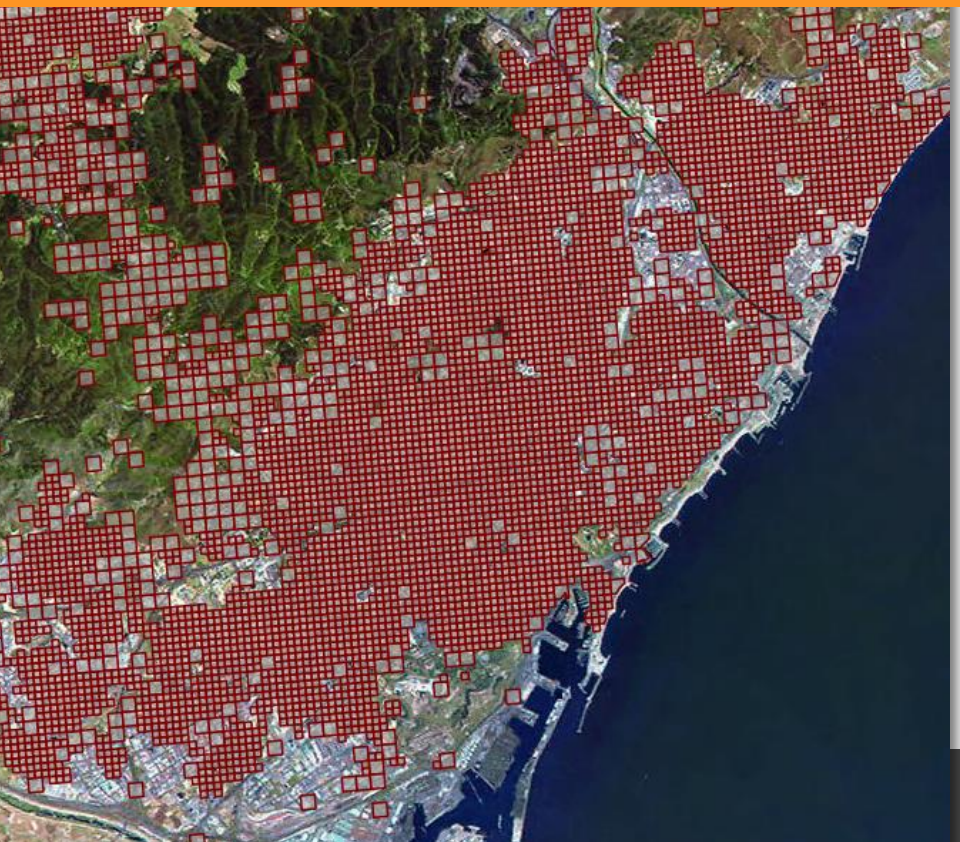
Generalitat de Catalunya
Institut d'Estadística de Catalunya

Statistical disclosure control on visualising geocoded population data using a structure in quadtrees



