



Mapeamento baseado na regressão em Random Forest na região RMVPLN-SP

Orientador: Dr. Antonio Miguel Vieira Monteiro
Orientadora: Dra. Tathiane Mayumi Anazawa

Propósito

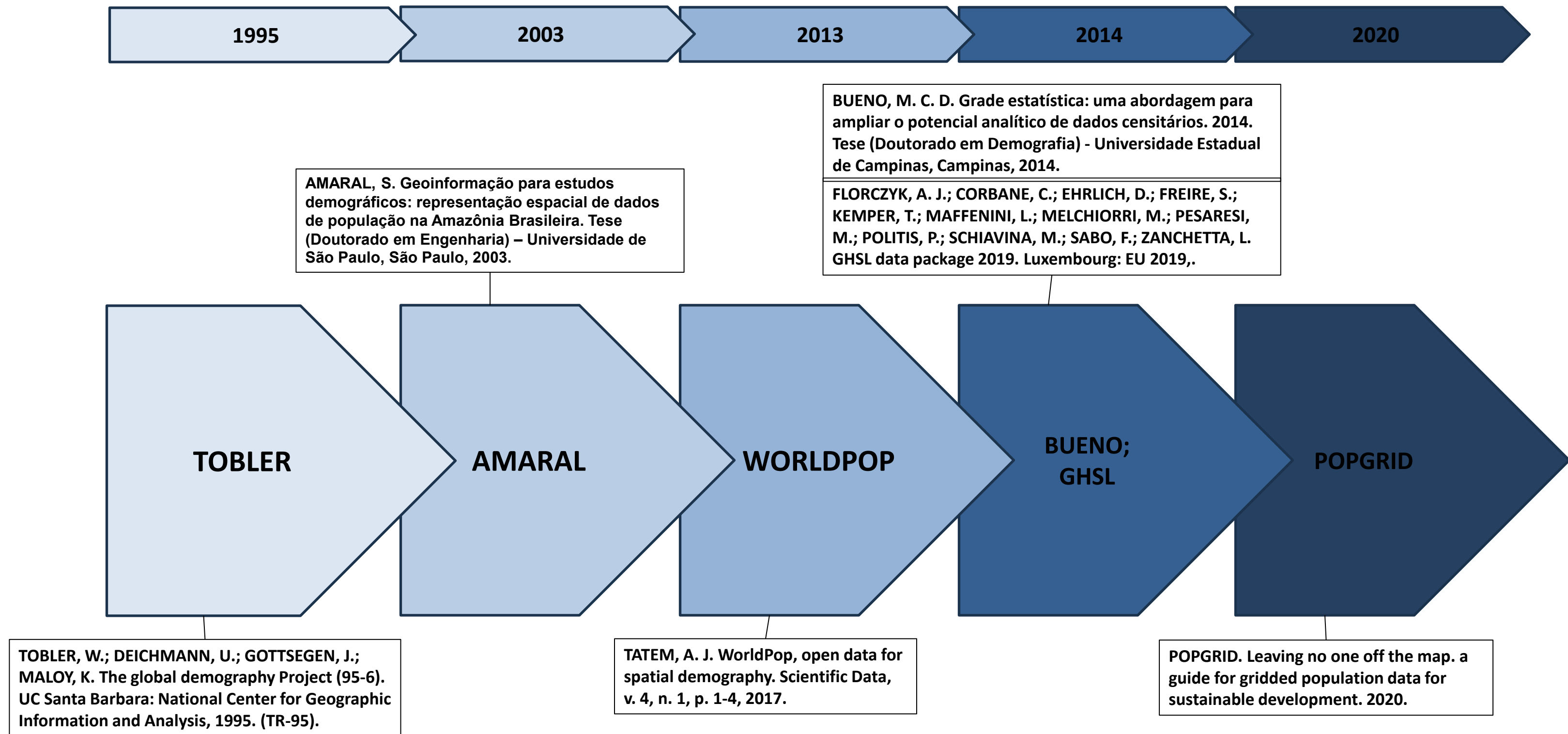
- Breve introdução: representação da população; integração; Grades Populacionais; e principais iniciativas no campo.
- Construção Grades Populacionais: *Buttom Up* e *Top Down*? ; métodos de desagregação.
- Metodologia: dimensões variáveis sintomáticas; pressupostos teóricos; objetivos; área de estudo.
- Resultados: principais resultados obtidos .
- Reflexões posteriores?

Introdução

Representação da distribuição populacional: evolução e métodos;

- Distribuição da população -> Padrões diversos;
- **Distribuição de espécies (não-humanas);**
- Evolução das grades populacionais globais(WaldoTobler, 1995) -> *Consortium for International Earth Science Information Network* (CIESIN) e *Joint Research Centre* (JRC);
- Sensoriamento Remoto e Dados Censitários -> Grade;
- Iniciativas de grades populacionais para o contexto local/regional.

Evolução das grades populacionais



Planejamento territorial e distribuição da população

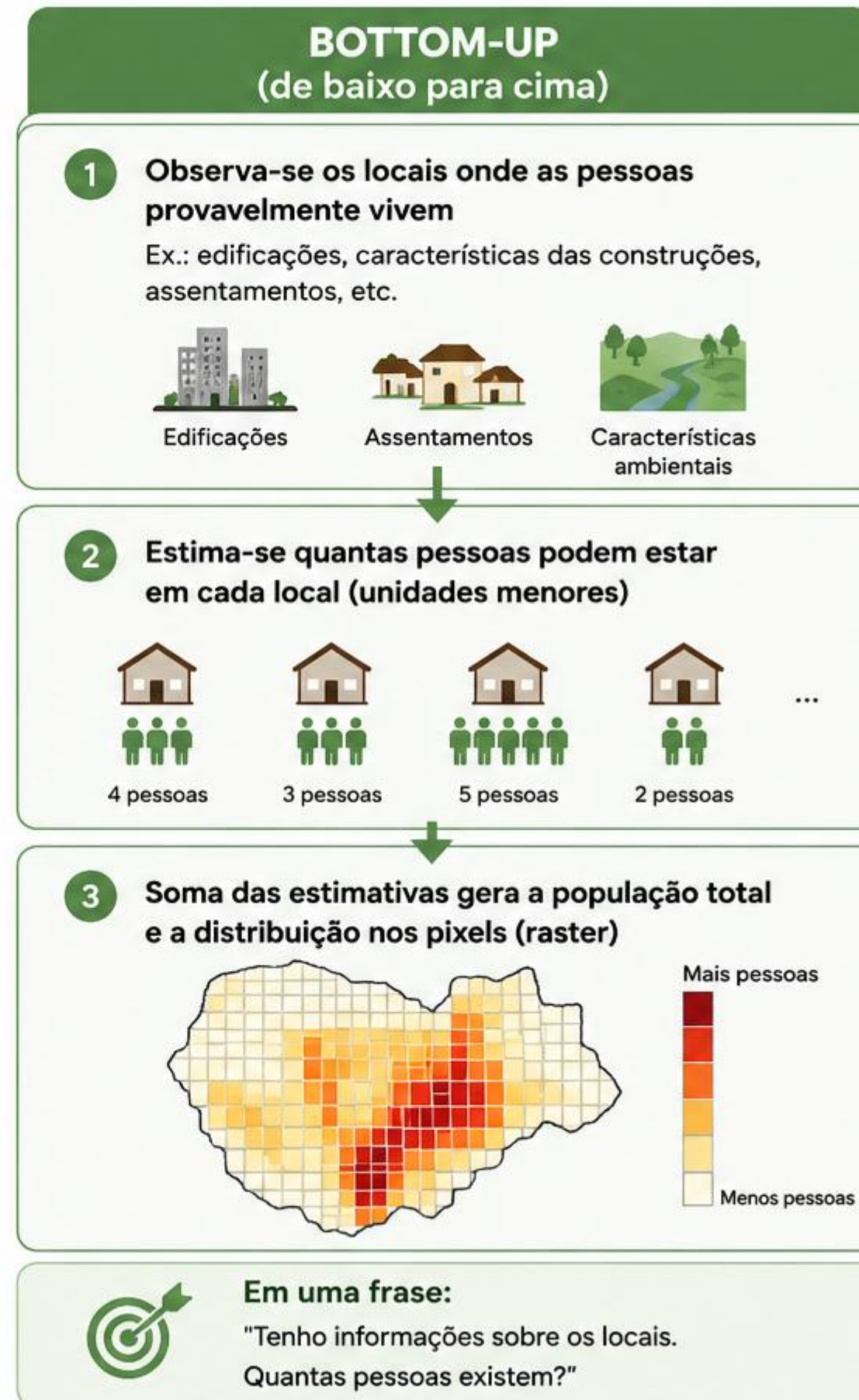
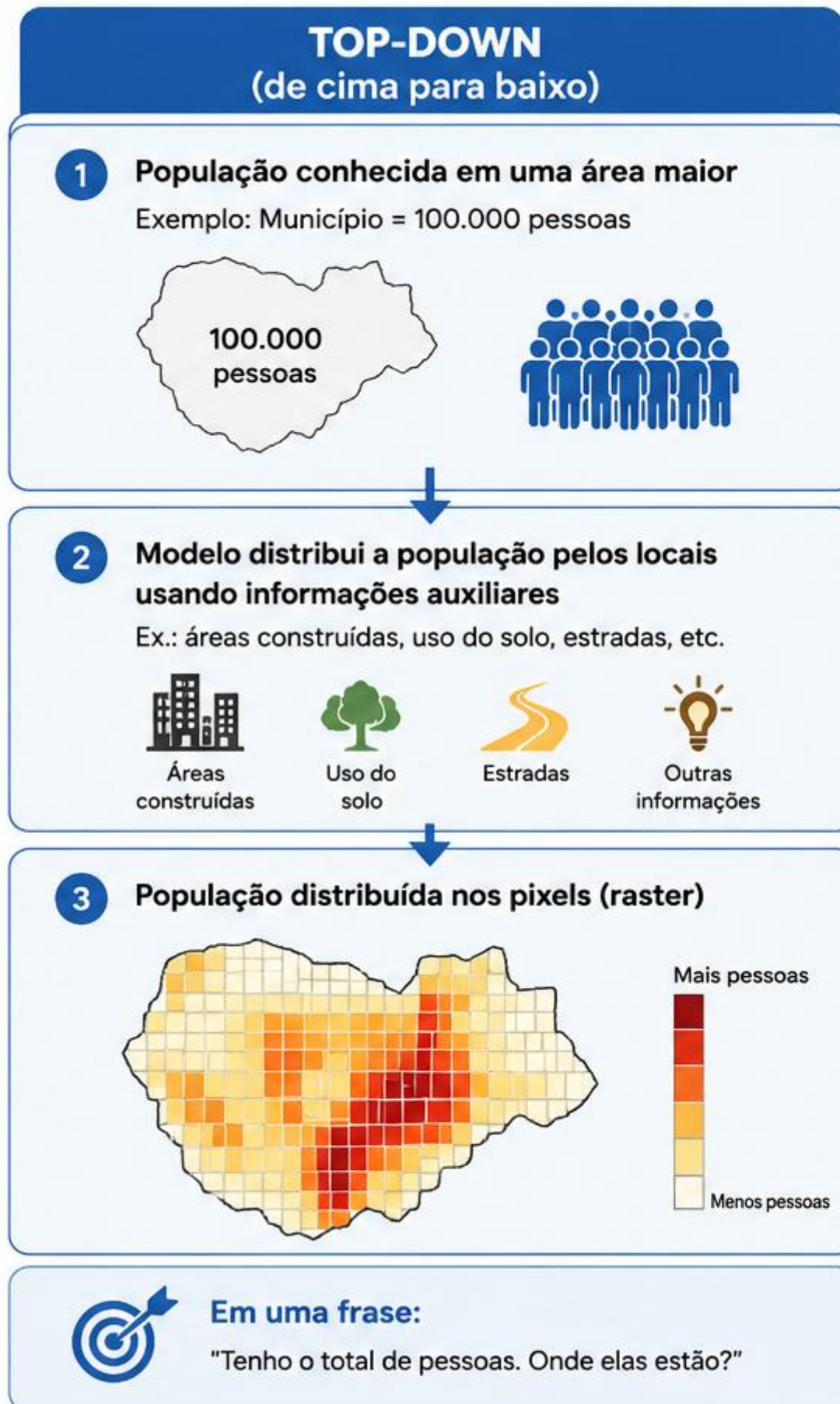
- Necessidade de obtenção de dados populacionais para pequenas áreas -> Volume e distribuição;
- Incorporar nos modelos aqueles elementos de localização de grupos ou indivíduos que não estão presentes nas grades globais: investigando os dados locais;
- Avaliar métodos de (re)distribuição mais atuais utilizados por essas grades globais: WorldPop, GHSL, GPWv4 e outros.

Iniciativas Globais

Produto	Nome	Fonte	Método	Limitações	Definição da contagem	Variáveis	Resolução espacial	Anos disponíveis
GPWv4	Population Density of Gridded Population of the World, version 4 Revision 11.	Center for International Earth Science Information Network (CIESIN), Columbia University	Ponderação por área	Pressupõe que a distribuição é uniforme	Pessoas/km ²	DEM, uso e cobertura, rede de transporte, dados de sensores remotos, comunicação móvel, outros.	1 km	2000, 2005, 2010, 2015, 2020
WorldPop	WorldPop Population Counts	Center for Spatial Sciences at the University of California	Desagregação baseado em machine learning (RF)	Células da grade que não são povoadas podem receber algum valor	Pessoas em cada célula	Elevação, declividade, uso e cobertura, infraestrutura, dados satelitais e dados de comunicação.	100 m	2000–2020
GHSL	GHS-POP of Global Human Settlement Layer	European Commission	Regressão linear	Ignora limites administrativo e não mensura populações em pequenas áreas	Pessoas em cada célula	Máscara de áreas construídas	100 m	1975, 1990, 2000, 2015

Fonte: Produção do autor

Bottom Up e Top Down?



Diferenças principais

Top-down

- Começa com o total conhecido
- Distribui para os lugares prováveis
- Útil quando temos dados populacionais confiáveis em áreas administrativas

Bottom-up

- Começa com informações locais
- Estima população localmente e soma
- Útil quando temos bons dados locais (ex.: edificações)

Em resumo



Não significa simplesmente que um método é melhor que o outro.

Eles respondem a situações diferentes e dependem dos dados disponíveis e da escala de análise.

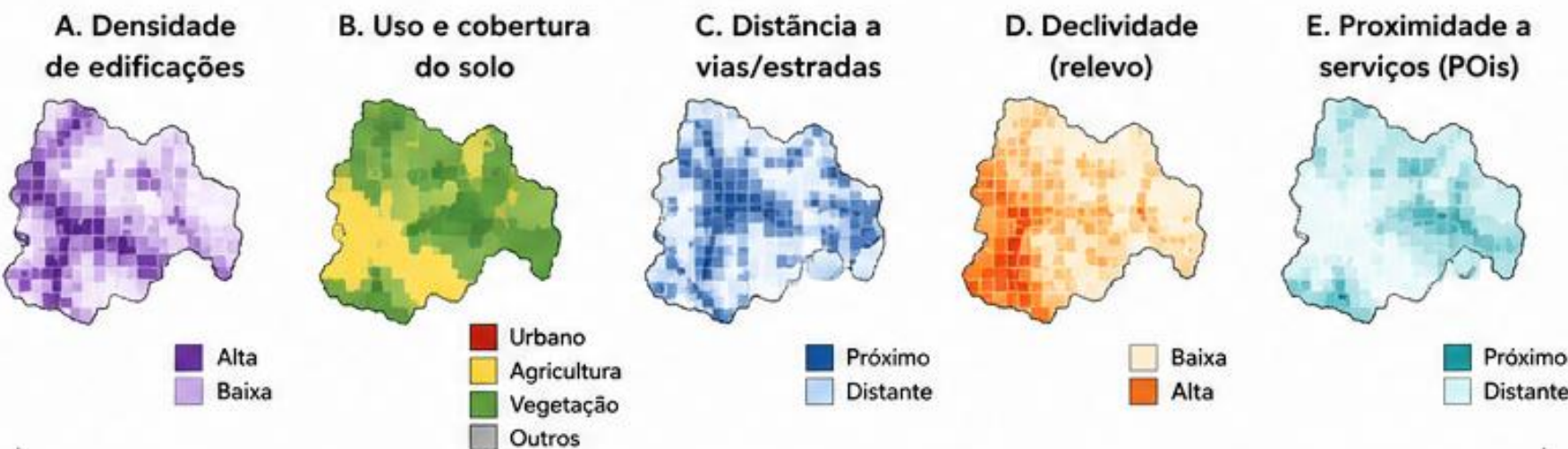
Método de desagregação

- As diferentes técnicas de transferência de informação de um ponto A, de origem para um ponto B, com base em dados auxiliares ou não;
- **Método *dasimétrico***: remonta de 1910, Rússia (Semenov-Tian-Shansky), em 1930, o primeiro artigo publicado sobre o tema (Wright, 1936);
- O método dasimétrico com múltiplas camadas integra múltiplas camadas de dados auxiliares que possuam associação com a presença de humanos (informações complementares).

Método de desagregação

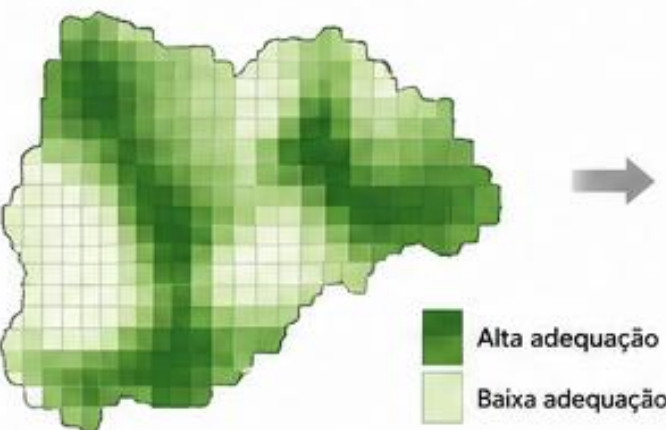
DASIMÉTRICO COM MÚLTIPLAS VARIÁVEIS AUXILIARES

1. Variáveis auxiliares utilizadas (exemplos)



Combinação ponderada das variáveis → superfície de adequação (probabilidade de presença de população)

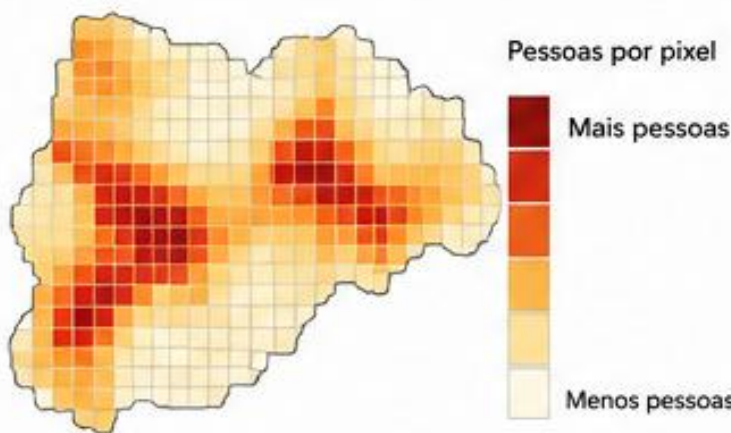
2. Superfície de adequação contínua



População total da área administrativa (ex.: 100.000 pessoas)



3. Redistribuição da população

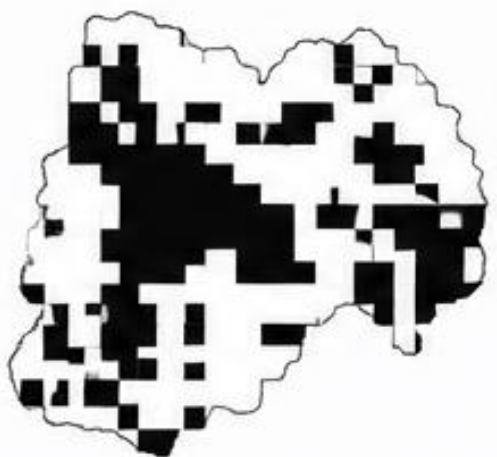


Características

Captura variações graduais. Identifica melhor onde as pessoas estão mais e menos concentradas.

DASIMÉTRICO BINÁRIA

1. Variável auxiliar binária (ex.: máscara de área construída)



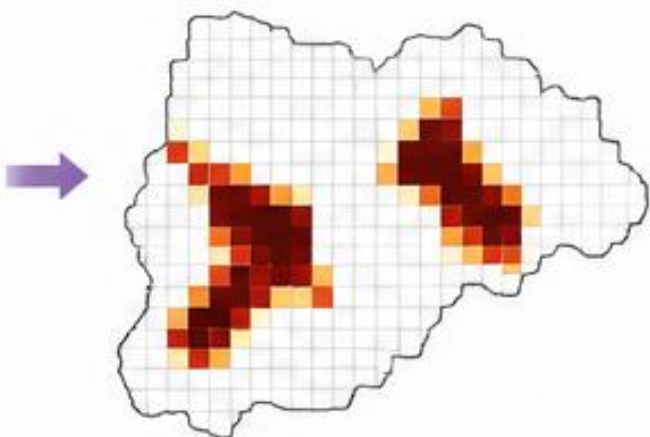
- Área construída (habitação possível)
- Área não construída (habitação não possível)



A variável binária apenas indica onde pode (1) ou não pode (0) existir população.

2. Redistribuição da população apenas nas áreas "1" (construídas)

População total da área administrativa (ex.: 100.000 pessoas)



Características

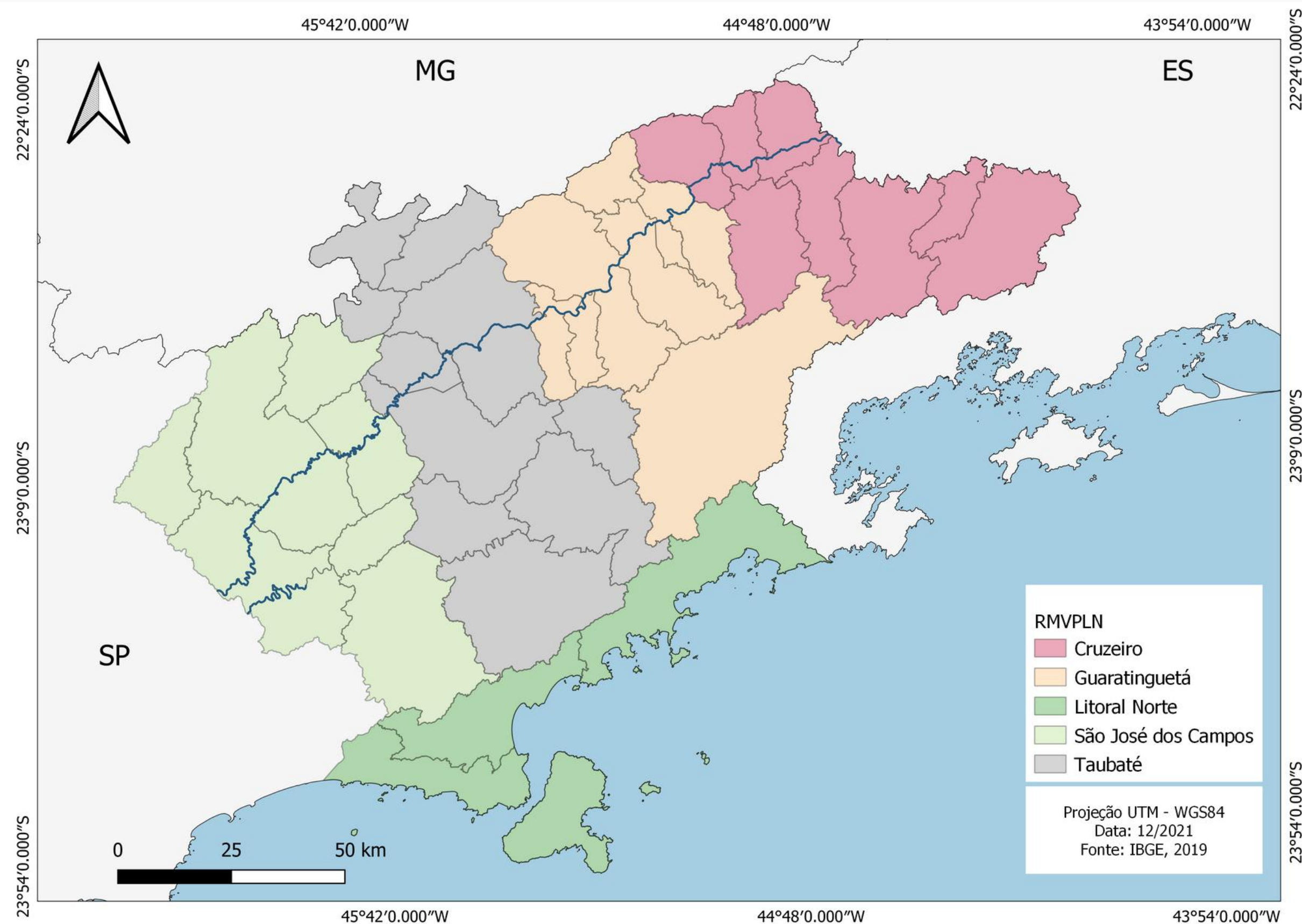
Simple e direto. Assume distribuição uniforme dentro das áreas construídas e zero fora delas.

Questões

- Como as variáveis sintomáticas locais refletem a distribuição espacial da população em comparação com as variáveis sintomáticas globais ?
- Quais variáveis possuem maior influência no modelo de distribuição espacial da população utilizando o algoritmo *Random Forest*?

Área de estudo

Mapa da Região Metropolitana Vale do Paraíba e Litoral Norte.



Sub-regiões (Emplasa, 2012):

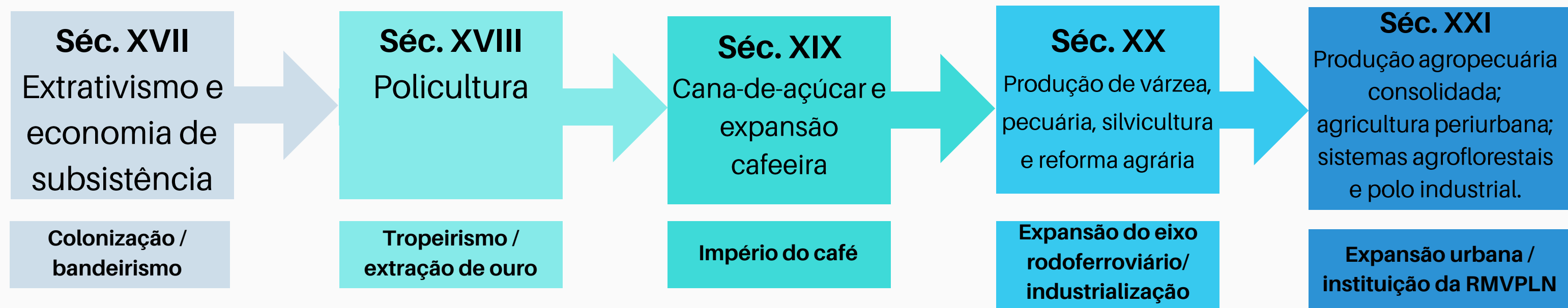
- Cruzeiro;
 - Guaratinguetá;
 - Litoral Norte;
 - São José dos Campos;
 - Taubaté.
-
- 2.583.607 habitantes estimados (IBGE, 2021)
 - 16.178 km² de extensão territorial (Emplasa, 2012);
 - Diversidade cultural e paisagística;
 - Liga as metrópoles de São Paulo e Rio de Janeiro

Objetivos

- i) Avaliação e seleção das variáveis sintomáticas de presença de população na RMVPLN (elemento principal) estabelecendo procedimentos para inclusão de variáveis associada ao histórico de ocupação regional;
- ii) Harmonização de bases de uso e cobertura, com foco nas classes indicadoras de presença humana em para a RMVPLN (i);
- iii) Introduzir o modelo dasimétrico, no contexto da regressão com *Random Forest*, para o cálculo dos pesos para a redistribuição;
- iv) Produzir grades de densidade populacional locais, com células de 100x100 m, geradas pelo modelo desenvolvido em (iii);
- v) Avaliar as grades de densidade populacional local, geradas frente as grades globais do WorldPop e GHSL (produto GHS-POP).

Fundamentação teórica

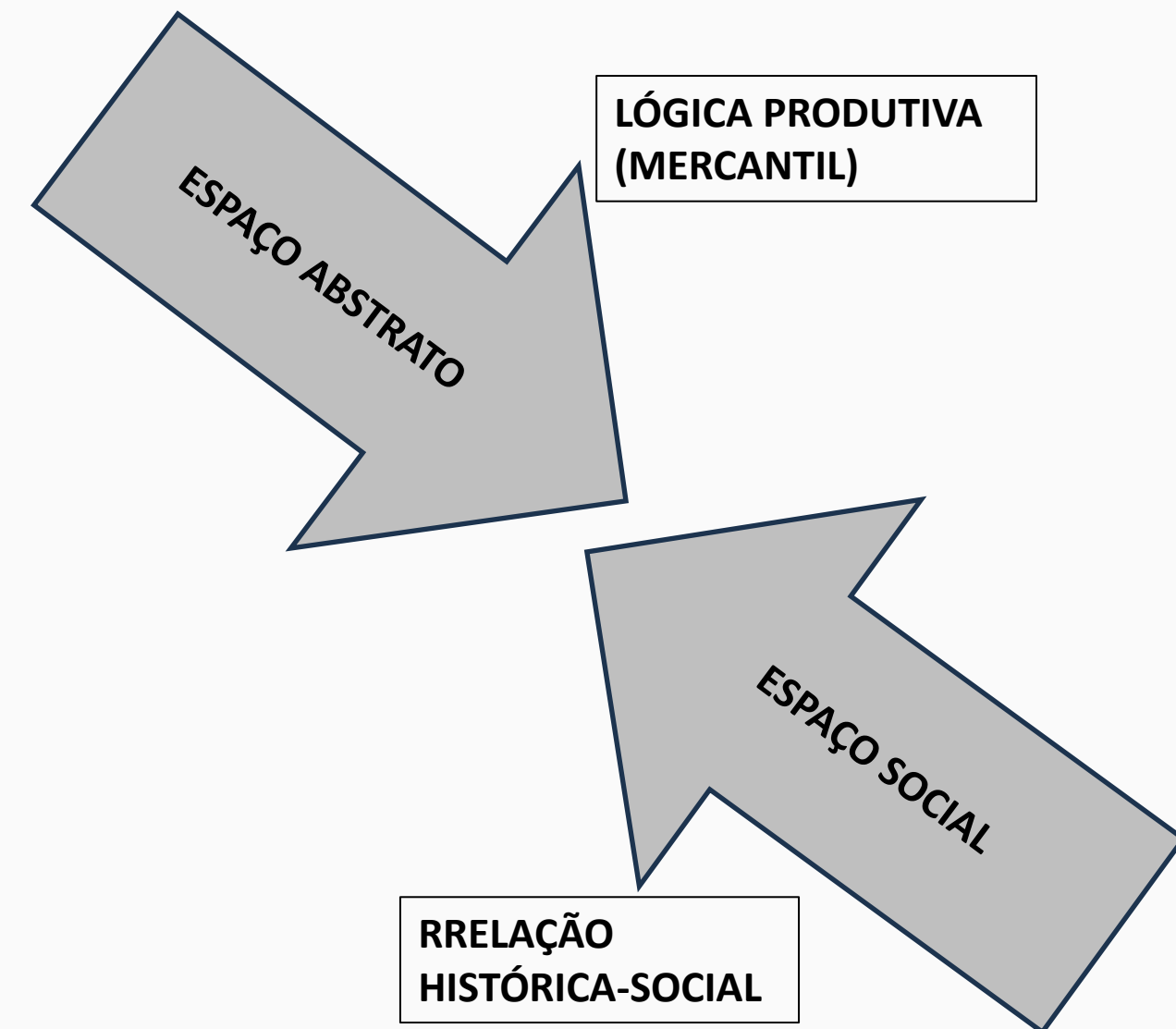
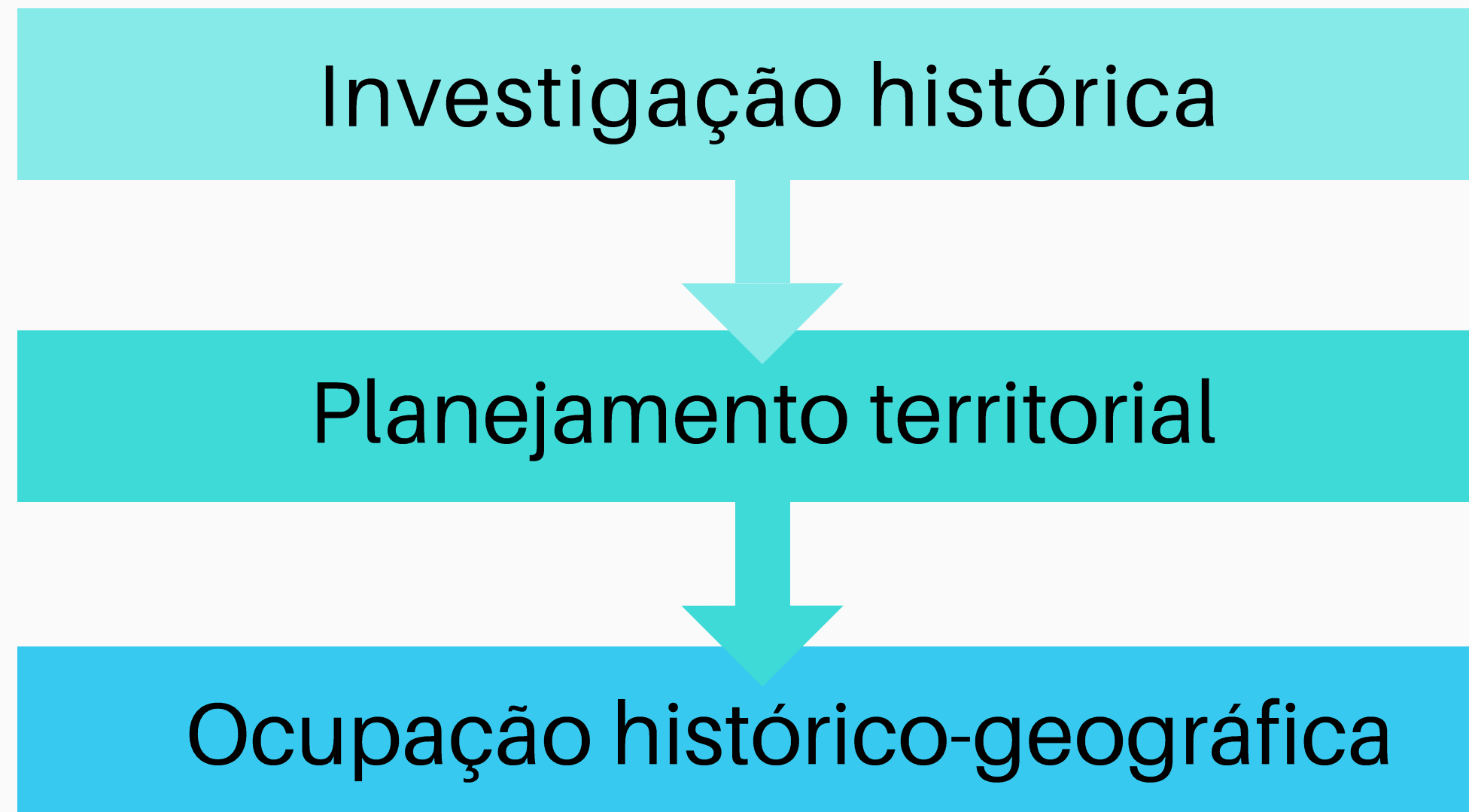
Linha das atividades produtivas associadas à formação territorial da RMVPLN.



Fonte: Adaptado de MARTINS (2020).

Fundamentação teórica

Pressuposto: Busca das variáveis locais/regionais para a RMVPLN?



LEFEBVRE, Henri. Prefácio: a produção do espaço. Estudos avançados, v. 27, n. 79, p. 123-132, 2013.

Dimensões das variáveis sintomáticas



Uso e cobertura

Bases de uso e cobertura
CCI e MAPBIOMAS



Redes

Tipologias de vias, CNEFE
Agro e CNEFE Faces



Elevação e Declividade

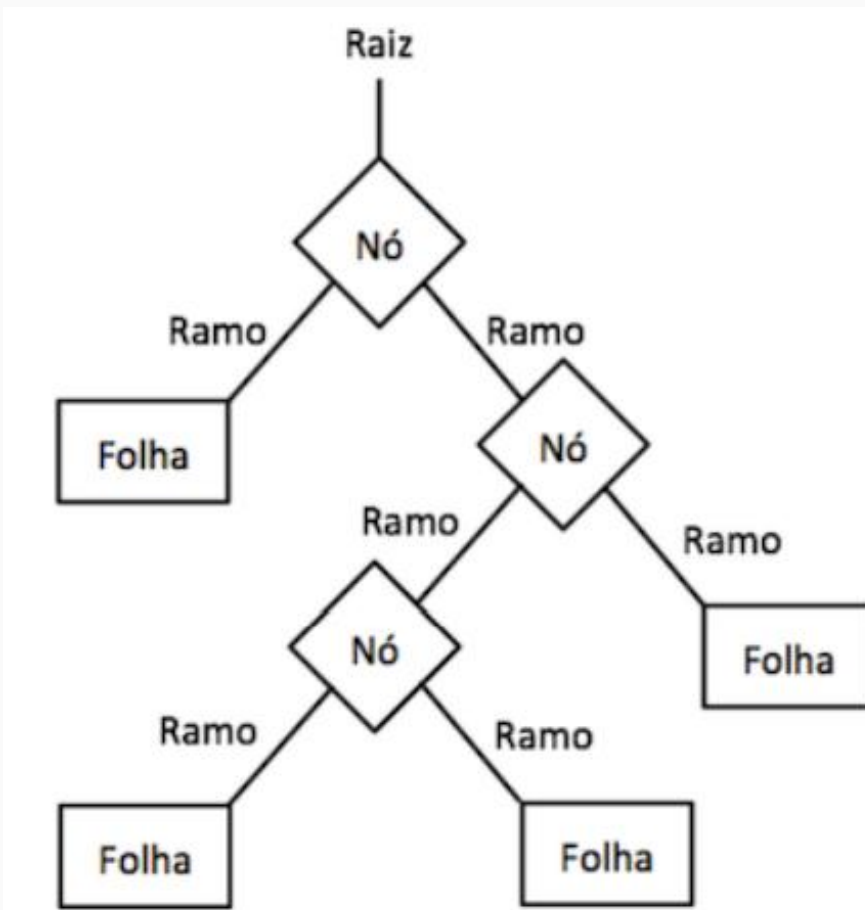
Declividade e elevação



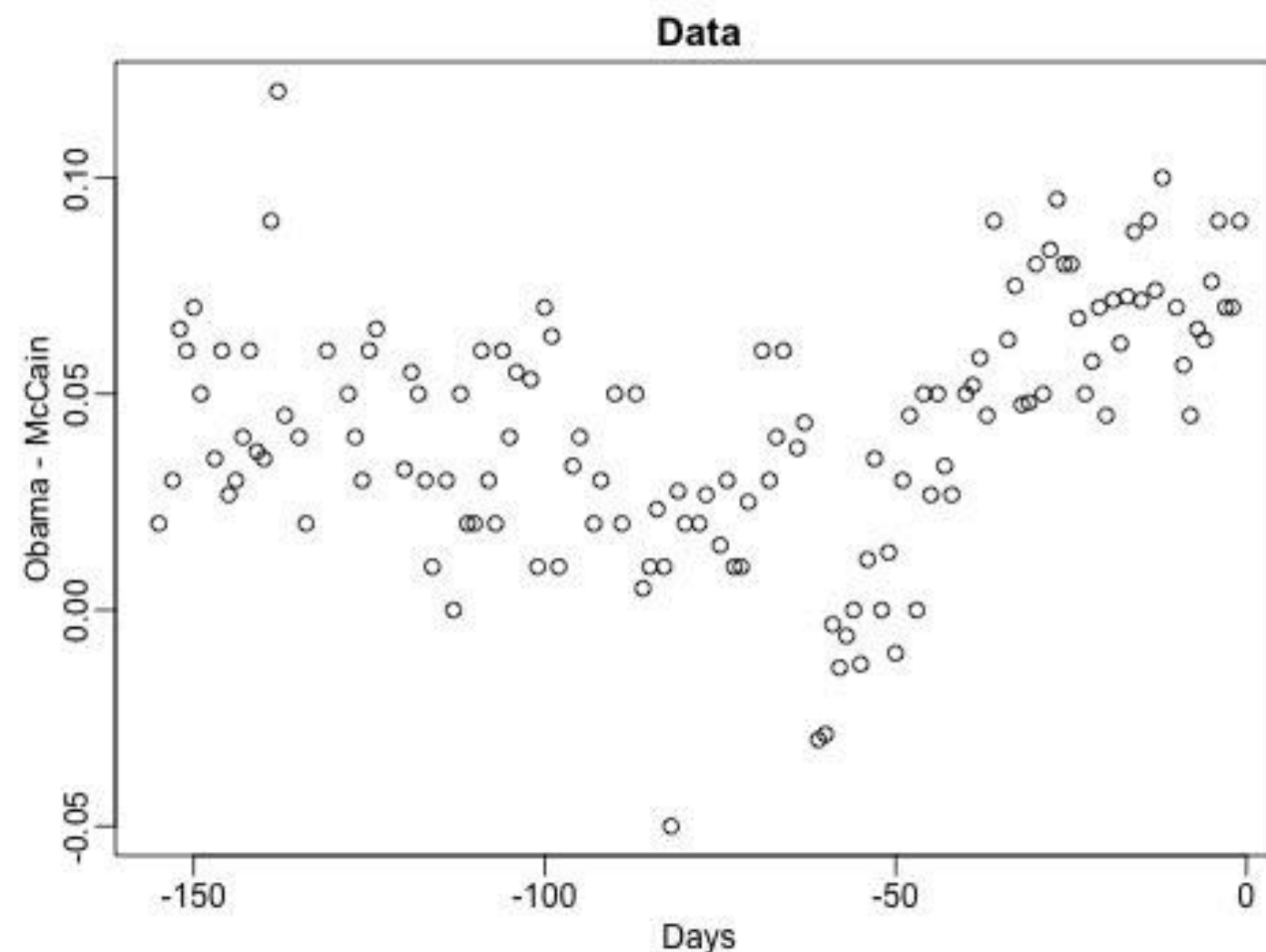
Infraestrutura

Black Marble, GHSL_LAND
e GHSL_BUILT_S

Random Forest



Fonte: MEDEIROS, 2014.

