise espacial de áreas. In: FUKS, S.D.; CARVALHO, M.S.; CÂMARA, G. (Orgs.) Análise espacial de dados geográficos. Disponível em: <<http://www.dpi.inpe.br/gilberto/livro/analise/cap5-areas.pdf>>. Acesso em: 4 jul. 2020.
- CDHU – COMPANHIA DE DESENVOLVIMENTO HABITACIONAL E URBANO DO ESTADO DE SÃO PAULO; UFABC – UNIVERSIDADE FEDERAL DO ABC. FEITOSA, F. F. (org.) Desenvolvimento e aplicação de metodologia para identificação, caracterização e dimensionamento de assentamentos precários. São Bernardo do Campo: UFABC, 2019a. Relatório de Pesquisa. Produto V.
- CDHU – COMPANHIA DE DESENVOLVIMENTO HABITACIONAL E URBANO DO ESTADO DE SÃO PAULO; UFABC – UNIVERSIDADE FEDERAL DO ABC. FEITOSA, F. F. (org.) Sumário Executivo: MAPPA “Metodologia para Identificação e Caracterização de Assentamentos Precários em Regiões Metropolitanas Paulistas”. São Bernardo do Campo: UFABC, 2019b. E-book - ISBN: 978-65-5040-006-4
- CONSÓRCIO INTERMUNICIPAL DA REGIÃO DO GRANDE ABC (CIGABC); UNIVERSIDADE FEDERAL DO ABC (UFABC); DENALDI, R. et al. Diagnóstico Habitacional Regional do ABC. São Bernardo do Campo: UFABC, 2016. Disponível em: <<http://lepur.com.br/wp-content/uploads/2017/12/03-Diagnostico-Habitacional-Regional-do-Grande-ABC-Relat%C3%B3rio-Final.pdf>>. Acesso em: 23 jun. 2020.
- FERRACINI, K. R. Localizando a vulnerabilidade social na dimensão infraestrutura (IVS-I) versus os investimentos do Programa de Aceleração do Crescimento (PAC) nos municípios pernambucanos. In *Ciência&Trópico / Fundação Joaquim Nabuco*. -Vol.1, no.1. Recife: Editora Massangana, 2018.
- FJP – FUNDAÇÃO JOÃO PINHEIRO. Déficit Habitacional no Brasil 2015. Belo Horizonte: Fundação João Pinheiro – Centro de Estatística e Informações, 2018. Disponível em: <<http://fjp.mg.gov.br/index.php/docman/direi-2018/871-6-serie-estatistica-e-informacoes-deficit-habitacional-no-brasil-2015291118/file>>. Acesso em: 20 jul. 2020.
- FRIZZI, G. Índice de acesso à moradia adequada como instrumento para o planejamento territorial: elaboração metodológica e análise. 2019. 228 p. Dissertação (Mestrado) - Programa de Planejamento e Gestão do Território, Universidade Federal do ABC, São Bernardo do Campo, 2019.
- FRIZZI, G.; PINHO, C. M. D. Índice de vulnerabilidade socioecológica para avaliação das remoções na Cidade de São Paulo. In: XVII Encontro Nacional da ANPUR. Desenvolvimento, crise e resistência: Quais os caminhos do Planejamento Urbano e Regional? 2017, São Paulo. Anais do XVII ENANPUR, 2017.
- FUNDAÇÃO BRASILEIRA PARA O DESENVOLVIMENTO SUSTENTÁVEL (FBDS). Dados de hidrografia, 2015.
- GARETH, J.; WITTEN, D.; HASTIE, T. e TIBSHIRANI, R. 2014. *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated. 2014.
- HIRSCH, Robert P. (2016). *Introduction to Biostatistical Applications in Health Research with Microsoft Office Excel*. [S.l.]: Wiley. p. 150. 392 páginas.



INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). Aglomerados Subnormais 2019: Classificação preliminar e informações de saúde para o enfrentamento à COVID-19. Rio de Janeiro: IBGE, 2020. Disponível em: <<https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=2101717>>. Acesso em: 23 jun. 2020.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). Censo Demográfico: Brasil, 2010. Rio de Janeiro: IBGE, 2010

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). Grade Estatística. Rio de Janeiro: IBGE, 2016.

