



Generalitat de Catalunya  
**Institut d'Estadística de Catalunya**

## ¿Son los grids de población realmente útiles?

Eduard Suñé, Daniel Ibáñez, Marc Castelló

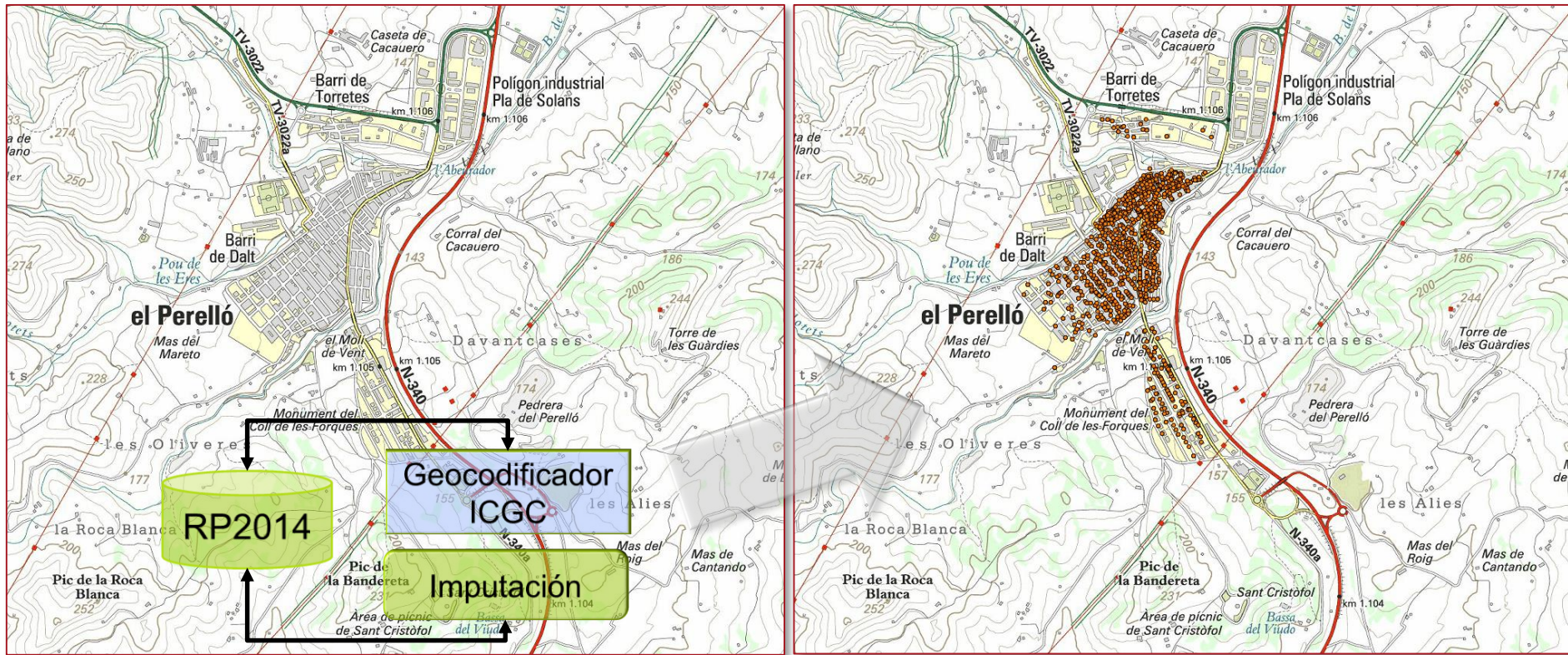
Introducción.

Quadtree. Construcción. Efecto frontera.

Estimación de errores.

Alternativa a la agregación: perturbación de coordenadas

Conclusiones.



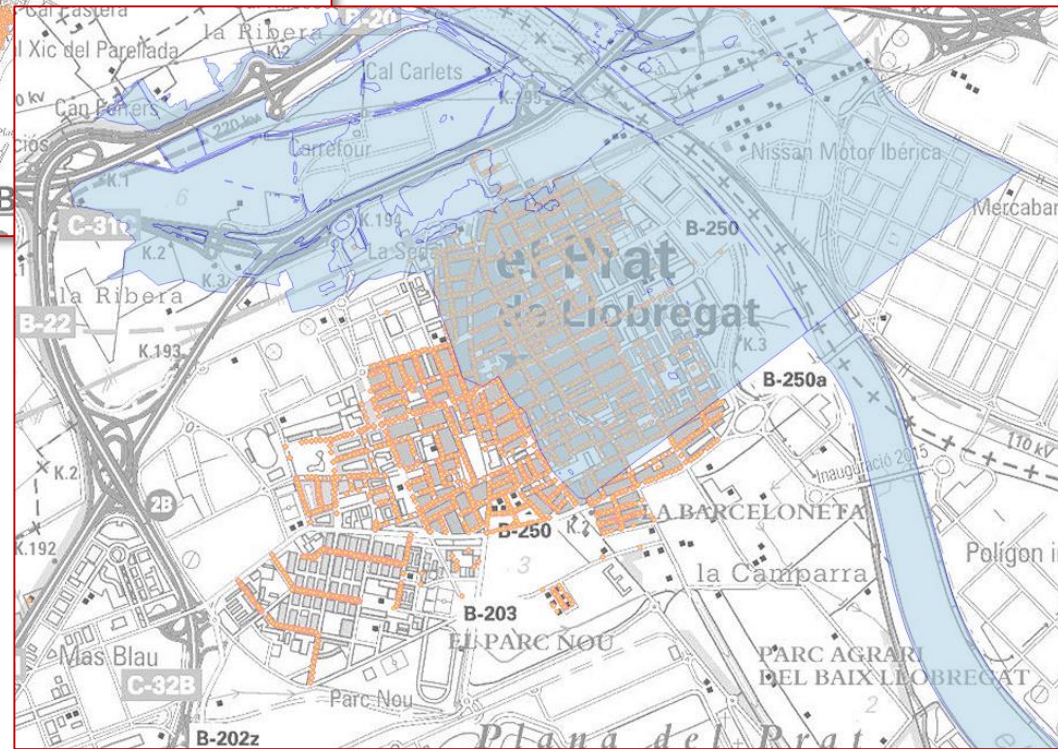
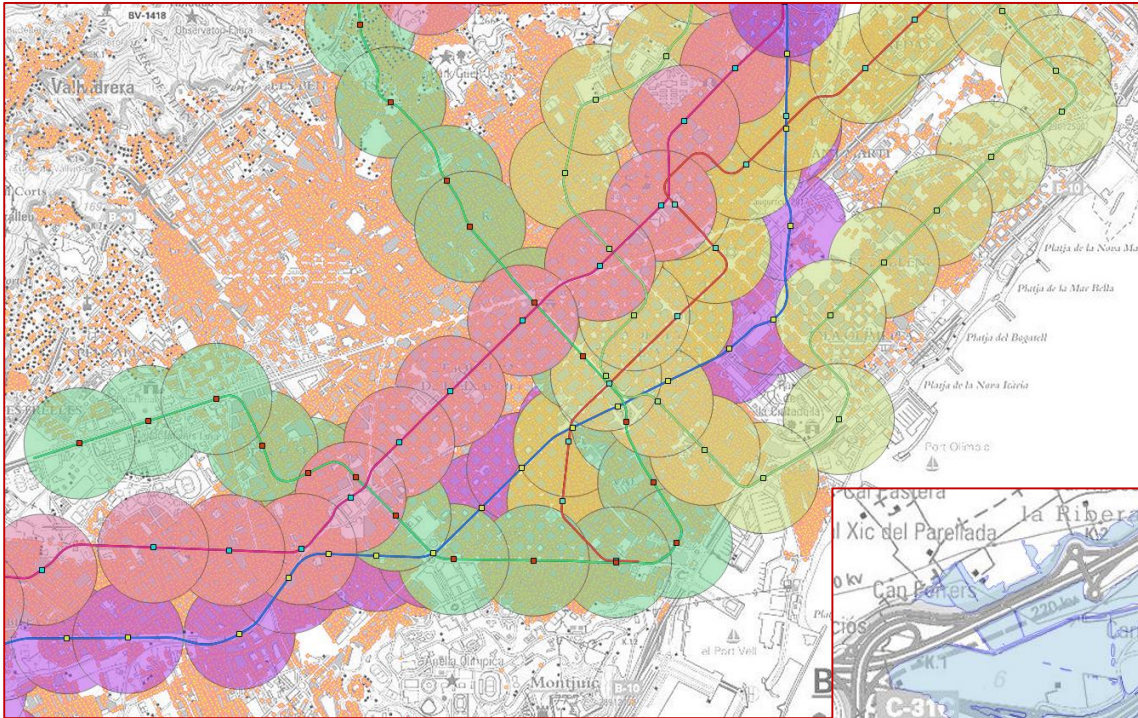
Exactitud (RP2014)	N	% imputados
Portal	2581826	
Portal interpolado	3204766	
Portal interpolado más cercano	1351764	
Rectángulo rodeando la calle	93852	
Portal asignado a finca vía tabla ine-dgc numeración coincidente	127047	1.68
Portal asignado aleat. a finca vía tabla ine-dgc numeración más cercana dentro del convex	8164	0.11
Portal asignado aleat. a finca vía tabla ine-dgc numeración más cercana fuera del convex	69161	0.91
Portal asignado aleat. a finca según tabla hogares-bien inmuebles dentro del convex	113861	1.50
Portal asignado aleat. a finca dentro del convex	16023	0.21
<b>Total</b>	<b>7566464</b>	<b>4.42</b>

## Usos:

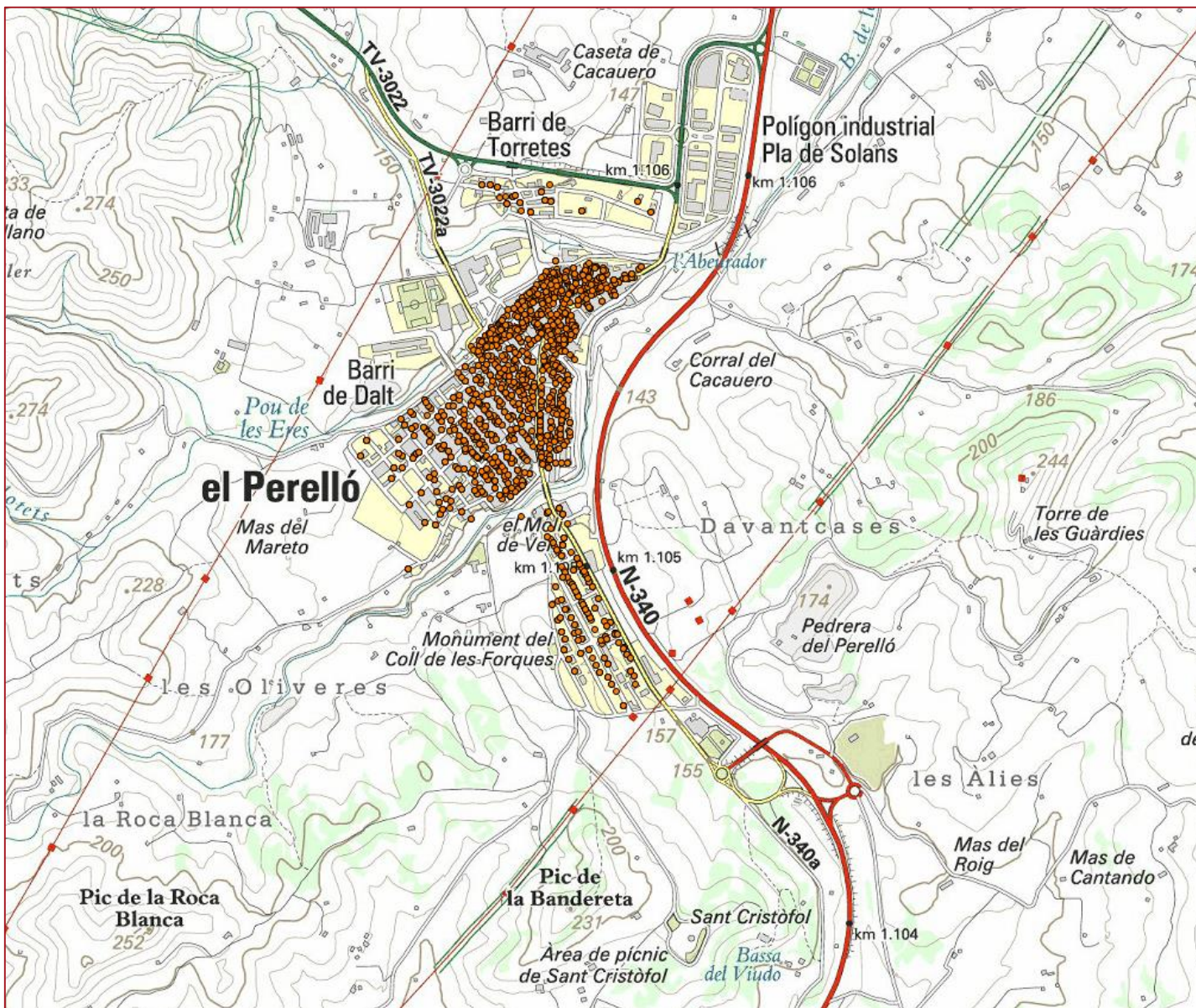
- Planificación del transporte
- Protección civil
- Localización de infraestructuras
- Modelos calidad del aire etc...

← transporte

Inundaciones



Usos:  
cálculo de población dentro de  
una superficie

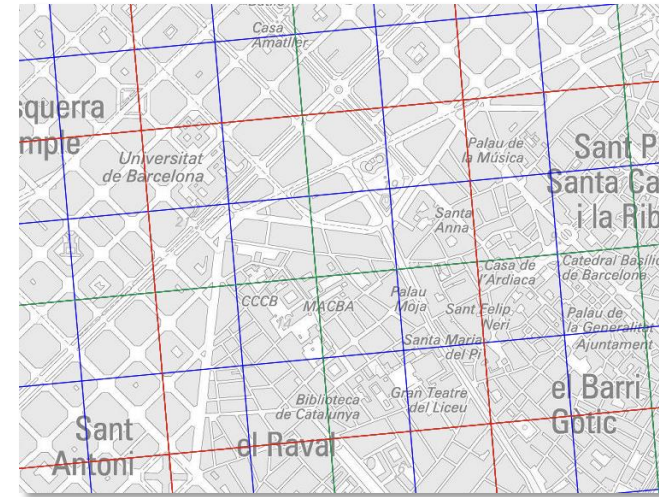
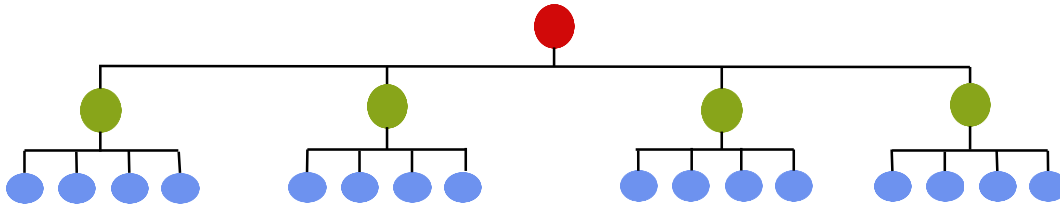


En el momento de la difusión,  
como preservar el secreto estadístico?

