

População, Espaço e Ambiente

Abordagens Espaciais em Estudos de População:
Métodos Analíticos e Técnicas de Representação

Parte IV – Integração de Dados

De áreas a superfícies

Silvana Amaral

Antonio Miguel V. Monteiro

{silvana@dpi.inpe.br, miguel@dpi.inpe.br}



Interpoladores *Geográficos*

Modelos de População em Superfícies

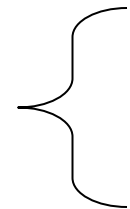
Baseados em Pontos – *Point-based*

Baseados em Áreas – *Areal-Based*

Informação Auxiliar

Conhecimento dos Processos

População

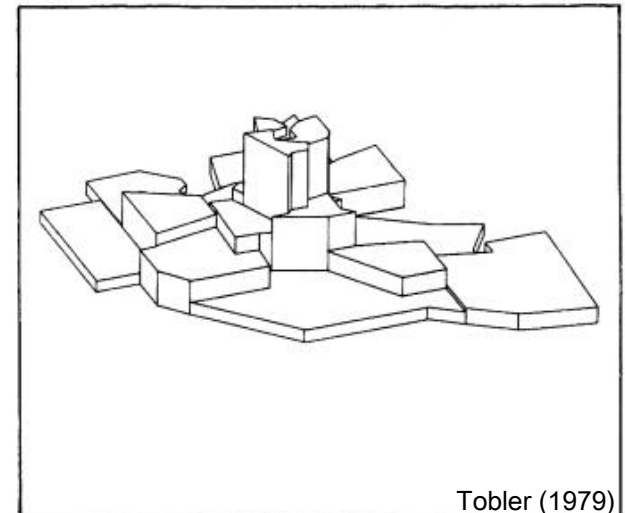
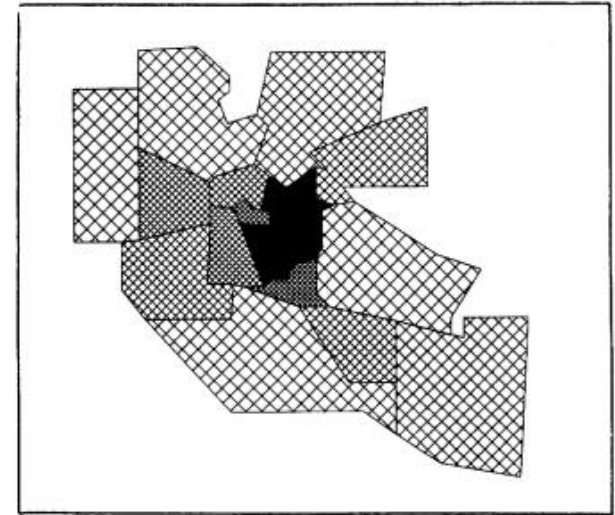


Preservar Volume
Restrições

Modelos de População “Univariados”

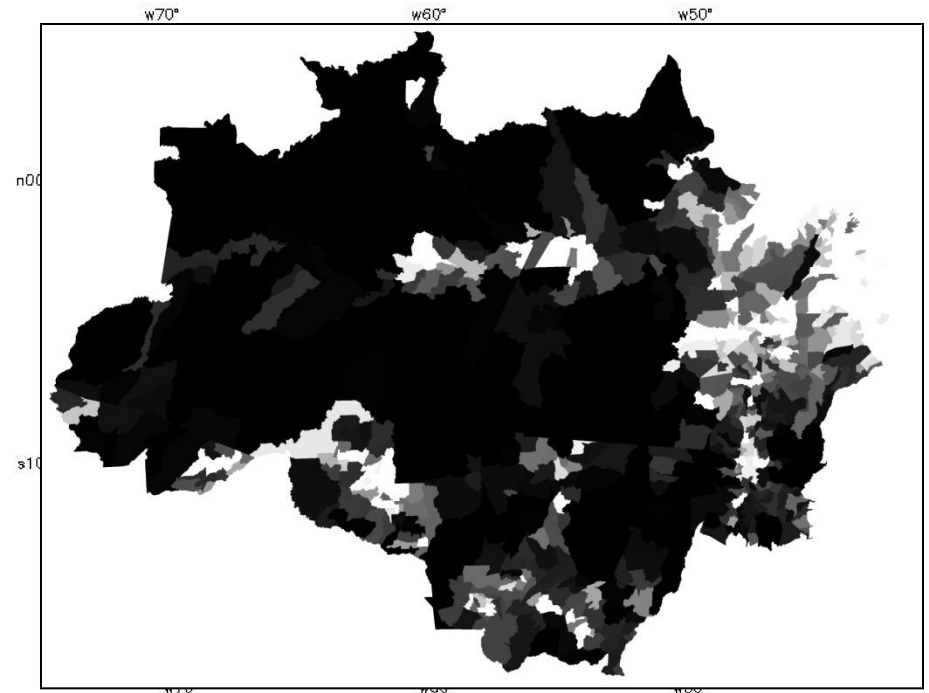
- Ponderado pela área
 - Valores de densidade de população proporcionais à área de intersecção das zonas com as células da grade.
 - Limites abruptos, nas regiões de fronteira e valores constantes no interior das unidades.
 - Os erros maiores quanto mais agrupada a distribuição da população e quanto menor as regiões de destino em relação às regiões de origem.

D. Population Densities in Ann Arbor Shown As Choropleths and As a Bivariate Histogram



Modelos de População “Univariados”

- Ponderado pela área
 - População agregada por município - representação matricial do mapa de densidade
 - Matriz do mapa de densidade



População, Espaço e Ambiente

Abordagens Espaciais em Estudos de População:
Métodos Analíticos e Técnicas de Representação

Parte IV – Integração de Dados

De áreas a superfícies: Método Pycnofilático de Tobler

Silvana Amaral

Antonio Miguel V. Monteiro

{silvana@dpi.inpe.br, miguel@dpi.inpe.br}



Referência



Tobler, W. R. Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association*, 74, 519-530, 1979

Geographical Interpolation

Waldo Tobler
Professor Emeritus
Geography Department
University of California
Santa Barbara, CA 93106-4060
<http://www.geog.ucsb.edu/~tobler>

- CSISS Summer Institute, 2001



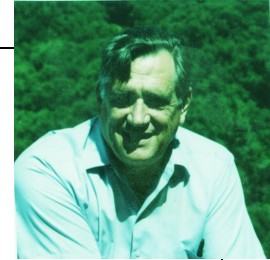
Exploring Geography Cartographically... (58)

http://www.ncgia.ucsb.edu/projects/tobler/Explore_Geography/sld032.htm

Amostragem aleatória seguida por interpolação seriam eficientes para representar esta superfície ??



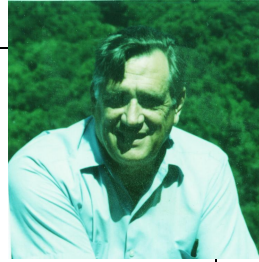
Pycnophylactic Reallocation



- Permite a produção de mapa de densidade ou de contorno a partir de dados de área
- É uma REALOCAÇÃO – ou um operador de desagregação
- Útil para a conversão de dados de um conjunto de unidades estatísticas para outro, como por exemplo de setores censitários para distritos escolares.

“My assertion is that it may actually improve the data.”

Densidade Populacional / município

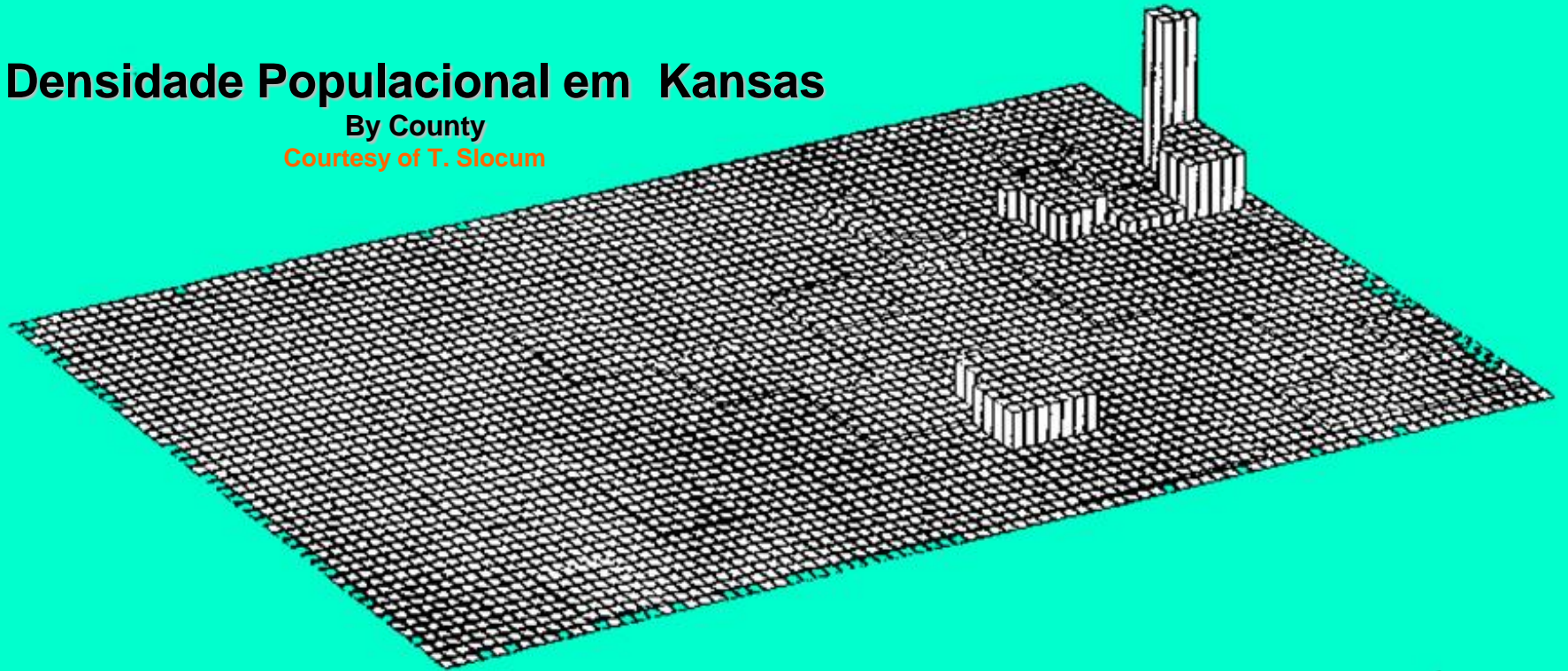


- Observe as discontinuidades dos limites municipais.
- Precisa de um mapa de densidade de população suavizado para desenhar os contornos.
- Procedimentos de interpolação usuais não vão funcionar a menos que se use centróides e esta “**Ficção**” permitiria a mudança das pessoas de um município para outro

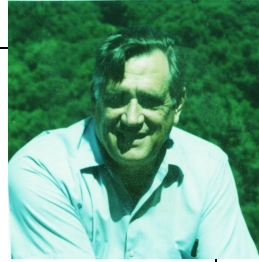
Densidade Populacional em Kansas

By County

Courtesy of T. Slocum



Densidade Populacional / município



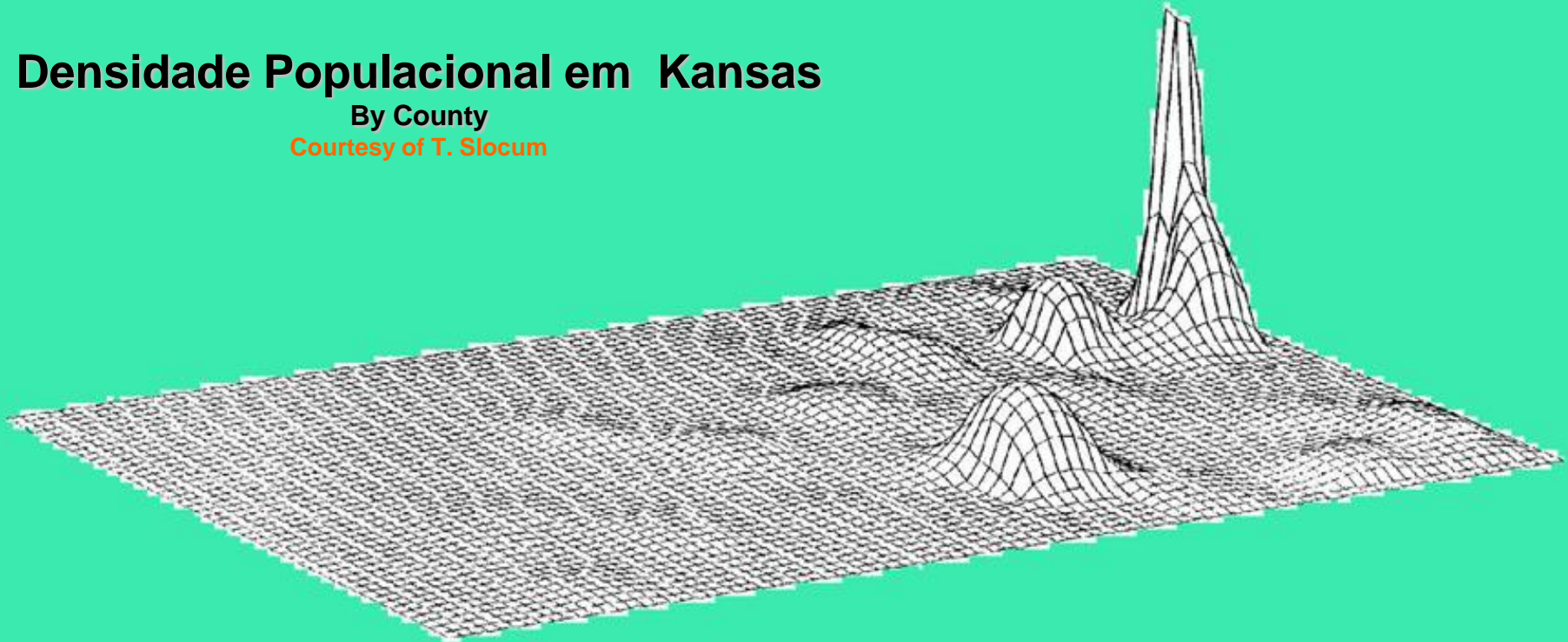
Superfície contínua suave, com população redistribuída
“pynophylactically”

Cada município ainda contém o mesmo número de pessoas

Densidade Populacional em Kansas

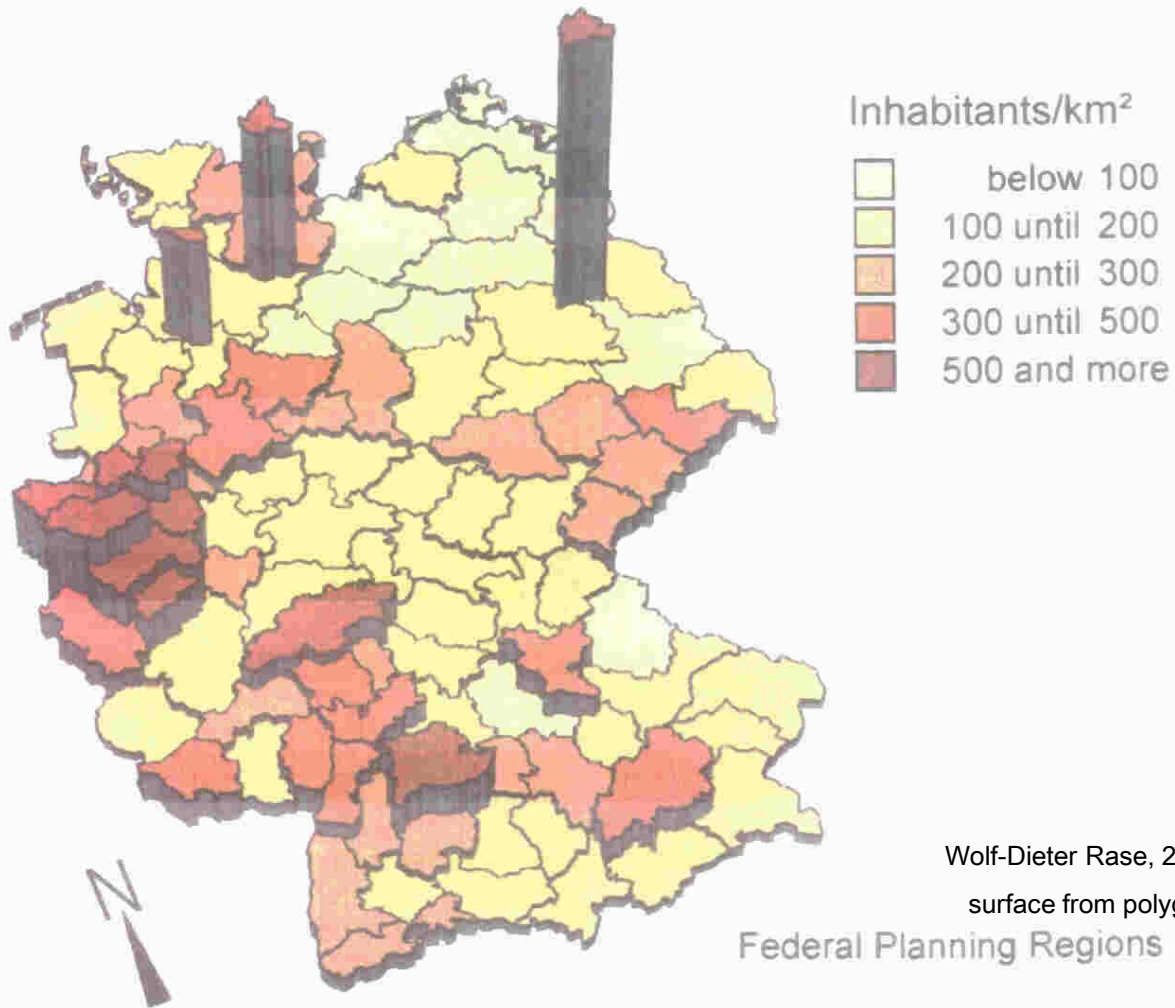
By County

Courtesy of T. Slocum



Exemplo 2

- Dados por regiões de planejamento federal para Alemanha
- Vista perspectiva do histograma bivariado

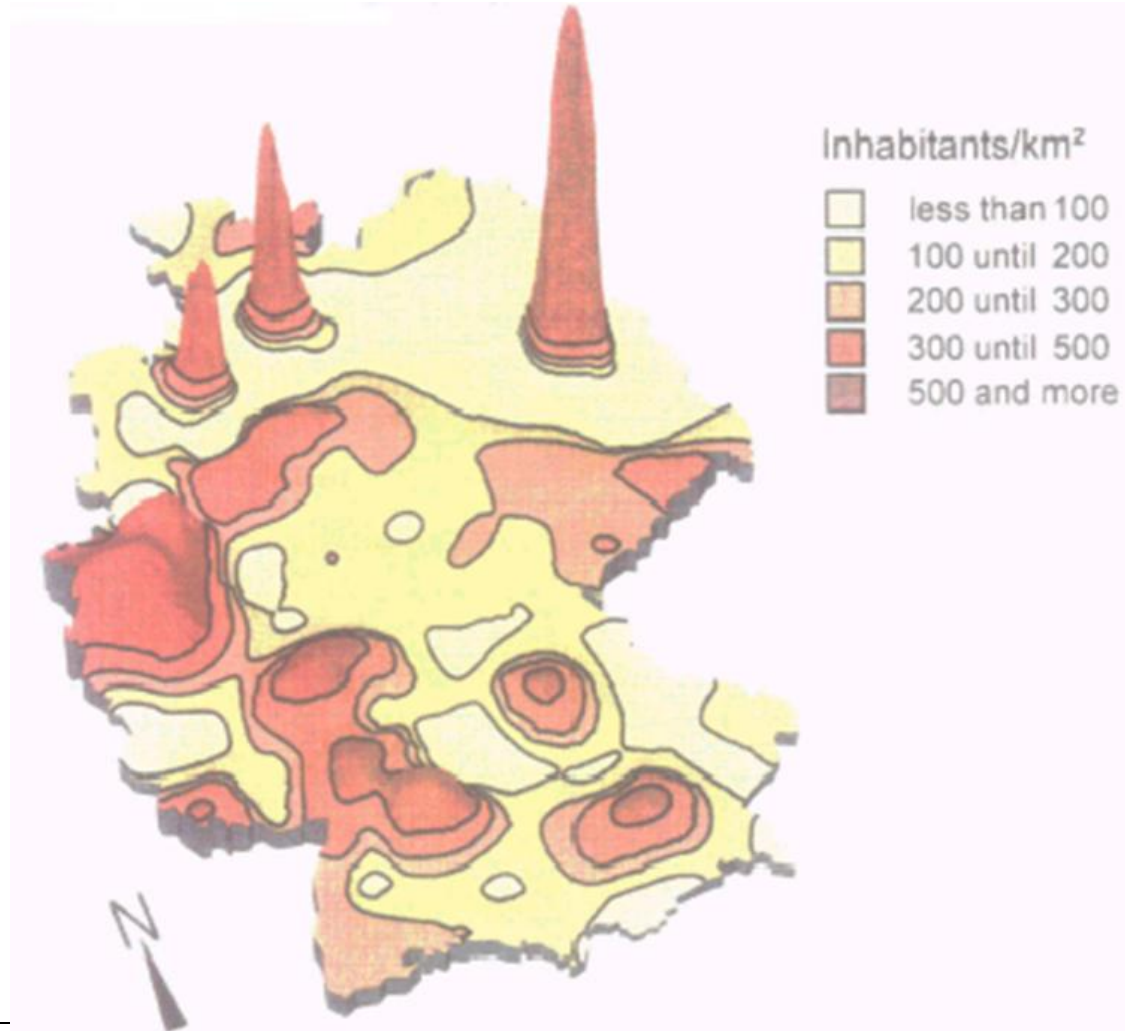


Cortesia de Wolf Rase in Bonn..

Wolf-Dieter Rase, 2001, "Volume-preserving interpolation of a smooth surface from polygon-related data", J. Geograph. Syst, 3:199-213.

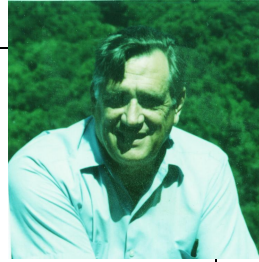
Exemplo 2

- Dados por regiões de planejamento federal para Alemanha
- Distribuição de densidade de população contínua

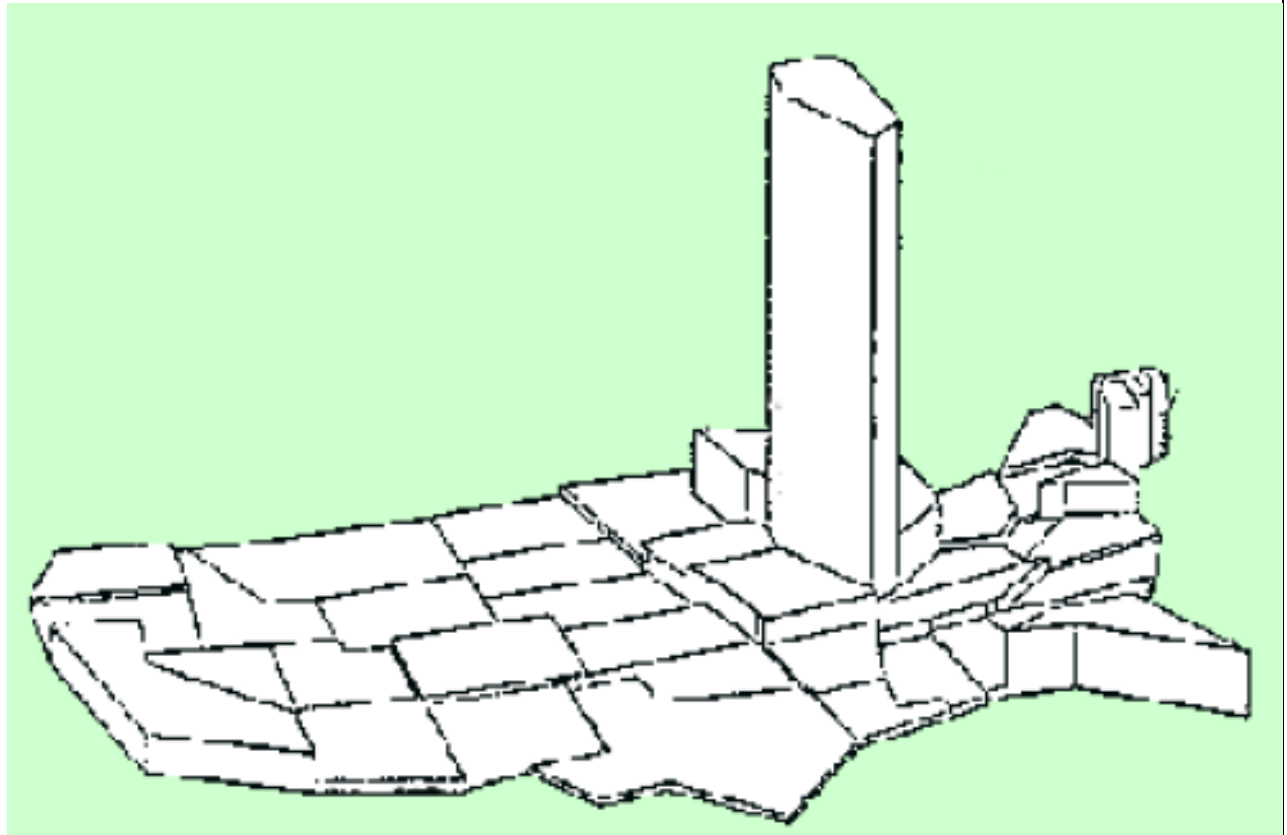


Waldo Tobler

Exemplo 3

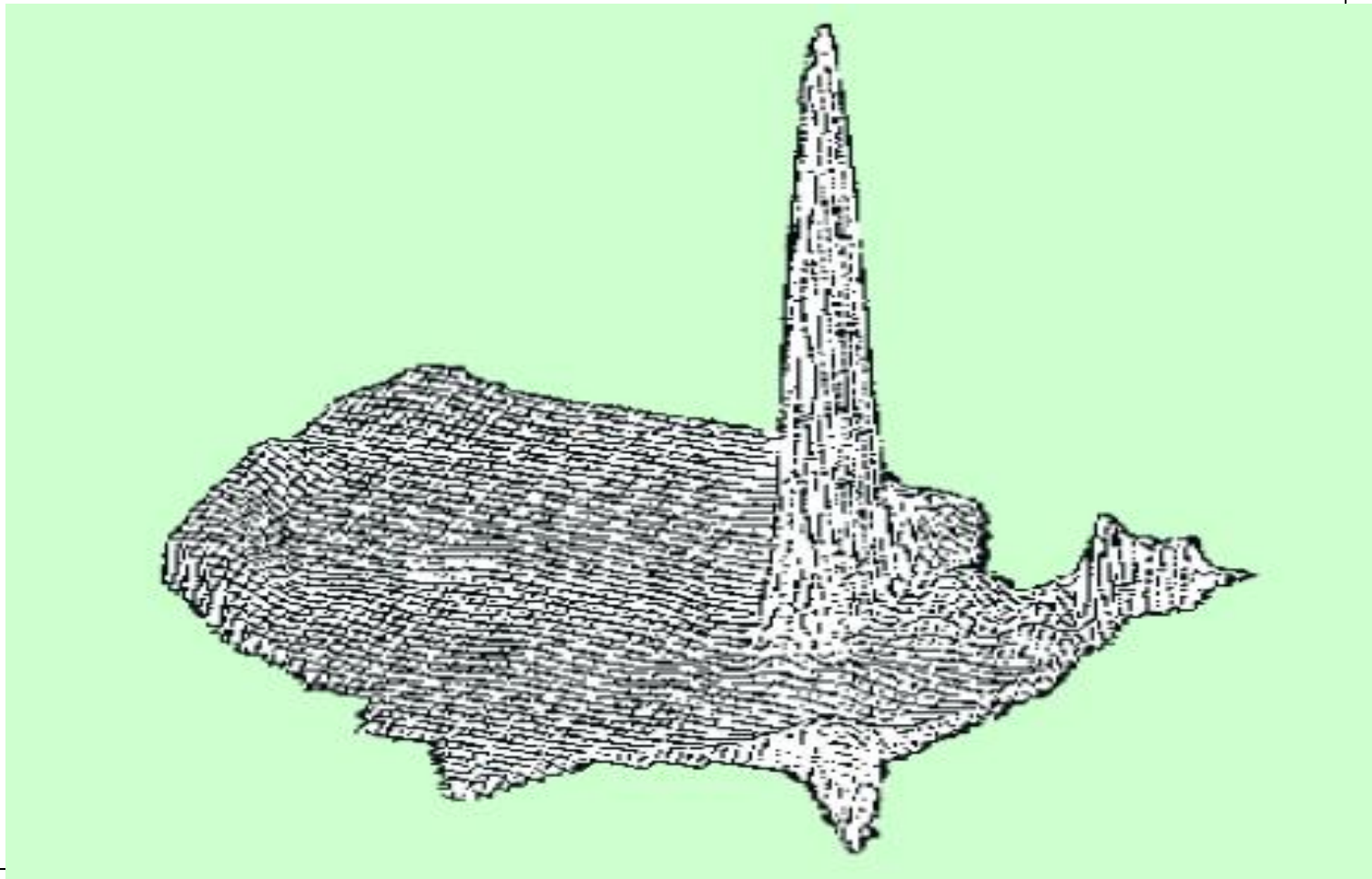
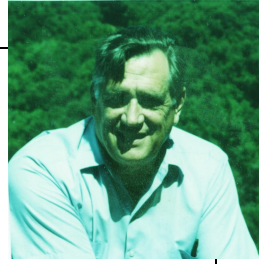


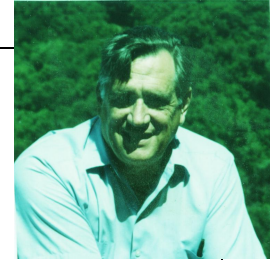
- Base nos contornos do Estado e volumes de acordo com a emigração de Illinois
- A maioria dos migrantes em Illinois são realocados no interior do próprio Estado



Exemplo 3

- Superfície suave interpolada pycnophylactically
- Pode ser particionada para produzir migração estimada por regiões arbitrárias - a bacia dos grandes lagos por exemplo

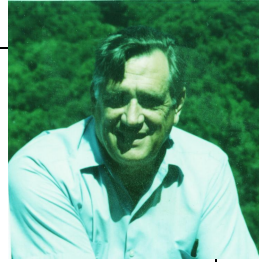




“How Pycnophylactic Reallocation Works”

- Baseado na noção de que as pessoas são gregárias, influenciam umas às outras, se movimentam, e tendem a congregar-se
- ➔ lugares vizinhos e adjacentes tendem a ser similares
- Matematicamente se traduz em um **critério de suavização** (*smoothness criterion*) com pequenas derivadas parciais.
- Aplica-se a qualquer dados que exiba **autocorrelação espacial positiva**.

Realocação de dados de área - preservar massa



1º) Definir a condição primária para a preservação de massa.

É a condição de “invertibilidade” necessária para todo método de redistribuição de informação / área, dada por:

$$\iint_{R_i} f(x, y) dx dy = V_i$$

Para todo i , onde V_i é o valor (população) na região R_i (polígono)

2º) Limitar a superfície resultante a ser suavizada requisitando que os locais vizinhos tenham valores similares.

É um pressuposto sobre o processo demográfico espacial, uma forma de capturar a noção de que a maioria das pessoas são “gregárias”

A forma mais simples, a suavização **Laplaciana** é obtida minimizando a equação:

$$\iint_R \left(\frac{\partial f^2}{\partial x} + \frac{\partial f^2}{\partial y} \right) dx dy$$

onde R é o conjunto de todas as regiões. A condição dos limites é:

$$\frac{\partial f}{\partial \eta} = 0$$

O mínimo da integral



- A solução da equação de suavização integral é dada pela equação de Laplace

$$\nabla^2 X + \nabla^2 Y = 0$$

- Isso significa que os locais vizinhos têm valores similares, ou num raster, ou matriz – que o valor central é a média dos que estão no entorno.
 - Se traduz em um algoritmo computacional

Traduzindo a Matemática



- Imagine que cada unidade é composta por argila colorida, cores diferentes para cada unidade
- O volume da argila representa o número de pessoas, e a altura representa a densidade.
- Para obter uma densidade suavizada usa-se uma espátula, mas nenhuma argila pode ser movida de uma para outra unidade.
- Não é permitido misturar cores!



Smoothing – processo iterativo

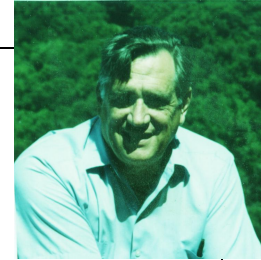


1 – Rasterizar a região;

Suavização é feita sobre este raster, mantendo a população

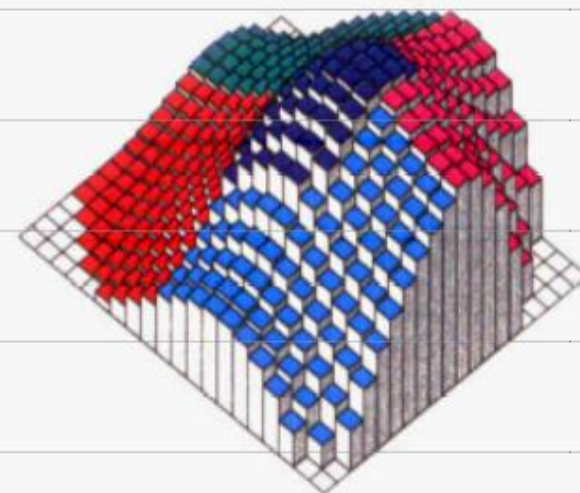
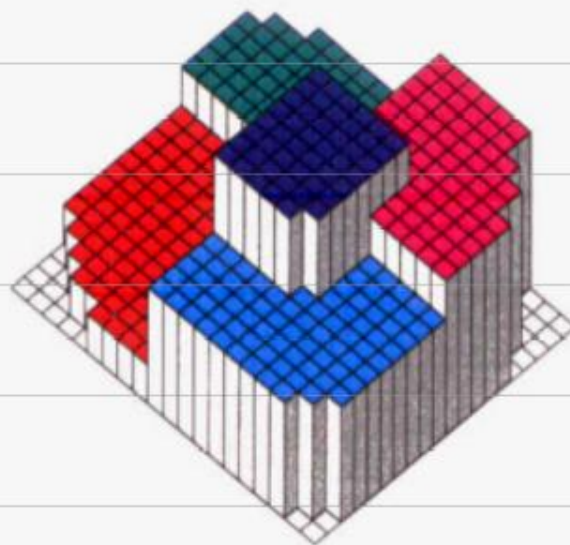
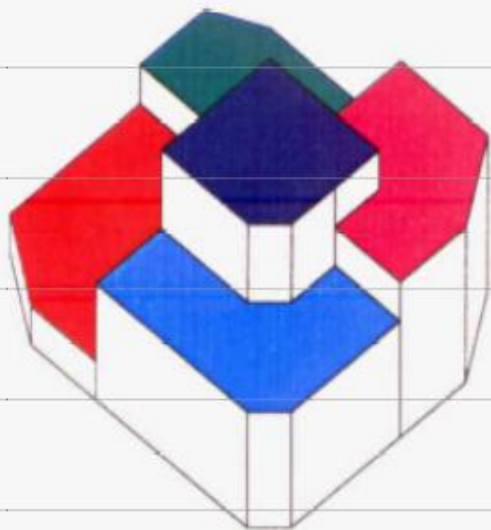
2 – O # de iterações depende do tamanho da maior região, em unidades raster

- A suavização deve cruzar de borda a borda da maior região
- Quanto mais detalhada grade (raster), maior a resolução e o tempo de iteração



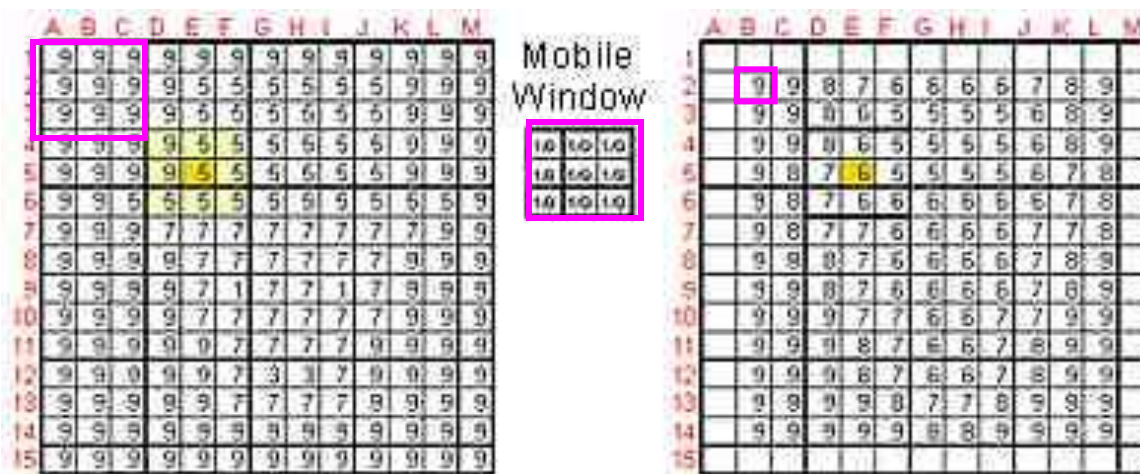
Left to right

1. Dado poligonal
2. Dado matricial (Rasterized)
3. Superfície suavizada (Smoothed)



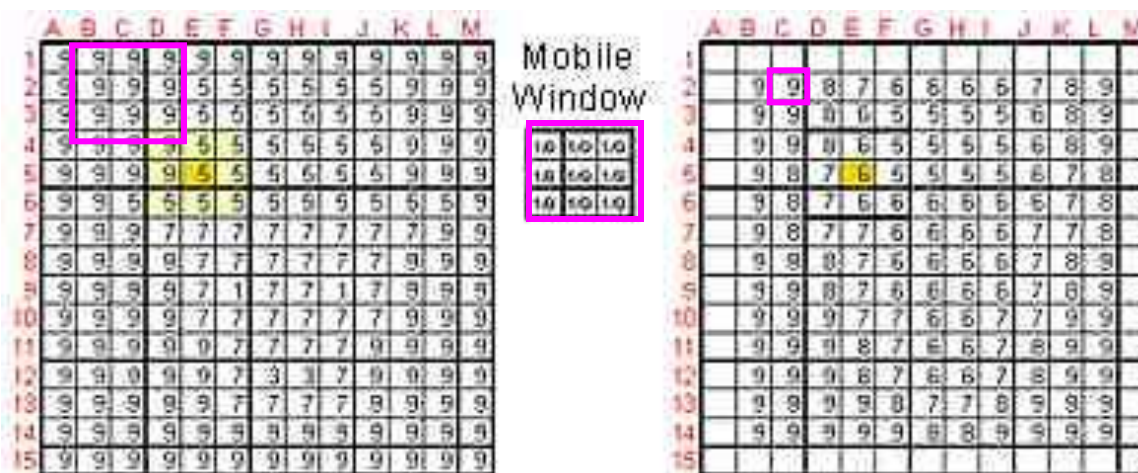
- **Picnofilático de Tobler**

- Superfície baseada nos centroides geométricos das unidades censitárias
- Superfície suave ~ “filtro de média”



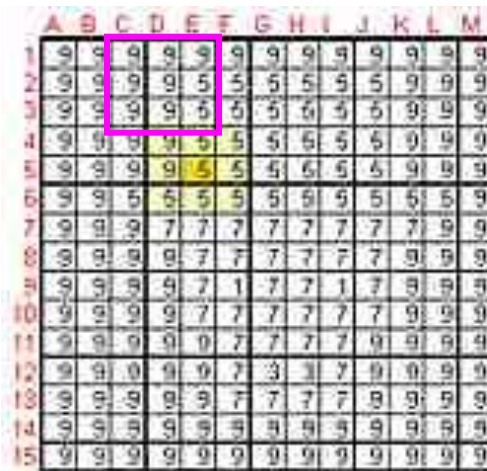
- **Picnofilático de Tobler**

- Superfície baseada nos centróides geométricos das unidades censitárias
- Superfície suave ~ “filtro de média”