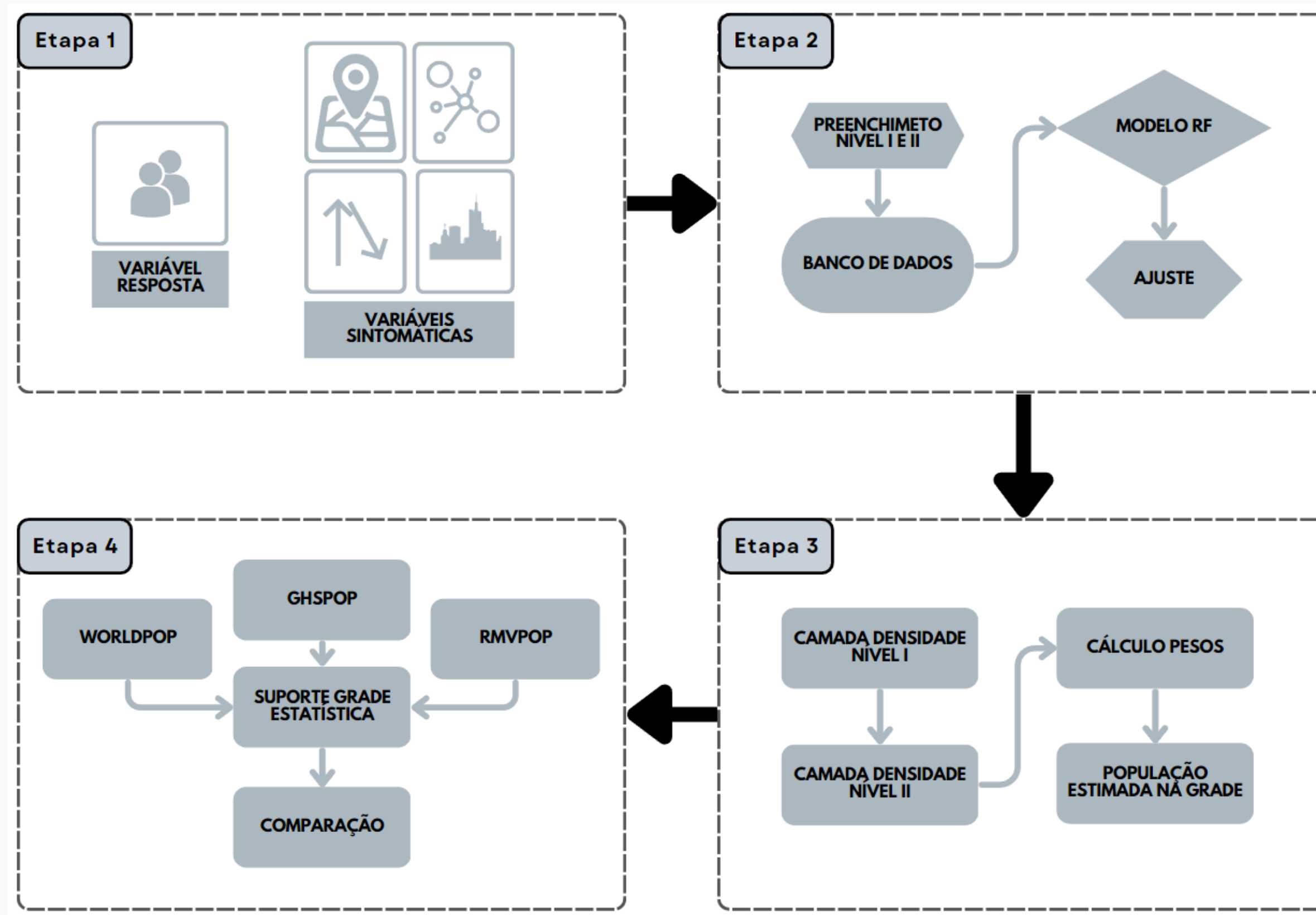


Possibilidades e limitações desta abordagem:

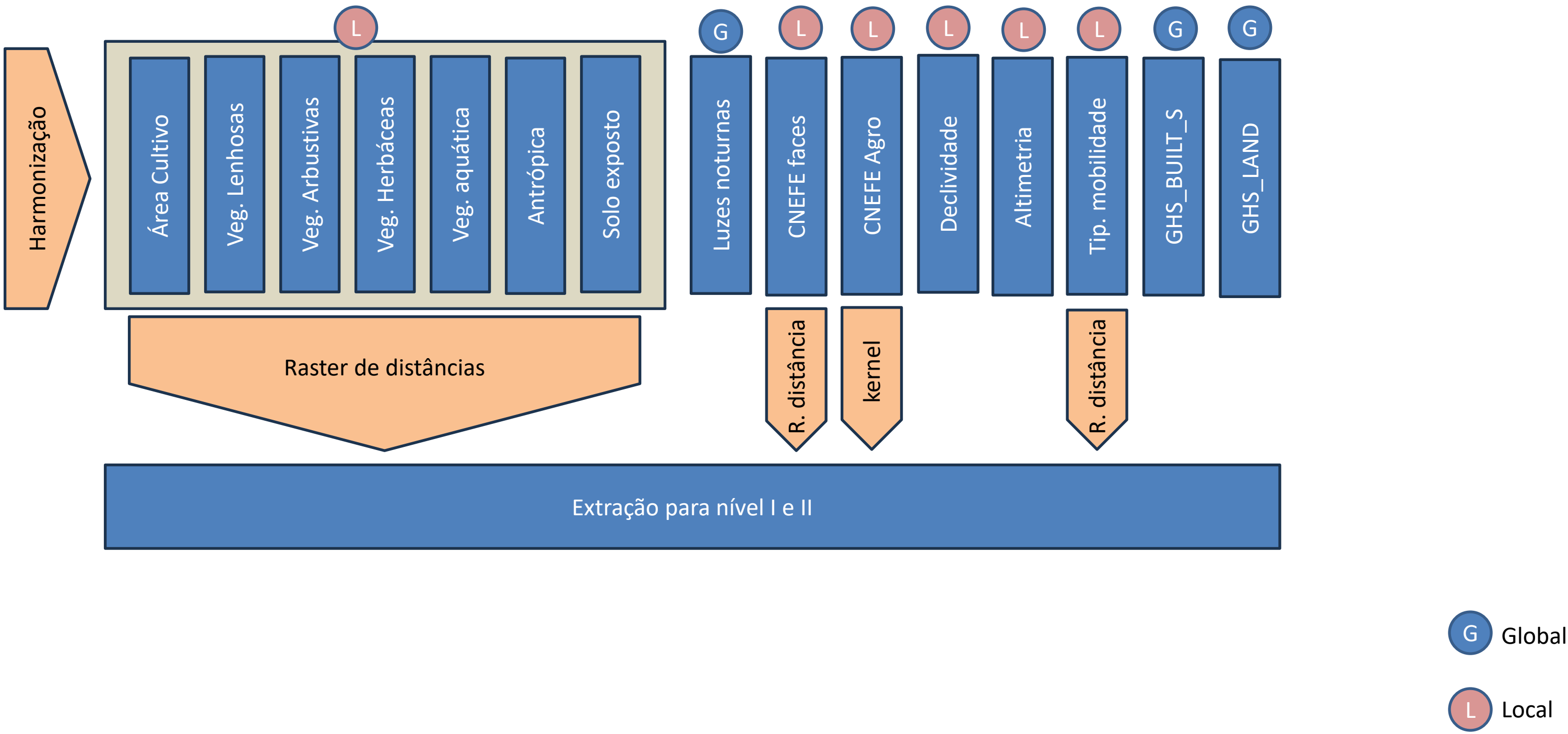
- Consegue lidar com elevadas quantidade de dados;
- Devido seu caráter aleatório, diminui a possibilidade de overfitting;
- Consegue lidar com variáveis que possuam informações faltantes.
- Não realiza extrapolação dos dados, ou seja, só prevê a média das variáveis previamente observadas;
- Penaliza *outliers*.

“A COMBINAÇÃO DOS PARÂMETROS DE VÁRIAS ÁRVORES, A PARTIR DE SUA MÉDIA, REDUZIRÁ A VARIABILIDADE FINAL” (BREIMAN, 2001).

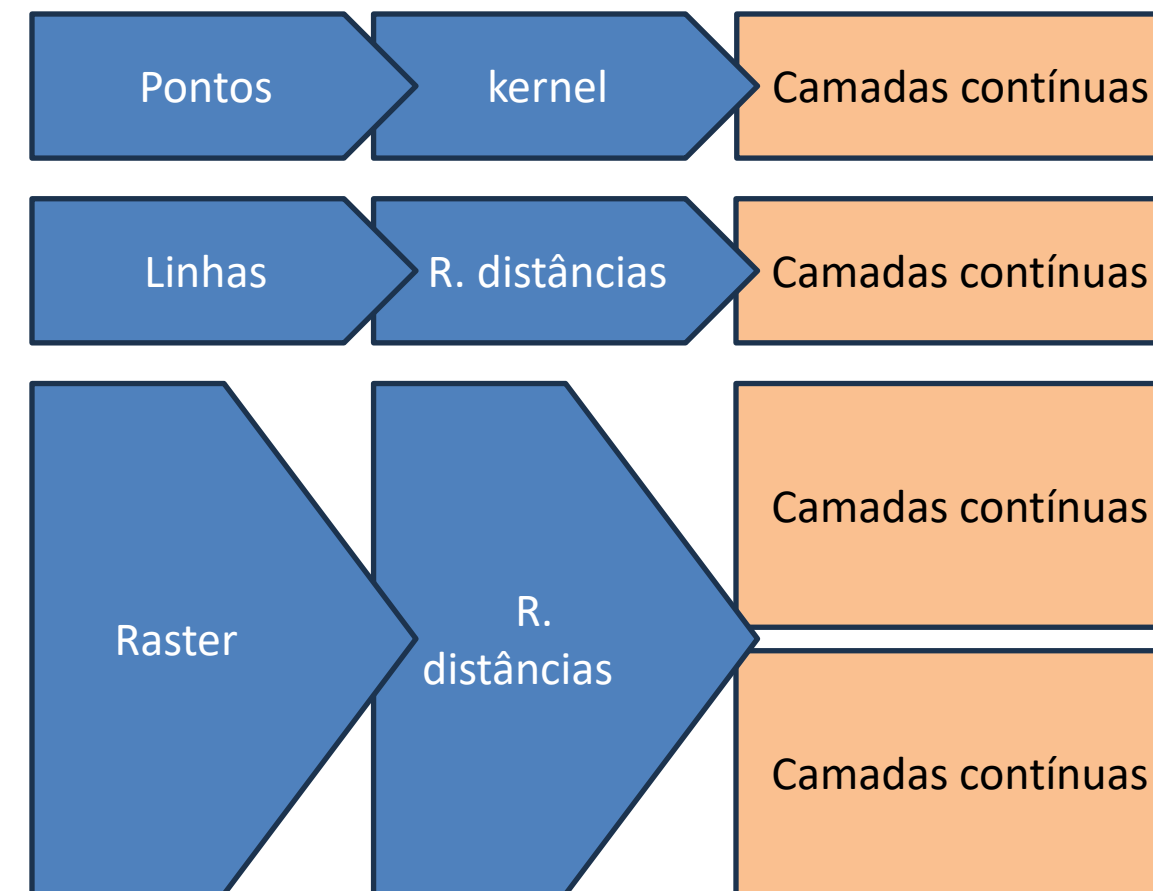
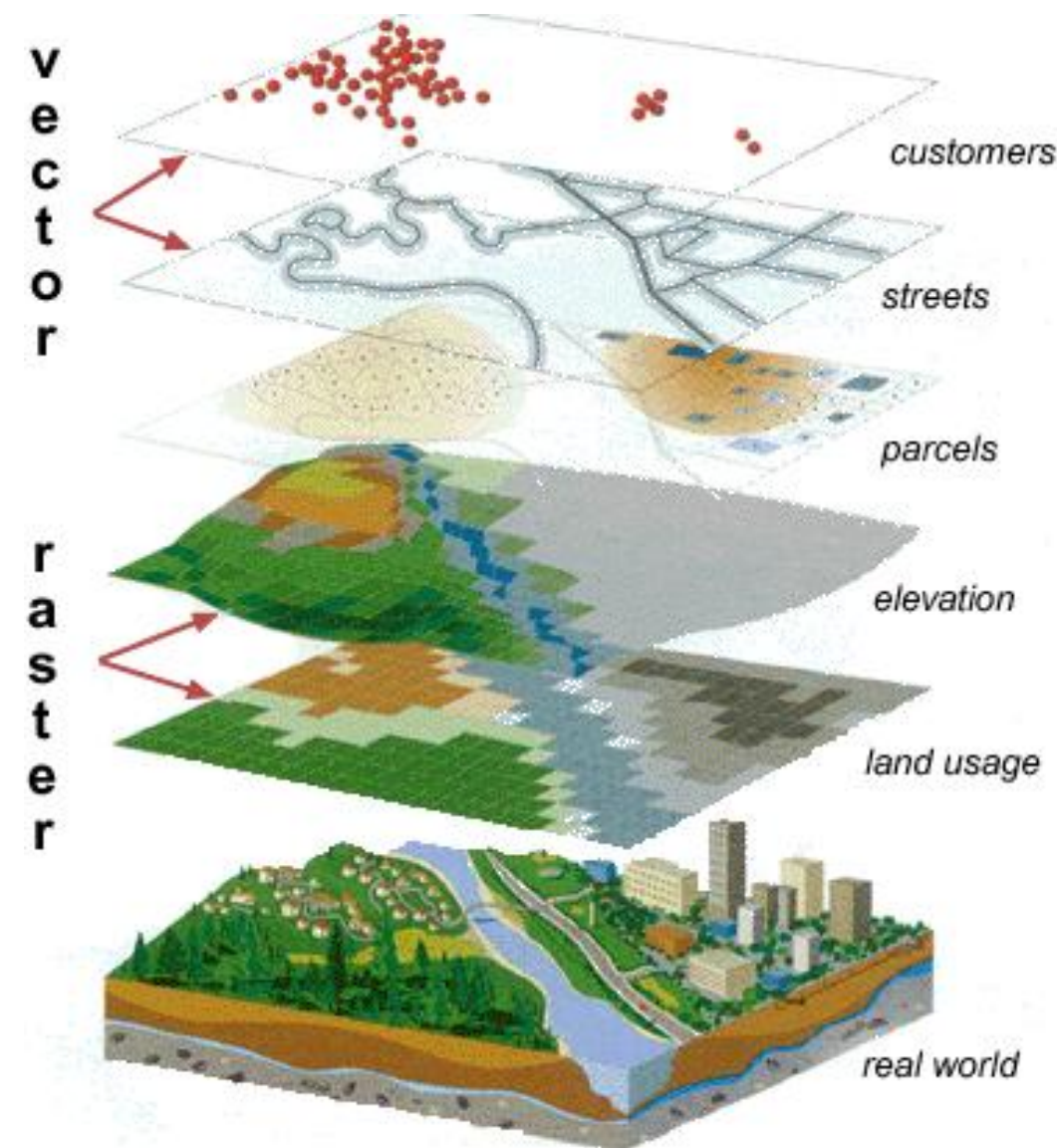
Metodologia geral



Etapa 1: Seleção e construção de camadas contínuas



Etapa 1: Seleção e construção de camadas contínuas



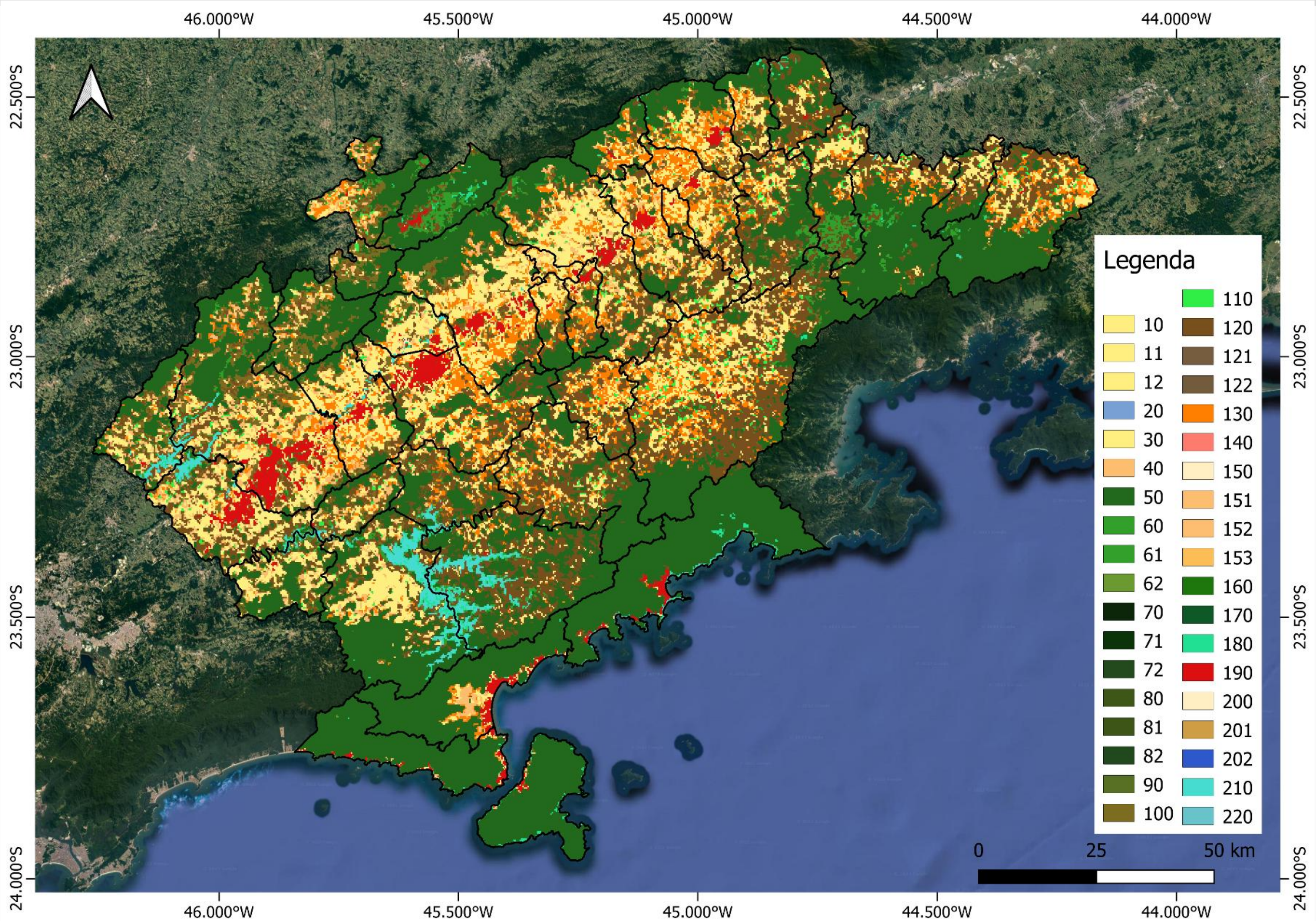
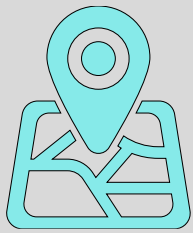
DADOS INPUT

Nomenclatura	Dados utilizados	Descrição	Fonte	Ano	Formato	Resolução Espacial	Contexto
BLACKM_100M	VNP46A4/VJ146A4	Black Marble: um produto melhorado do sensor DNB de composição anual livre de nuvens	NASA	2015	raster - hierarchical Data Format	~ 500 metros	global
CNEFE_FACES_100M	CNEFE Faces	Faces de logradouros	IBGE	2010	linhas/vértices	—	local
CNEFE_AGR_100M	CNEFE Agro	Pontos georreferenciados de domicílios e estabelecimentos rurais e urbanos	IBGE	2017	pontos de localização	—	local
TOPO_DECL_100M	Declividade	Ângulo de inclinação do terreno em relação ao plano horizontal	TOPODATA	2010	raster contínuo	30 metros	local
TOPO_ALT_100M	Altimetria	Elevação em metros do terreno	TOPODATA	2010	raster contínuo	30 metros	local
MAPB_LULC_011_100M	MAPBIOMAS	Áreas de cultivo	Projeto MAPBIOMAS	2015	raster/categorico	30 metros	local
MAPB_LULC_040_100M	MAPBIOMAS	Árvores lenhosas	Projeto MAPBIOMAS	2015	raster/categorico	30 metros	local
MAPB_LULC_150_100M	MAPBIOMAS	Áreas de vegetação esparsa	Projeto MAPBIOMAS	2015	raster/categorico	30 metros	local
MAPB_LULC_160_100M	MAPBIOMAS	Áreas de vegetação aquática	Projeto MAPBIOMAS	2015	raster/categorico	30 metros	local
MAPB_LULC_190_100M	MAPBIOMAS	Superfície antrópica	Projeto MAPBIOMAS	2015	raster/categorico	30 metros	local

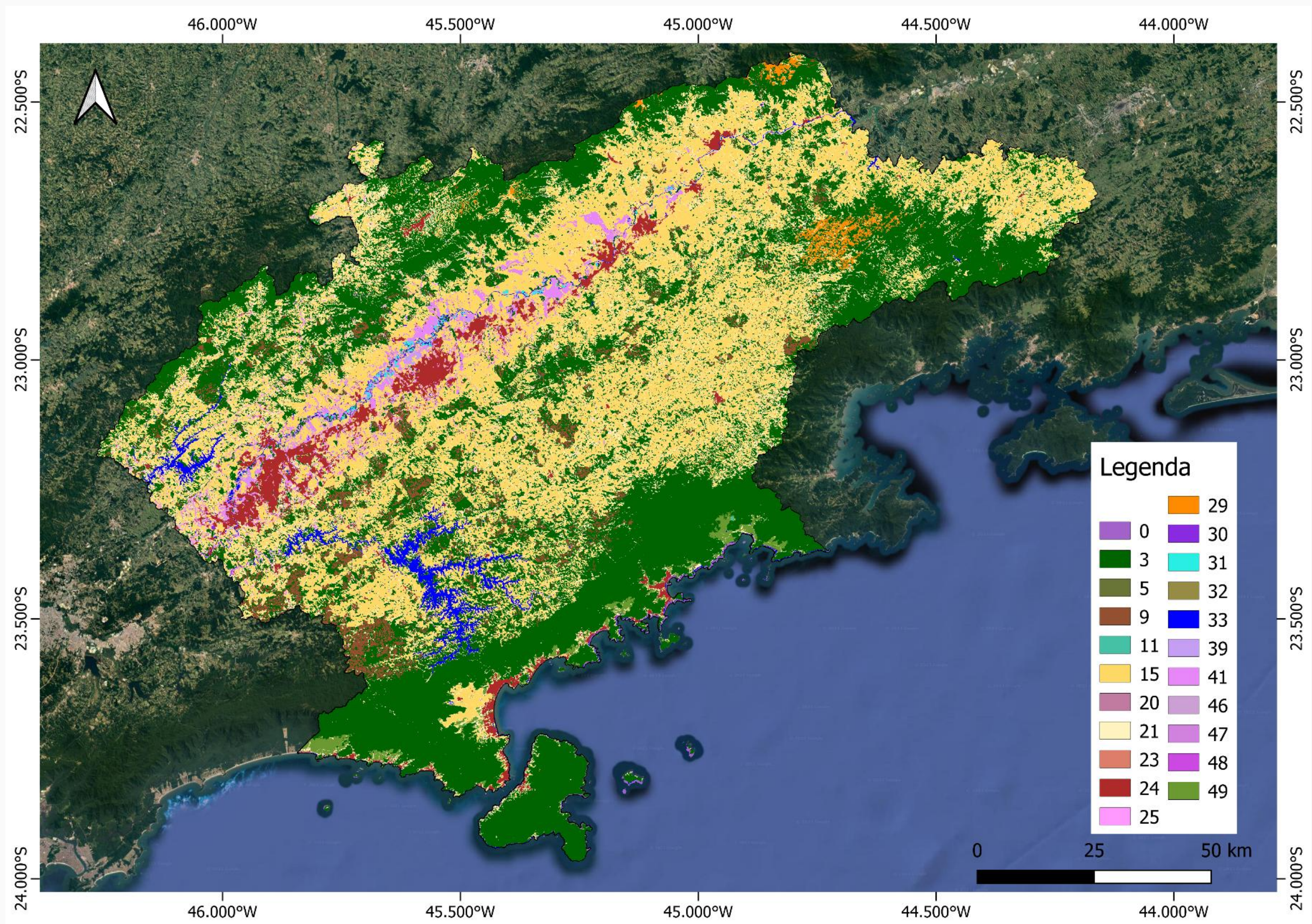
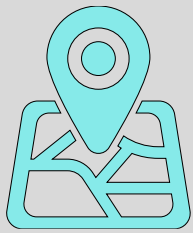
DADOS INPUT

Nomenclatura	Dados utilizados	Descrição	Fonte	Ano	Formato	Resolução Espacial	Contexto
MAPB_LULC_200_100M	MAPBIOMAS	Solo exposto	Projeto MAPBIOMAS	2015	raster/categórico	30 metros	local
SOS_LULC_100M	SOS Mata Atlântica	Classes de remanescentes florestais	SOS MA	2015	vetorial	—	local
TIPO_MOB_100M	Tipologias de mobilidade	Classificação de tipologias de mobilidade para a RMVPLN	ARRUDA e colaboradores (2021)	2021	linhas/vértices	—	local
GHS_BUILT_100M	GHS_BUILT_S	Total de superfície construída e áreas residenciais e não residenciais	JRC (Comissão Europeia)	2015	raster categórico	10 metros	geral
GHS_LAND_100M	GHS_LAND	Distribuição da superfície da terra estimada pela presença/ausência de população	JRC (Comissão Europeia)	2018	raster binário	10 metros	geral
CENSO		População de fato	IBGE	2010	tabular		local

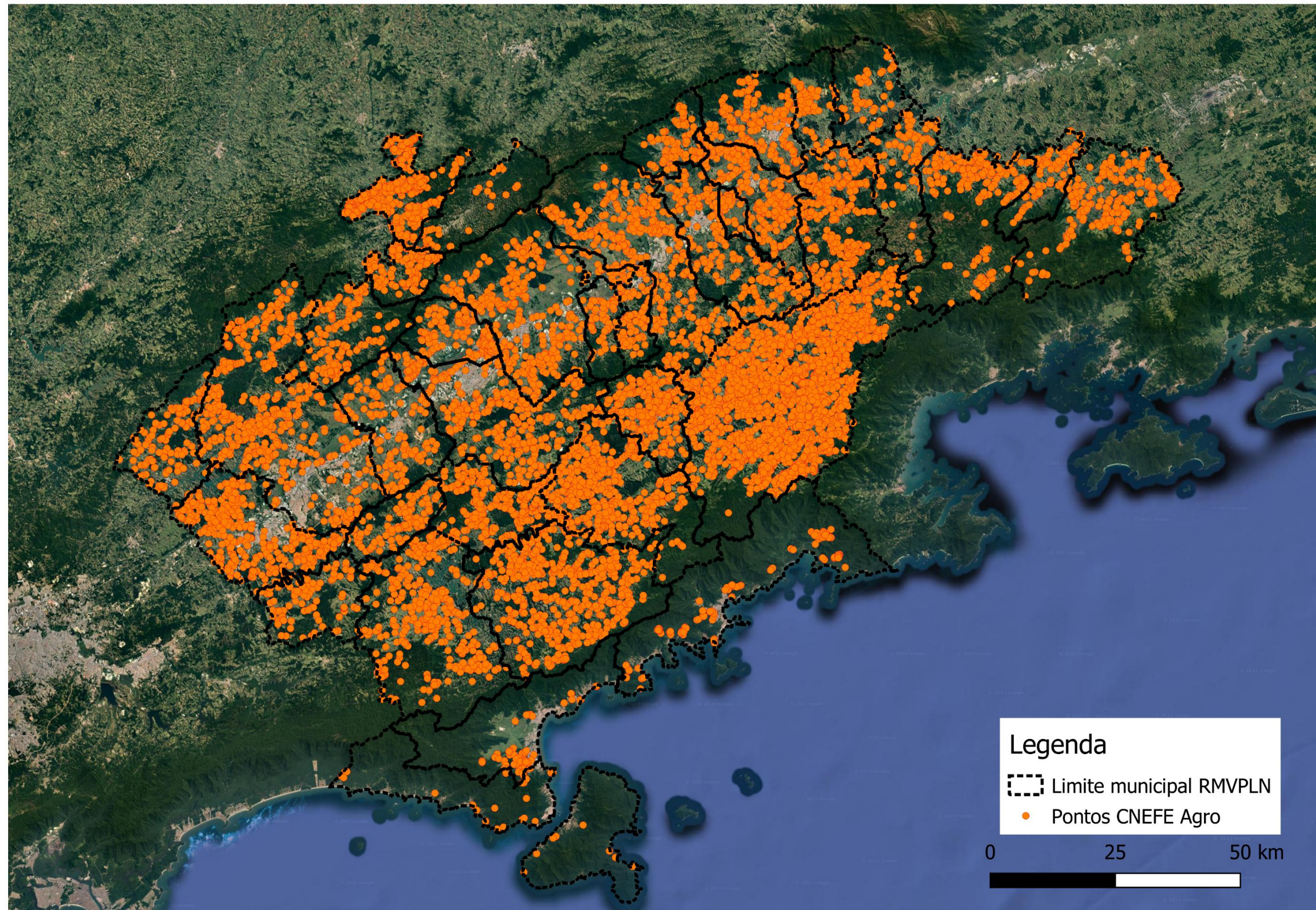
LULC - CCI



LULC - MAPBIOMAS



Pontos CNEFE Agro



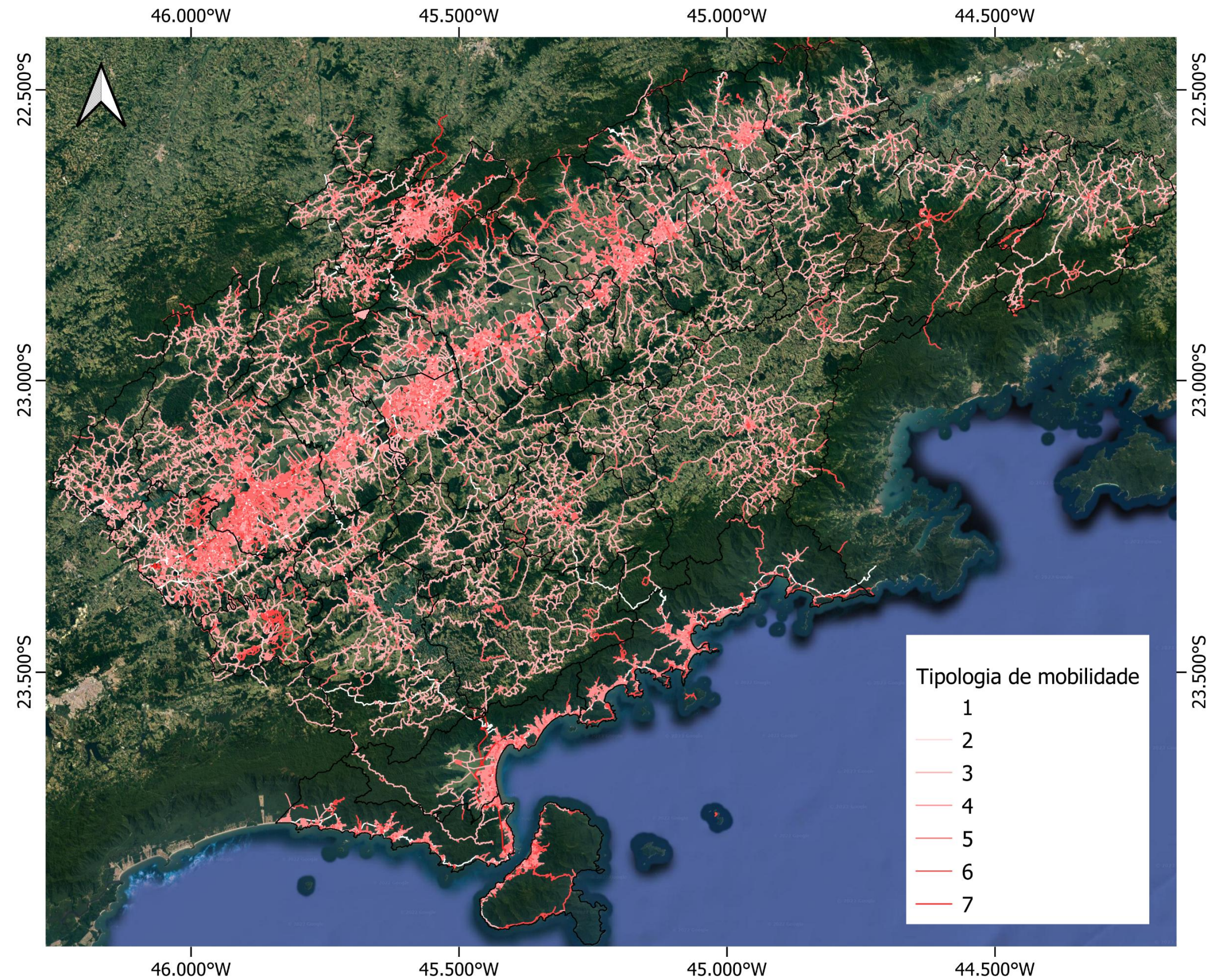
Tipologias de mobilidade



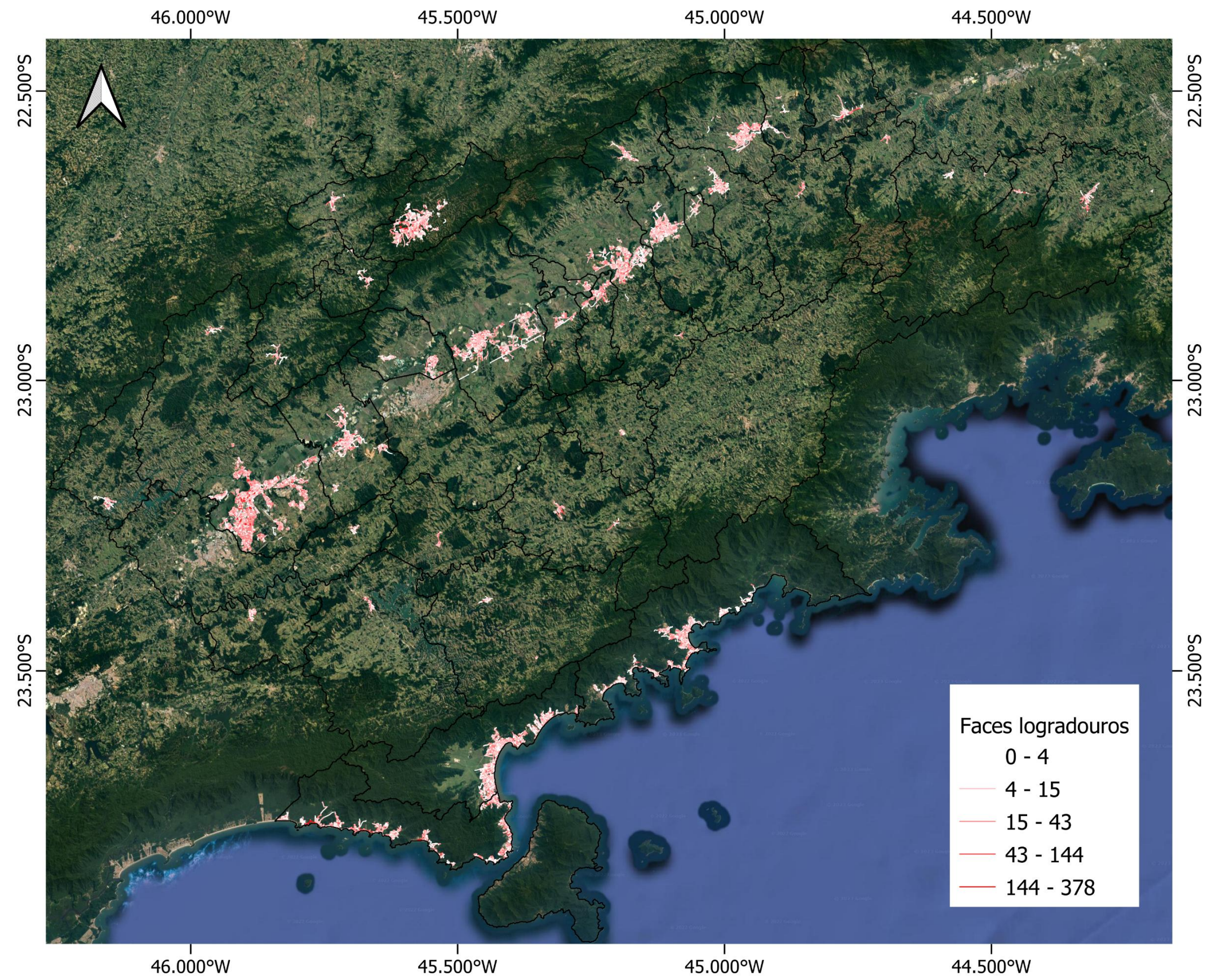
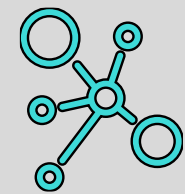
Tabela 1. Descrição das Tipologias de Mobilidade, sua potencial circulação e a as categorias utilizadas do OSM.

Circulação	Tipologias	Categorias OSM	Descrição	Tipologia (n)
Circulação regional	Mobilidade por vias rápidas entre cidades	<i>Motorway, Motorway - Link, Trunk, Trunk - Link, Primary, Primary - Link</i>	As vias terrestres rápidas correspondem as rodovias e vias expressas, com velocidade de circulação $\geq 80\text{km/h}$, reservadas a veículos motorizados e Vias terrestres arteriais, com velocidade de circulação $\geq 60\text{km/h}$, conectando cidades	1
Circulação municipal	Mobilidade por vias coletoras na cidade	<i>Secondary, Secondary - Link</i>	Vias terrestres coletoras, com velocidade de circulação média $\geq 40\text{km/h}$, conectando cidades e povoados (Secundárias)	2
Circulação local	Mobilidade por vias locais	<i>Tertiary, Tertiary -Link, Unclassified</i>	Vias terrestres locais, com menor fluxo de veículos, podendo ser vias não classificadas (vias locais sem uso residencial), conectando bairros e/ou permitindo uma rede de circulação secundária em cidades pequenas.	3
Circulação local residencial	Mobilidade por vias residenciais	<i>Residential</i>	Vias locais, de baixo fluxo de veículos, dentro de área residencial	4
Circulação local de serviços e outros	Mobilidade por vias de acesso	<i>Road, Service, Raceway, Living Street, Track</i>	Estradas rústicas (<i>track</i> – vias para uso agrícola, florestal), de caráter de acesso ou o interior de propriedades industriais, acampamentos, estacionamentos (vias de serviço), e pistas de corrida (<i>raceway</i>), vias de preferência de pedestres, e vias de tipo desconhecido (<i>road</i>)	5
Circulação local não motorizado	Mobilidade ativa	<i>Steps, Path, Cycleway, Pedestrian, Footway, Bridleway</i>	Vias terrestres de uso não motorizado, podendo ser do tipo escadas, caminho, ciclovias, calçadão, pista de caminhada e hipovias	6
Circulação potencial futura	Mobilidade futura	<i>Abandoned, Construction, Proposed</i>	Vias terrestres abandonadas, em construção ou propostas	7

Tipologias de mobilidade

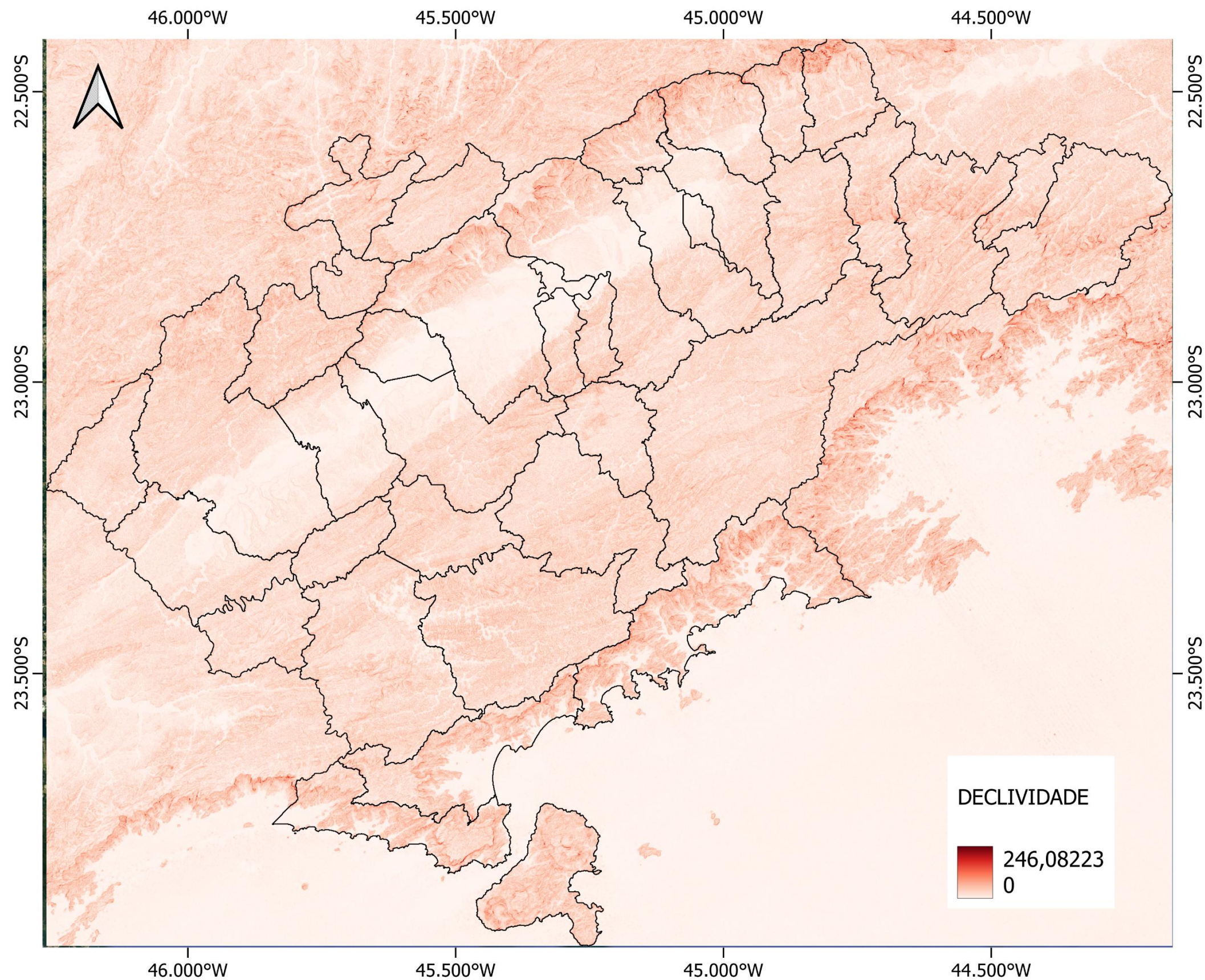


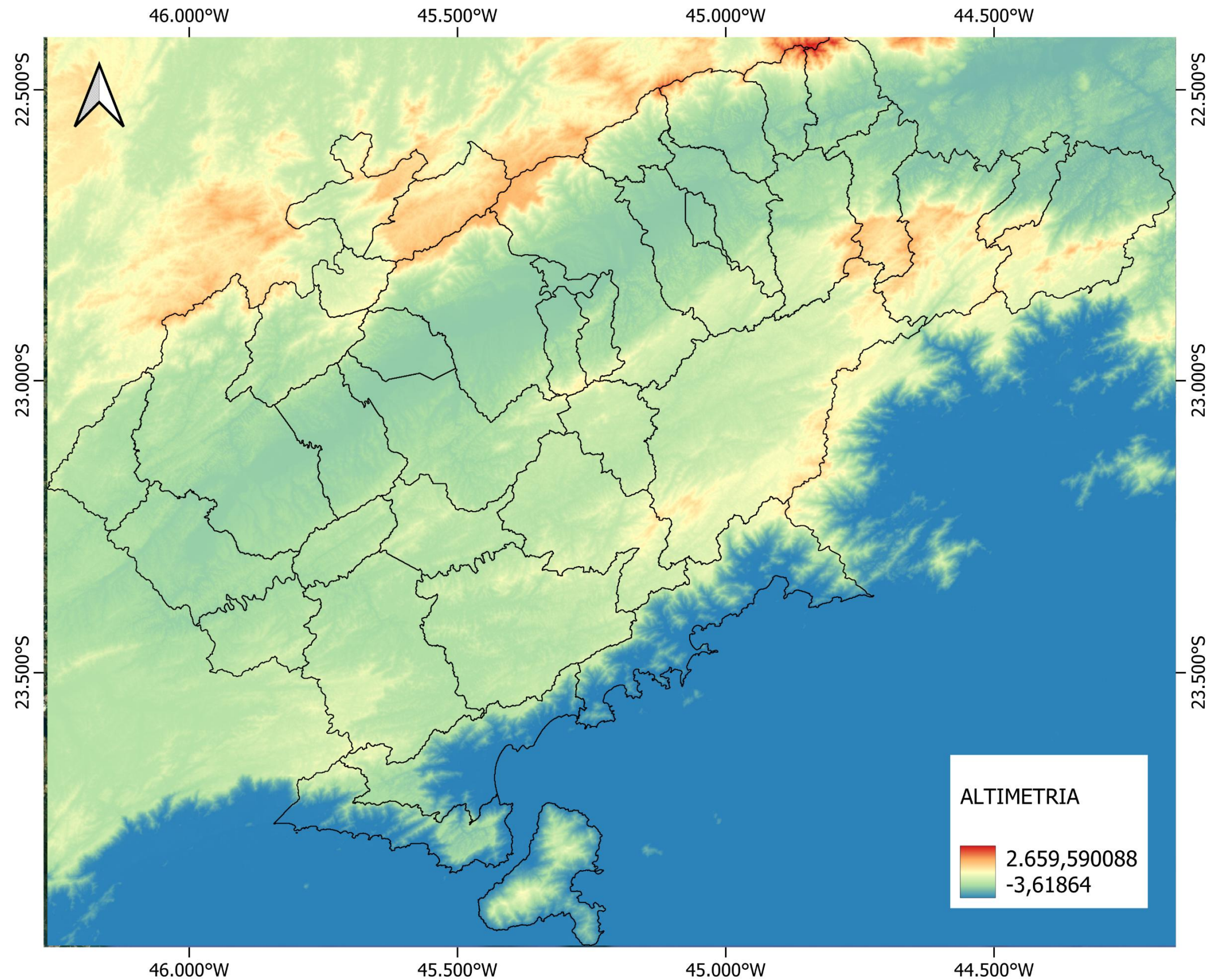
Faces logradouros



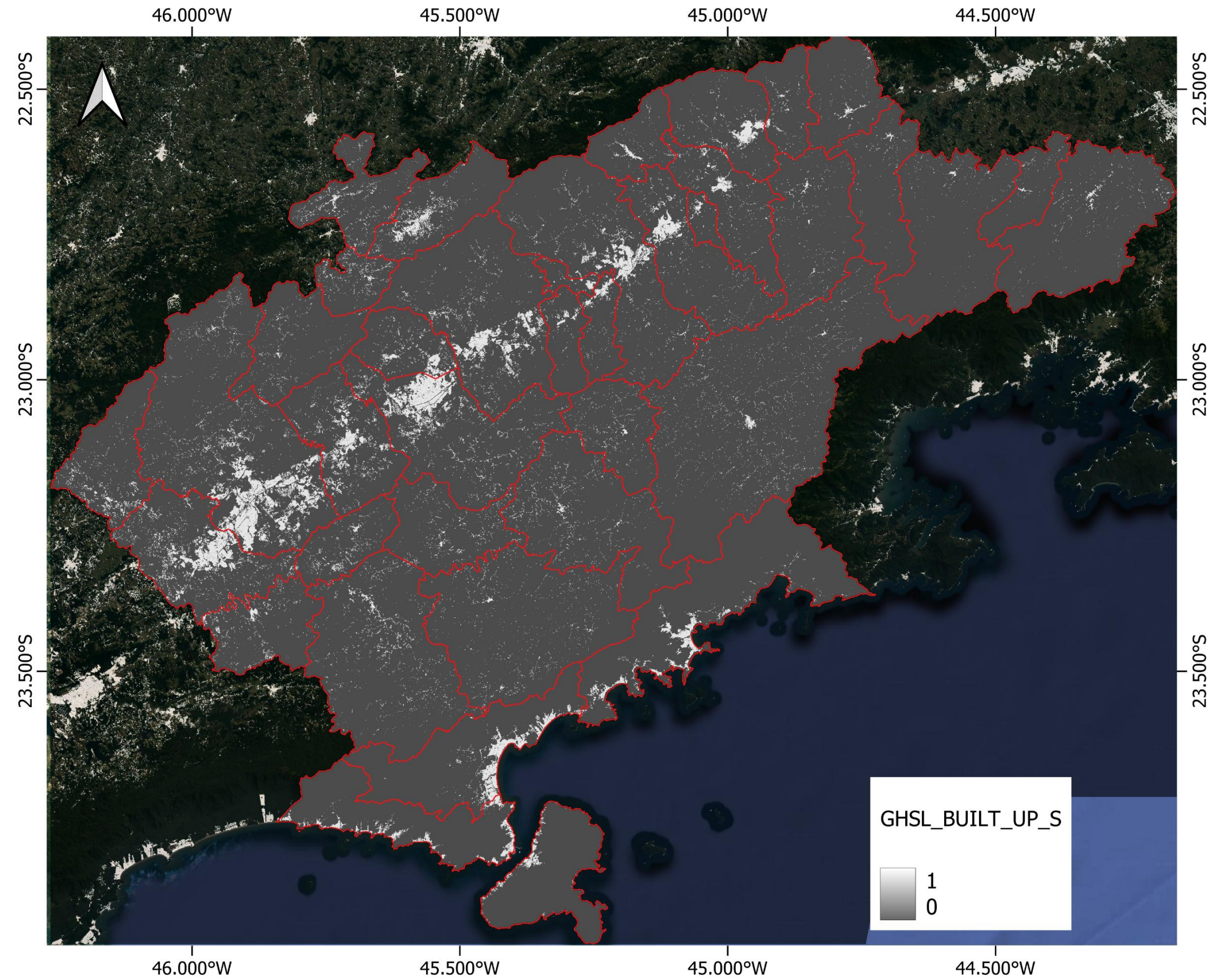
CD_GEO – código que agrega o código do setor com quadra e face;
CD_SETOR – código do setor censitário, o setor censitário é único no distrito;
CD_QUADRA – código da quadra dentro do setor censitário, que é único no setor;
CD_FACE – código da face dentro da quadra, que é único dentro da quadra;
NM_TIPO_LO – nome do tipo de logradouro
NM_TITULO_LO – nome do titulo do logradouro
NM_NOME_LO – nome do nome do logradouro
TOT_RES – total de endereços de residência existente na face;
TOT_GERAL – total de endereços existente na face (não é contabilizado terrenos, vazios sem edificação, mesmo que tenham endereçamento).