Perturbando posiciones  
(puntos → puntos)

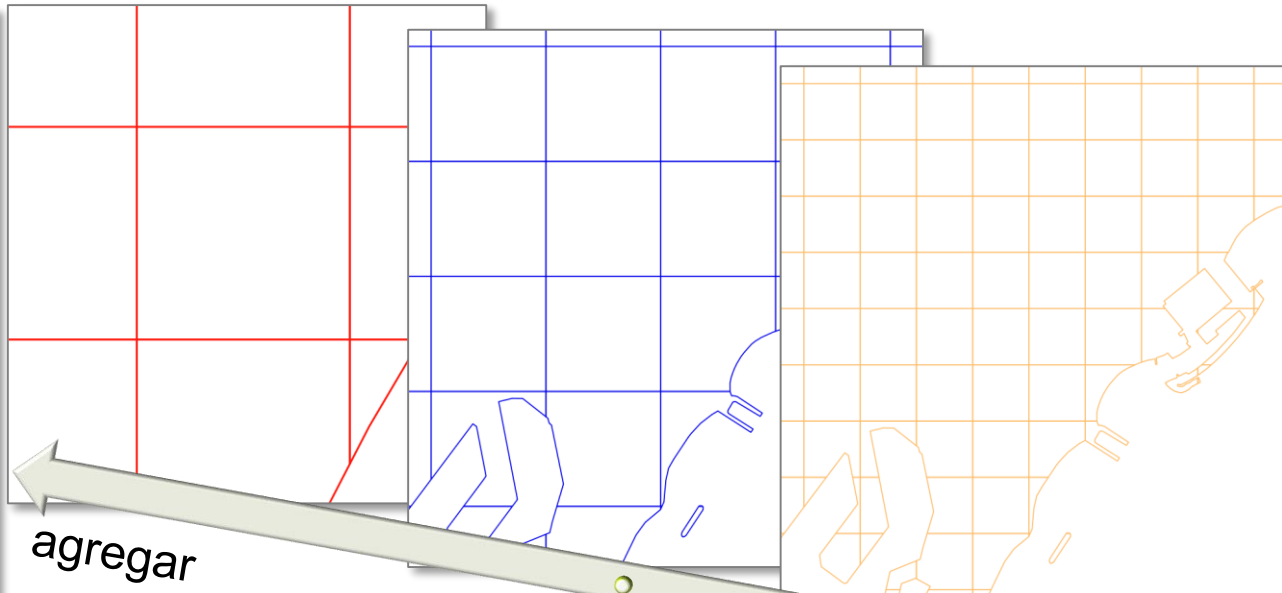
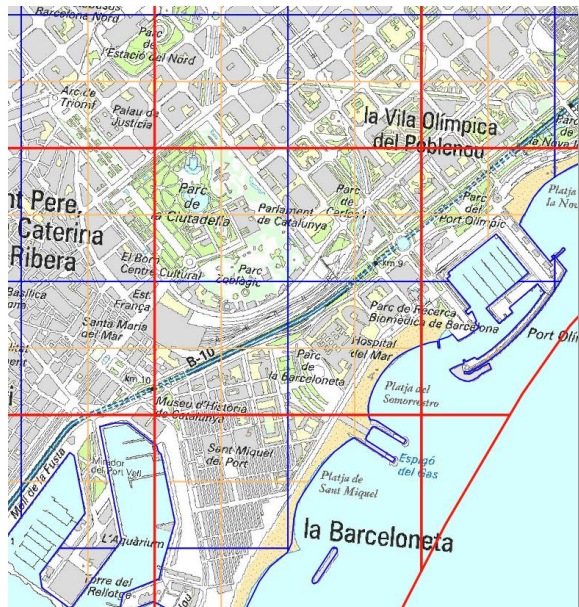
Agregando espacialmente  
(puntos → polígonos)

Estructura jerárquica en la que los padres tienen cuatro hijos.



la resolución se adapta localmente para preservar el secreto estadístico: *si hay suficiente población se divide el área...* y así recursivamente..

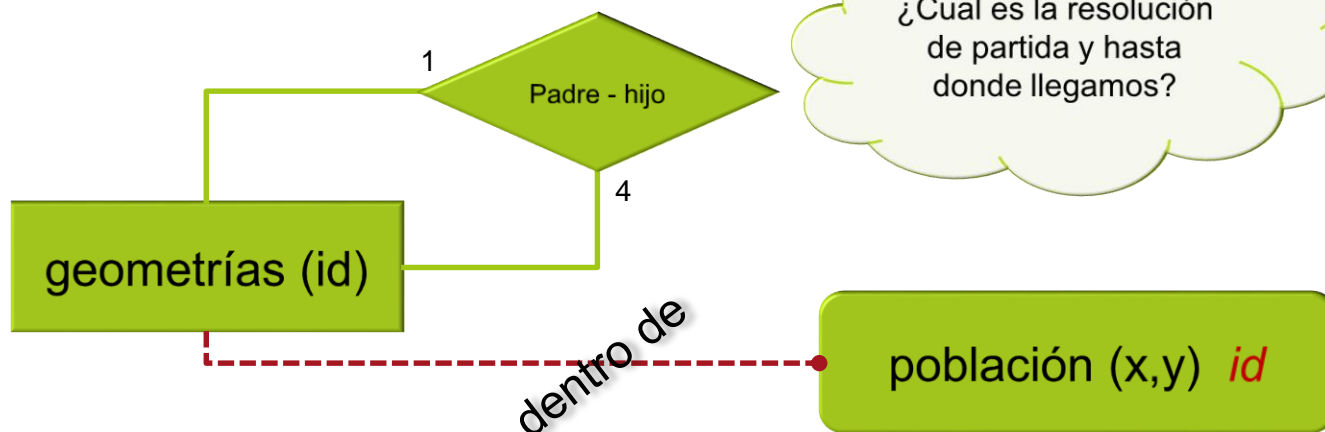


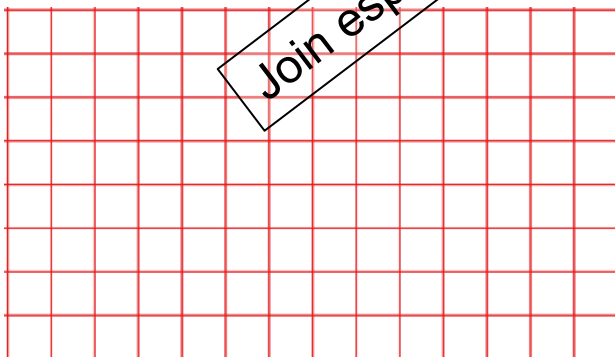
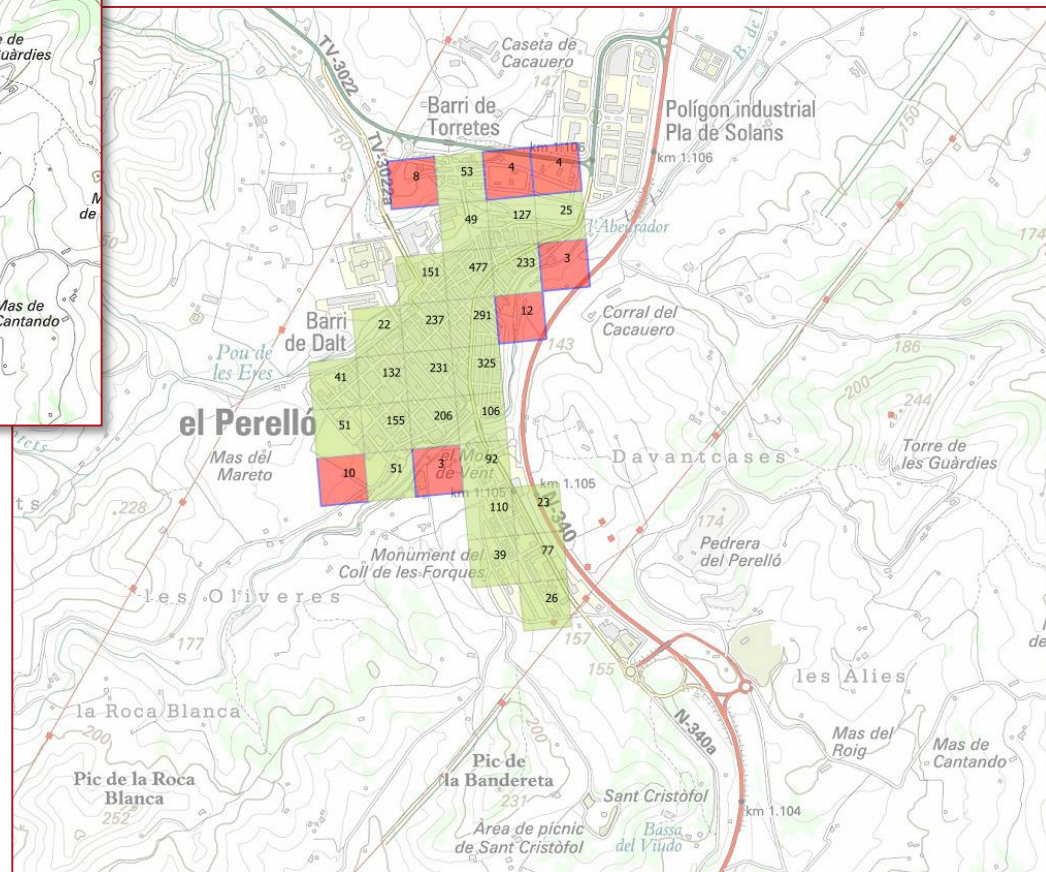
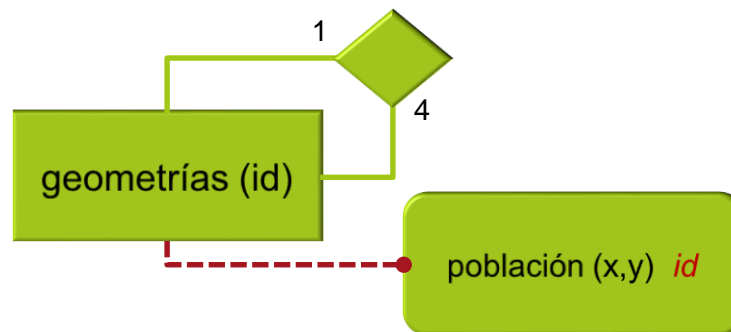
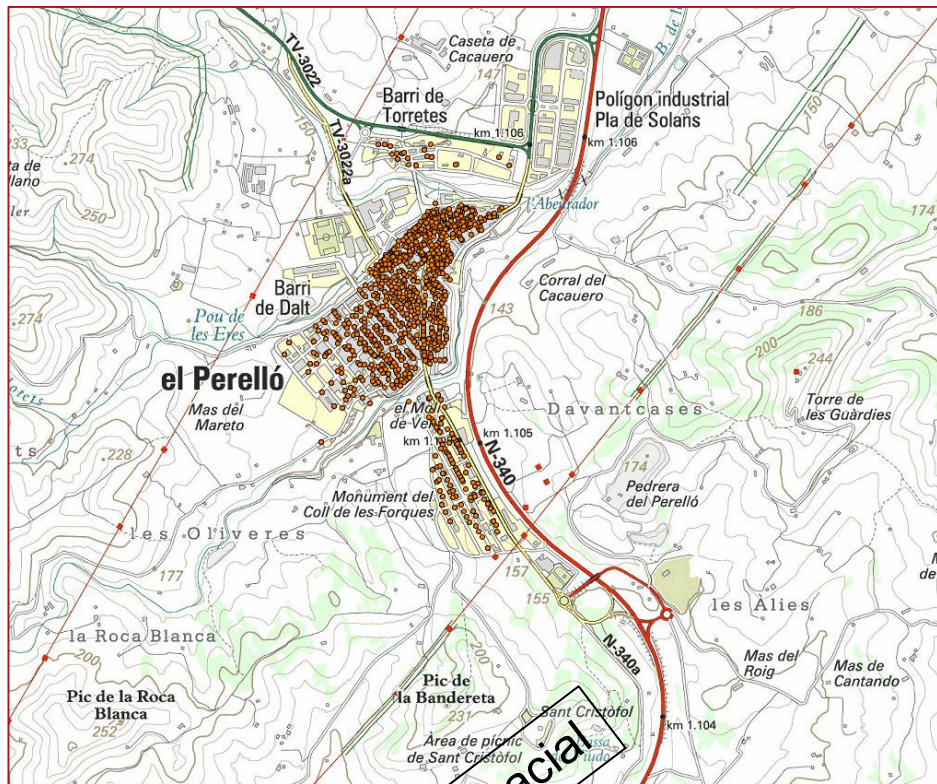


agregar

dividir

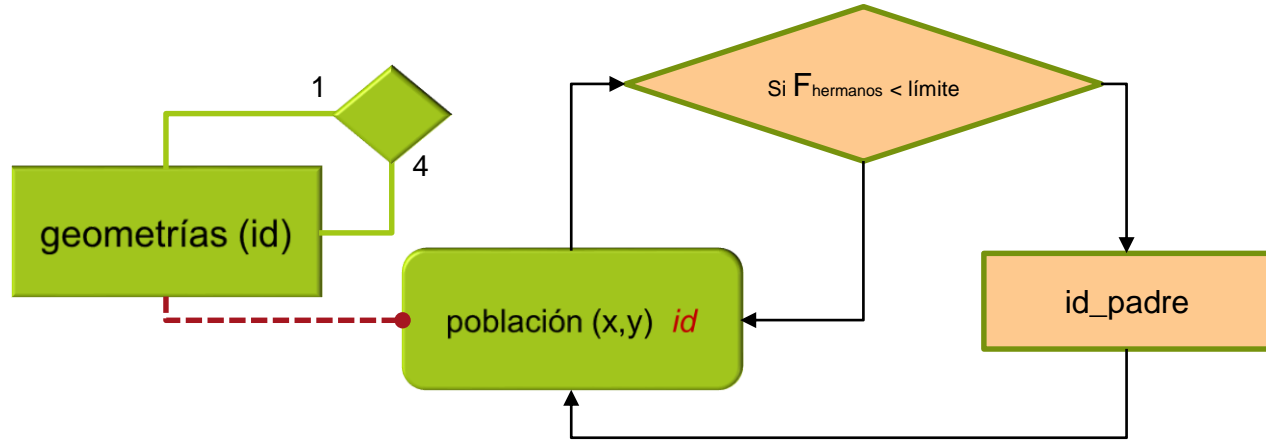
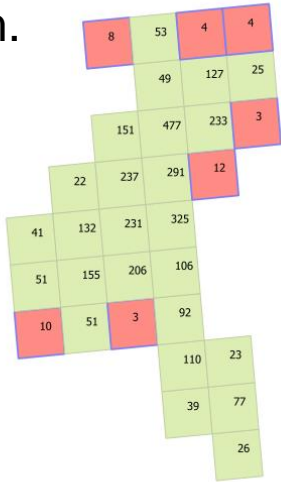
¿Cual es la resolución de partida y hasta donde llegamos?



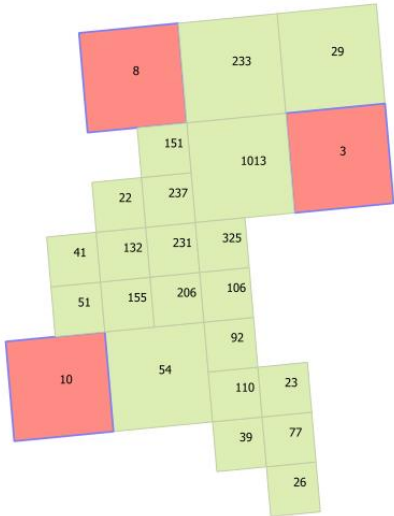


Join espacial

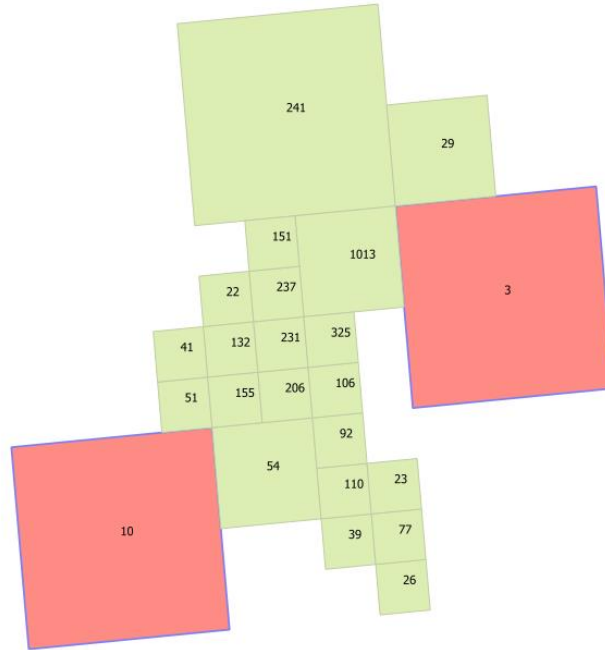
125 m.