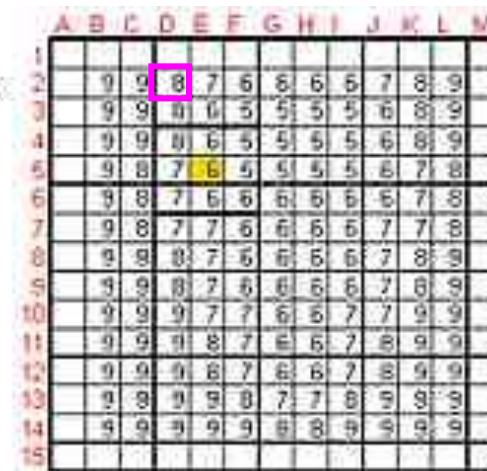
- **Picnofilático de Tobler**

- Superfície baseada nos centróides geométricos das unidades censitárias
- Superfície suave ~ “filtro de média”



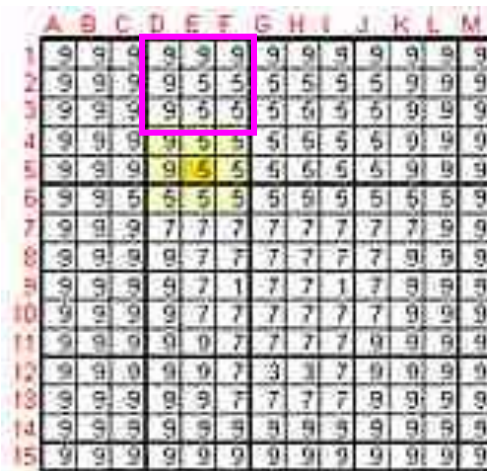
Mobile Window

10	10	10
10	10	10
10	10	10



- **Picnofilático de Tobler**

- Superfície baseada nos centróides geométricos das unidades censitárias
- Superfície suave ~ “filtro de média”



Mobile Window

10	10	10
10	10	10
10	10	10



- **Picnofilático de Tobler**

- Superfície baseada nos centróides geométricos das unidades censitárias
- Superfície suave ~ “filtro de média”

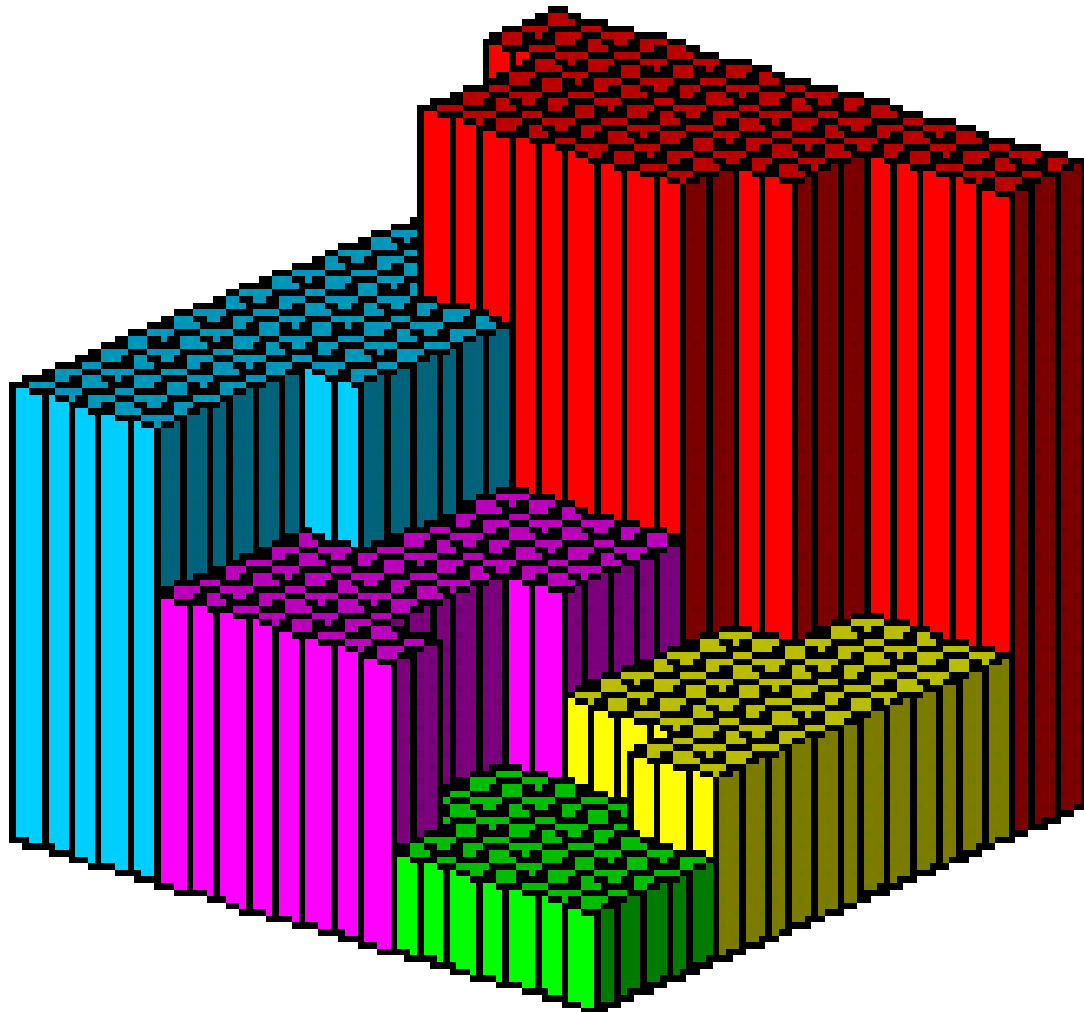


Mobile Window

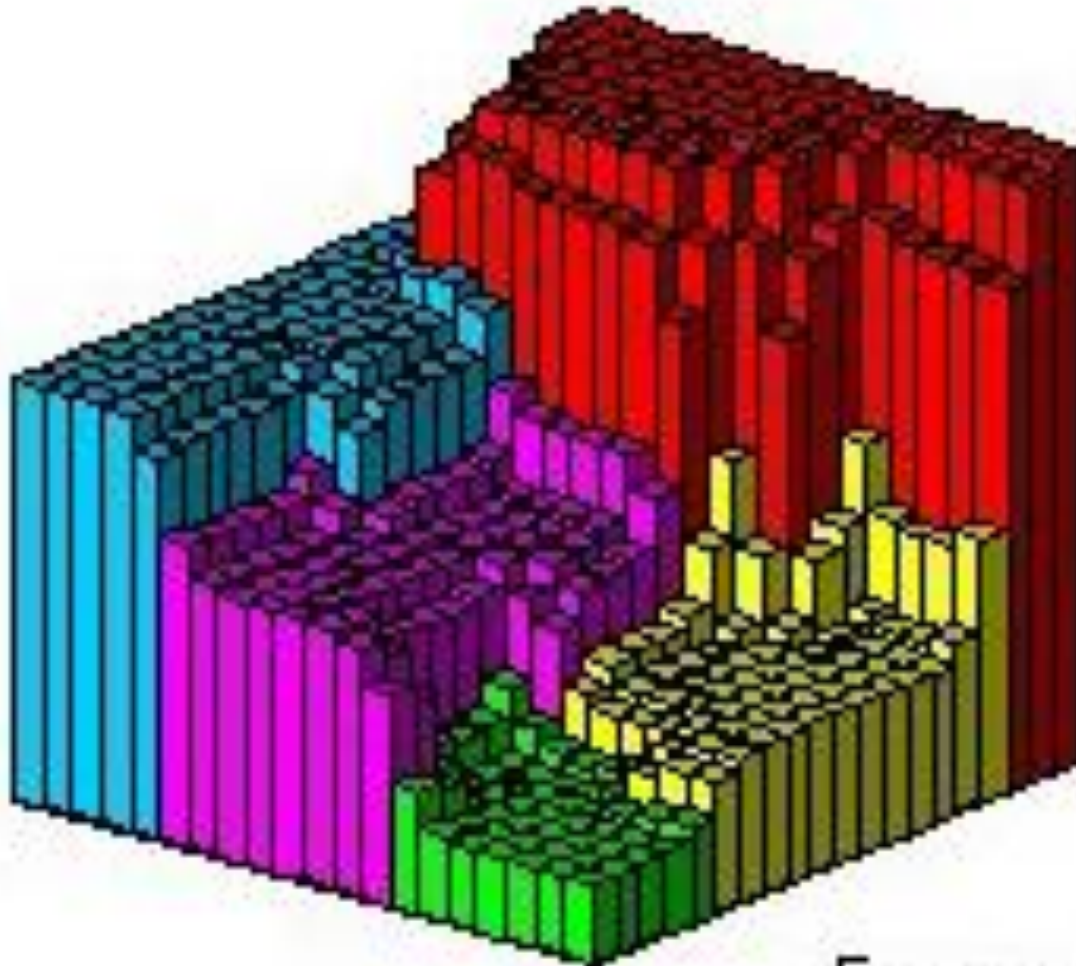
10	10	10
10	10	10
10	10	10



Zero Iterations

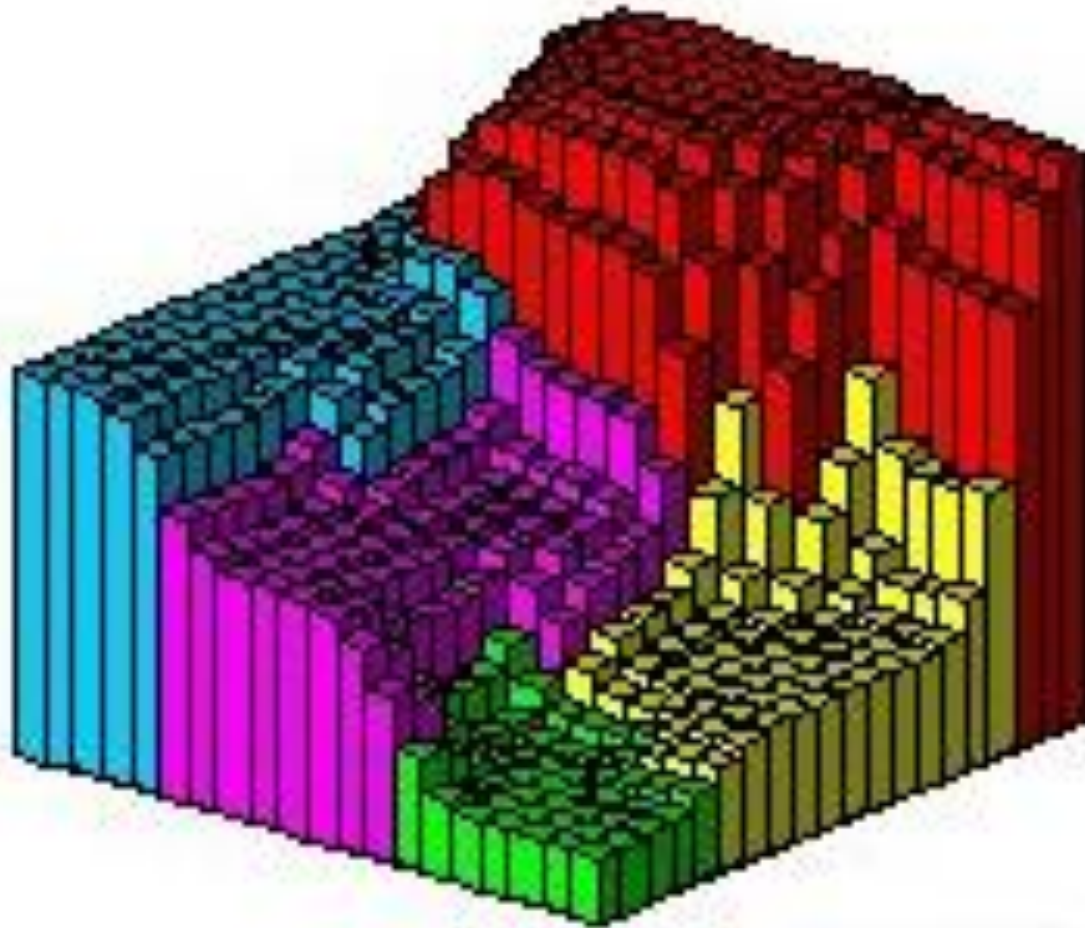


5 Iterations



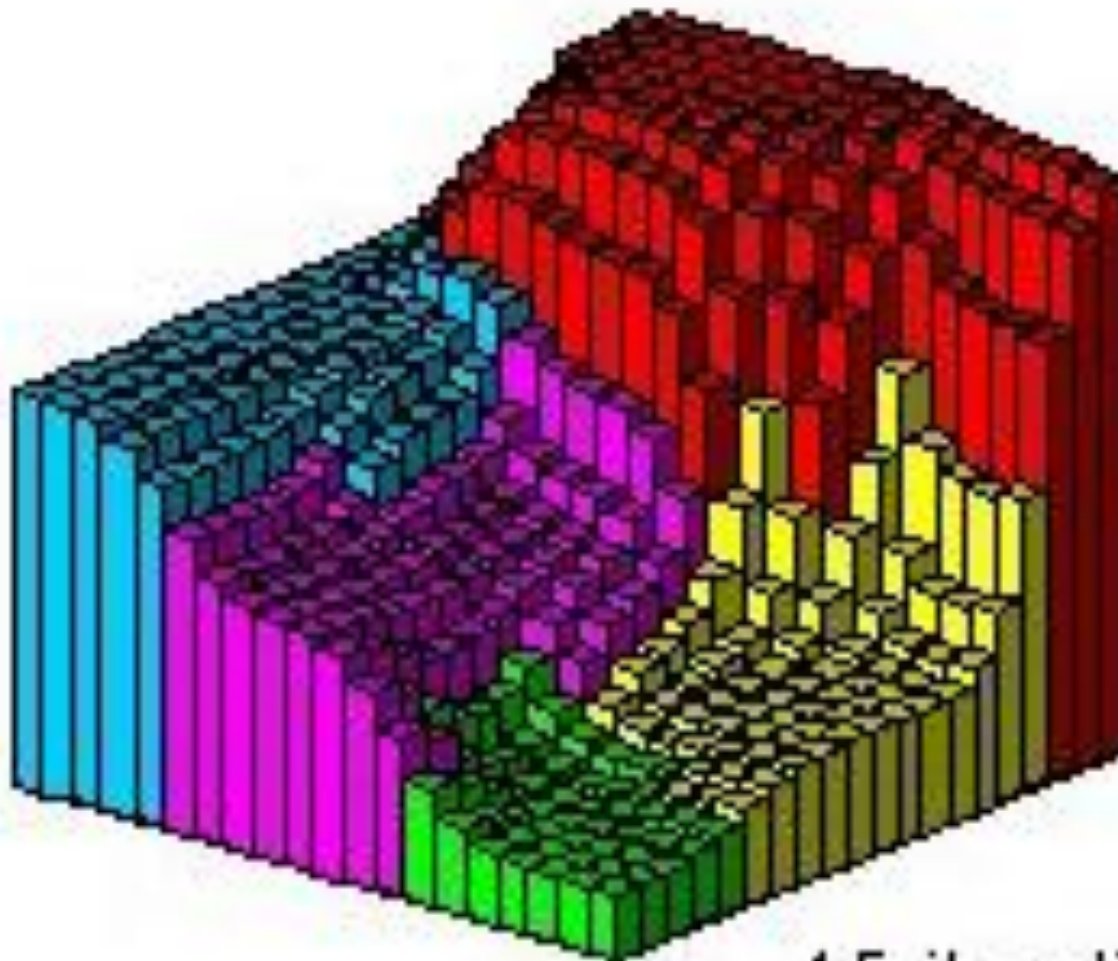
5 iterations

10 Iterations



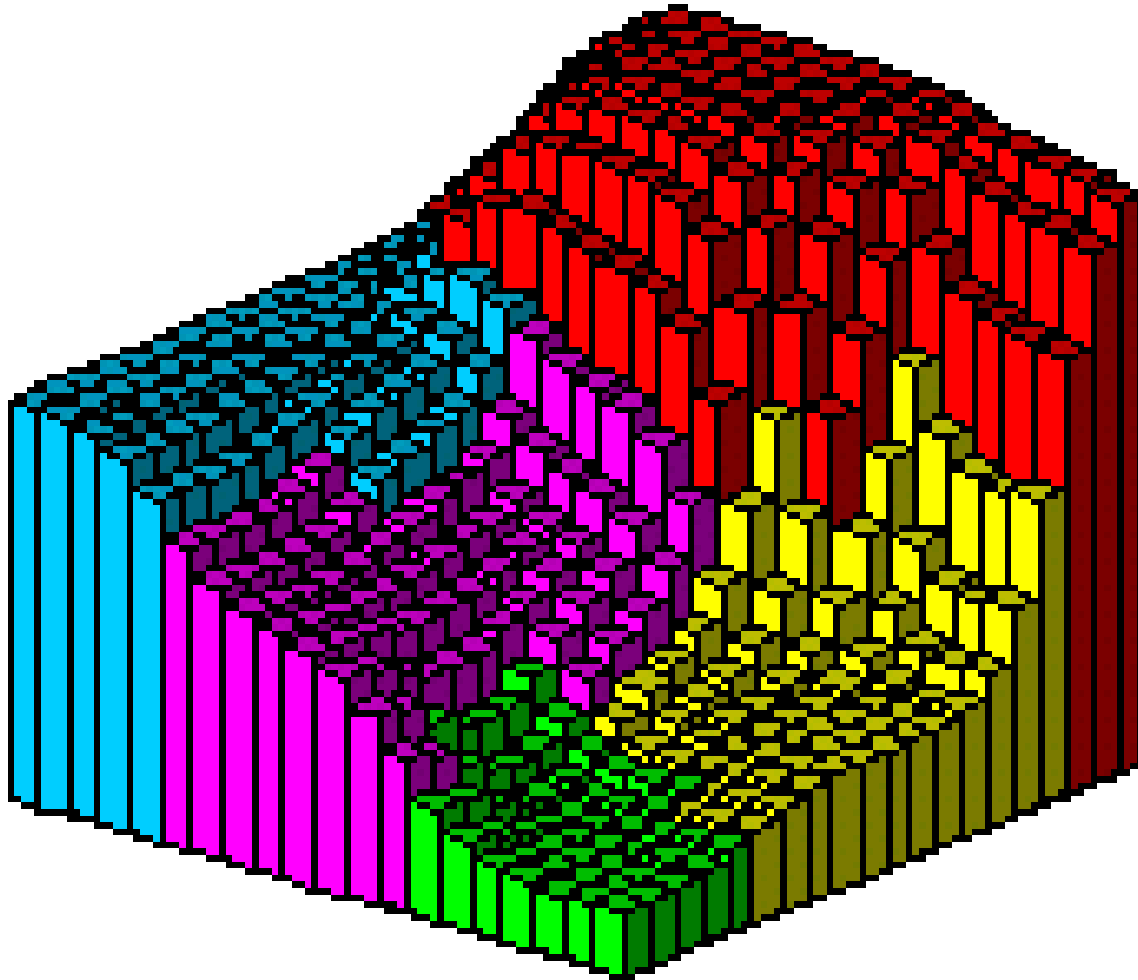
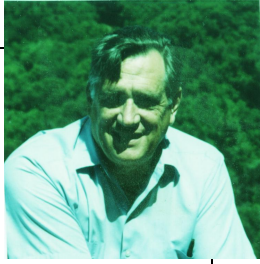
10 iterations

15 Iterations

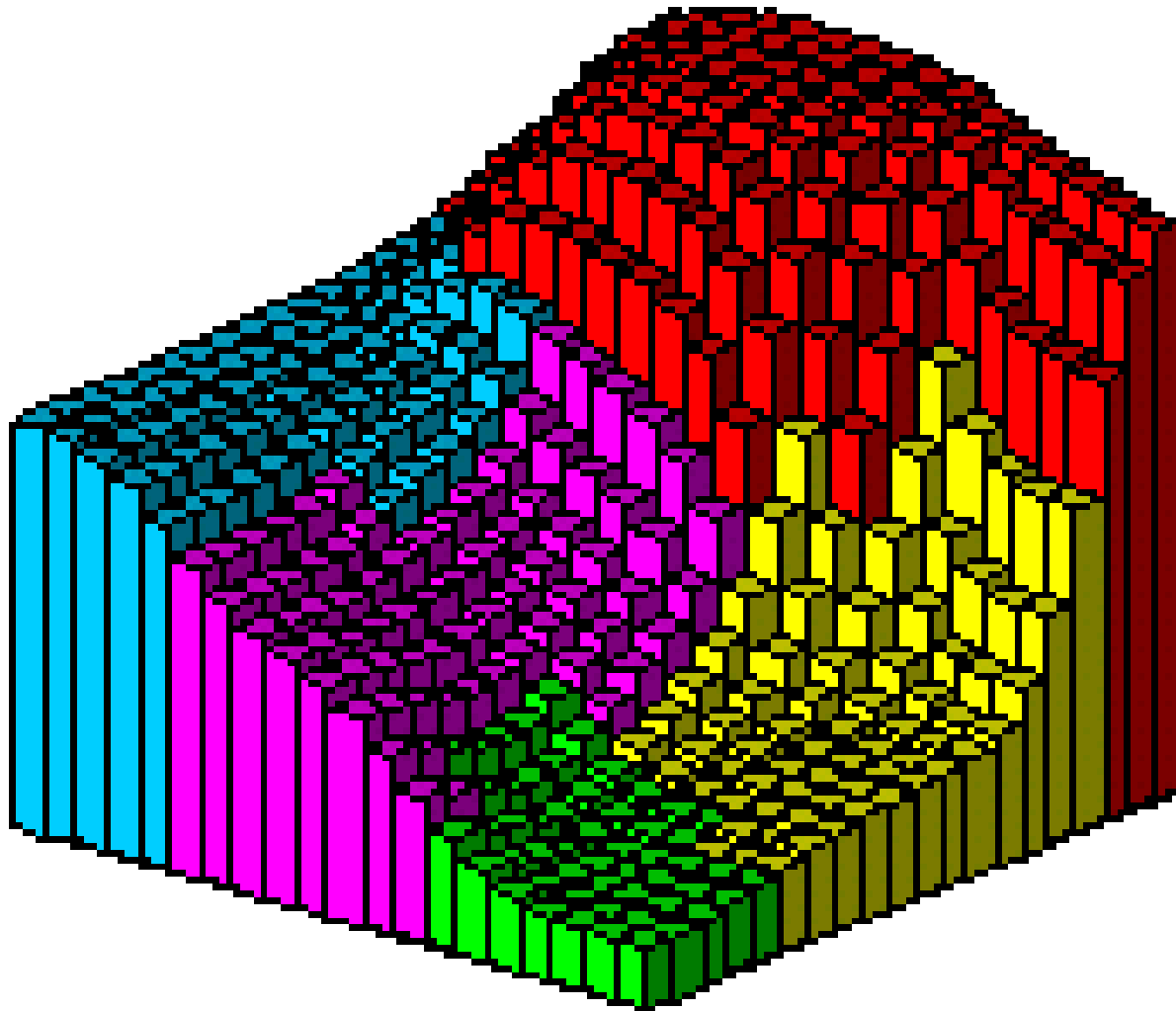


15 iterations

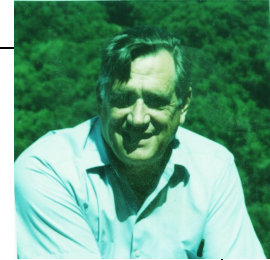
20 Iterations



25 Iterations



25 iterations

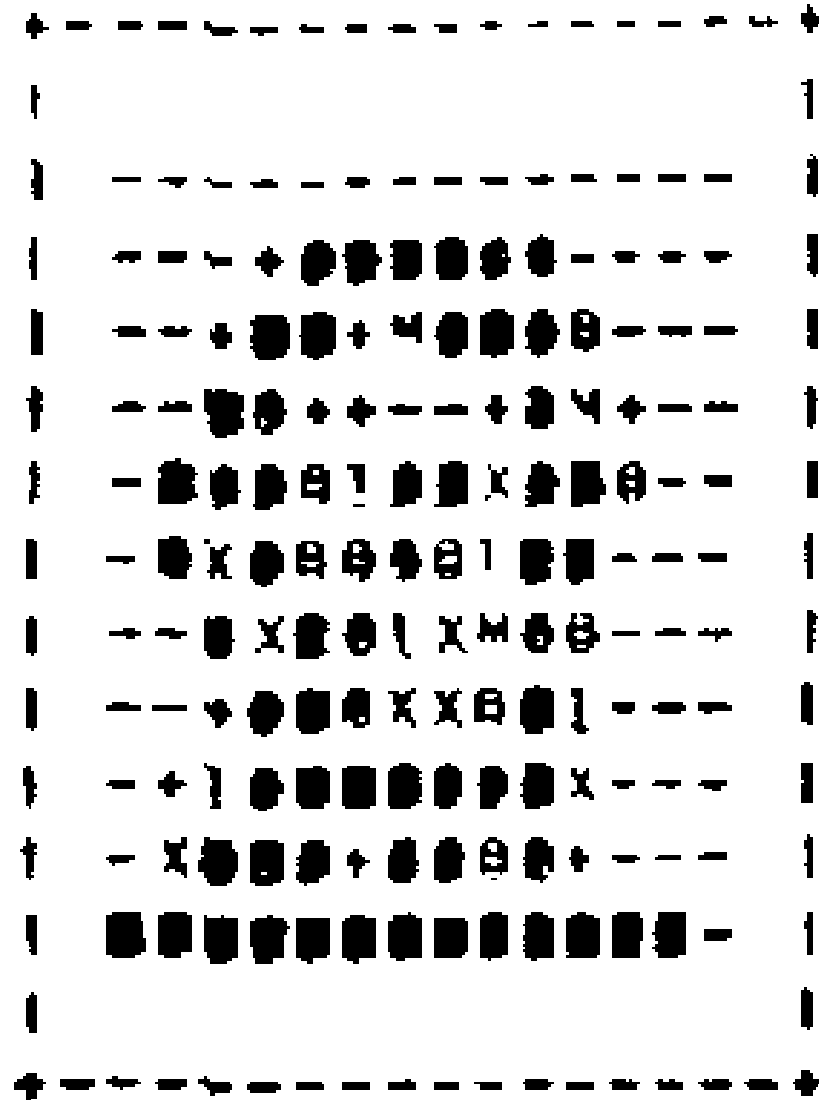


Realocação Pycnophylactic também funciona para dados coletados em células/grades

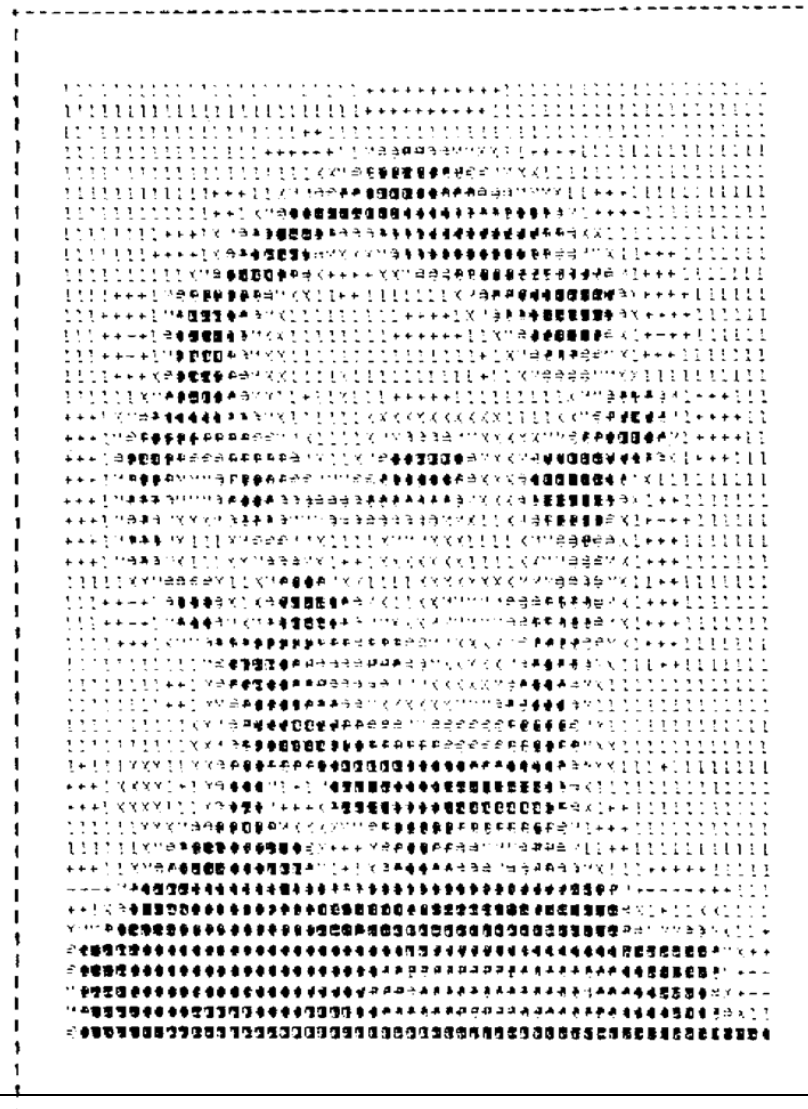
(não foi desenhado para este objetivo)

- Ex. Dados em Pixels
- Não entre pixels - o que resultaria em efeitos diferentes
- Mas valores nas vizinhanças dos pixels são considerados para um pixel, pelo critério de suavização

Ex Imagem processada (20 by 14)



Quadruplicada (80 x 56) mas com a mesma "massa" total



Condições de limites e suavização



Pode usar diferentes critérios de suavização:

- **LaPlacian e biharmonic.**
- Assim como para a equação diferencial parcial, é necessário especificar as condições de contorno/limite
 - A Condição **Dirichlet** especifica o valor no limite, e
 - A condição de **Neumann** especifica o gradiente no limite.

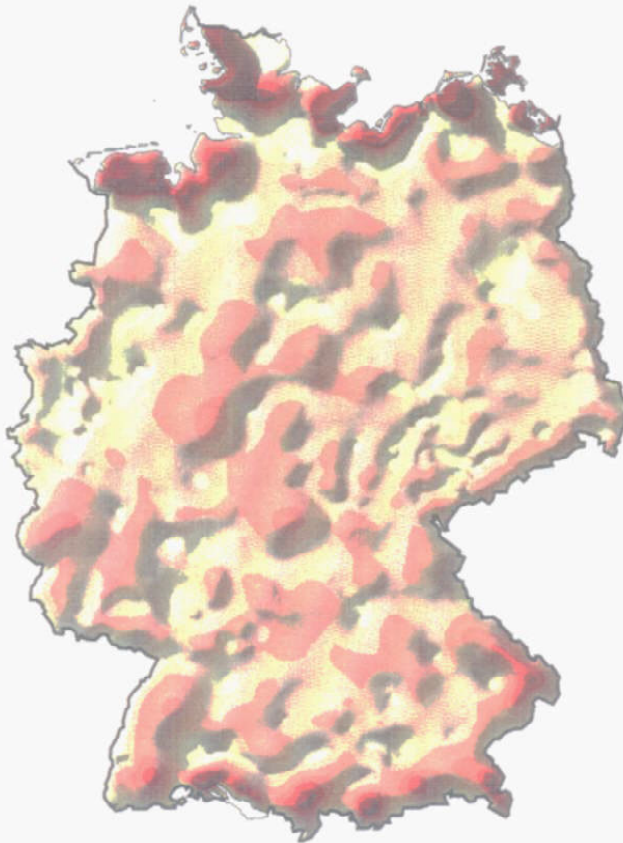
LaPlacian & Biharmonic Smoothing

Dirichlet Boundary Condition



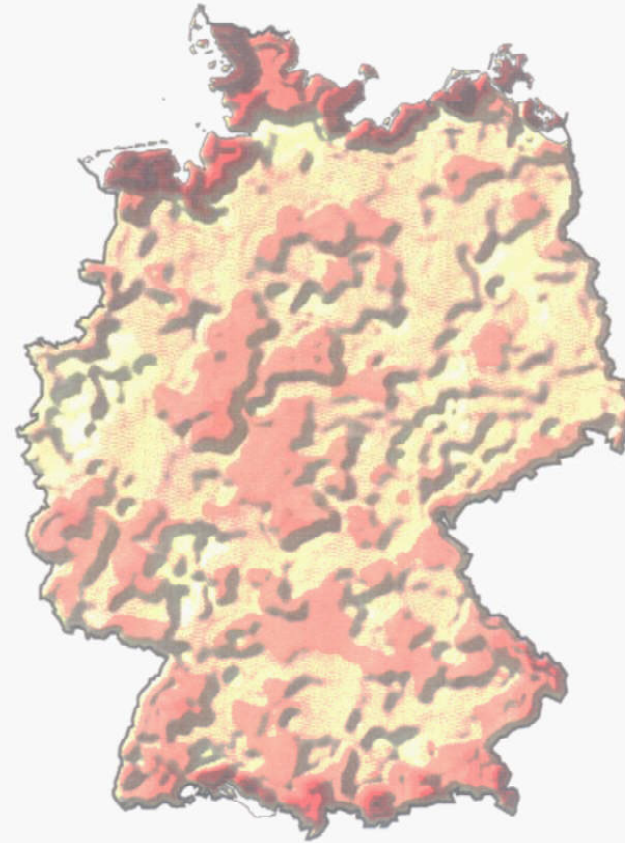
Landschaftliche Attraktivität

Suavização
Contorno: Laplace
Dirichlet
Iterationen: 100
Restfehler: 0.3142 %



a) Stetigkeit nach Laplace-Gleichung

Suavização
Contorno: Biharmonisch
Dirichlet
Iterationen: 50
Restfehler: 0.0099 %



b) Stetigkeit nach biharmonischer Gleichung

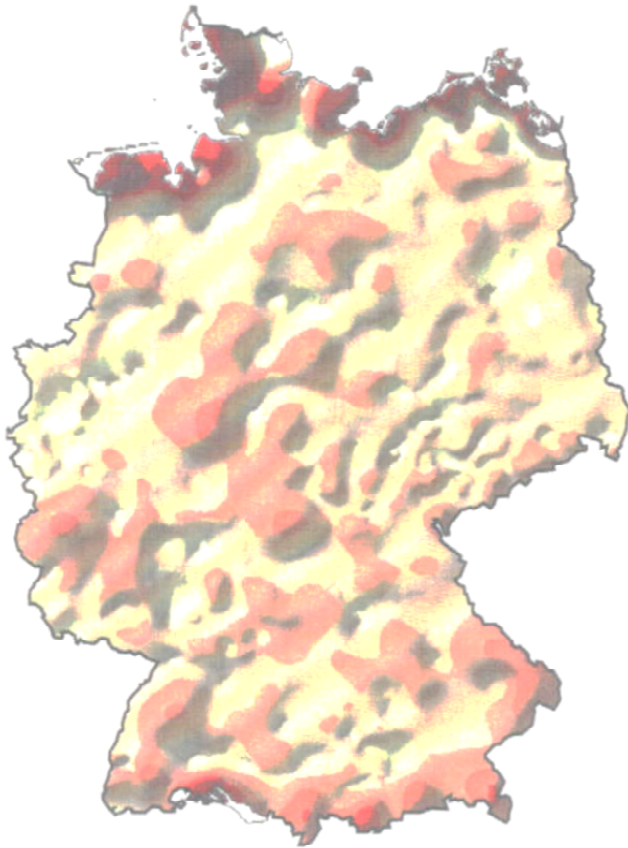
LaPlacian & Biharmonic Smoothing

Neumann Boundary Condition



Landschaftliche Attraktivität

Suavização
Contorno: Laplace
Neumann
Iterationen: 100
Restfehler: 0.0123 %



a) Laplace-Gleichung, Neumann-Grenze

Suavização
Contorno: Biharmonisch
Neumann
Iterationen: 13
Restfehler: 0.0099 %



b) Biharmonische Gleichung, Neumann-Grenze

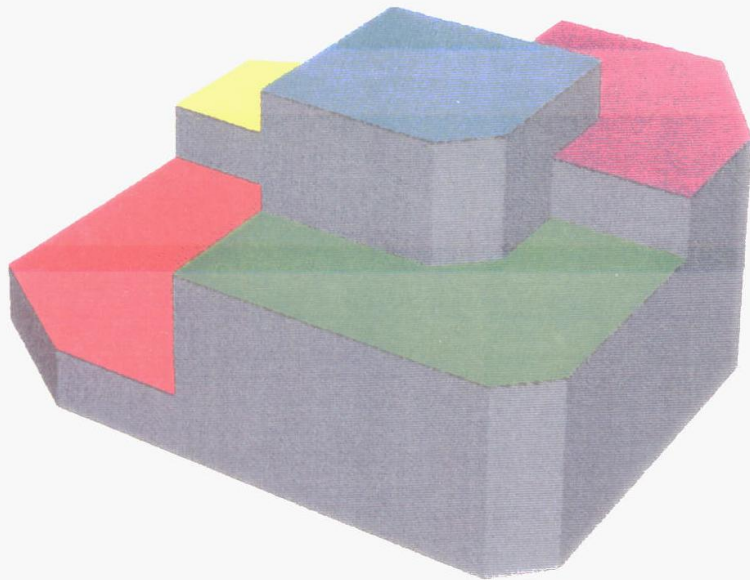
Abb. 2-30 Pyknophylaktische Interpolation mit Neumann-Grenzbedingung

Elementos Finitos também funcionam...

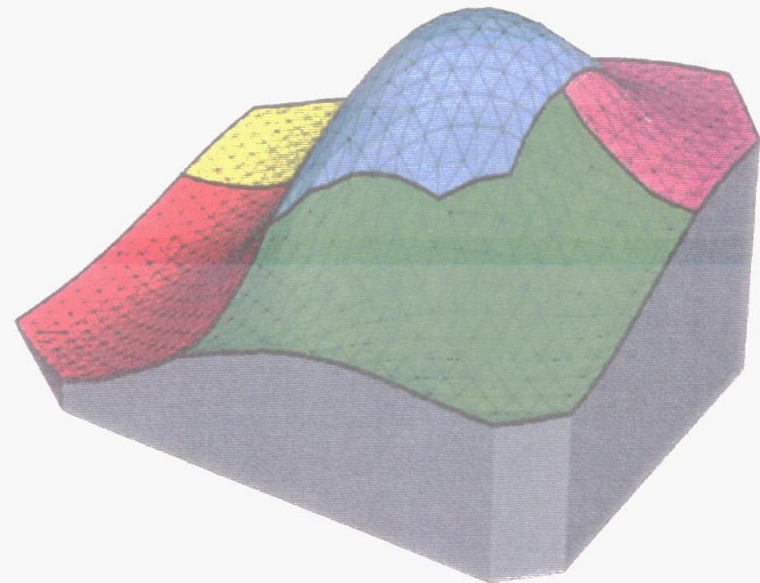


Superfície Original e suavizada usando triângulos

Wolf-Dieter Rase, 2001, "Volume-preserving interpolation of a smooth surface from polygon-related data", J. Geograph. Syst., 3: 199-213



a) Perspective view of choropleth map



b) Regular grid of triangles

Pode incluir outras fontes de dados



- Imagens ou foto aéreas

- podem ser apropriadas para redefinir polígonos, excluindo-se algumas áreas.

- No caso do fenômeno ser população:

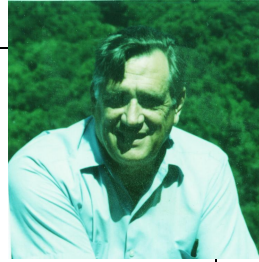
- Lagos, florestas e zonas industriais devem ser isentos de população

- Neste caso, usar condições de Dirichlet para os limites destas áreas de valor zero, e depois proceder com os novos polígonos como usual.

- Pode-se refinar o procedimento para incluir polígonos com valores de população diminuídos ou aumentados

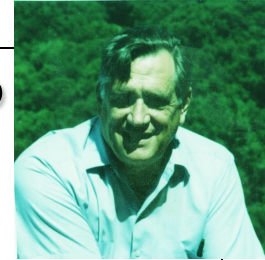
Em cartografia conhecidos como **métodos dasimétricos**.