Declividade

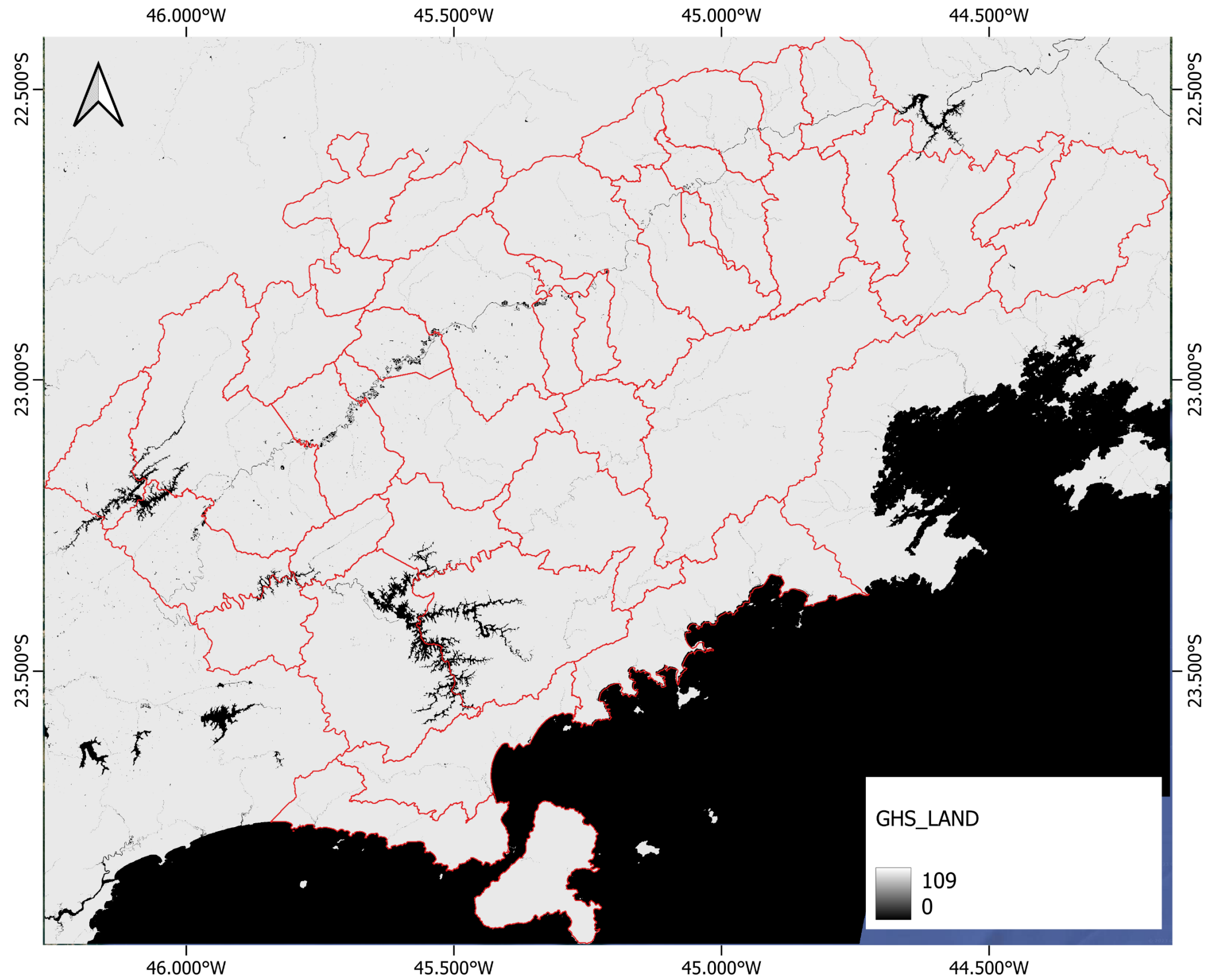




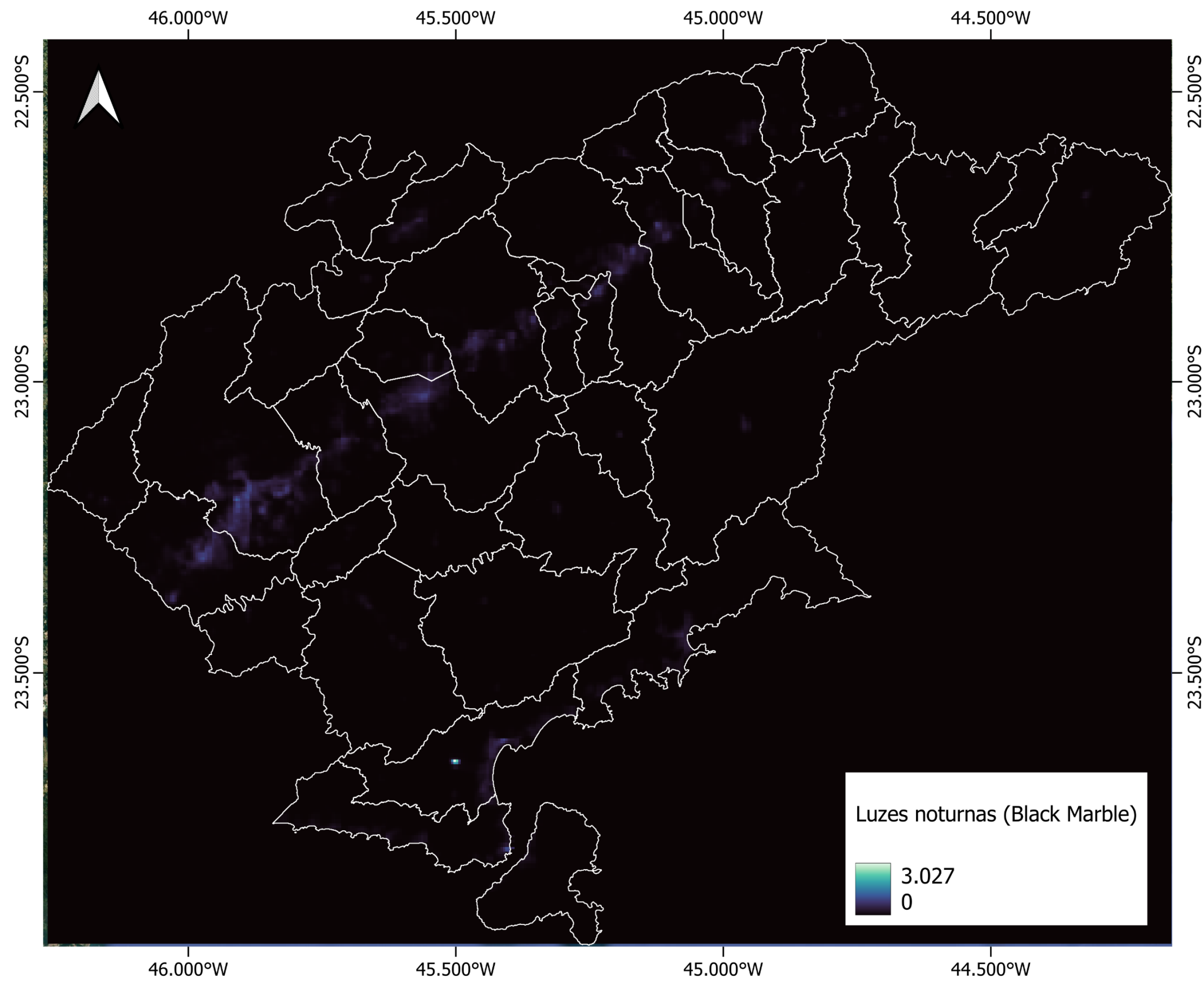
GHS_BUILT_S



GHS_LAND



Black Marble



Etapa 1: Harmonização LULC

Classes reclassificadas	Classes MAPBIOMAS/código	SOS Mata Atlântica
Área de cultivo	Café / 46, outras lavouras temporárias / 41, permanentes / 48, Cana / 20 e Citrus / 47	-
Área de floresta lenhosa	Formação florestal / 3 e Silvicultura / 9	Mata
Vegetação de arbustos	-	Restinga; refúgios vegetacionais
Vegetação de herbáceas	-	-
Áreas de vegetação esparsa	Mosaicos de usos / 21 e Pastagem / 15	-
Vegetação do tipo aquática	Mangue / 5, Campo alagado e área pantanosa / 11	Mangue; vegetação de várzea
Superfície antrópica	Área urbanizada / 24 e Mineração / 30	-
Solo exposto	Praia, duna e areal / 23 e Afloramento rochoso / 29	-

Etapa 1: Procedimentos - densidade kernel

1. Fórmula do Kernel Quártico

$$K(x_i, x_j) = \left(1 - \|x_i - x_j\|^2\right)^2$$

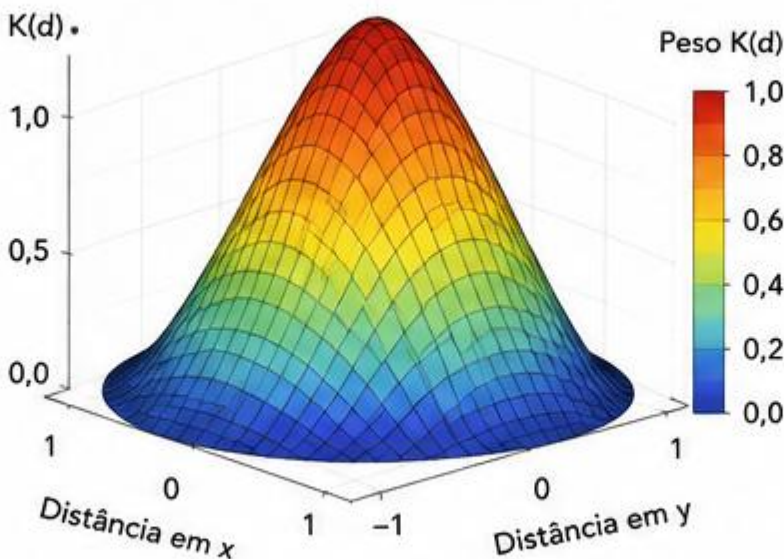
onde:

- x_i = localização onde a densidade está sendo calculada
- x_j = localização do ponto que contribui
- $\|x_i - x_j\|$ = distância **NORMALIZADA** entre os pontos (distância real d dividida pelo raio de influência h)
- $K(x_i, x_j)$ = peso/influência do ponto x_j sobre x_i

O kernel é diferente de zero apenas até a distância 1 (ou seja, até o raio h). Fora disso, o peso é 0.

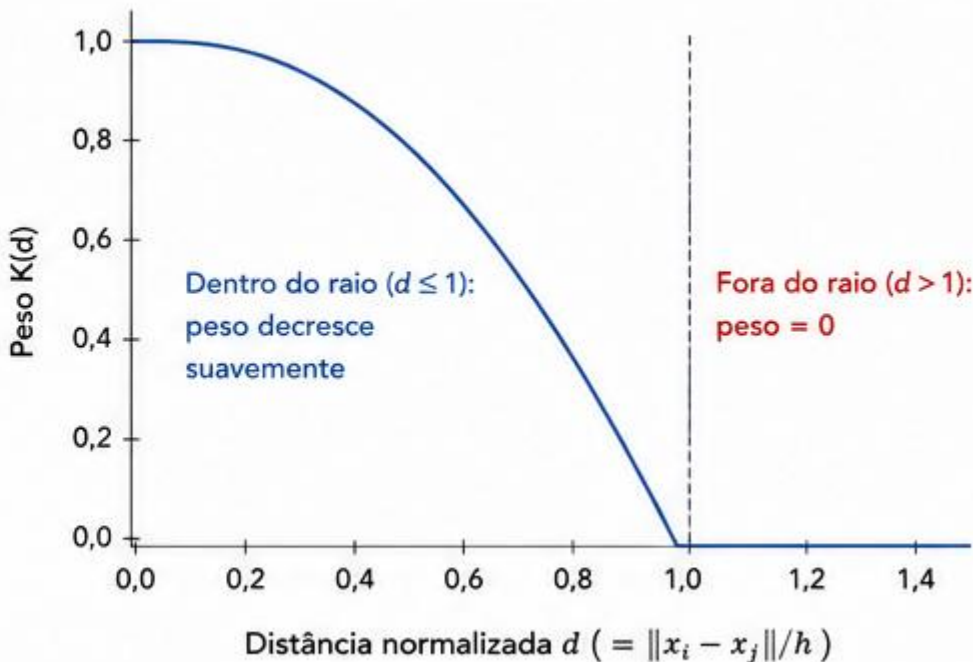
$$K(x_i, x_j) = \begin{cases} (1 - d^2)^2, & 0 \leq d \leq 1 \\ 0, & d > 1 \end{cases}, \text{ com } d = \frac{\|x_i - x_j\|}{h}$$

2. Superfície 3D do Kernel Quártico

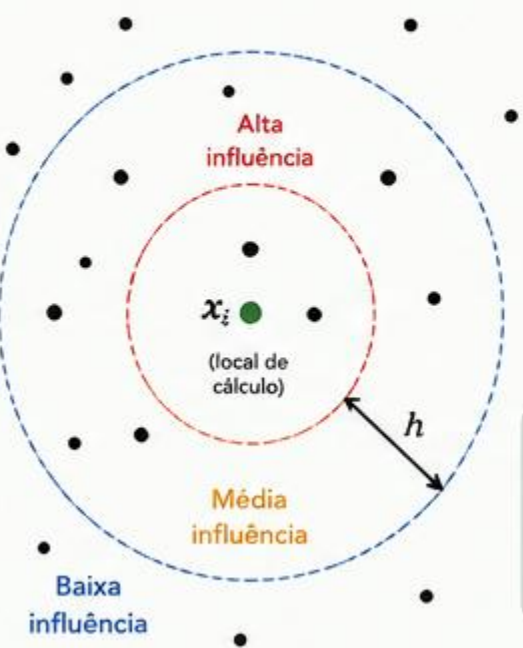


O peso é máximo no centro ($d = 0$) e decresce suavemente até zero quando $d = 1$.

3. Perfil do Kernel (1D radial)



4. Interpretação espacial



- Pontos próximos (d pequeno) recebem alto peso.
- Pontos distantes (d próximo de 1) recebem baixo peso.
- Pontos além do raio h não contribuem (peso = 0).

Raio de influência: h
Todos os pontos a uma distância maior que h têm influência zero sobre x_i .

5. Valores do Kernel Quártico

Distância normalizada $d = \frac{\ x_i - x_j\ }{h}$	Kernel $K(d) = (1 - d^2)^2$	Peso relativo (%)
0,0	1,0000	100%
0,1	0,9801	98,0%
0,2	0,9216	92,2%
0,3	0,8281	82,8%
0,4	0,7056	70,6%
0,5	0,5625	56,3%
0,6	0,4096	41,0%
0,7	0,2601	26,0%
0,8	0,1296	13,0%
0,9	0,0361	3,6%
1,0	0,0000	0%
> 1,0	0,0000	0%

6. Densidade Kernel Estimada

A densidade em x_i é a soma dos pesos dos pontos dentro do raio h , podendo incluir um valor ou população w_j associado a cada ponto:

$$\hat{f}(x_i) = \frac{1}{h^2} \sum_{j=1}^n w_j K\left(\frac{\|x_i - x_j\|}{h}\right) \quad \text{para } \|x_i - x_j\| \leq h$$

$(K = 0 \text{ para } \|x_i - x_j\| > h)$

onde:

- w_j = peso ou população do ponto j
- h = bandwidth (raio de influência)
- n = número de pontos que caem dentro do raio h

Características principais



Suave e contínuo

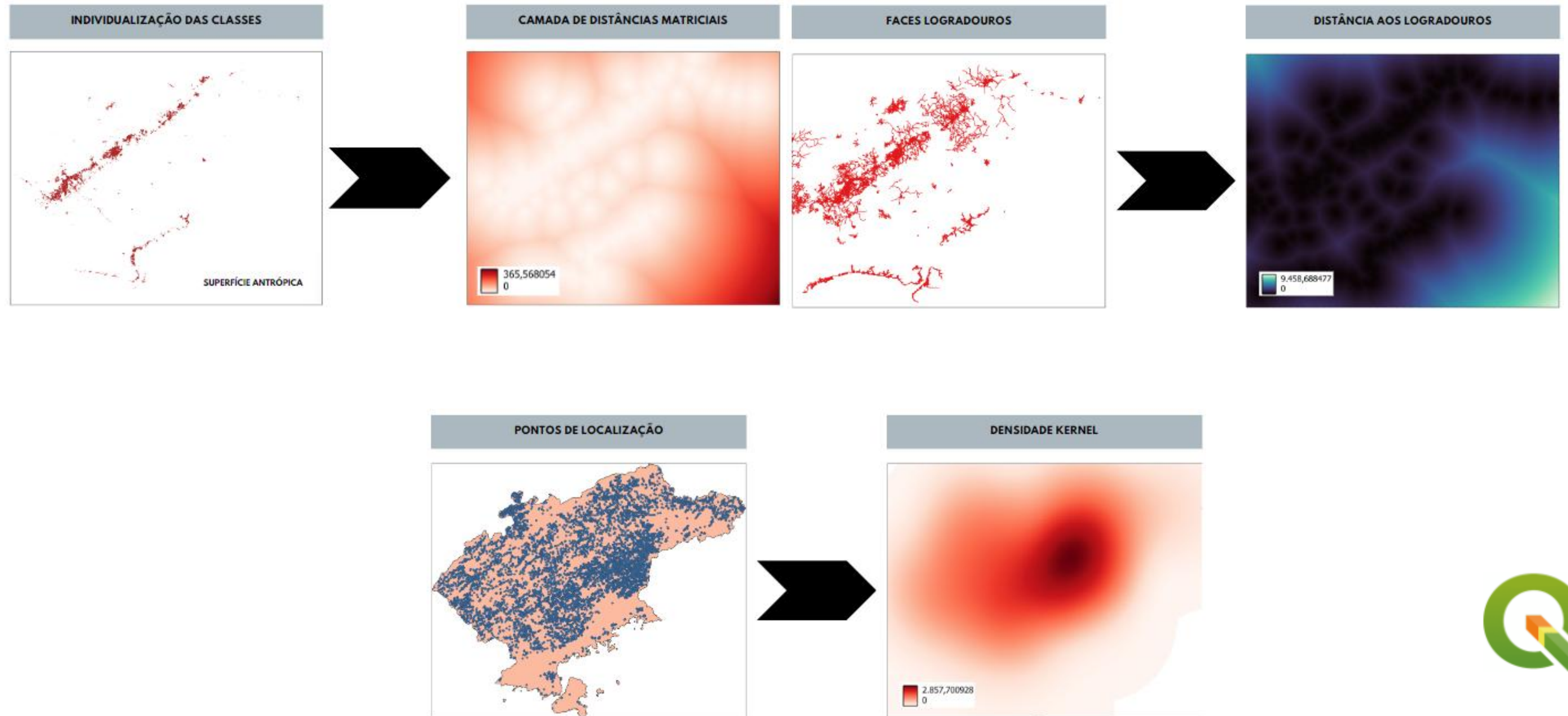


Peso máximo no centro



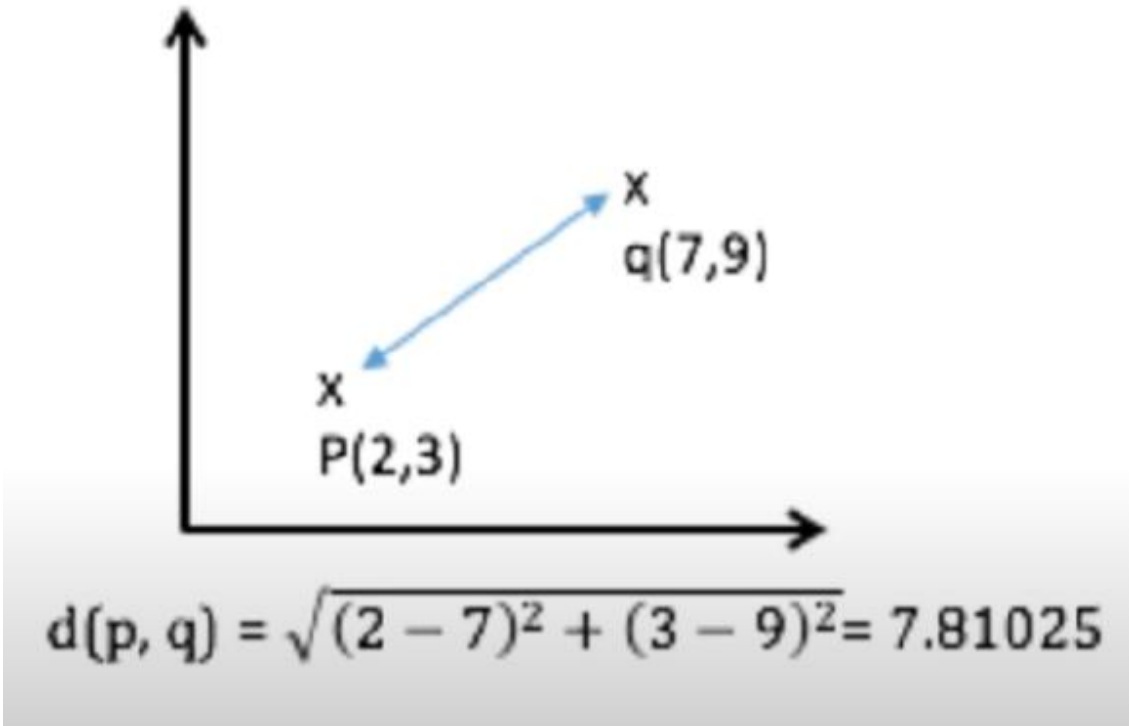
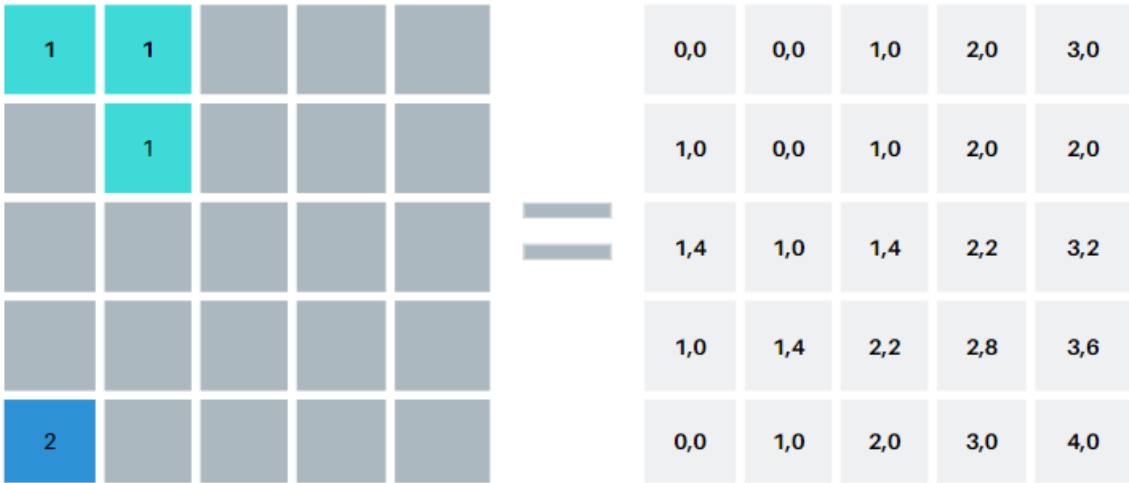
Peso zero fora do raio

Etapa 1: Procedimentos - Distância Raster

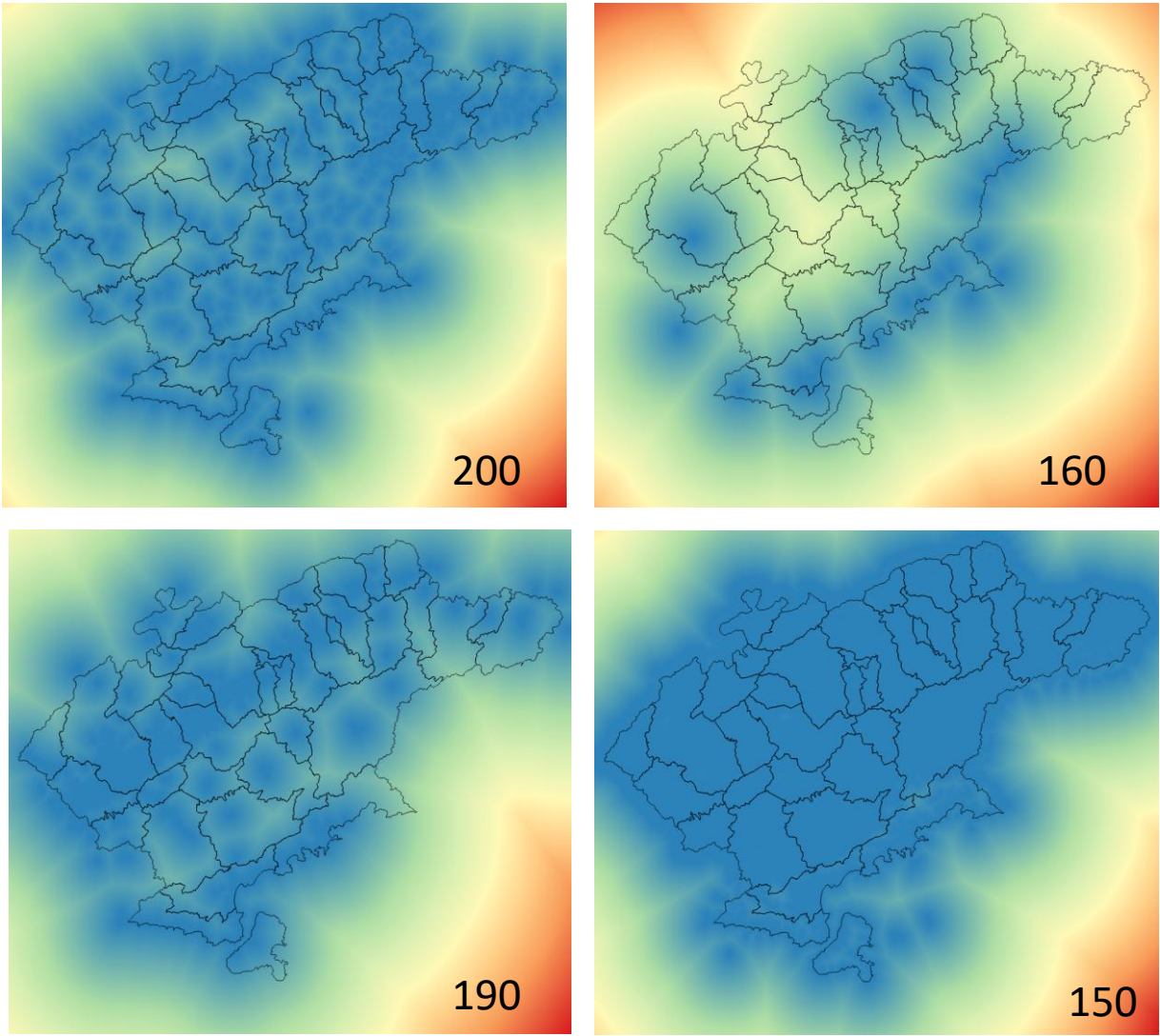


Etapa 1: Procedimentos - Distância Raster

$$d_{i,j} = \min_{p \geq 1, q \leq N} \left\{ \sqrt{(i - p)^2 + (j - q)^2} \mid b_{pq} = 0 \right\}, \text{ for all } i, j.$$

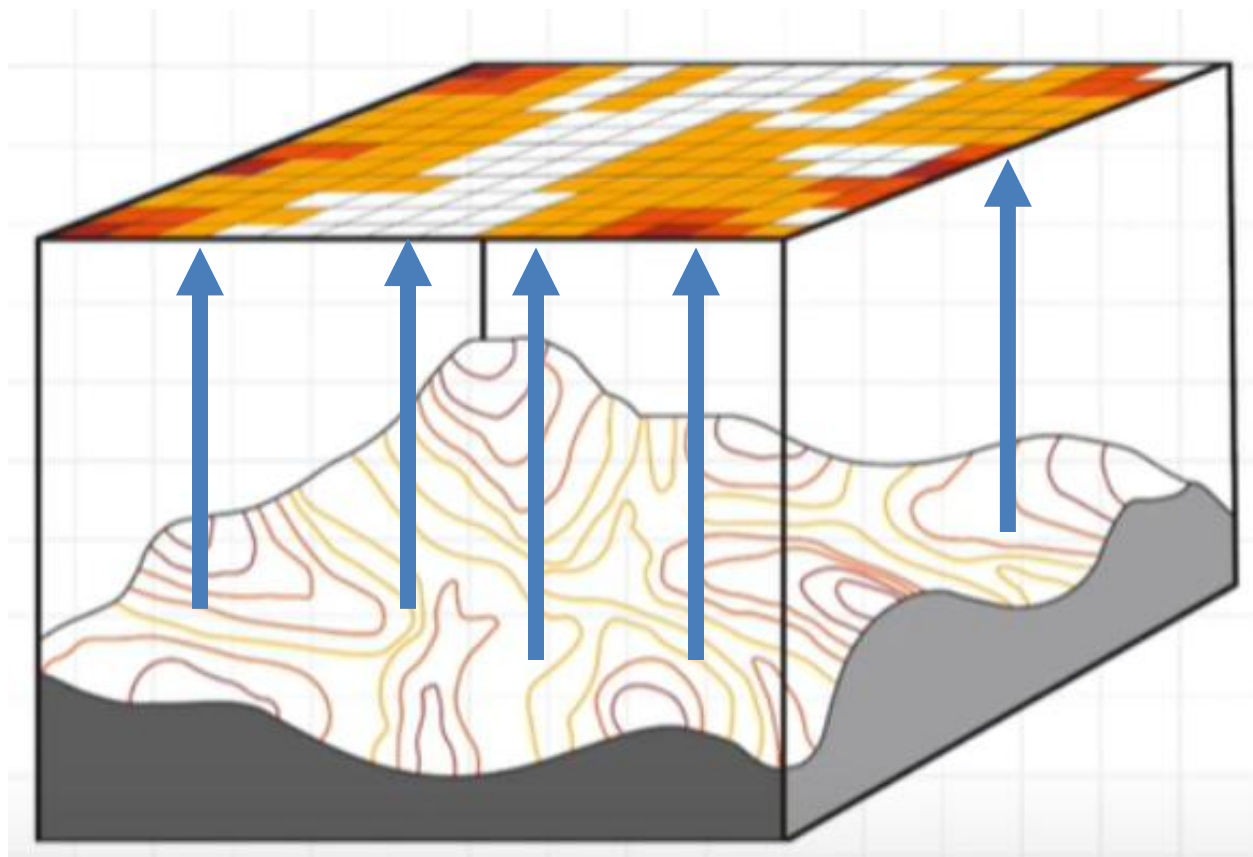


- 150 - Áreas de vegetação esparsas
- 160 - Áreas de vegetação aquática
- 190 - Áreas de superfície artificial
- 200 - Áreas de solo exposto



Etapa 1: Banco de dados RMVPop

- Extração dos valores das camadas de variáveis sintomáticas através das **médias estatística** para a camada alvo e fonte;
- Padronização dos rasters para o alvo adequado;
- Empilhamento das camadas de variáveis sintomáticas.

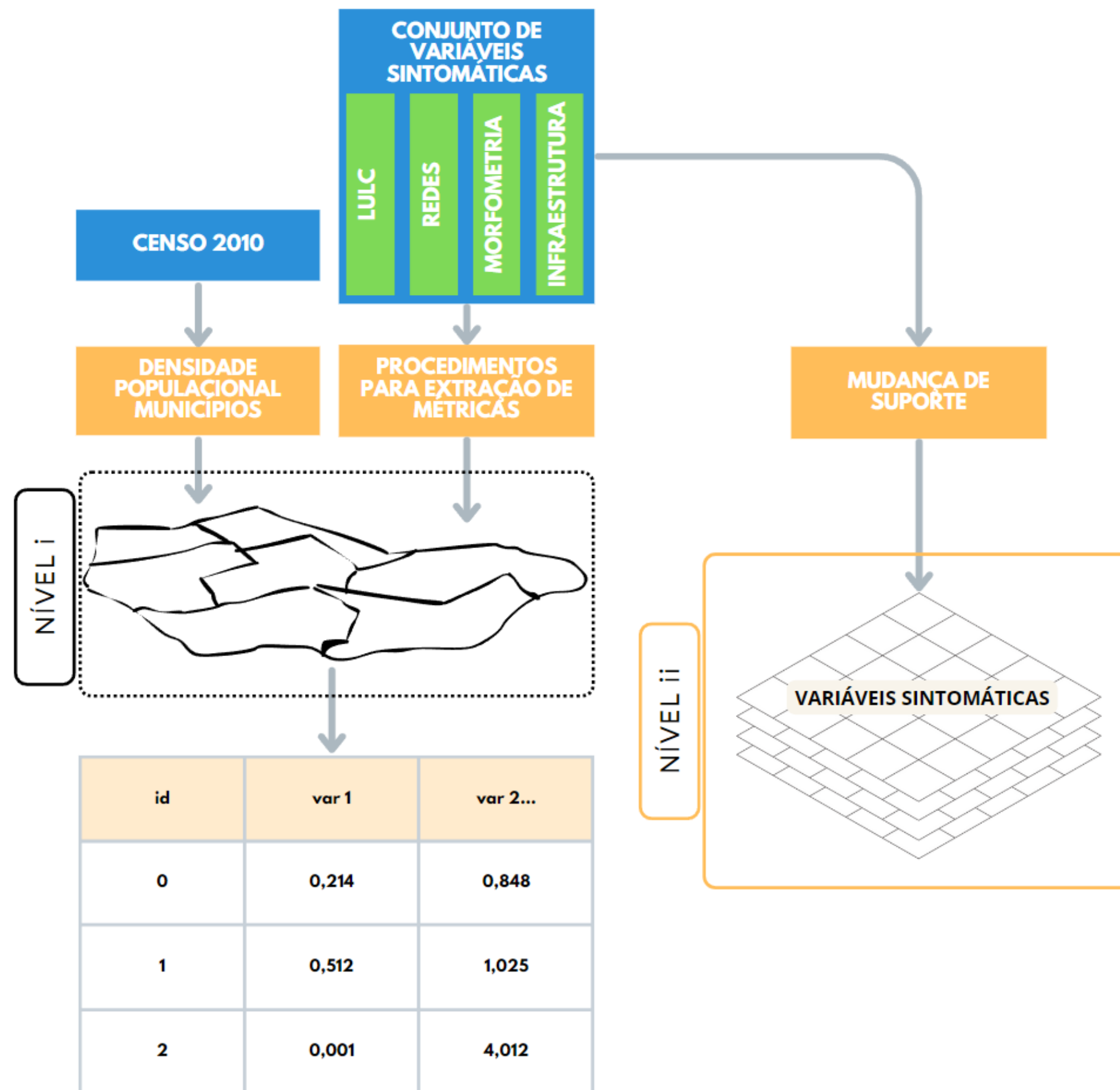


- Variável dependente: População 2010.
- Densidade populacional:
- População/Área em metros;
- Transformação BOX COX

```
(densidade ^ lambda - 1)/lambda
```

```
((1 + lambda * matriz_predicao)^(1/lambda))
```

Etapa 2: Banco de dados RMVPop



Etapa 3: Ajuste do Modelo RMVPop

- Nível I – treinamento da camada de municípios da RMVPLN.

Hiperparâmetros	Funções
<i>tuneRF</i>	ajuste <i>randomForest</i> para o parâmetro <i>mtry</i> ideal
<i>mtryStart</i>	especifica o número das variáveis aleatoriamente a serem amostradas em cada uma das árvores de decisão geradas. (x_data/3)
<i>ntreeTry</i>	define o valor de árvores da floresta (500)
<i>improve</i>	controla a melhoria relativa no OOB a ser selecionada como divisão do nó com base no limiar de impureza (0,0001)
<i>stepFactor</i>	define o tamanho do passo para busca de melhoria dos parâmetros (1,5)
<i>trace</i>	plota a relação da melhoria do modelo e o número de árvores (TRUE)
<i>doBest</i>	seleciona o último modelo considerado melhor (TRUE)
<i>nodesize</i>	controla o tamanho mínimo dos nós terminais
<i>importance</i>	variáveis mais importantes dentro do modelo: do tipo 1 e 2 (TRUE)
<i>proximity</i>	cálculo da matriz de proximidade (TRUE)
<i>sampsize</i>	tamanho da amostra para cada árvore da floresta (y_data)
<i>replace</i>	seta a reposição das amostras (TRUE)

Etapa 4: Atribuição dos Pesos para o Modelo *Dasimétrico* de Distribuição Espacial da População e Desagregação

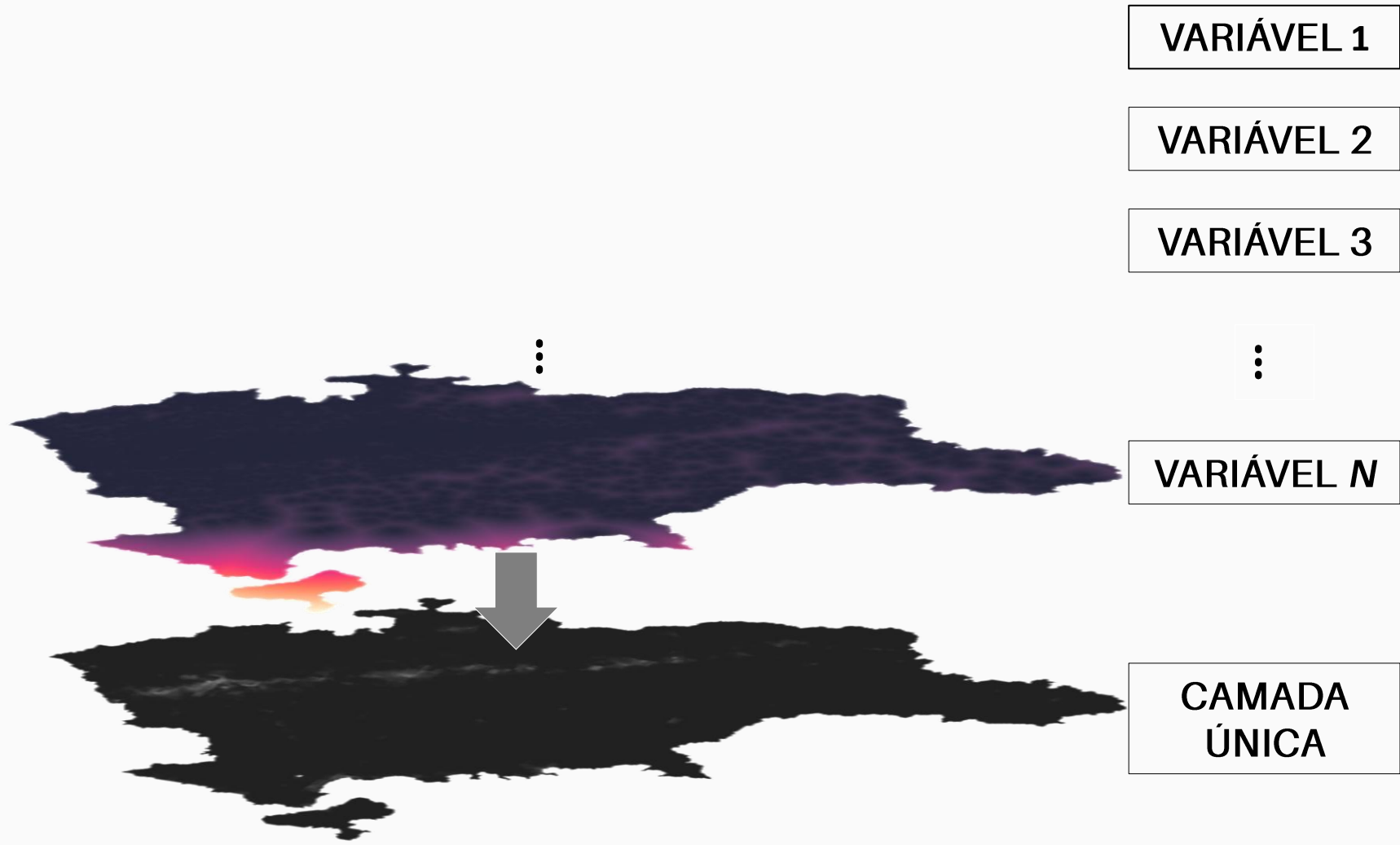
$$peso_i = \frac{\exp (estimado_i)}{\sum_{i=1}^{l_j}(estimado_i)}$$

$estimado_i$ é o modelo de predição para as áreas de cada setor e para cada célula da grade;
 l_j é o número total dos setores censitários ou células em cada município.

$\exp (estimado_i)$		
0,0194	0,0128	0,0111
0,0246	0,0174	0,0169
0,0032	0,0092	0,0244

/

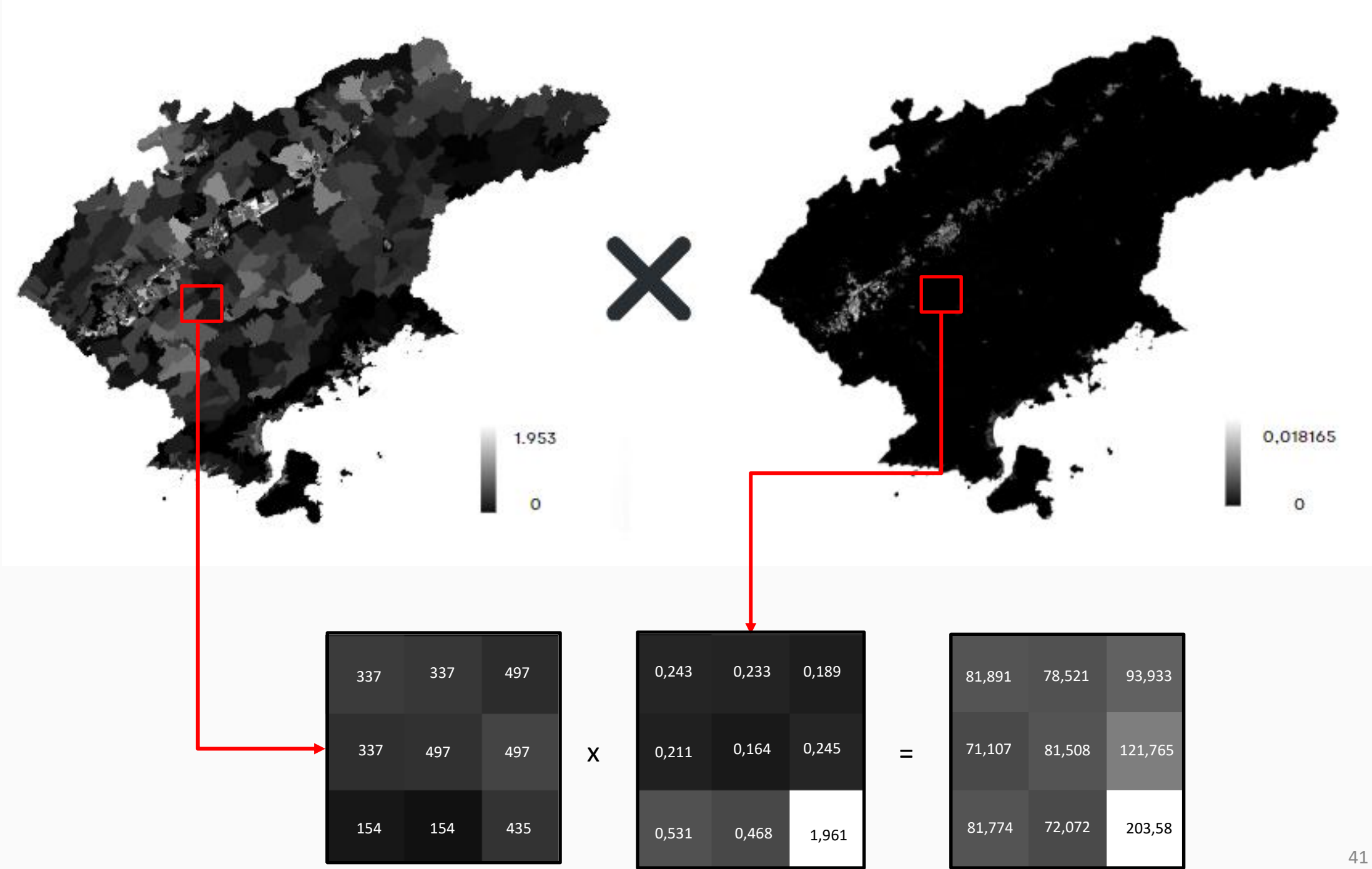
$\sum_{i=1}^{l_j}(estimado_i)$		
0,010	0,2658	0,2658
0,2658	0,2658	0,0137
0,283	0,2658	0,2658



Etapa 3: Atribuição dos Pesos para o Modelo *Dasimétrico* de Distribuição Espacial da População e Desagregação

$$população_i = total_j \times peso_i$$

$total_j$ é a população total para cada setor censitário j ;
 $peso_i$ é a proporção da população que vive em cada setor censitário ou célula da grade.