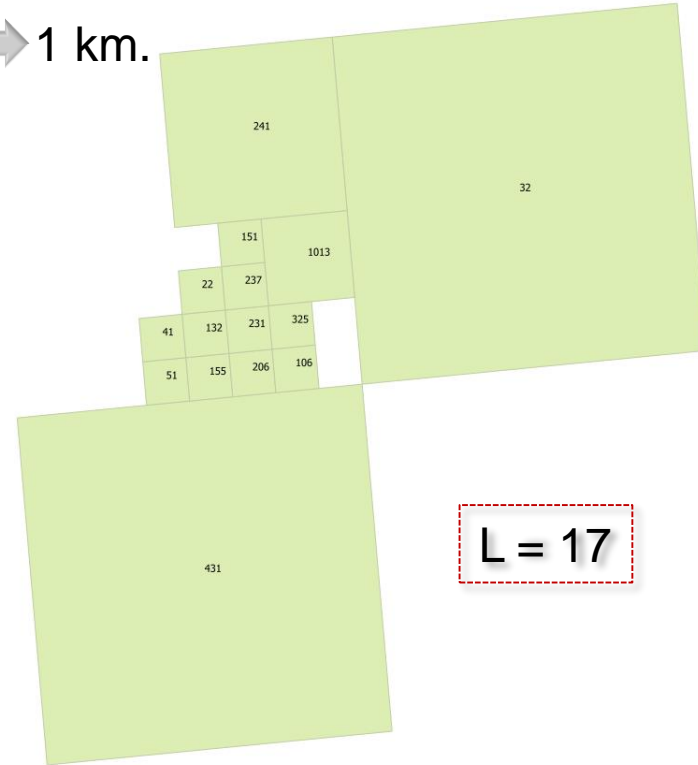
250 m.



500 m.



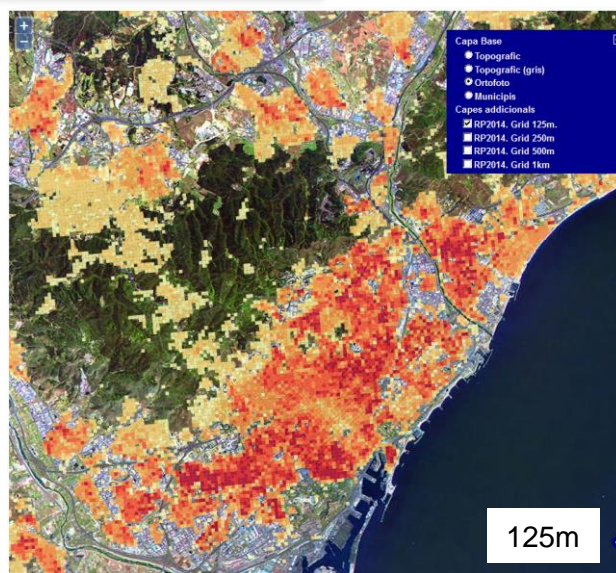
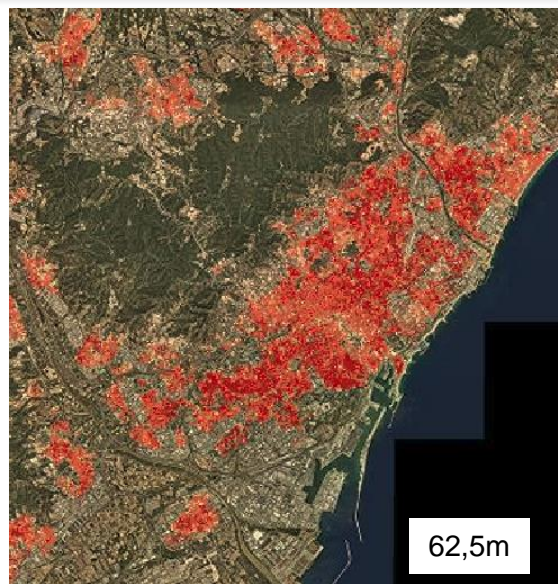
1 km.



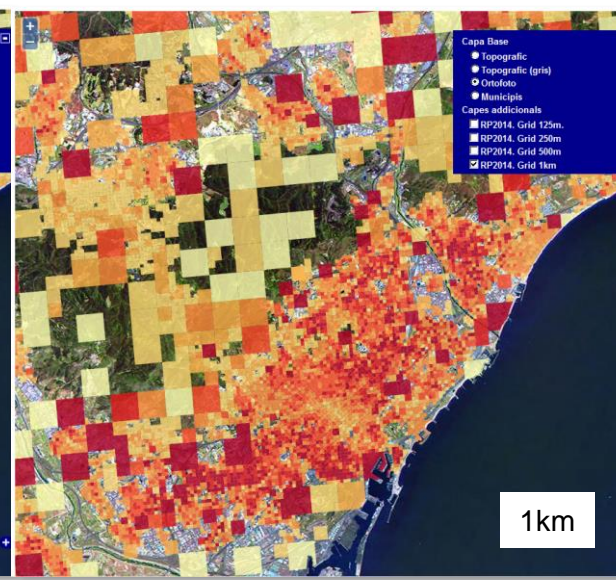
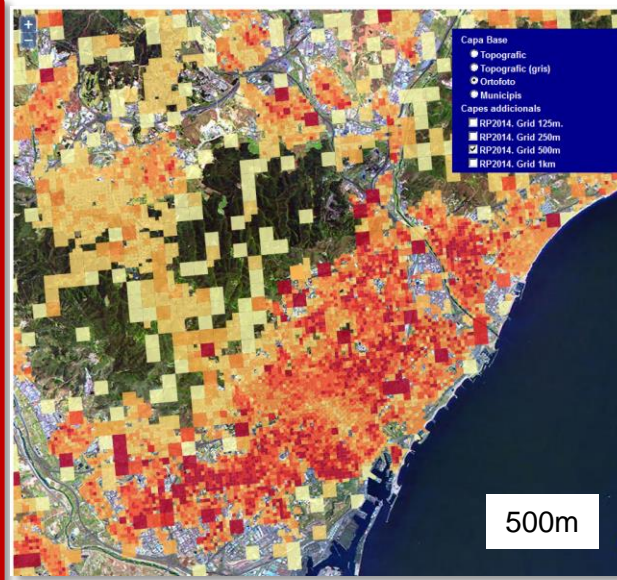
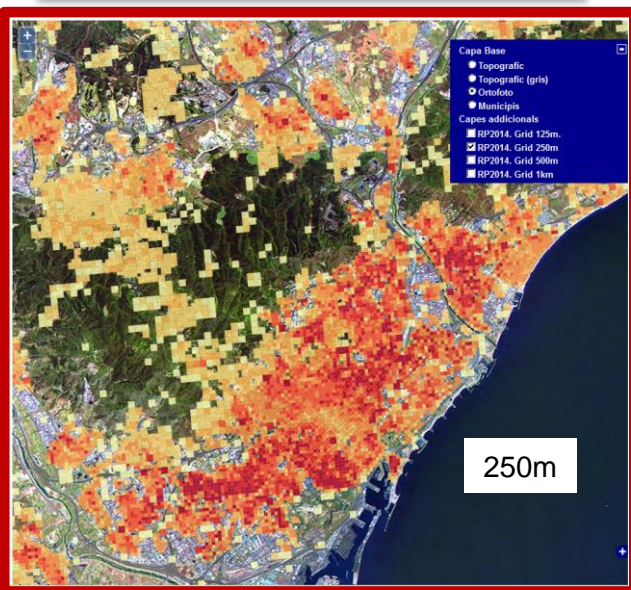
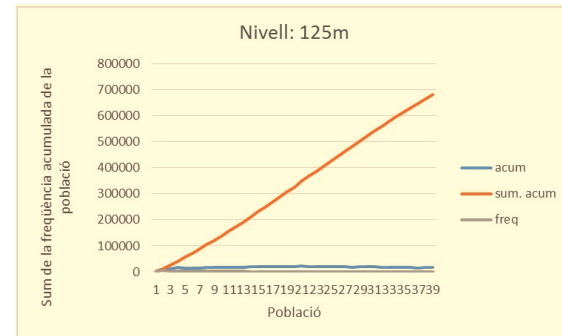
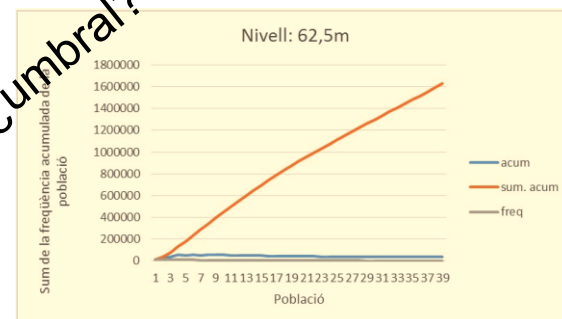
**L = 17**



QT {Res. Max. ; Res. Min.; Datos; umbral}

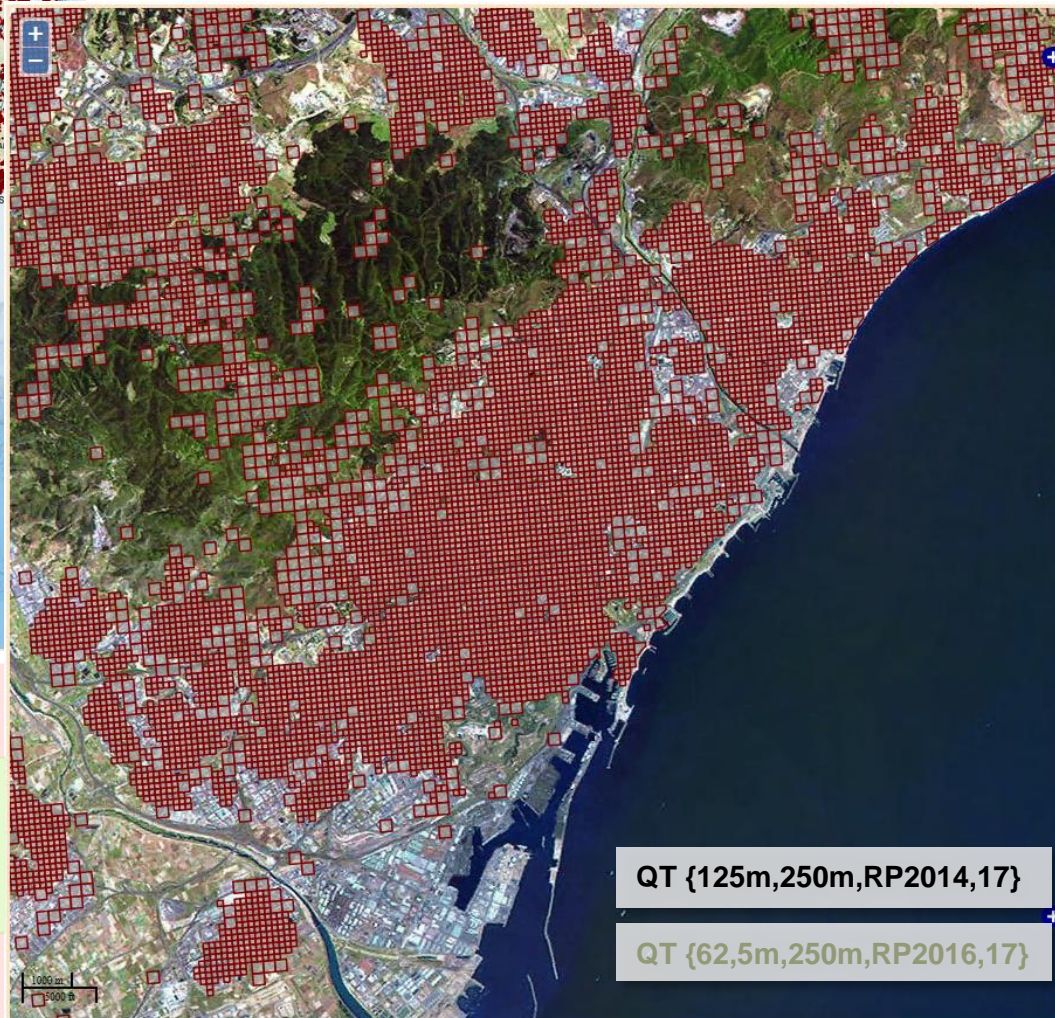


¿umbral?