Vantagens da Preservação de Massa



- Problemas de realocar observações em unidades de coleta diferentes, quando os dois conjuntos não são nem aninhados nem compatíveis
 - Ex. converter o # de crianças observado por setor censitário em um município para uma contagem por distrito escolar.
 - Limites mudam com o tempo, requerem realocação para compatibilização
 - Valores de densidade obtidos pelo método pycnophylatic permitem uma estimativa simples.
 - Um “cookie cutter” pode cortar a superfície de contínua, em novas zonas e fazer para obter a soma do valor.





Faz diferença?

Interpolação procedimento	Desvio Absoluto			Relativa
	Valor da Variável (%)			Variância
	MIN	MÉD	MAX	%
Malha, sem particionamento				
Shepard	0.01	10.65	1013.39	3.51
Hardy (Q = 0,5, R = 0,21)	0	11.07	1515.45	4.23
Mesh, casco convexo				
Renka, linear	0	17.39	3192.25	10.66
Renka, global Gradientenschätzung	0.01	13.26	1113.36	6.38
Renka, local Gradientenschätzung	0.04	14.65	2463.1	7.79
Renka com QUADSF (Spath)	0	12.77	998.08	5.95
Montefusco Casciola / Akima	0.01	11.52	1114.73	4.82
	0	11.59	1389.1	5.08
Grille, casco côncavo				
Renka global Gradientenschätzung	0.01	10.45	1112.81	3.6
Renka, local Gradientenschätzung	0.01	12.32	2658.03	6.49
Renka com QUADSF (Spath)	0	10.93	997.6	4.81
Continuação divisão do TIN				
Shepard	0.03	6.64	2147.73	0.52
Renka global Gradientenschätzung	0.03	2.75	2644.66	0.16
Renka, local Gradientenschätzung	0.01	0.58	4111.18	0.16
Renka com QUADSF (Spath)	0.01	4.13	2632.2	0.24
Pyknophylaktische Interpolação	0.1	1.89	364.18	0.02

- Poucas comparações entre realocação de área com preservação de massa e interpolação baseada em pontos
- Tabela Wolf Rase – German Data

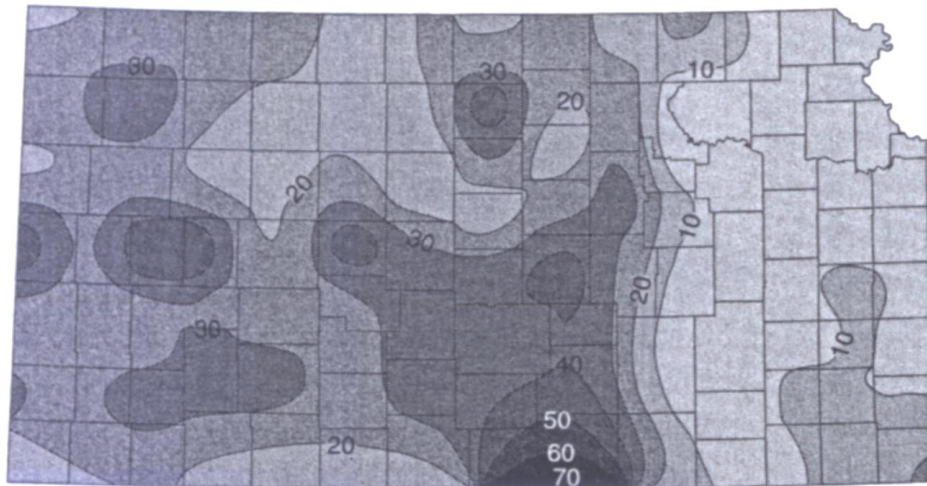
Comparação



Figure 8.13 (p. 150) of T. Slocum, “Thematic Cartography and Visualization”,
Prentice Hall, 1999.

Pycnophylactic reallocation

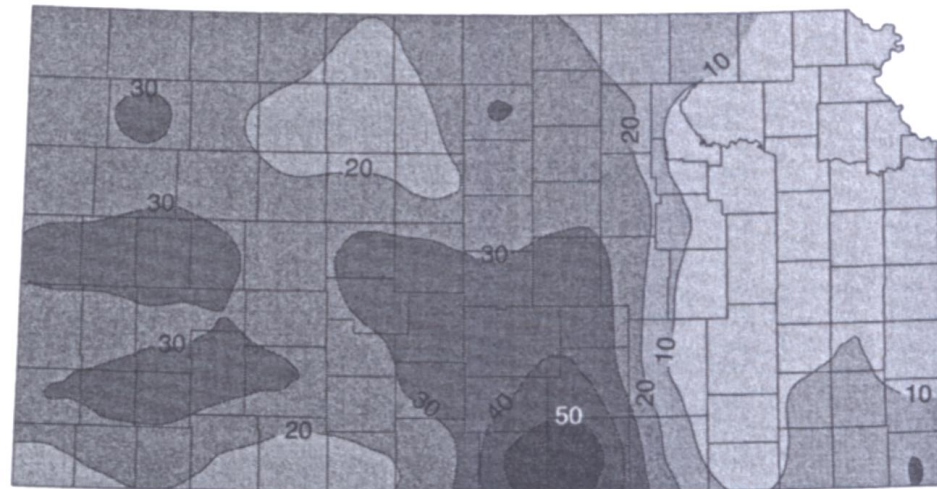
Wheat Harvested in Kansas, 1993
(pycnophylactic)



Contour Lines Represent Percent of Land Area

Punctual Kriging from Centroids

Wheat Harvested in Kansas, 1993
(kriging)



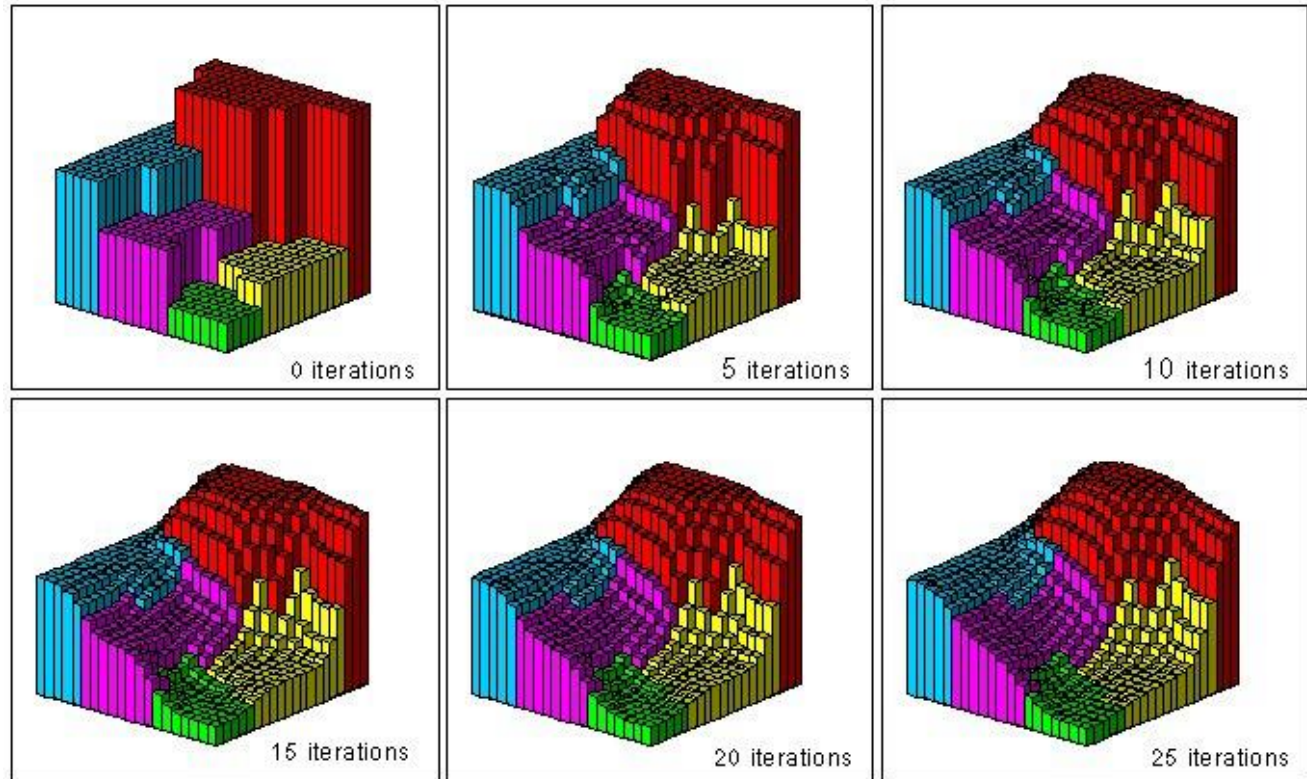
Contour Lines Represent Percent of Land Area

Modelos de População “Univariados”

Picnofilático de Tobler

- Pondera pela distância ao centróide, função de densidade populacional concêntrica em torno do centróide geométrico
- Indica população em toda a superfície (não há zeros)

Considera os valores das áreas adjacentes e mantém o total de população

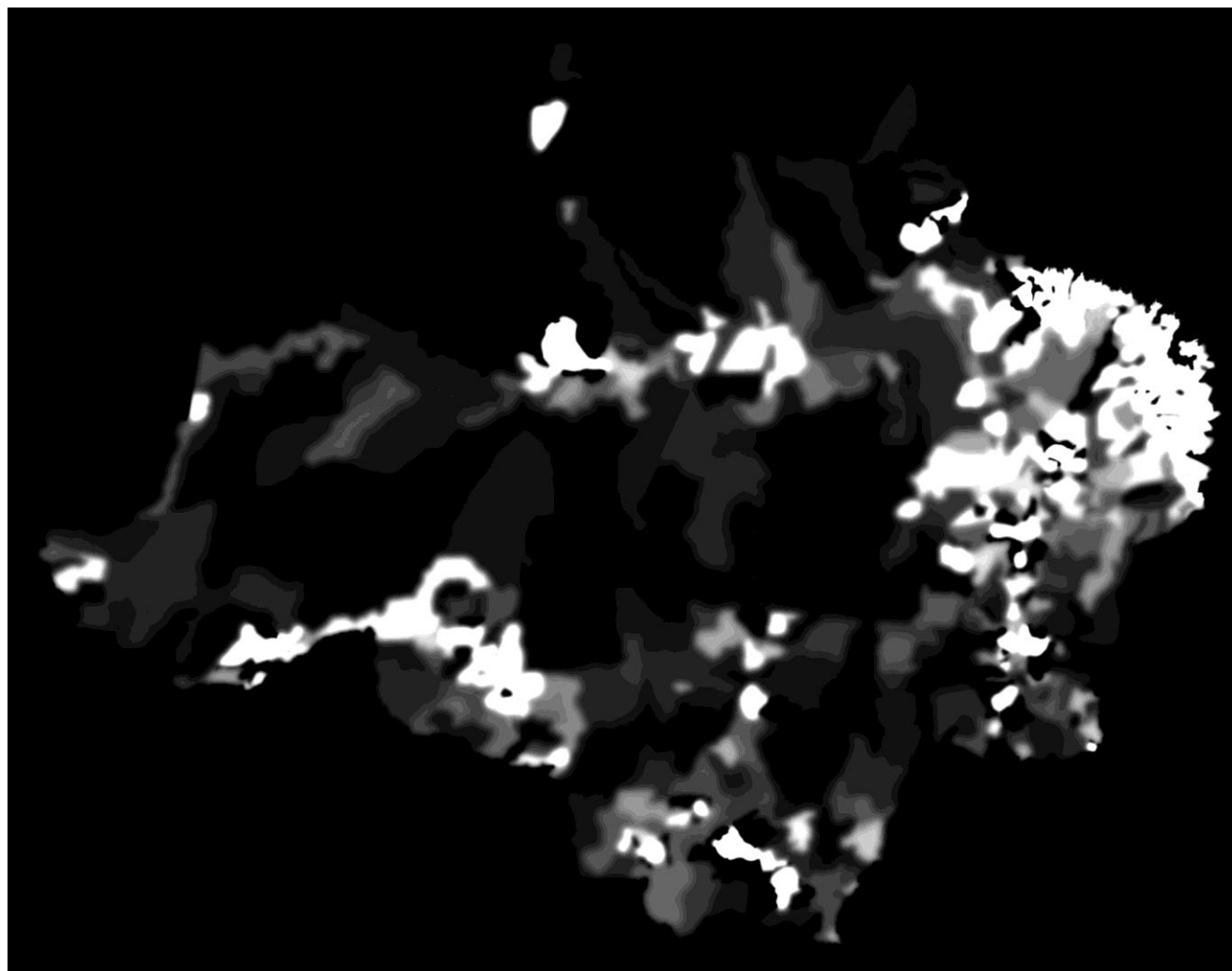


Modelos de População “Univariados”

Para Amazônia - 2000

Picnofilático
de Tobler

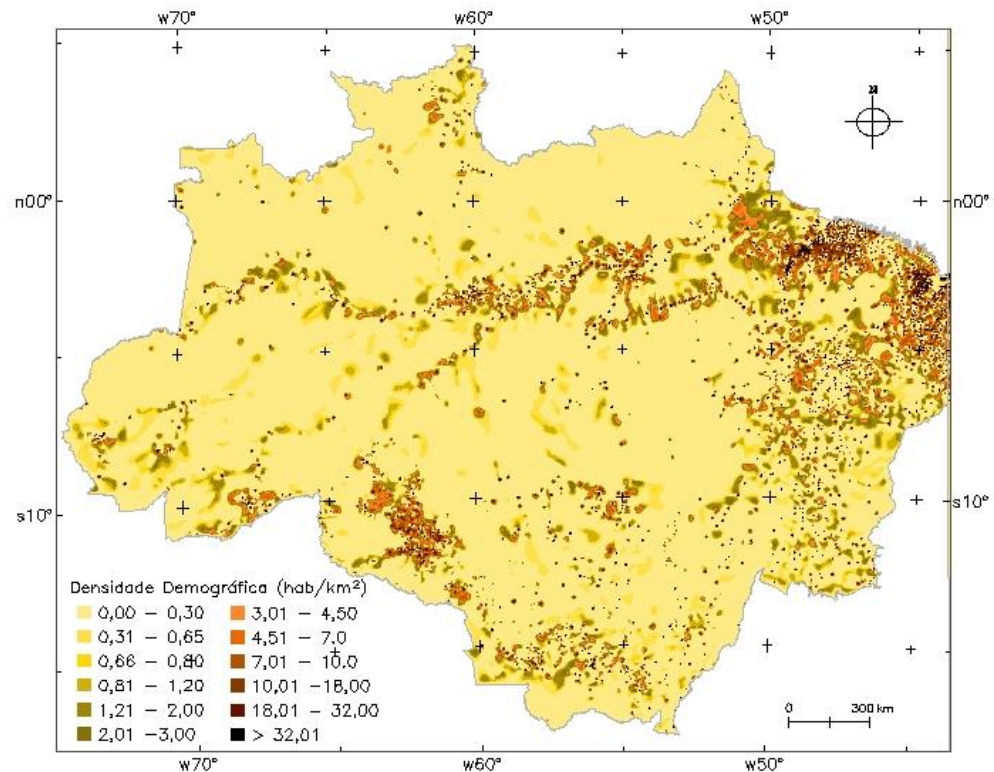
- MM97
- TobLux*



População na AmzL

Superfícies de Densidade – Picnofilático de Tobler

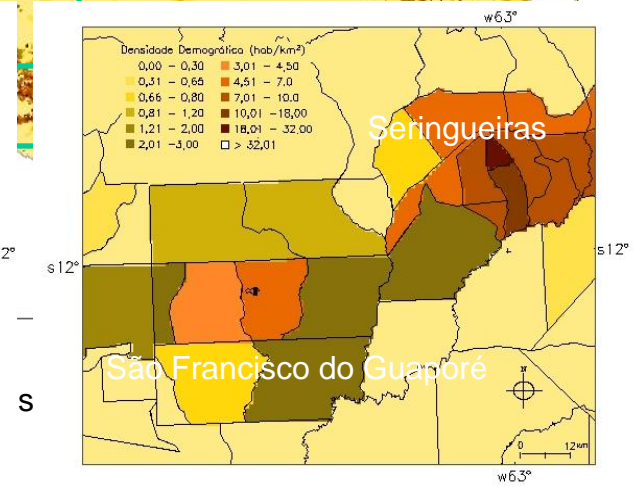
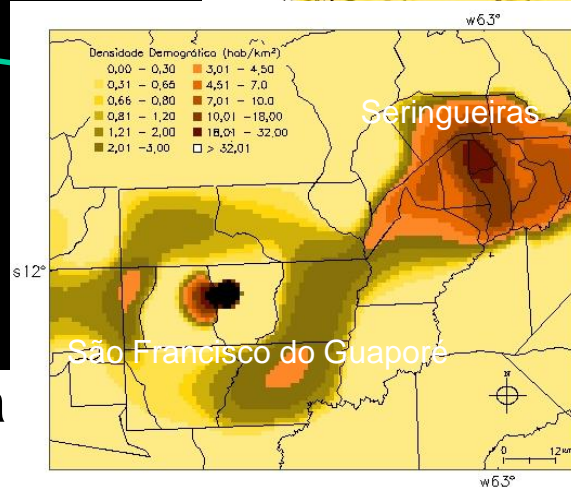
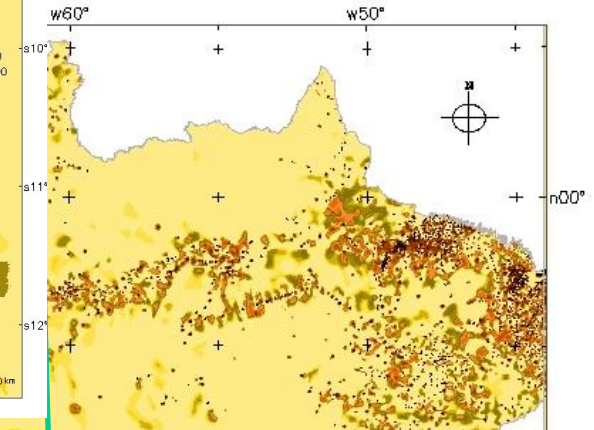
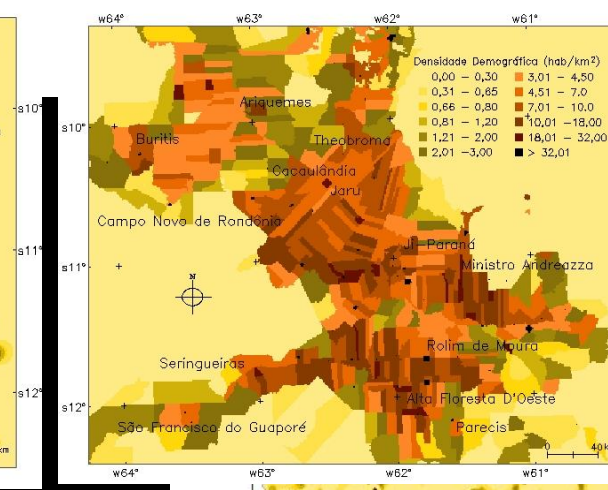
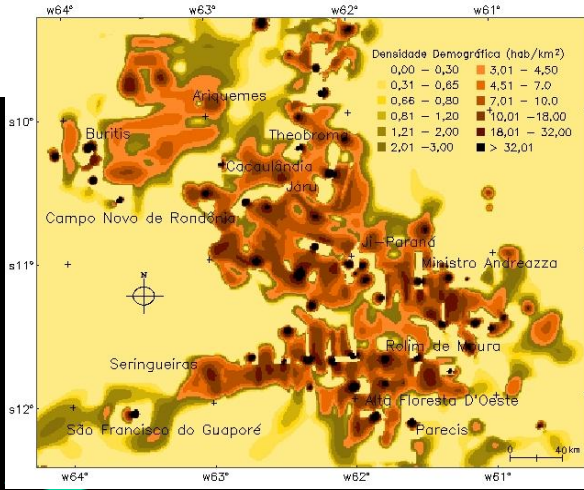
- Preserva volumes, suaviza limites poligonais
- Malha municipal 1997, população total 2000
- Das superfícies-> Mapas temáticos para análise
- Continuidade para áreas pequenas e densidades similares
- Por setor censitário 2000



Densidade População Total 2000

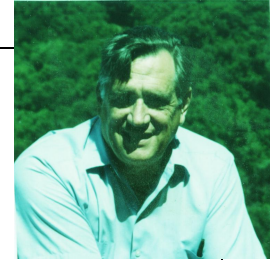
População na AmzL

Picnofilático de Tobler



densidades simila

- Por setor censitário 2000



More Tests Needed

Tests are also needed to determine the viability of pycnophlyactic reallocation when converting from one set of areas to another.

To date no adequate tests of this procedure have been reported.

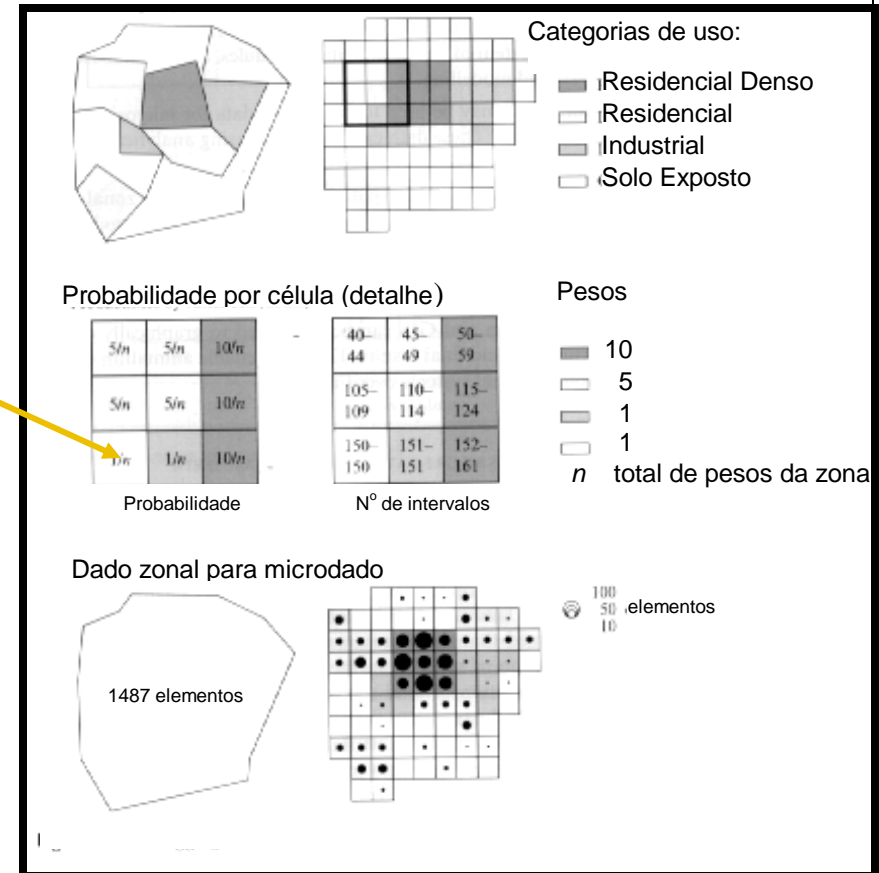
In essence these would test the adequacy of the hypothesis implying spatial autocorrelation - that neighboring places have similar values.

I appreciate your attention and thank you.



Modelos de População “Multivariados”

- Variáveis indicadoras da presença humana para distribuir a população nas superfícies de densidade.
- **Método Dasimétrico** – uso de classificações de imagens de SR: classes de uso do solo – pesos para desagregação
- Interpoladores Inteligentes: informação espacial de outras fontes para orientar interpolação:
 - superfície de ponderação “mapeia” dados originais na superfície de saída
 - Variáveis preditoras x var. de interesse



População, Espaço e Ambiente

Abordagens Espaciais em Estudos de População:
Métodos Analíticos e Técnicas de Representação

Parte IV – Integração de Dados

Método Dasimétrico

Silvana Amaral

Antonio Miguel V. Monteiro

{silvana@dpi.inpe.br, miguel@dpi.inpe.br}



Origem



- **dasymetry**

[física] s. dasimetria, f.; parte da Física que estuda a determinação da **densidade** do ar nas diferentes camadas da atmosfera

Mapeamento Dasimétrico

Termo criado por Benjamin (Veniamin) Petrovich Semenov-Tyan-Shansky e popularizado por Wright (1936).

Técnica cartográfica para a apresentação de dados, que podem ser conceituados e representados como variações contínuas, mas irregulares de dimensões como tempo, espaço, elevação ou temperatura (adaptado de Mennis, 2009).

Origem



- **dasymetry**

[física] s. dasimetria, f.; parte da Física que estuda a determinação da densidade do ar nas diferentes camadas da atmosfera

Mapeamento Dasimétrico

Pode ser definido como um tipo de interpolação para áreas que utiliza dados auxiliares (adicionais ou relacionados) para facilitar o processo de interpolação

IS NOT= to choropleth mapping:

- Os limites da representação cartográfica não são arbitrários, mas refletem a distribuição espacial da variável que está sendo mapeada (Eicher and Brewer 2001).

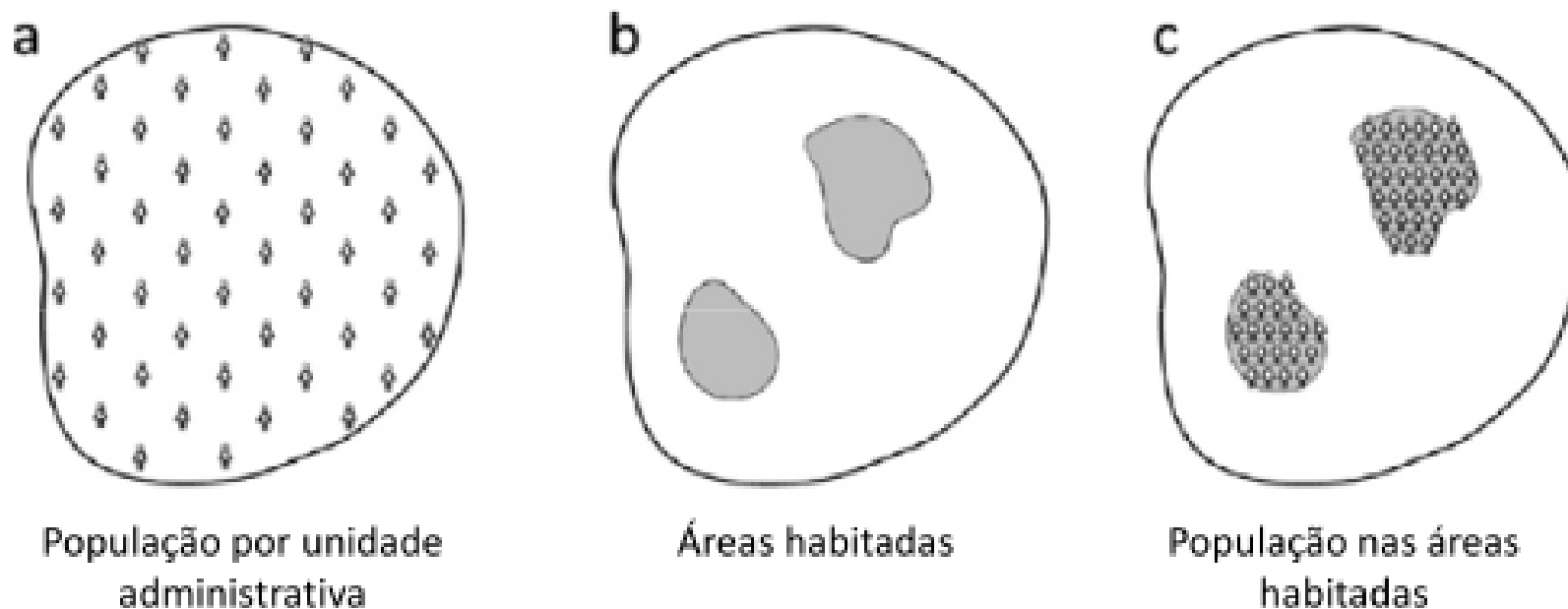
Objetivos - aplicações



- Suporte para elaboração de superfícies de densidade populacional
- Levantamento de dados
- Reconhecimento dos processos e sua heterogeneidade espacial
- Interação com outros aspectos e interesses
- Suporte para Modelagem

Método de desagregação: dasimétrico

Recorre a informação geográfica auxiliar, fortemente correlacionada com a variável a representar, de forma a melhorar a sua representação cartográfica, limitando a distribuição às áreas em que de facto ela ocorre.



Mapeamento Dasimétrico

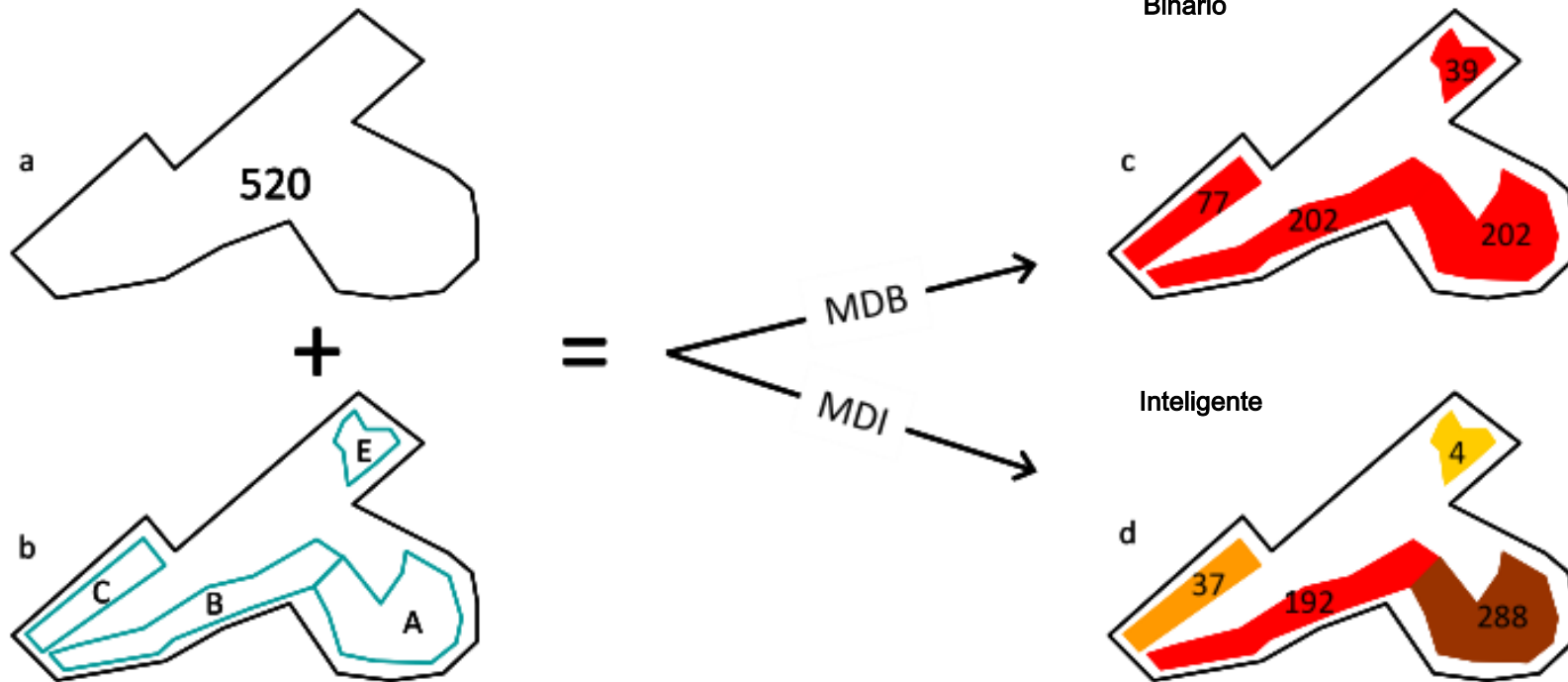
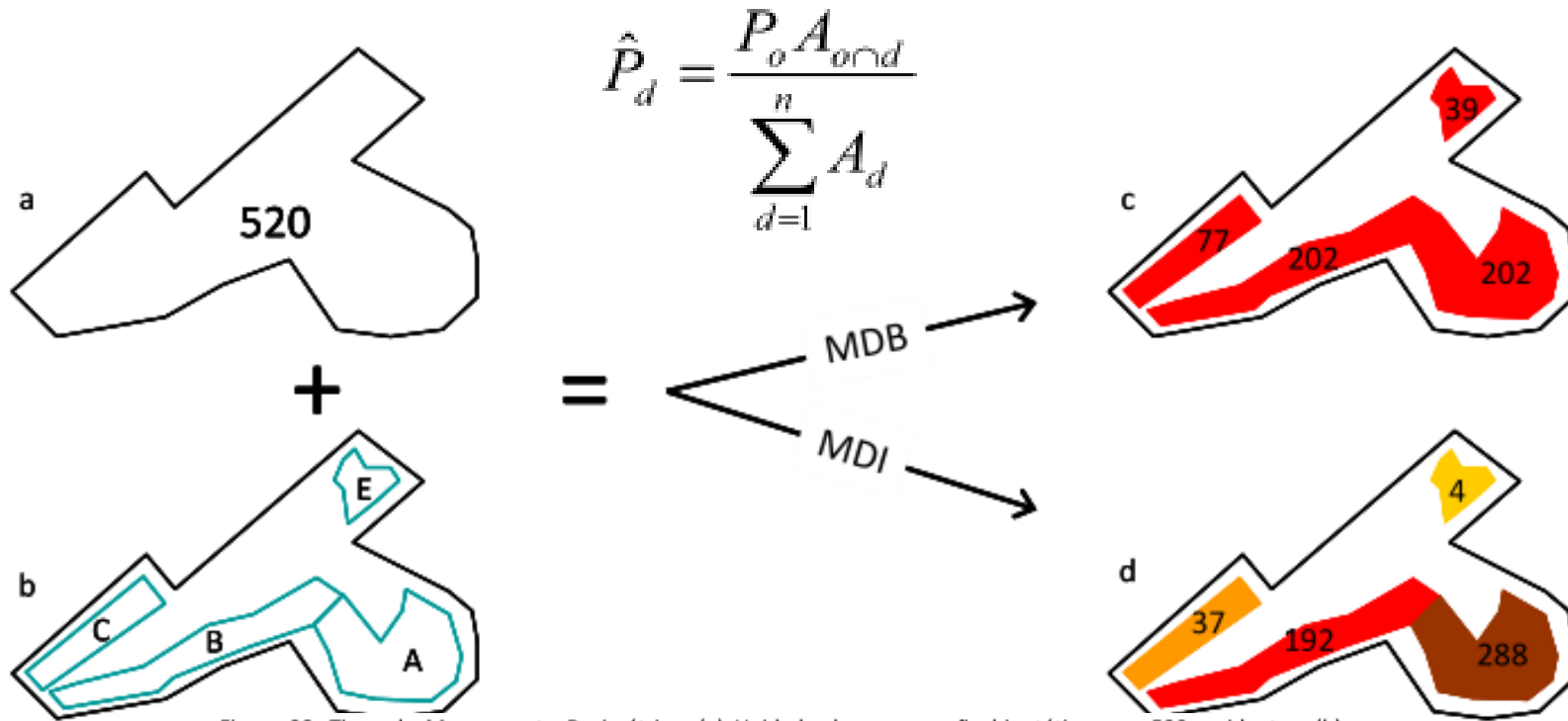


Figura 28. Tipos de Mapeamento Dasimétrico. (a) Unidade de enumeração hipotética com 520 residentes; (b) Classes de ocupação do solo hipotéticas existentes no interior da unidade de enumeração. (c) e (d) Modelos de distribuição da população produzidos por Mapeamento Dasimétrico Binário (c) e Inteligente (d).

- A – Área edificada compacta
- B – Área edificada linear contínua
- C – Área edificada linear descontínua
- D – Área edificada em espaço rural
- E – Área de edificação dispersa

Mapeamento Dasimétrico Binário (MDB)



$$\hat{P}_d = \frac{P_o A_{o \cap d}}{\sum_{d=1}^n A_d}$$

Figura 28. Tipos de Mapeamento Dasimétrico. (a) Unidade de enumeração hipotética com 520 residentes; (b) Classes de ocupação do solo hipotéticas existentes no interior da unidade de enumeração. (c) e (d) Modelos de distribuição da população produzidos por Mapeamento Dasimétrico Binário (c) e Inteligente (d).

\hat{P}_d = população estimada para a zona de destino a

P_o = população conhecida na zona de origem o

$A_{o \cap d}$ = área zona de destino d

A_d = área da zona de destino d

n = número de zonas de destino

Mapeamento Dasimétrico Binário (MDB)

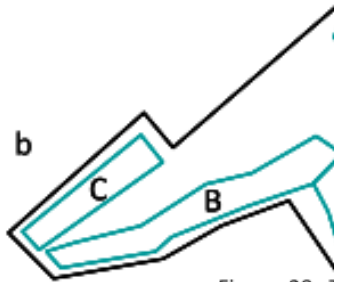
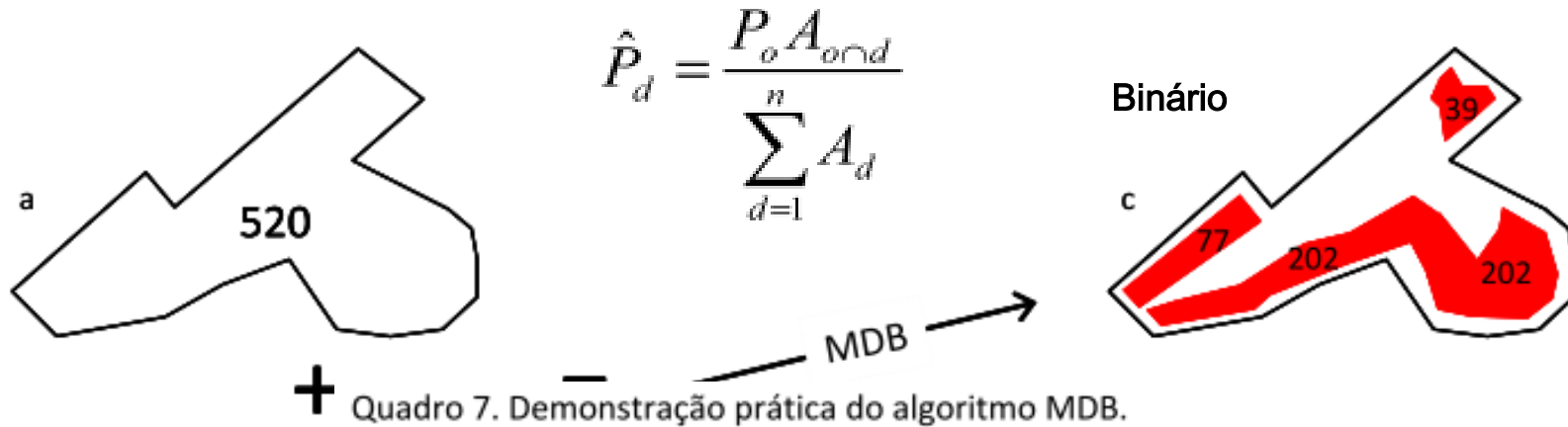


Figura 28.1
Classes de
distribuição

	1	2	3	4	5	6	7	8
	Classe de ocupação do solo	População agregada à secção	Nr. total de pixéis na secção	Nr. pixéis na secção	$\Sigma 4$	4 / 5	2 x 6	7 / 4
A				21		0.39	202	9.630
B				21		0.39	202	9.630
C		520	100	8	54	0.15	77	9.630
D				0		0.00	0	0.000
E				4		0.07	39	9.630

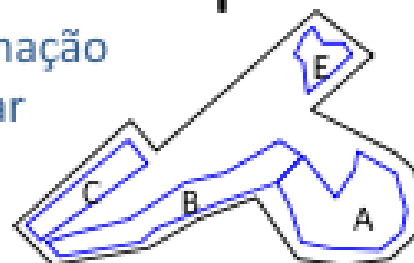
- 1 - Classes de ocupação do solo com presença de função residencial
- 2 - População agregada à secção estatística
- 3 - Número total de pixéis na secção (área total da secção)
- 4 - Número de pixéis de cada classe de ocupação do solo na secção estatística
- 5 - Número total de pixéis com função residencial
- 6 - Percentagem de população a alocar a cada classe de ocupação do solo
- 7 - População alocada por classe de ocupação do solo
- 8 - População atribuída a cada pixel

Método de desagregação: dasimétrico

População por subsecção

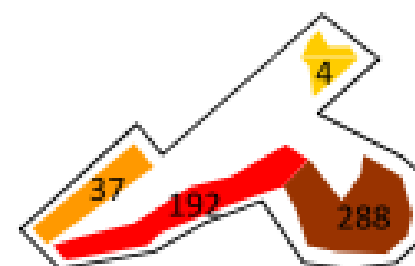


Informação auxiliar



=

População desagregada



- A – Área edificada compacta
- B – Área edificada linear contínua
- C – Área edificada linear descontínua
- D – Área edificada em espaço rural
- E – Área de edificação dispersa

$$\hat{P}_d = P_o \left(\frac{A_d \hat{D}_c}{\sum_{d \in o} (A_d \hat{D}_c)} \right)$$

D_c é um parâmetro obtido por amostragem;

Corresponde a uma **densidade populacional relativa** entre as classes do zoneamento auxiliar, ou **Fração de densidade**.