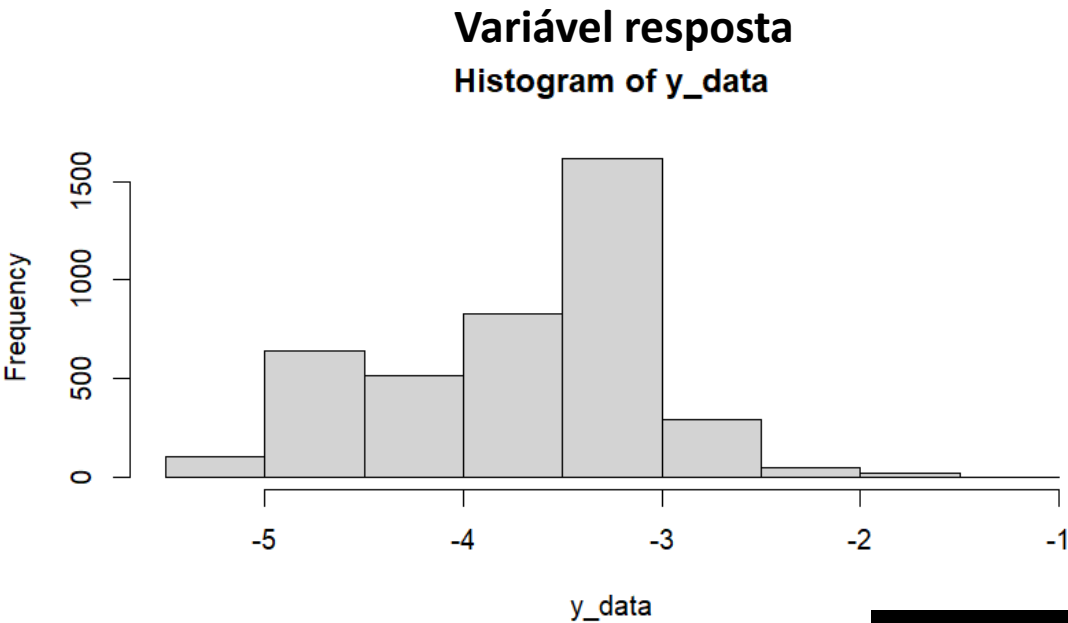
Avaliação dos protocolos

MODELOS	MAE	RMSE	% var. exp.
V003 (sem GHS e CCI)	141,86	269,53	80,28
V004 (sem CCI)	4,134001	33,06609	82,36
V005 (sem GHS)	4,187275	33,02187	80,44

GHS: GHS_BUILT_S; GHS_LAND

CCI ESA: Mapa de LU

Protocolo 1 - modelo_rmv_v003

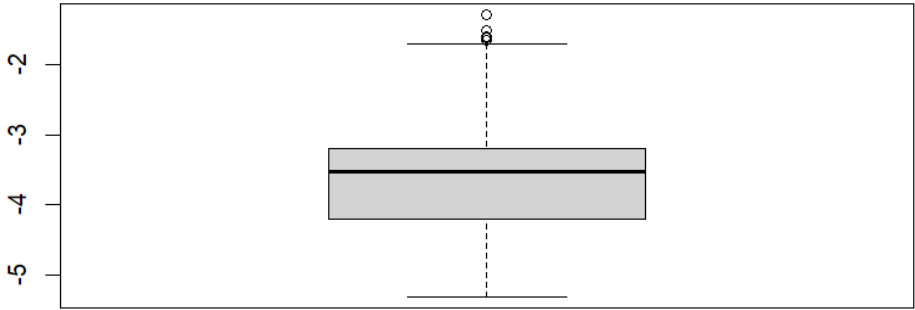


MAPB_LULC_011_100M	MAPB_LULC_040_100M	MAPB_LULC_150_100M	MAPB_LULC_160_100M
Min. : 4.5	Min. : 0.0	Min. : 0.0234	Min. : 622.7
1st Qu.: 188.1	1st Qu.: 185.9	1st Qu.: 60.9601	1st Qu.:10664.7
Median : 435.0	Median : 436.8	Median : 162.1740	Median :19408.5
Mean : 3424.7	Mean : 547.7	Mean : 257.1704	Mean :21190.7
3rd Qu.: 1266.4	3rd Qu.: 786.4	3rd Qu.: 357.5278	3rd Qu.:30520.1
Max. :46056.9	Max. :3877.4	Max. :2551.9810	Max. :54294.5

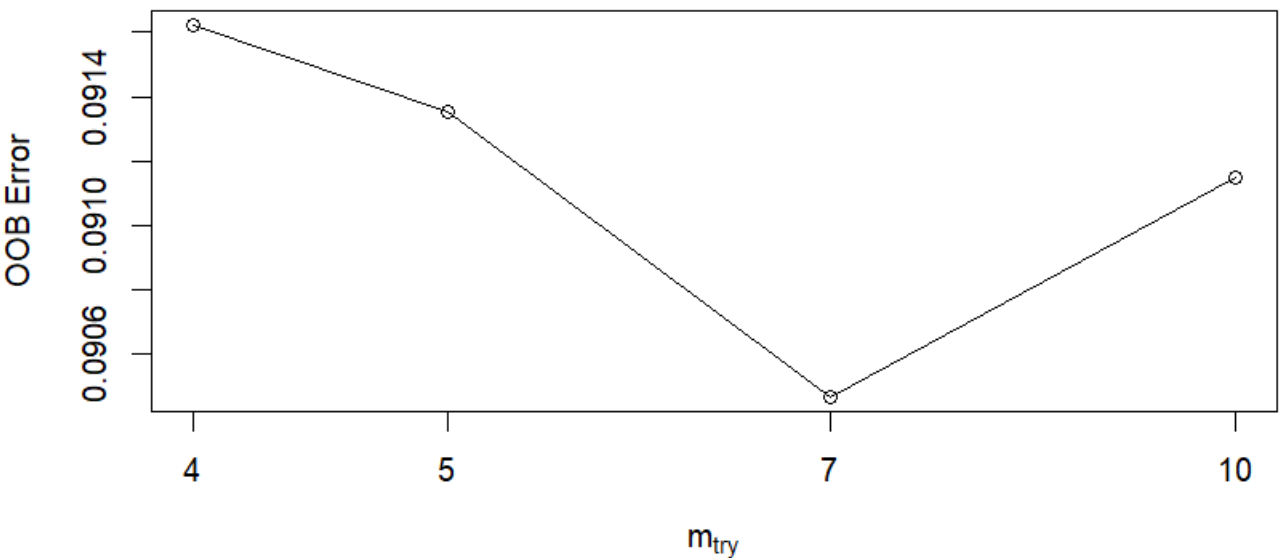
BLACKM_100M	CNEFE_FACES_100M	CNEFE_AGR_100M	TOPO_DECL_100M	TOPO_ALT_100M
Min. : 0.00	Min. : 0.5271	Min. : 46.44	Min. : 0.3892	Min. : 0.0844
1st Qu.: 64.39	1st Qu.: 2.4863	1st Qu.:1068.01	1st Qu.: 4.0077	1st Qu.: 542.3293
Median : 195.54	Median : 15.1664	Median :1305.86	Median : 7.6624	Median : 576.9869
Mean : 245.81	Mean : 208.8055	Mean :1335.19	Mean :11.6385	Mean : 552.4707
3rd Qu.: 376.48	3rd Qu.: 330.1620	3rd Qu.:1743.81	3rd Qu.:15.6741	3rd Qu.: 619.3418
Max. :1024.35	Max. :2970.0386	Max. :2854.60	Max. :53.1647	Max. :1827.5750

MAPB_LULC_190_100M	MAPB_LULC_200_100M	TIPO_MOB_100M	SOS_LULC_040_100M	SOS_LULC_130_100M
Min. : 0.000	Min. : 51.45	Min. : 0.000	Min. : 0.0	Min. : 0
1st Qu.: 0.005	1st Qu.: 2670.60	1st Qu.: 5.777	1st Qu.: 771.1	1st Qu.:24141
Median : 8.677	Median : 5510.39	Median : 22.508	Median :1990.0	Median :52144
Mean : 727.057	Mean : 7400.95	Mean : 129.176	Mean :2275.0	Mean :41803
3rd Qu.: 159.568	3rd Qu.:10270.74	3rd Qu.: 89.762	3rd Qu.:3365.5	3rd Qu.:59506
Max. :29496.990	Max. :22808.73	Max. :14743.755	Max. :8329.3	Max. :93702

SOS_LULC_160_100M
Min. : 295.8
1st Qu.: 22272.9
Median : 38468.6
Mean : 39374.7
3rd Qu.: 59156.4
Max. :100752.8

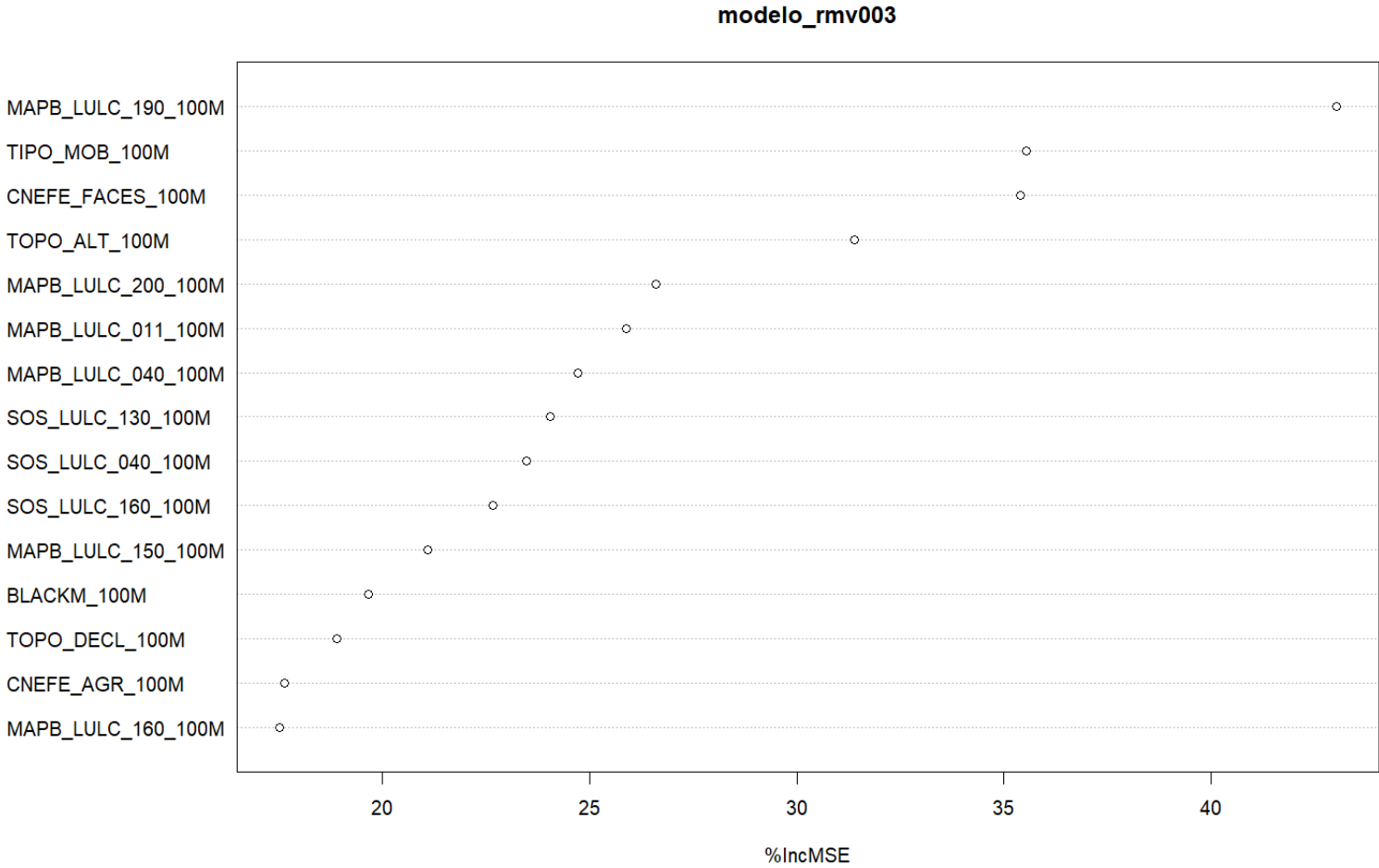
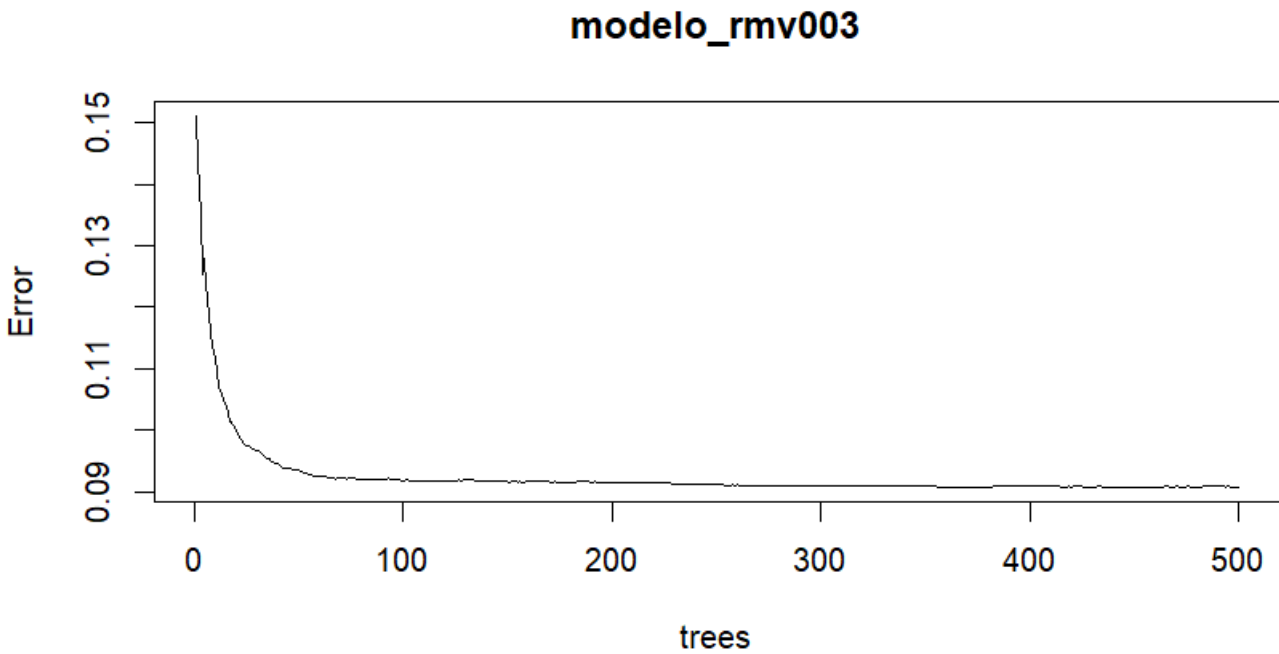


Protocolo 1 – Sem GHS e CCI (003)

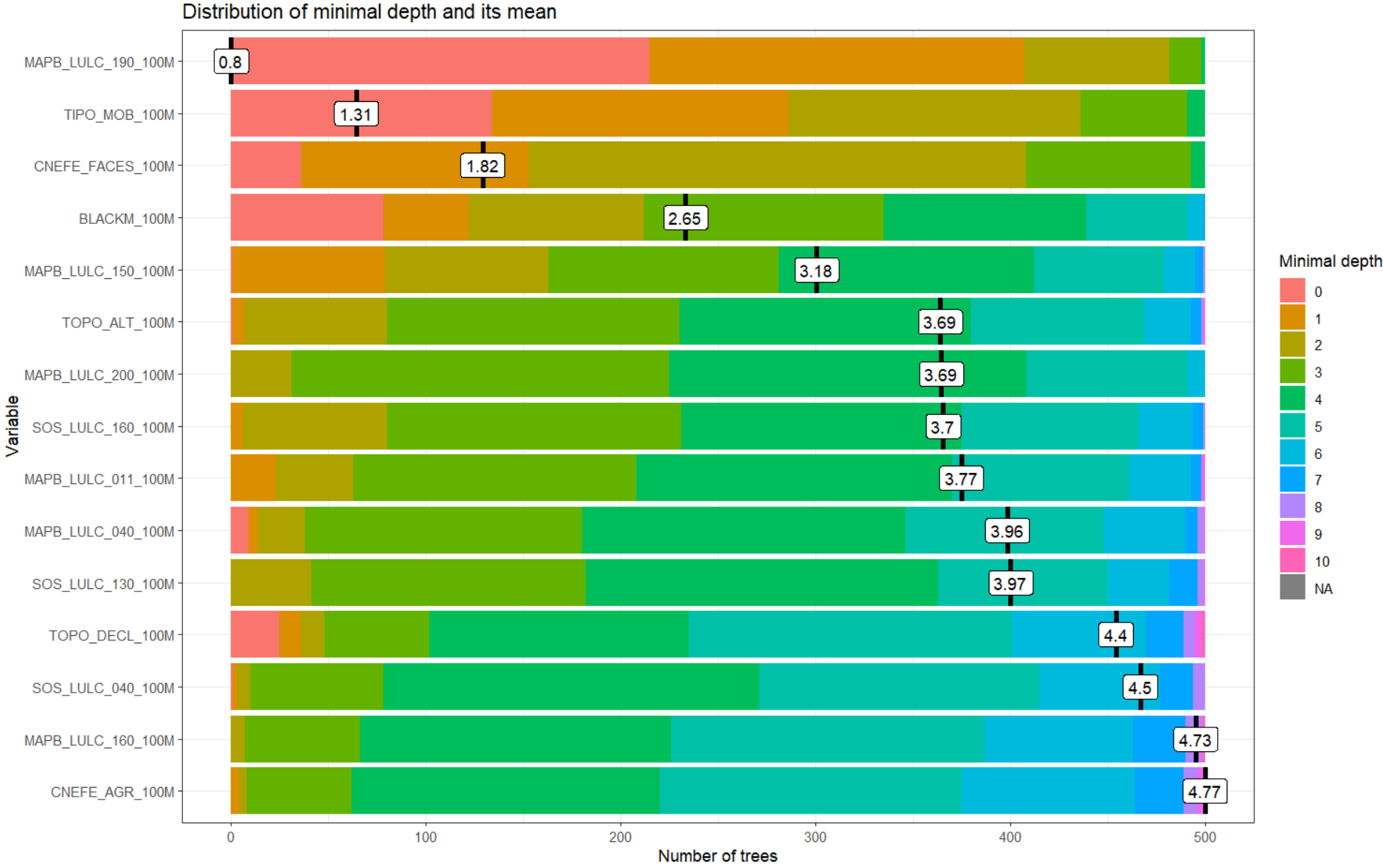
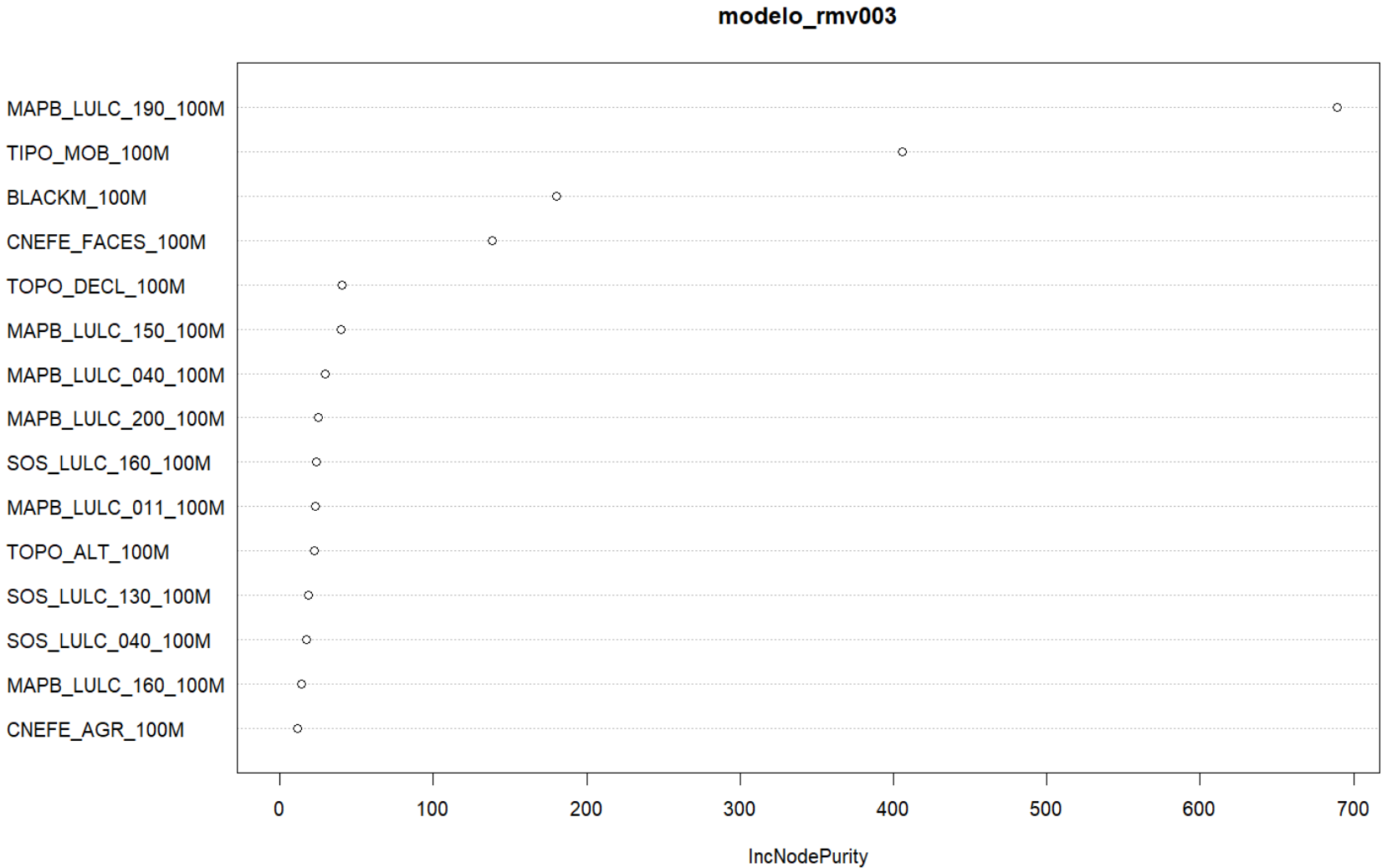


```
Call:
  randomForest(x = x, y = y, mtry = res[which.min(res[, 2]), 1],      replace = TRUE, sampsize = ..5,
    nodesize = ..1, importance = TRUE,      proximity = TRUE, na.action = ..2)
    Type of random forest: regression
    Number of trees: 500
No. of variables tried at each split: 7

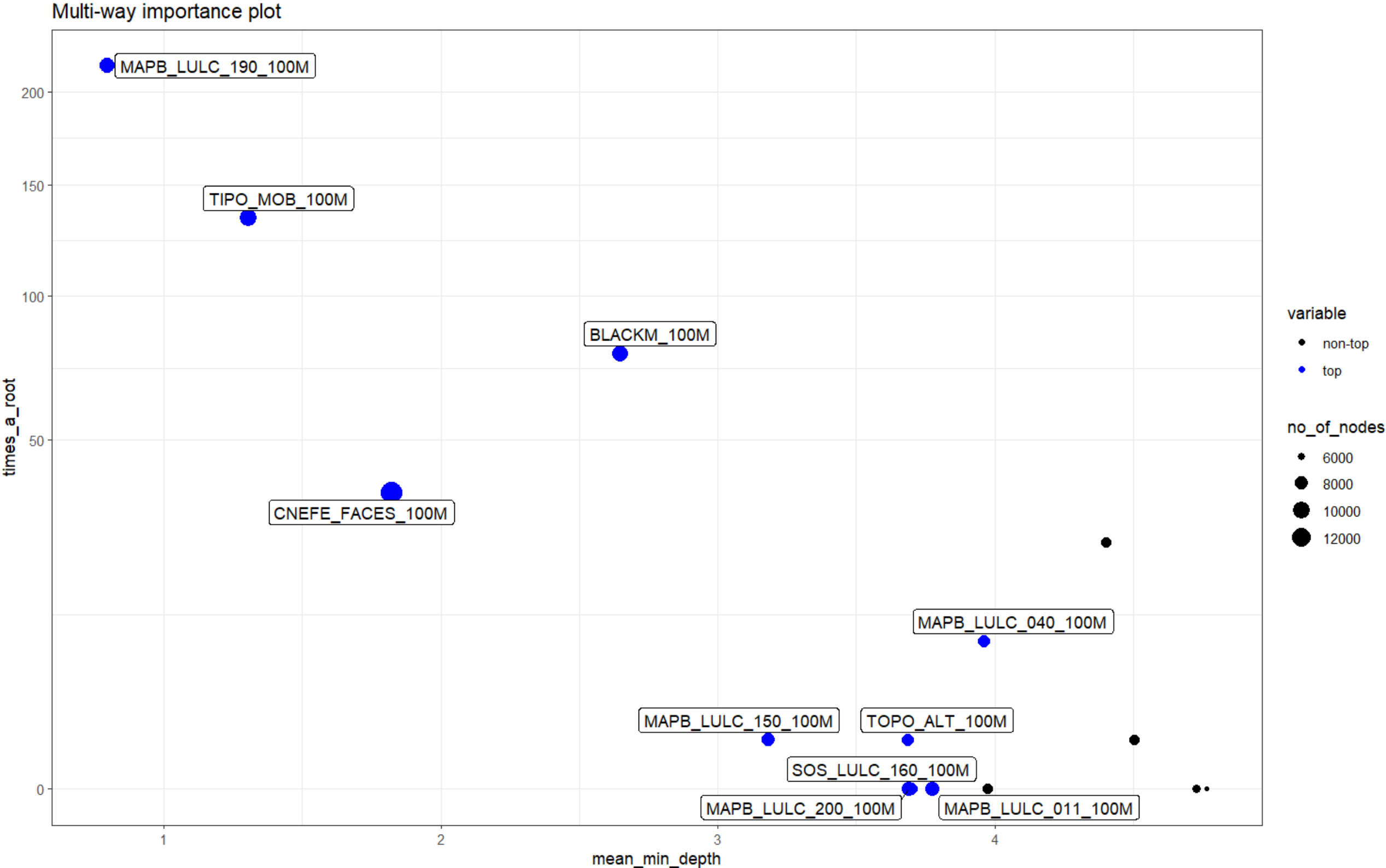
    Mean of squared residuals: 0.09074721
      % Var explained: 80.28
```



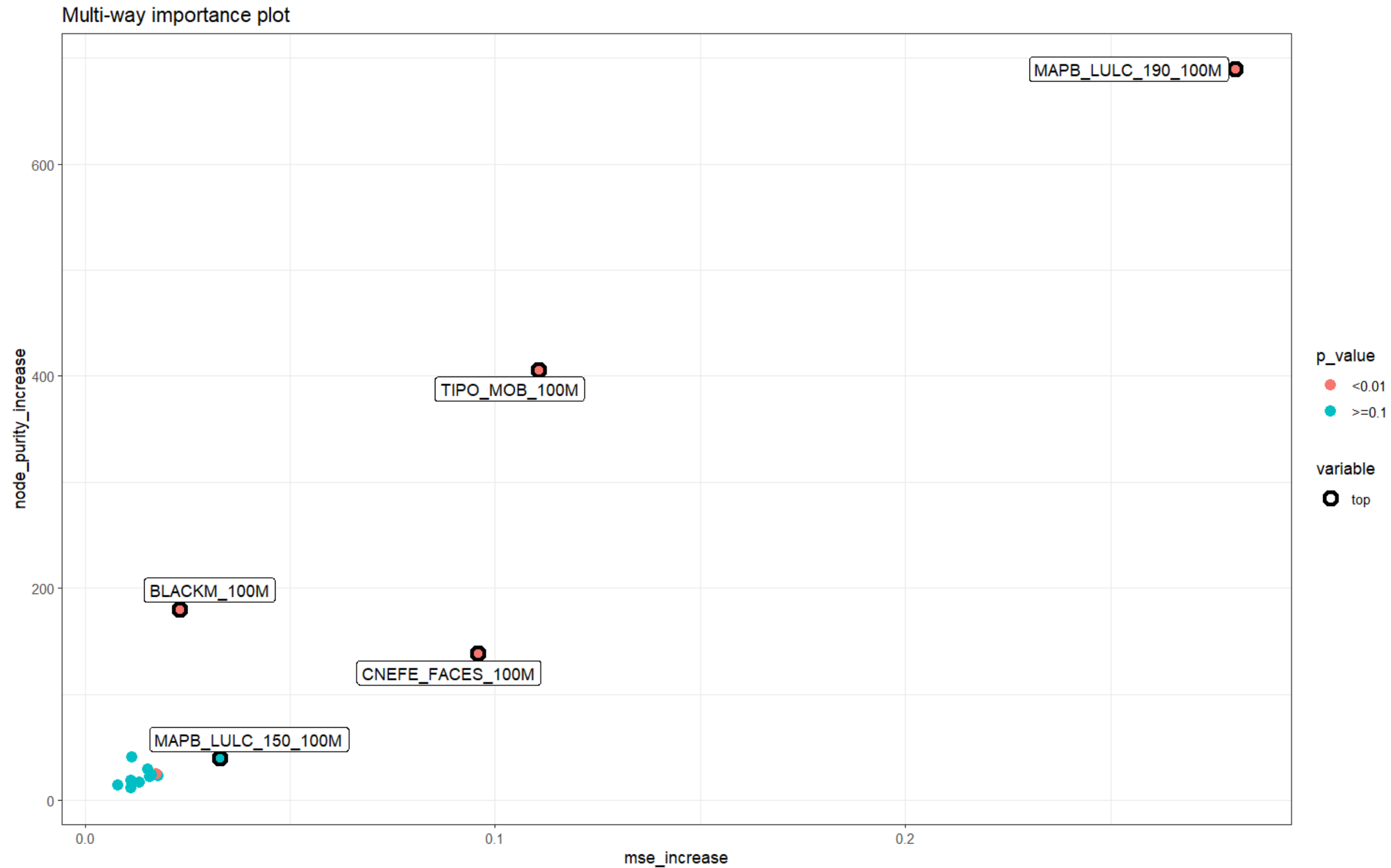
Protocolo 1 – Sem GHS e CCI (003)



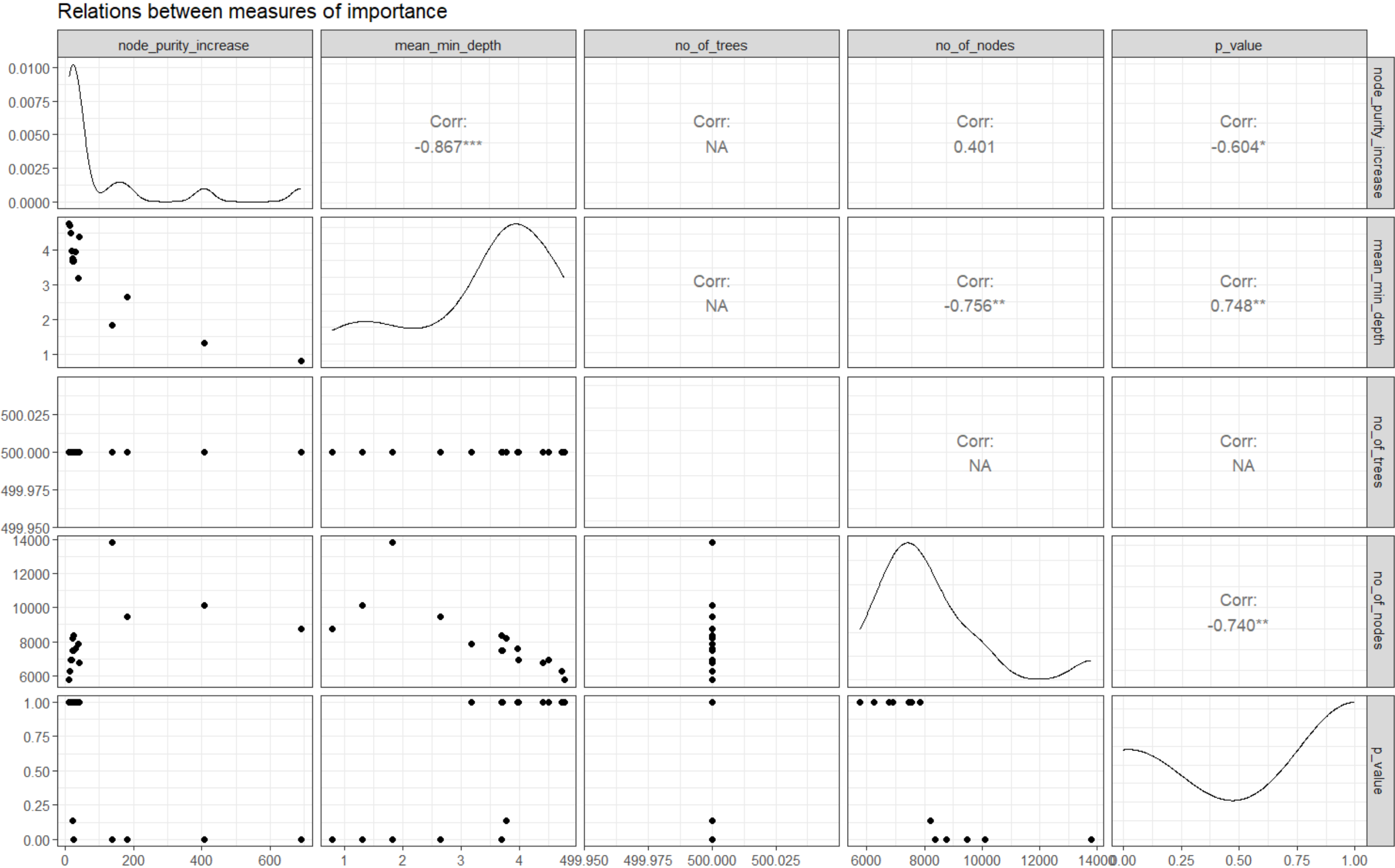
Protocolo 1 – Sem GHS e CCI (003)



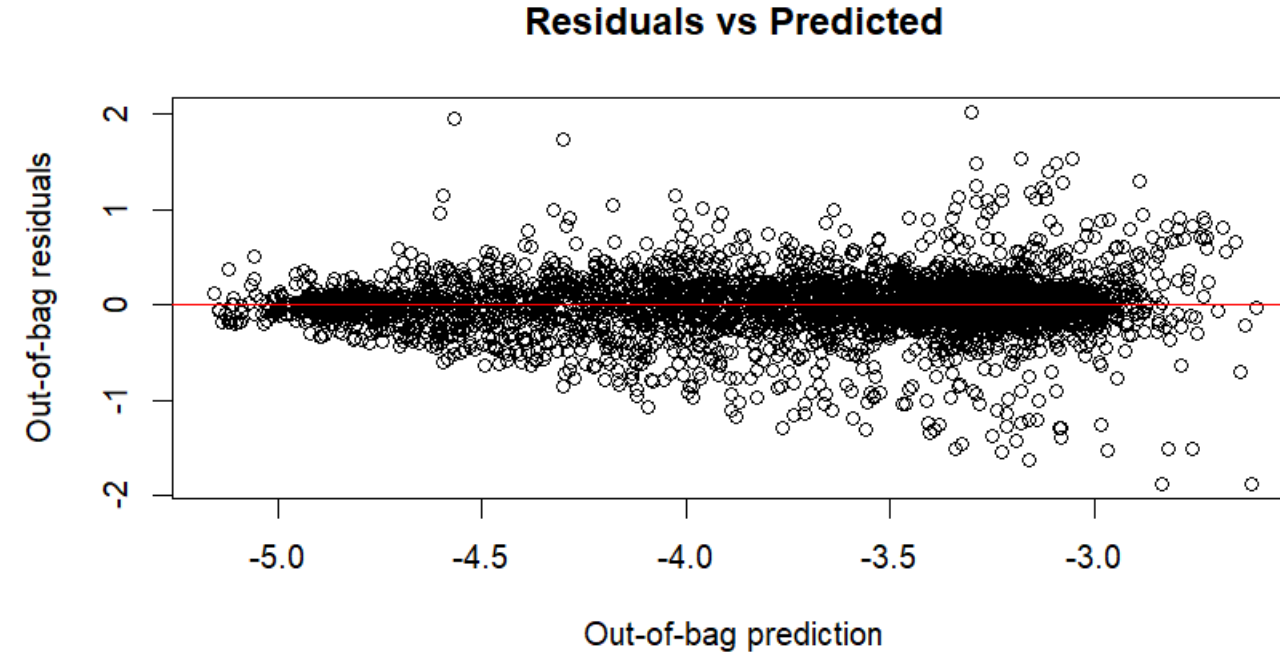
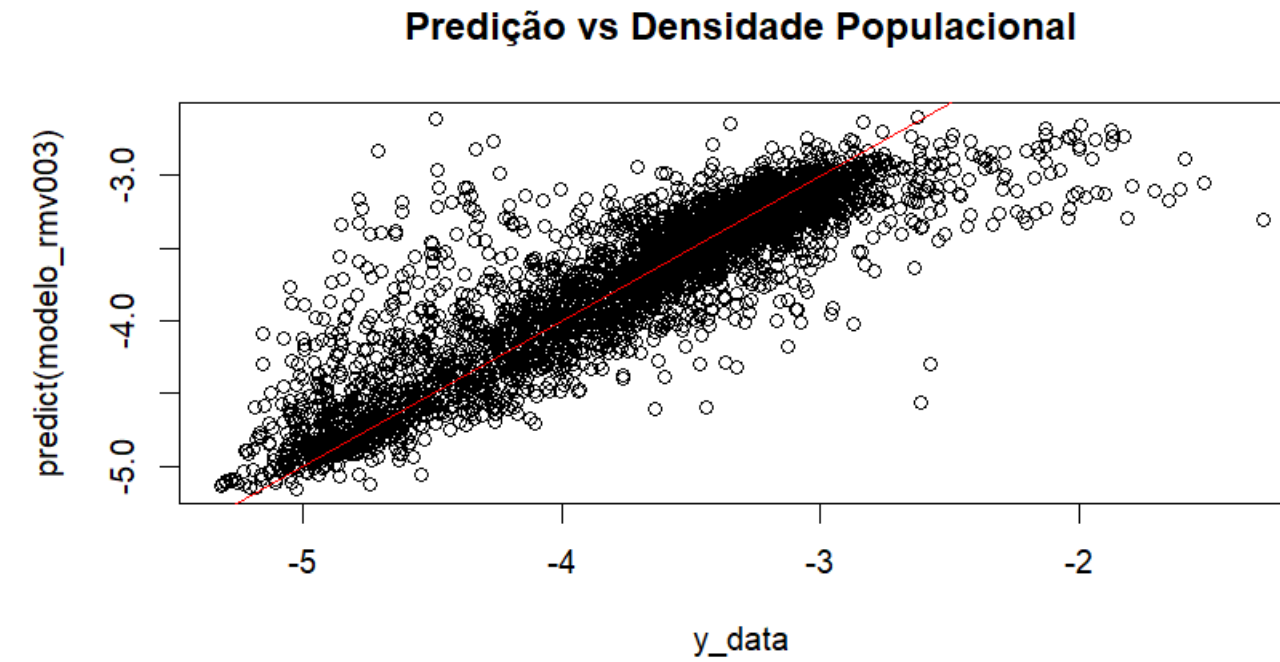
Protocolo 1 - Sem GHS e CCI (003)



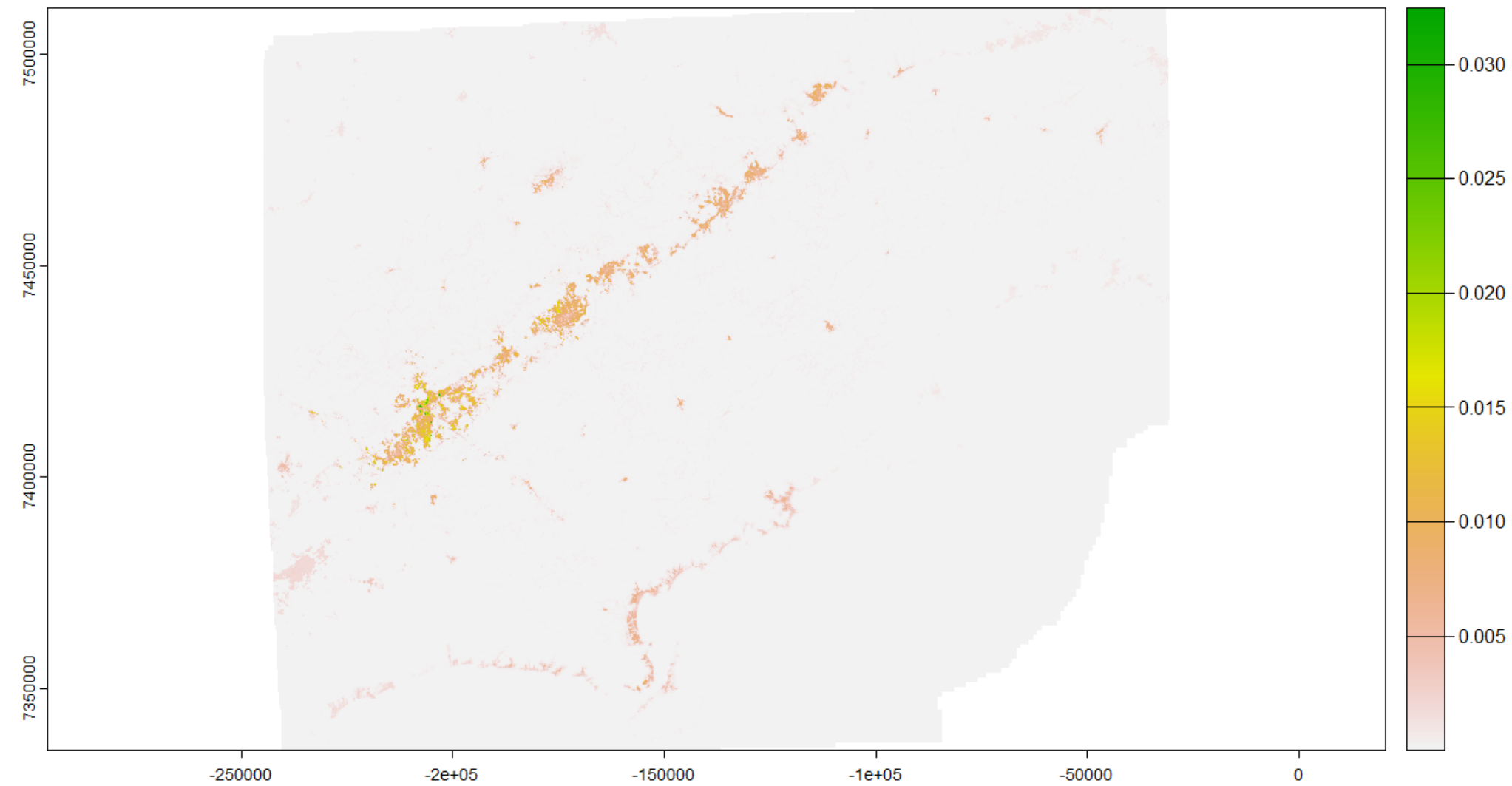
Protocolo 1 - Sem GHS e CCI (003)



Protocolo 1 - Sem GHS e CCI (003)

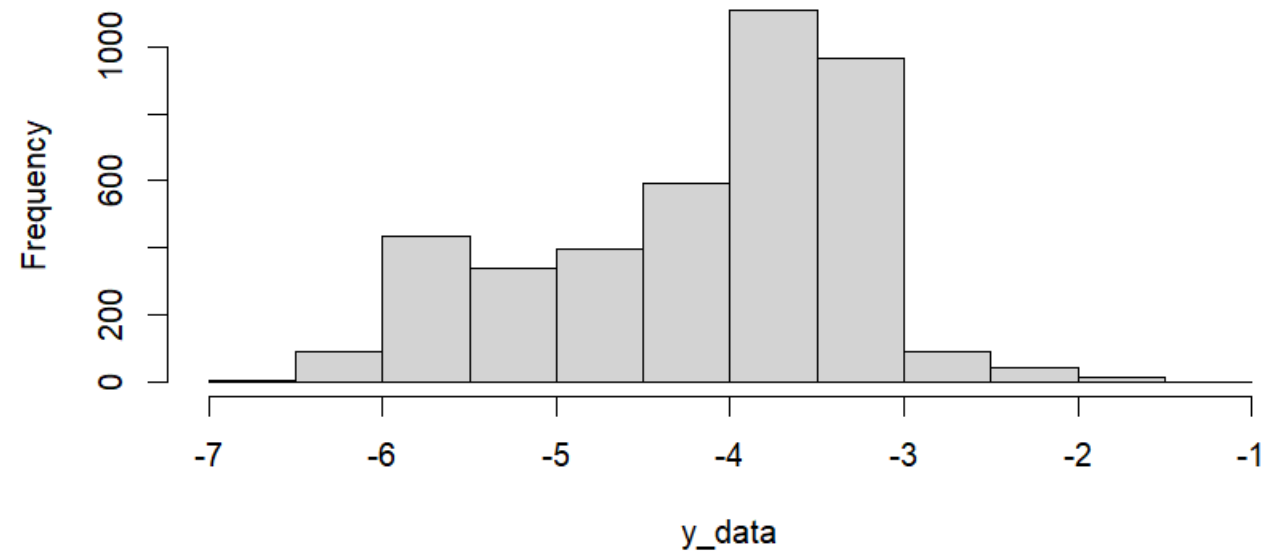


Protocolo 1 - Sem GHS e CCI (003)