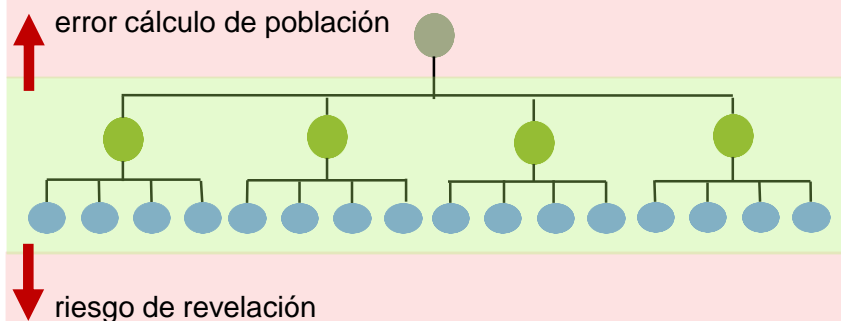
Quadtree queda definido por

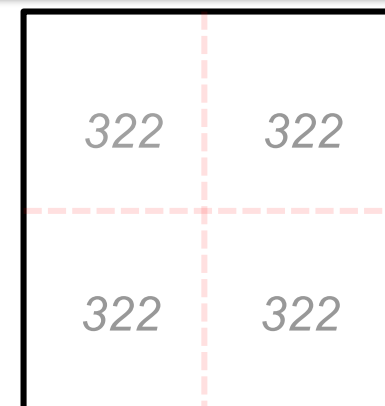
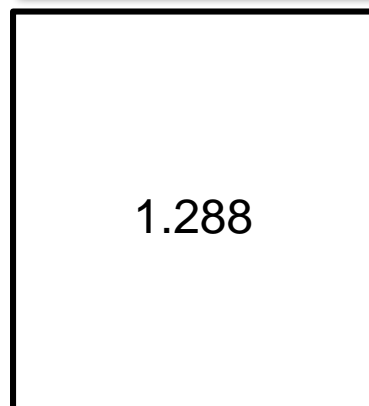
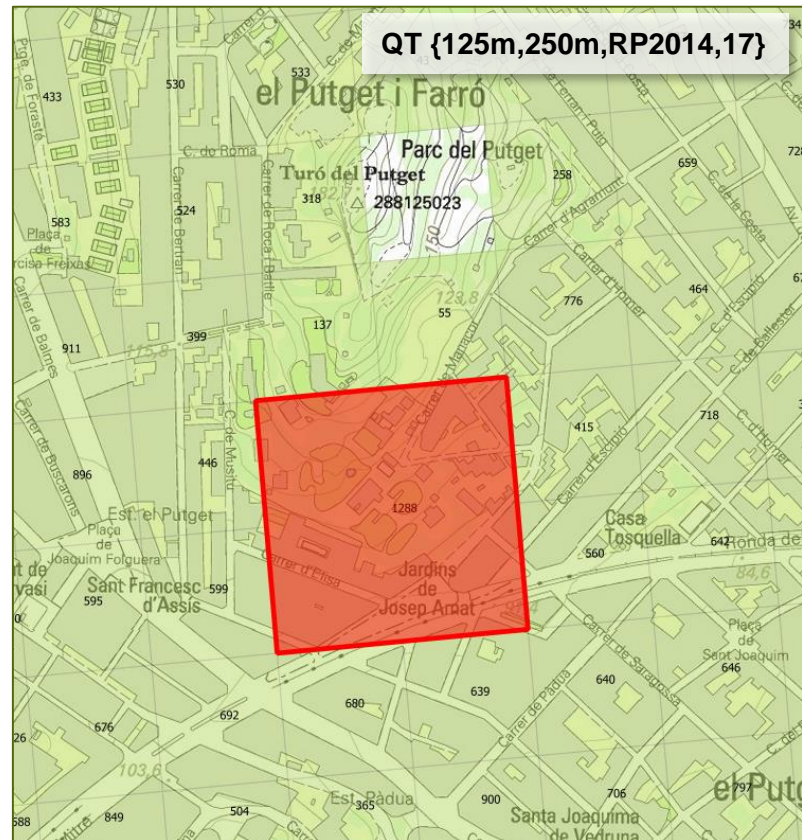
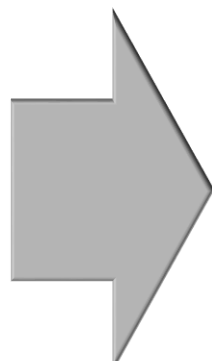
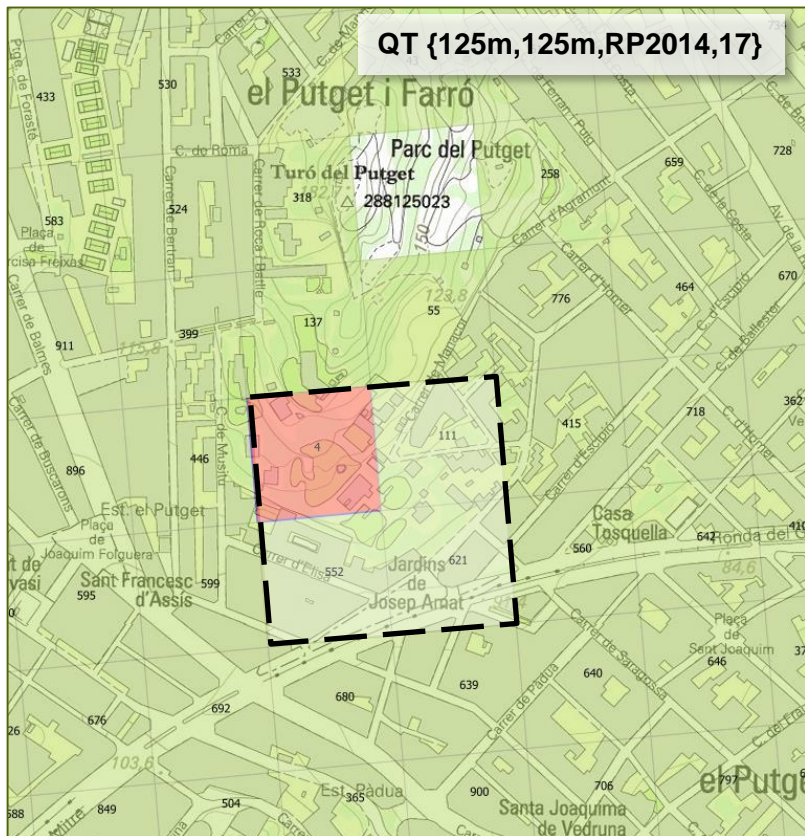
- Límite
- Resolución máxima
- Resolución mínima
- Datos a difundir



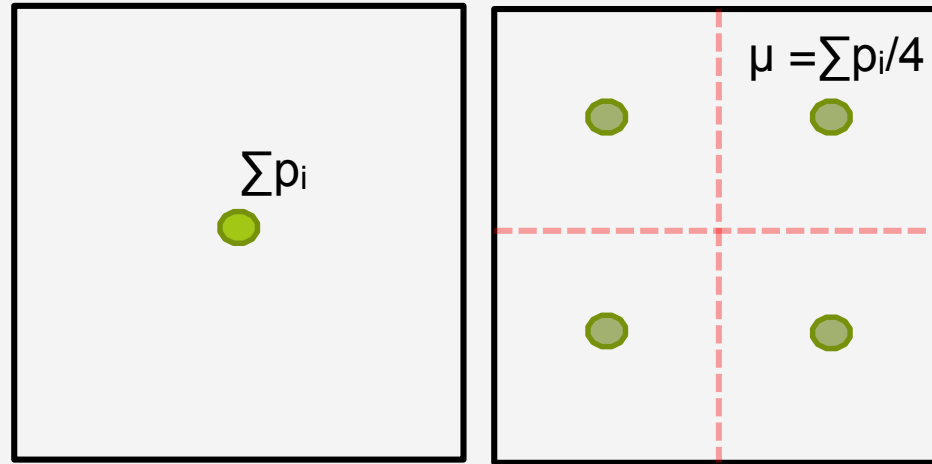
QT {125m,250m,RP2014,17}

QT {62,5m,250m,RP2016,17}





¿Qué ve el usuario con el quadtree?



Error absoluto



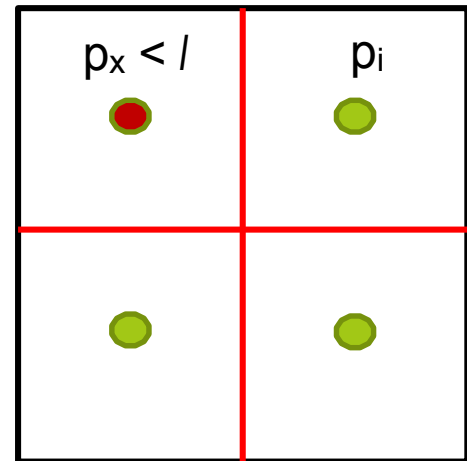
$$\epsilon = \sum |p_i - \mu|$$

Si

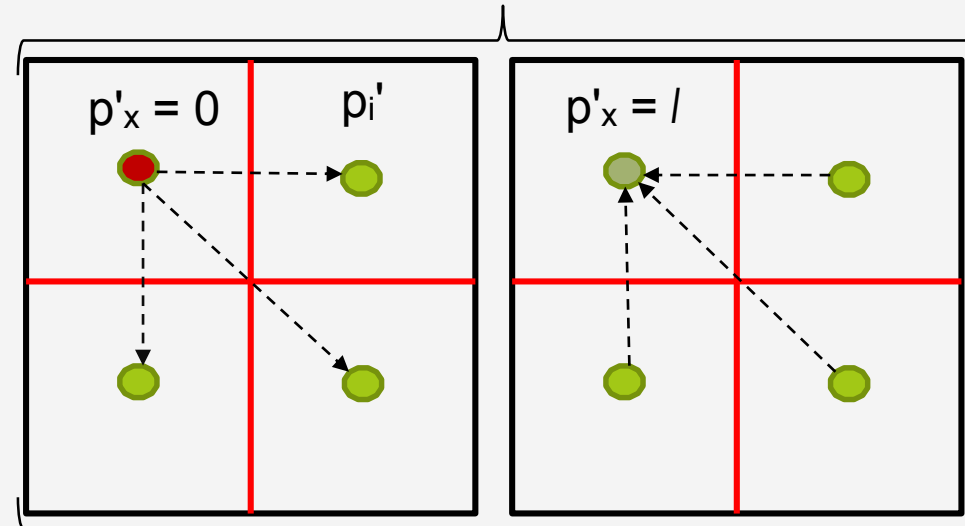
$$\epsilon' < \epsilon$$

es mejor desplazar

equivale a



¿Qué vería el usuario si desplazáramos algunos puntos?



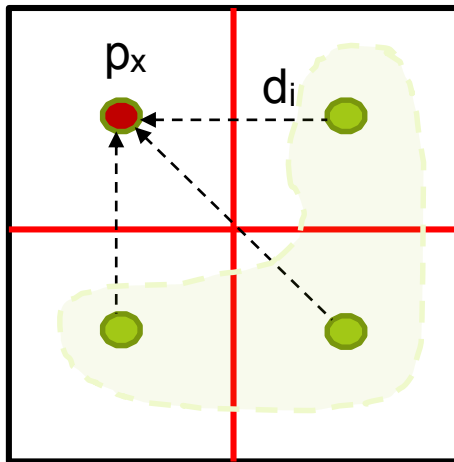
$$\epsilon' = \sum |p_i - p'_i|$$



Error absoluto

✓ Si  $\epsilon > 2*N$  (N : número de elementos trasladados)

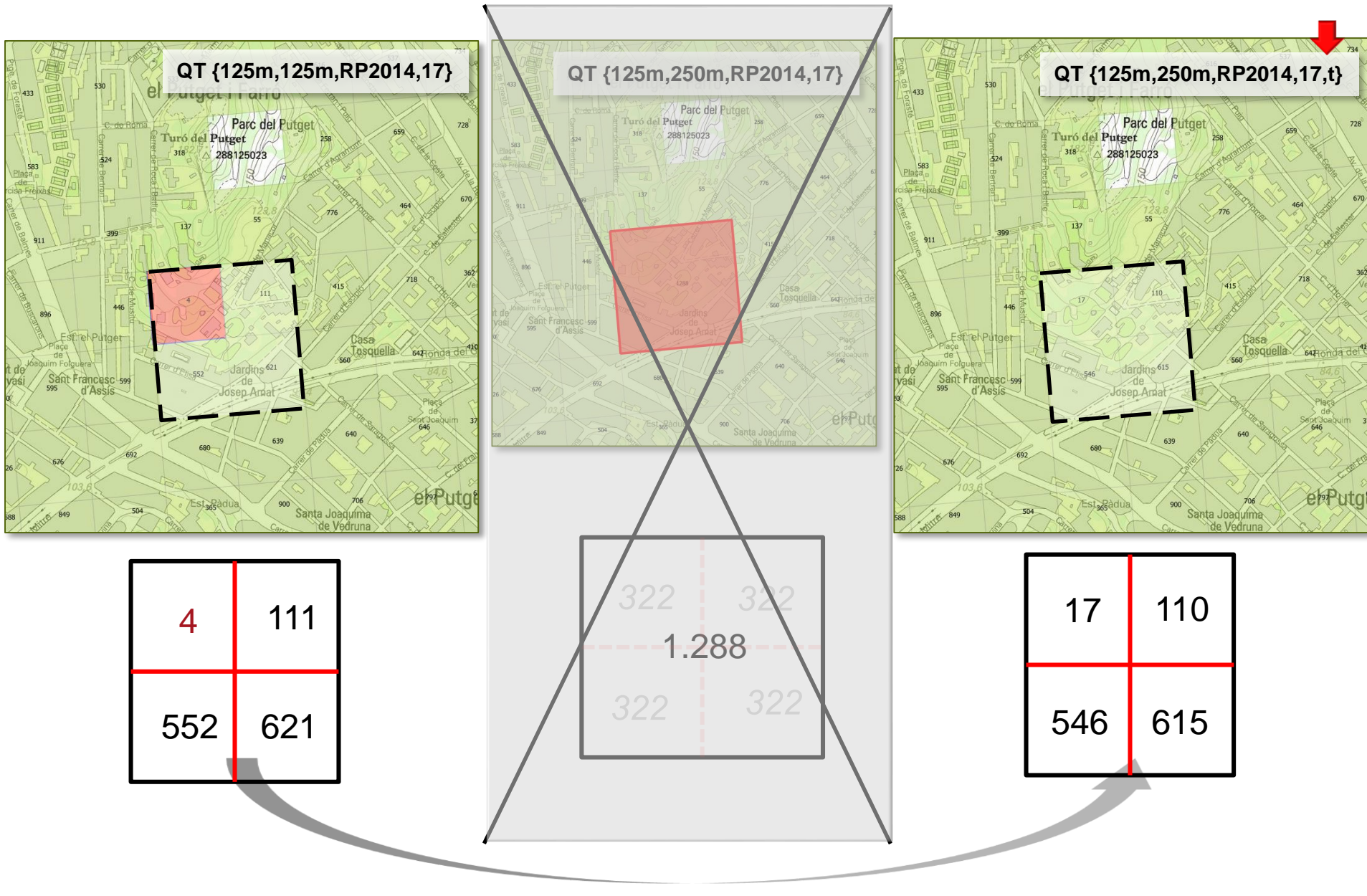
- Aplicar siempre añadiendo hasta llegar al límite  $l$
- Repartir los movimientos según las frecuencias relativas del subconjunto de donantes