Mapeamento Dasimétrico Inteligente

$$\hat{P}_d = P_o \left(\frac{A_d \hat{D}_c}{\sum_{d \in o} (A_d \hat{D}_c)} \right)$$



Quadro 8. Demonstração prática do algoritmo MDI.

1	2	3	4	5	6	7	8	9	10	11
Classe de ocupação do solo	População agregada à secção	Nr. total de pixéis na secção	Fracção de densidade	Nr. pixéis na secção	$\sum 5$	$4 * 5$	$\sum 7$	$7 / 8$	$2 * 9$	$10 / 5$
A	520	100	0.45	21	54	9.45	17.07	0.55	288	13.708
B			0.3	21		6.3		0.37	192	9.139
C			0.15	8	1.2	0.07		37	4.569	
D			0.07	0	0	0.00	0	0.000		
E			0.03	4	0.12	0.01	4	0.914		

- A – Área edificada compacta
- B – Área edificada linear contínua
- C – Área edificada linear descontínua
- D – Área edificada em espaço rural
- E – Área de edificação dispersa

1 - Classes de ocupação do solo com presença de função residencial

2 - População agregada à secção estatística

3 - Número total de pixéis na secção (área total da secção)

4 - Densidade populacional relativa amostrada para cada classe de ocupação do solo

5 - Número de pixéis de cada classe de ocupação do solo na secção estatística

6 - Número total de pixéis com função residencial

7 - Ponderação da área de cada classe de ocupação do solo pela fracção de densidade

8 - Somatório de 7

9 - Percentagem de população a alocar a cada classe de ocupação do solo

10 - População alocada por classe de ocupação do solo

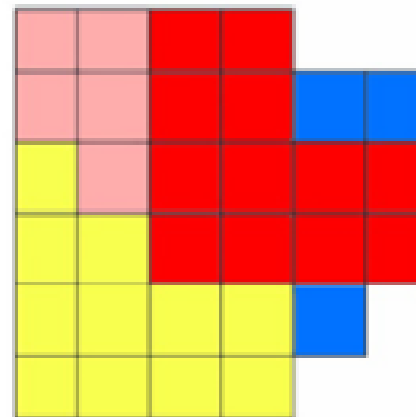
11 - População atribuída a cada pixel

Método de desagregação: dasimétrico matricial

População por subsecção



Informação auxiliar



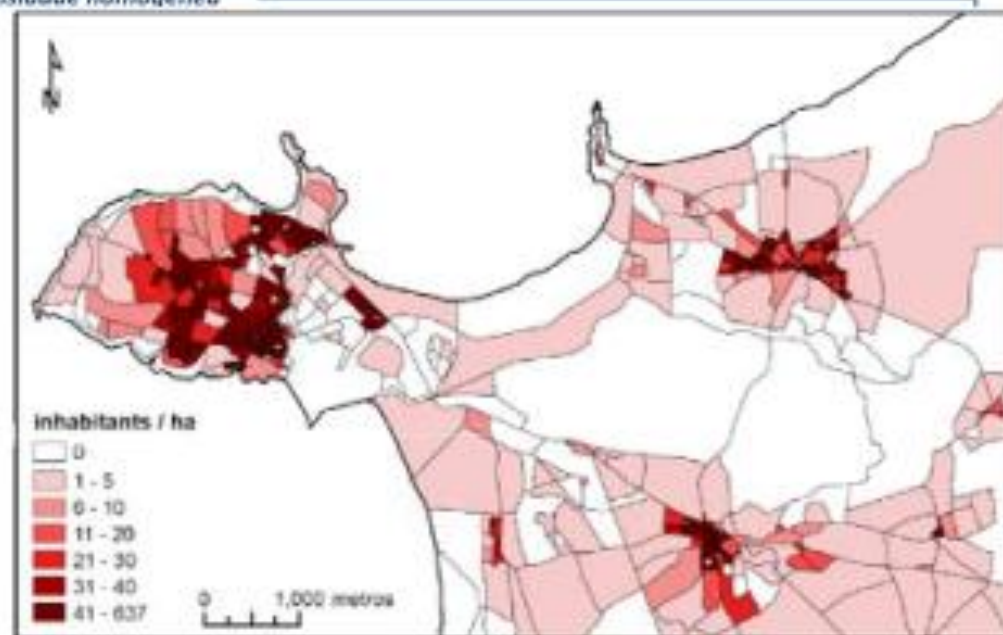
População desagregada

9.64	9.64	12.65	12.65		
9.64	9.64	12.65	12.65	0	0
3.73	9.64	12.65	12.65	12.65	12.65
3.73	3.73	12.65	12.65	12.65	12.65
3.73	3.73	3.73	3.73	0	
3.73	3.73	3.73	3.73		

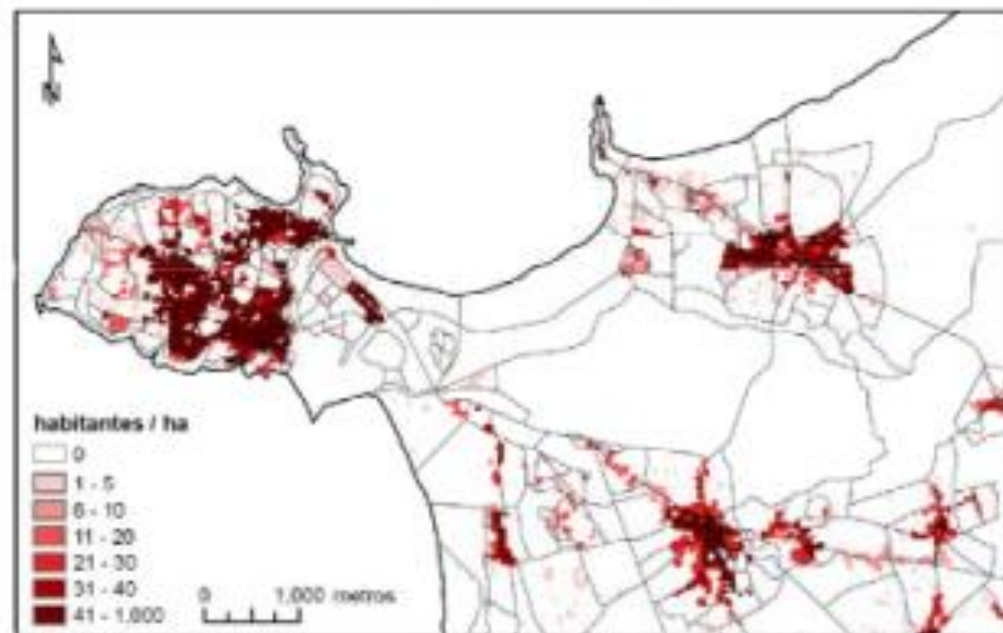
“Modelo digital da População”

Resultados: distribuição populacional

Mapa
coropleto



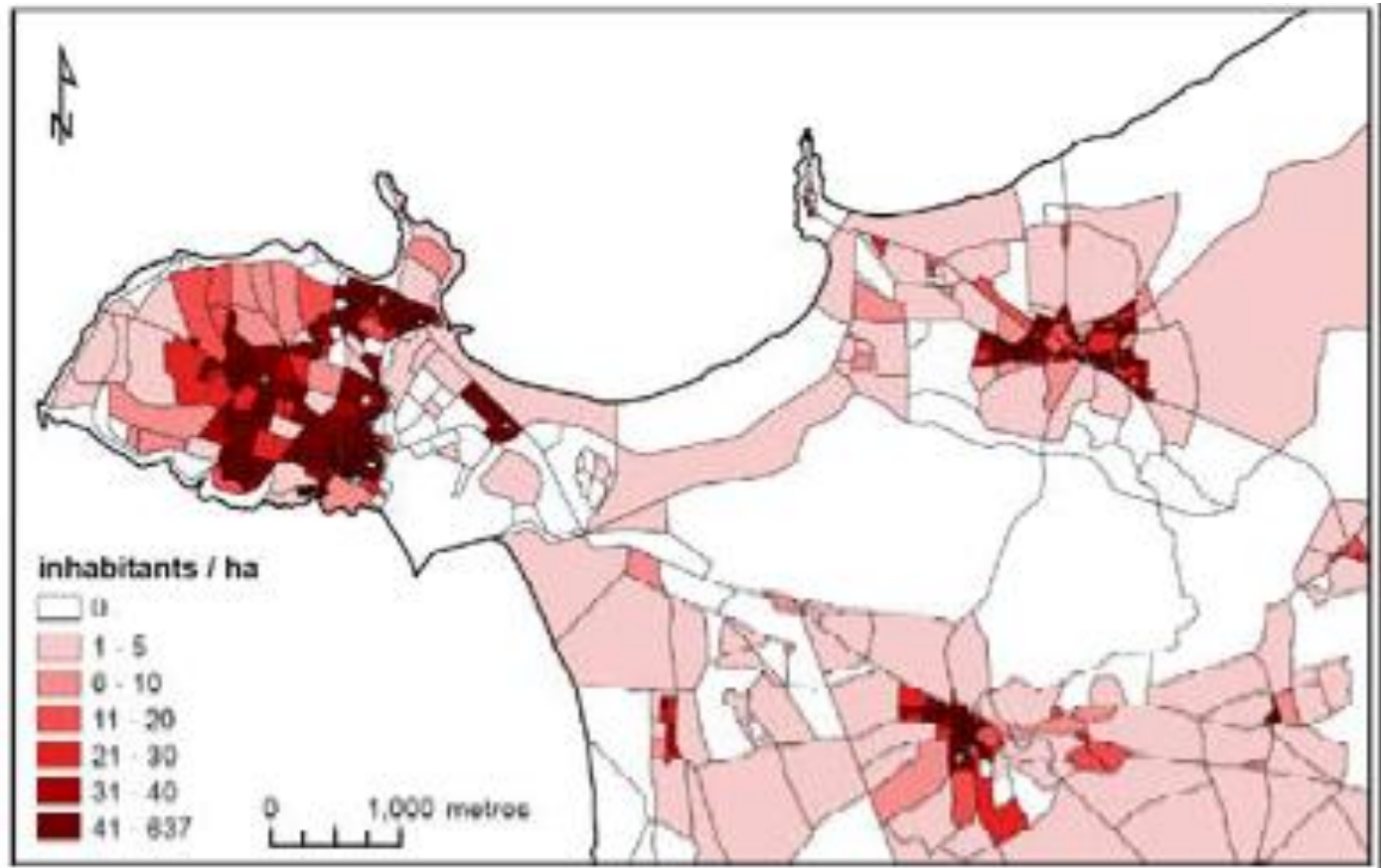
Mapa
dasimétrico



Marques, Silva & Delgado - A ocupação edificada: delimitação de áreas de densidade homogênea – Seminário “A ocupação dispersa no quadro das

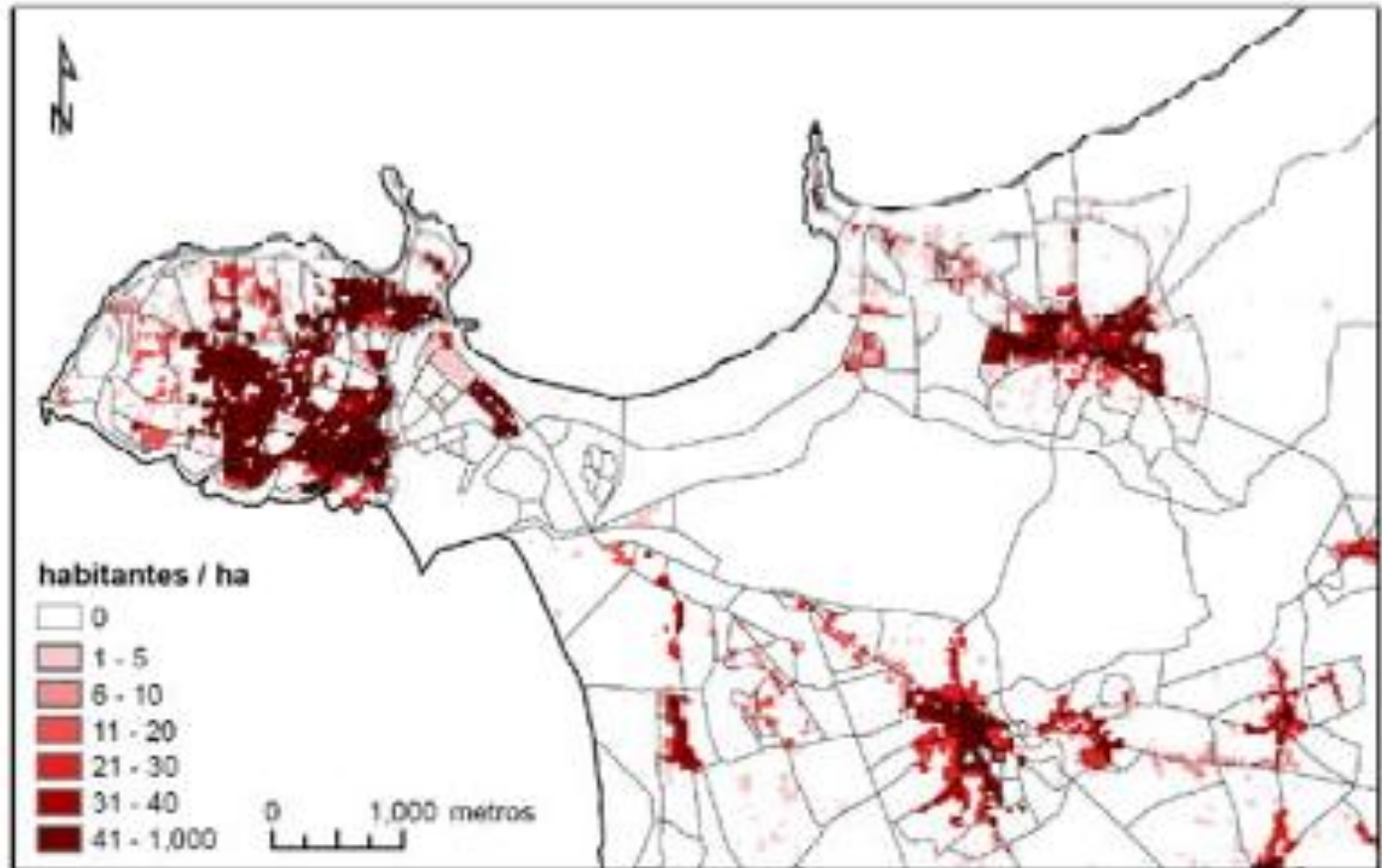
Resultados – Distribuição populacional

Mapa coroplético



Resultados – Distribuição populacional

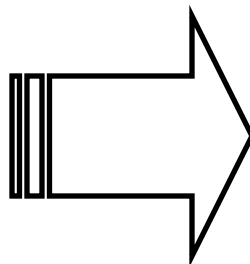
Mapa Dasimétrico



Exemplo didático...

- Perspectiva e Objeto de estudo

**Dinâmica
Populacional e
Assentamentos
Humanos**



Geração de
Território

DP & DPHA – Pensando Marabá

- Para entender mobilidade:
 - Núcleos urbanos - ocorrência e conectividade
 - Informações demográficas (migração) – INCRA, ONGs?
 - Estradas a acessos / vicinais
 - População ribeirinha
 - Indicadores das redes técnicas, físicas e sociais
 - Energia elétrica – produção e consumo (*)

DP&AHA – O que temos...

- Marabá

- 15.137,4 km² município

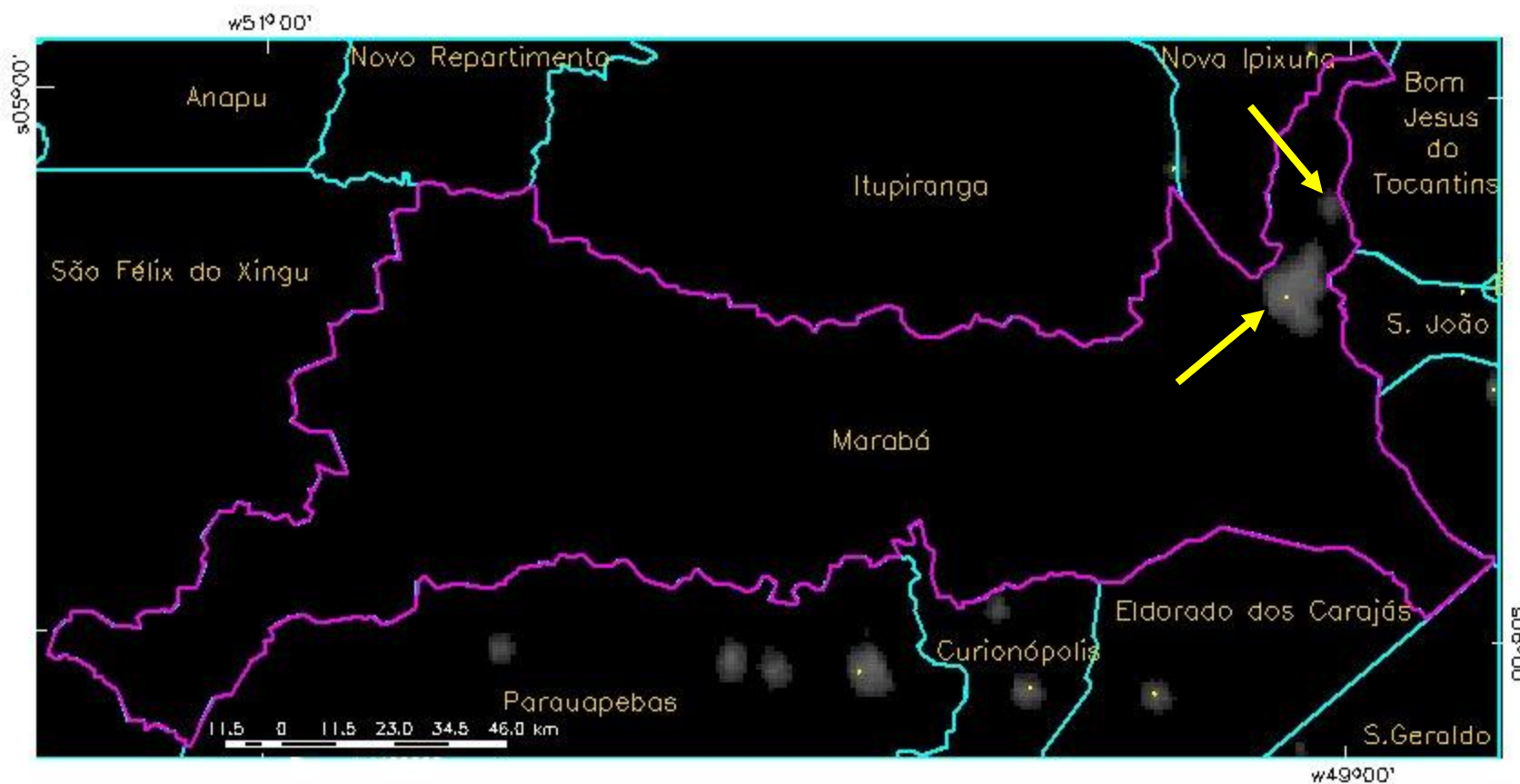
- População

	Pop Urbana	Pop Rural	Pop Total
1991	102.435	21.233	123.668
1996	123.378	26.717	150.095
2000	134.258	33.615	167.873
2000*	134.373	33.647	168.020

- Como encontra-se distribuída esta população?

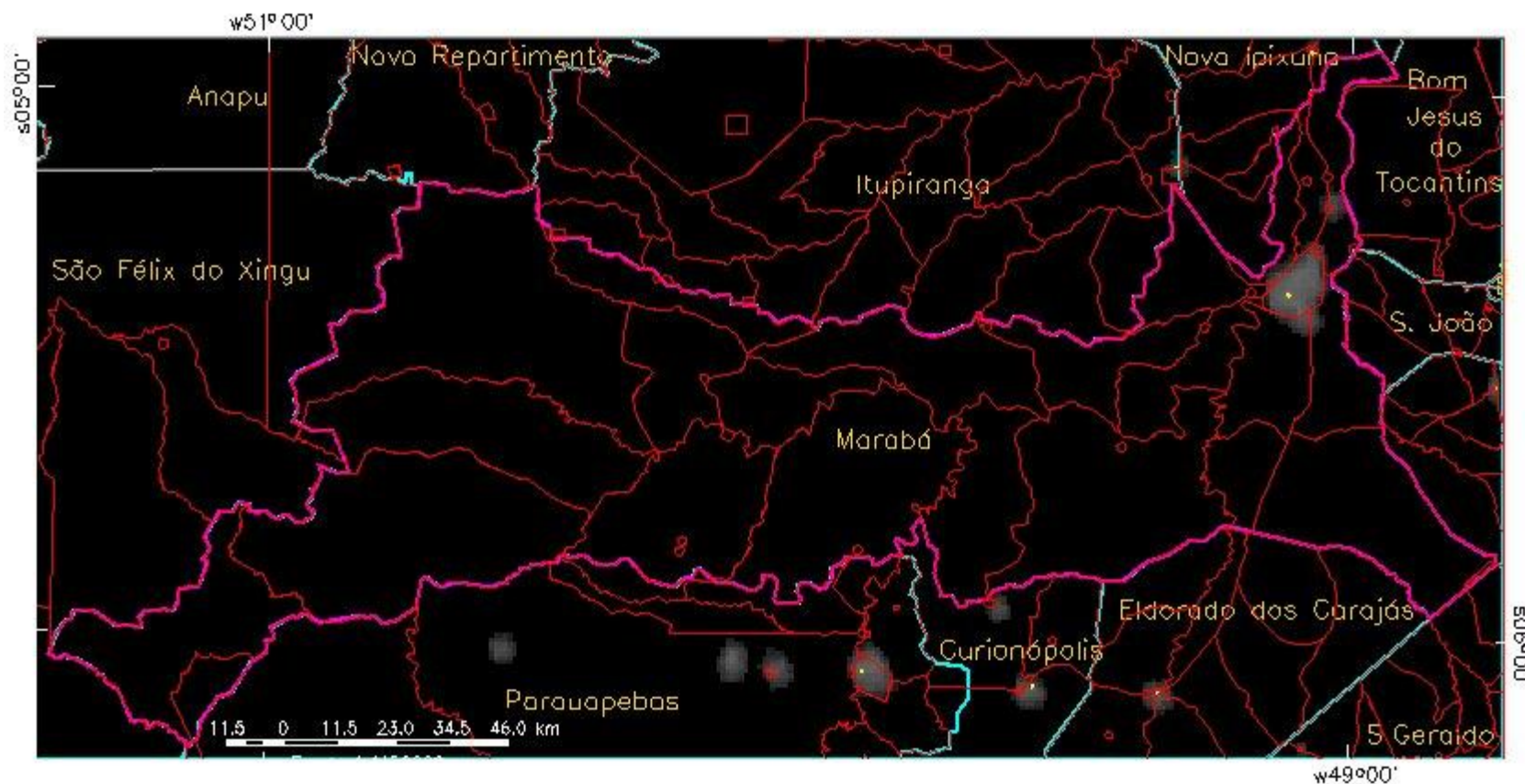
DP&AHA – O que temos...

- Luzes Noturnas – Escala Global



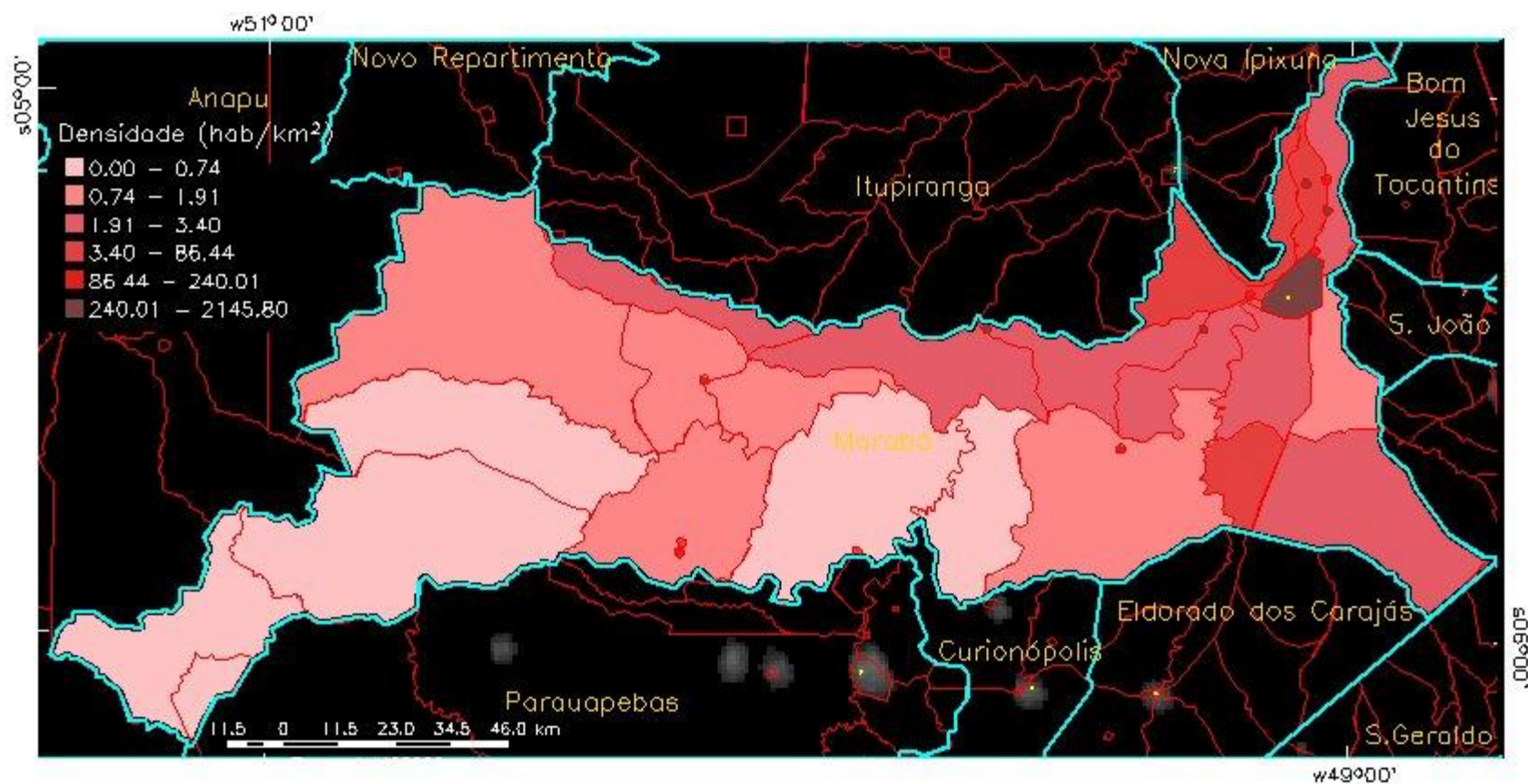
DP&AHA – O que temos...

- Setores Censitários 2000 *



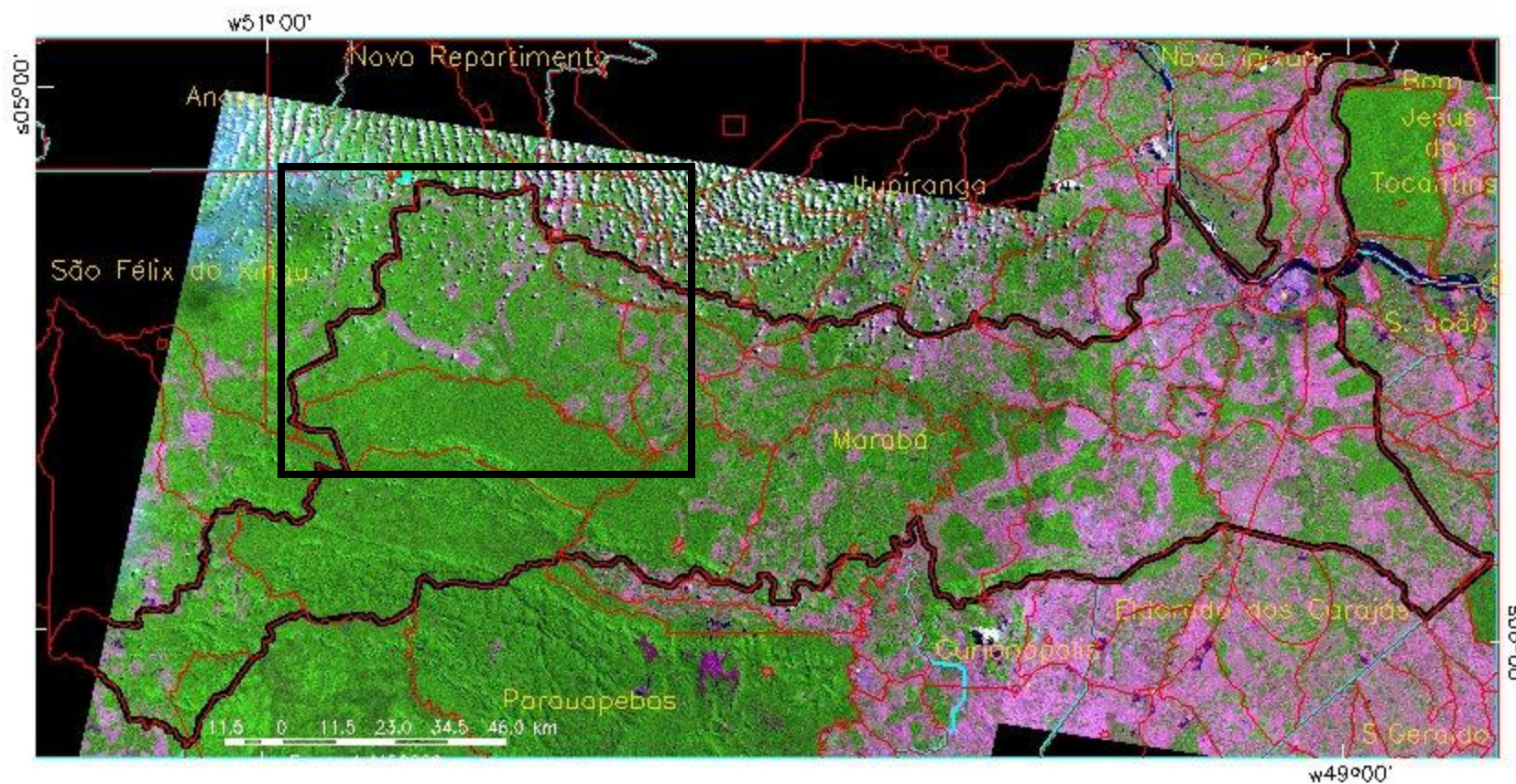
DP&AHA – O que temos...

- Setores Censitários 2000 * - Densidade Populacional

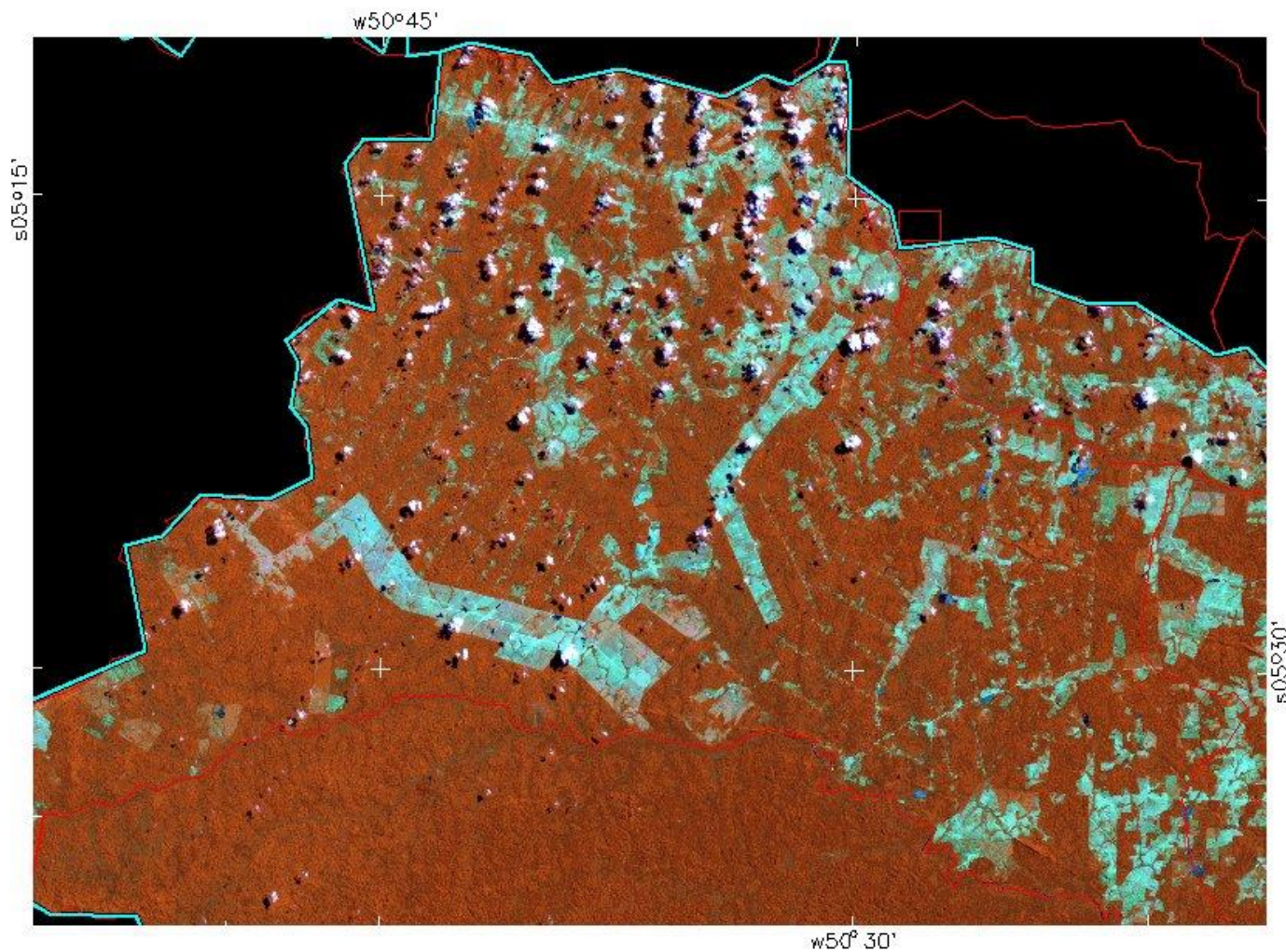


DP&AHA – O que temos...

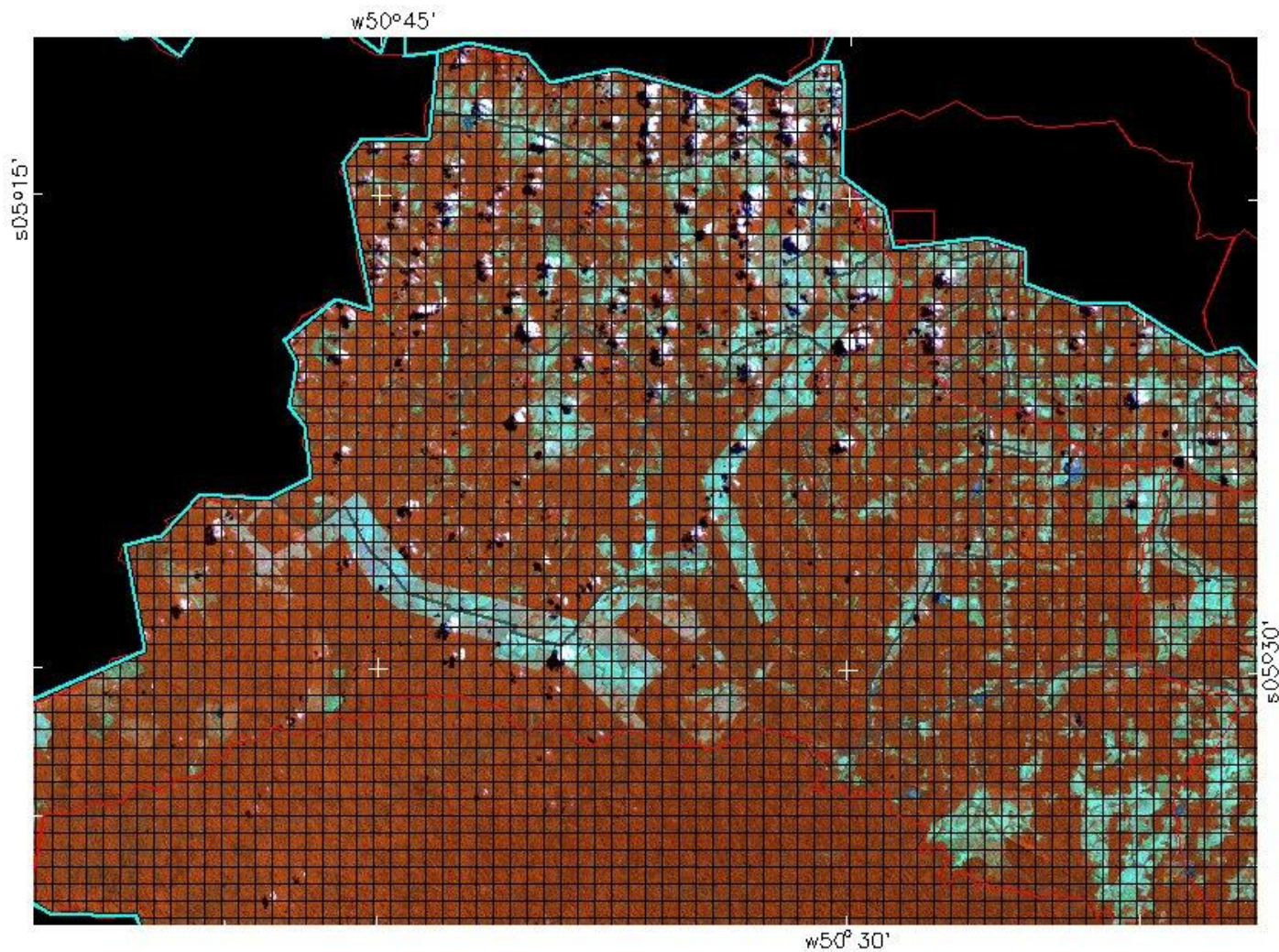
- Setores Censitários 2000 * - Heterogeneidade



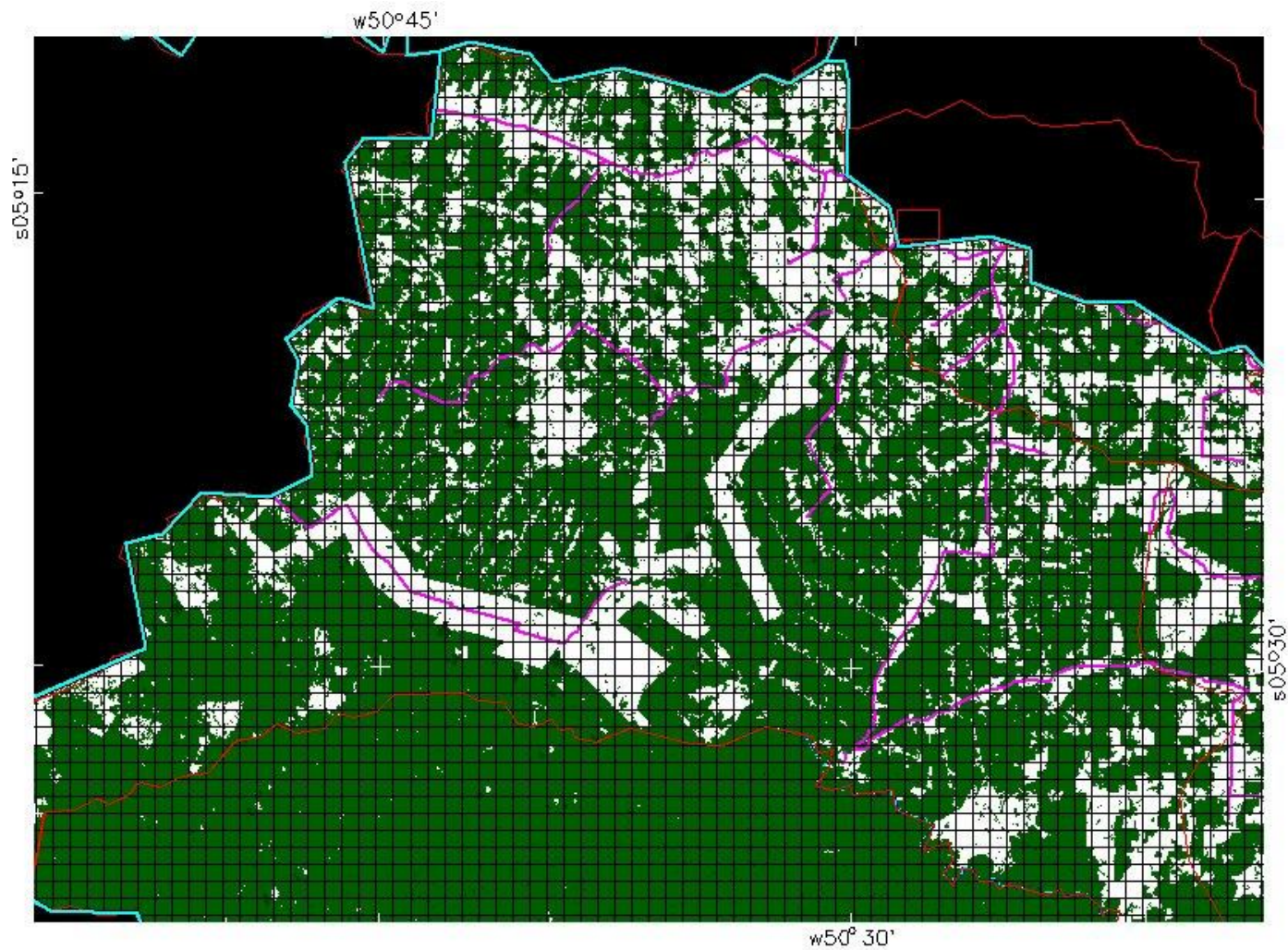
DP&AHA – O que temos...