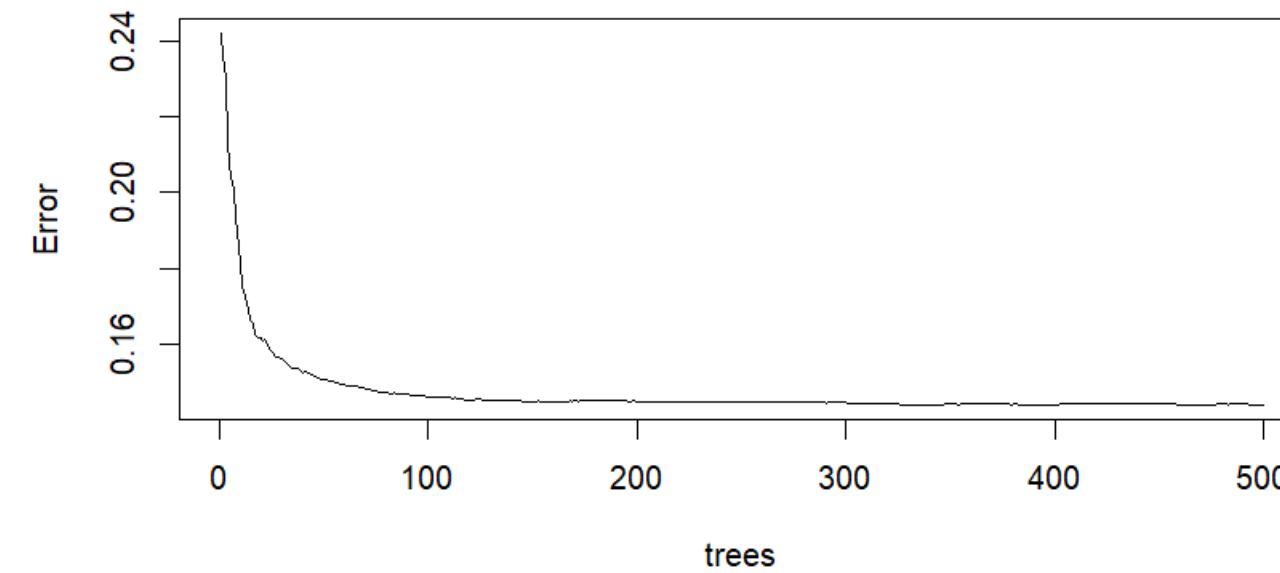


Protocolo 2 - Sem CCI (004)

Histogram of y_data



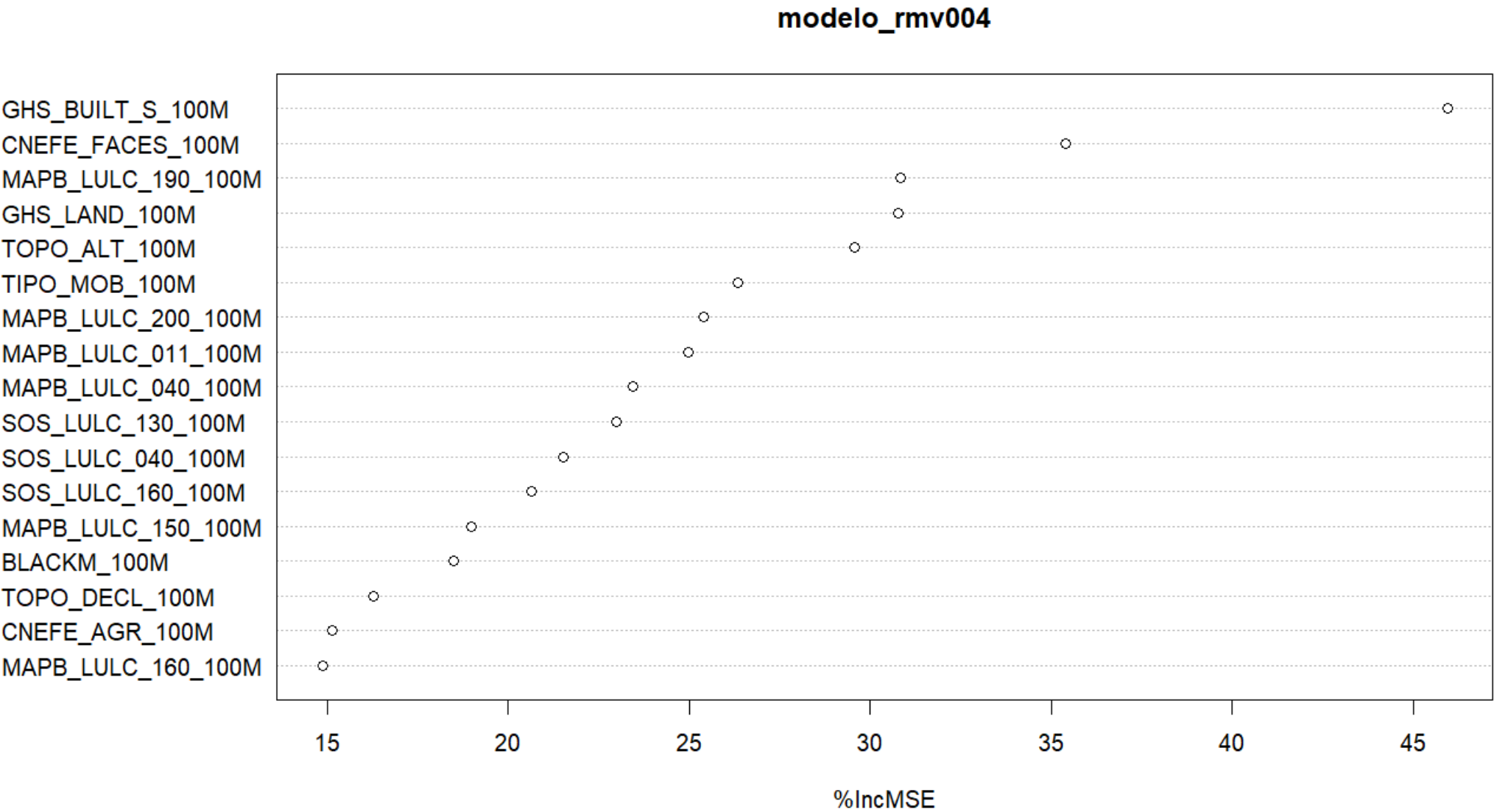
modelo_rmv004



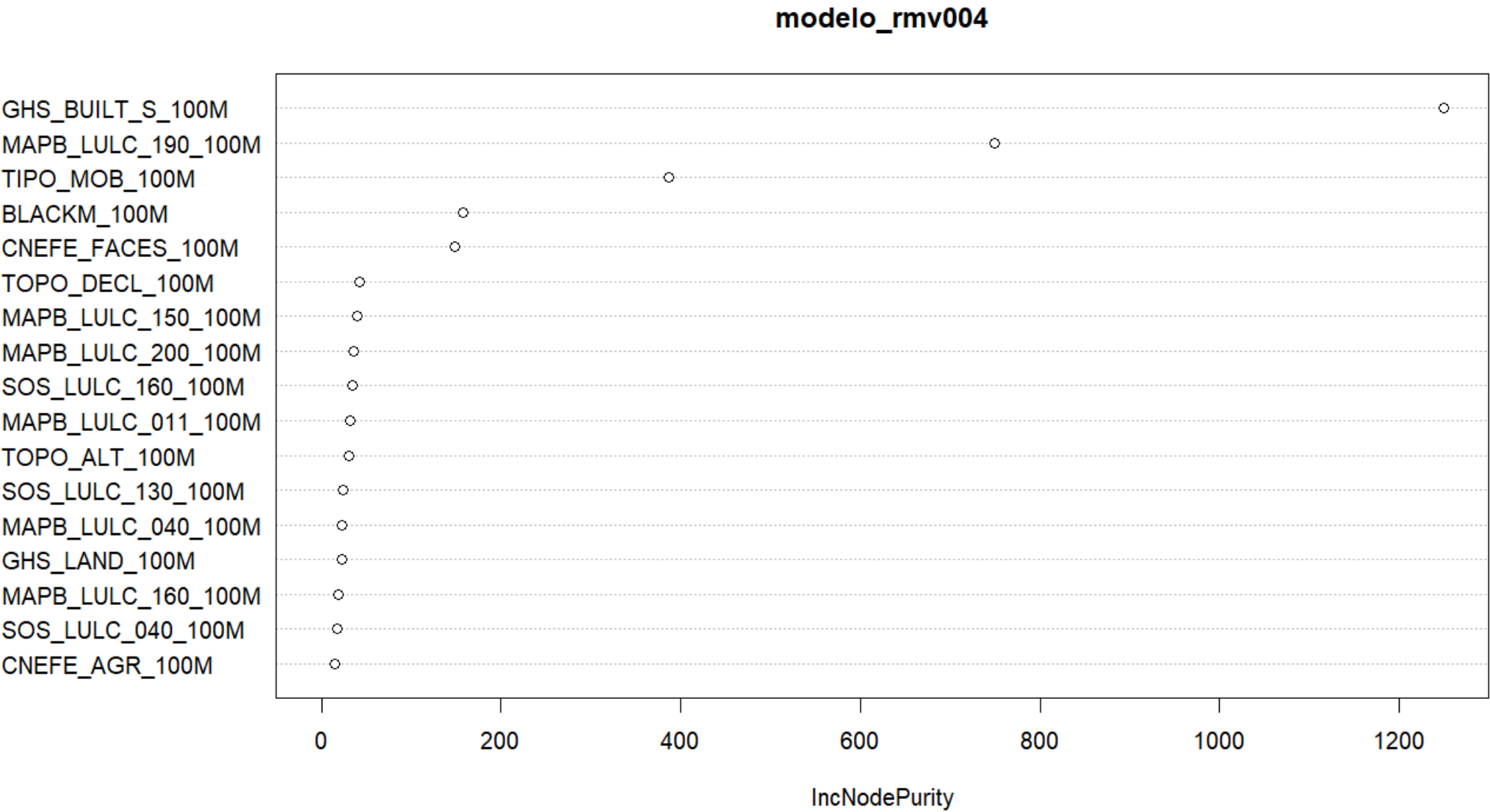
```
Call:
randomForest(x = x, y = y, mtry = res[which.min(res[, 2]), 1],      replace = TRUE, sampsize = ..5,
nodesize = ..1, importance = TRUE,      proximity = TRUE, na.action = ..2)
      Type of random forest: regression
      Number of trees: 500
No. of variables tried at each split: 8

      Mean of squared residuals: 0.1441582
      % Var explained: 82.36
```

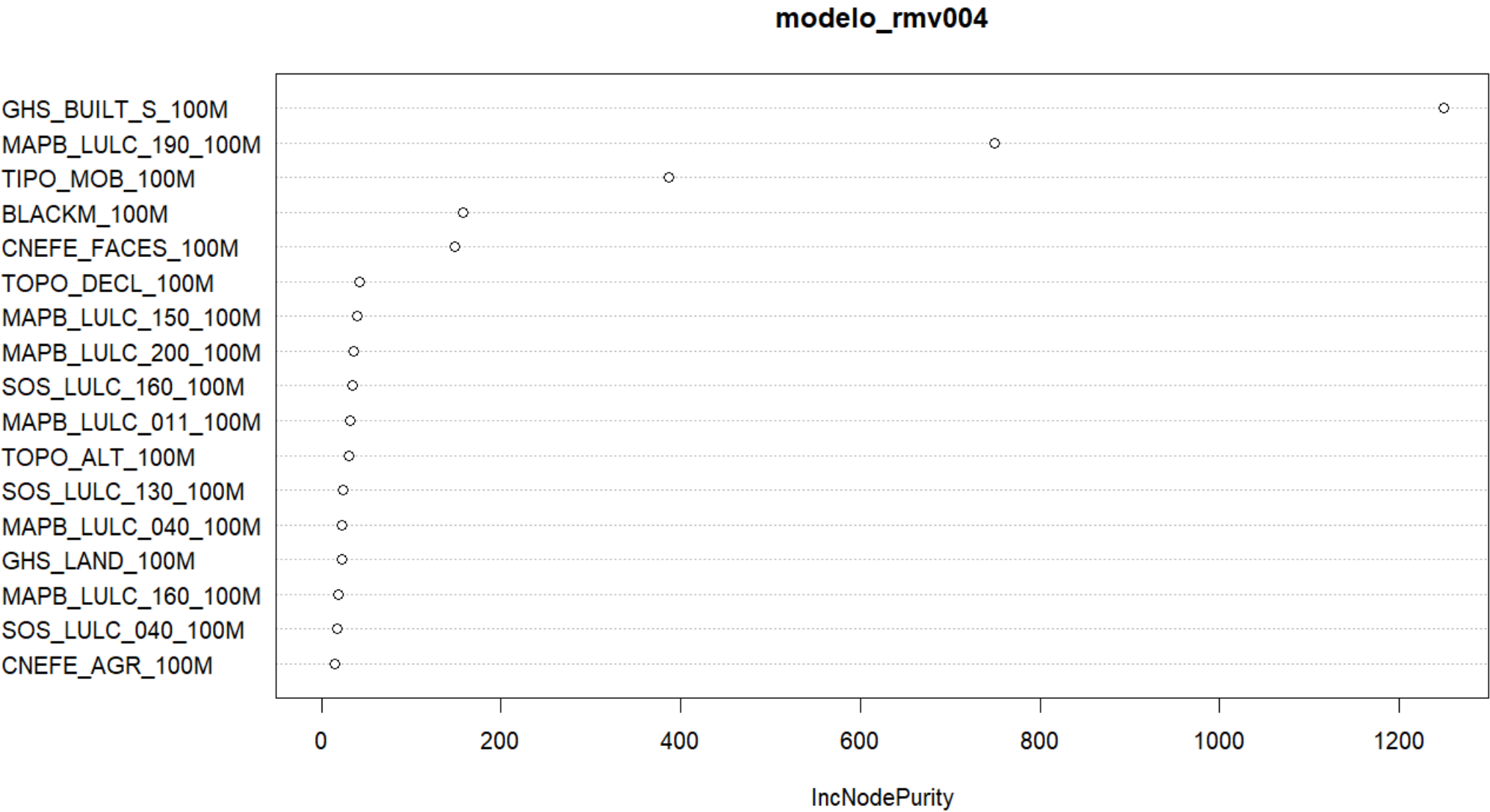
Protocolo 2 - Sem CCI (004)



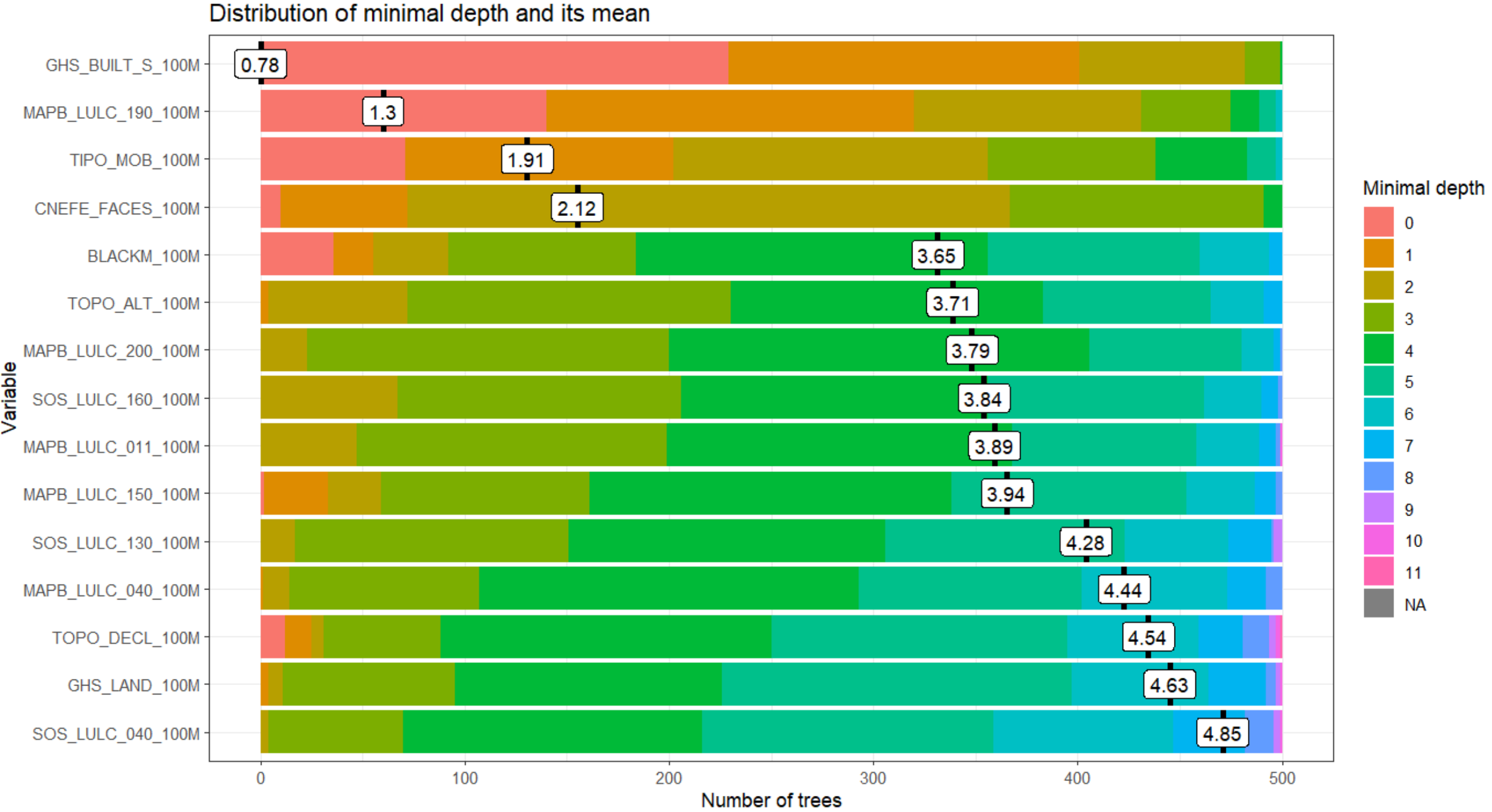
Protocolo 2 - Sem CCI (004)



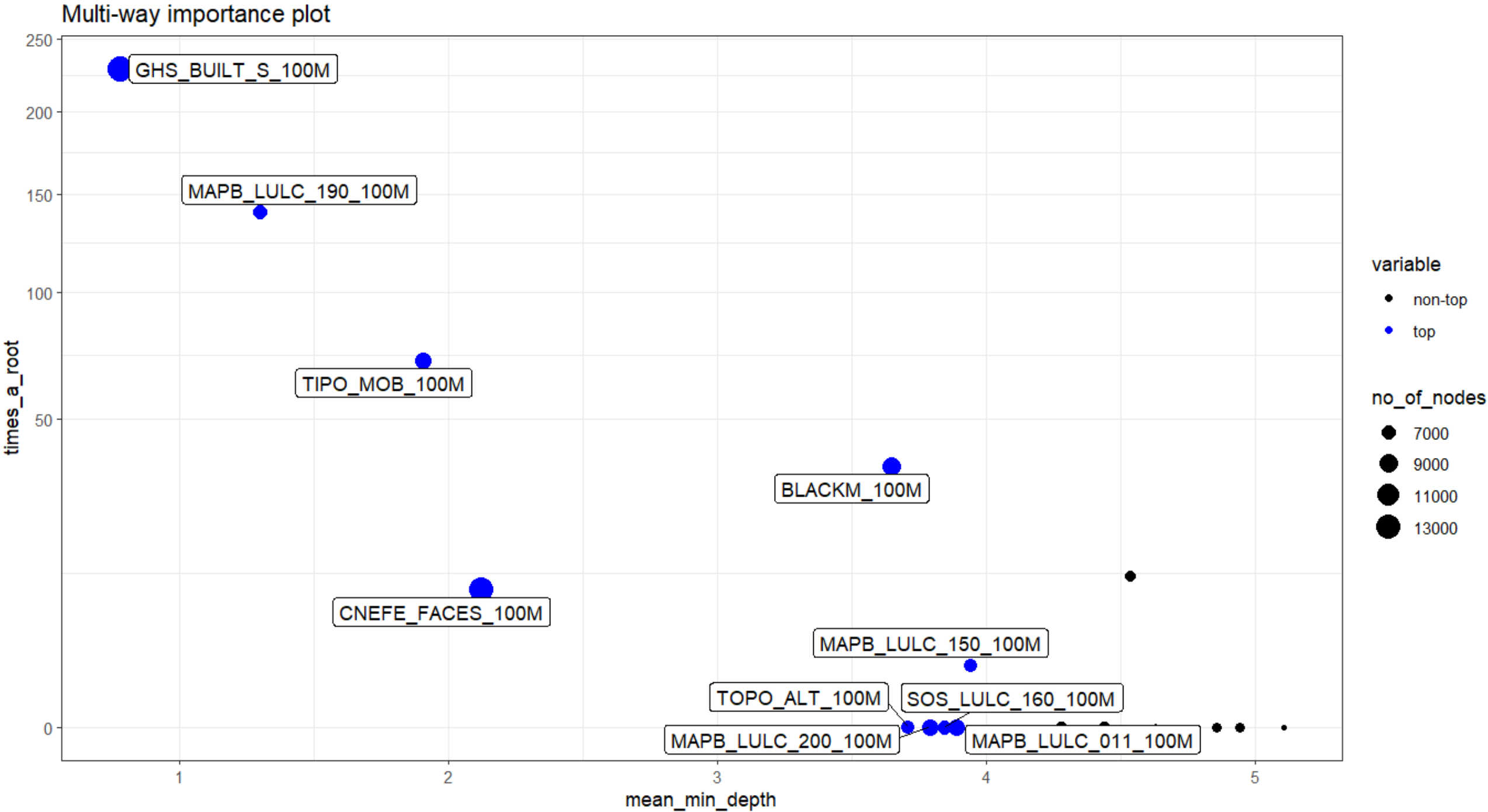
Protocolo 2 - Sem CCI (004)



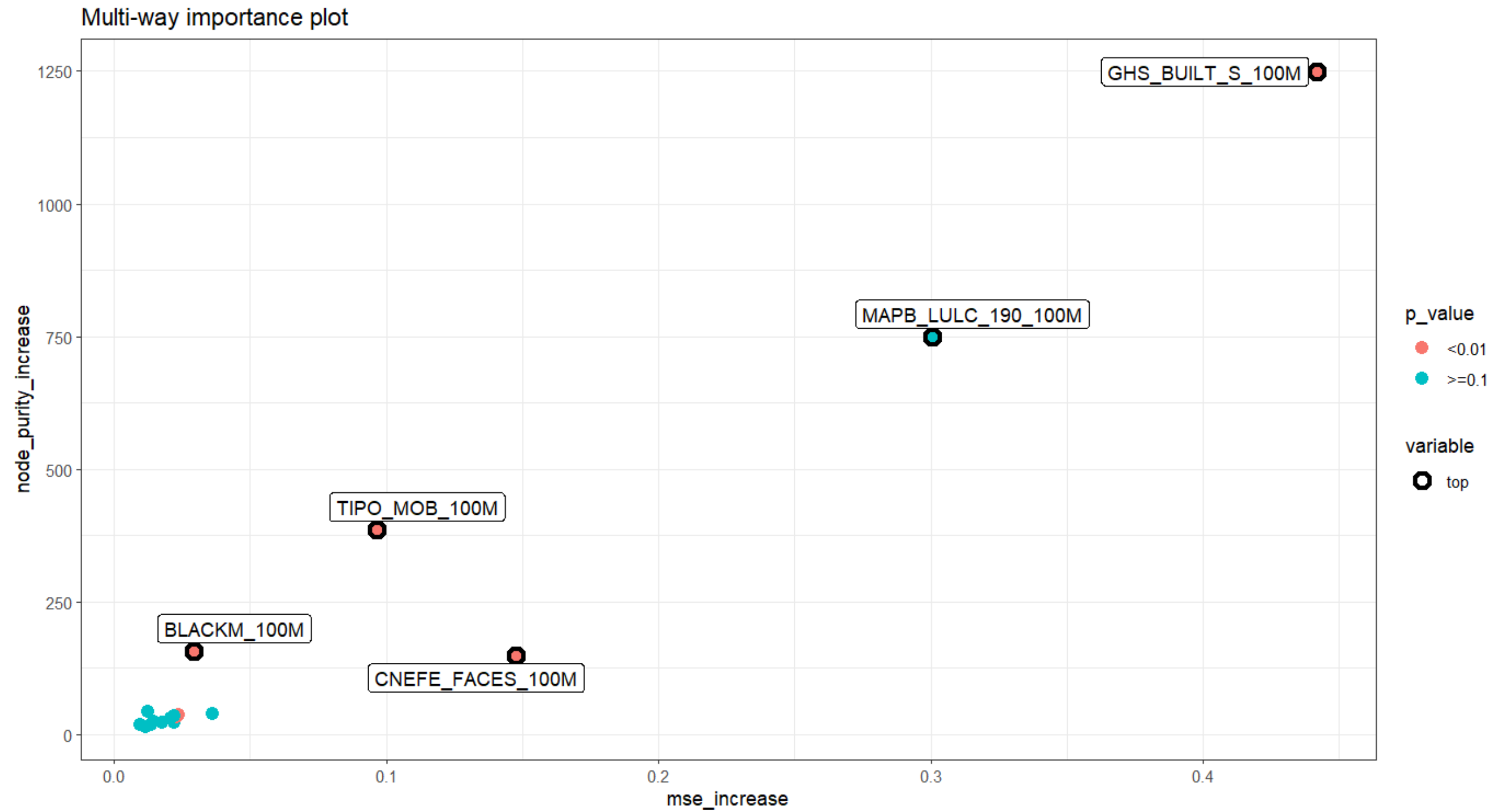
Protocolo 2 - Sem CCI (004)



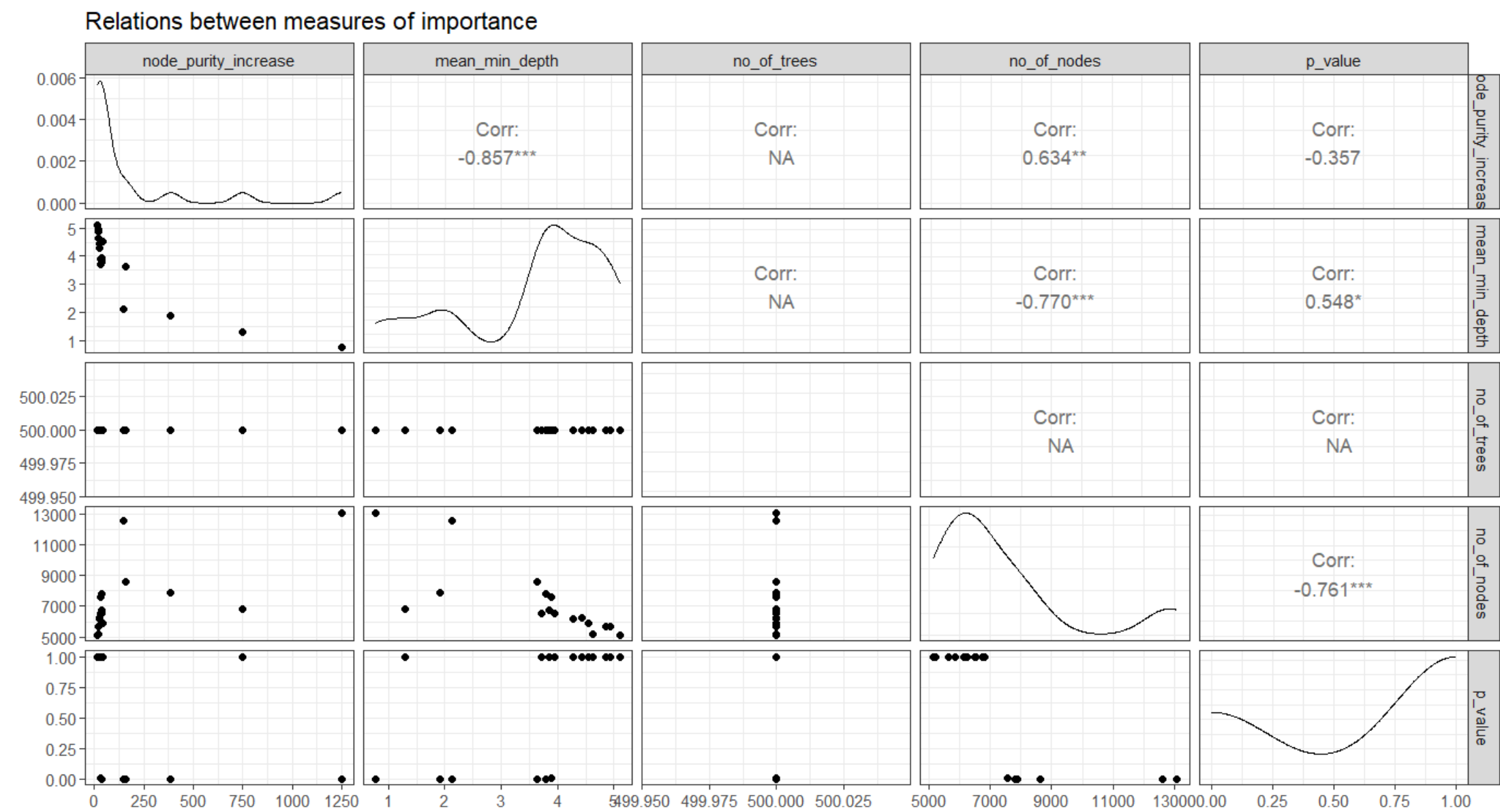
Protocolo 2 - Sem CCI (004)



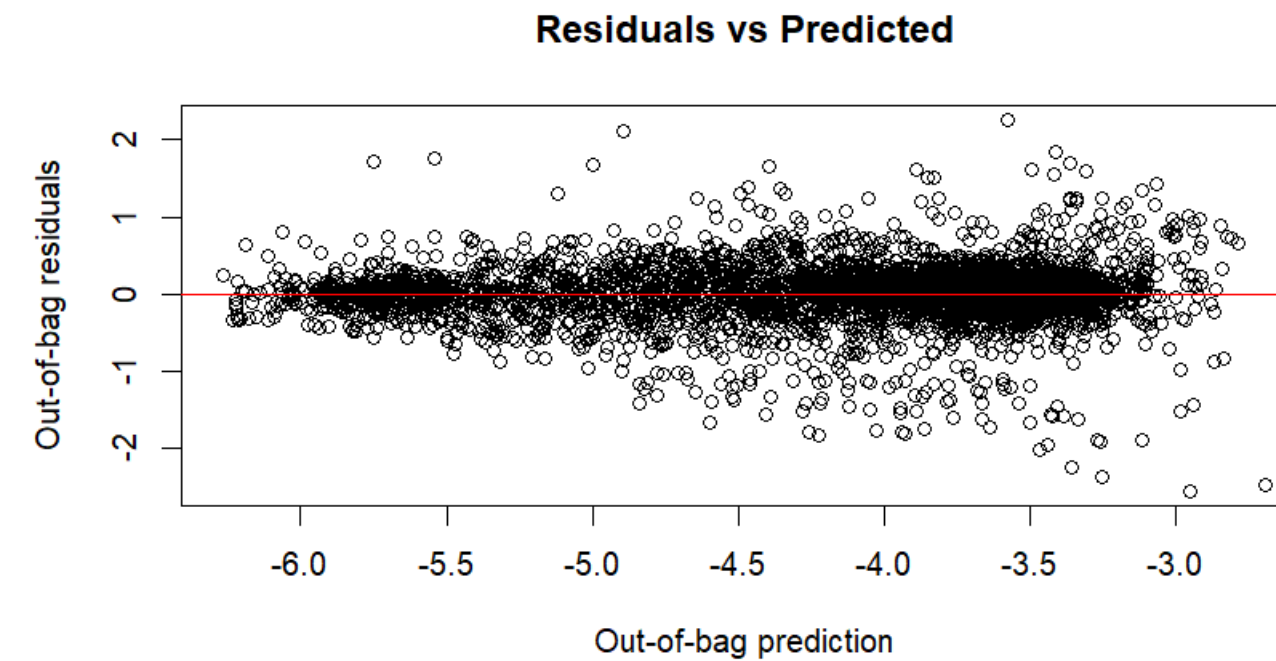
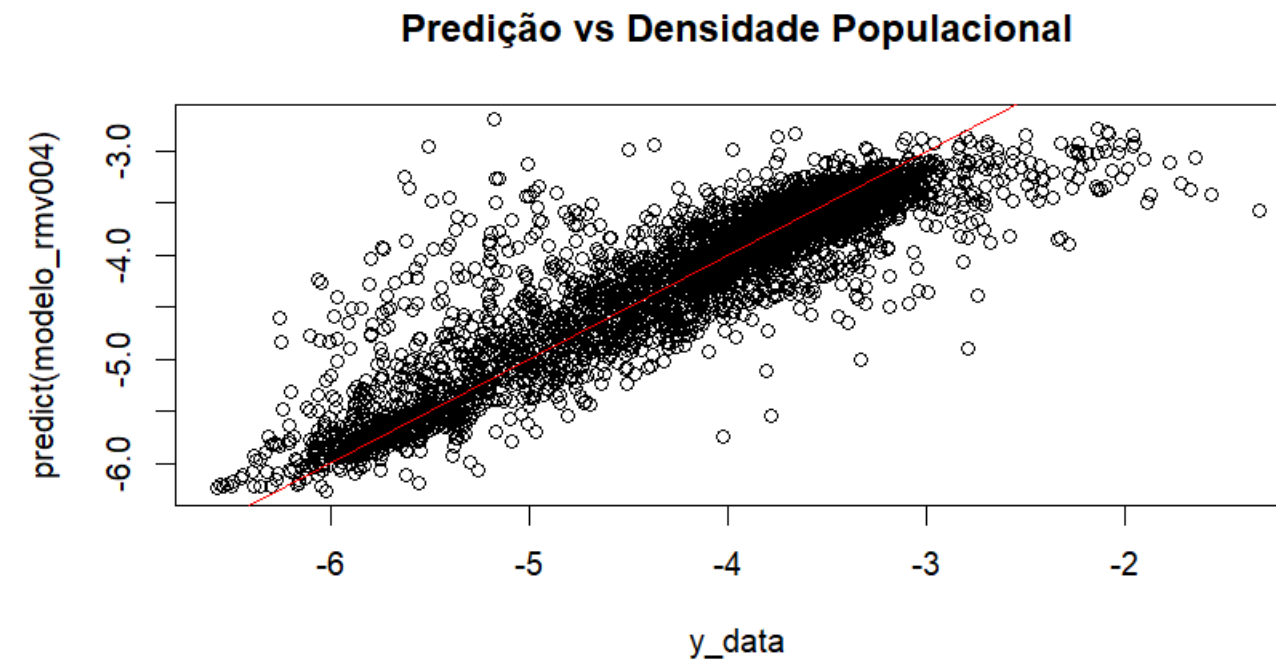
Protocolo 2 - Sem CCI (004)



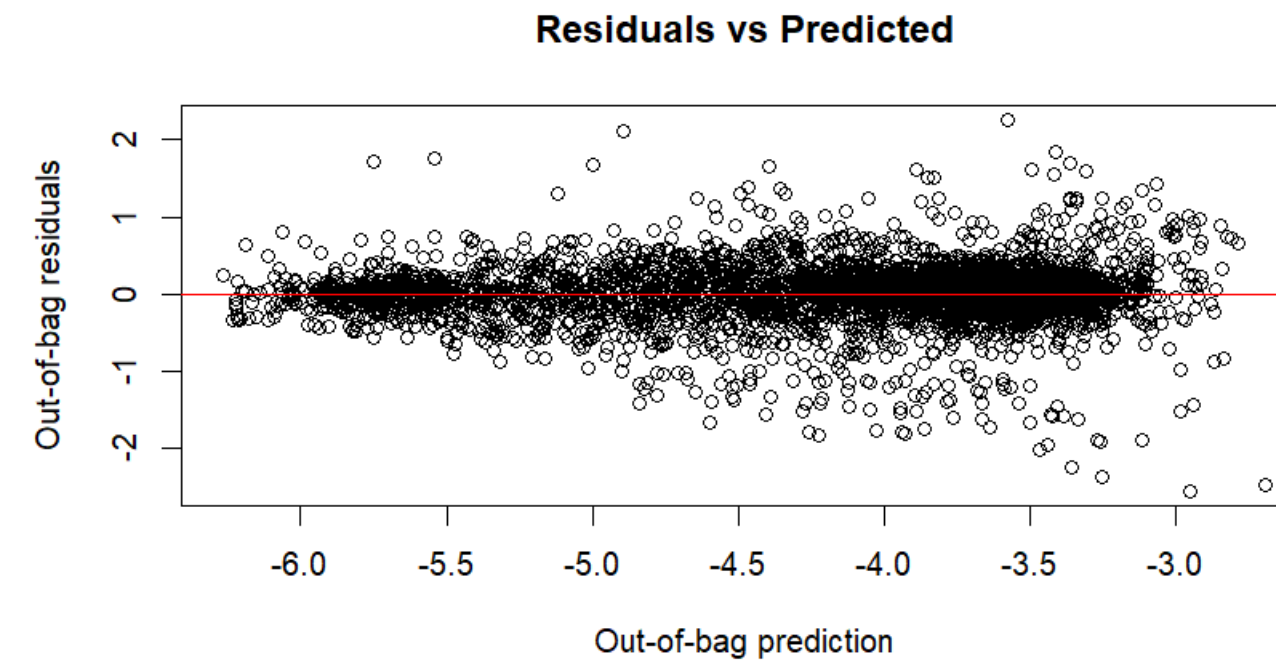
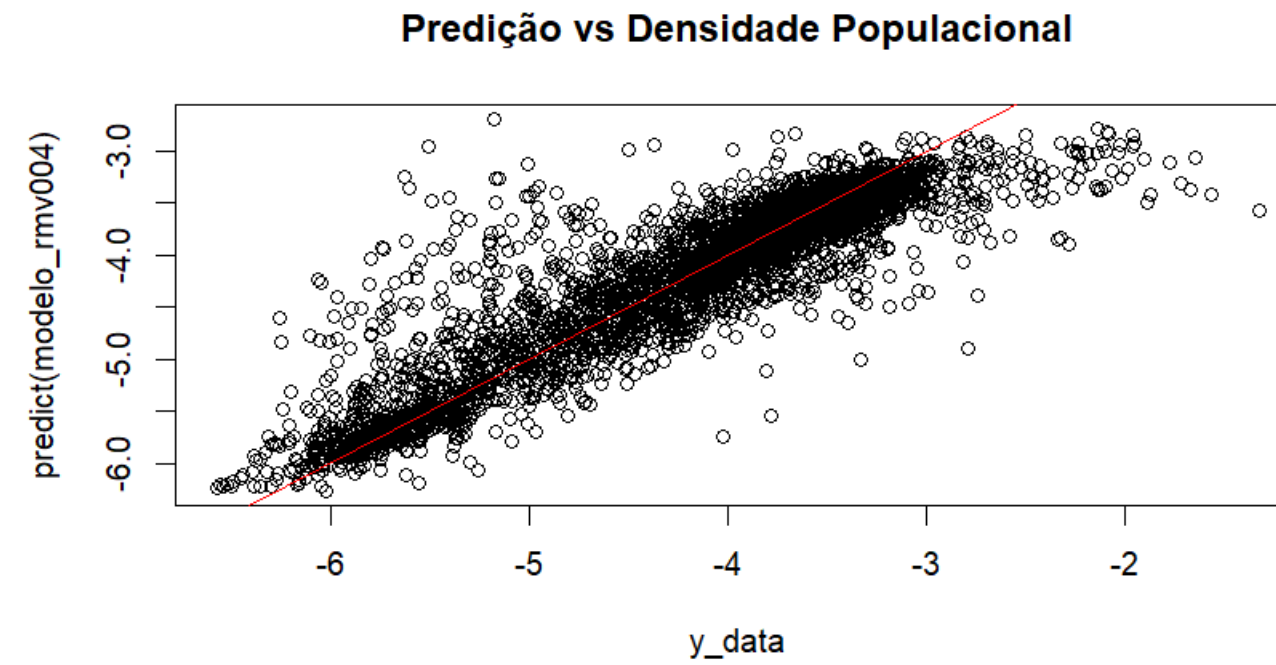
Protocolo 2 - Sem CCI (004)



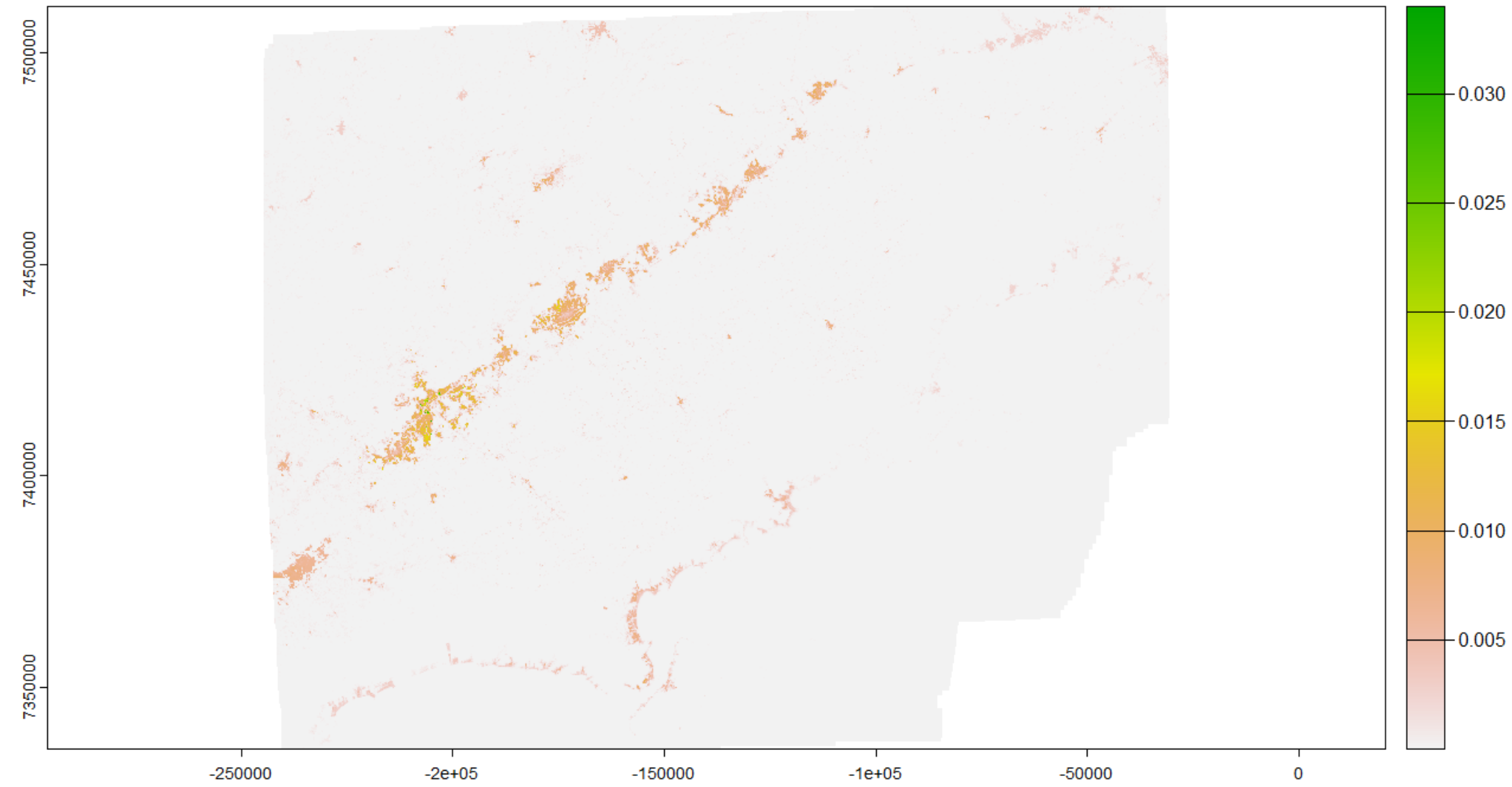
Protocolo 2 - Sem CCI (004)



Protocolo 2 - Sem CCI (004)

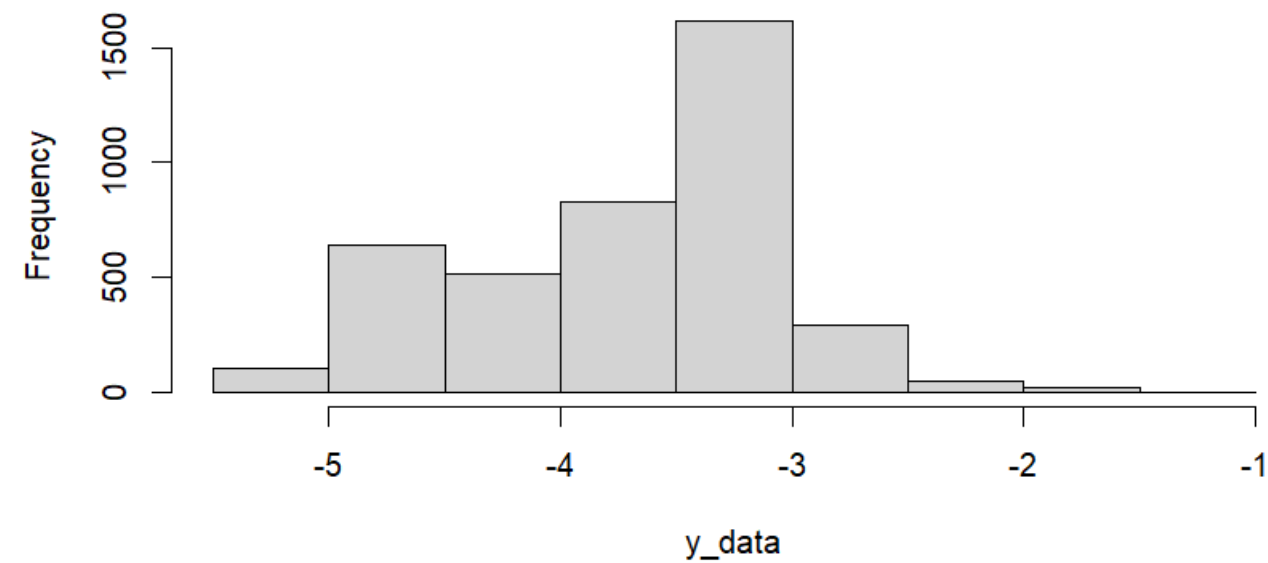


Protocolo 2 - Sem CCI (004)