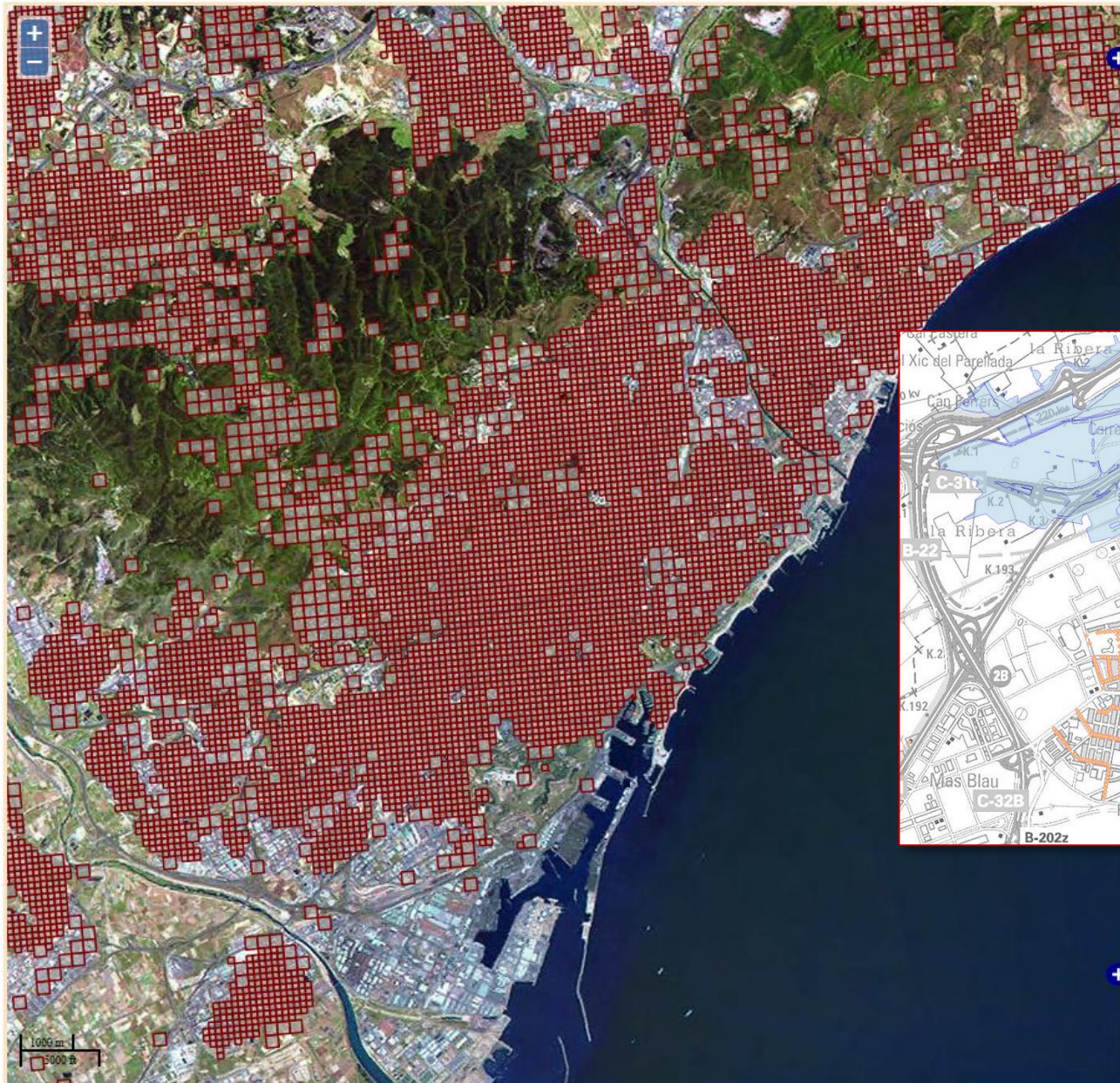


$$\sum d_i = \sum l - p_x$$

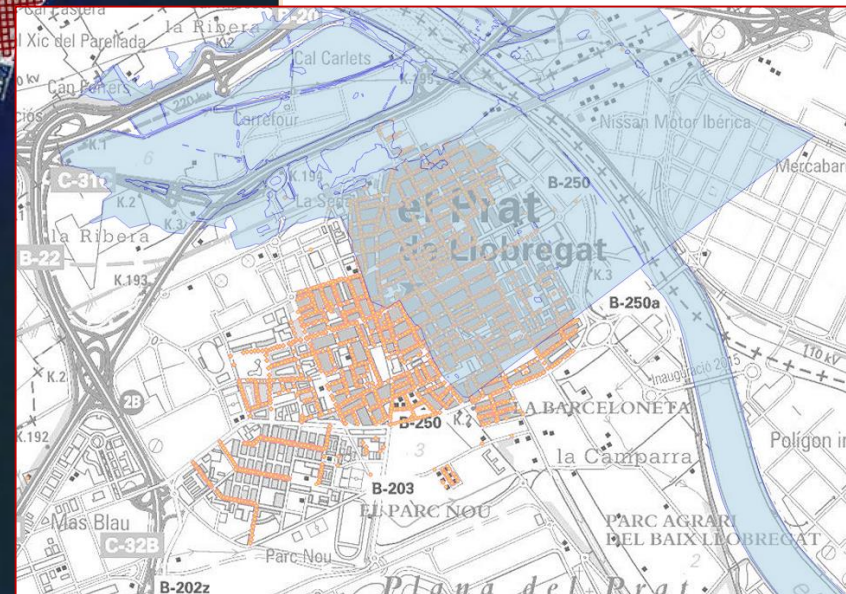
$$d_i = |l - p_x| * (p_i / \sum p_i) \quad \text{para todo } i \text{ tal que } p_i - d_i \geq l$$

- Escoger aleatoriamente los elementos a mover del subconjunto de donantes





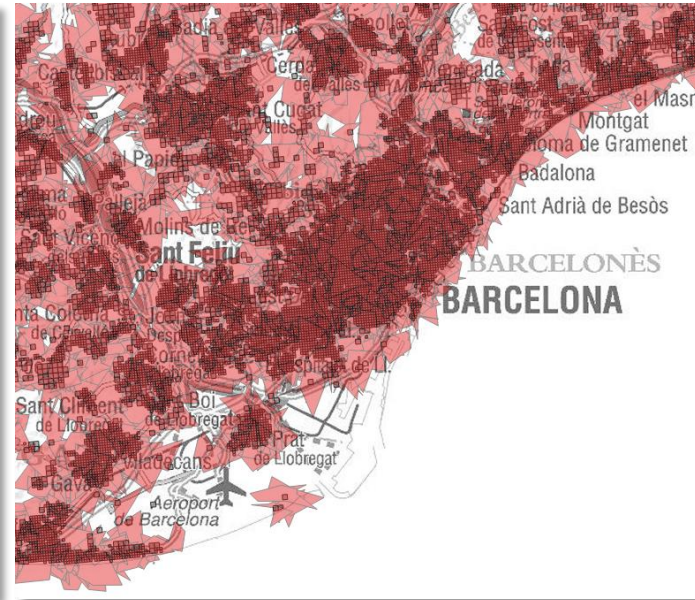
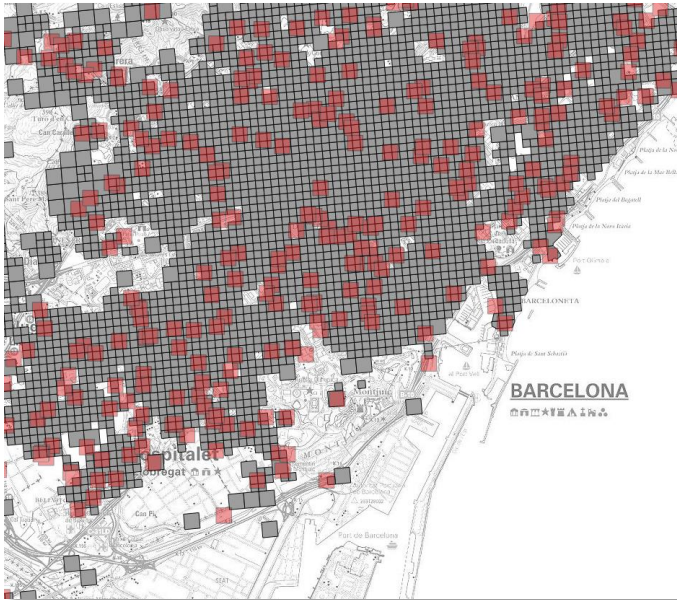
¿Son los grids de población realmente útiles?



Error en el calculo en superficies de interes

## Experimentos:

1. Generar n cuadrados aleatoriamente (controlando el área).
2. Generar n polígonos aleatoriamente.

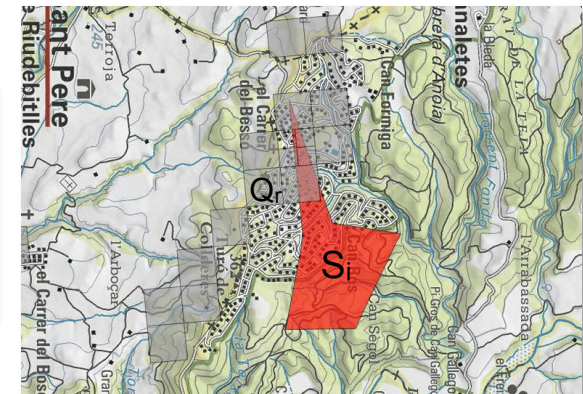


➤ Calcular en cada caso  $\epsilon_i$  para cada quadtree:

- QT {125m,125m,RP2014,17}
- QT {125m,250m,RP2014,17}
- QT {125m,250m,RP2014,17,t}
- etc..

$p_i$  = número de puntos dentro de la geometría  $S_i$   
 $p'_i = \sum p_r * ST\_AREA(Q_r \cap S_i) / ST\_AREA(Q_r)$

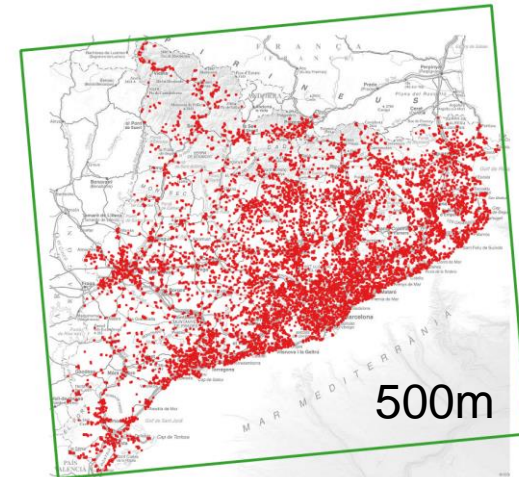
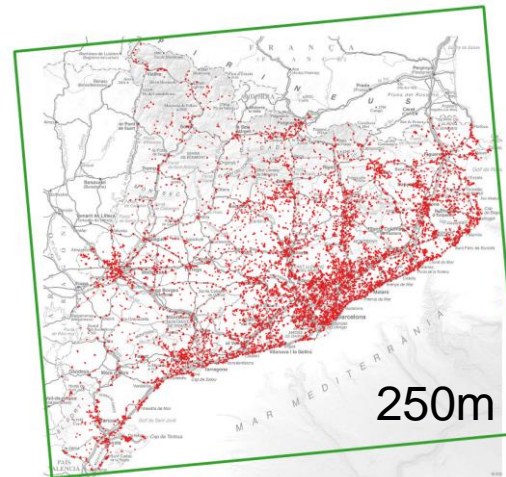
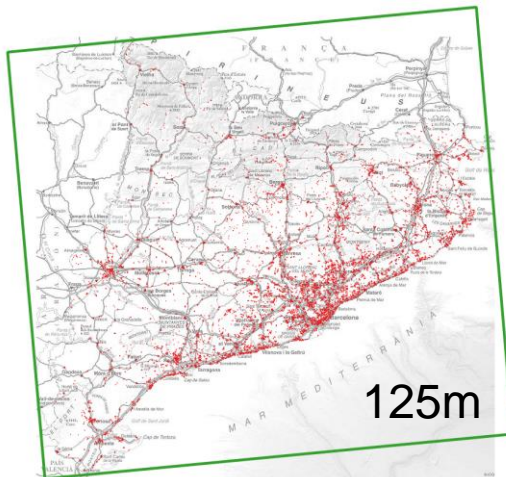
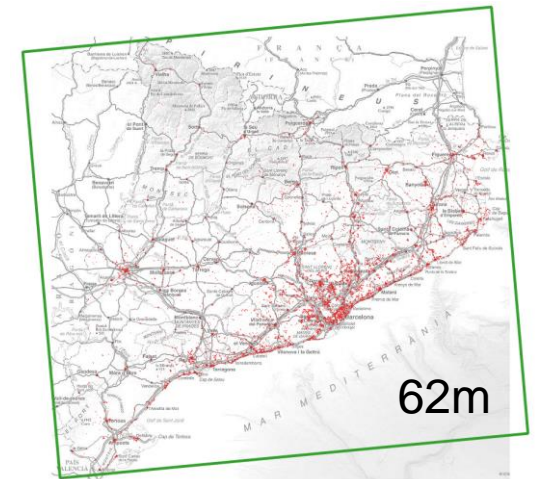
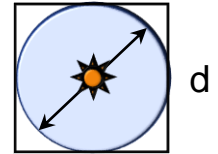
$$\epsilon_i = \frac{|p'_i - p_i|}{p_i}$$





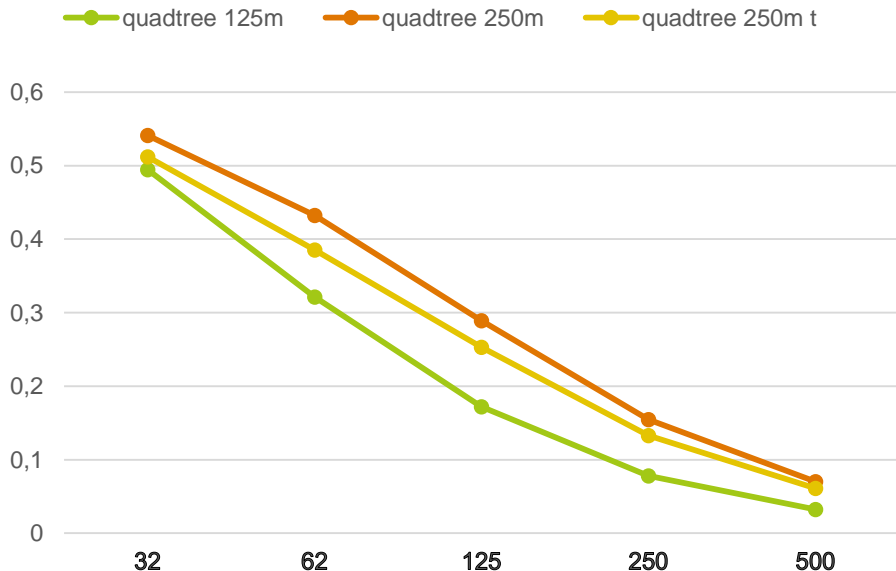


50.000 cuadrados con posiciones aleatorias, de diferente tamaño, y con la restricción de que exista población dentro de él.

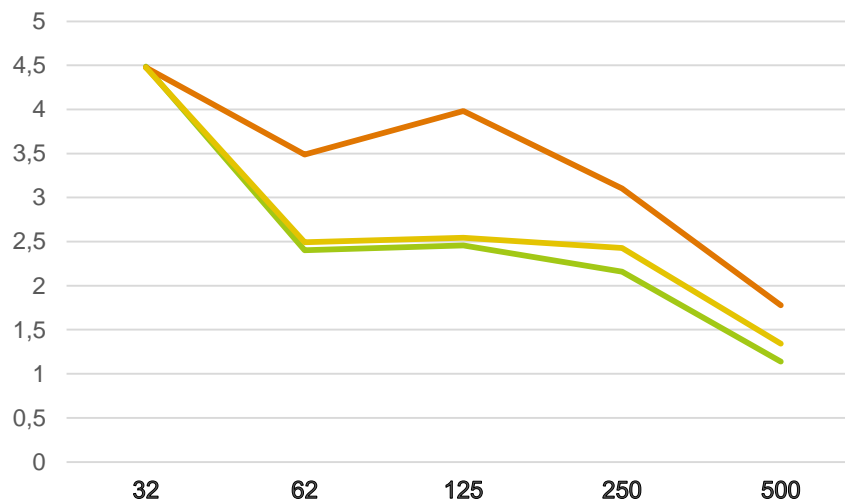


Mediana $\epsilon_i$		lado				
		32	62	125	250	500
1	QT {125m,125m,RP2014,17}	0,495	0,321	0,172	0,078	0,032
2	QT {125m,250m,RP2014,17}	0,541	0,432	0,289	0,154	0,070
3	QT {125m,250m,RP2014,17,t}	0,512	0,385	0,253	0,133	0,061
Stddev $\epsilon_i$						
1	QT {125m,125m,RP2014,17}	4,488	2,408	2,457	2,160	1,140
2	QT {125m,250m,RP2014,17}	4,476	3,487	3,981	3,104	1,778
3	QT {125m,250m,RP2014,17,t}	4,474	2,491	2,547	2,427	1,343