DP&AHA – O que temos...



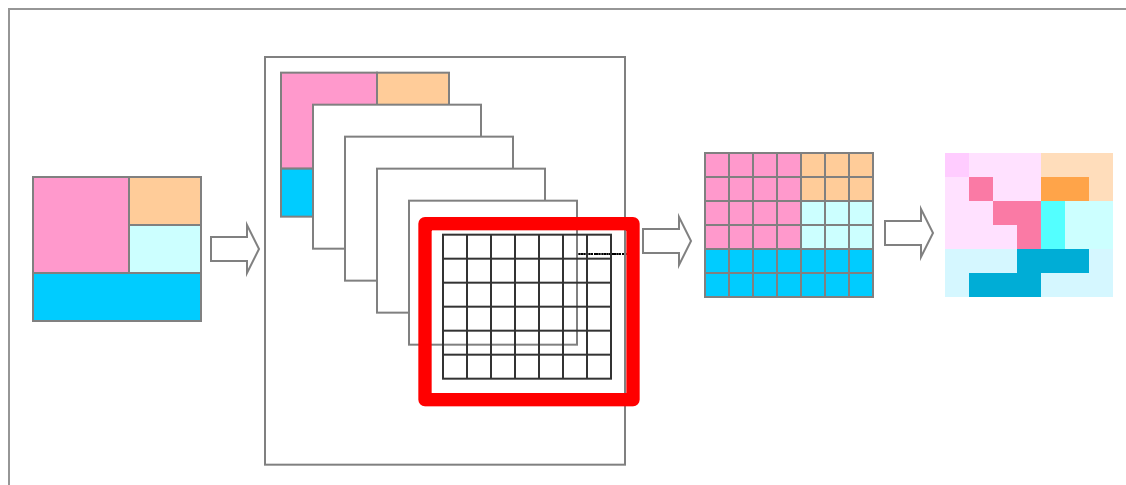
DP&AHA – O que temos...



População em Marabá

Desagregando no município...

- Incluir heterogeneidade
- Setores Censitários de um município
- Considerações / Método
 - Água e floresta
Restrição de células
 - Variáveis para indicar presença ->
Superfície de população
 - Relação entre as variáveis ->
Redistribuição



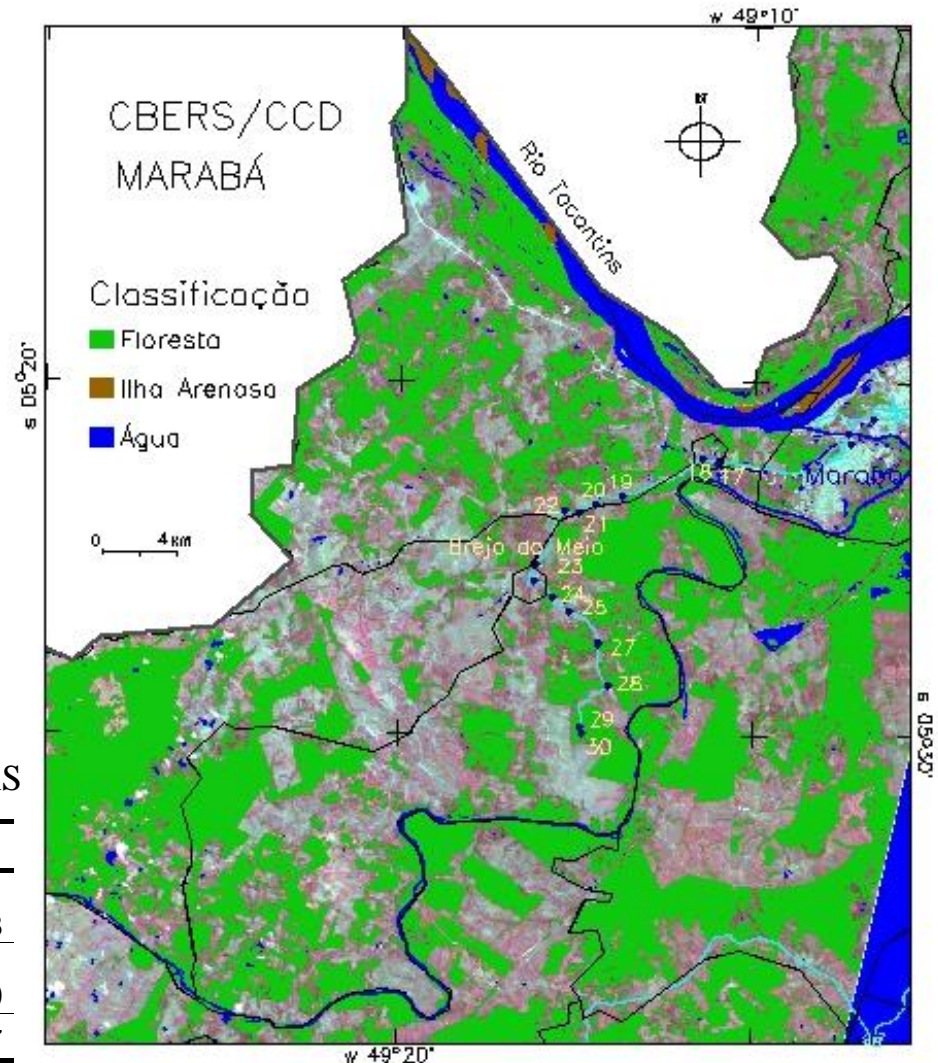
Redistribuição de setores censitários para células

Método Dasimétrico

- Imagens de satélite
- Classes Água e Floresta
- CBERS para região
- Landsat para município
- Técnicas simples de classificação digital
- Células 95%

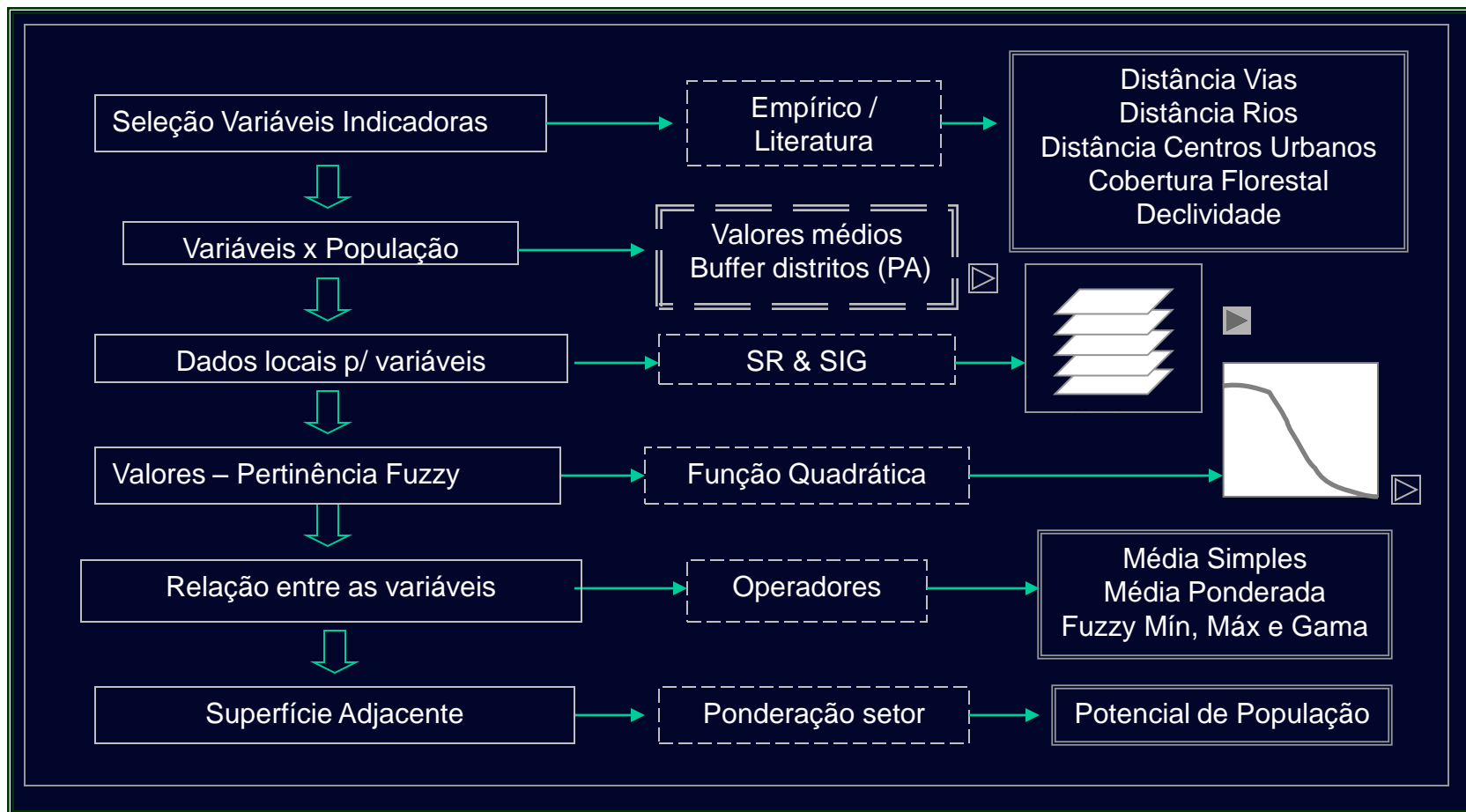
Limiaries para detecção de classes nas bandas CCD-CBERS

Classe	Banda Espectral	Limiar
Corpos D'água	Banda CCD4 Infravermelho Próximo (0,77 - 0,89 μm)	ND < 118
Ilhas arenosas	Banda CCD3 Vermelho (0,63 - 0,69 μm)	ND > 200
Floresta	Razão = $(\text{CCD4}/\text{CCD3})/(\text{CCD4} + \text{CCD3})$	ND > 107



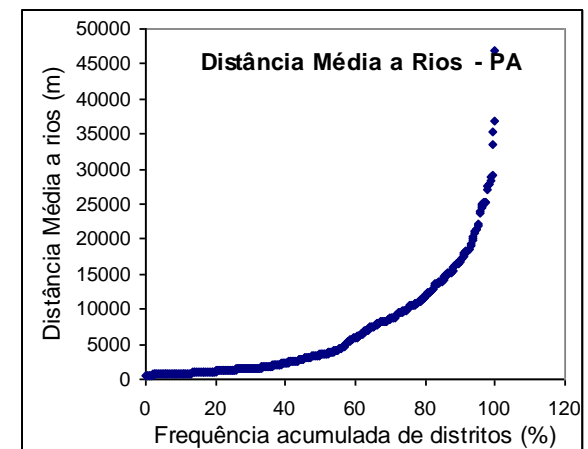
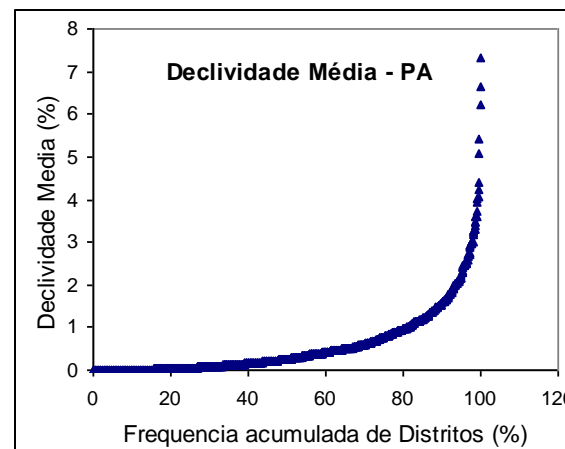
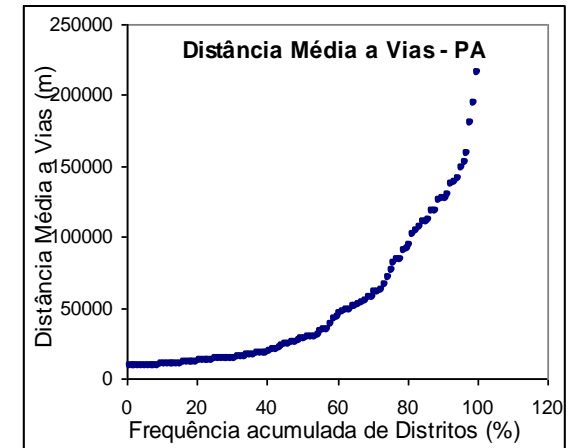
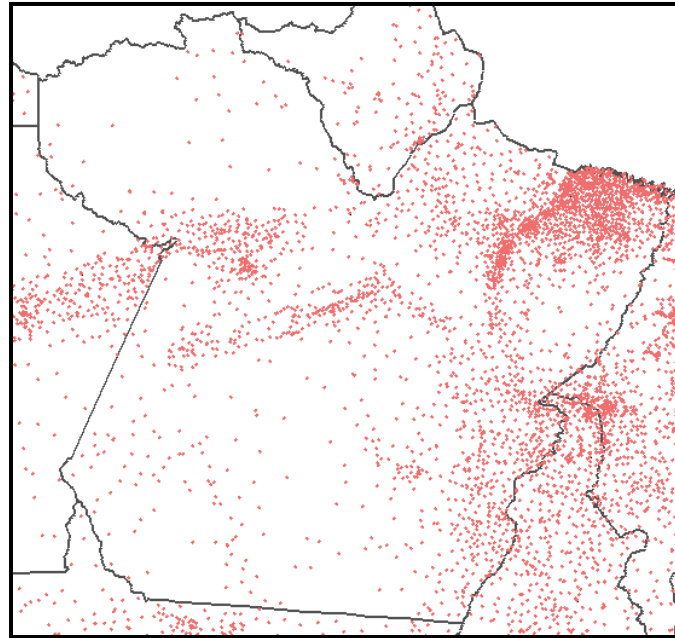
Método Multivariado

Inferir superfície que descreva distribuição



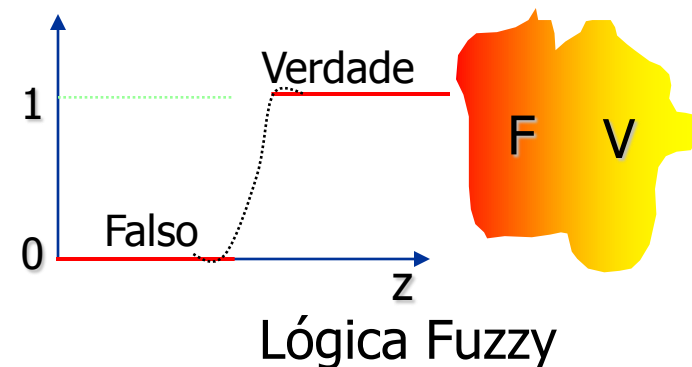
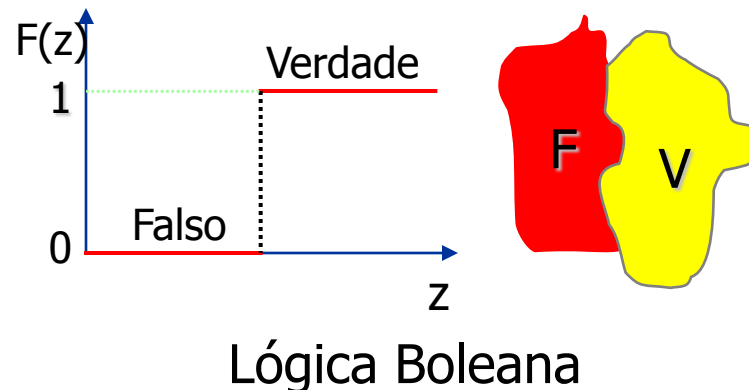
Método Multivariado – Contribuição relativa das variáveis preditoras

- Distritos do PA
- Área de Influência
- Média das distâncias a vias, distância a rios, declividade
- Distância a centros urbanos – Viz + Próx.
- Percentagem Floresta
 $5\% < x < 99\%$



Classificação contínua

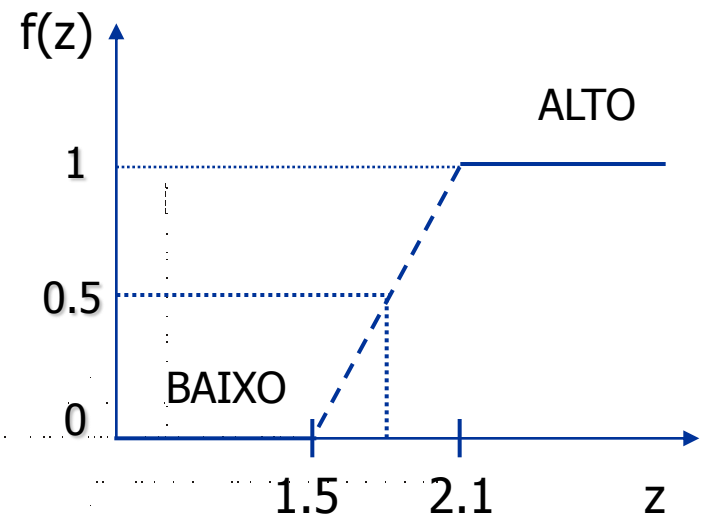
- A análise espacial em SIG será muito melhor realizada com uso das técnicas de classificação contínua: os dados são transformados para o espaço de referência $[0,1]$ e processados por combinação numérica, através de média ponderada ou inferência “fuzzy”
- Isto nos permite construir cenários (por exemplo, risco de 10%, 20% ou 40%), que indicam os diferentes compromissos de tomada de decisão => maior flexibilidade e um entendimento muito maior sobre os problemas espaciais
- Lógica Fuzzy: Introduzida por Lofti Zadeh (1960s), como um meio de modelar incertezas da linguagem natural
- Lógica Fuzzy é uma extensão da lógica Booleana: “verdade parcial”, valores entre “completamente verdadeiro” e “completamente falso”



Conjuntos Fuzzy

- Exemplo: Altura de Pessoas
 - S um conjunto fuzzy ALTO, que responderá a pergunta:
 - " a que grau uma pessoa "z" é alta?
- $Z : S = (z, f(z))$ especialistas

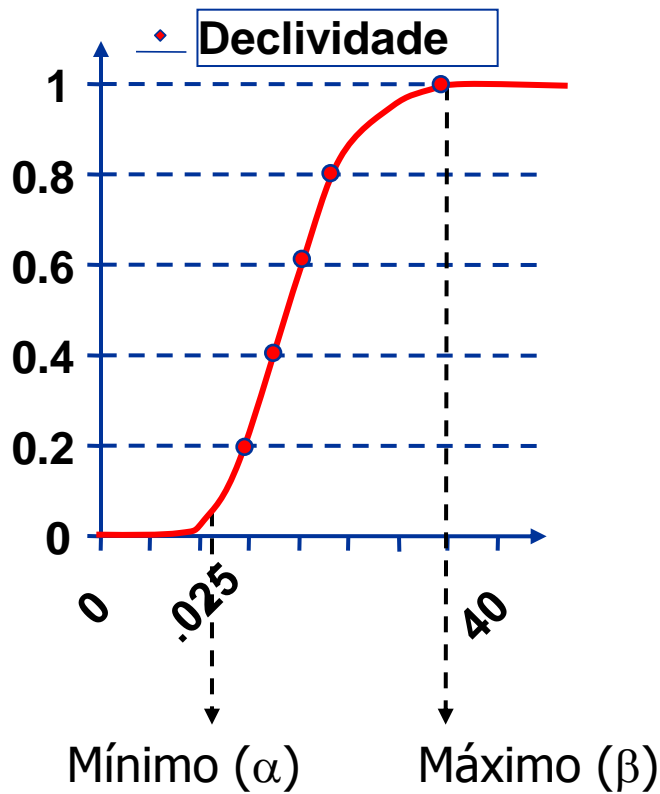
$$f(z) = \begin{cases} 0, & \text{se } z \leq 1.5 \\ (z - 1.5) / 0.6 & \text{se } 1.5 < z < 2.1 \\ 1, & \text{se } z \geq 2.1 \end{cases}$$



- Exemplo: "João é ALTO" = 0.38

Conjuntos Fuzzy

- Outro exemplo - Declividade



$$f(z) = 0 \quad \text{se } z \leq \alpha$$

$$f(z) = 1/[1 + \alpha(z - \beta)^2] \quad \text{se } \alpha < z < \beta$$

$$f(z) = 1 \quad \text{se } z \geq \beta$$

$$f(z) = 0 \quad \text{se } z \leq 0.025$$

$$f(z) = 1/[1 + 0.025(z - 40)^2] \quad \text{se } \alpha < z < 40$$

$$f(z) = 1, \quad \text{se } z \geq 40$$

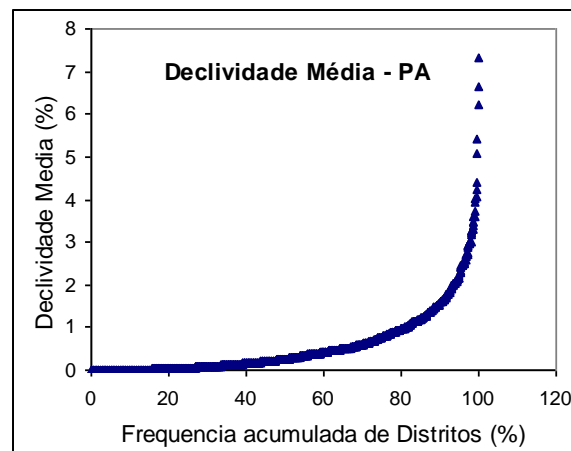
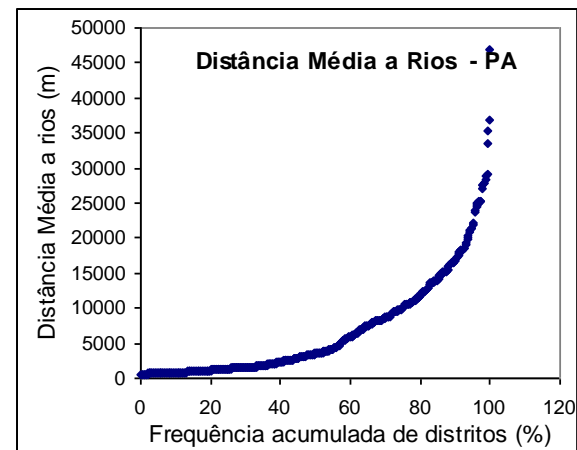
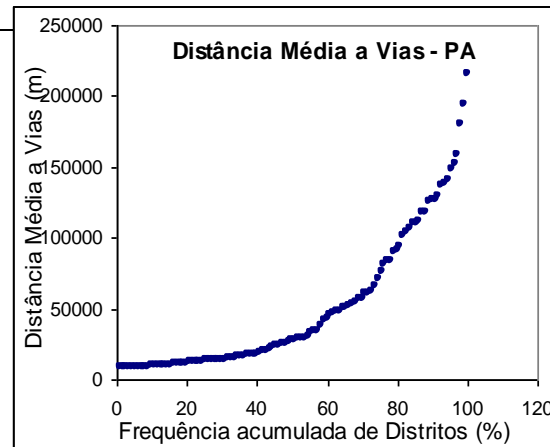
Variáveis preditoras

Parâmetros

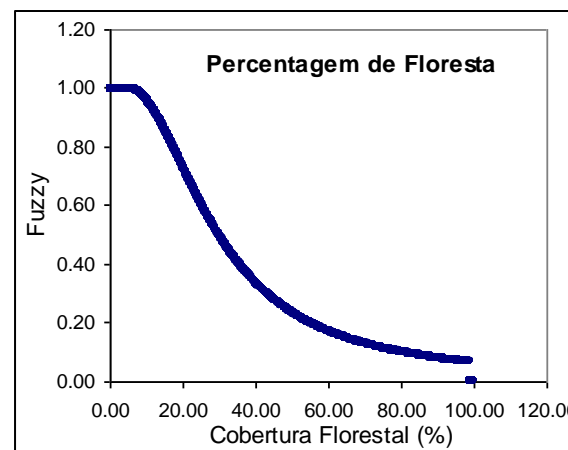
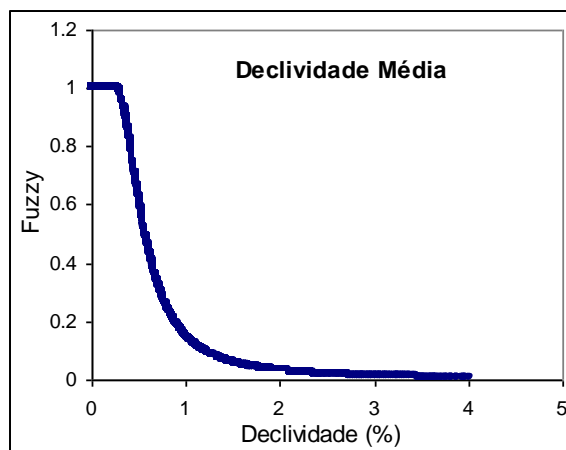
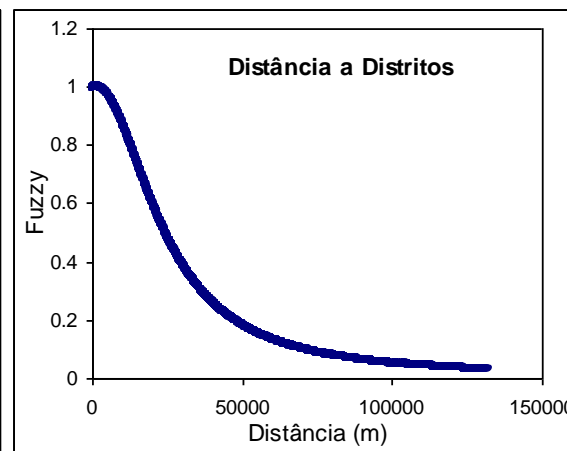
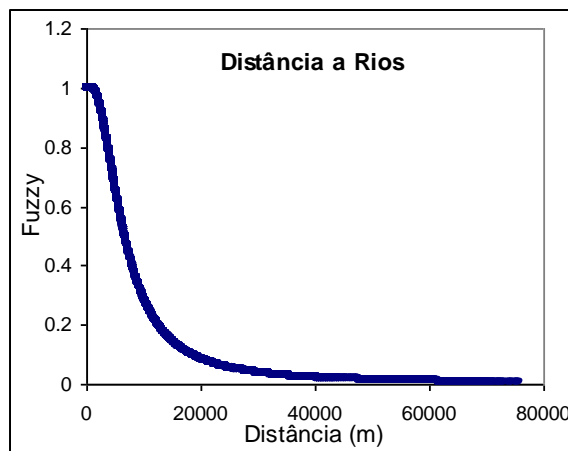
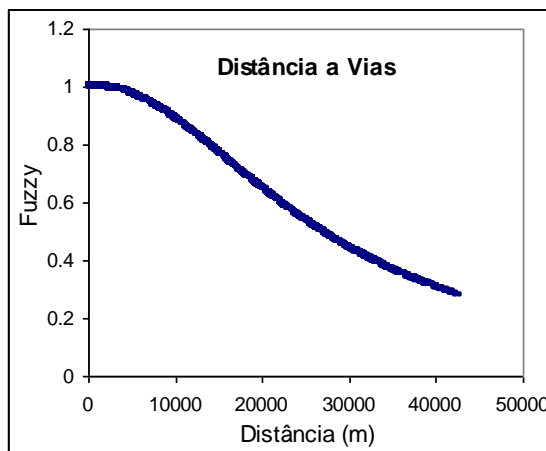
Parâmetros utilizados na definição das funções de pertinência *Fuzzy*.

Variável		Valores	f(z)	alfa	beta
Distância a vias (m) <i>Dvias</i>	<=	1000	1	1,48 E-09	1000
	=	27000	0,5		
	>	40000	0		
Distância a rios (m) <i>Drios</i>	<=	1000	1	2,96 E-08	1000
	=	6810	0,5		
	>	17000	0		
Distância a Setores Urbanos (m) <i>Durb</i>	<=	1500	1	1,89 E-09	1500
	=	24500	0,5		
	>	140000	0		
Floresta (% de área) <i>Pflor</i>	<=	0,05	1	16	0,05
	=	0,30	0,5		
	>	0,95	0		
Declividade Média (%) <i>Decl</i>	<=	0,27	1	10,4058	0,27
	=	0,58	0,5		
	>	3,5	0		

$$f(z) = \begin{cases} 0 & \text{se } (z > 40.000 \text{ m}) \\ 1/(1 + \alpha (z - \beta)^2) & \\ 1 & \text{se } (z \leq 1.000 \text{ m}) \end{cases}$$



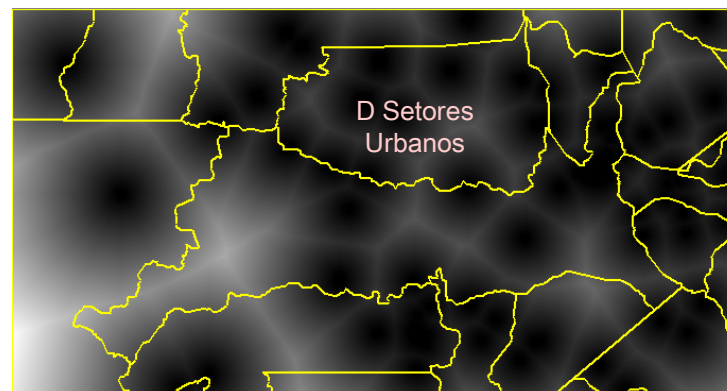
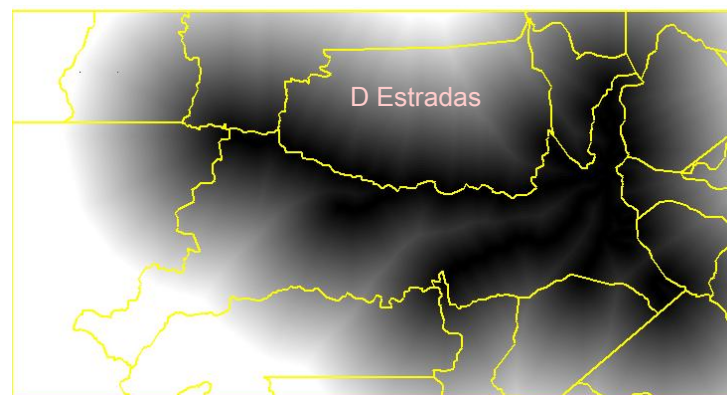
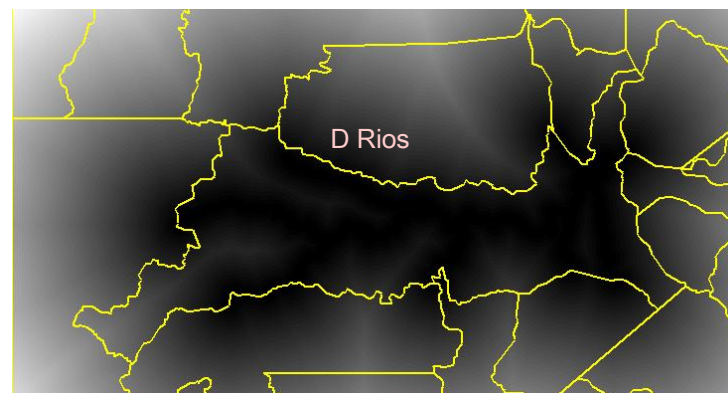
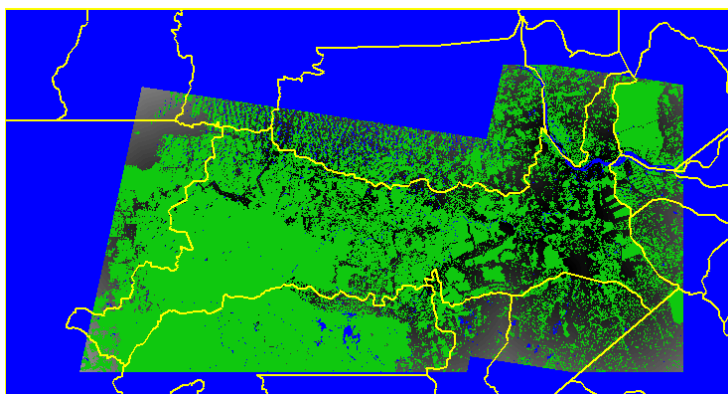
Método Multivariado – Função de Pertinência Fuzzy



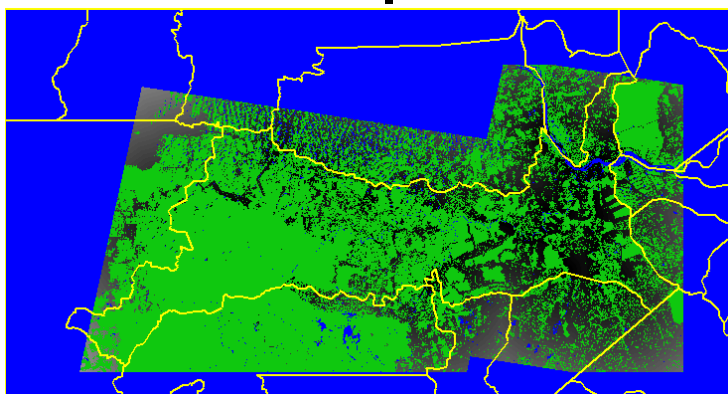
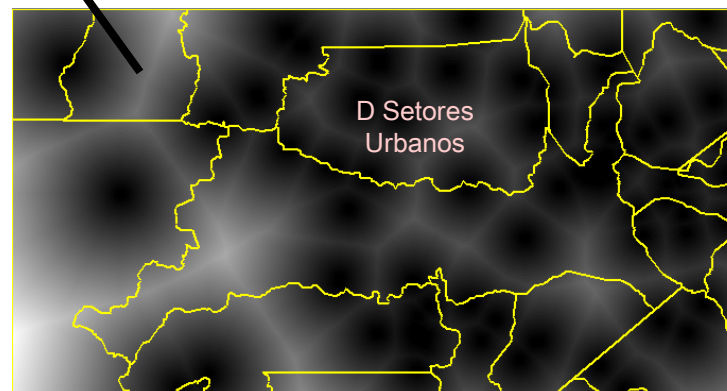
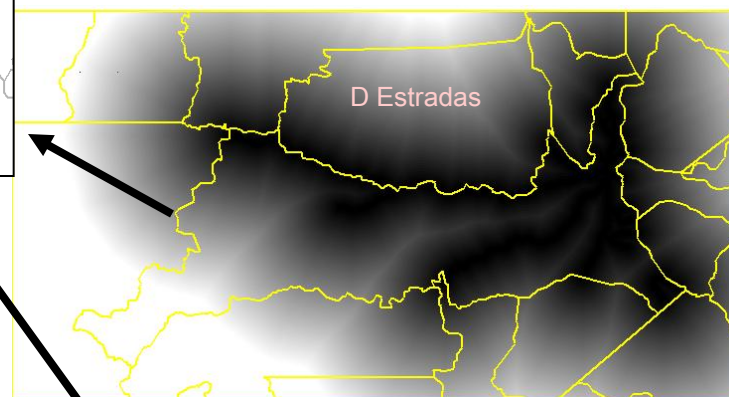
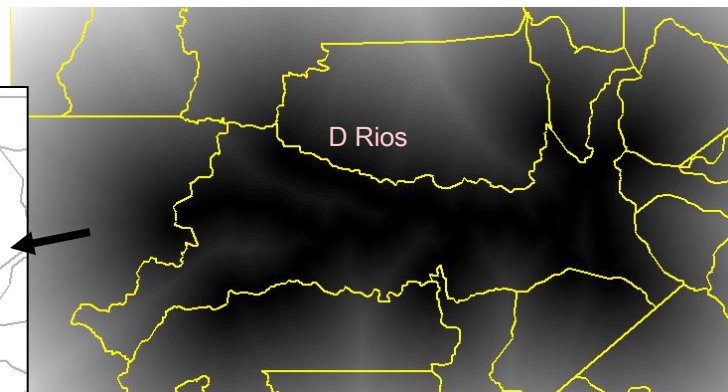
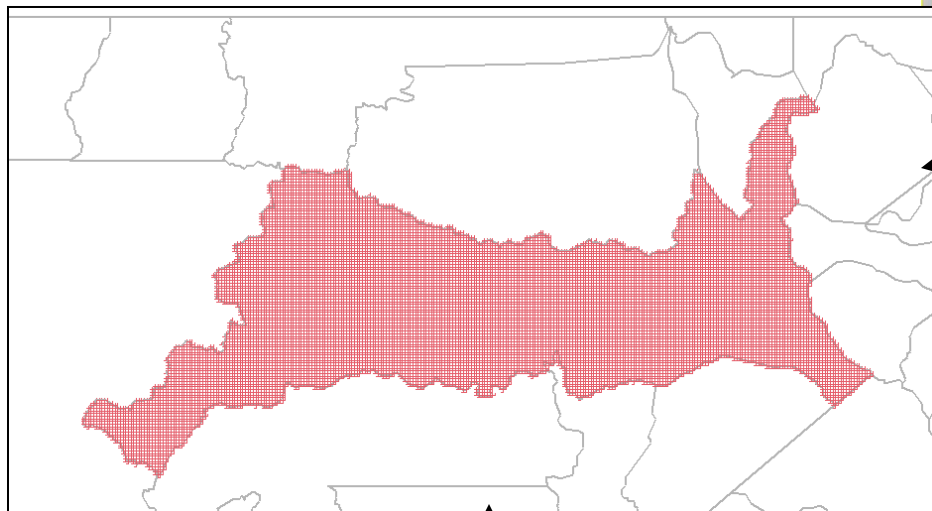
Método Multivariado – Variáveis indicadoras