Eduard Suñé, Cristina Rovira, Daniel Ibáñez, Mireia Farré

www.idescat.cat

Contents

1. Disclosure control by spatial aggregation using quadtrees

2. Quadtree parameters

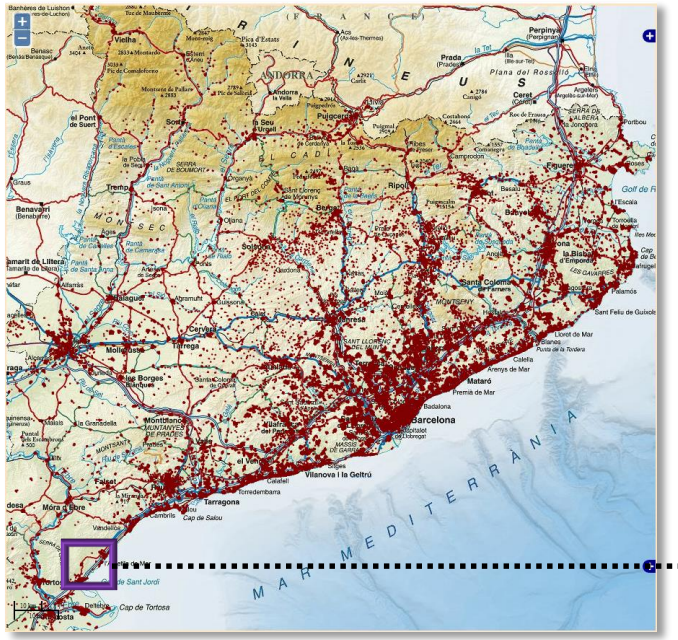
3. Border effect. Proposed solution

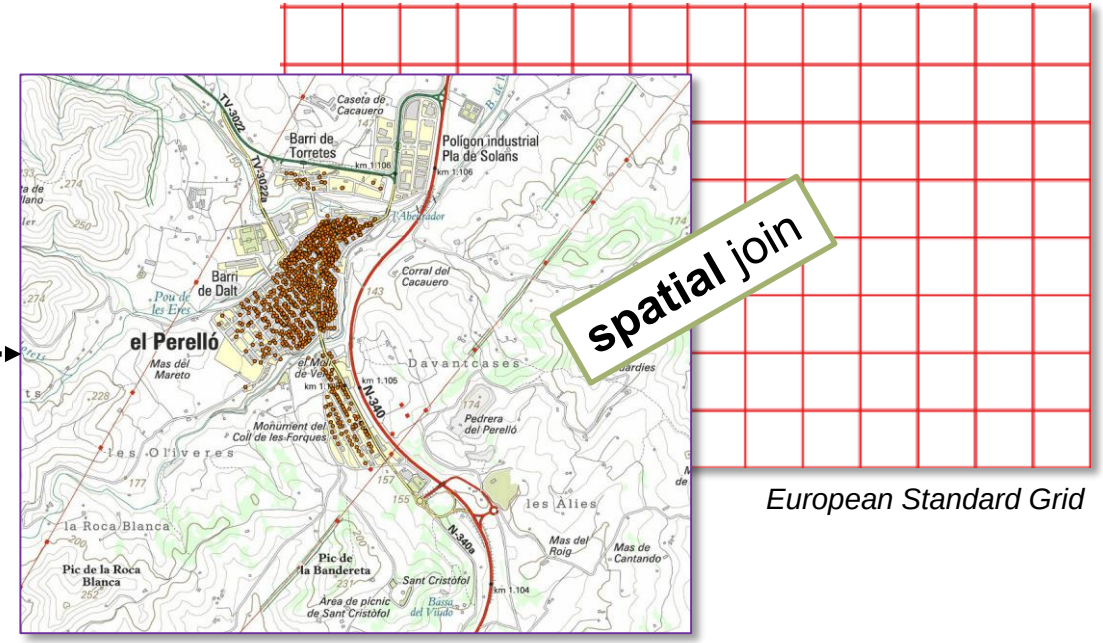
4. Errors in population calculations

5. Estimations of errors. Monte Carlo experiment

6. Conclusions

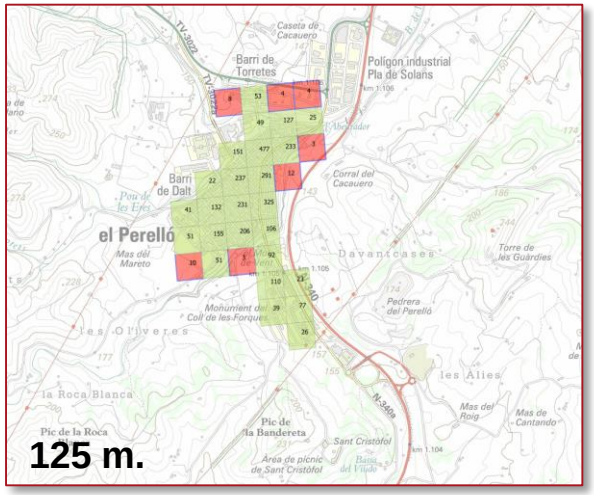
1. Disclosure control by spatial aggregation using quadtrees