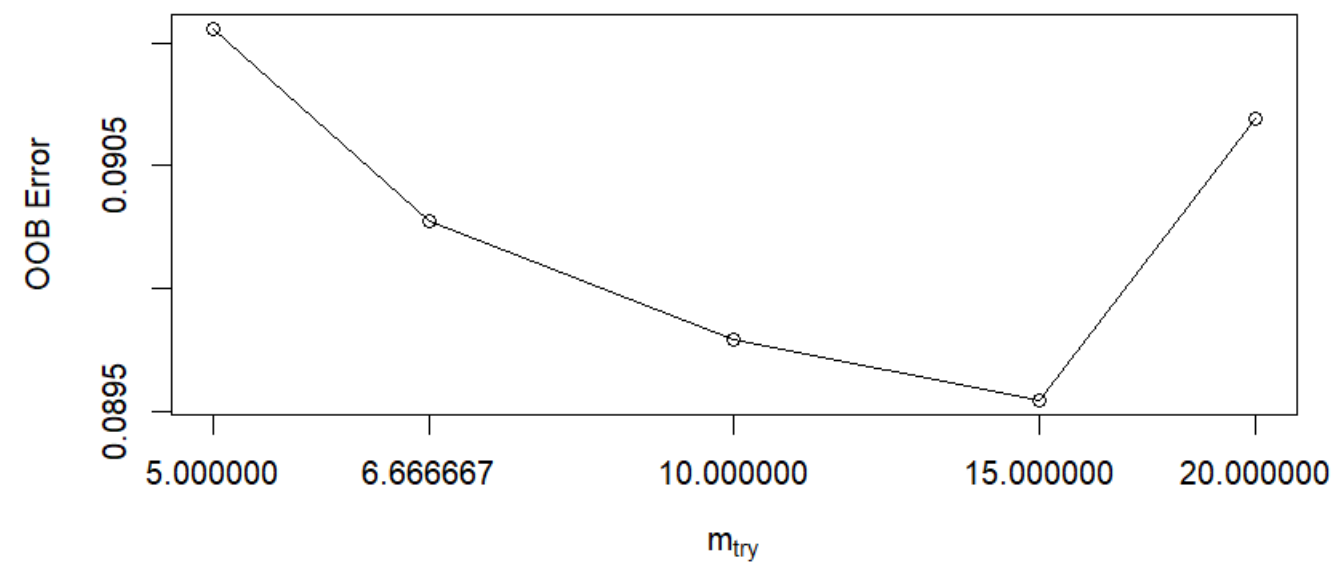
Protocolo 3 - Sem GHS (005)

Histogram of y_data

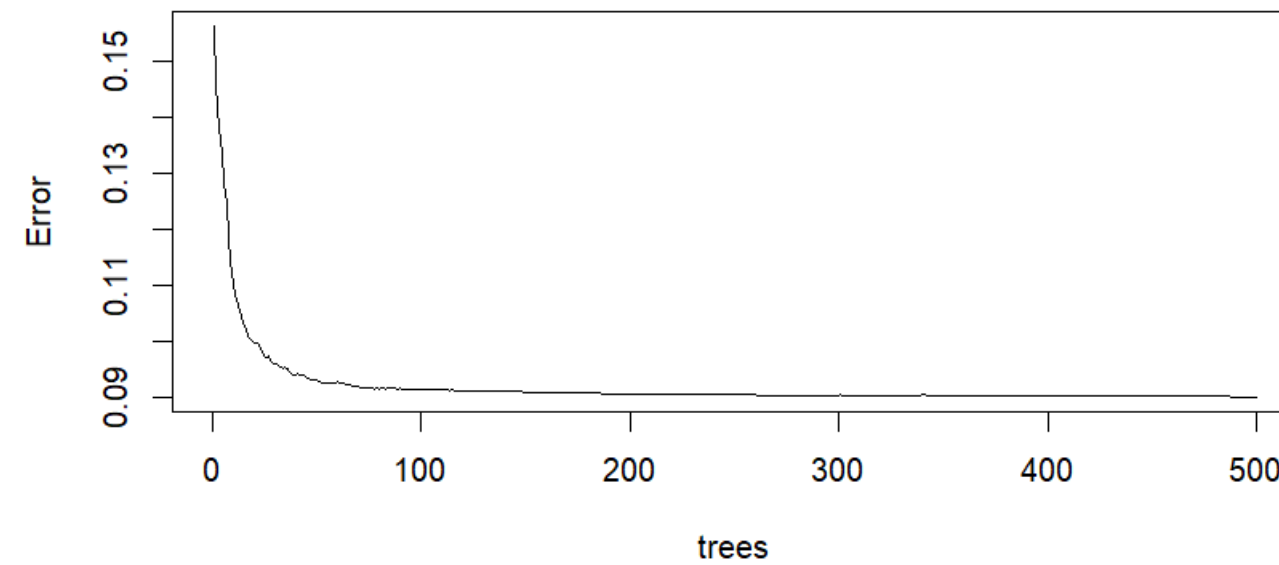


```
Call:
  randomForest(x = x, y = y, mtry = res[which.min(res[, 2]), 1],      replace = TRUE, sampsize = ..5,
    nodesize = ..1, importance = TRUE,      proximity = TRUE, na.action = ..2)
    Type of random forest: regression
    Number of trees: 500
No. of variables tried at each split: 15

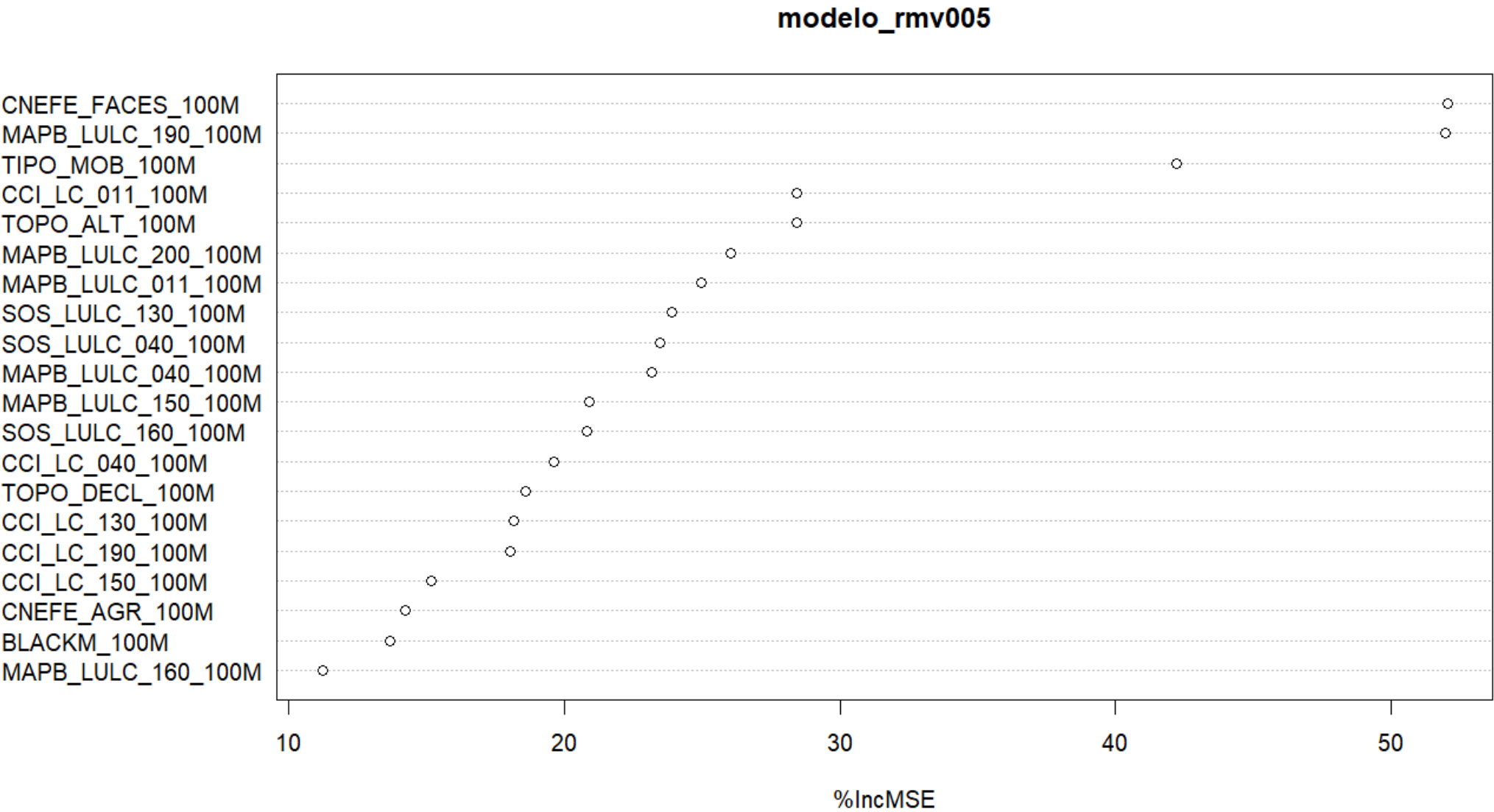
    Mean of squared residuals: 0.08999407
      % Var explained: 80.44
```



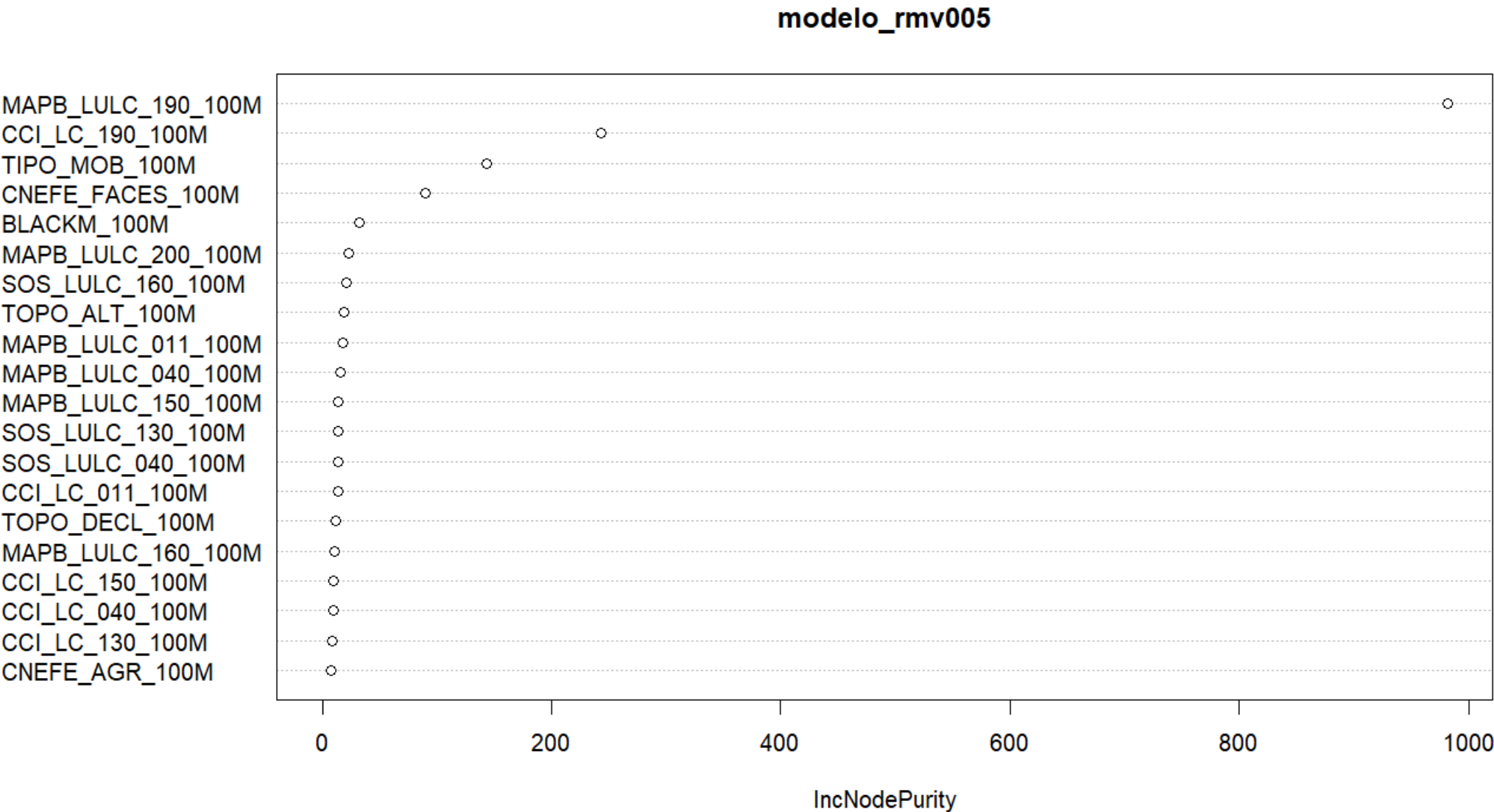
modelo_rmv005



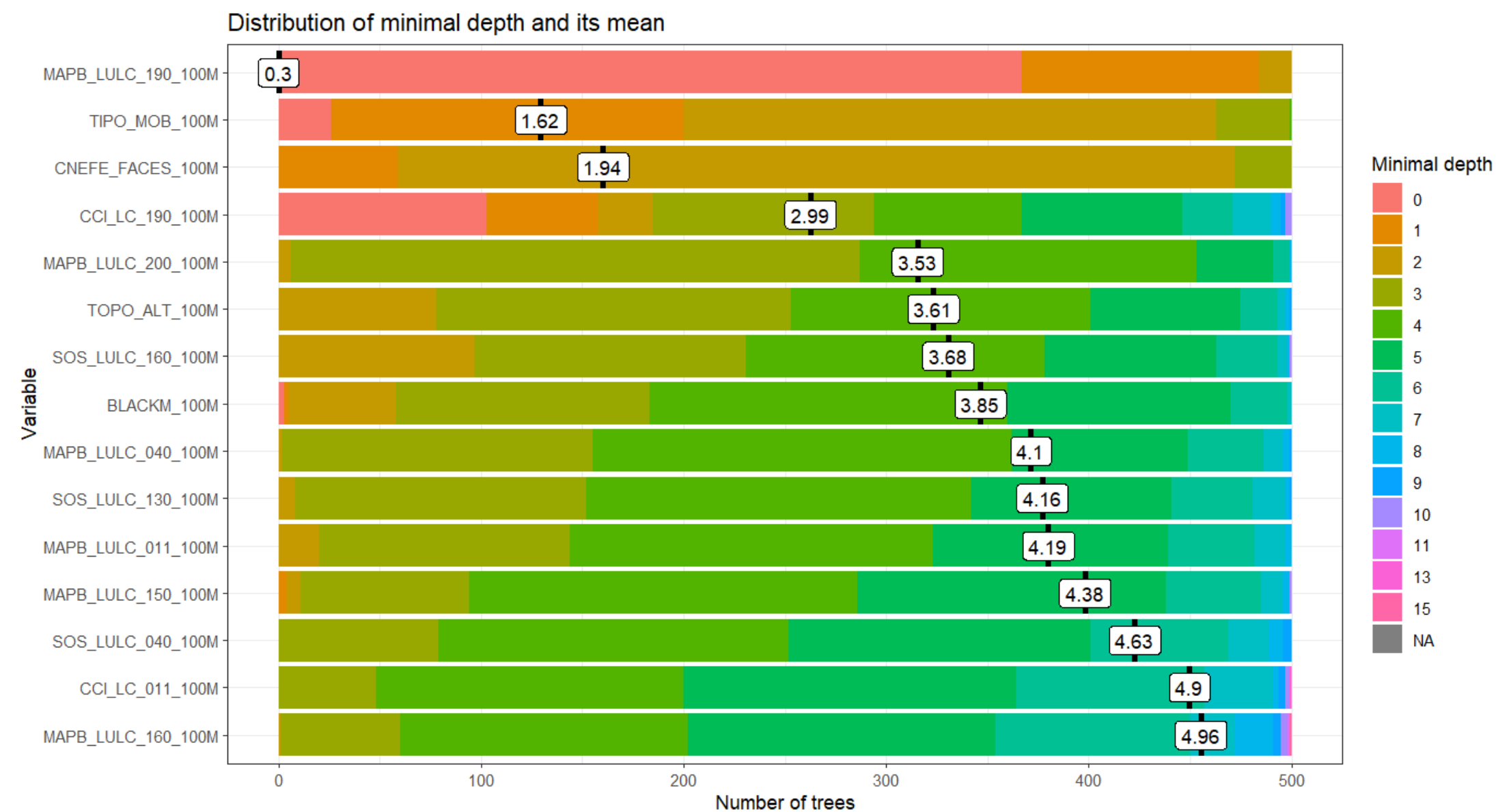
Protocolo 3 - Sem GHS (005)



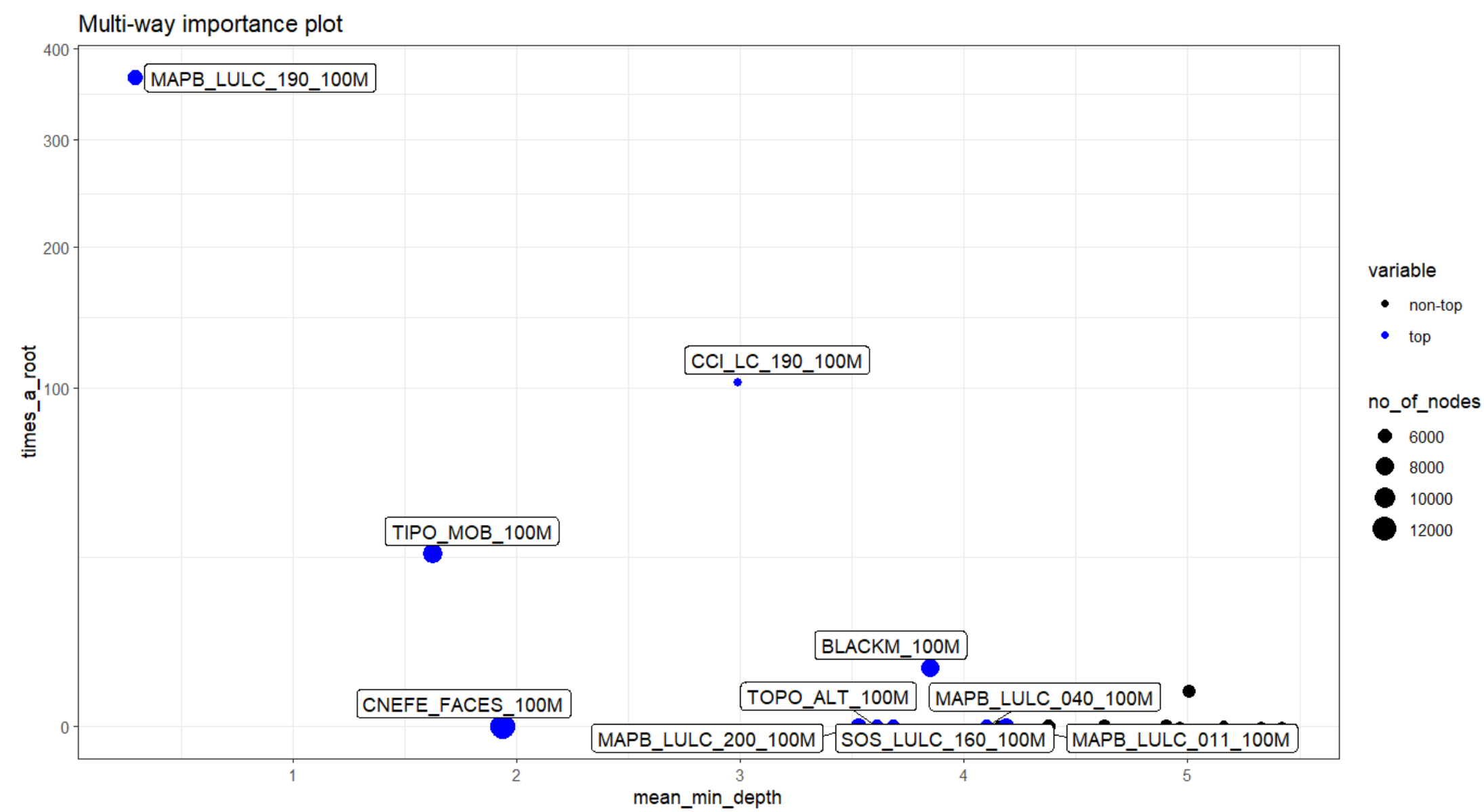
Protocolo 3 - Sem GHS (005)



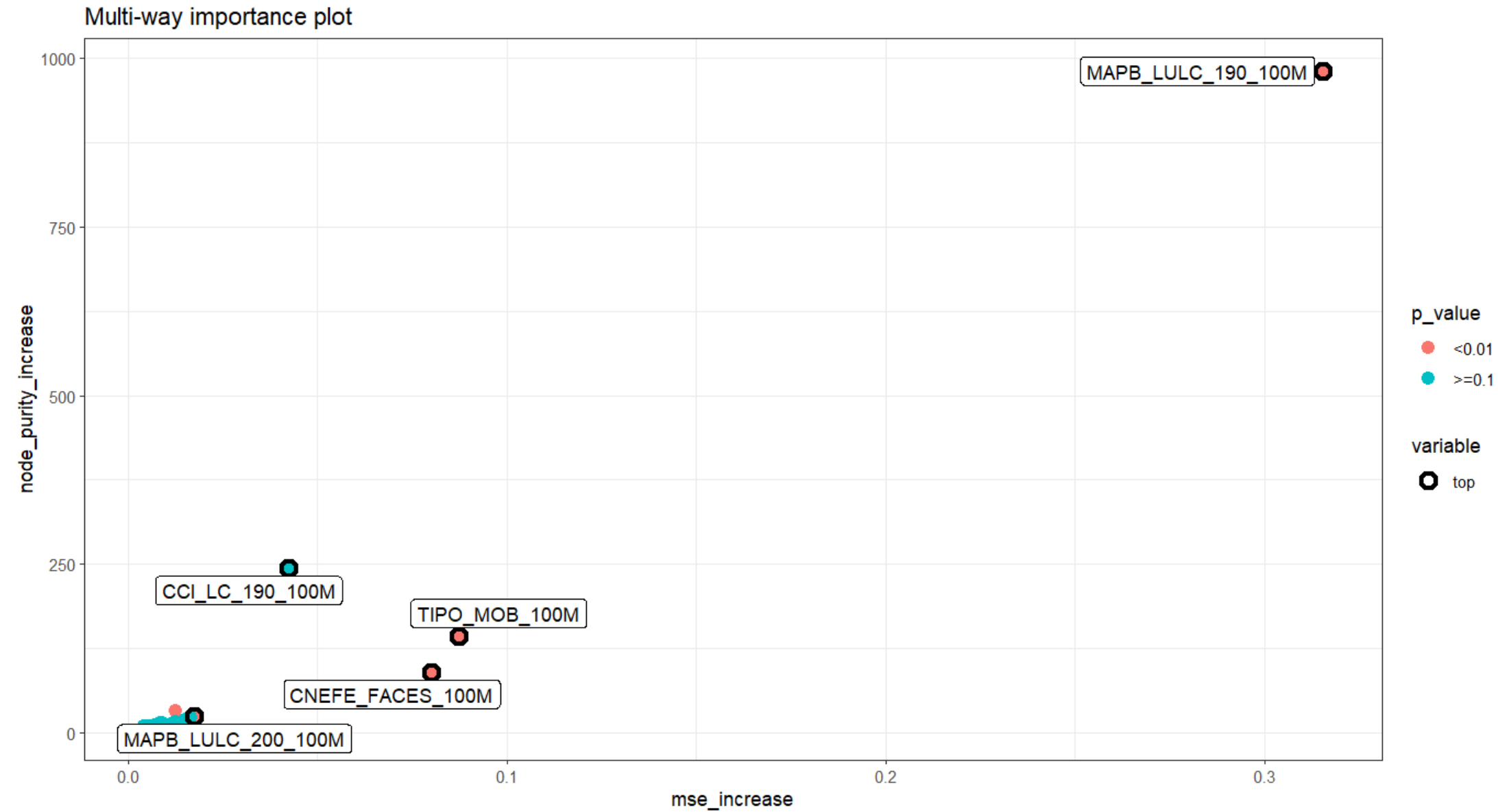
Protocolo 3 - Sem GHS (005)



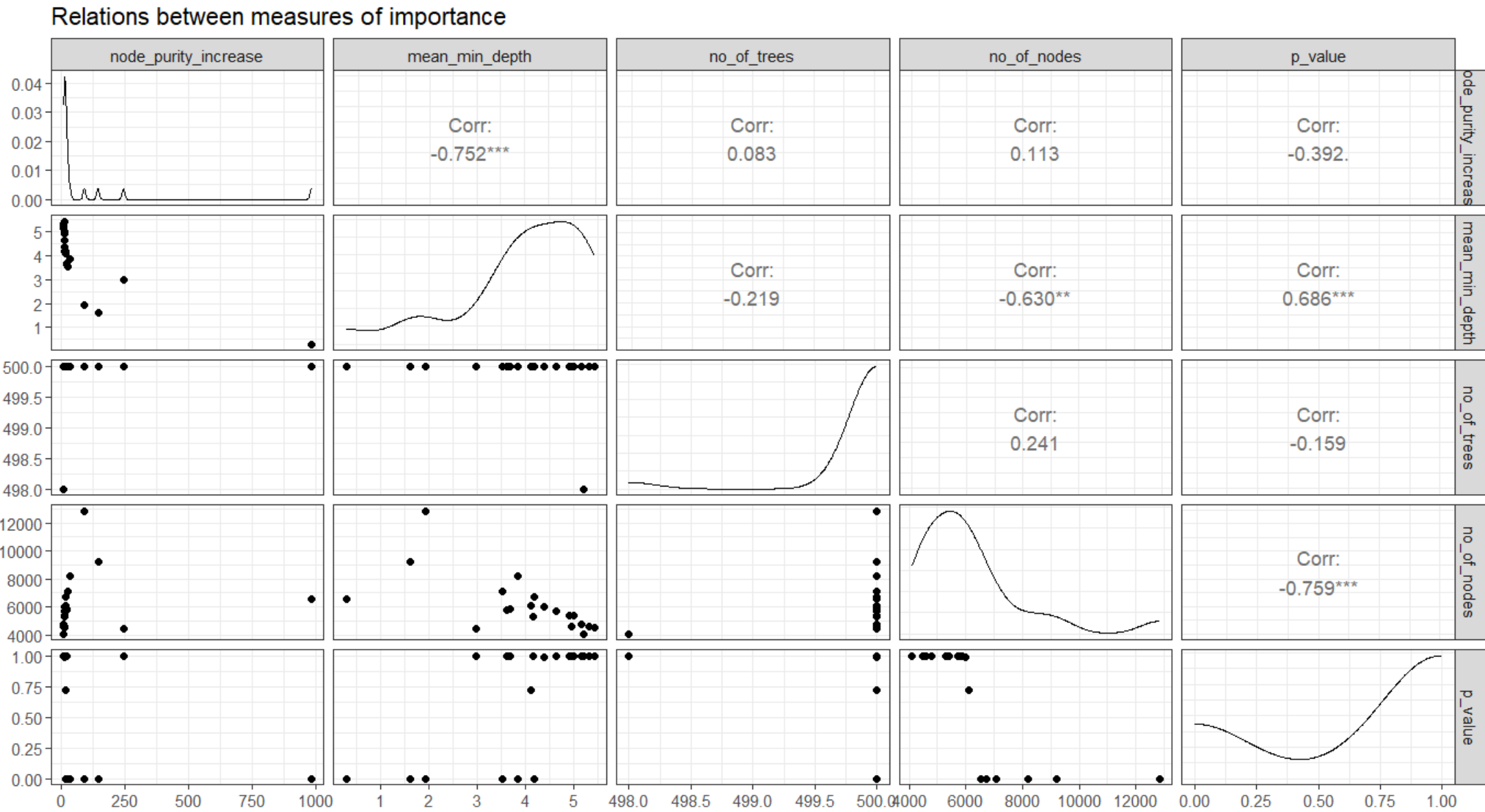
Protocolo 3 – Sem GHS (005)



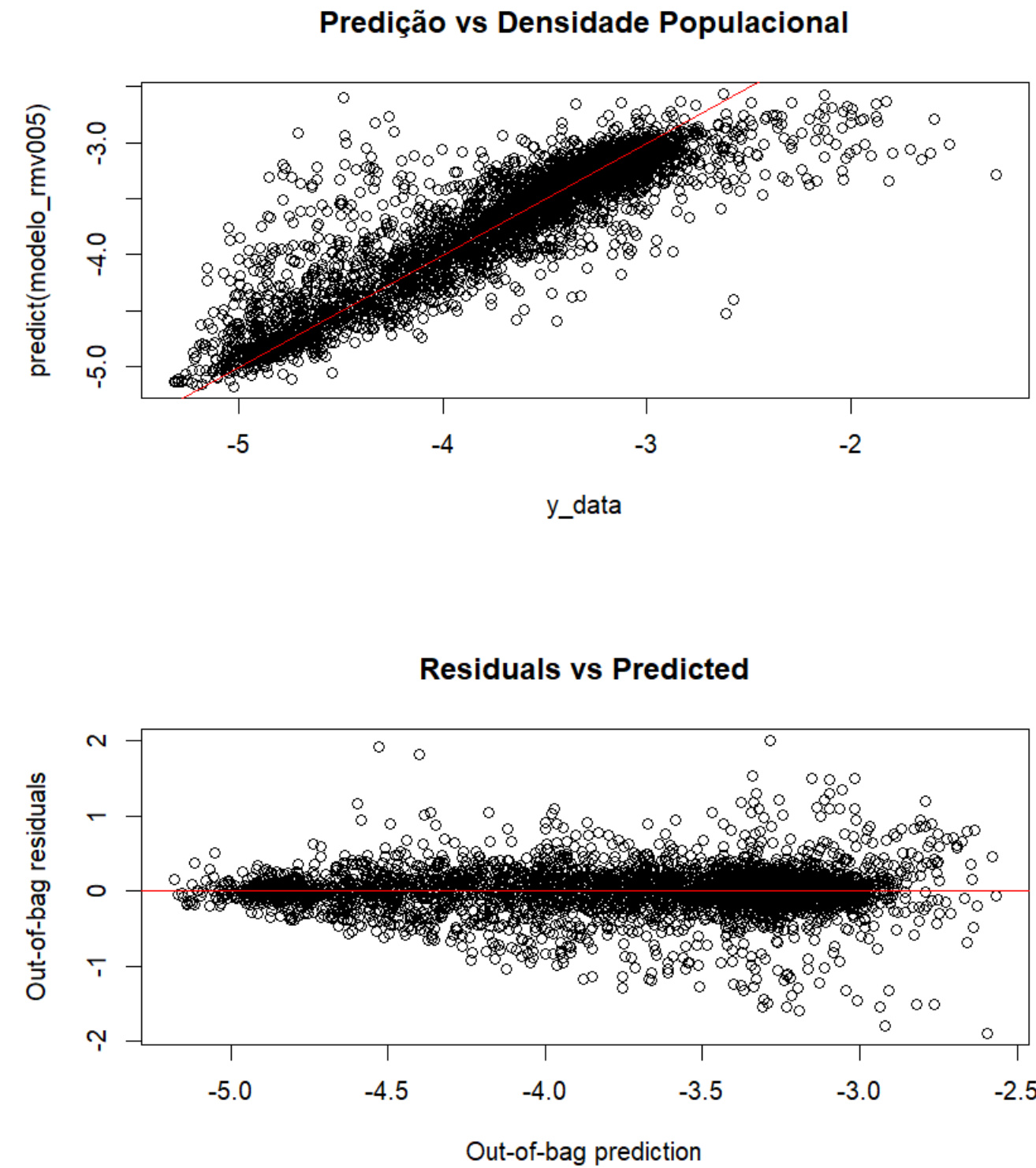
Protocolo 3 - Sem GHS (005)



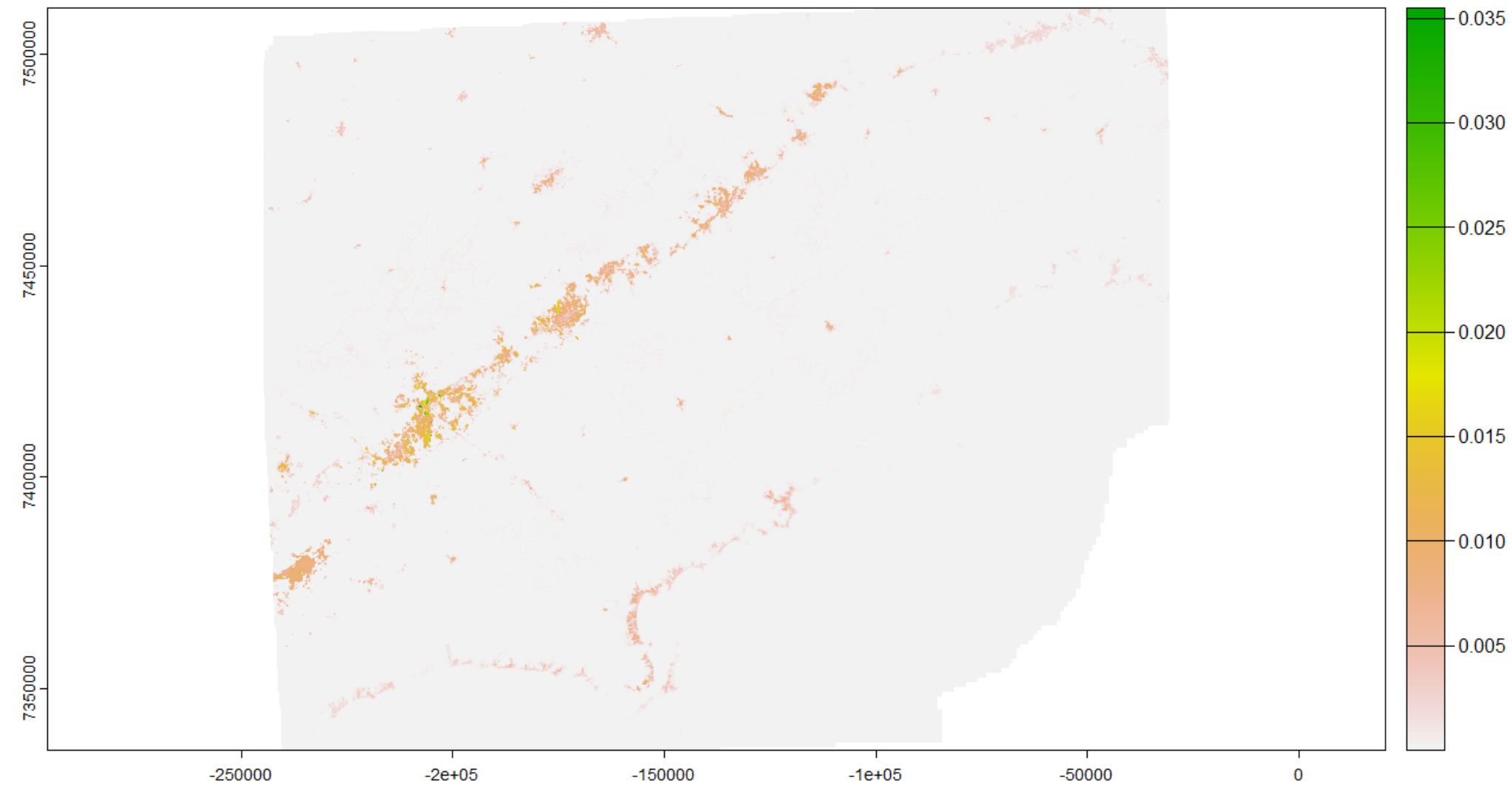
Protocolo 3 - Sem GHS (005)



Protocolo 3 - Sem GHS (005)

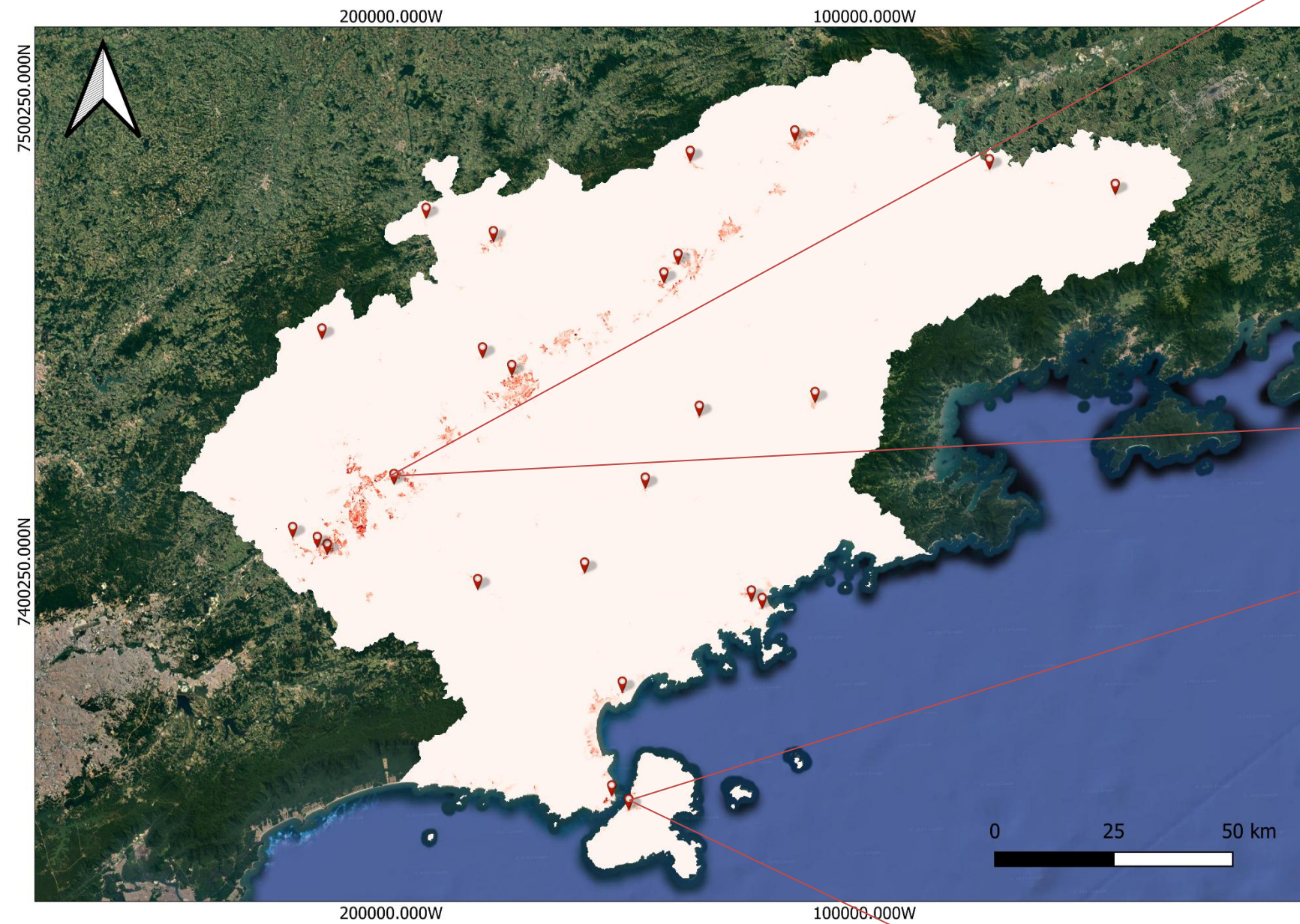


Protocolo 3 - Sem GHS (005)



```
dens.t <- (densidade ^ lambda - 1)/lambda  
matriz_pred <- ((1 + lambda * matriz_predicao)^(1/lambda))
```