Mapas

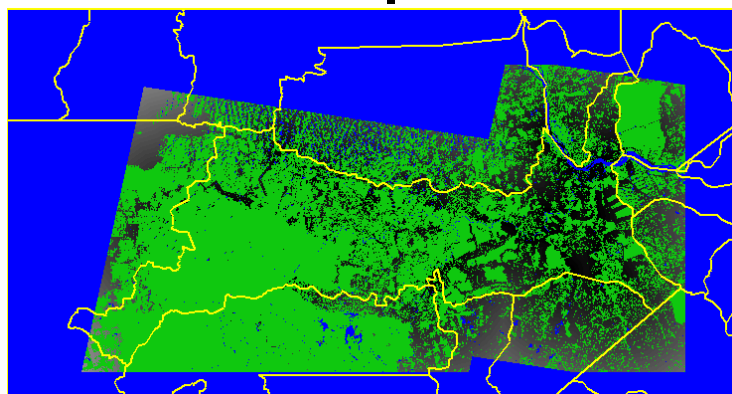
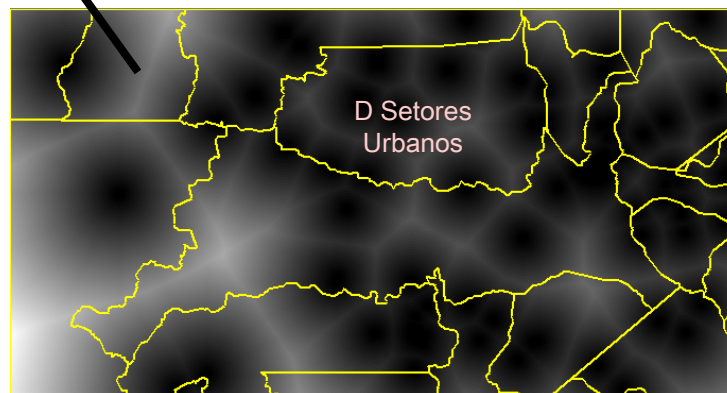
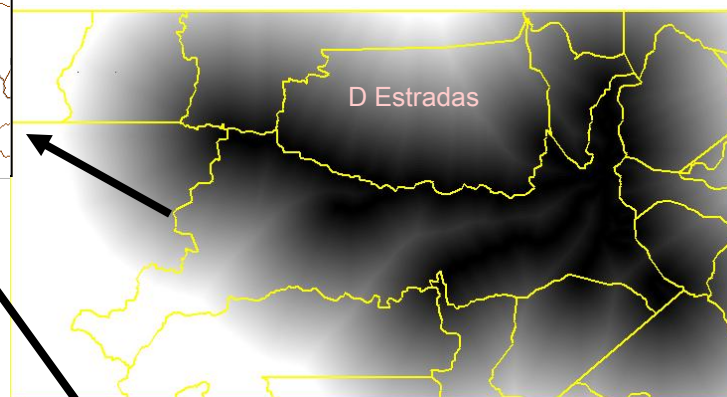
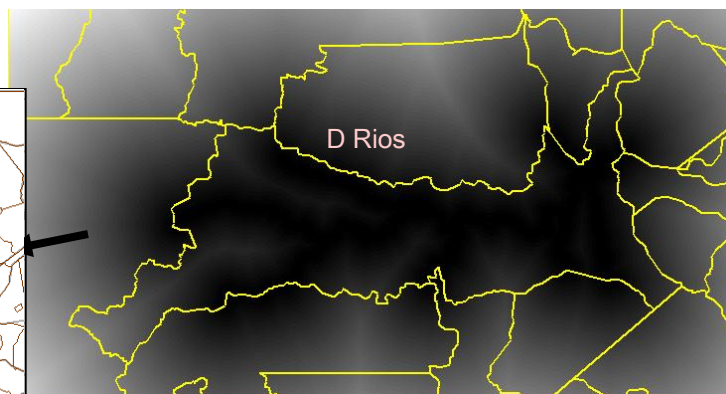
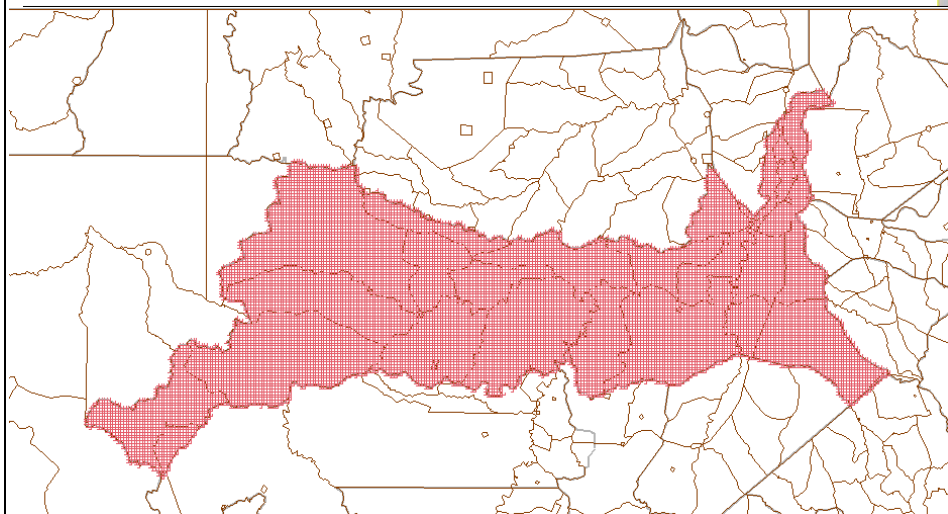
- Distância Rios
- Distância Estradas
- Distâncias setores urbanos
- Classificação Floresta e água



Método Multivariado – Variáveis indicadoras



Método Multivariado – Variáveis indicadoras



Método Multivariado – Variáveis indicad

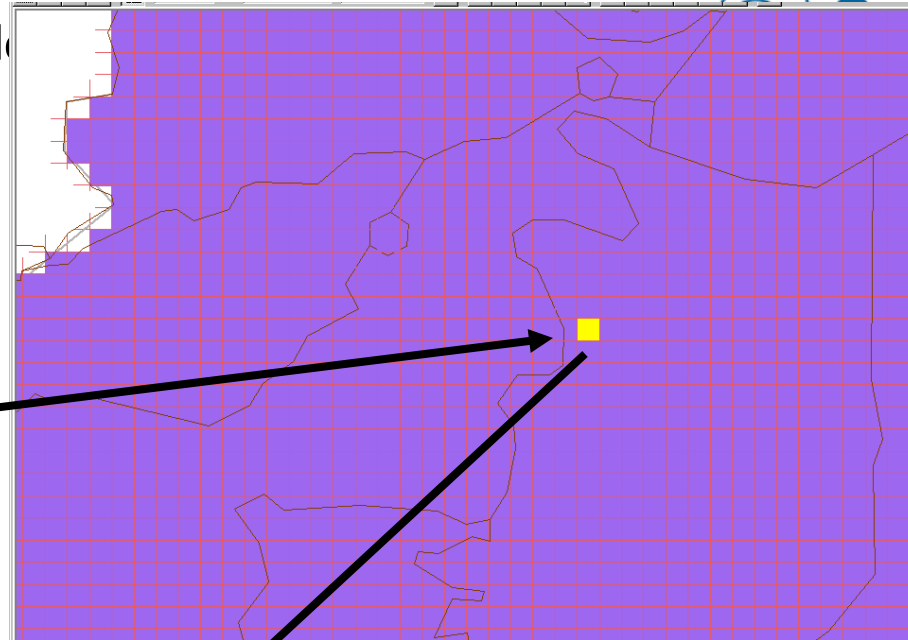
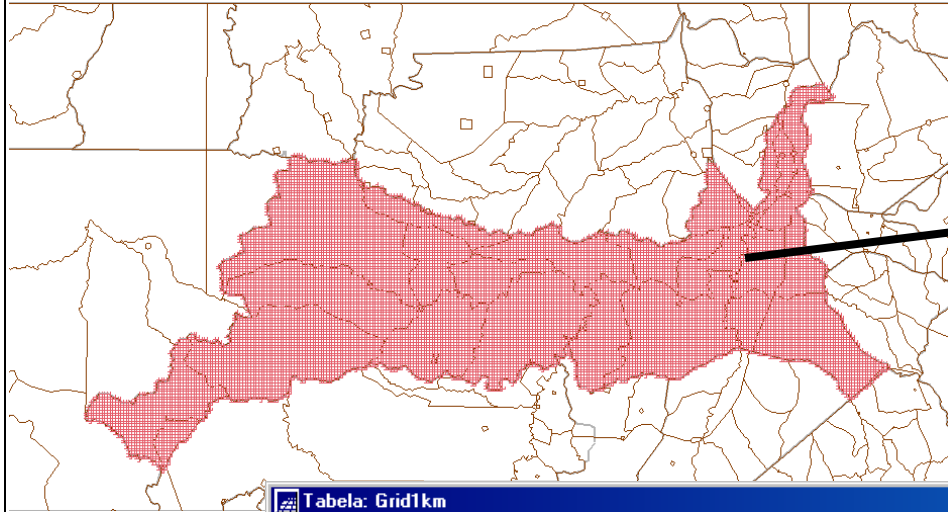


Tabela: Grid1km

	SPERTM	F_FLOR_C	F_DECLI	F_DSURB	F_DRIOS	F_PFLOTM	F_DVIAS	C_TID	SC_POP00	GRID_00
15185	0.20000	1.00000	0.99999	0.90838	0.99806	0.73529	1.00000	999.0000	1060.0000	1.00000
15186	0.39000	1.00000	1.00000	0.92063	1.00000	0.60975	0.99998	999.0000	1060.0000	1.00000
15187	0.83000	1.00000	1.00000	0.93782	0.99991	0.09315	0.99946	008.0000	963.0000	1.00000
15188	0.73000	1.00000	1.00000	0.95310	0.99100	0.11907	0.99701	008.0000	963.0000	1.00000
15189	0.59000	1.00000	0.99999	0.96642	0.96759	0.17650	0.99176	008.0000	963.0000	1.00000
15190	0.38000	1.00000	0.99951	0.97631	0.93216	0.36464	0.98359	008.0000	963.0000	1.00000
15191	0.18000	1.00000	0.99960	0.98339	0.87306	0.81274	0.97270	008.0000	963.0000	1.00000
15192	0.44000	1.00000	1.00000	0.98840	0.78894	0.29123	0.95938	008.0000	963.0000	1.00000
15193	0.31000	1.00000	0.99940	0.99098	0.69624	0.48039	0.94521	008.0000	963.0000	1.00000
15194	0.02000	1.00000	0.98697	0.99236	0.60277	1.00000	0.94180	008.0000	963.0000	1.00000
15195	0.22000	1.00000	0.96949	0.99098	0.51355	0.70942	0.94552	008.0000	963.0000	1.00000
15196	0.11000	1.00000	0.95265	0.98840	0.44042	0.94553	0.94935	008.0000	963.0000	1.00000
15197	0.00000	1.00000	0.99985	0.96642	0.45602	1.00000	0.96030	009.0000	902.0000	1.00000
15198	0.34000	1.00000	1.00000	0.95310	0.46045	0.42633	0.96498	009.0000	902.0000	1.00000
15199	0.00000	1.00000	0.96708	0.97631	0.44476	1.00000	0.95521	009.0000	902.0000	1.00000
15200	0.01000	1.00000	0.90976	0.98339	0.43018	1.00000	0.95267	008.0000	963.0000	1.00000
15201	0.07000	1.00000	0.99978	0.98339	0.43018	1.00000	0.95267	008.0000	963.0000	1.00000

F (Z)

Mundo Celular...

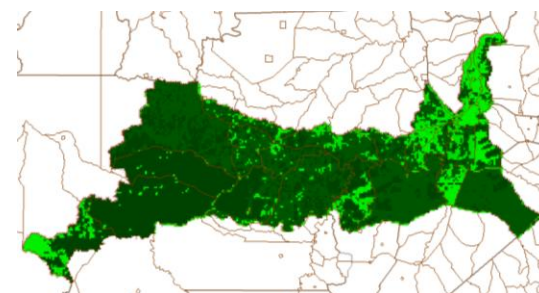
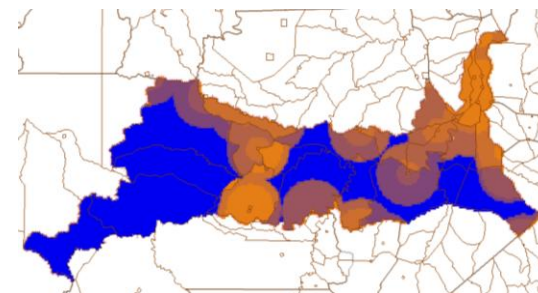
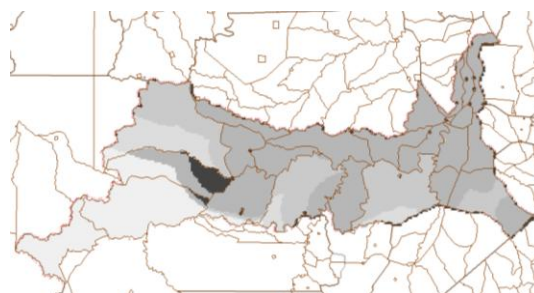
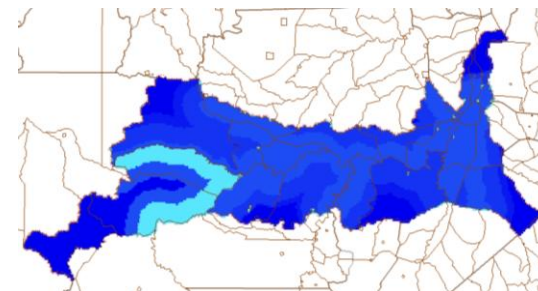
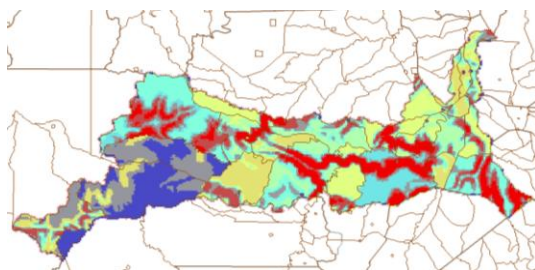


Método Multivariado – Variáveis indicadoras



Mapas (F(z))

- Declividade
- Distância Rios
- Distância Estradas
- Distâncias setores urbanos
- Classificação Floresta e água



E como estas variáveis podem resultar em um valor de possibilidade de ocorrência de população???





Inferência Fuzzy

- Dados em conjuntos Fuzzy manipulados com métodos lógicos da lógica fuzzy ou operadores fuzzy

- AND, OR, Soma algébrica, Produto Algébrico, Operador Gama e Soma Convexa

- OR (otimista)

$$\text{res} = \text{MAX} (\text{crit1}, \text{crit2}, \dots)$$

- AND (pessimista)

$$\text{res} = \text{MIN} (\text{crit1}, \text{crit2}, \dots)$$

- Produto Algébrico

$$\text{res} = \prod (\text{crit1}, \text{crit2}, \dots)$$

- Soma Algébrica

$$\text{res} = 1 - \prod (\text{crit1}, \text{crit2}, \dots)$$

- Operador Gama (compromisso)

$$\text{res} = [\text{soma algébrica} (\text{crit1}, \text{crit2}, \dots)]^\gamma * [\text{produto algébrico} (\text{crit1}, \text{crit2}, \dots)]^{1-\gamma}$$

Método Multivariado – Variáveis indicadoras

Mapas (F(z))

- Declividade
- Distância Rios
- Distância Estradas
- Distâncias setores urbanos
- Classificação Floresta e água

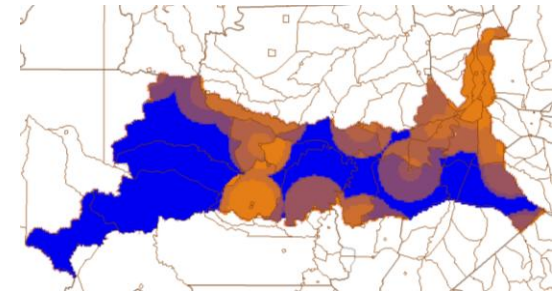
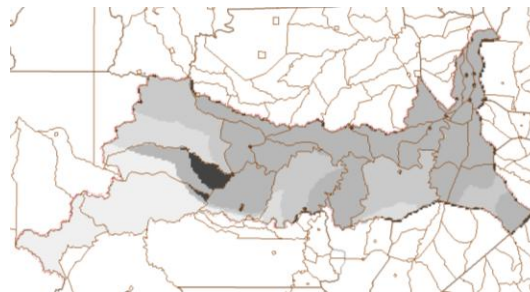
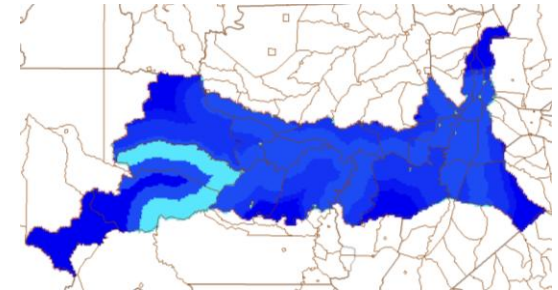
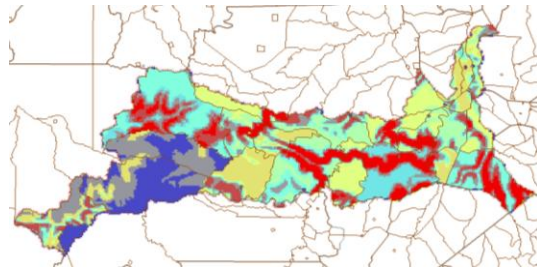
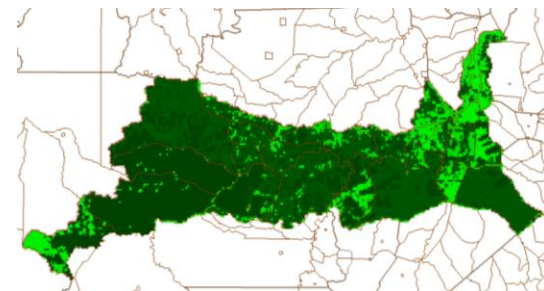


TABELA 6.3 – Operadores para relacionar as variáveis indicadoras.

Operador	Expressão
Média Simples	$(Dvias + Drios + Decl + Durb + Pflor)/5$
Fuzzy Mínimo	Mínimo ($Dvias, Drios, Decl, Durb, Pflor$)
Fuzzy Máximo	Máximo ($Dvias, Drios, Decl, Durb, Pflor$)
Fuzzy Gama	$(1 - (Dvias * Drios * Decl * Durb * Pflor))^{0,2} * (Dvias * Drios * Decl * Durb * Pflor)^{0,8}$
Média Ponderada	$Dvias*0,589 + Drios*0,161 + Durb*0,099 + Pflor*0,082 + Decl*0,032$ #



$Dvias$ = distância a vias; $Drios$ = distância a rios; $Decl$ = declividade média; $Durb$ = distância a centros urbanos; $Pflor$ = percentagem de florestas

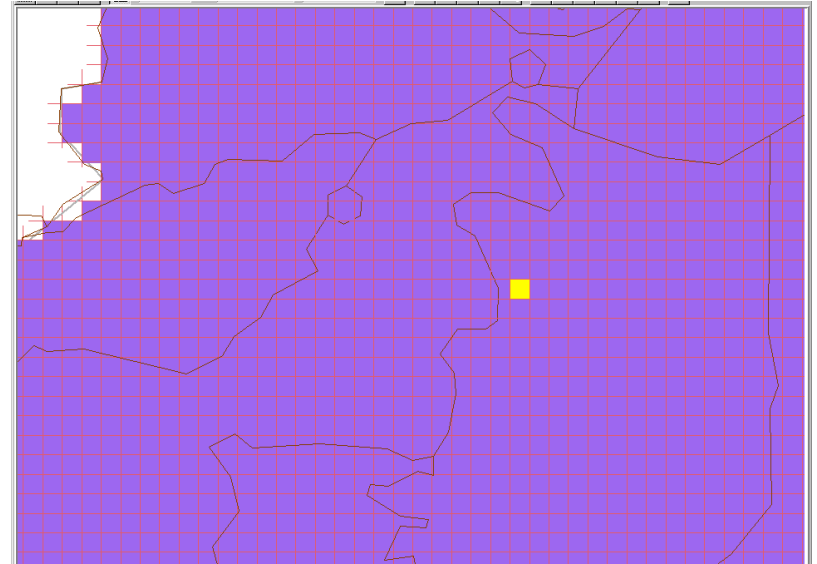
Razão de Consistência igual a 0,056



Método Multivariado – Variáveis indicadoras

E a distribuição da população / célula????

$$DP_{grid} = P_{CS} * \left(\frac{F_{grid}}{\sum_{i=0}^j F_{grid}} \right)$$



Onde:

DP_{grid} é a densidade demográfica da célula,

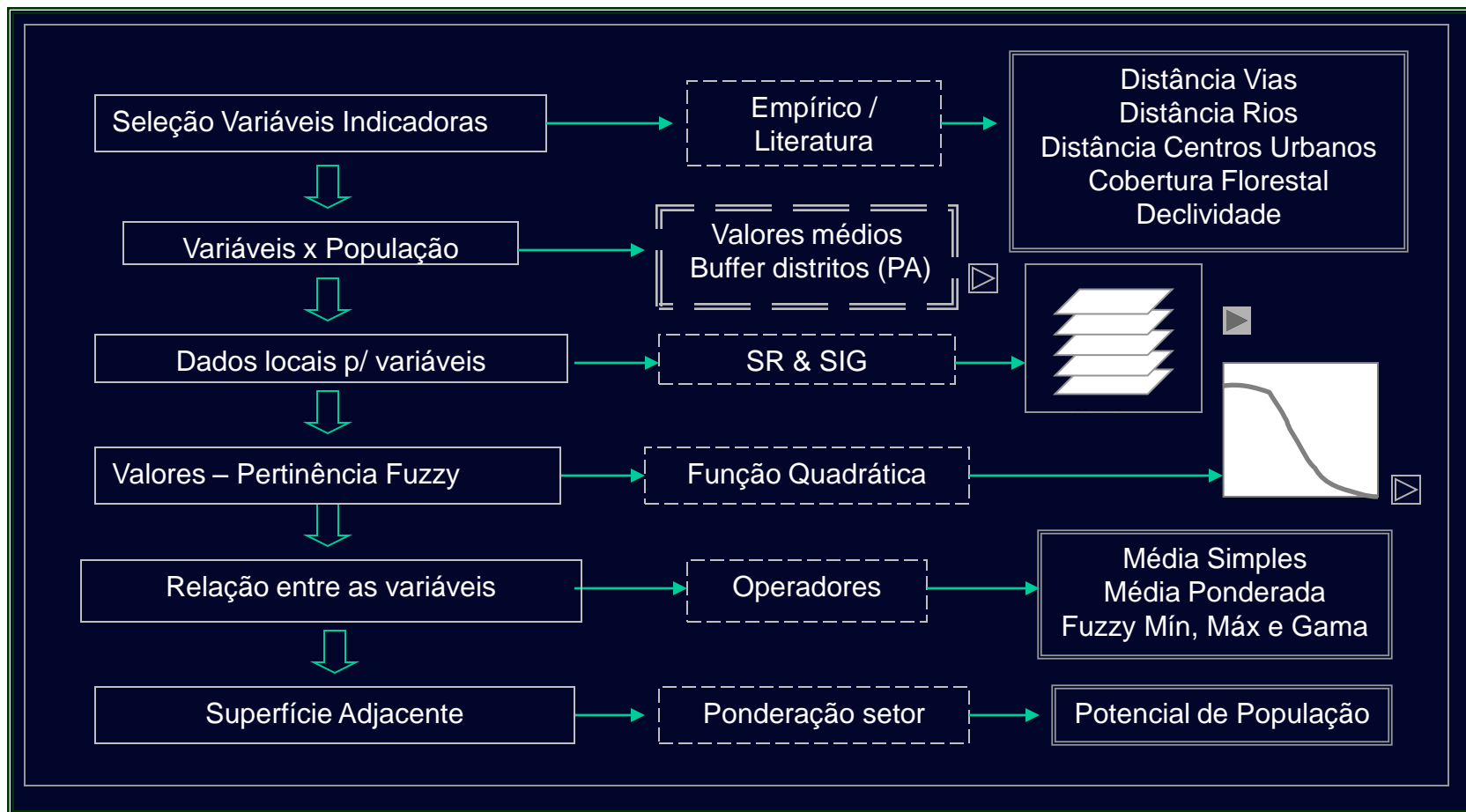
P_{CS} é a população do setor censitário,

F_{grid} é o valor Fuzzy de possibilidade de ocorrência de população para a célula ponderado pela somatória dos valores Fuzzy obtidos para o setor censitário e considerando-se apenas as células válidas, ou seja, com percentagem de floresta e corpos d'água maior que 95%.



Método Multivariado

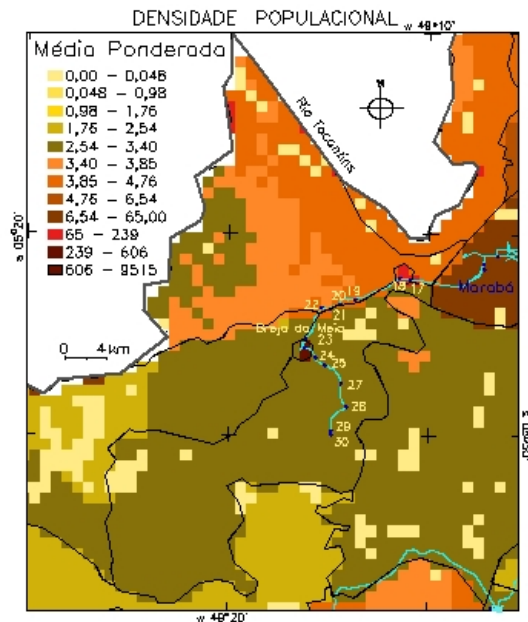
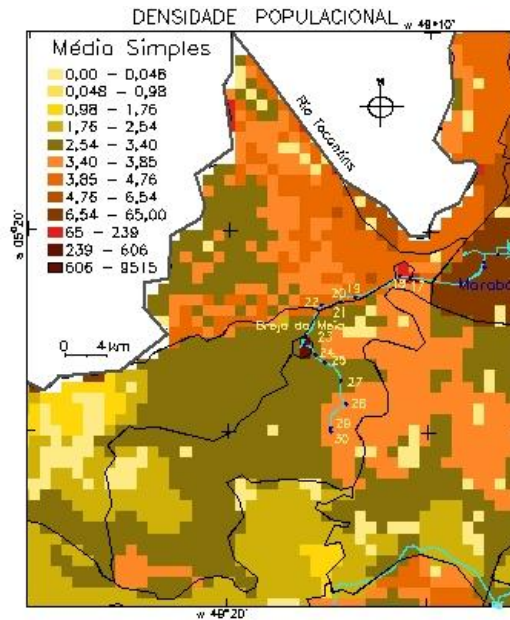
Inferir superfície que descreva distribuição



População em Marabá

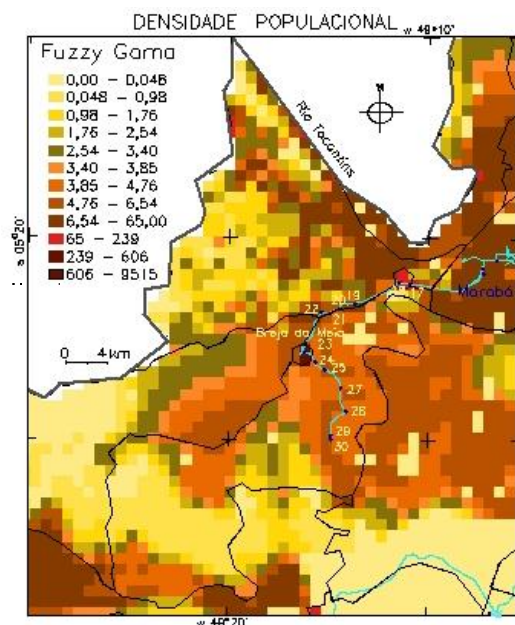
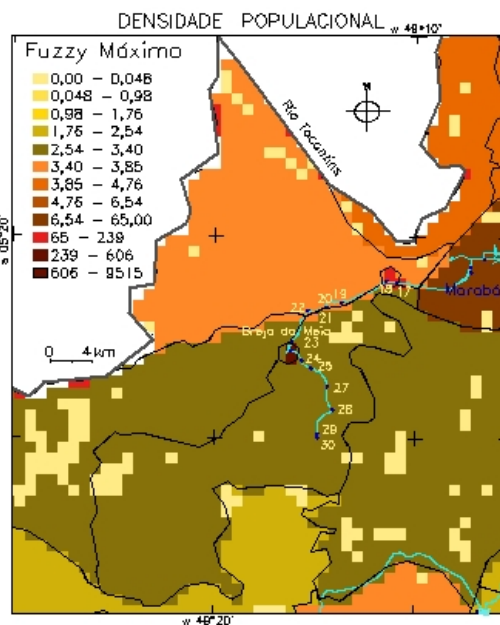
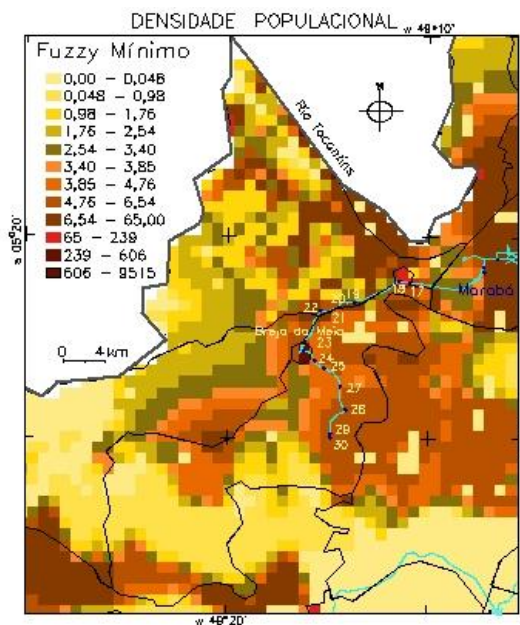
Redistribuição – Superfícies resultantes - Região

- Restrição floresta e água ineficiente
- Percurso de campo
- Média Simples – mais variabilidade
- Média Ponderada superfície mais suave