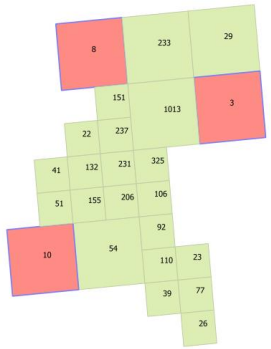


European Standard Grid

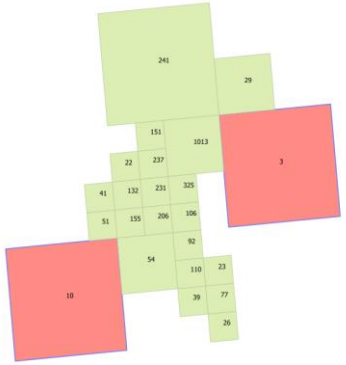
division — **aggregation**



125 m.



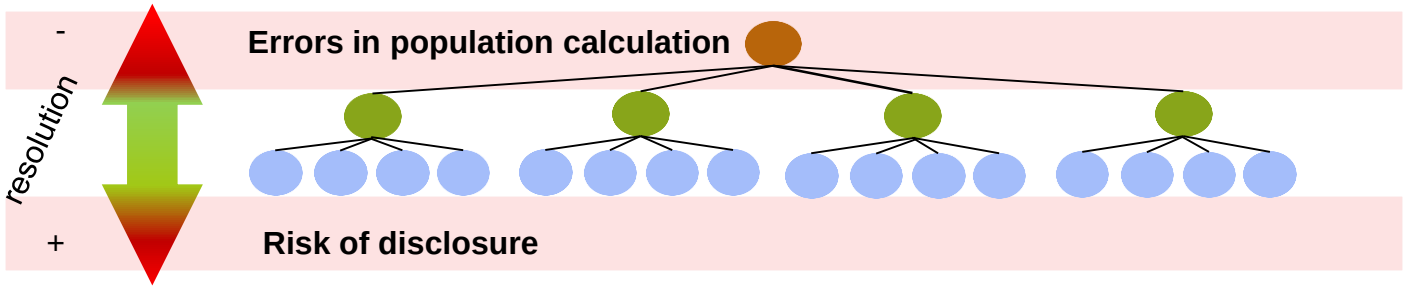
250 m.



500 m.

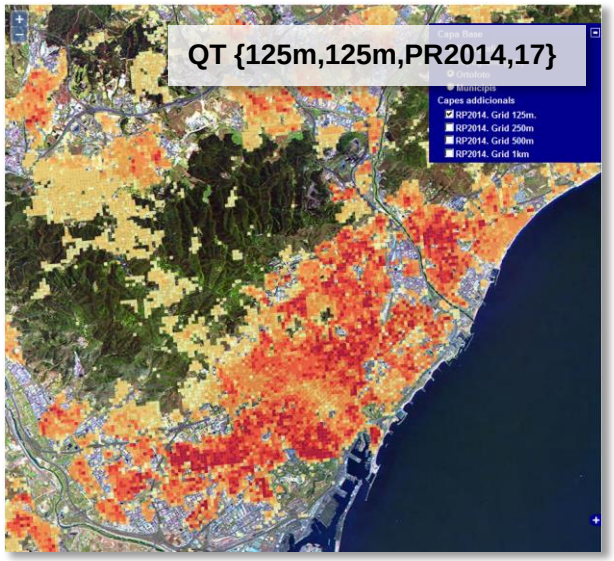
1 Km.