Avaliação: distribuição populacional

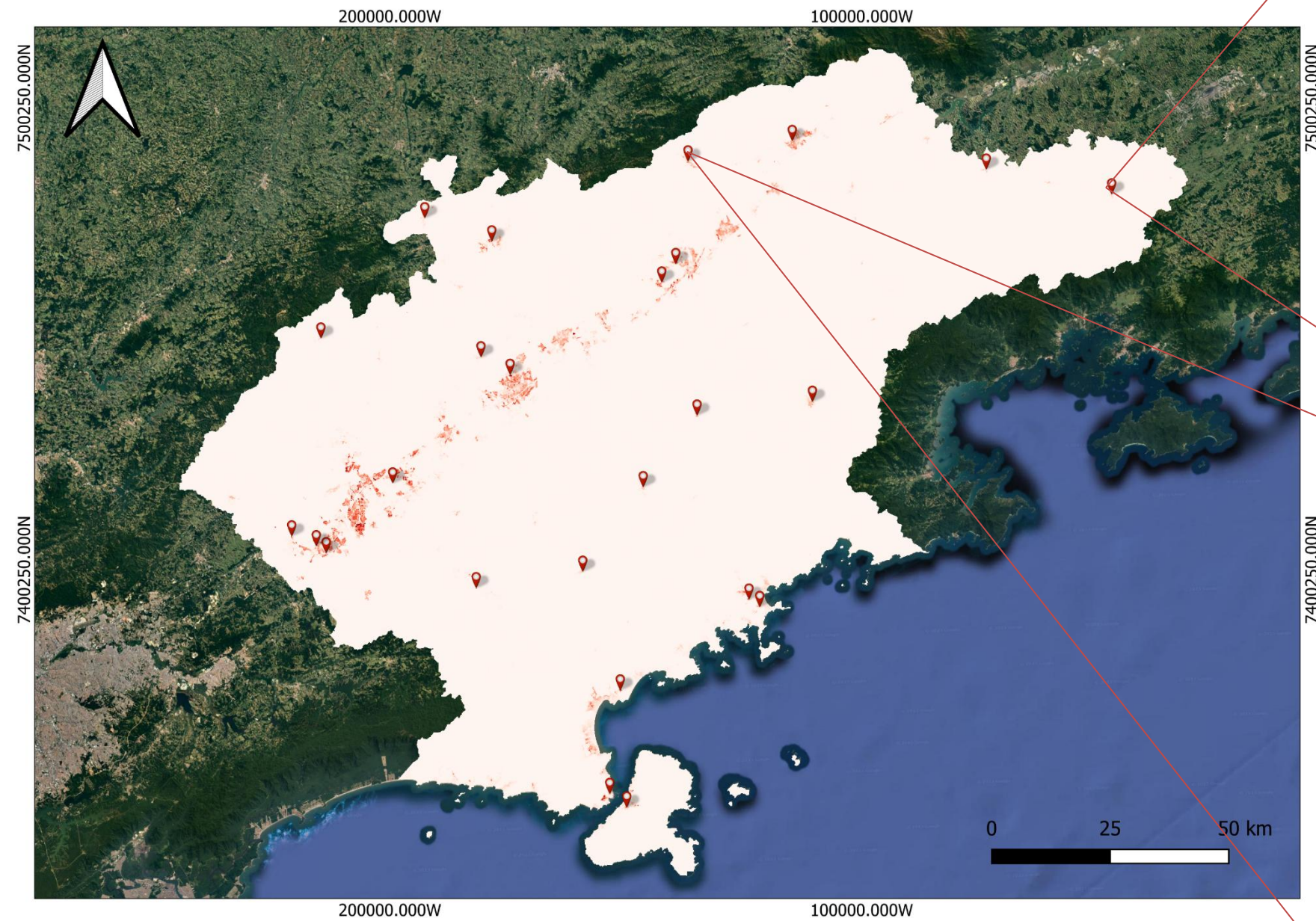


Revap



Ilhabela

Avaliação: distribuição populacional



Bananal



Piquete

Avaliação: Sub-região Guaratinguetá

Pq. Santa Clara - Guaratinguetá



Piquete



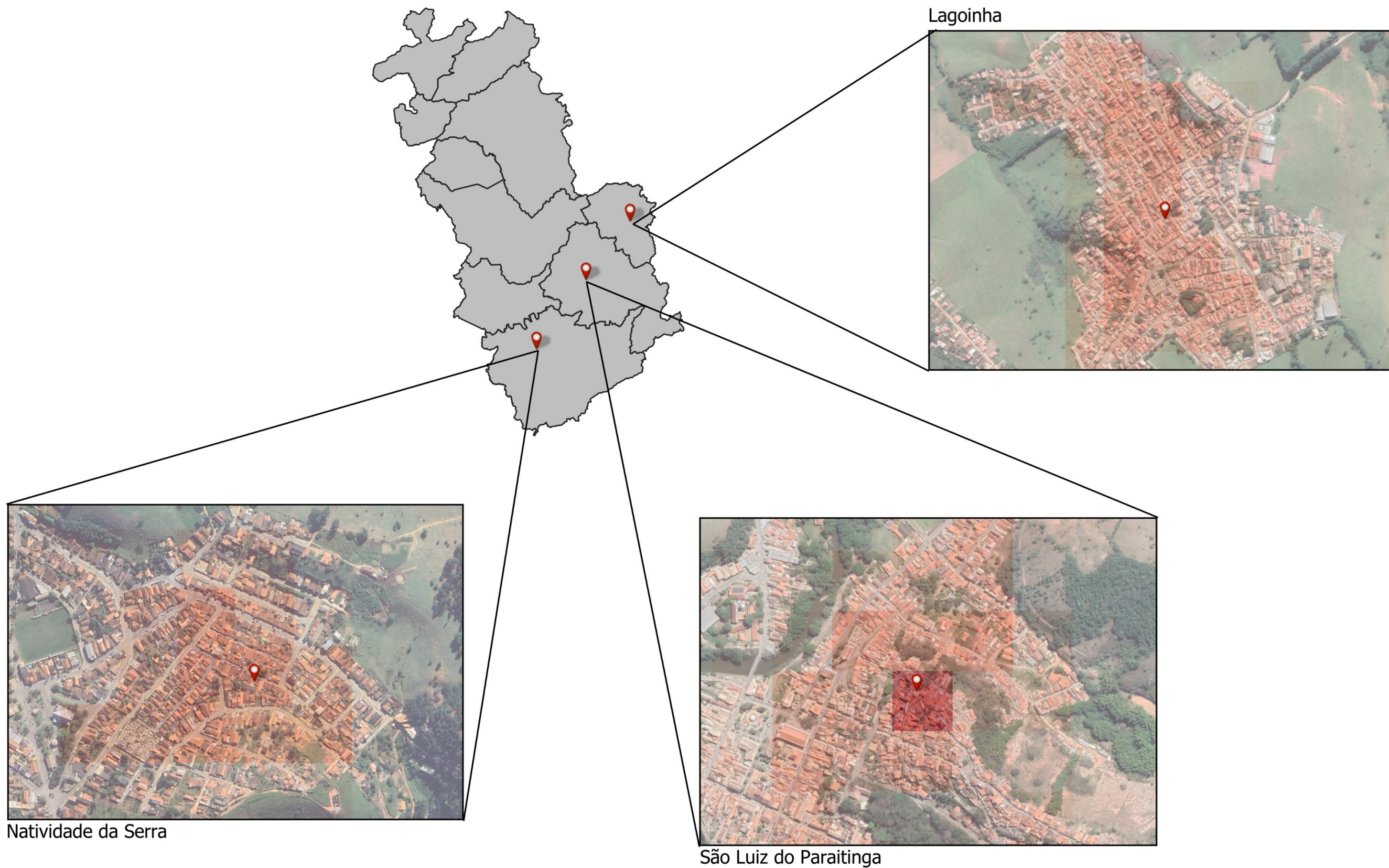
Potim



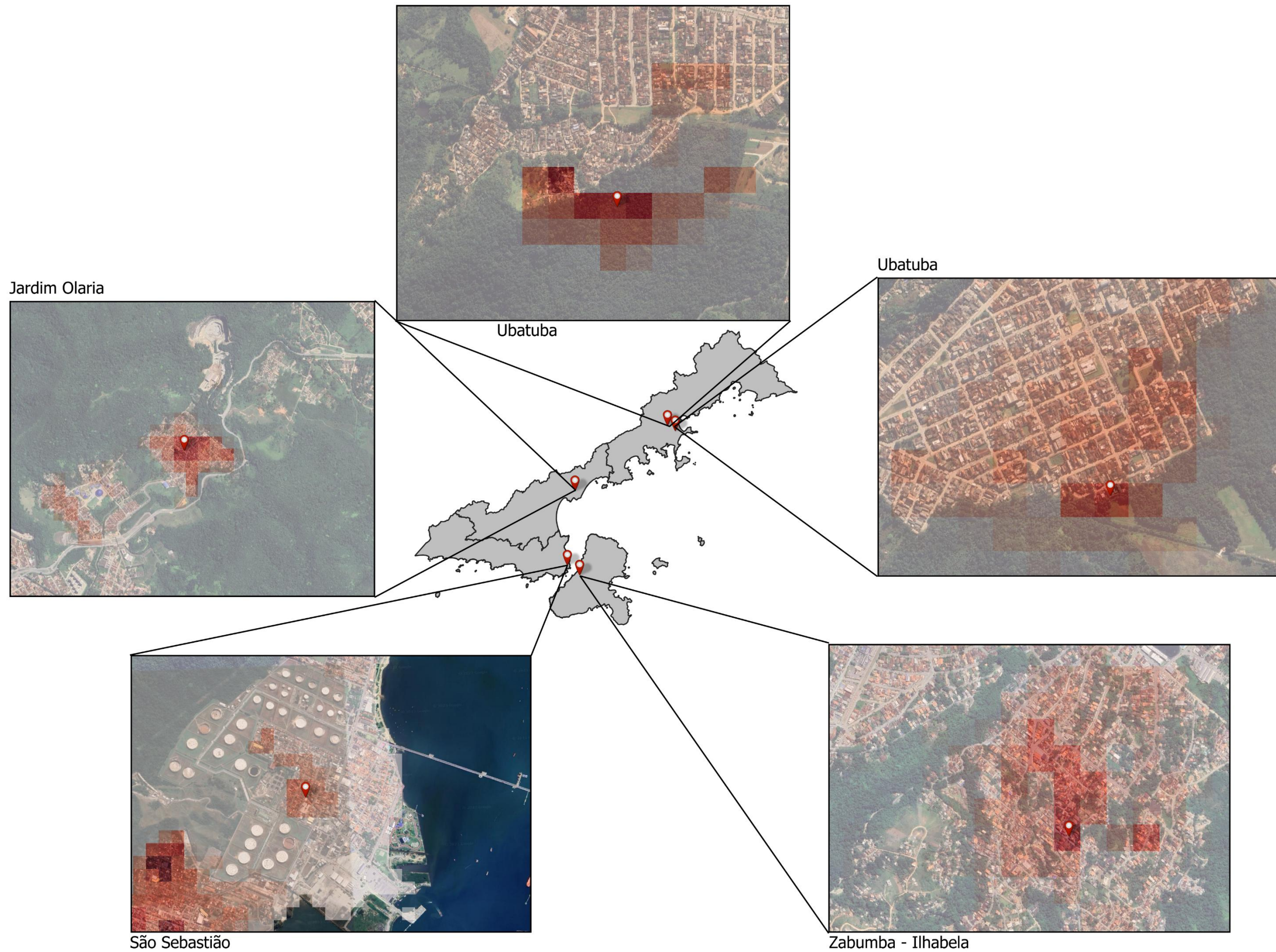
Cunha



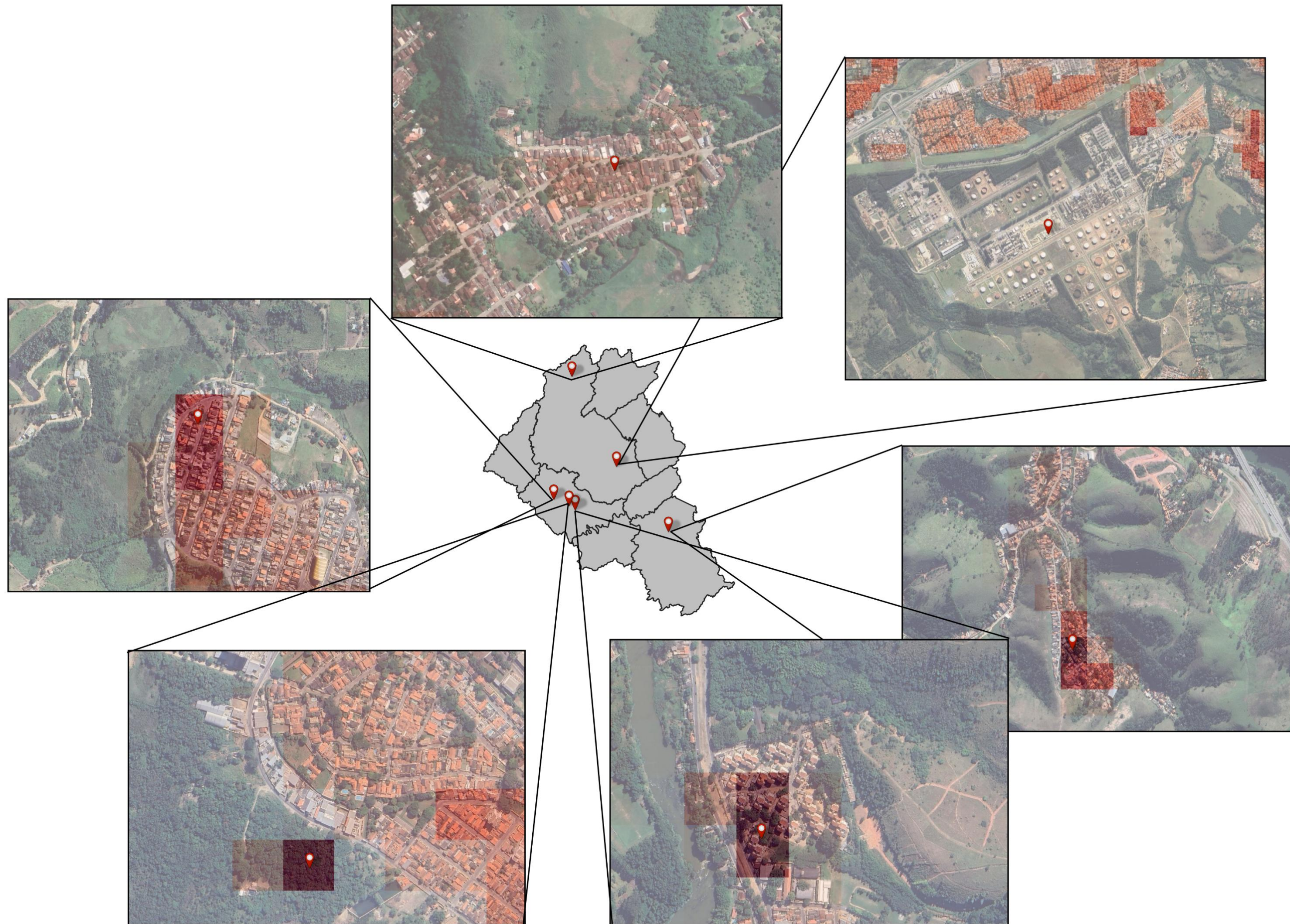
Avaliação: Sub-região Taubaté



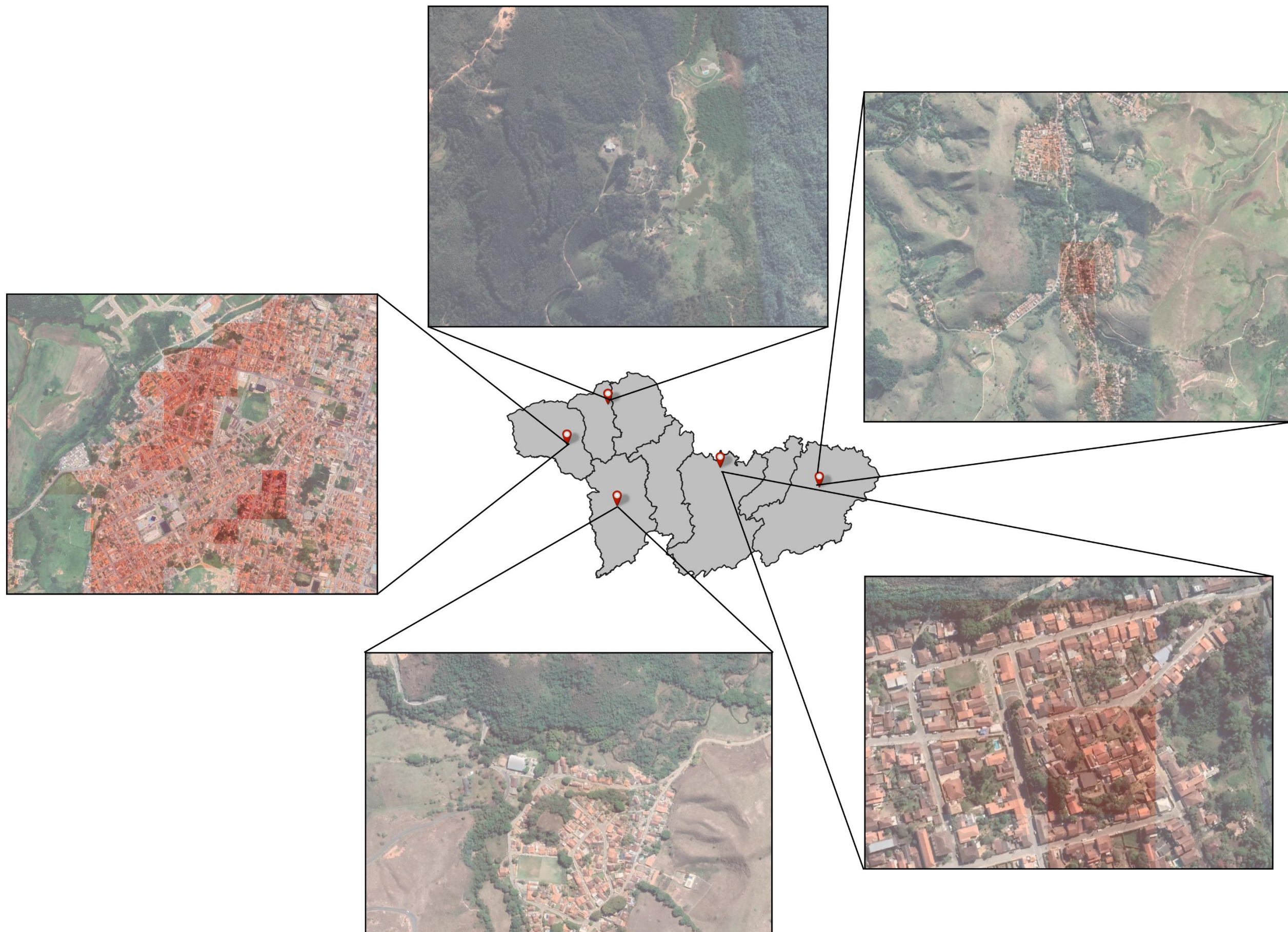
Avaliação: Sub-região Litoral Norte



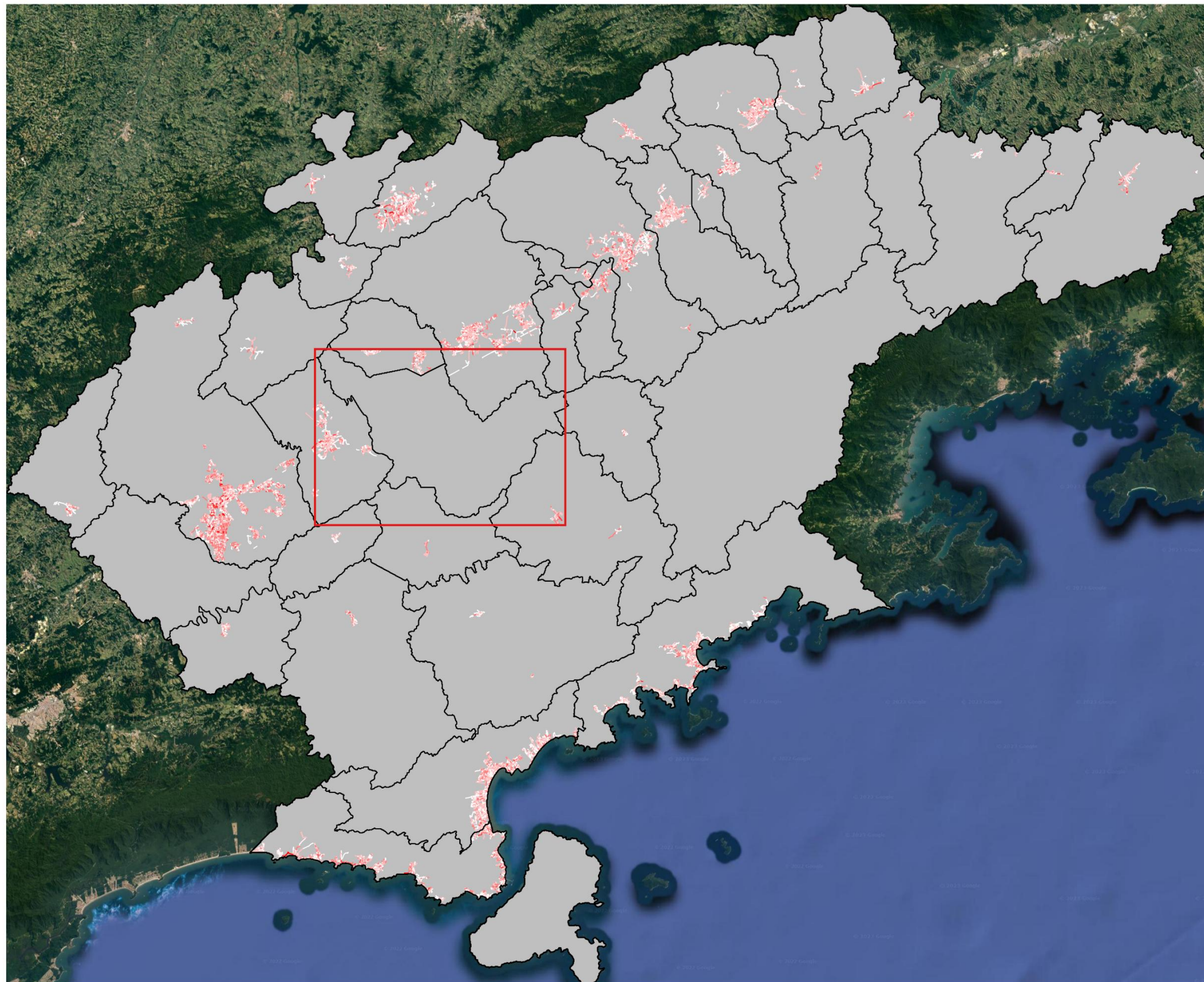
Avaliação: Sub-região São José dos Campos



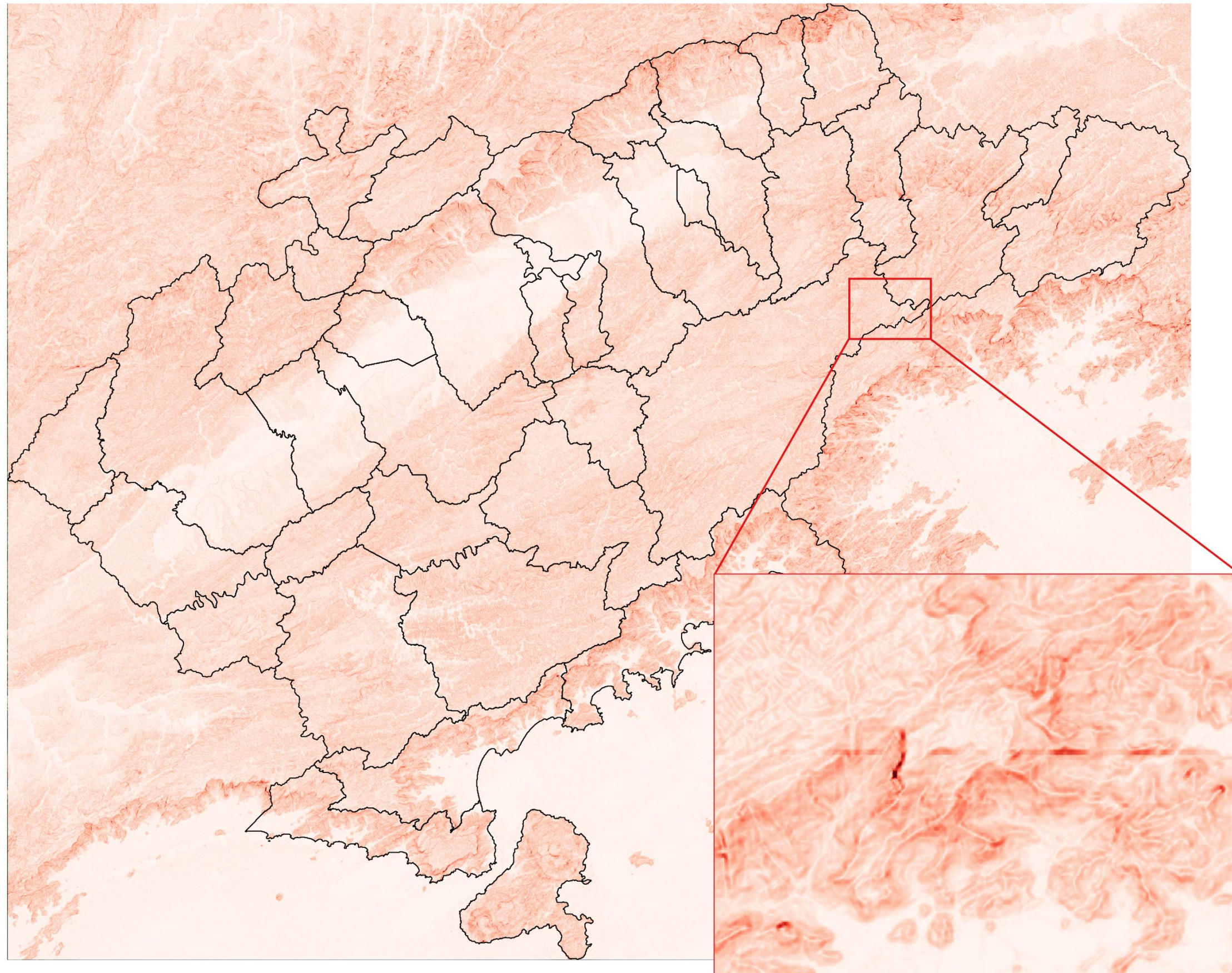
Avaliação: Sub-região Cruzeiro



Avaliação: CNEFE faces logradouros



Avaliação: Declividade



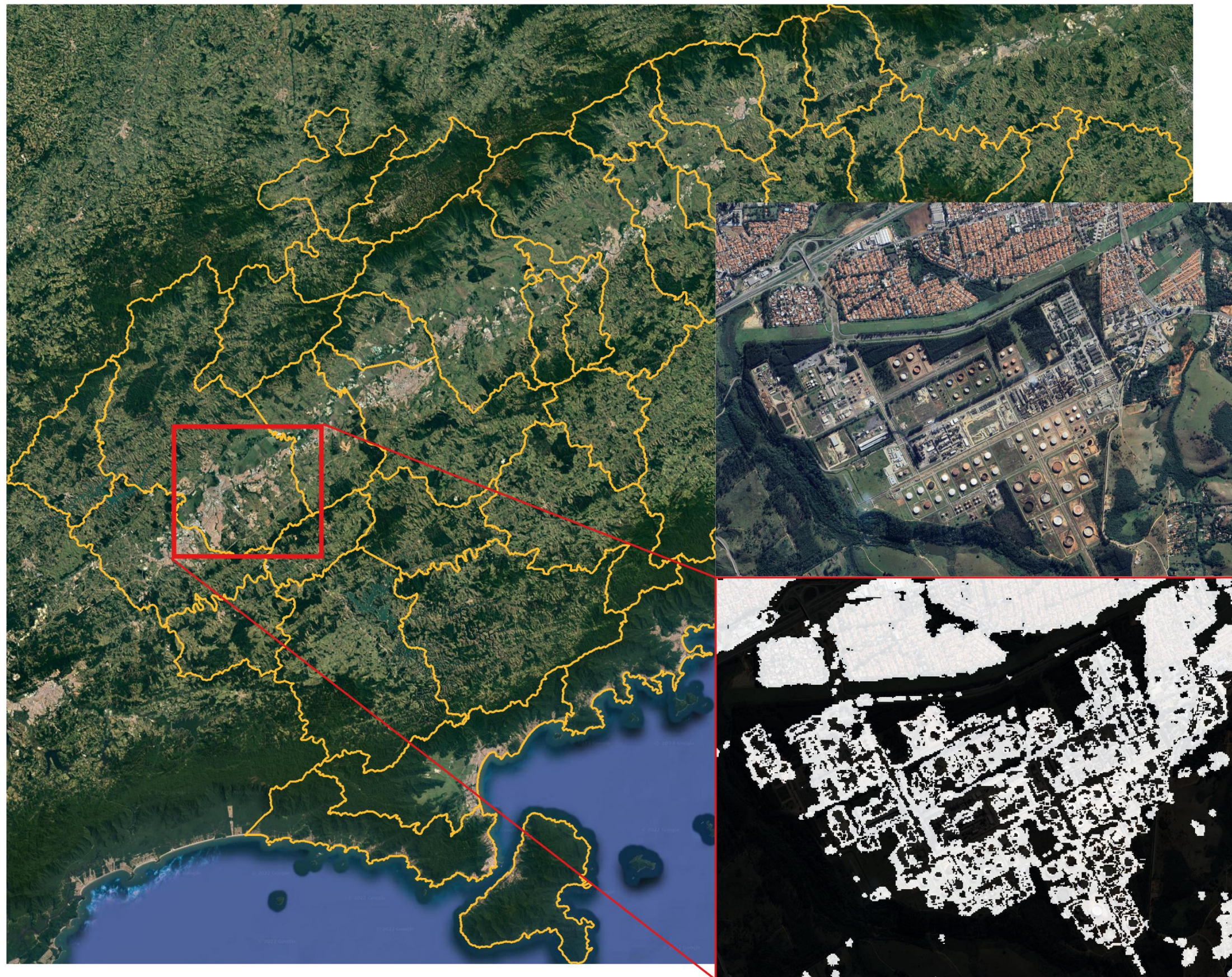
Avaliação: GHS_Land

- Representa a distribuição da superfície da terra estimada por meio da união dos dados do OSM e dados de satélite.

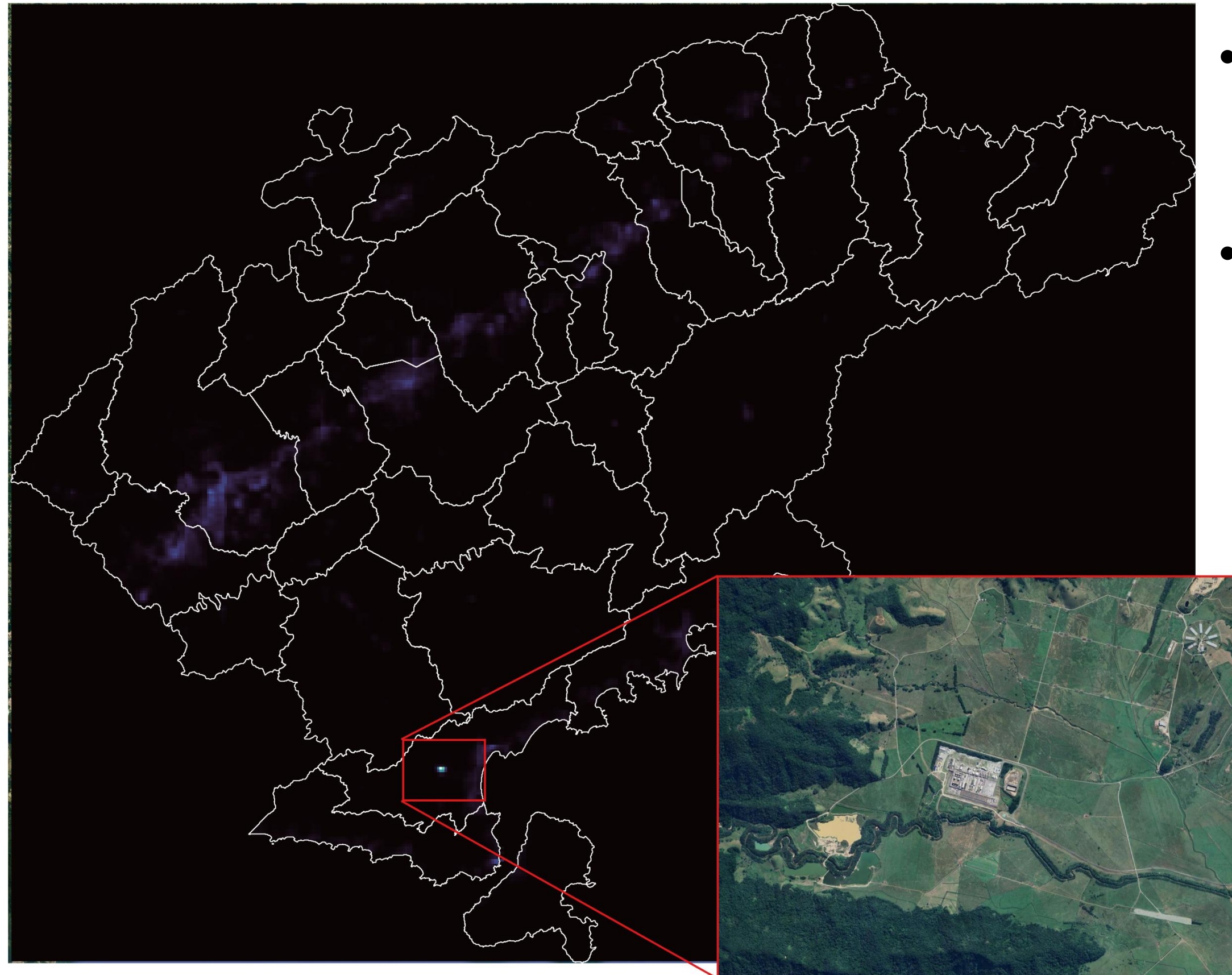


Avaliação: GHS_Built_S

- O total de superfície construída; não residencial construída;
- A classe não residencial (NRES) é estimado da imagem do Sentinel 2 pela observação dos atributos radiométricos, texturais e morfológicos, baseados no método localmente adaptativo por meio de índice de pixels espectrais, avaliação de campo espectral multiescalar e análise de forma por orientada objeto;
- A superfície direcionada para o uso das residências e áreas construídas.



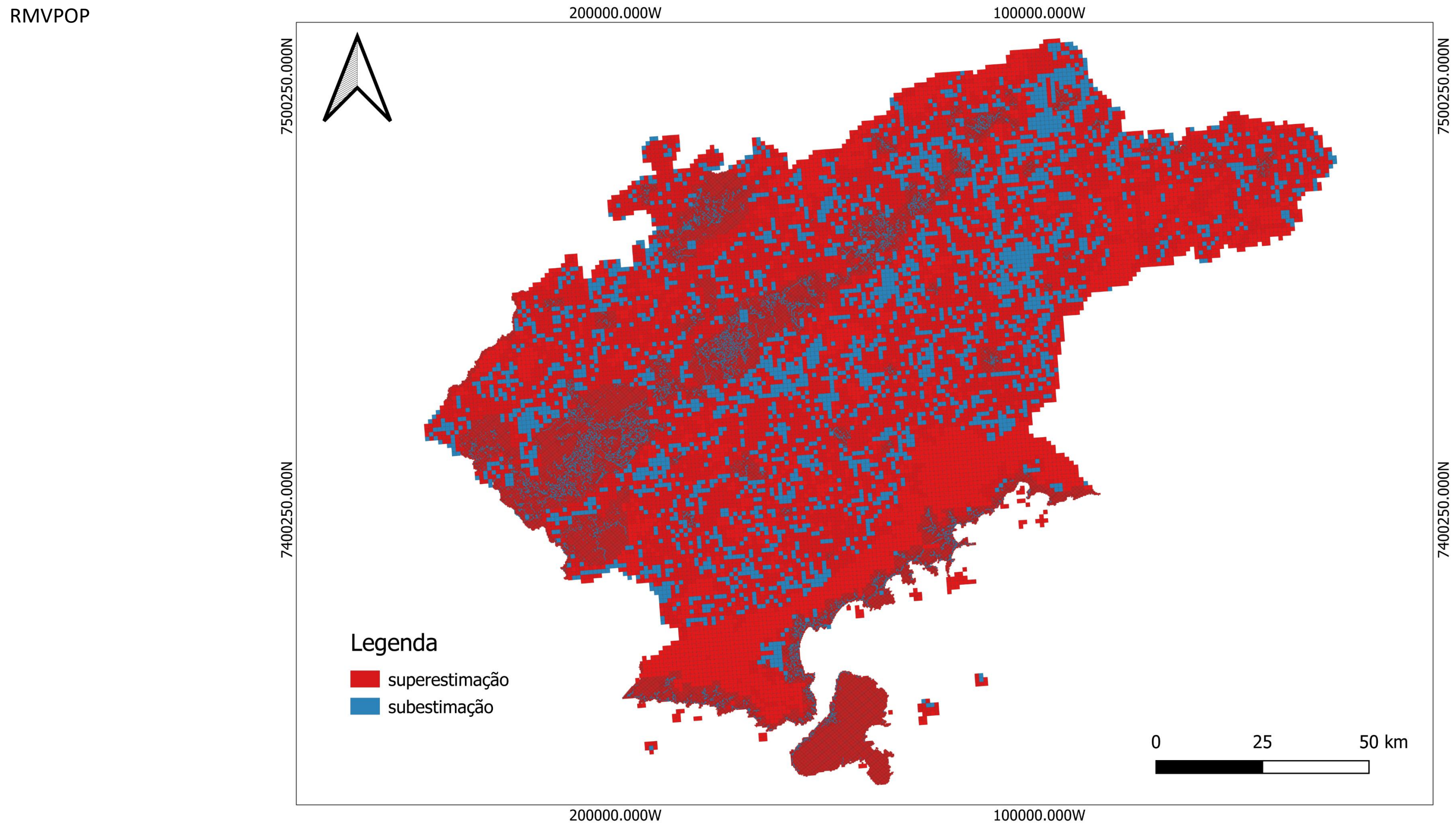
Avaliação: Black Marble



- Correção dos valores de reflectância proveniente das águas marinhas.
- Reclassificação das áreas oceânicas = 0.