Redistribuição – Superfícies resultantes - Região

- Fuzzy Mínimo representou melhor a heterogeneidade espacial



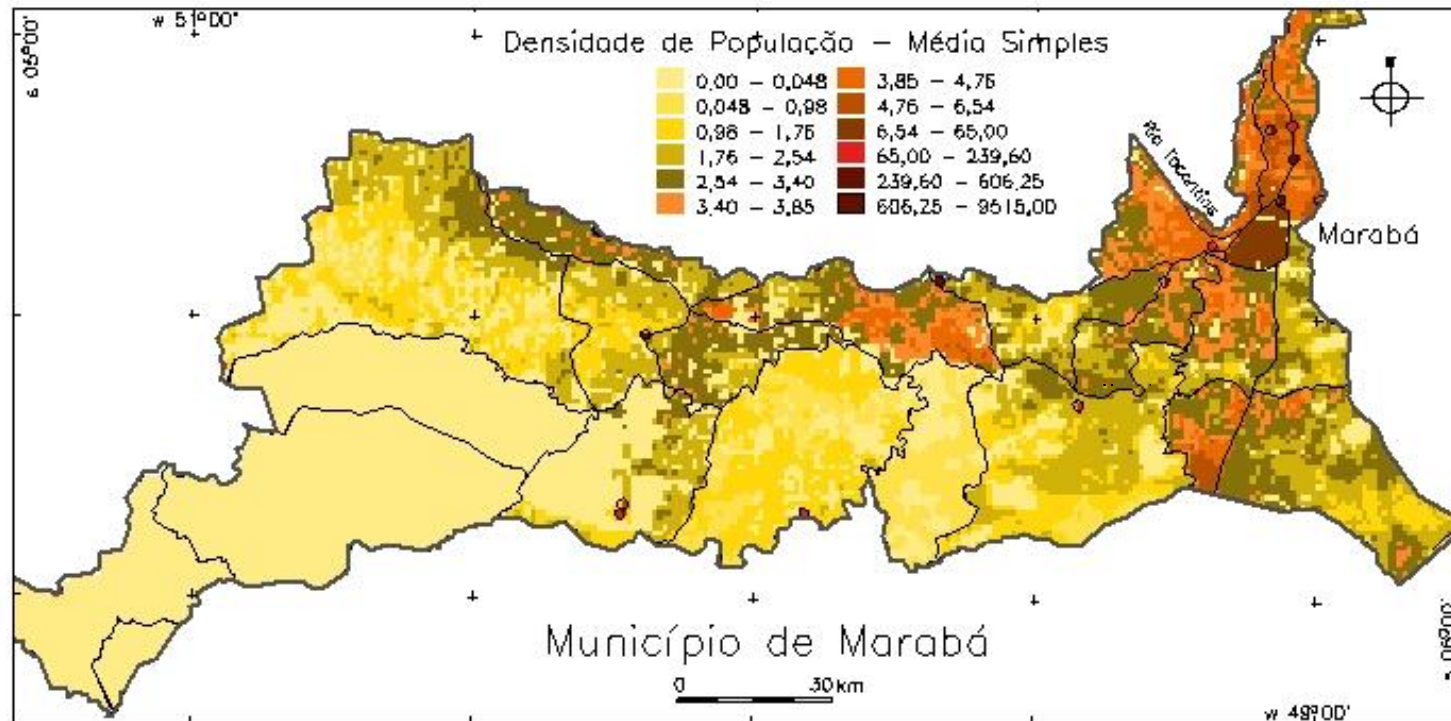
População em Marabá

Redistribuição – Superfícies resultantes – Município



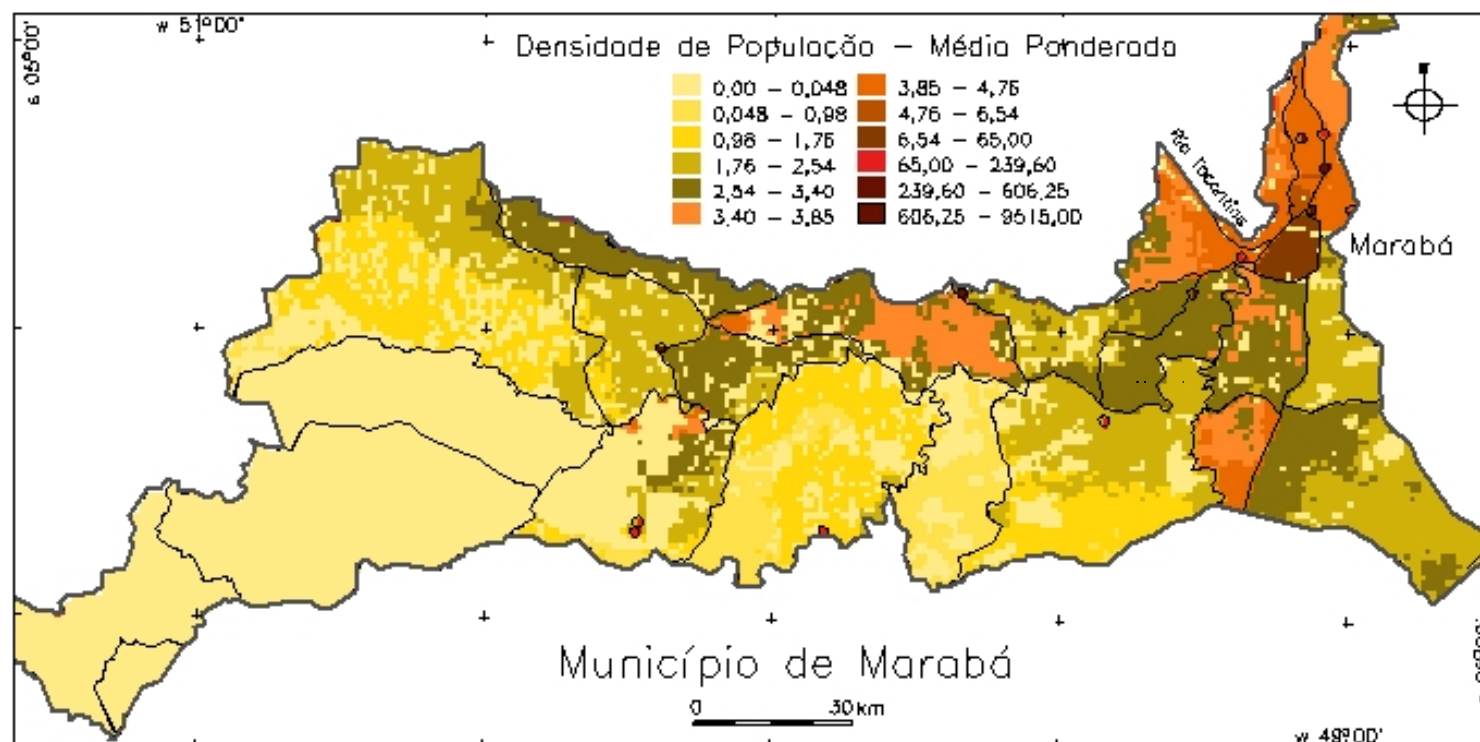
População em Marabá

Redistribuição – Superfícies resultantes – Município



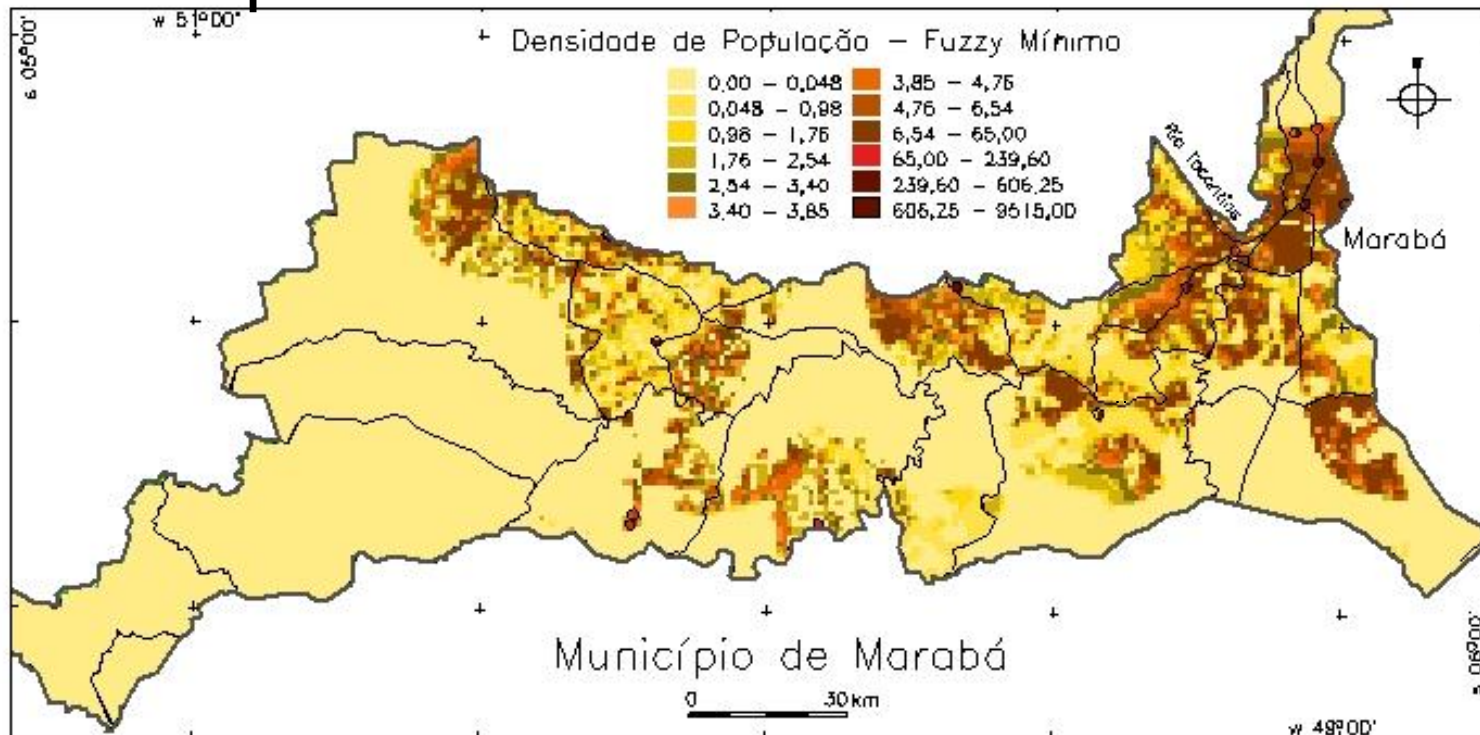
População em Marabá

Redistribuição – Superfícies resultantes – Município



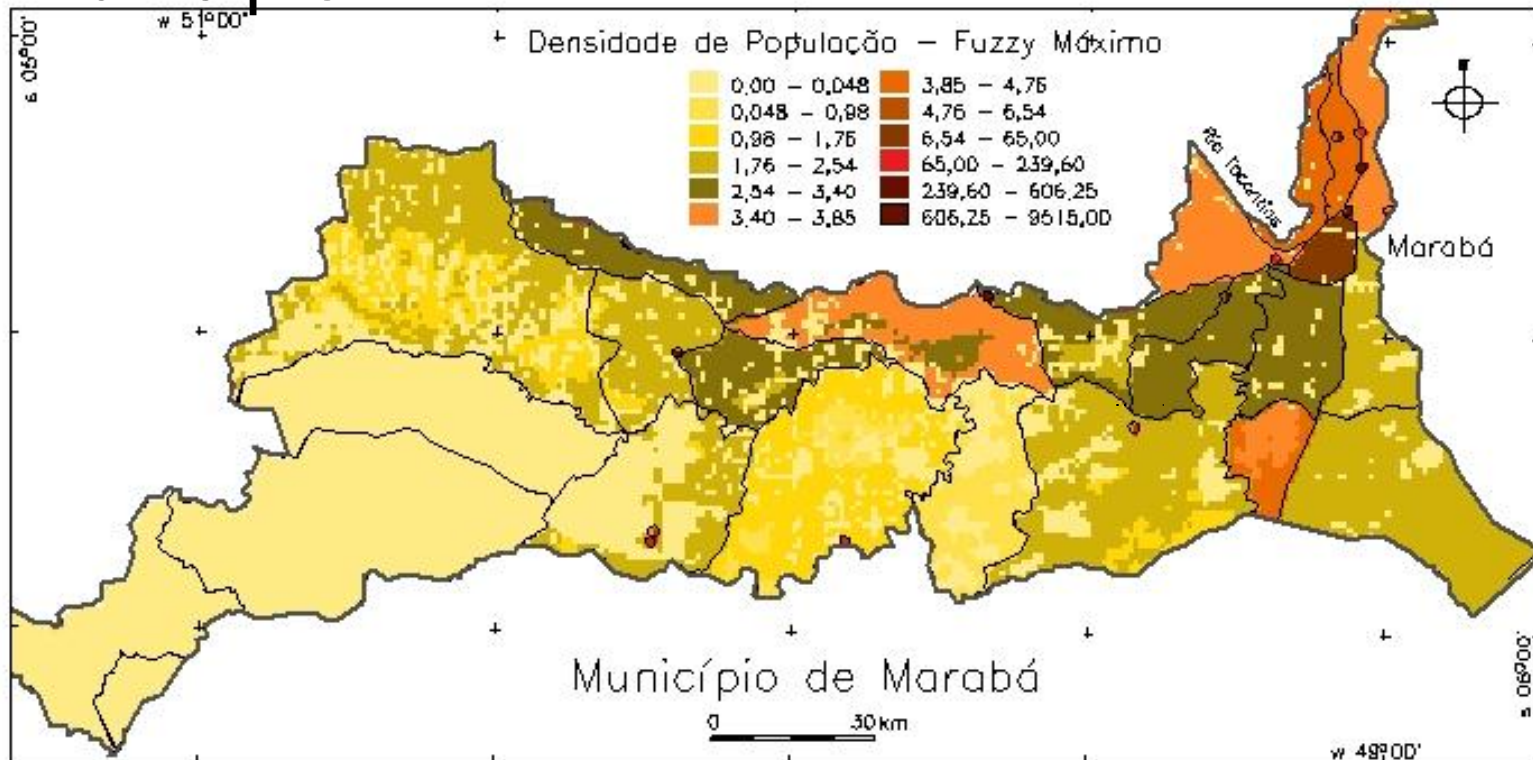
População em Marabá

Redistribuição – Superfícies resultantes – Município



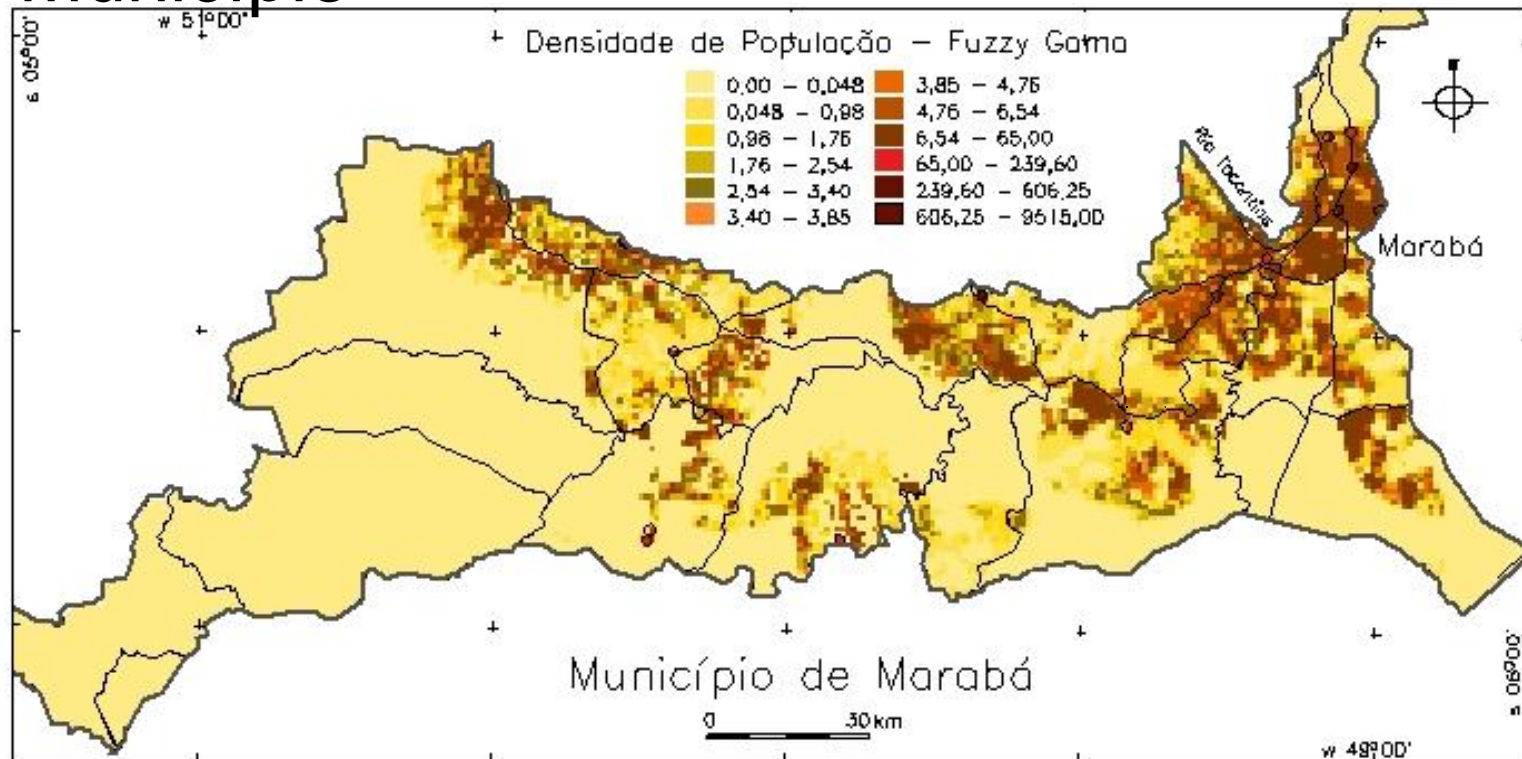
População em Marabá

Redistribuição – Superfícies resultantes – Município



População em Marabá

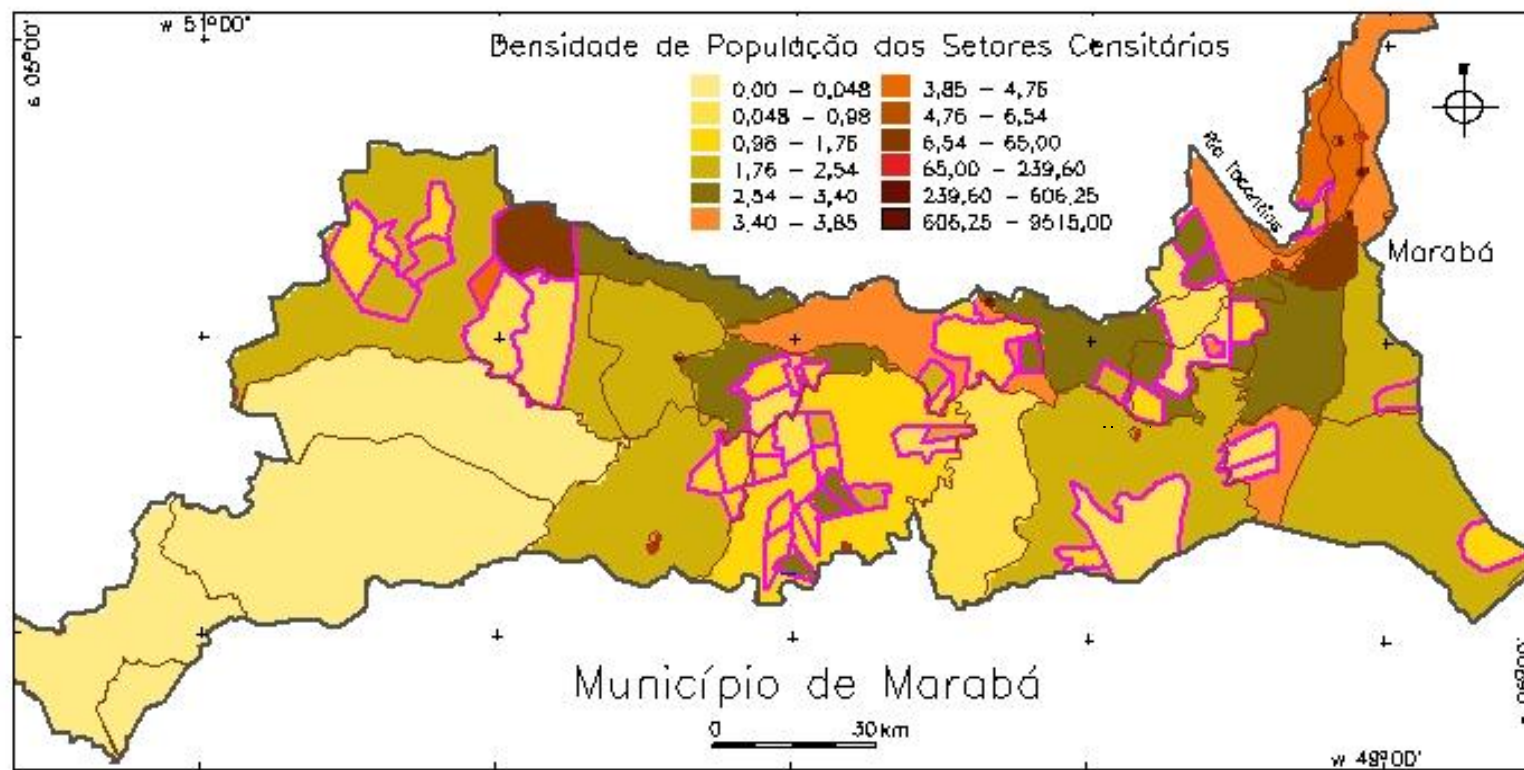
Redistribuição – Superfícies resultantes – Município



População em Marabá



Redistribuição – Superfícies resultantes – Município



População em Marabá

Redistribuição – Superfícies resultantes – Município

- Restrição floresta e água – setores sem população
- Média Simples – mais variabilidade que Média Ponderada (peso para % floresta)
- *Fuzzy* Mín e Gama similares
- *Fuzzy* Máx semelhante Setores com restrição inicial
- Dados de pessoas nos PAs do INCRA para análise global

Superfície de Densidade	Exatidão Global (%)
Média Simples	14,3
Média Ponderada	10,4
<i>Fuzzy</i> Mínimo	10,4
<i>Fuzzy</i> Máximo	9,5
<i>Fuzzy</i> Gama	18,8
Setores Censitários	11,8

- **Fuzzy gama – acertos nas classes de densidade extremas**
- **Média Simples – acerto nas classes intermediárias**

TO READ....

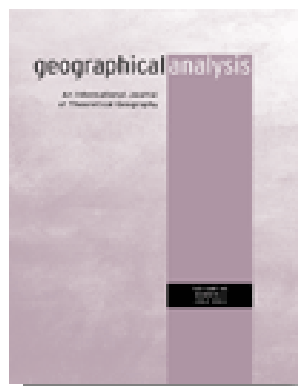


geographical analysis

Geographical Analysis



© The Ohio State University



Special Issue: Areal Interpolation and Dasymetric Modeling

July 2013

Volume 45, Issue 3

Pages 213–344

[Previous Issue](#) | [Next Issue](#)

Modelos de População “Multivariados”

- Interpoladores Inteligentes:

- *Smarter SIM*

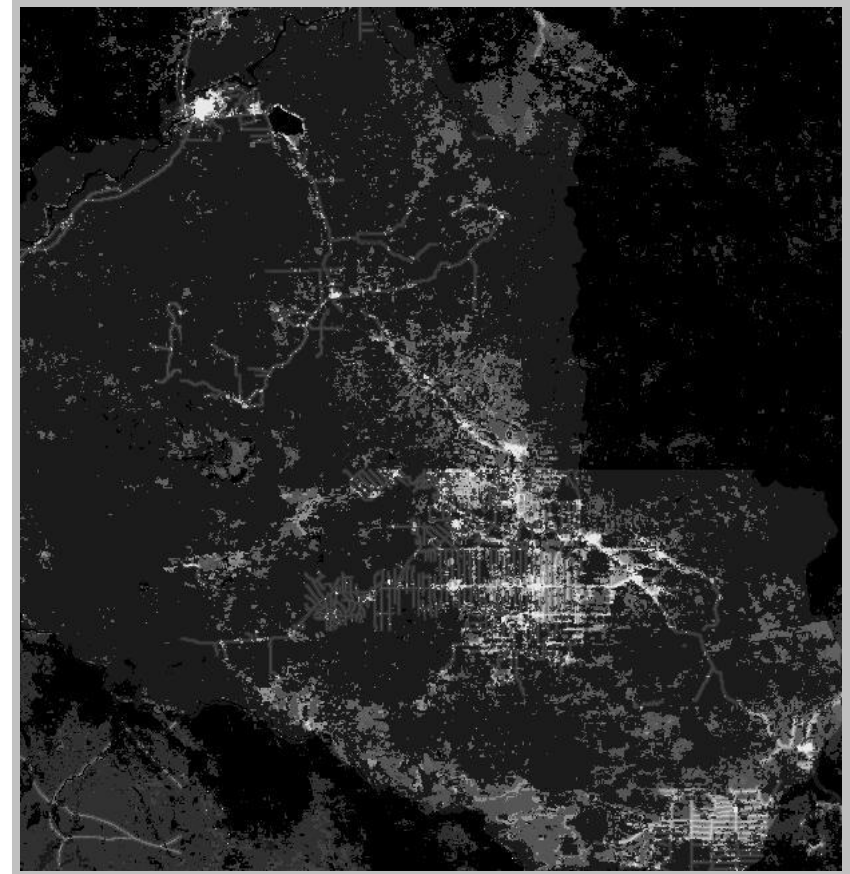
- uso de variáveis de posição, densidade e distância como indicadoras de população
 - uso de redes neurais para mapear as variáveis preditoras de entrada nas variáveis de interesse

- *Clever SIM – Smart SIM* melhorado:

- Pré-processamento mais sofisticado das variáveis espaciais preditoras
 - *Bootstrap* no treinamento- melhor resultado volta para o treinamento, recursivo
 - Treinamento: zero, alta e baixa densidade e combinando-as por inferência fuzzy.

Modelos de População “Multivariados”

- Interpoladores Inteligentes:
 - Exemplo: *LandScan* – grade 1km, 1995
 - Modelo de População: uso do solo, proximidade de estradas, luzes noturnas => **coeficiente de probabilidade**
 - População em risco: Informação para medidas emergenciais – desastres naturais ou antropogênicos





GIST



Geographic Information Science and Technology

Computational Sciences & Engineering Division

- Home
- Data Availability
- Documentation
- Frequently Asked Questions
- Related References
- Sample Dataset
- LandScan Usage
- References
- Credits/Disclaimers

FOR TRIAL ACCESS

Contact East View



LandScan™

Using an innovative approach with Geographic Information System and Remote Sensing, ORNL's LandScan™ is the community standard for global population distribution. At approximately 1 km resolution (30" X 30"), LandScan is the finest resolution global population distribution data available and represents an **ambient population** (average over 24 hours). The LandScan algorithm, an R&D 100 Award Winner, uses spatial data and imagery analysis technologies and a multi-variable dasymmetric modeling approach to disaggregate census counts within an administrative boundary. Since no single population distribution model can account for the differences in spatial data availability, quality, scale, and accuracy as well as the differences in cultural settlement practices, LandScan population distribution models are tailored to match the data conditions and geographical nature of each individual country and region.

HIGHLIGHTS

Mapping Technology Aids Tsunami Recovery. . .
Relief agencies assist victims of the Asian tsunami using a demographic database developed at ORNL.

RESEARCH AREAS

- Download Citation in .RIS Format
- Geographic Information Science and Technology (GIST)

DOWNLOADS

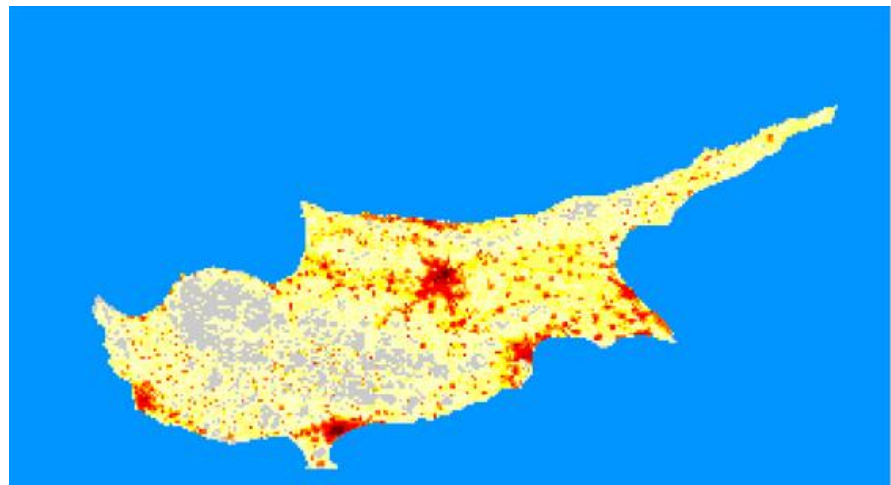
- Download Citation in .RIS Format

LS2012 | LS2011 | LS2010 |
LS2009 | LS2008 | LS2007 |
LS2006 | LS2005 | LS2004 |
LS2003 | LS2002 | LS2001 |
LS2000 | LS1998

Landscan - <http://web.ornl.gov/sci/landscan/index.shtml>

- Abordagem GIS & RS
- ORNL's LandScan™ - padrão para distribuição de população global.
- Resolução ~ 1 km resolution (30" X 30")
- Dados representando população ambiental (média 24 h).
- Algoritmo LandScan: dados espaciais e imagens em modelo dasimétrico multi-variável para desagregar contagens(censo) dos limites administrativos
- Ajusta as condições dos dados e natureza geográfica de cada país e região
- Ex: dado – Chipre (2011)
(N de pessoas/células)
30 arc sec; Geographic WGS84

- Fees – caso-a-caso !



Modelos de População “Multivariados”

- Interpoladores Inteligentes e Variantes:
 - Importância da escolha de variáveis e conhecimento das inter-relações – “modelo”
 - Existência e qualidade dos dados auxiliares determinam a exatidão da superfície resultante.



Para Visitar...

- <http://web.ornl.gov/sci/gist//landscan/>
- <http://www.grida.no/db/gis/prod/html/glohuma1.htm>
- <http://www.ciesin.org/datasets/gpw/globldem.doc.html>
- <http://www.geog.leeds.ac.uk/people/a.turner/projects/medalus3/home.htm>
- <http://www.geog.ucsb.edu/~tobler/publications/reprints.html>
- <http://www.geog.soton.ac.uk/users/martindj/davehome/software.htm>
- <http://census.ac.uk/cdu/software/surpop/>



Referências

- Tobler, W.R. (1979). Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association*, 74, 519-530.
- Tobler, W.R., Deichmann, U., Gottsegen, J. & Maloy, K. (1995). *The Global Demography Project*. Santa Barbara, CA: National Center for Geographic Information and Analysis.
- Martin, D (1989). Mapping population data from zone centroid locations. *Transactions of the Institute of British Geographers NS*, 14, 90-97.
- Martin, D. (1996). *Geographic Information Systems and their Socioeconomic Applications*, London: Routledge.
- Martin, D. (2002). Census Population Surfaces. In Rees, P., Martin, D., Williamson, P. (Eds.), *The Census Data System* (pp. 139-148). Chichester - England: John Wiley & Sons.
- Martin, D., Langford, M. & Tate, N.J. (2000). Refining Population Surface Models: Experiments with Northern Ireland Census Data. *Transactions in GIS*, 4, 343-360.
- Deichmann, U., Balk, D. & Yetman, G. (2001). *Transforming Population Data for Interdisciplinary Usages: From census to grid*. Palisades, N.Y.: Center for International Earth Science Information Network (CIESIN) - Columbia University.
- Dobson, J.E., Bright, E.A., Coleman, P.R., Duree, R.C. & Worley, B.A. (2000). LandScan: A Global Population Database for Estimating Populations at Risk. *Photogrammetric Engineering and Remote Sensing*, 66, 849-857.
- Faure, J.F., Tran, A., Gardel, A. & Polidori, L. (2003). *Sensoriamento remoto das formas de urbanização em aglomerações do litoral Amazônico: elaboração de um índice de densidade populacional*. Paper presented at XI Simpósio Brasileiro de Sensoriamento Remoto, Belo Horizonte, 5-10 Abril 2003.

Krig x Pycnophylactic



E. Yoo, P.C. Kyriakidis, W. Tobler

Reconstructing Population Density Surfaces from Areal Data: A Comparison of Tobler's Pycnophylactic Interpolation Method and Area-to-Point Kriging

Geographical Analysis 42 (2010) 78–98 2010

GOAL: Abordagem geoestatística de krigagem area-para-ponto (ATP) para distribuir dados de população coletados por unidades de área em 18 setores censitários (census tracts) in Ann Arbor de 1970 para reconstruir uma superfície de densidade de população.

Krig x Pycnophylatic



E. Yoo, P.C. Kyriakidis, W. Tobler

Reconstructing Population Density Surfaces from Areal
Data: A Comparison of Tobler's Pycnophylactic
Interpolation Method and Area-to-Point Kriging

Geographical Analysis 42 (2010) 78–98 r 2010

Krig x Pycnophylatic

Estabelecer relações entre a variografia logarítmica utilizada em geoestatística e o critério de suavização do método de Tobler.

Construir uma superfície:

- A regularidade da superfície desconhecida é governada por uma função determinística;
- Equação que soluciona a superfícies é governada pela diferencial parcial de Laplace (~Tobler)
- Apresenta uma solução para aproximar a superfície contínua a partir dos dados de área baseado na premissa que a regularidade (smoothness) da superfície adjacente é conhecida, e não necessariamente é a descrita pela equação diferencial de Laplace.

Krig x Pycnophylatic



Geographical Analysis 42 (2010) 78–98 r 2010

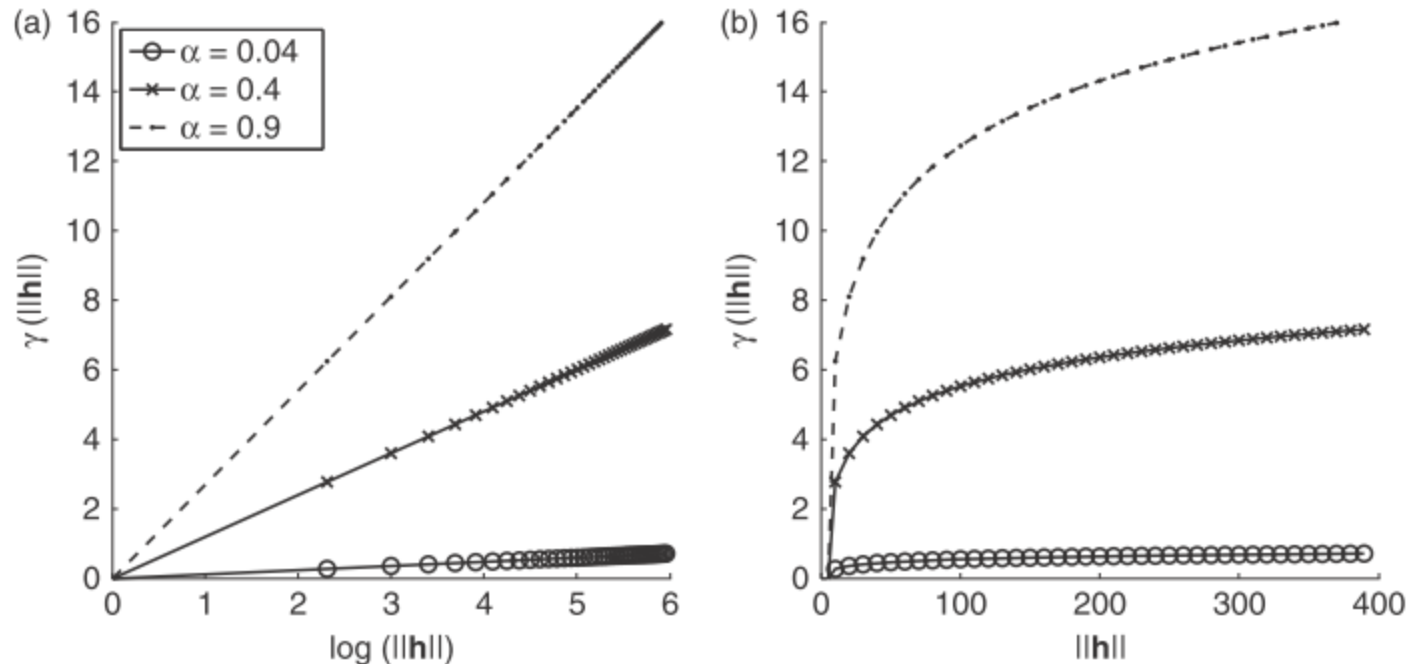


Figure 1. Illustration of logarithmic or de Wijsian variograms with different *absolute dispersion* parameters. The solid lines with circles and x-marks denote, respectively, the de Wijsian model with absolute dispersion parameters 0.04 and 0.4; the dashed-dotted line denotes the variogram model with $\alpha = 0.9$. (a) Plots of de Wijsian variograms versus distance on log scale. (b) Plots of de Wijsian variograms versus distance.

Krig x Pycnophylatic

Superfície de densidade de população a partir de dados coletados em áreas

18 census tracts, Ann Arbor 1970

Eun-Hye Yoo et al.

Reconstructing Population Density Surfaces from Areal Data

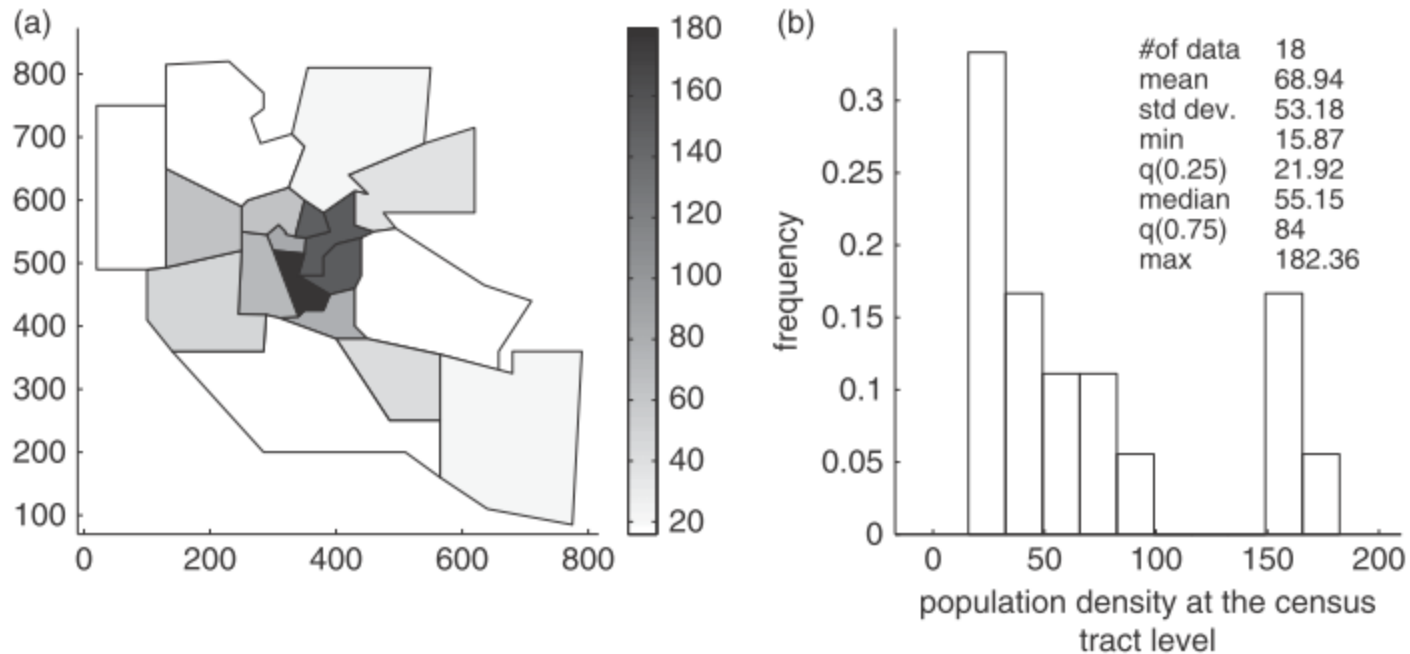


Figure 2. (a) Choropleth map representation of population density at the census tracts in Ann Arbor (1970). (b) Histogram and summary statistics of the mean population density values at census tracts.

Krig x Pycnophylatic



Comparação Tobler x ATP Kriging

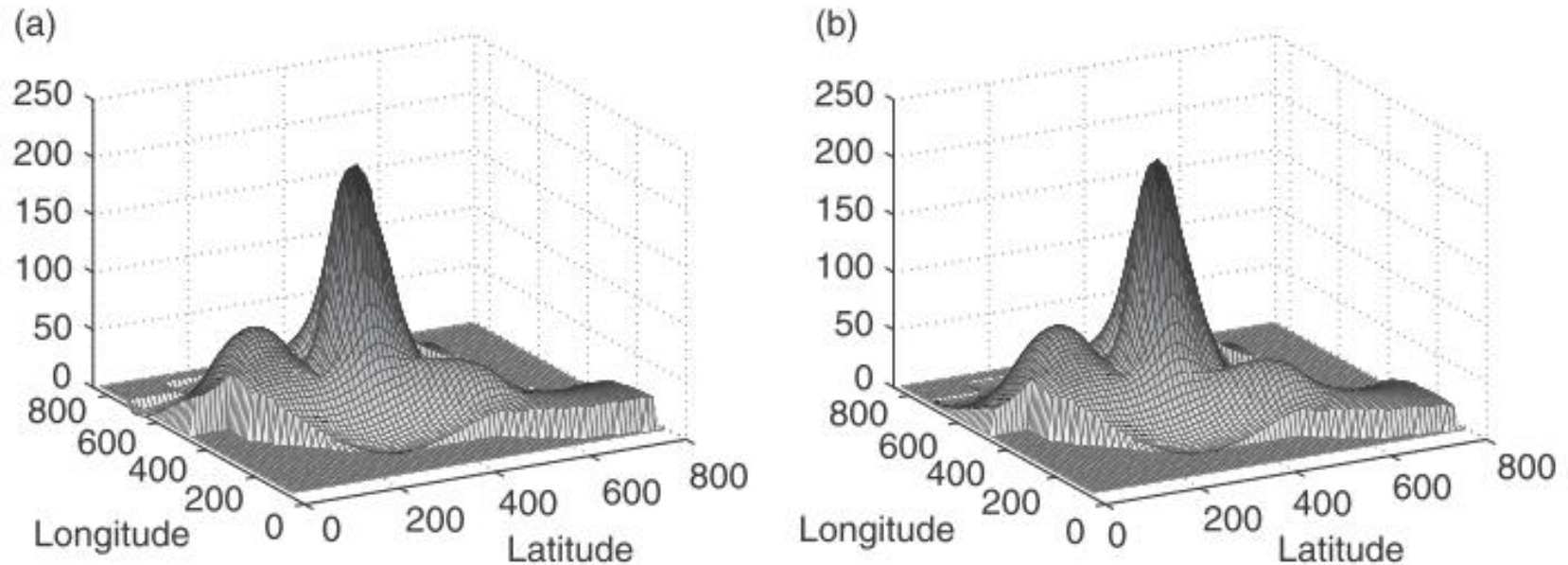
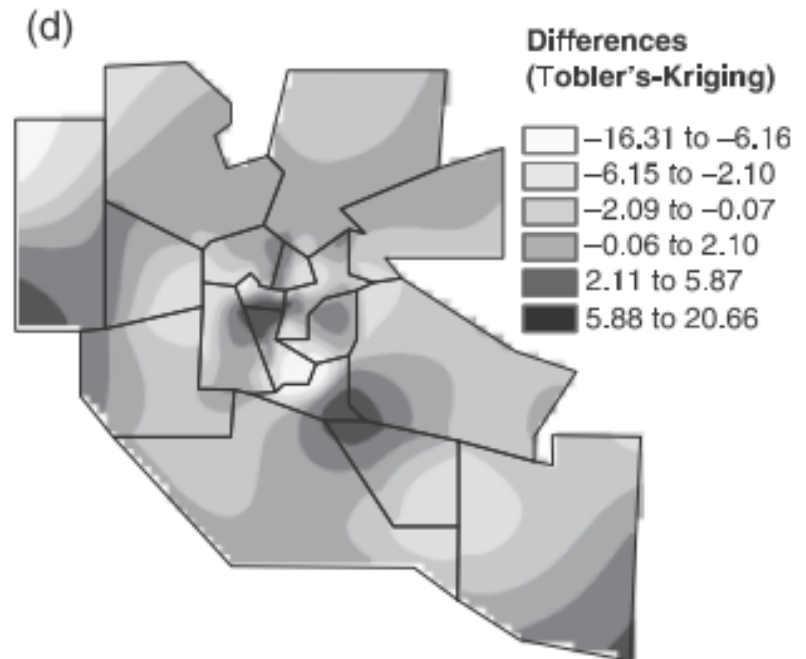
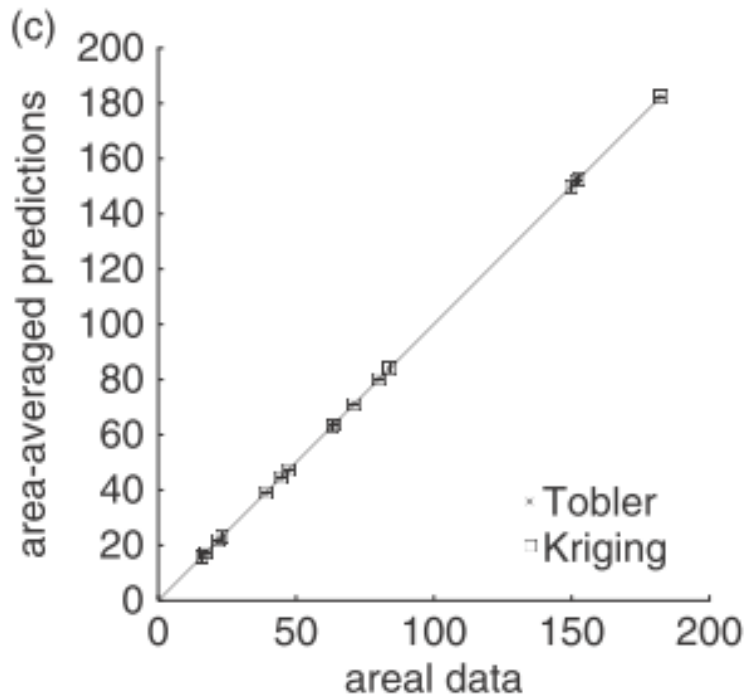


Figure 3. (a) Tobler's solution surface under a Neumann-type boundary condition.
(b) Area-to-point (ATP) kriging prediction surface with boundary conditions at infinity.

Krig x Pycnophylatic



Comparação Tobler x ATP Kriging



(c) Scatter-plot of original areal data (population density at the census tract level of Fig. 2a) versus area-averaged point predictions derived by Tobler's method (asterisks) and by kriging (squares).

(d) Map of differences between point predictions obtained by Tobler's interpolation method (a) and ATP kriging (b).

Krig x Pycnophylatic



Comparação Tobler x ATP Kriging

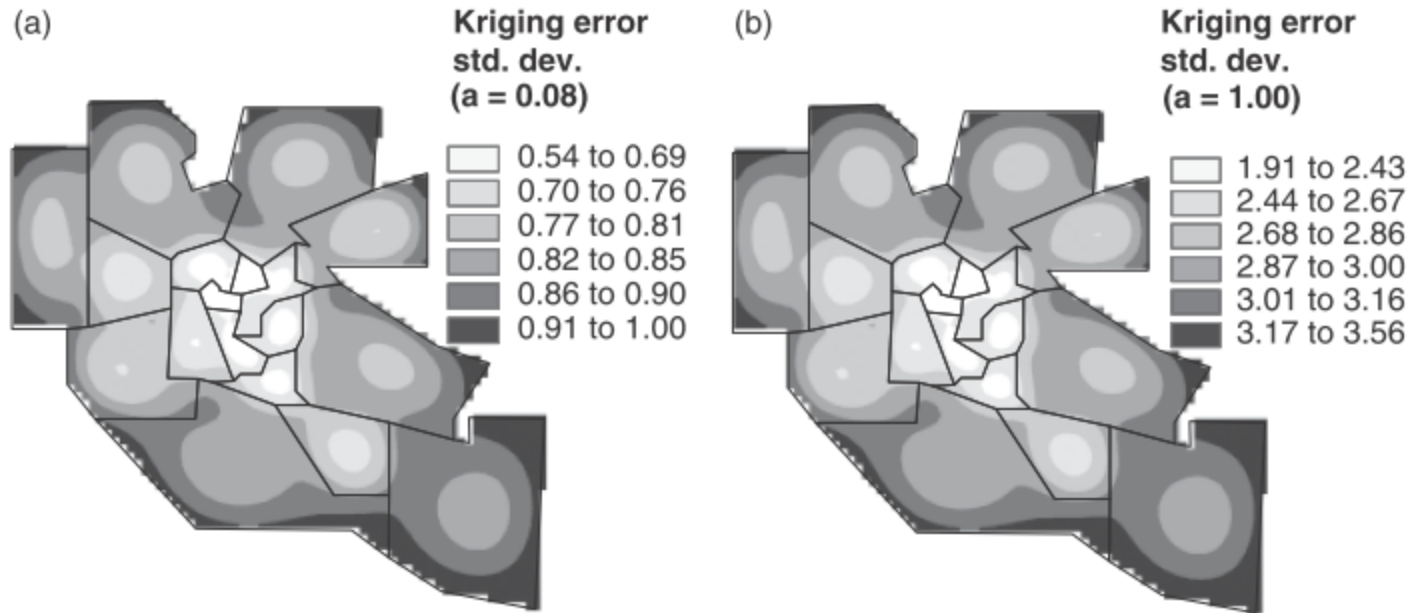


Figure 4. Error standard deviations associated with the area-to-point kriging predictions of Fig. 3b with absolute dispersion $\alpha \approx 0.08$ (a) and $\alpha \approx 1$ (b).