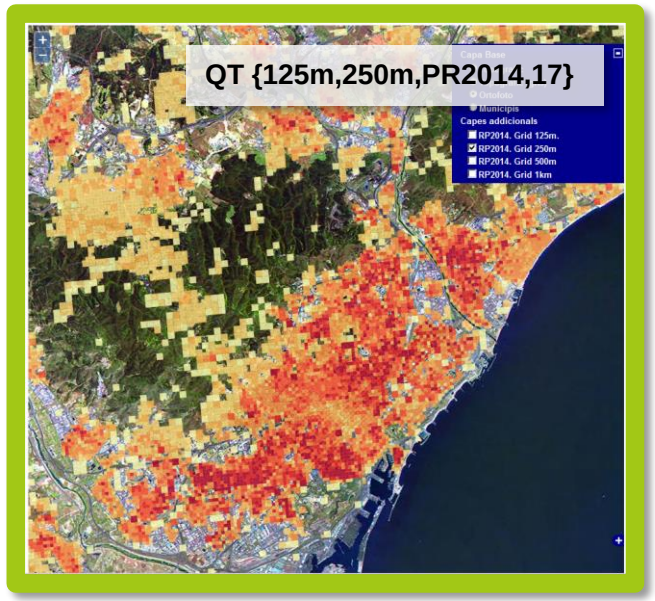
2. Quadtree parameters

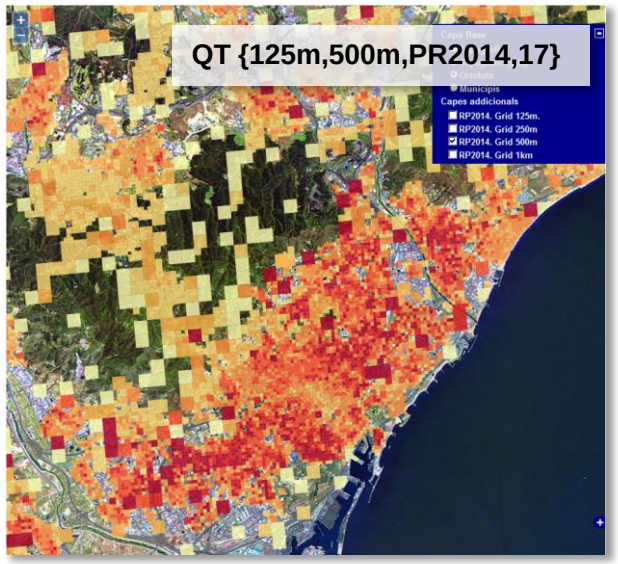


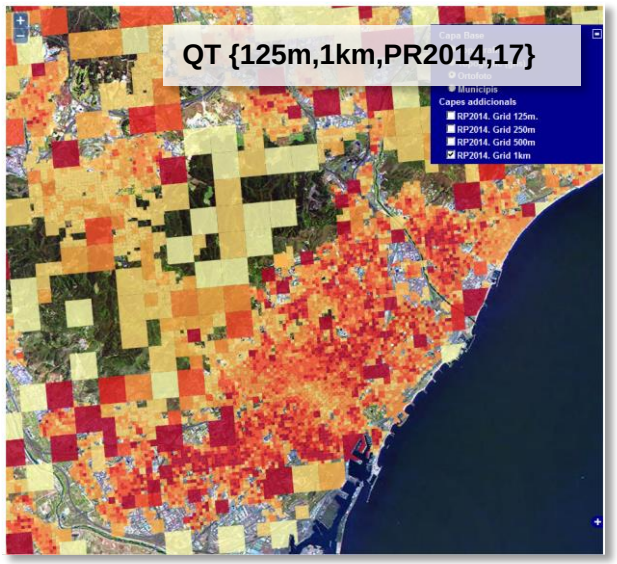
A quadtree is defined by {maximum resolution, minimum resolution, georeferenced data, threshold}