Mediana  $\epsilon_i$  vs tamaño cuadrado



Desviación estándar vs tamaño cuadrado



$$F \left( \frac{p'_i - p_i}{p_i} \right)$$



$l$

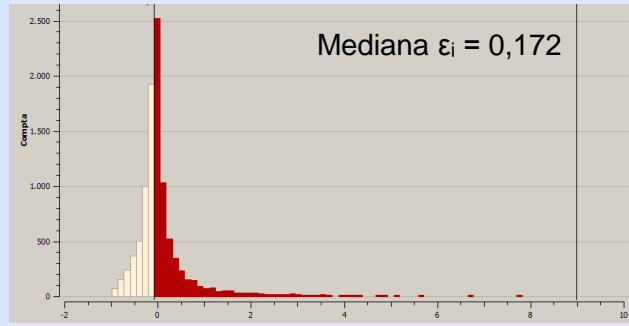


125

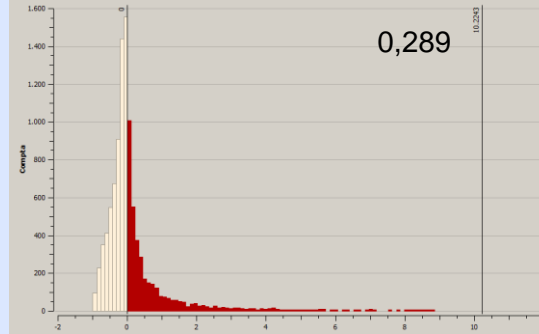
250

500

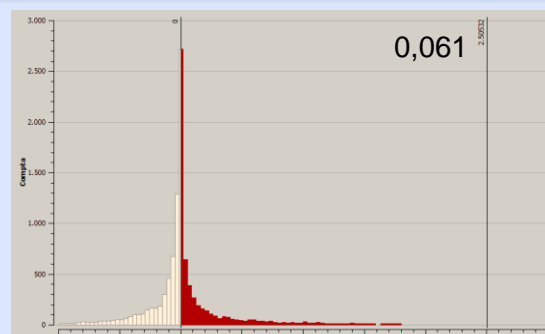
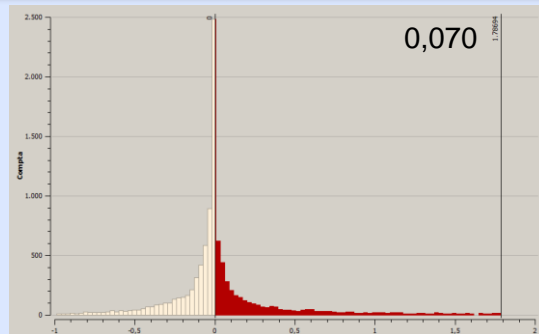
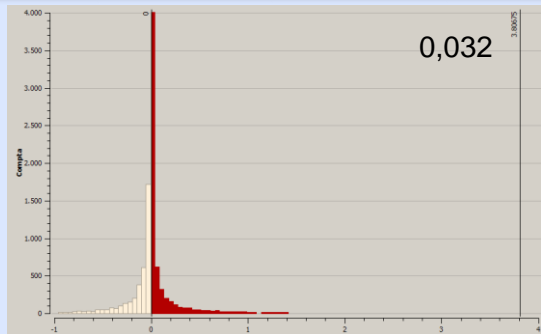
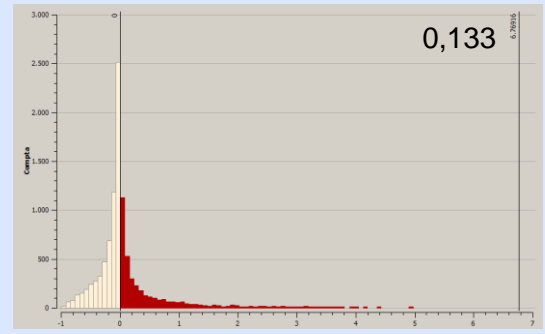
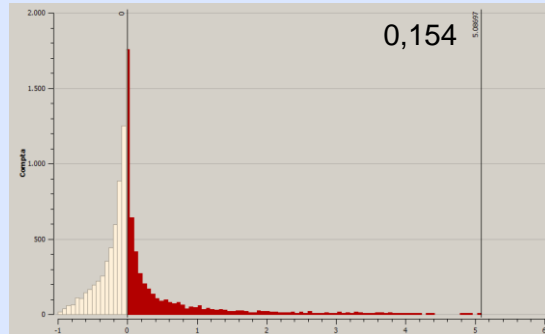
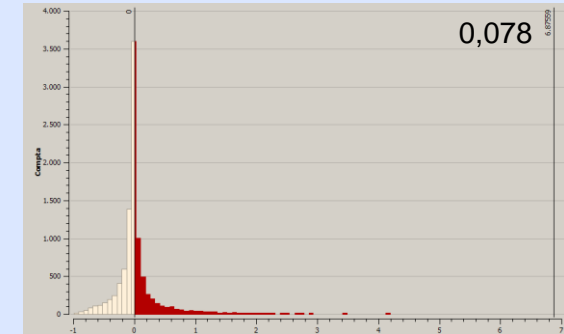
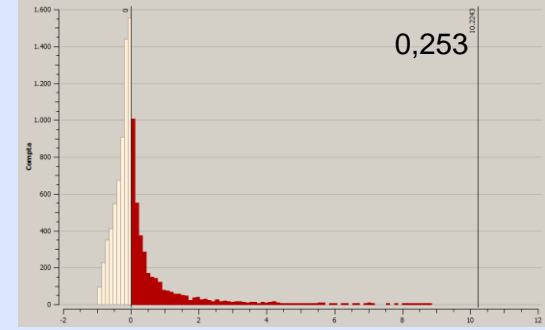
QT {125m,125m,RP2014,17}

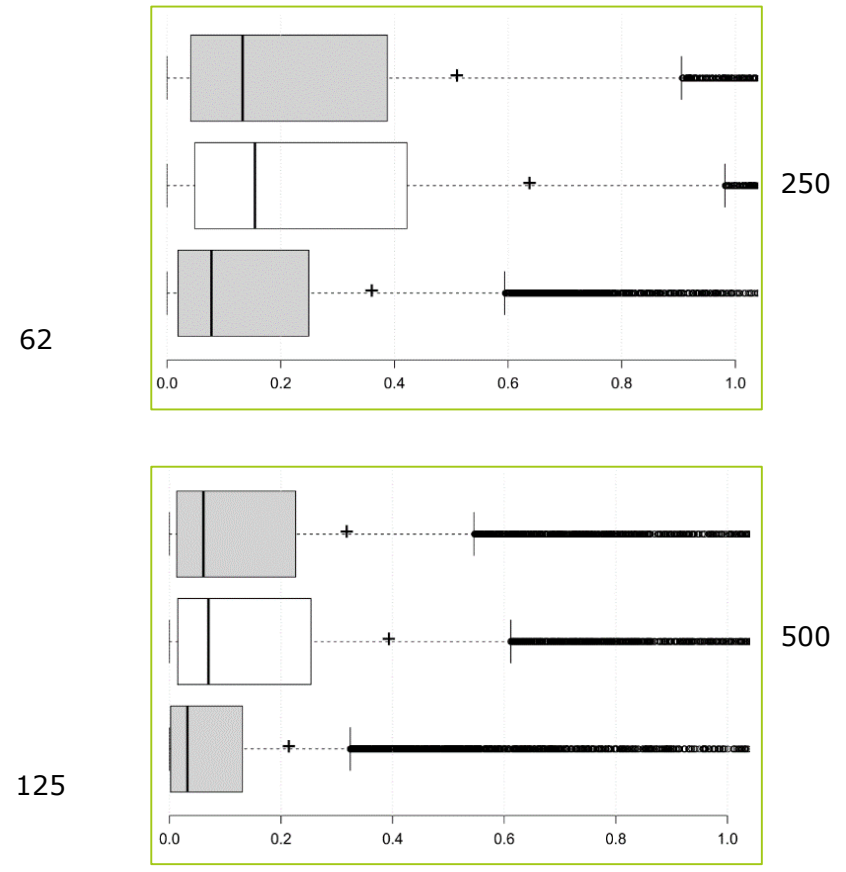
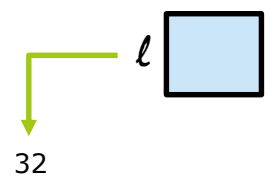
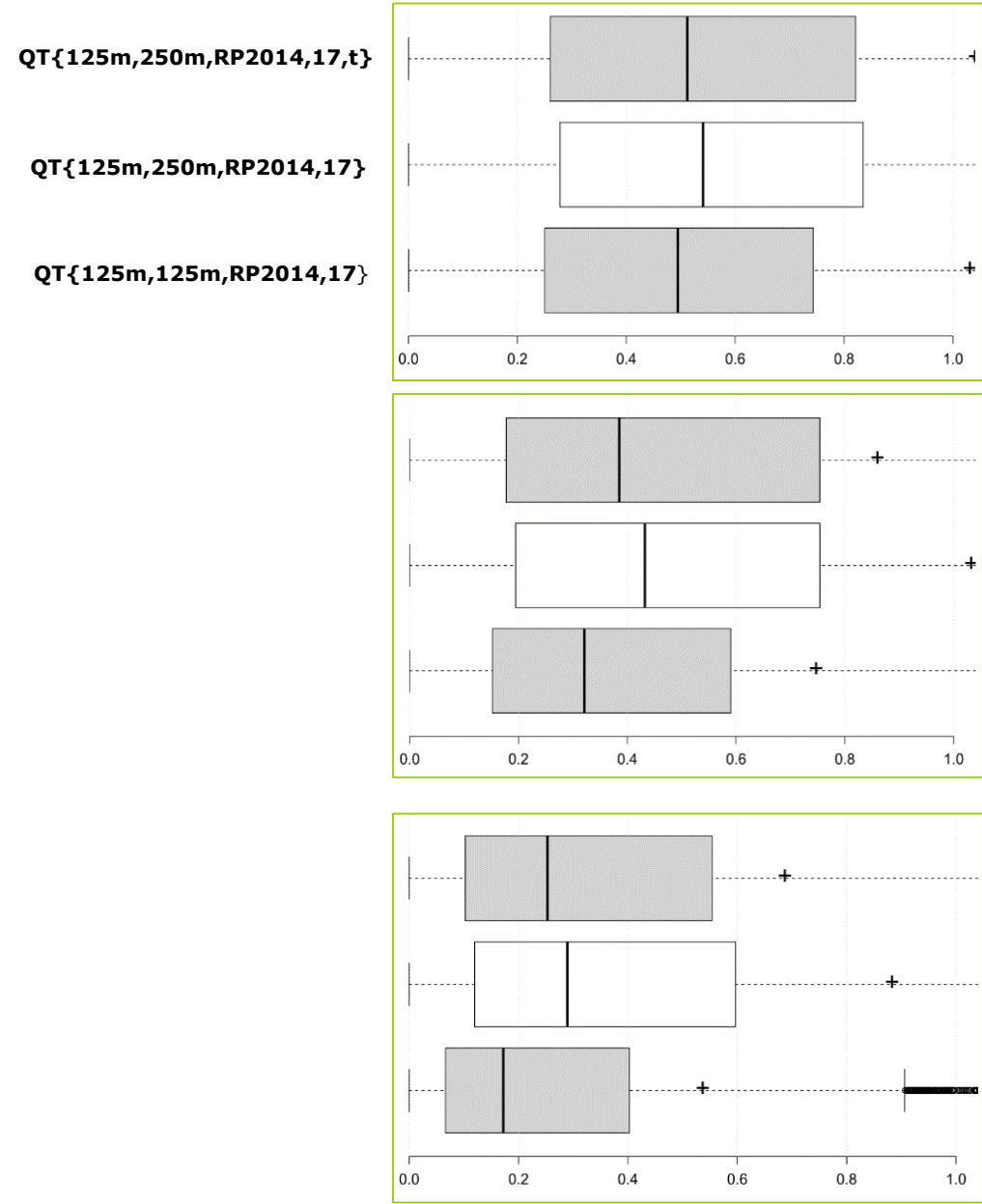


QT {125m,250m,RP2014,17}



QT {125m,250m,RP2014,17,t}

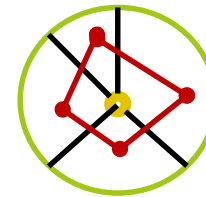




125

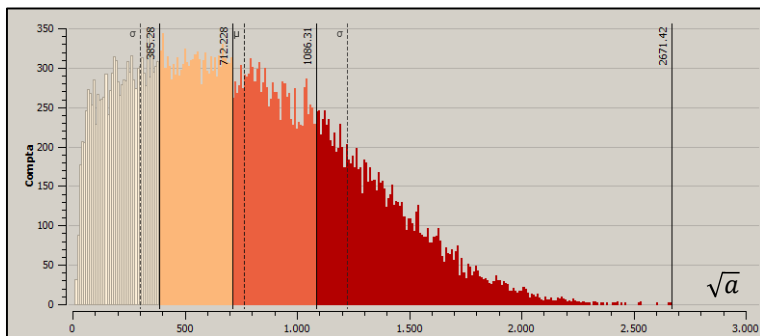
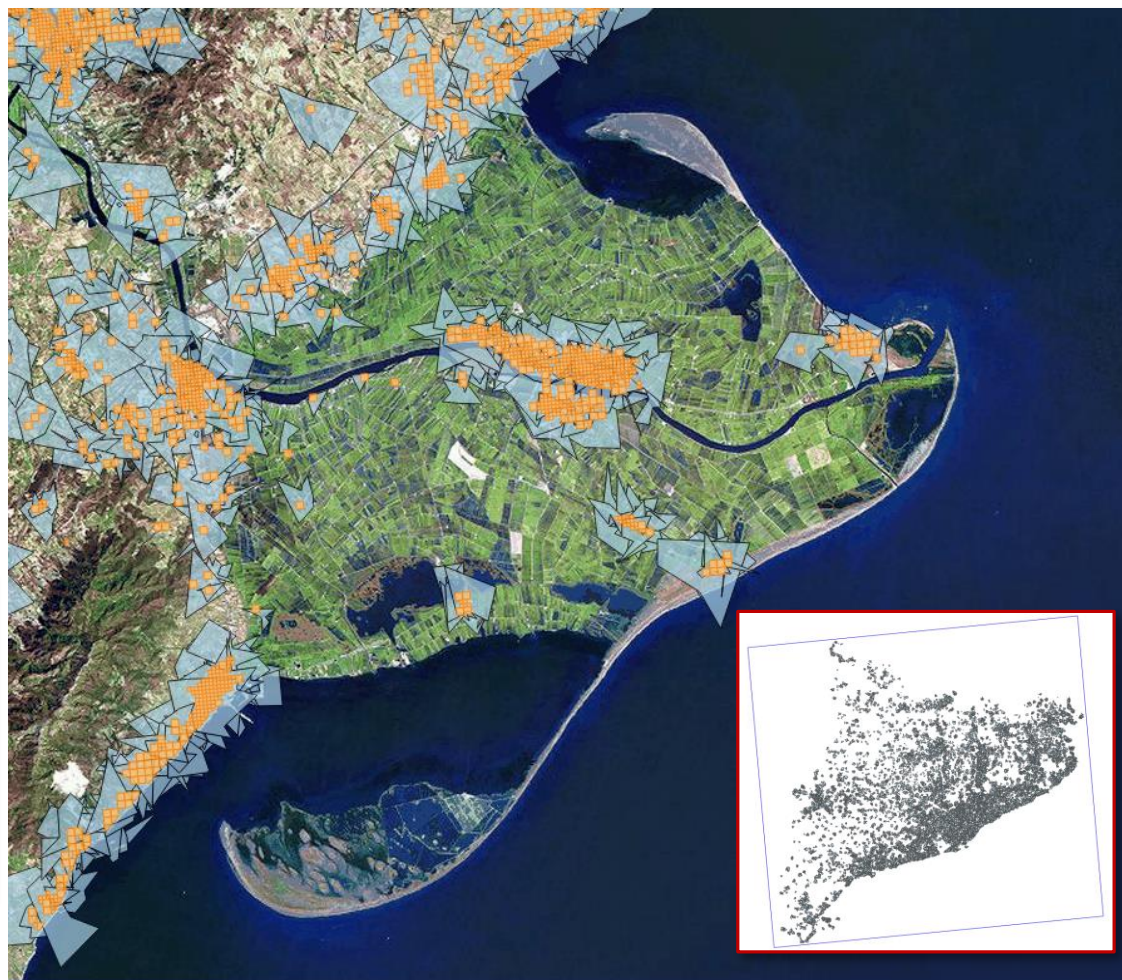
50000 polígonos generados aleatoriamente:

- reutilización centros experimento anterior
- selección aleatoria de un radio y división en n sectores aleatorios
- selección aleatoria de 1 punto en cada sector
- ✓ unión de los puntos, comprobación de que la geometría es válida

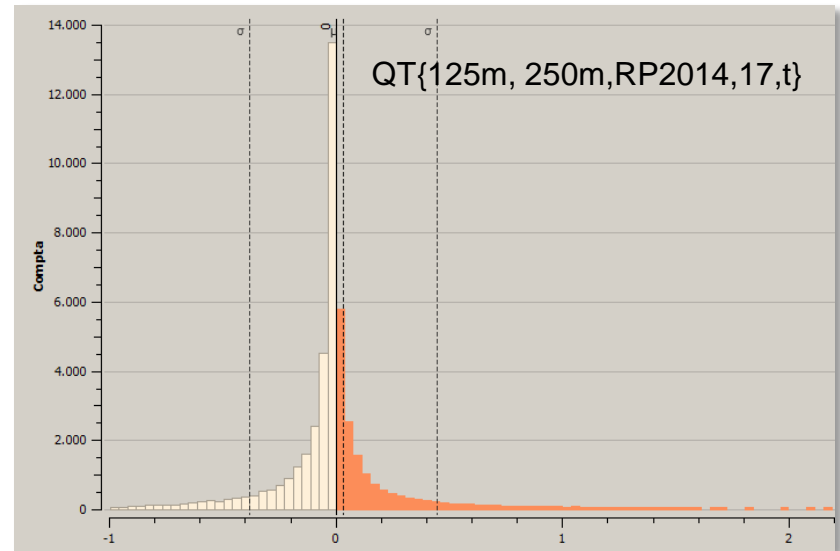
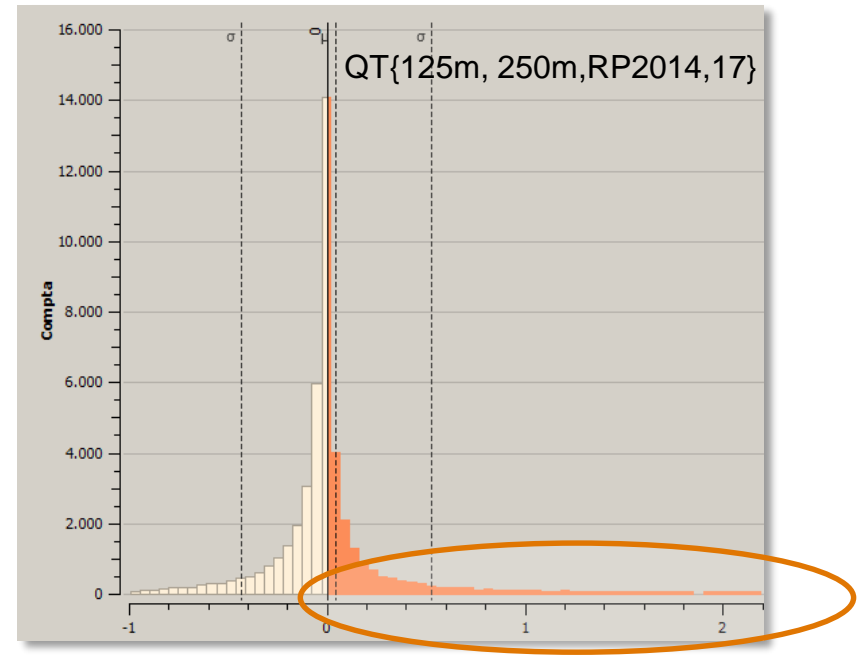
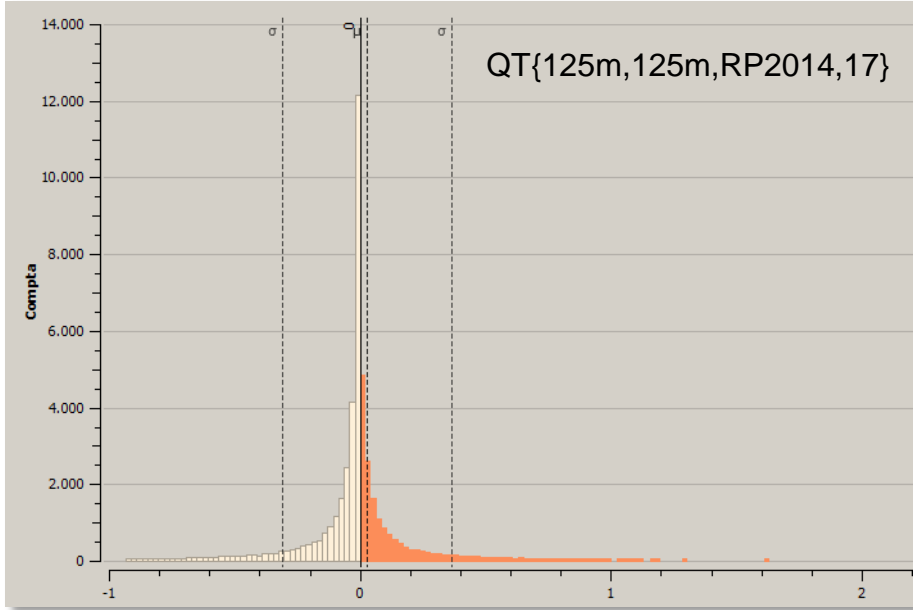


➤ Cálculo de  $\epsilon_i$  para cada uno de los quadrees

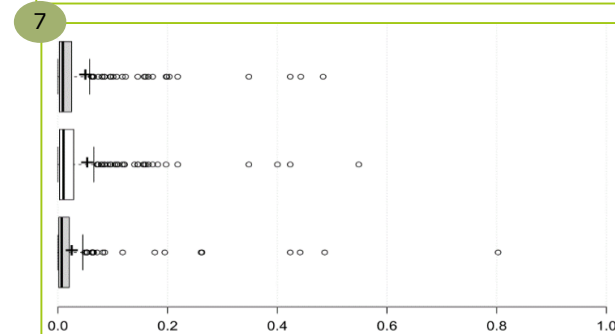
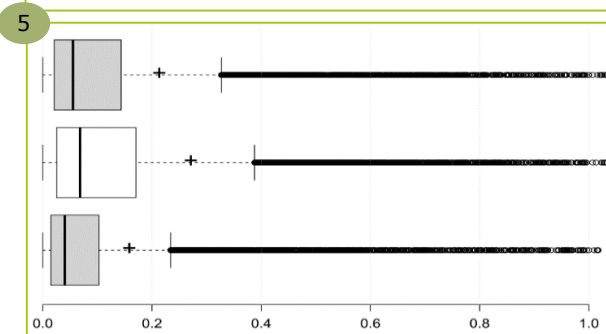
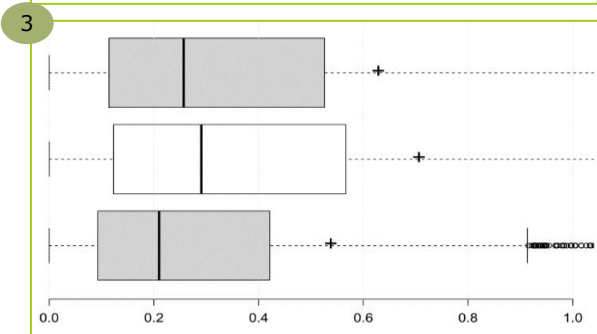
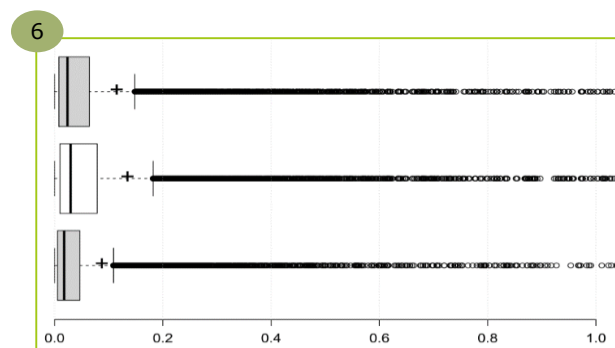
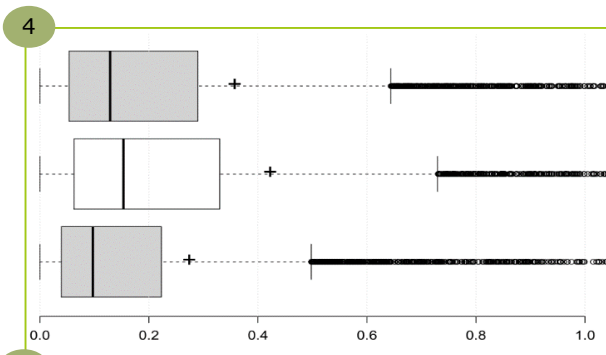
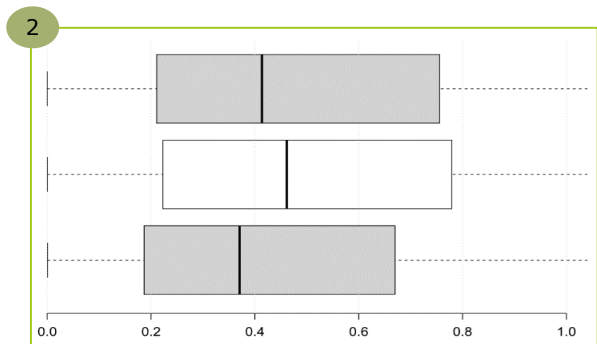
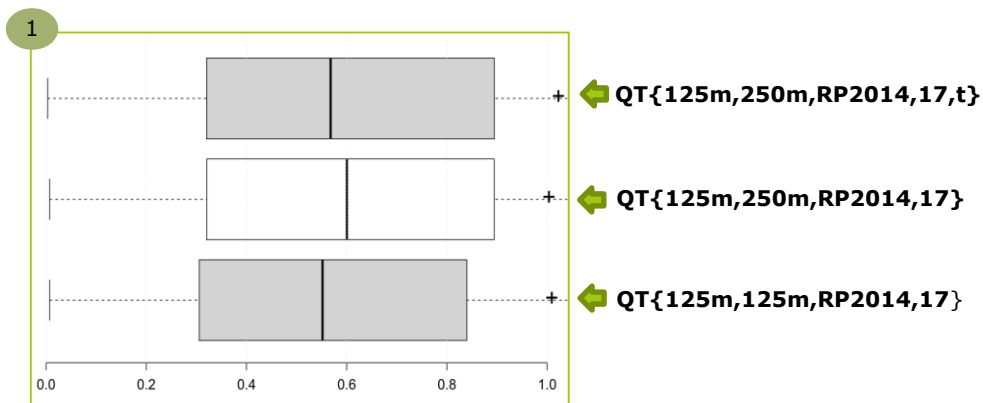
Vertices	F	media àrea	maximo area	min area
4	484	388104.479	3523856.364	23.454
5	5349	605138.829	5000220.933	11.347
6	12200	689349.712	6420900.238	20.074
7	13313	741150.445	6768701.343	143.618
8	9732	776475.394	5985818.791	76.234
9	5350	854488.028	7044323.070	81.341
10	2286	864726.190	7136507.463	423.589
11	910	858276.262	4785315.530	865.238
12	262	833356.607	4119275.589	611.671
13	88	791260.992	4084807.041	954.528
14	25	730933.474	2762205.941	1287.796
16	1	1453.833	1453.833	1453.833

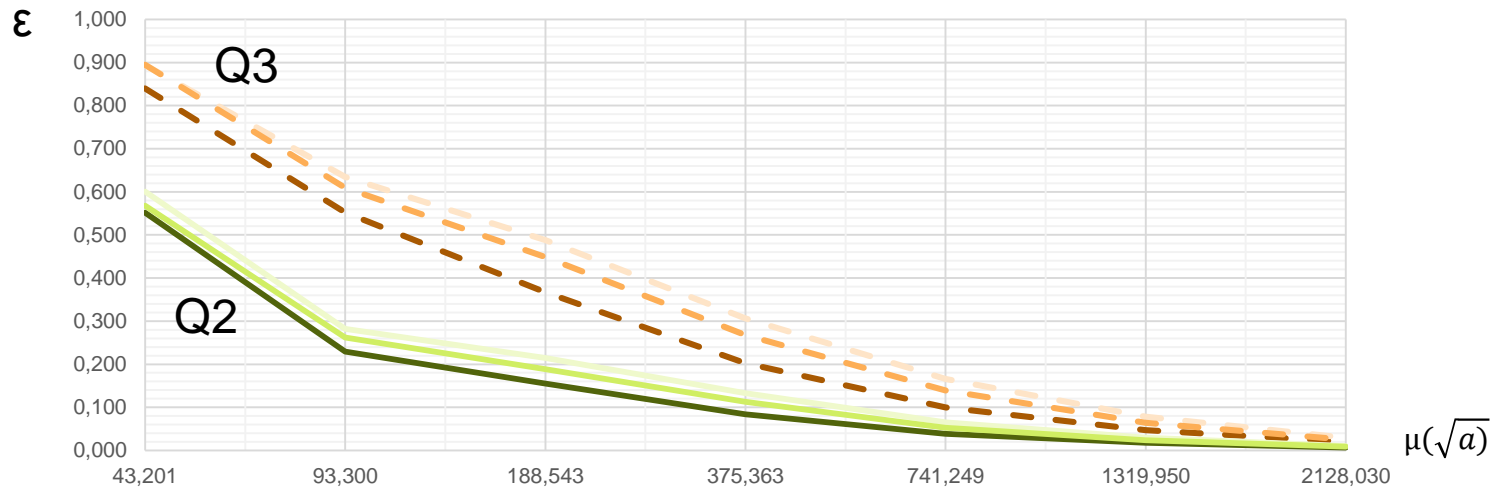


$$F\left(\frac{p'_i - p_i}{p_i}\right)$$



	Grupo		
	Intervalo área	media sqrt(area)	casos
1	< 3906.25	43,201	906
2	3906.25- 15625	93,300	1848
3	15625 - 62500	188,543	3955
4	62500 - 250000	375,363	8517
5	250000 - 1000000	741,249	16242
6	1000000-4000000	1319,950	13407
7	> 4000000	2128,030	245
<b>Total (*)</b>		762,151	45120