IPEA. Atlas da vulnerabilidade social nos municípios brasileiros. Ed. COSTA, M. A. MARGUTTI, B. O. Brasília: IPEA, 2015

IPEA. Dados IVS. Disponível: [http://ivs.ipea.gov.br/images/publicacoes/ivs/publicacao\\_atlas\\_ivs.pdf](http://ivs.ipea.gov.br/images/publicacoes/ivs/publicacao_atlas_ivs.pdf). Acesso em: 10 ago. 2020.

IPEA. Desenvolvimento humano nas macrorregiões brasileiras: 2016. – Brasília: PNUD: IPEA : FJP, 2016. Disponível em:<[http://www.atlasbrasil.org.br/2013/data/rawData/19FEV\\_IDHM\\_WEB.pdf](http://www.atlasbrasil.org.br/2013/data/rawData/19FEV_IDHM_WEB.pdf)>. Acesso em: 4 set. 2020.

IPEA. Métodos e conceitos para o cálculo do índice de vulnerabilidade social com base nas PNADS e desagregações. Brasília, 2018.

IPEA. Territórios em números: insumos para políticas públicas a partir da análise do IDHM e do IVS de municípios e Unidades da Federação brasileira, livro 1 / organizadores: Bárbara Oliveira Marguti, Marco Aurélio Costa, Carlos Vinícius da Silva Pinto. – Brasília: IPEA: INCT, 2017.

LUZARDO, A. J. R.; CASTAÑEDA FILHO, R. M.; RUBIM, I. B. Análise espacial exploratória com o emprego do índice de Moran. *GEOgraphia*, Niterói – RJ, v. 19, n. 40, 2017.

MARQUES, E. (ed.). Assentamentos precários no Brasil urbano. São Paulo/Brasília: Centro de Estudos da Metrópole / CEBRAP ; Secretaria Nacional de Habitação / Ministério das Cidades, 2007.

NASA, National Aeronautics and Space Administration. U.S. Releases Enhanced Shuttle Land Elevation Data. Disponível em <<https://www2.jpl.nasa.gov/srtm/>> Acesso em set. de 2020.

*OpenStreetMap*®, contribuidores do *OpenStreetMap*, 2020.

PNUD, IPEA, FJP. Índice de Desenvolvimento Humano Municipal Brasileiro. – Brasília: PNUD, Ipea, FJP, 2013. Disponível em: <[https://www.ipea.gov.br/portal/images/stories/PDFs/130729\\_AtlasPNUD\\_2013.pdf](https://www.ipea.gov.br/portal/images/stories/PDFs/130729_AtlasPNUD_2013.pdf)>. Acesso em: 4 set. 2020.

SÃO PAULO (Estado). Empresa Paulista de Planejamento Metropolitano SA (EMPLASA). Sistema de Informações Metropolitanas do Estado de São Paulo (SIM), s.d. Disponível em: <http://www.sim.emplasa.sp.gov.br/mapa>. Acesso em: 15 nov. 2018.

Secretaria do Meio Ambiente (SMA) / Coordenadoria de Planejamento Ambiental (CPLA). UNIDADES HOMOGÊNEAS DE USO E OCUPAÇÃO DO SOLO URBANO (UHCT) - Ficha Técnica. São Paulo: SMA, 2010.

SOUZA, M.C.P.; ROSSI, M.T.B.; RUDGE, M.S. Mapeamento colaborativo de assentamentos precários em regiões metropolitanas paulistas. In: III Seminário sobre Urbanização de Favelas, Salvador, BA, Brasil, 21 a 23 de novembro de 2018. Anais..., Salvador: Universidade Católica de Salvador, 2018.

TASHNER, S. P. O desafio da mensuração. In: BRASIL; MINISTÉRIO DAS CIDADES. Política Habitacional e Integração Urbana de Assentamentos Precários: Parâmetros conceituais, técnicos e metodológicos. Brasília: Ministério das Cidades, 2008. p. 93–110.

UNITED NATIONS – UN *Committee on Economic, Social and Cultural Rights* (CESCR). General Comment No. 4: The Right to Adequate Housing (Art. 11 (1) of the Covenant, 1991.

WHO – WORLD HEALTH ORGANIZATION. *Health principles of housing*. Report of WHO Consultation. Geneva: 1989. Disponível em:  
<[https://apps.who.int/iris/bitstream/handle/10665/39847/9241561270\\_eng.pdf?sequence=1&isAllowed=y](https://apps.who.int/iris/bitstream/handle/10665/39847/9241561270_eng.pdf?sequence=1&isAllowed=y)> Acesso em: 08 jul. 2020.