Decision → QT{125m,250m,PR2014,17}



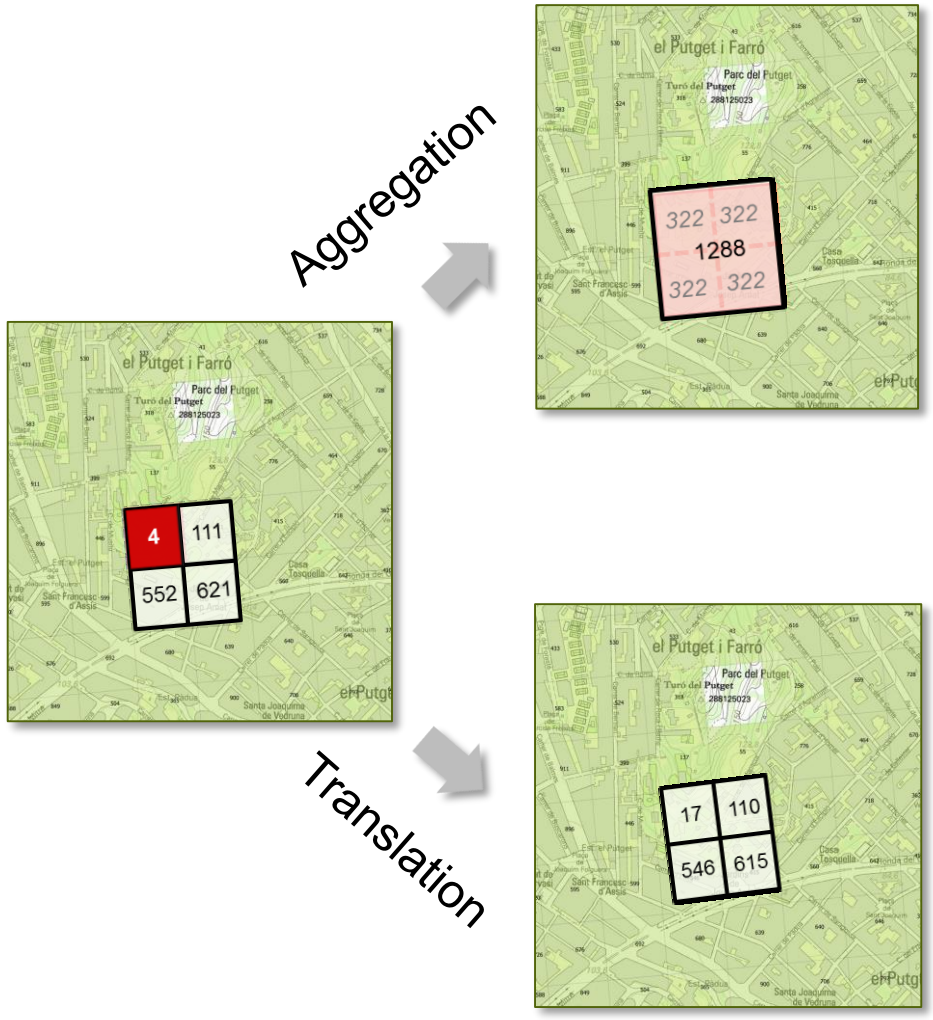






3. Border effect. Proposed solution

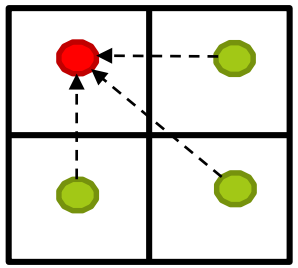
Border effect: undesired aggregation due to high values of the variance



Absolute error when **aggregating**

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

Can be avoided by **translations** when the absolute error is less than the aggregation



Absolute error when **translating**

$$\epsilon' = \sum | n_i - n_i' |$$

Squares of	% population	
	Aggregation	Translation
250 m	15.97	3.52
125 m	84.03	96.48
Total	100	100