Resultados

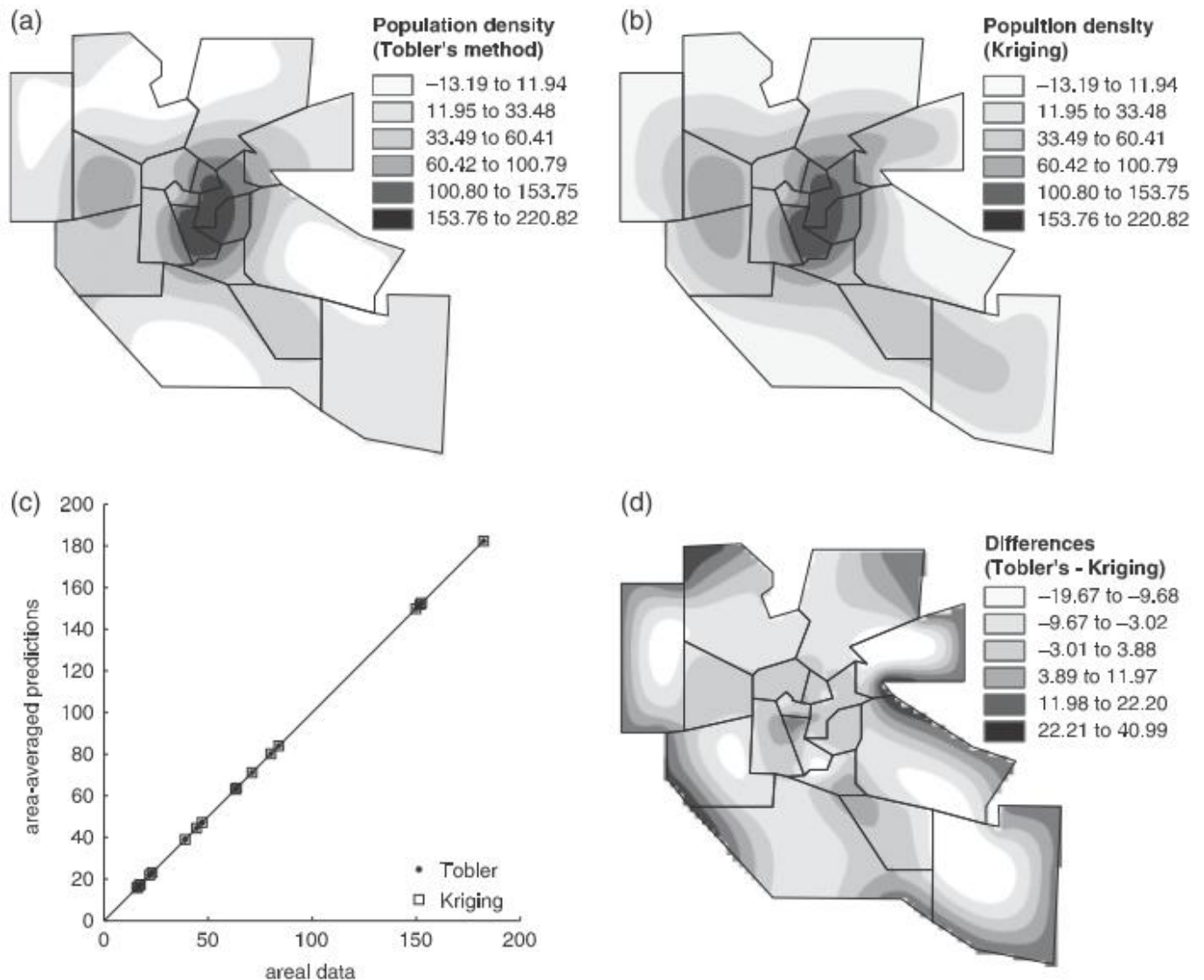


Figure 5. Population density surface derived from Tobler's interpolation method (a) and from area-to-point (ATP) cokriging (b), both under a Dirichlet-type boundary condition. (c) Scatter plot of original areal data (of Fig. 2a) versus area-averaged point predictions derived by Tobler's method (asterisks) and by kriging (squares). (d) Map of differences between point predictions obtained by Tobler's interpolation method (a) and ATP cokriging (b).

Krig x Pycnophylatic



Comparação Tobler x ATP Kriging

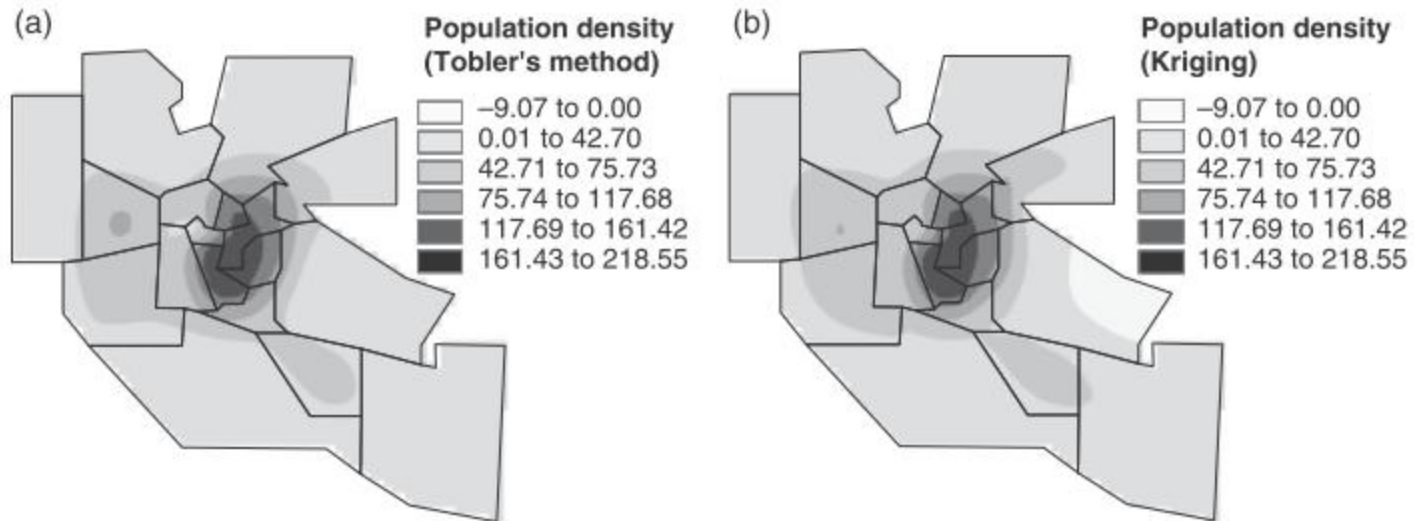


Figure 7. Population density surface predicted from Tobler's interpolation method (a) and from areal-to-point cokriging (b), both subject to a Dirichlet-type boundary condition and non-negativity constraint.

Resultados

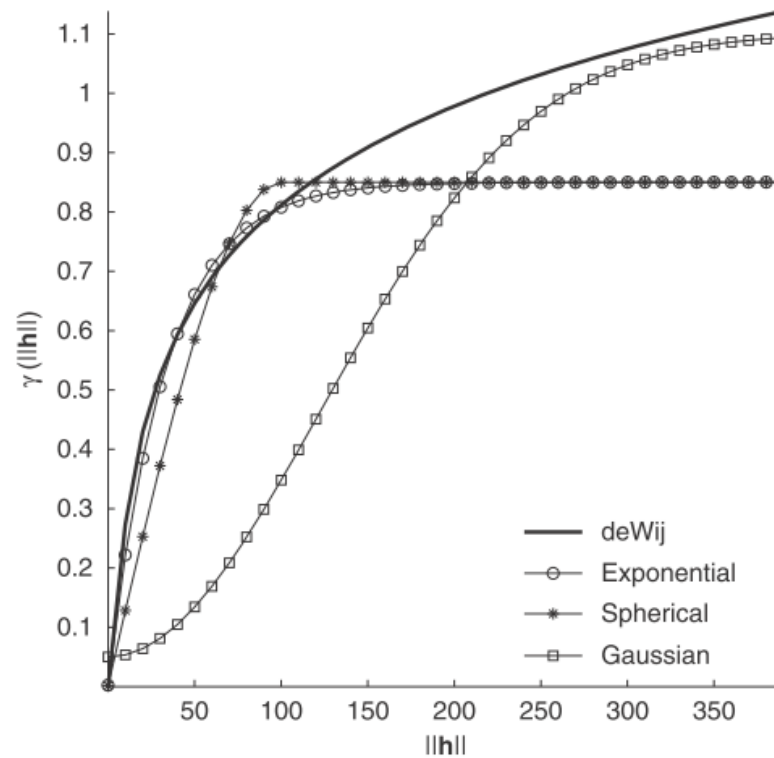


Figure 8. Alternative point variogram models used for area-to-point cokriging: de Wijsian model (solid line), exponential model with effective range 100 and sill 0.85 (solid line with circles), spherical model (solid line with stars) with same parameters as the exponential, and Gaussian model with effective range 300 and sill 1.05 plus a very small nugget effect (solid line with squares).

Table 1 Summary Statistics Comparing Tobler’s Solution Surface with Alternative Kriging-Derived Surfaces Constructed Using Different Point Variogram Models

Statistic/Model	de Wijsian	Exponential	Spherical	Gaussian
Correlation coefficient	0.97	0.96	0.95	0.84
Mean absolute error	7.51	7.75	9.74	18.24
Root mean squared error	9.60	9.97	12.21	23.22

Krig x Pycnophylatic

Resultados:

Geoestatística com Modelo de variograma específico – (Wijsian – correspondente a função “free-space Green” da equação de Laplace) e condições e restrições semelhantes a abordagem de Tobler →

Superfície de densidade de população predita é quase idêntica !!!

“results imply that the ATP ordinary kriging with the de Wijsian model provides the closest approximation to Tobler’s solution surface among other point variogram models”

Krig x Pycnophylatic



“we believe it is this Laplacian smoothness criterion that should be the defining feature of Tobler’s method, not its pycnophylactic property”

IMPLICAÇÕES

(1) Superfícies de atributos múltiplos a partir de dados de área podem ser construídas a partir de geoestatística, dependendo do modelo de variograma adotado:

- precisa ser relacionado à solução de Tobler

(2) É responsabilidade do analista justificar se o critério de Tobler empregado é relevante para sua aplicação

(3) Vantagem da geoestatística: permite reportar as incertezas / confiabilidade sobre a região interpolada, com implicações críticas para propagação de incertezas em operações de análise espacial.

Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data

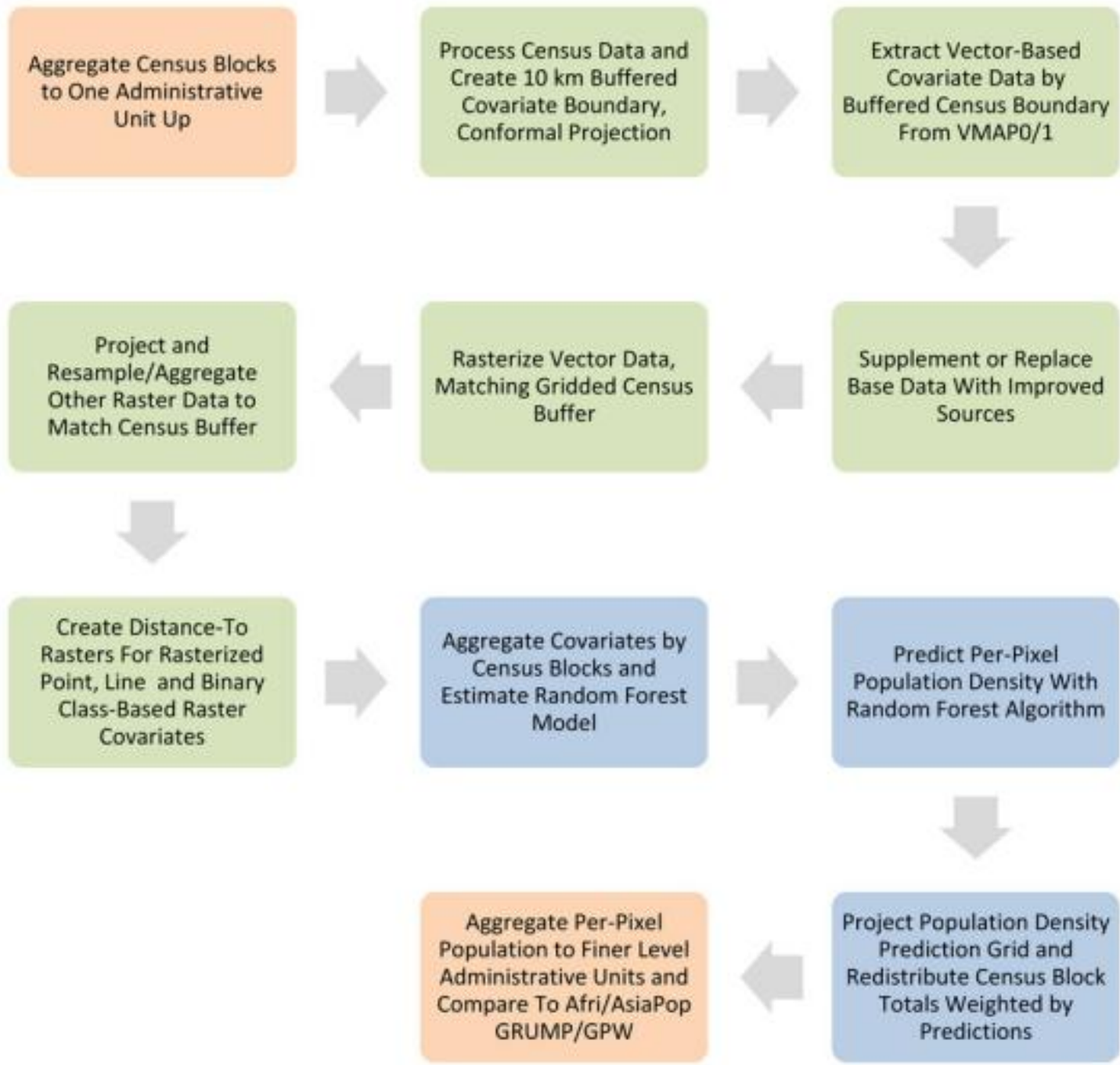
PLOS ONE | DOI:10.1371/journal.pone.0107042 February 17, 2015

Forrest R. Stevens^{1*}, Andrea E. Gaughan¹, Catherine Linard^{2,3}, Andrew J. Tatem^{4,5}

- Justificativa: população e em grade → ausência de dados
- Redistribuição dasimétrica – pop counts/dados de censo e ponderação baseado em modelagem preditiva não-paramétrica “Randon Forest”
- Dados geoespaciais e de SR de múltiplas escalas → 100m
- Estudo de Caso: Kenya (KEN), Vietnã (VNM) e Camboja (KHM)

Metodologia

- Dados de censo e limites administrativos de níveis mais detalhados de cada país
 - Log da densidade de população como variável resposta
 - Distr Pop é sempre relacionado com tipos de cobertura da terra – Mapeamentos de LandCover local, do GeoCover(30m) e GlobCover (300m) → classes compatibilizadas e todos para pixels de ~100m
 - SRTM – DEM;
 - MOD17A3 – produção primária líquida (NPP)
 - Imagens de luzes noturnas – VIIRS
 - Clima: pp anual (BIO12) e temperatura média anual (BIO1)
 - Vetores>> redes de estradas, corpos d'água, locais de população ou assentamentos, áreas de proteção e facilidades (hospitais, escolas, clínicas de saúde)
- Pesos → Random Forest Trees → superfície de densidade



Laranja – este trabalho

Verde – pre-processamento

Azul – estimativa do modelo Random Forest, predição por-pixel e redistribuição dasimétrica dos counts do censo

Mapas para usuários finais → estimativas para outros anos 2010, 2015 e 2020

Projeções baseados na proporção de pop urbana e rural

P_{-} → ano de projeção

P_d → pop no ano do dado de entrada

t → n de anos entre o dado de entrada e o ano projetado

r → é a taxa de crescimento da pop urbana ou rural

$$P_{\{2010\}} = P_{\{d\}} e^{\{rt\}}$$

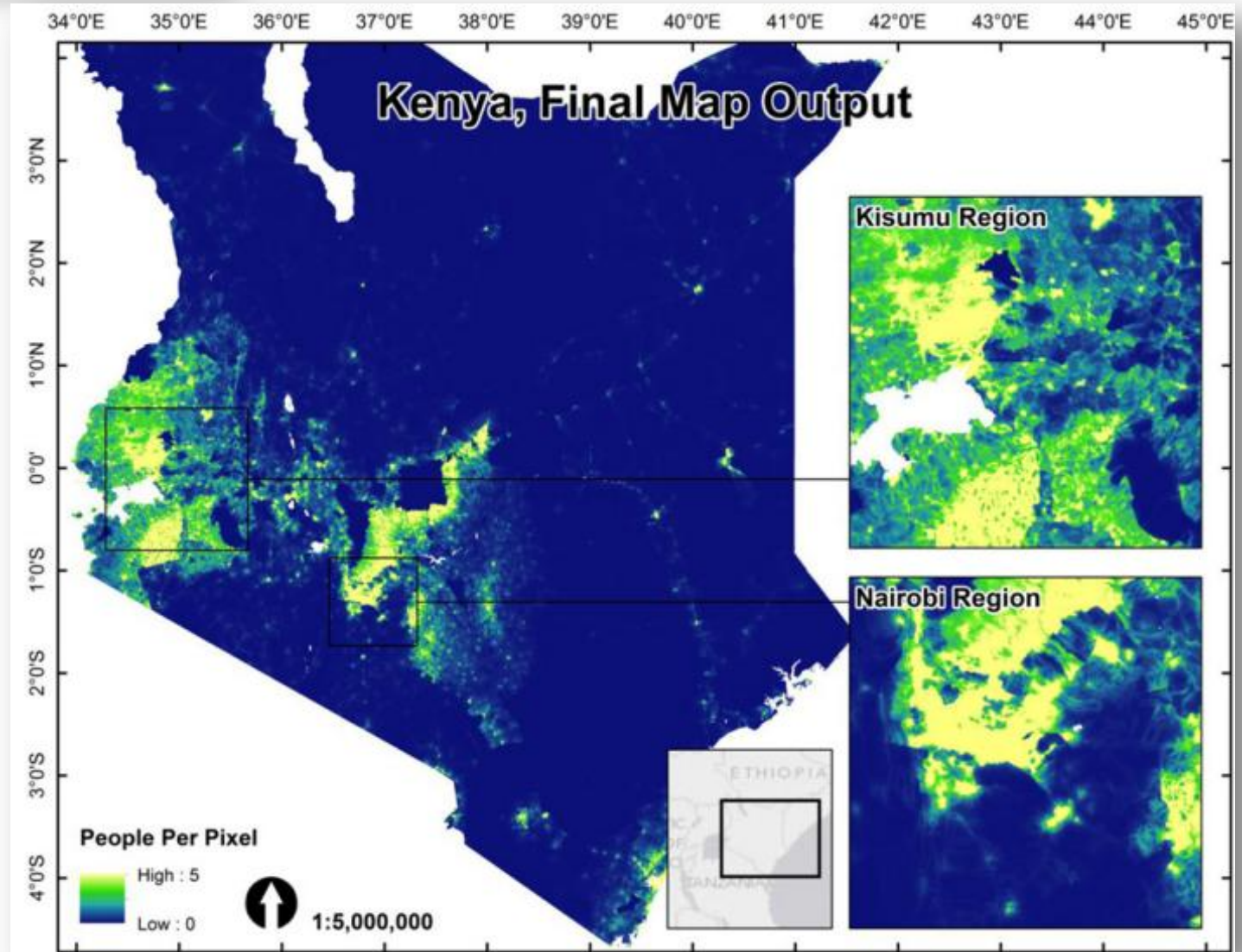
Comparações dos resultados do Random Forest (RF) com outras bases de pop em grades

- Global Rural Urban Mapping Project (GRUMP)
- Gridded Population of the World (GPW)
- AfriPop - freely-available gridded population data for Africa

Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data

Forrest R. Stevens^{1*}, Andrea E. Gaughan¹, Catherine Linard^{2,3}, Andrew J. Tatem^{4,5}

Kenya



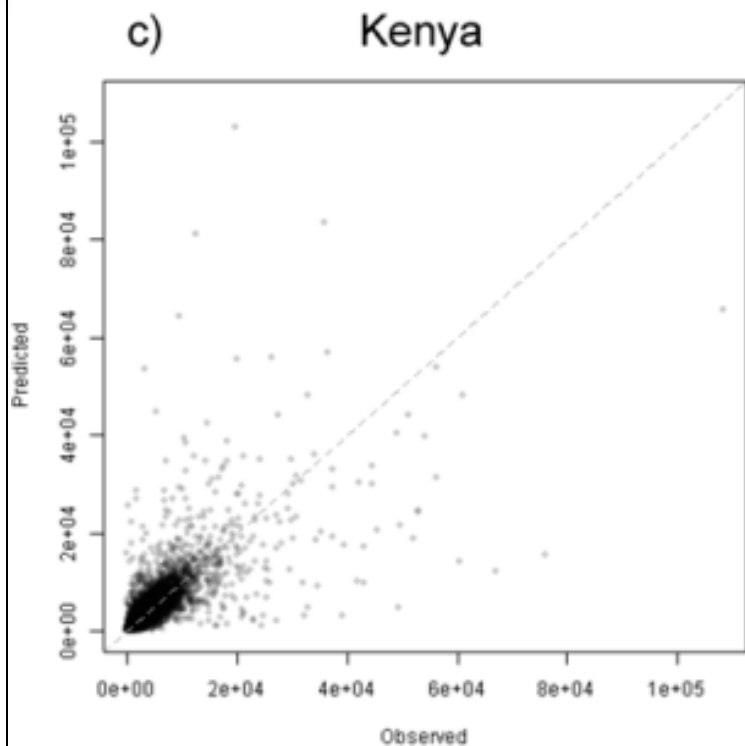


Fig 4. Observed census counts plotted from the finer census administrative units versus the summed grid cell values from the population map estimated using coarser administrative units for a) Cambodia, b) Vietnam, and c) Kenya.

Random Forest Variable Importance for Kenya Population Map Production

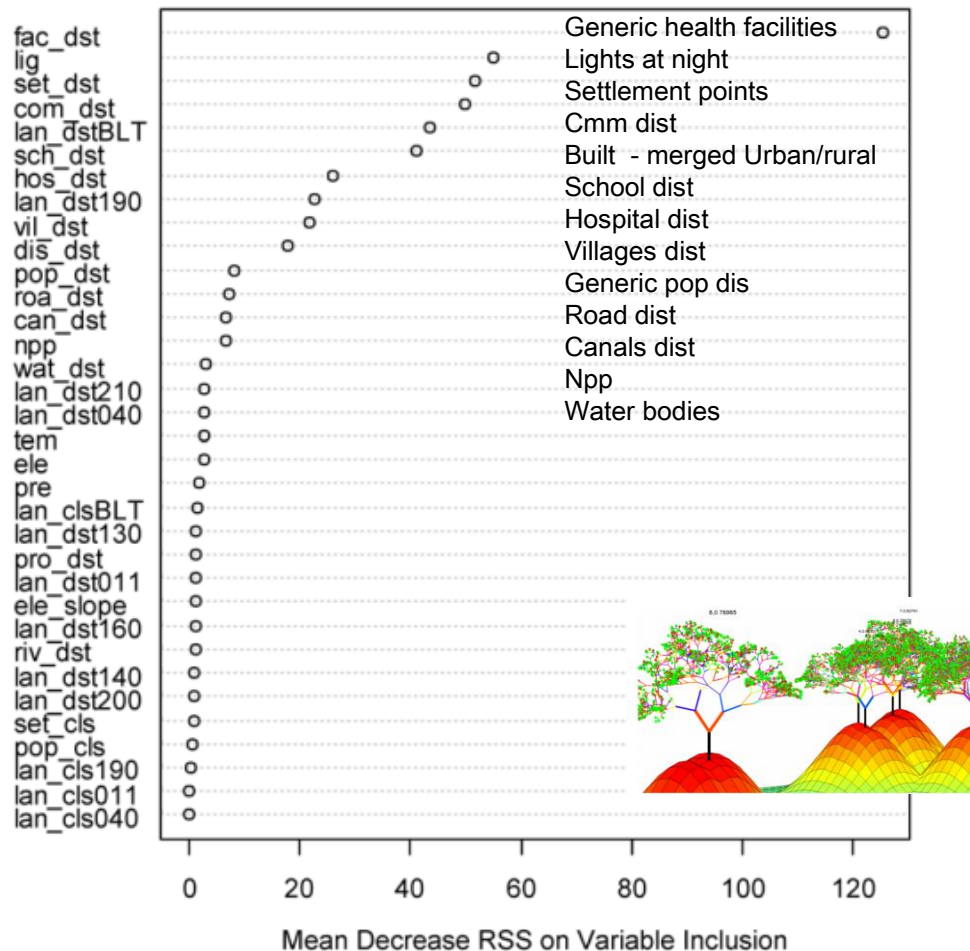


Fig 3. Variable importance for Random Forest regression, presented as the mean decrease in residual sum of squares when the variable is included in a tree split. The model including variables here was used to produce the density weighting layer for the dasymmetrically distributed population map in Fig 2. Variable names are defined and described in Table 1.

Comparaç o

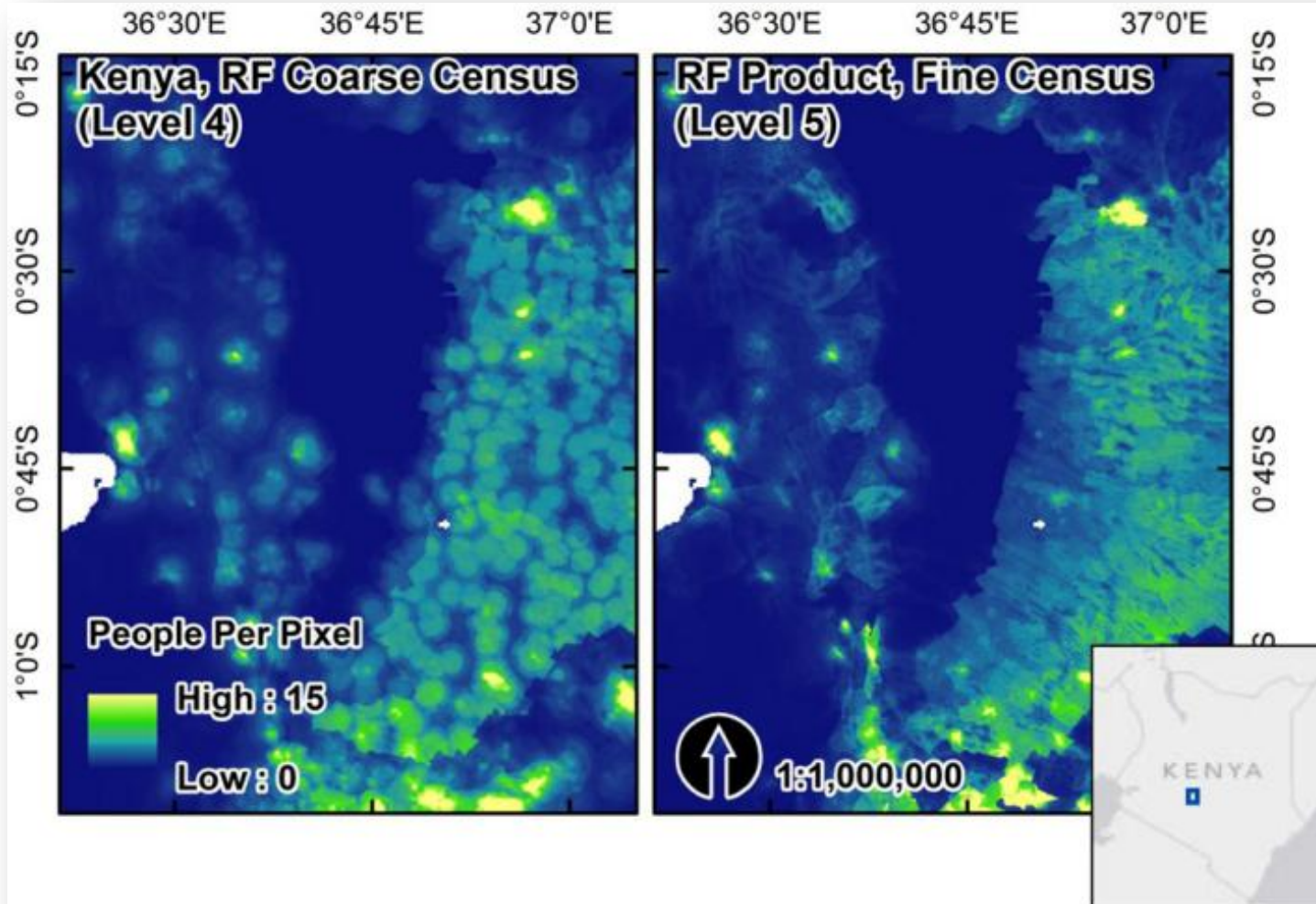


Fig 5. A visual comparison of Kenyan population maps for census data in 1999 produced at coarser administrative unit (Level 4) and finer-scale administrative unit (Level 5). The difference illustrates the finer gradations of RF model predictions for the density weighting layer when there are larger ranges of observed population densities present in training data (N = 505 for Level 4, N = 6622 for Level 5).

Comparaçã

Table 2. Accuracy assessment results for the RF, Afri/AsiaPop, GRUMP and GPW modeling methods for Cambodia, Vietnam and Kenya.

Country	Method	RMSE	% RMSE	MAE
Cambodia	RF	3209.40	38.84	2026.73
	AsiaPop	3834.51	46.40	2494.32
	GRUMP	6767.39	81.89	3889.39
	GPW	6794.88	82.22	4021.41
Vietnam	RF	4367.00	61.96	2778.80
	AsiaPop	4943.31	70.13	3007.04
	GRUMP	6523.77	92.56	3771.64
	GPW	7081.76	100.47	3844.47
Kenya	RF	3956.74	91.36	1685.82
	AsiaPop	5208.79	120.28	2184.64
	GRUMP	6294.61	145.35	2383.64
	GPW	6327.80	146.11	2304.70

Two different error assessment methods are presented: root mean square error (RMSE), also expressed as a percentage of the mean population size of the administrative level (% RMSE); and the mean absolute error (MAE).

Global Rural Urban Mapping Project (GRUMP)

Gridded Population of the World (GPW)

AsiaPop - freely-available gridded population data for Asia

AfriPop - freely-available gridded population data for Africa

Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data

Forrest R. Stevens^{1*}, Andrea E. Gaughan¹, Catherine Linard^{2,3}, Andrew J. Tatem^{4,5}

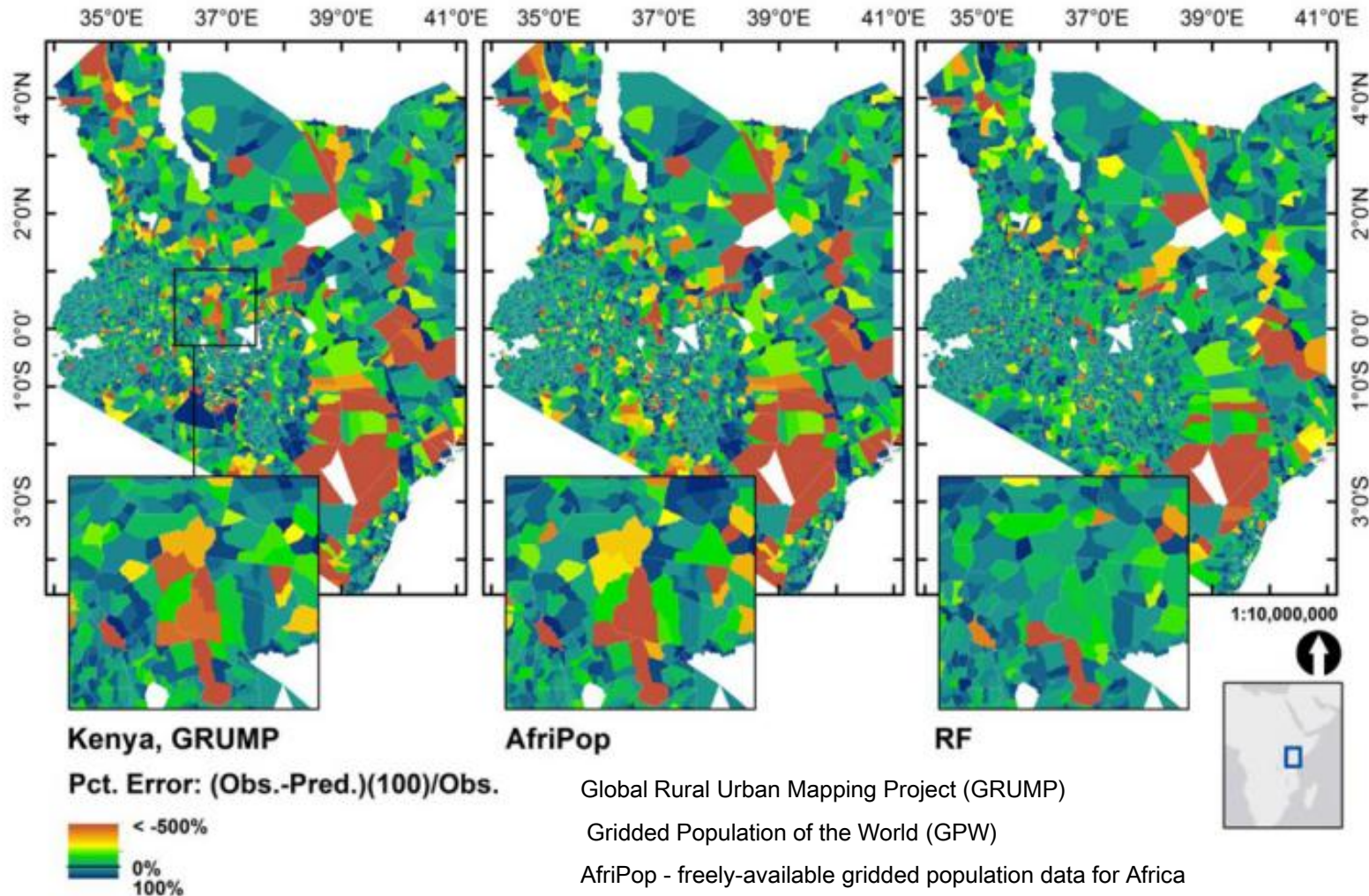


Fig 6. For Kenya we compared the summed predictions of population maps estimated using coarse census data (Level 4, "Division" level) to population counts at Level 5 (finer scale) units. We present prediction errors (observed minus predicted) as percentage values of the observed Level 5 census counts. The RF approach results in far fewer census units with extreme over predictions (negative percent residuals in yellows and oranges) and under predictions (positive percent residuals in dark blues).

Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data

Forrest R. Stevens^{1*}, Andrea E. Gaughan¹, Catherine Linard^{2,3}, Andrew J. Tatem^{4,5}

Kenya

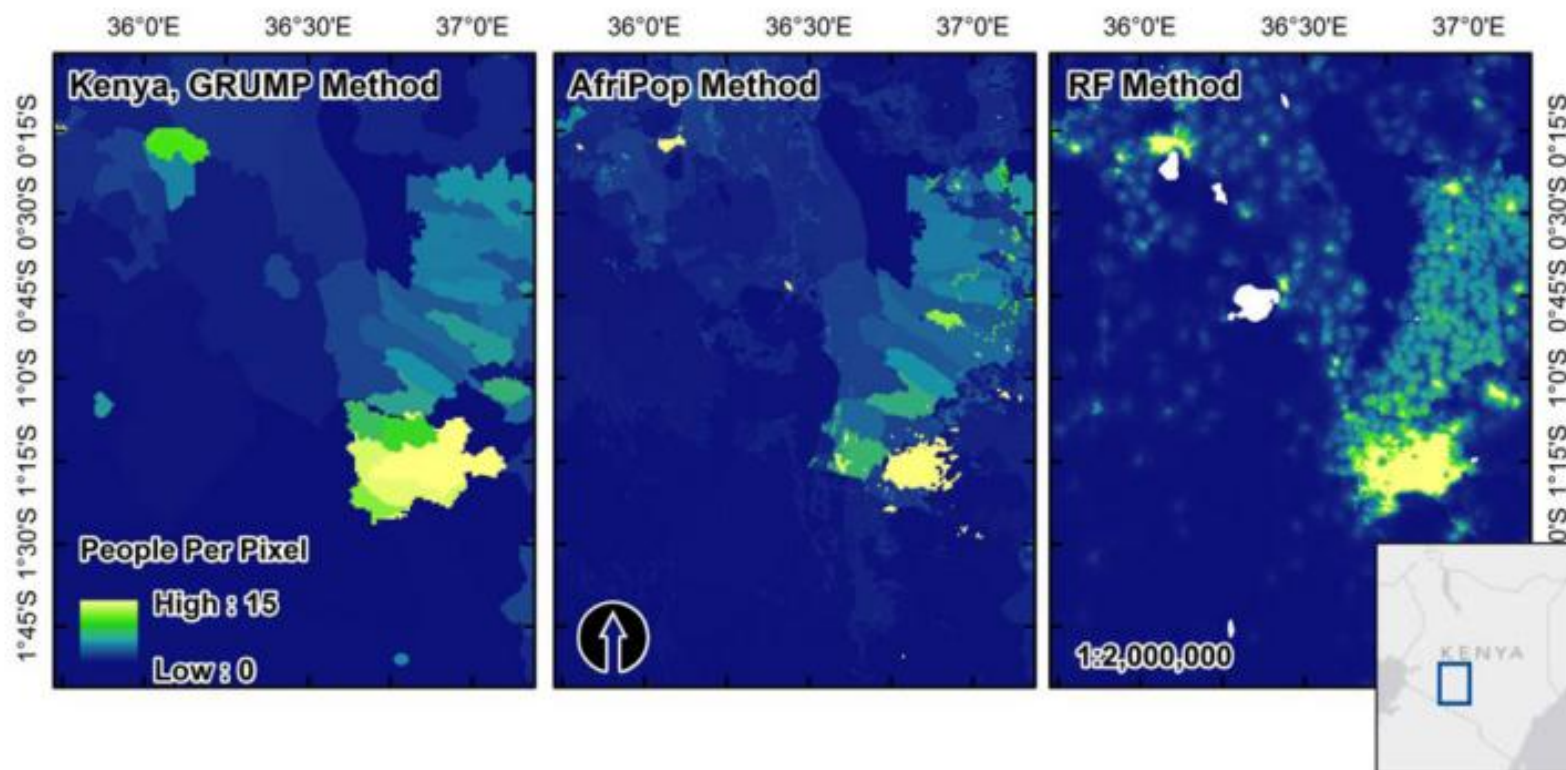


Fig 7. Visual comparisons of GRUMP, AfriPop and the RF-based population map from this study. Though this region northwest of Nairobi, Kenya is not a highly populated region this figure shows the results of the more detailed RF weighting layer versus the use of just urban areas (GRUMP) and land cover plus urban areas (AfriPop). The distinct edges in estimated people per pixel between census units are almost eliminated by the RF approach and it achieves greater consistency in predicted population density after census count redistribution.

Population datasets for 2010, 2015, and 2020 are freely available for download from the WorldPop Project (<http://www.worldpop.org.uk/>)