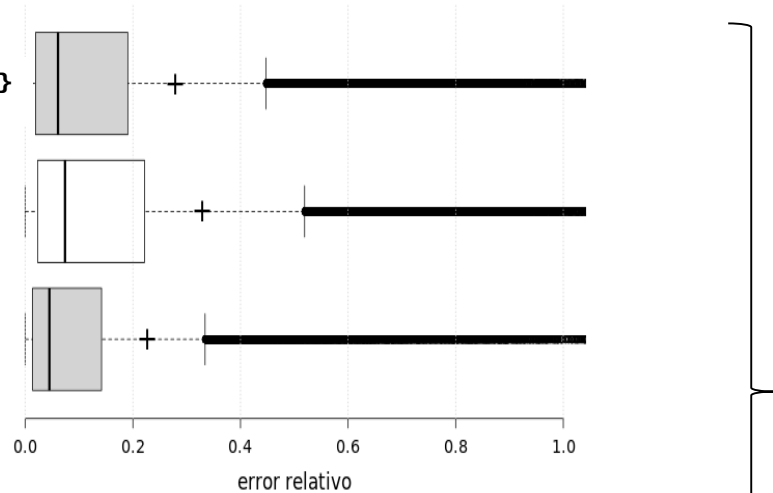


Todos los casos

QT{125m,250m,RP2014,17,t}

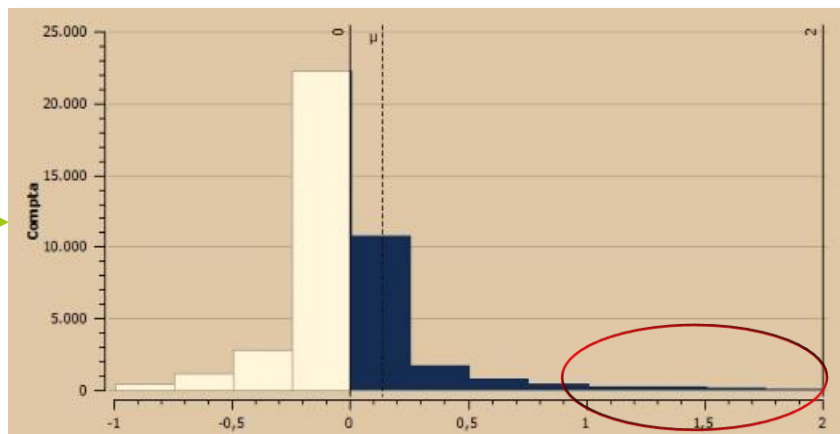
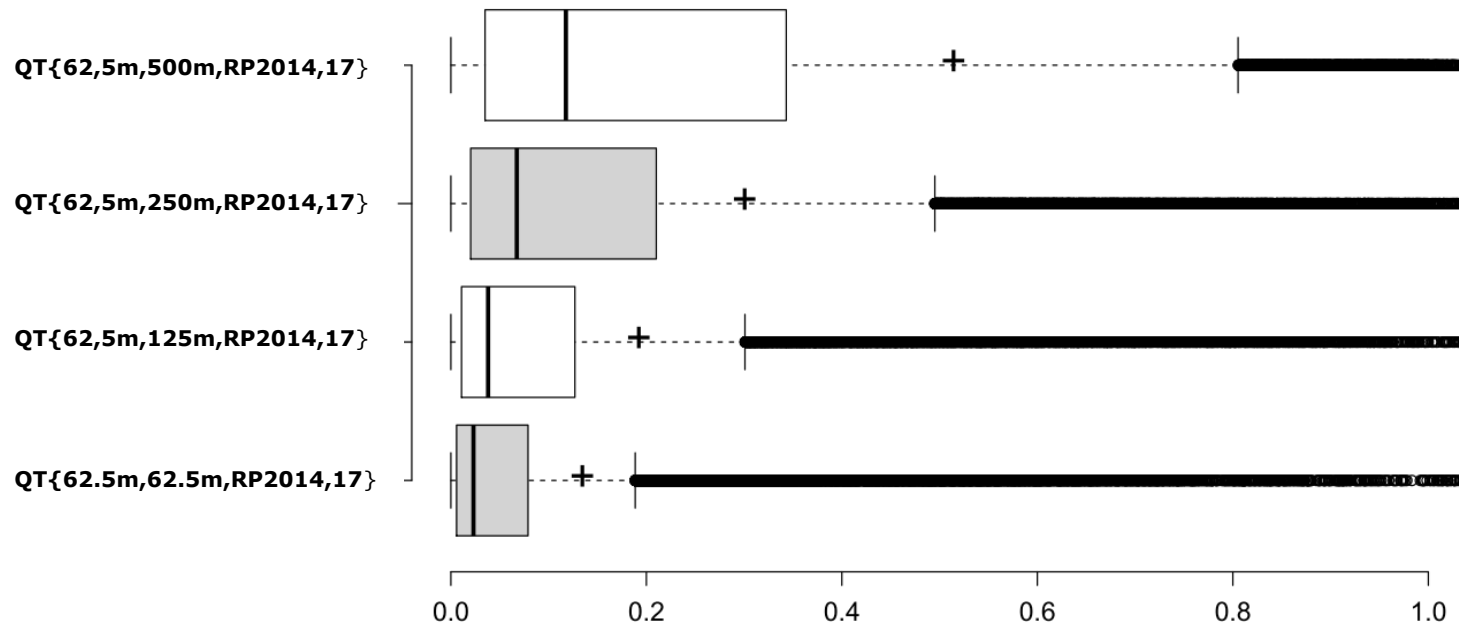
QT{125m,250m,RP2014,17}

QT{125m,125m,RP2014,17}



	QT{125m,125m,RP2014,17}	QT{125m,250m,RP2014,17}	QT{125m,250m,RP2014,17,t}
Q3	0,14	0,22	0,19
Q2	0,04	0,07	0,06
Q1	0,01	0,02	0,02
Media	0,23	0,33	0,28

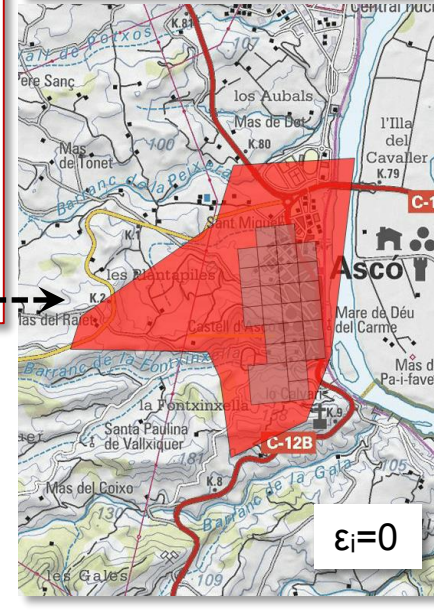
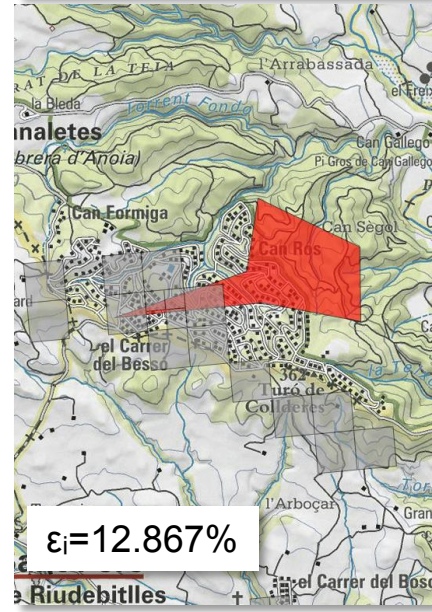
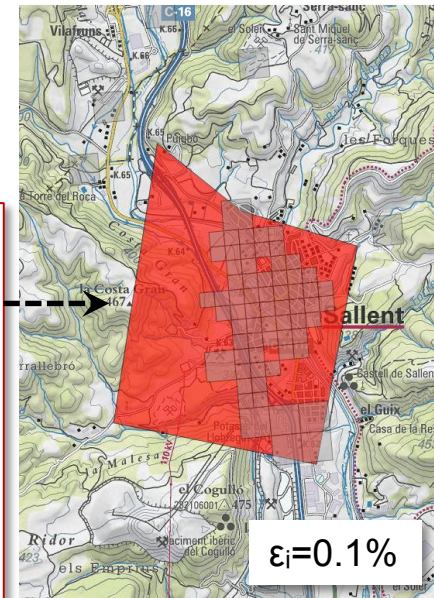
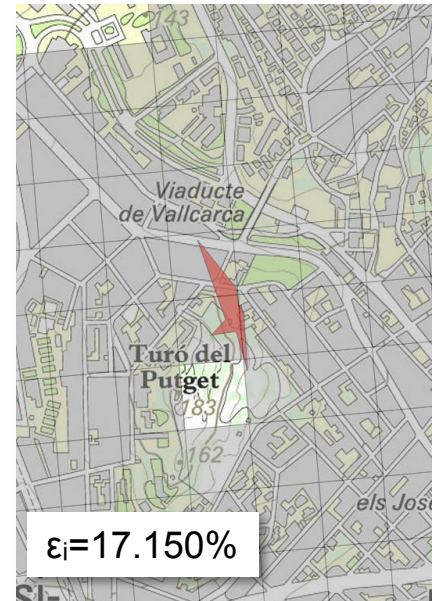
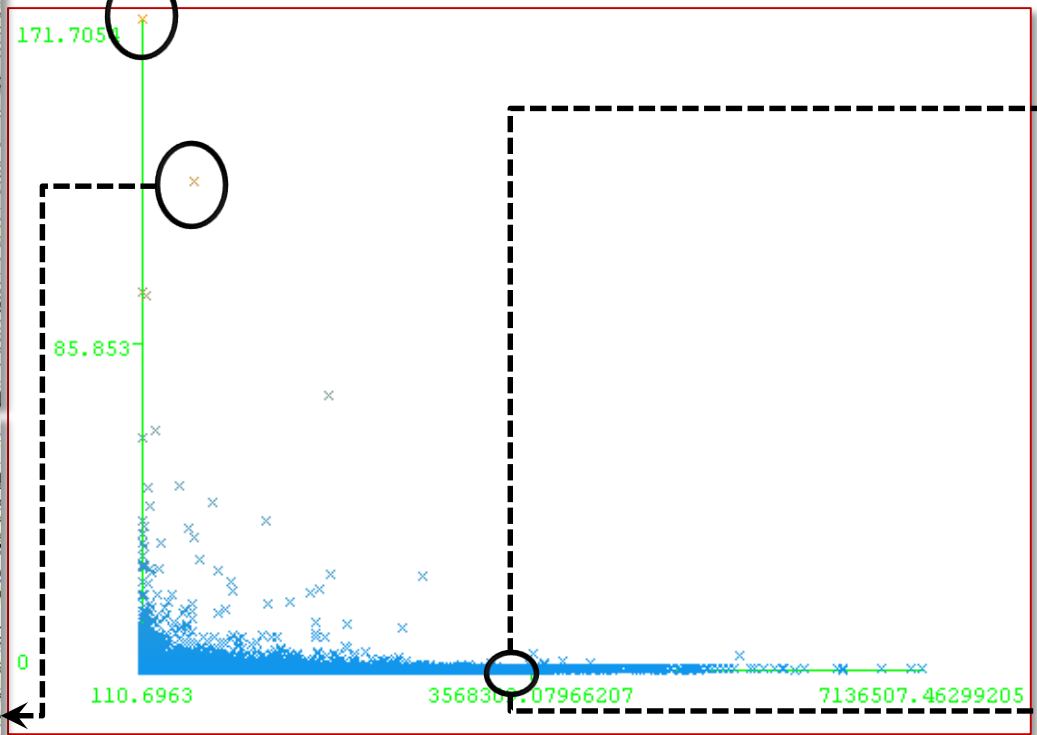




	Mediana $\epsilon$ Resolución mínima			
umbral	62,5	125	250	500
13	0,02	0,04	0,06	0,11
17	0,02	0,04	0,07	0,12
19	0,02	0,04	0,07	0,12
23	0,02	0,04	0,07	0,13

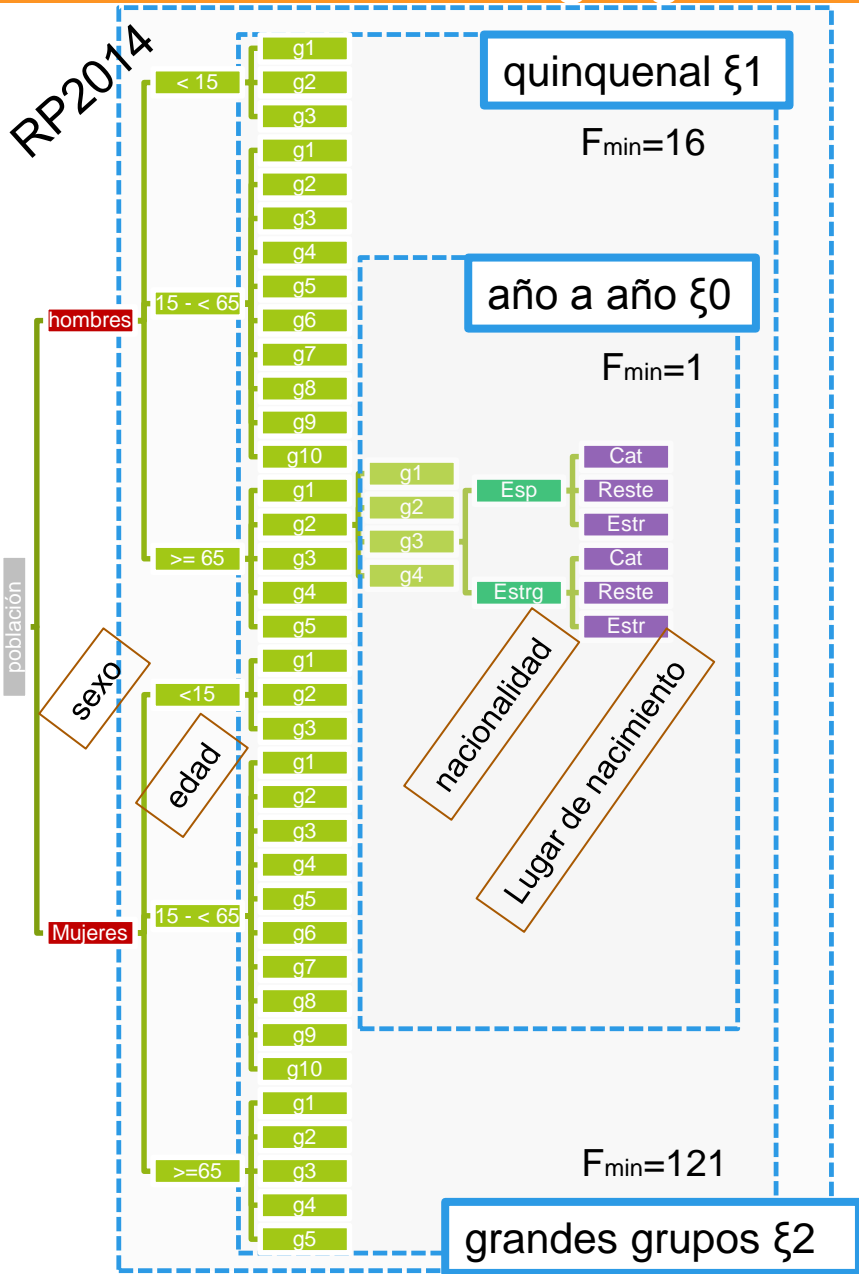
QT{125m; 250m,RP2014,17,t}

$\epsilon_i$  vs área del polígono



$$p_i = \sum p_r * ST\_AREA(Q_r \cap S_i) / ST\_AREA(Q_r)$$

Si  $Q_r \cap S_i = Q_r$  para todo  $r \rightarrow \epsilon_i = 0$



## Función *intercambio* de coordenadas $h_a$ i $h_e$

$$f(h_a, h_e) = \begin{cases} 0 & \text{si } h_a = h_e \text{ en } \xi_1 \\ \frac{1}{d(p_a, p_e)} & \text{si } h_a = h_e \text{ en } \xi_2 \text{ i } d(h_a, h_e) > 0 \\ 0 & \text{si } h_a = h_e \text{ en } \xi_2 \text{ i } d(h_a, h_e) = 0 \\ 0 & \text{si } h_a \neq h_e \text{ en } \xi_2 \end{cases}$$

En  $\xi_2$  mover les coordenadas no tiene ningún efecto pero si en  $\xi_1$ . A medida que  $\xi_2 \rightarrow \xi_1$  la perturbación disminuye

$$F \rightarrow \begin{bmatrix} 0 & \dots & f(h_e, h_a) \\ \vdots & \ddots & \vdots \\ f(h_a, h_e) & \dots & 0 \end{bmatrix}$$

$$P \rightarrow \begin{bmatrix} 0 & \dots & p(h_e, h_a) \\ \vdots & \ddots & \vdots \\ p(h_a, h_e) & \dots & 0 \end{bmatrix} \quad p(h_a, h_e) = \frac{f(h_a, h_e)}{\sum_1^n f(h_a, h_e)}$$