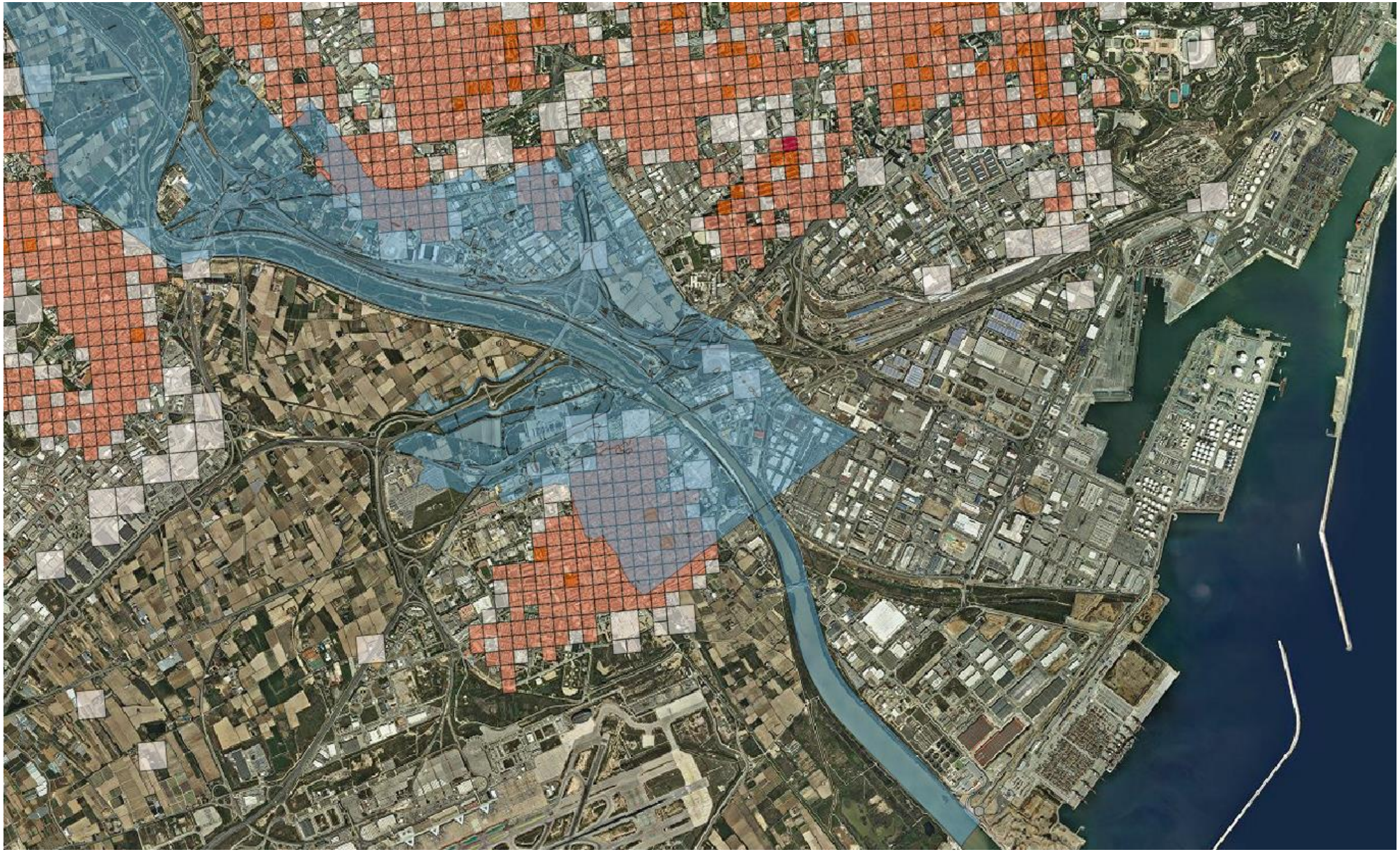
Only **0.85%** of the total population has been translated

If $\epsilon' < \epsilon$ it is better to translate

New quadtree with translations:

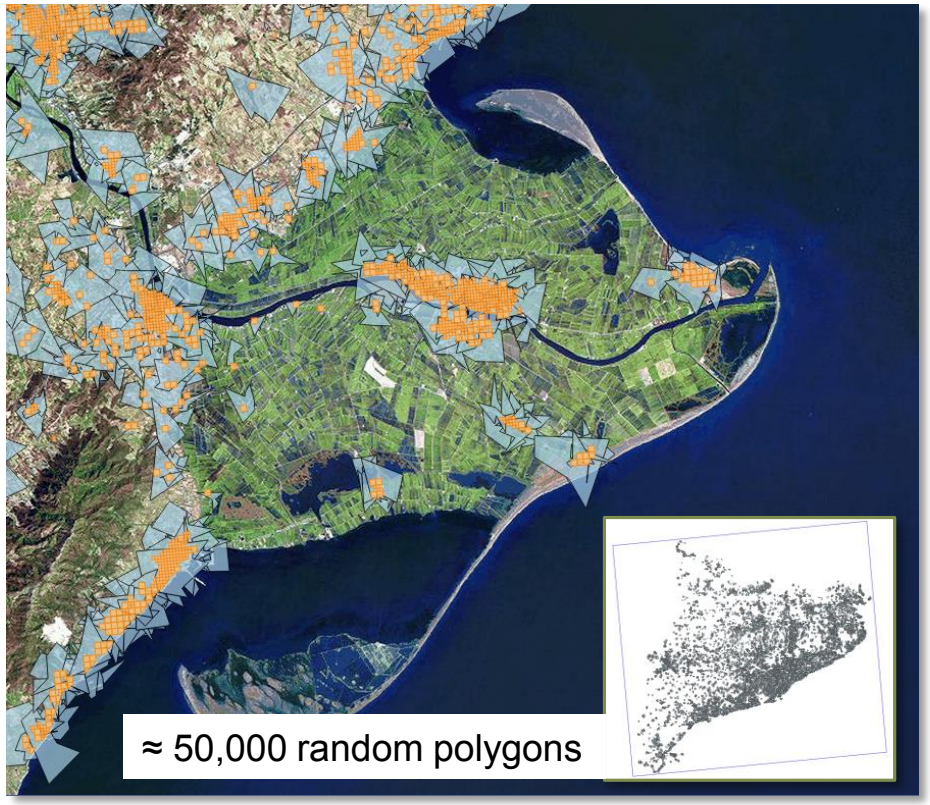
QT{125m,250m,PR2014,17,t}

4. Errors in population calculations



What is the error when calculating the population within a certain polygon, for example in a flood area, using a quadtree?

5. Estimations of errors. Monte Carlo experiment



For each polygon S_i, relative error is

$$\varepsilon_i = \frac{|n'_i - n_i|}{n_i} \qquad [1]$$

n_i = Population within the X geometry S_i

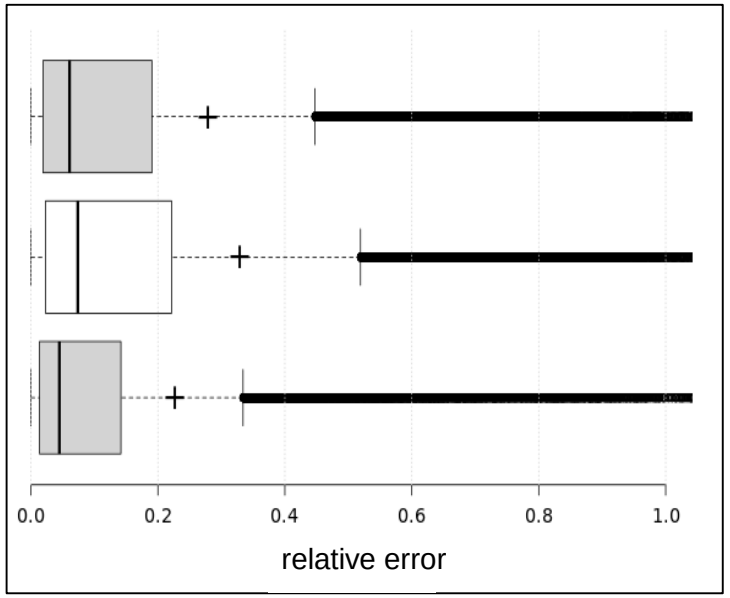
$n'_i = \sum n_r * \text{AREA}(Q_r \cap S_i) / \text{AREA}(Q_r)$

	Quartile 1	Median	Quartile 3	Mean
QT{125m, 250m, PR2014, 17, t}	0.02	0.05	0.19	0.28
QT{125m, 250m, PR2014, 17}	0.02	0.07	0.22	0.33
QT{125m, 125m, PR2014, 17}	0.01	0.04	0.14	0.23

QT{125m,250m,PR2014,17,t} ⇨

QT{125m,250m,PR2014,17} ⇨

QT{125m,125m,PR2014,17} ⇨



6. Conclusions

- The use of **quadtrees** for the dissemination of georeferenced data is a good method for the preservation of statistical confidentiality, as a certain balance between security and accuracy is achieved.

- This preservation method may lead to undesirable aggregations in areas which correspond to siblings in the hierarchy, due to the high values of population variance (**border effect**). A solution to the border effect consists of translating microdata under the condition that the absolute error of the aggregation is greater than that of the translation.

- **Monte Carlo** techniques allow the estimation of the relative error distribution for the population calculated within the quadtree structure **QT{125m,250m,PR2014,17,t}**. We have obtained a value of 5.3% for the median of these errors.