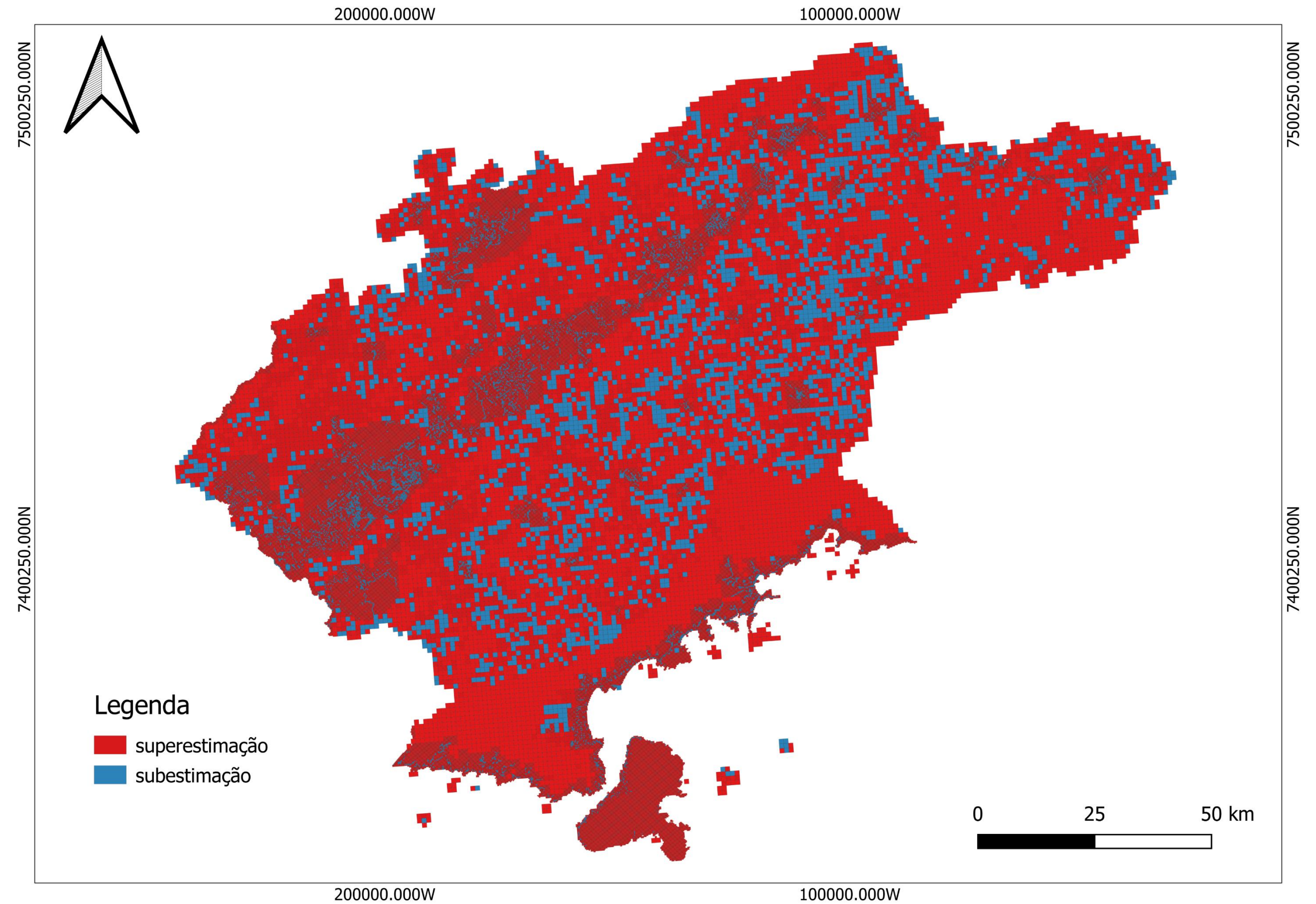
Prédio de Engenharia UTGCA

Avaliação: Superestimamções e subestimamções

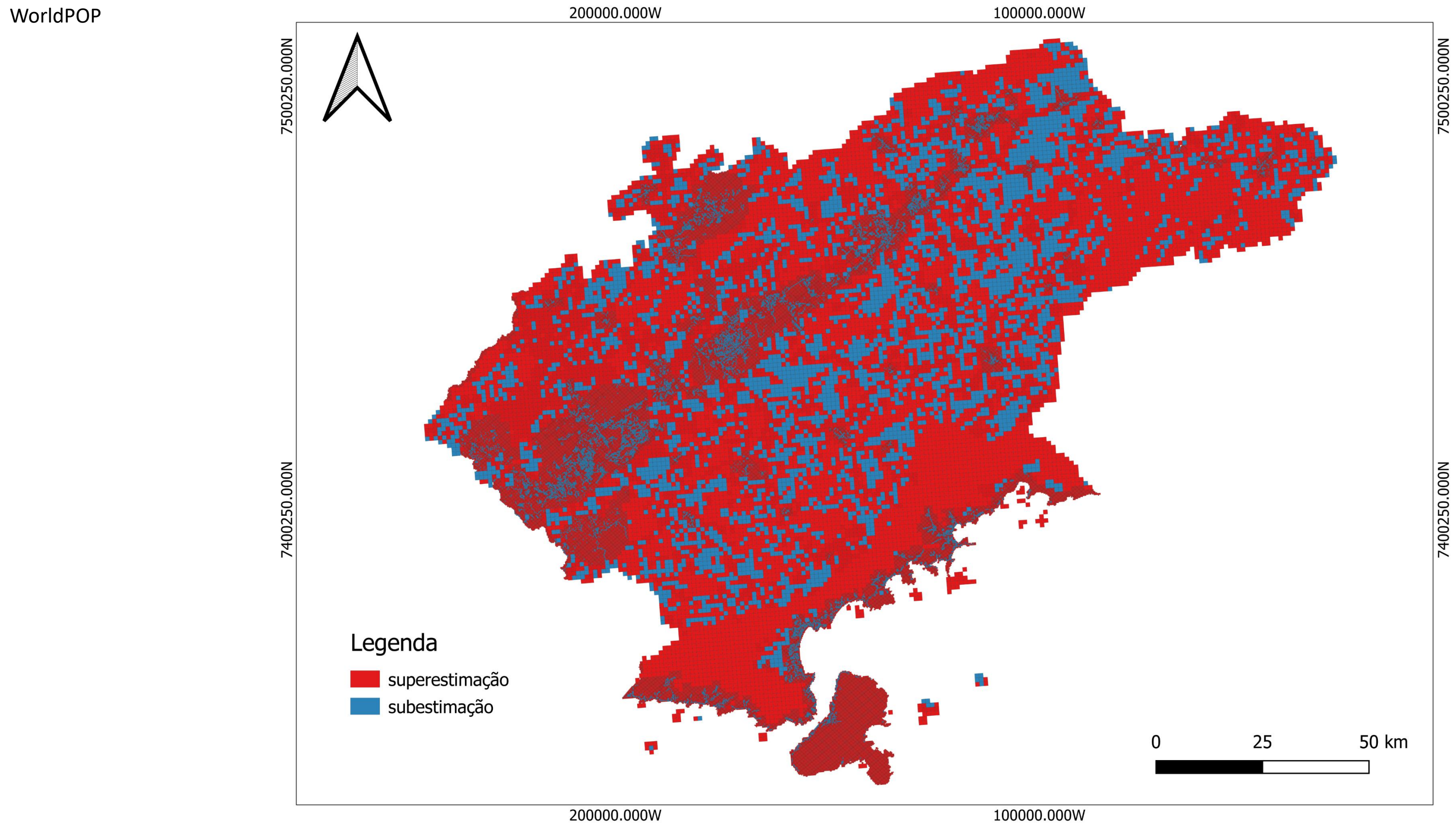


Avaliação: Superestimamções e subestimamções

GHS

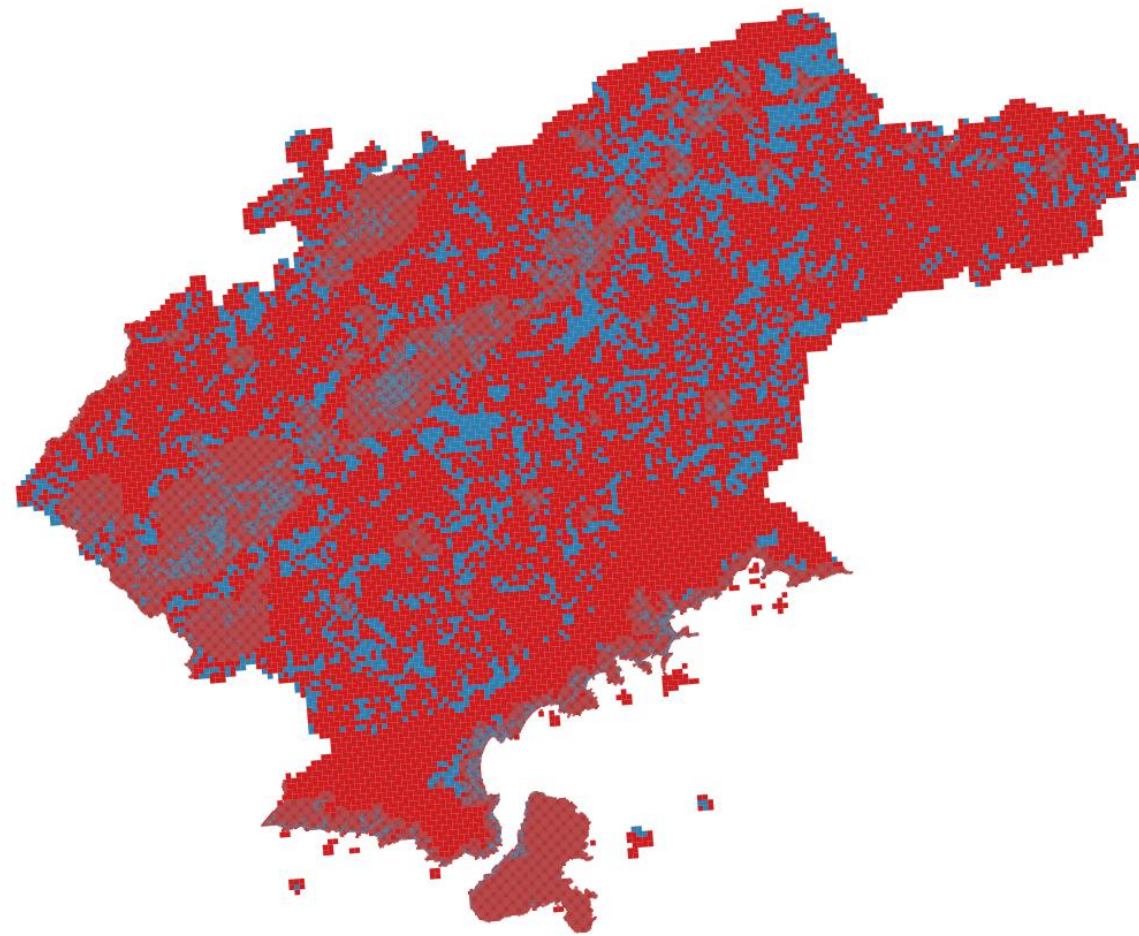


Avaliação: Superestimamções e subestimamções

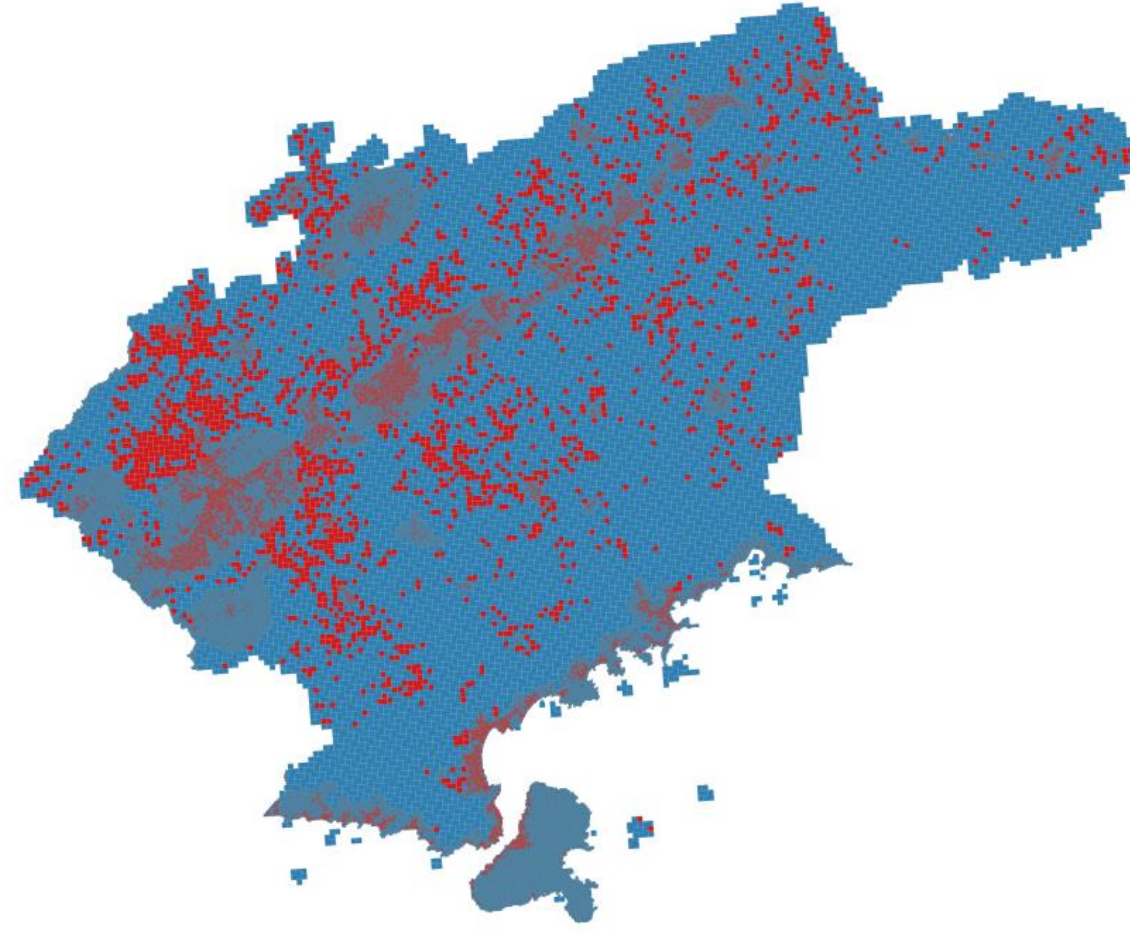


Avaliação: desvio padrão

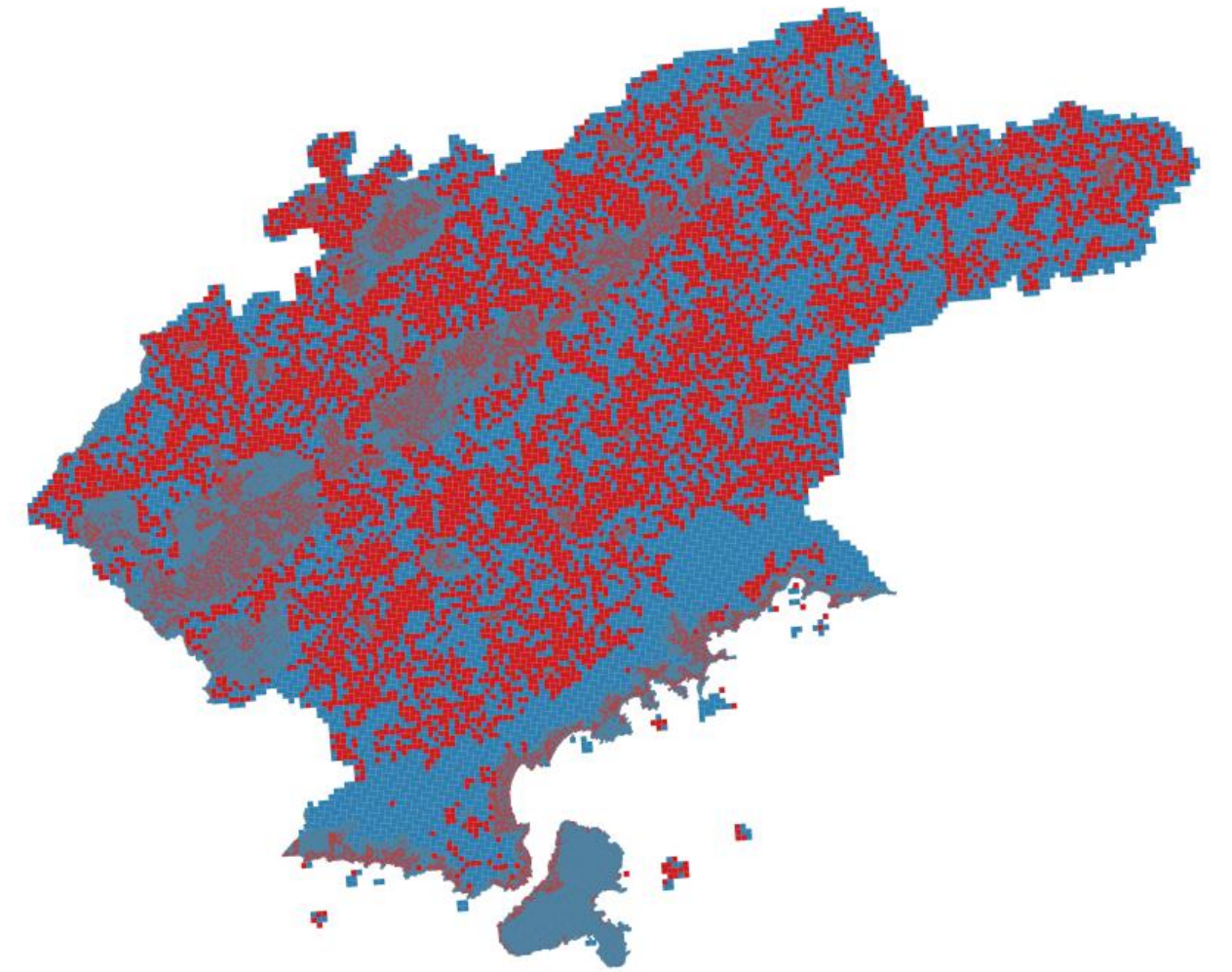
WORLDPOP



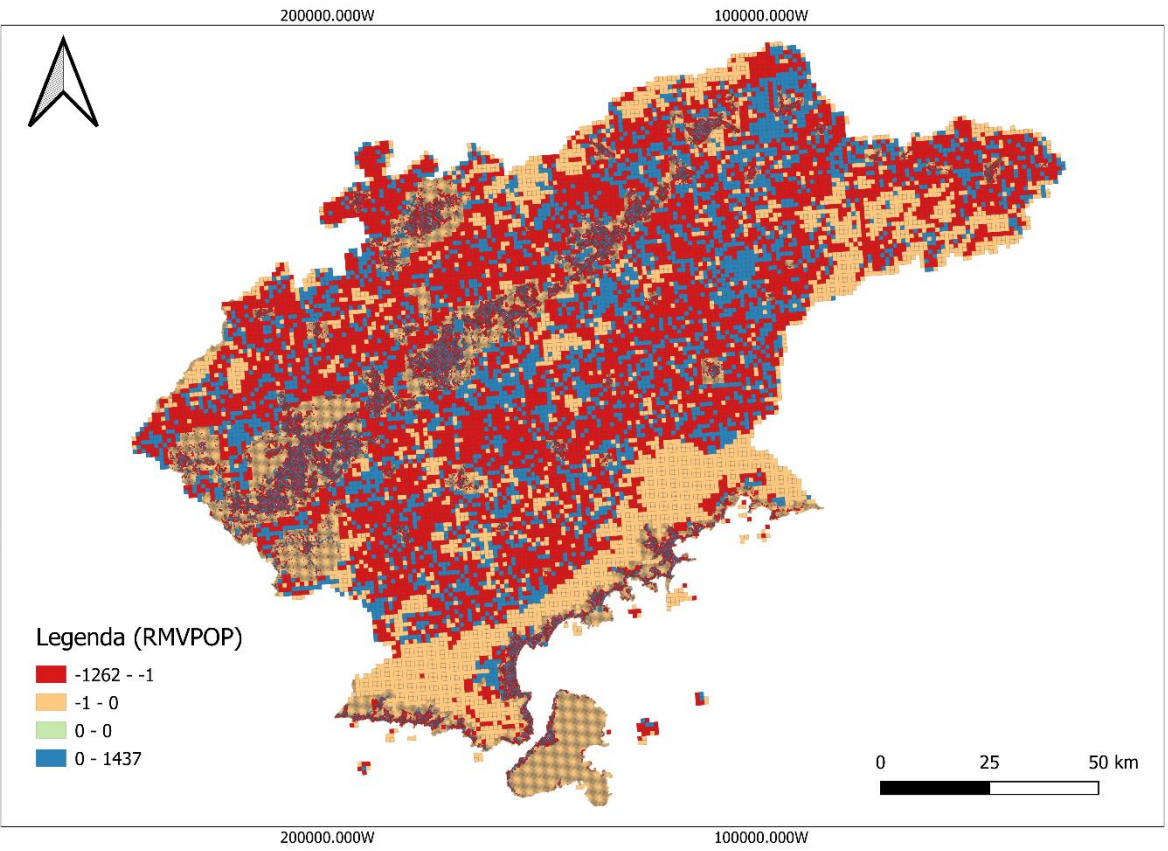
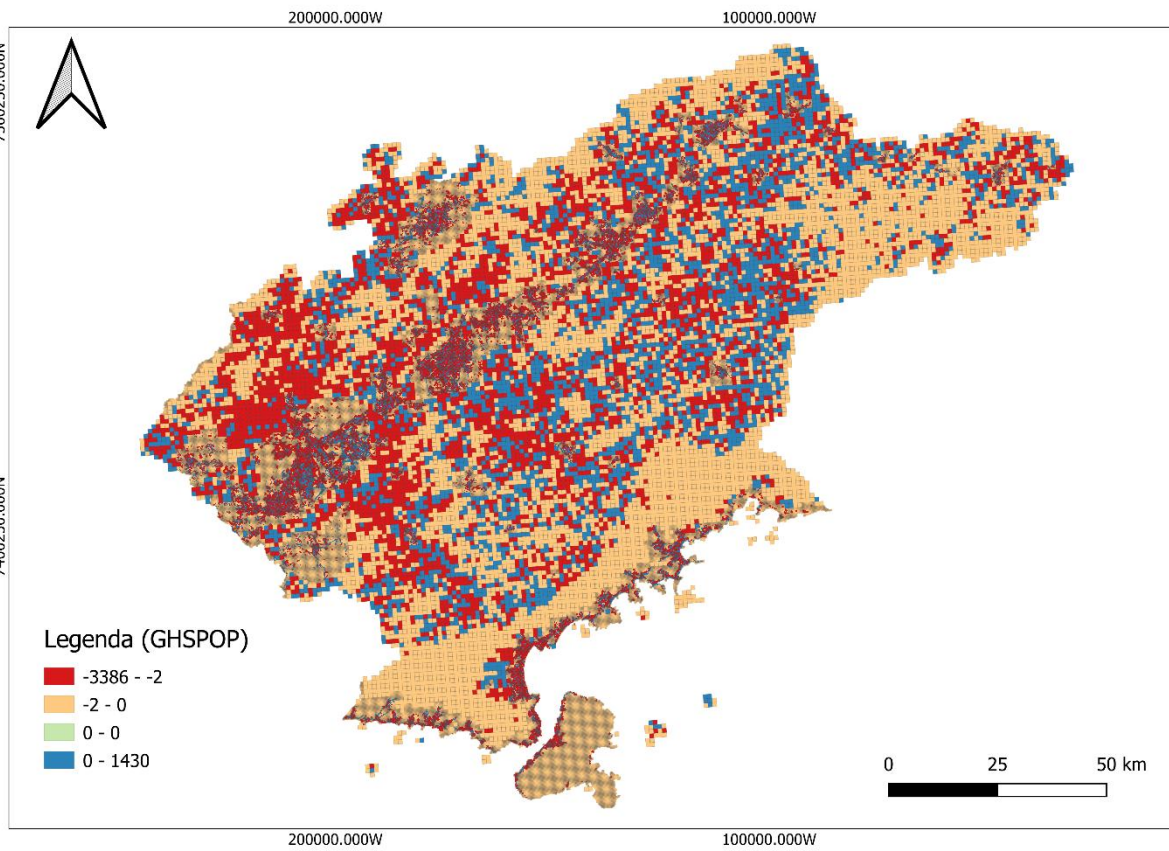
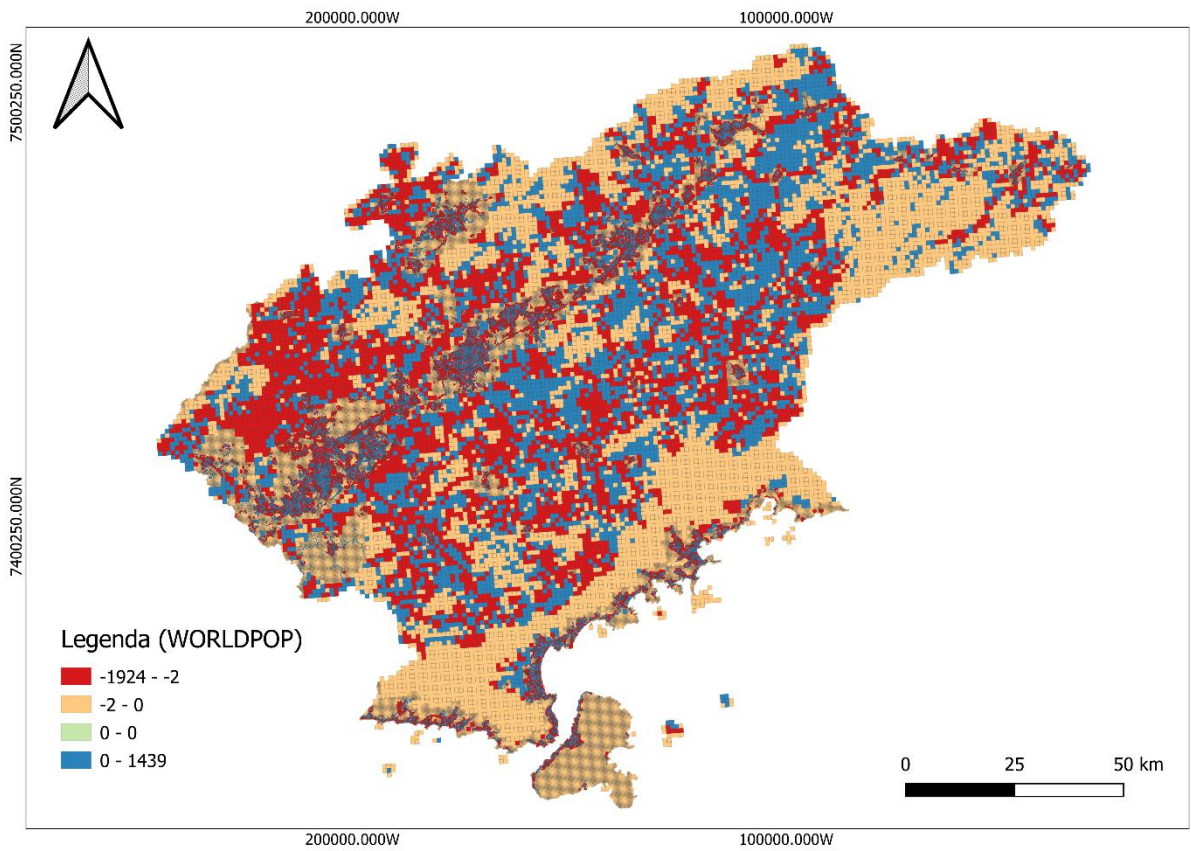
GHSPOP



RMVPOP



Avaliação: QUARTIL



Avaliação: agregado municípios - SJC

Sub-região	Municípios	IBGE 2010	WORLD DP OP	GHSP POP	RMVPO P	WP	GHS	RMV
São José dos Campos	Caçapava	81957	67438	91541	84684	14519	-9584	-2727
São José dos Campos	Igaratá	8583	7008	8678	8799	1575	-95	-216
São José dos Campos	Jacareí	208934	168290	248719	210594	40644	-39785	-1660
São José dos Campos	Jambeiro	5338	4264	8364	5330	1074	-3026	8
São José dos Campos	Monteiro Lobato	3968	3285	4688	4095	683	-720	-127
São José dos Campos	Paraibuna	16542	13820	15784	17310	2722	758	-768
São José dos Campos	Santa Branca	13589	10780	13291	13720	2809	298	-131
São José dos Campos	São José dos Campos	597425	501364	796883	623257	96061	-199458	-25832

Avaliação: agregados municípios - TAU

Sub-região	Municípios	IBGE 2010	WORLDPO P	GHSPOP	RMVPO P	WP	GHS	RMV
Taubaté	Campos do Jordão	47504	38147	48621	47300	9357	-1117	204
Taubaté	Lagoinha	4835	3870	4034	4801	965	801	34
Taubaté	Natividade da Serra	6667	5285	5345	6672	1382	1322	-5
Taubaté	Pindamonhangaba	146117	117102	183556	146467	29015	-37439	-350
Taubaté	Redenção da Serra	3850	3064	3099	3853	786	751	-3
Taubaté	Santo Antônio do Pinhal	6405	5160	5981	6468	1245	424	-63
Taubaté	São Bento do Sapucaí	10430	8340	9311	10394	2090	1119	36
Taubaté	São Luiz do Paraitinga	10375	8234	8990	10179	2141	1385	196
Taubaté	Taubaté	275218	222063	328712	274456	53155	-53494	762
Taubaté	Tremembé	40751	32615	49486	36817	8136	-8735	3934

Avaliação: agregados municípios - GUA

Sub-região	Municípios	IBGE 2010	WORLDPOP	GHSPop	RMVPOP	WP	GHS	RMV
Guaratinguetá	Aparecida	34817	27886	35876	34650	6931	-1059	167
Guaratinguetá	Cachoeira Paulista	29800	24056	31972	29923	5744	-2172	-123
Guaratinguetá	Canas	4382	3498	5608	4384	884	-1226	-2
Guaratinguetá	Cunha	21842	17384	17021	21758	4458	4821	84
Guaratinguetá	Guaratinguetá	111322	89527	132422	111383	21795	-21100	-61
Guaratinguetá	Lorena	81295	65774	80396	82197	15521	899	-902
Guaratinguetá	Piquete	14092	11283	10551	14013	2809	3541	79
Guaratinguetá	Potim	19392	15375	33999	16769	4017	-14607	2623
Guaratinguetá	Roseira	9594	7656	10446	9549	1938	-852	45

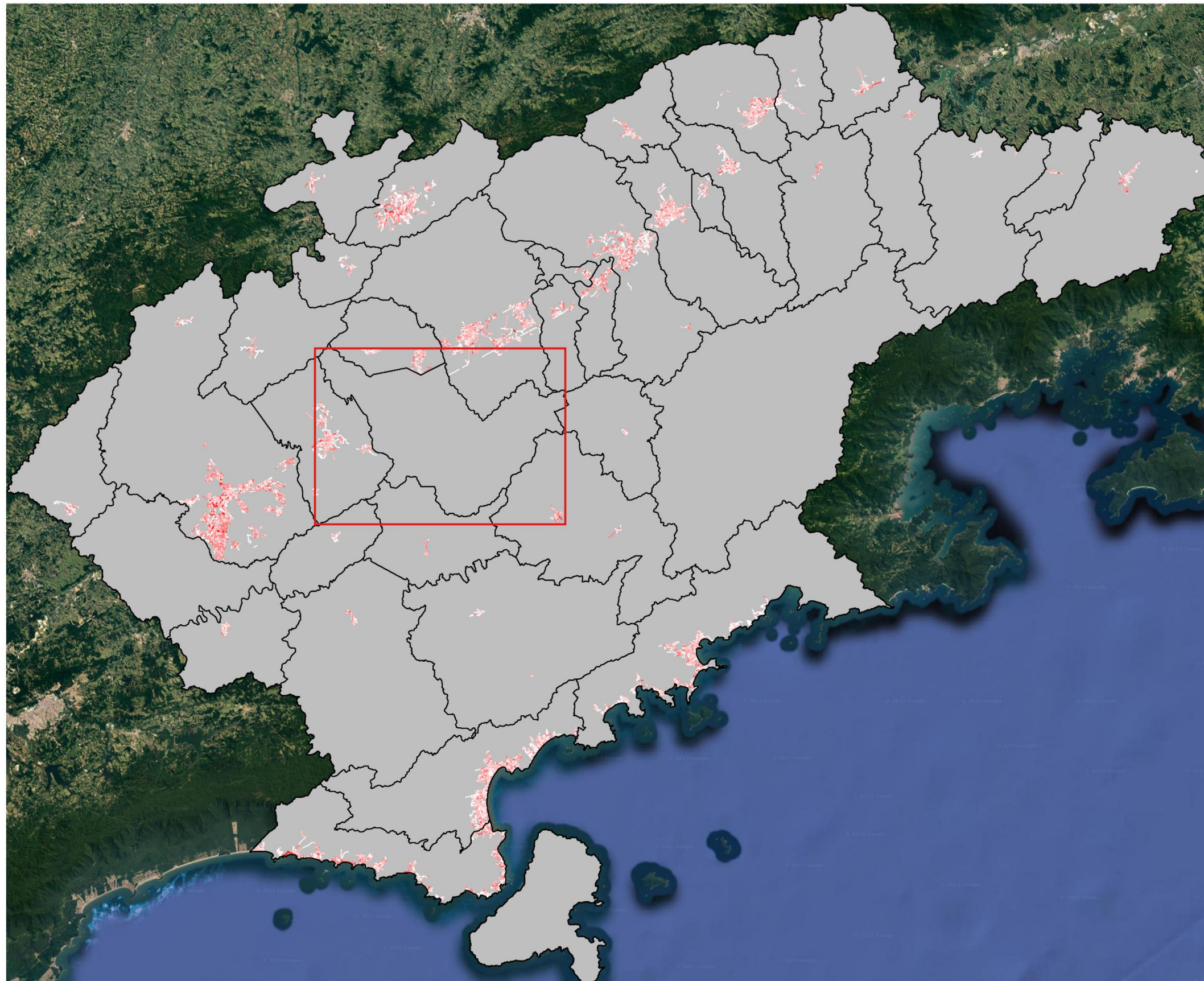
Avaliação: agregados municípios - LN

Sub-região	Municípios	IBGE 2010	WORLD DP OP	GHSPOP	RMVPOP	WP	GHS	RMV
Litoral Norte	Caraguatatuba	99540	79419	143015	99913	20121	-43475	-373
Litoral Norte	Ilhabela	27956	20834	43511	28071	7122	-15555	-115
Litoral Norte	São Sebastião	72219	56970	102245	73700	15249	-30026	-1481
Litoral Norte	Ubatuba	76456	62032	94607	78574	14424	-18151	-2118

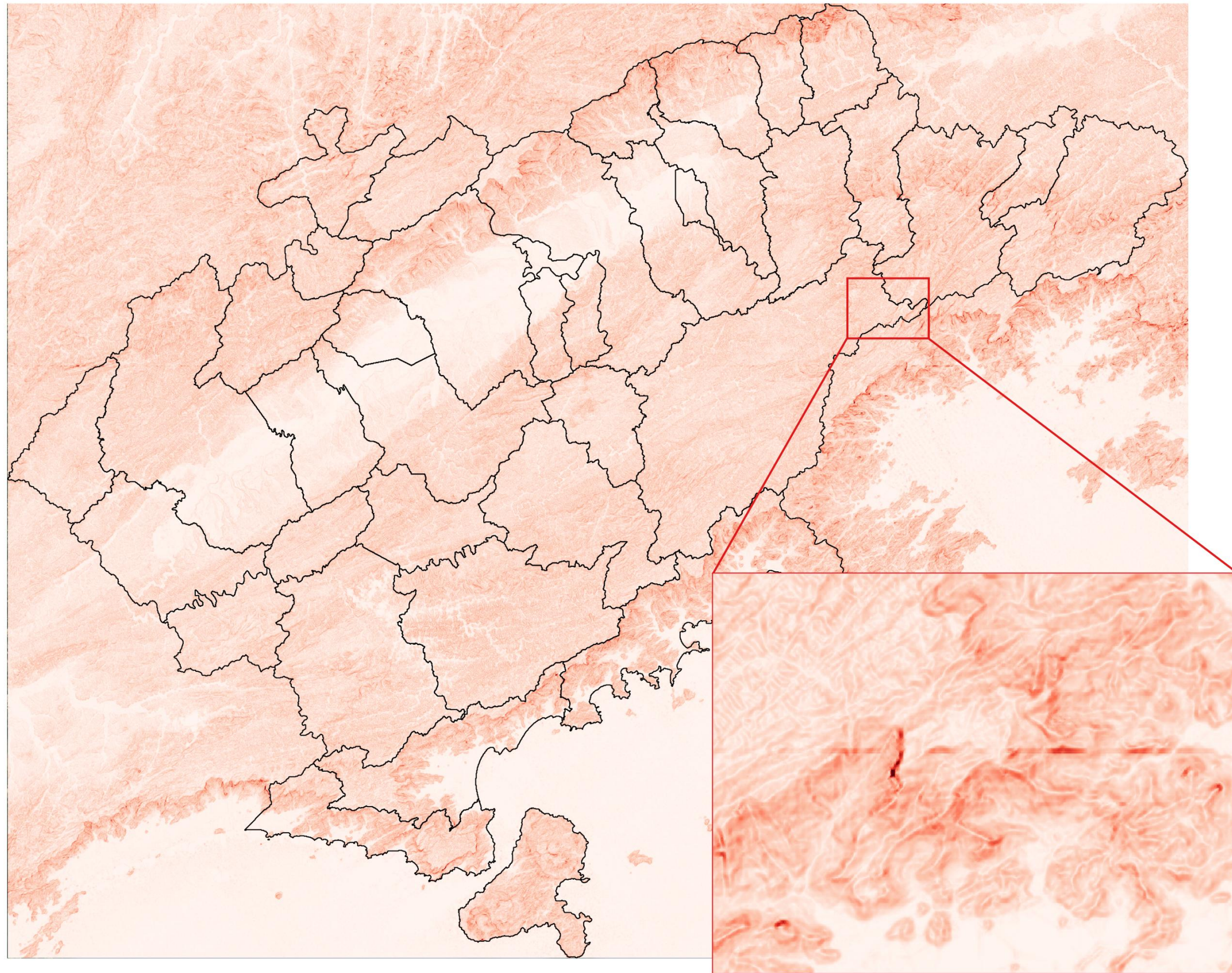
Avaliação: agregado municípios - CRZ

Sub-região	Municípios	IBGE 2010	WORLDPOP	GHSPOP	RMVPOP	WP	GHS	RMV
Cruzeiro	Arapeí	2491	1980	1959	2480	511	532	11
Cruzeiro	Areias	3692	2948	3376	3646	744	316	46
Cruzeiro	Bananal	10138	8130	9818	10152	2008	320	-14
Cruzeiro	Cruzeiro	77001	61543	73599	76855	15458	3402	146
Cruzeiro	Lavrinhas	6572	5293	6924	6511	1279	-352	61
Cruzeiro	Queluz	11325	8814	15145	10642	2511	-3820	683
Cruzeiro	São José do Barreiro	4032	3233	3417	4063	799	615	-31
Cruzeiro	Silveiras	5786	4612	5835	5780	1174	-49	6

Avaliação: CNEFE faces logradouros



Avaliação: Declividade



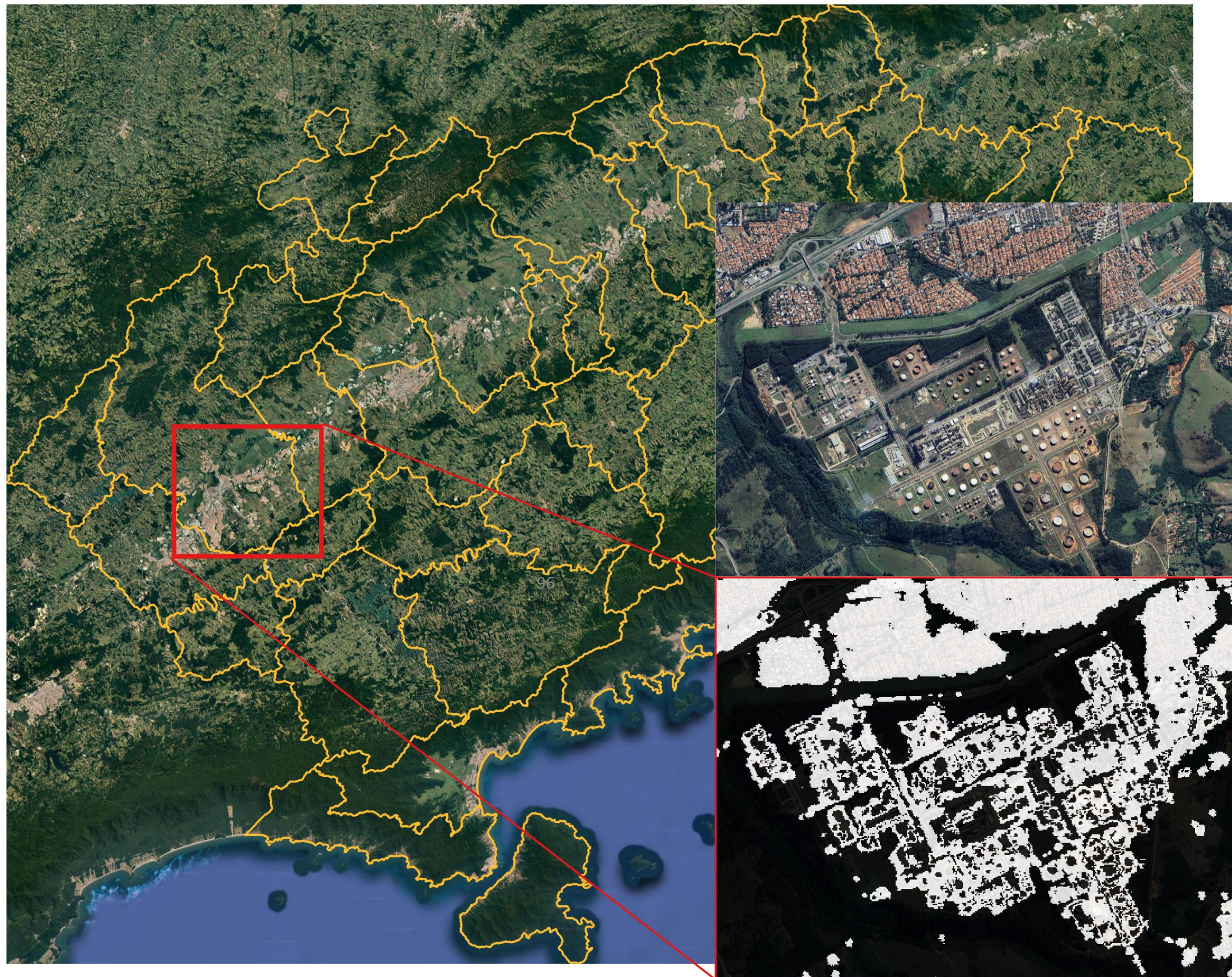
Avaliação: GHS_Land

- Representa a distribuição da superfície da terra estimada por meio da união dos dados do OSM e dados de satélite.

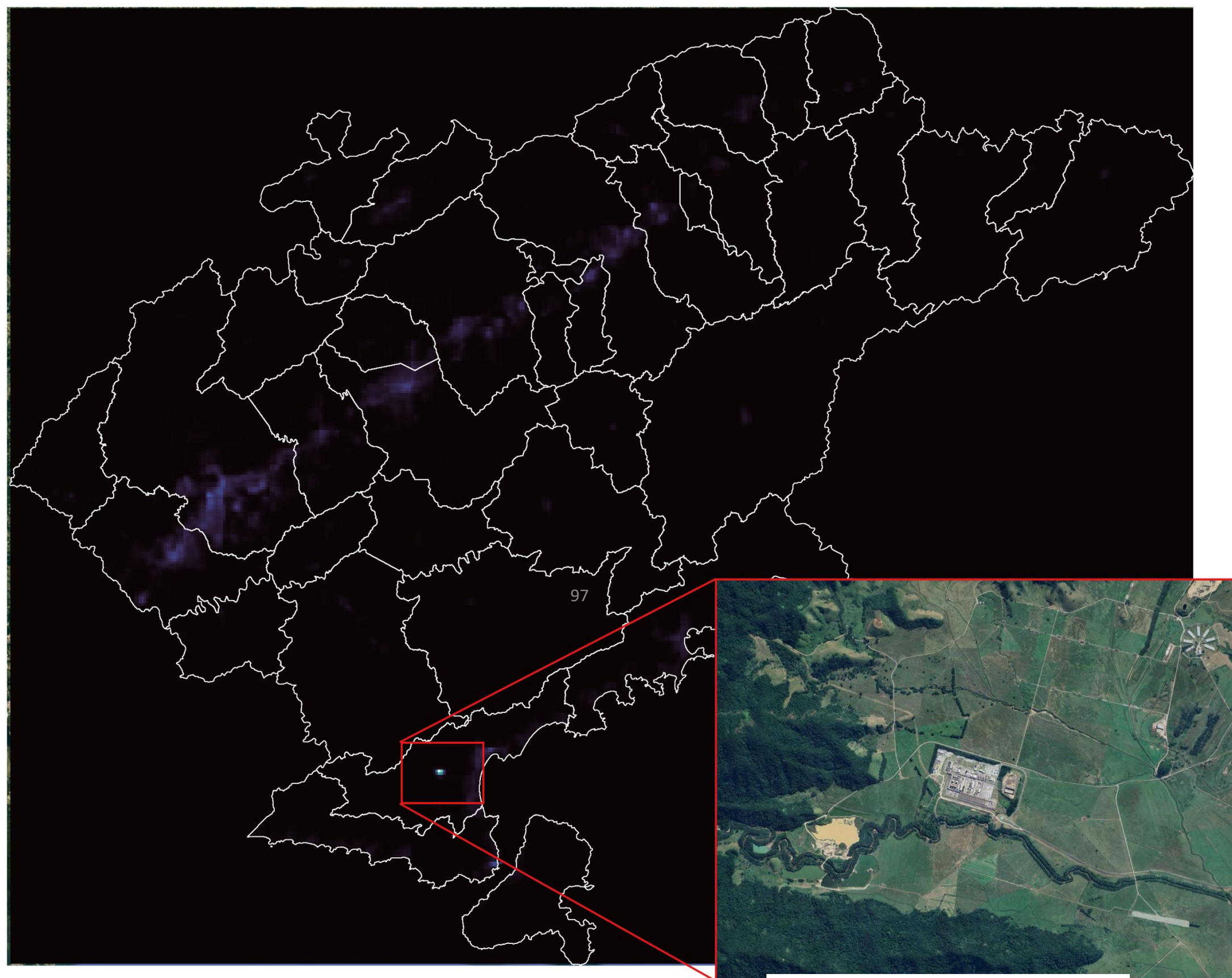


Avaliação: GHS_Built_S

- O total de superfície construída; não residencial construída;
- A classe não residencial (NRES) é estimado da imagem do Sentinel 2 pela observação dos atributos radiométricos, texturais e morfológicos, baseados no método localmente adaptativo por meio de índice de pixels espectrais, avaliação de campo espectral multiescalar e análise de forma por orientada objeto;
- A superfície direcionada para o uso das residências e áreas construídas.

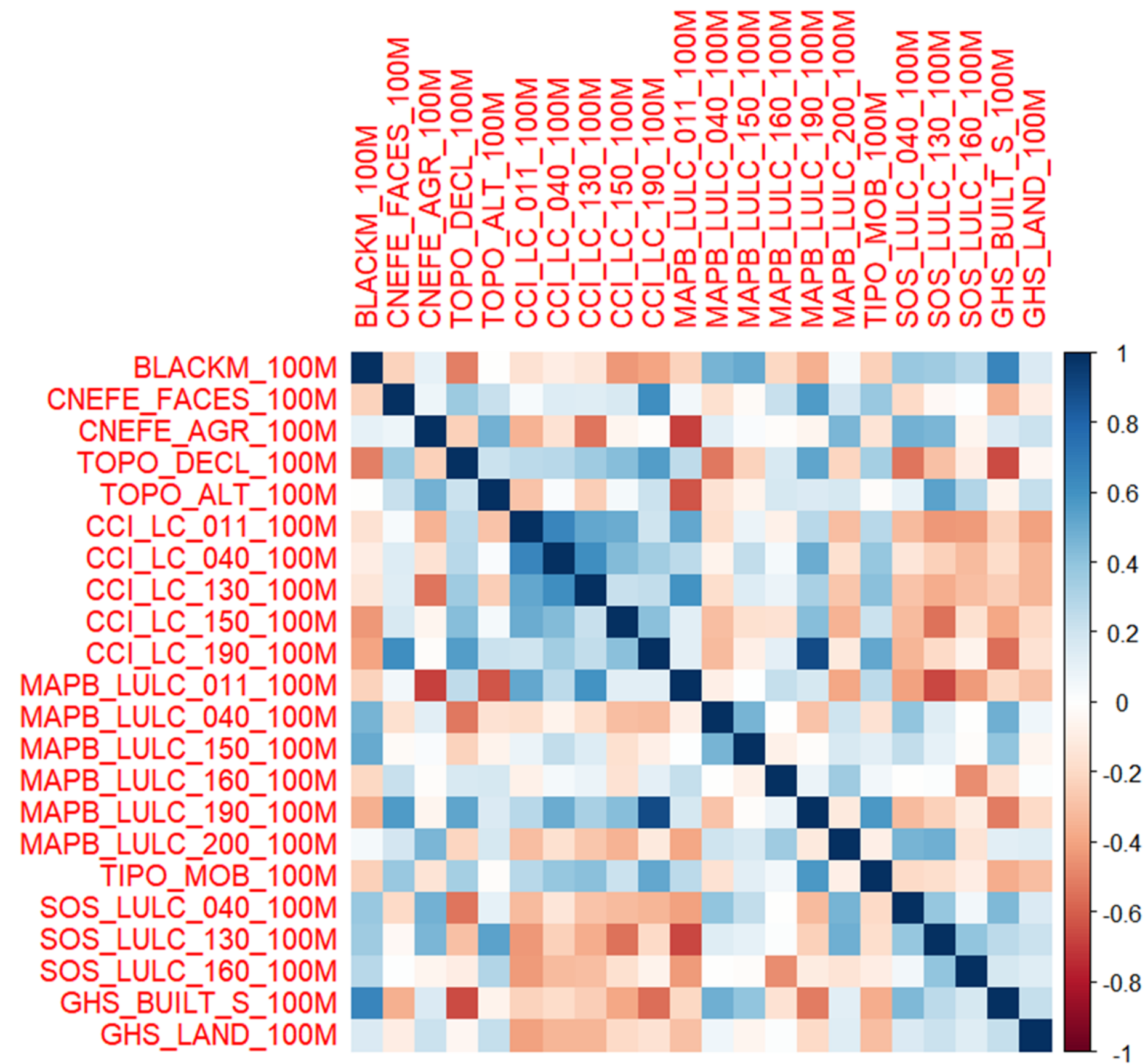


Avaliação: Black Marble



Prédio de Engenharia UTGCA

Avaliação: Correlação entre variáveis



Discussões

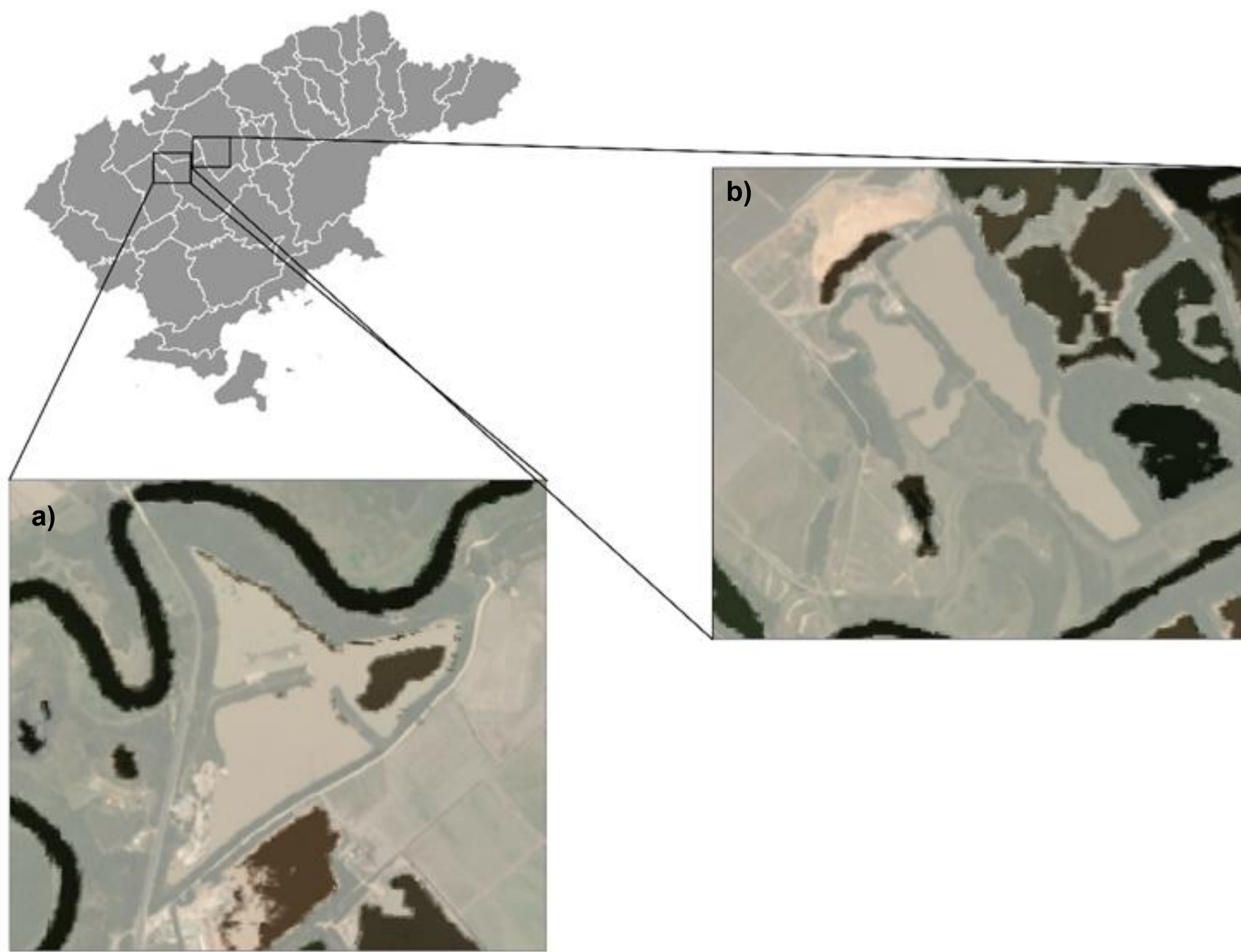
- RMVPop possui os melhores resultados em relação às medidas de MAE, RMSE;
- WorldPop tinha os valores mais baixos em áreas de baixa densidade;
- GHS-POP em áreas de alta densidade;
- Células da RMVPop não incluídas na área urbana captaram a presença de população;
- A variável GHS_Built_S compõe uma variável importante para a construção do mapa de densidade populacional da RMV.

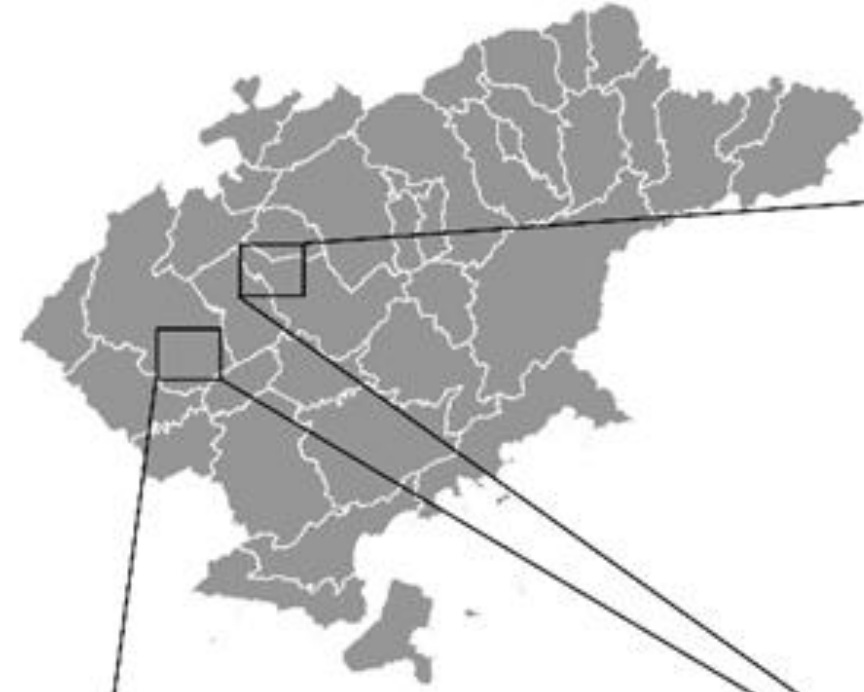
Discussões

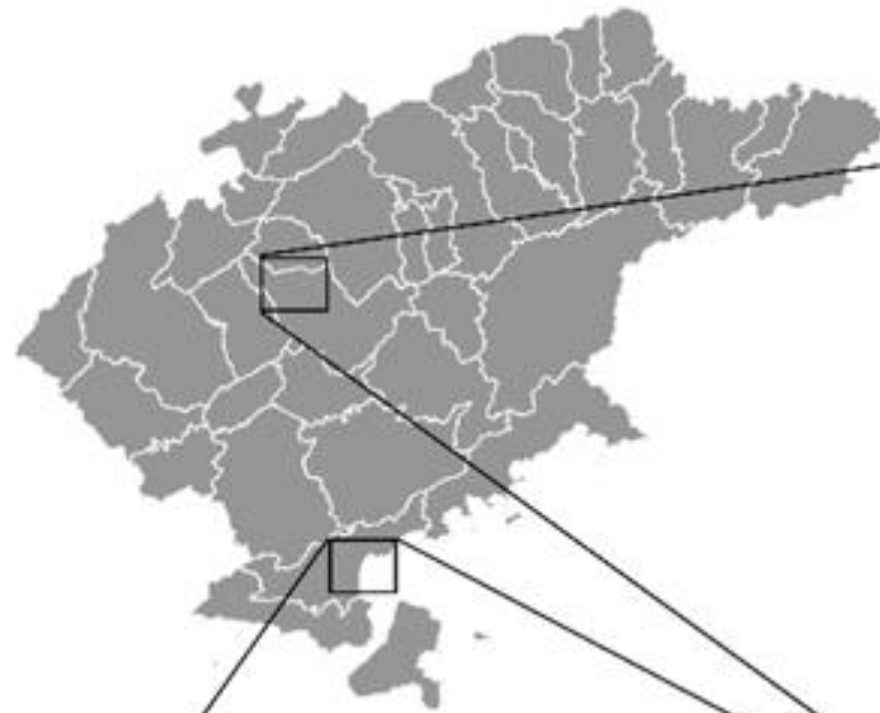
- Os dados de GHS_BUILT_S possuiu maior importância;
- A utilização da máscara da variável GHS_Land possibilitou a exclusão das áreas não habitadas;
- Ausência de atributos para representar determinado fenômenos pode influenciar o resultado de forma negativa;
- Considerando os protocolos realizados, o treinamento do modelo utilizando o conjunto de do protocolo 2 (v004) realizou as melhores predições e melhor resultado do modelo 82%.

Reflexões

- Modelo populacional apresenta tendências no espaço;
- Nova fronteira: como representar os invisíveis?;
- Estratificação perfil populacional?.









Obrigado pela atenção.

LinkedIn: <https://www.linkedin.com/in/diego-moreira-silva-a62a31153/>
E-mail: silvamoreiradiego0@gmail.com