



*Lessons learned from
disaggregating population data by
using different data sources.*

Radoslaw Jablonski
Central Statistical Office of Poland
RJablonski@stat.gov.pl



Agenda

- Aim of the project
- Data types used in project
- Test area
- Review of used method for solving the disaggregation problem
- Conclusions



Aim of the project

- Aggregate demographic data using the top-down approach for the entire country.
- Deaggregate demographic data using the bottom-up approach for the entire country.



Data types used in project.

1) Bottom-up method:

- Address points data set with population from Census 2011.
- Grid data set series 1x1km (GRID_ETRS89_LAEA_1K)

2) Top-down method:

- Enumeration area districts with population from Census 2011.
- High resolution soil sealing layer for the year 2006.

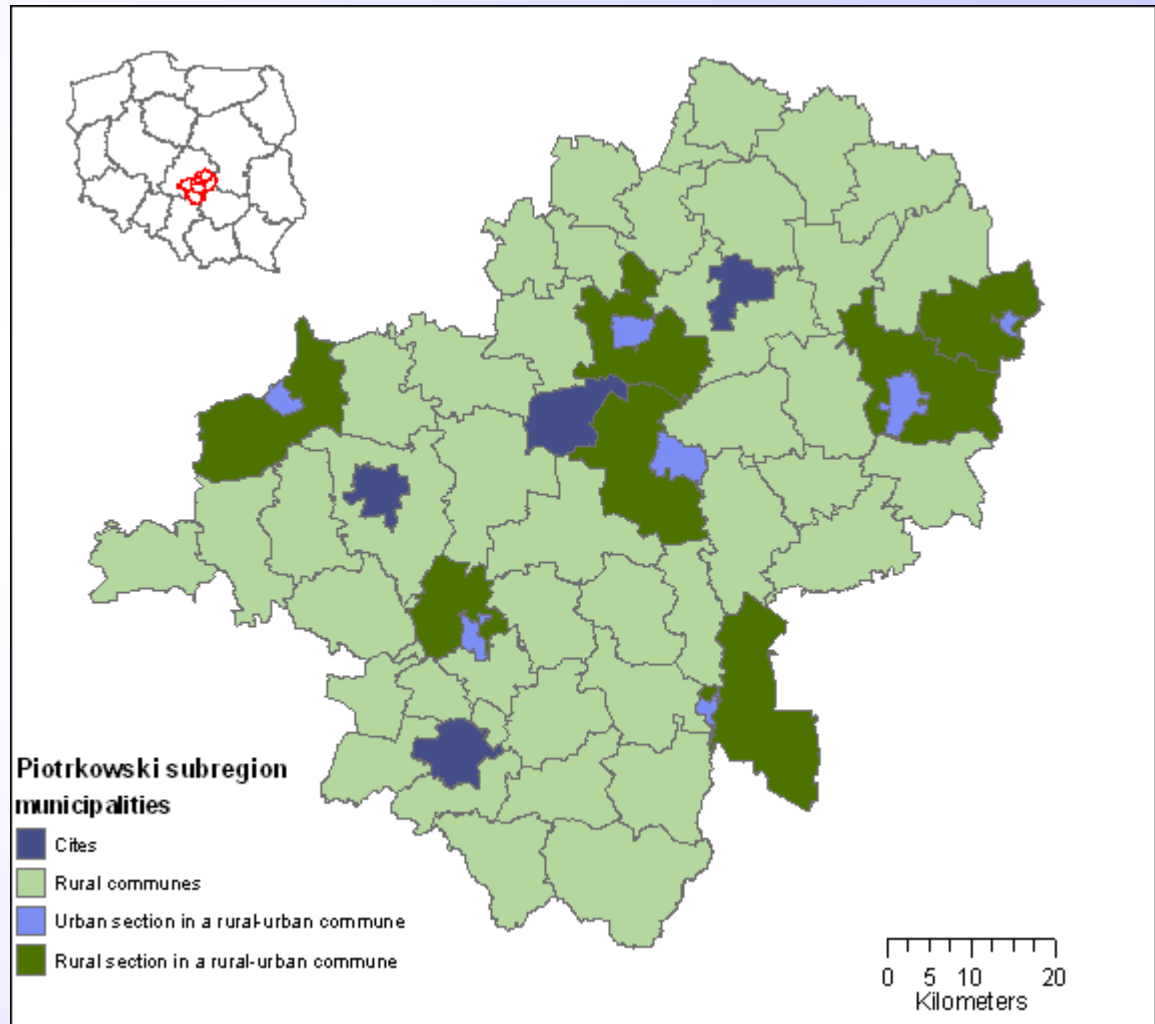


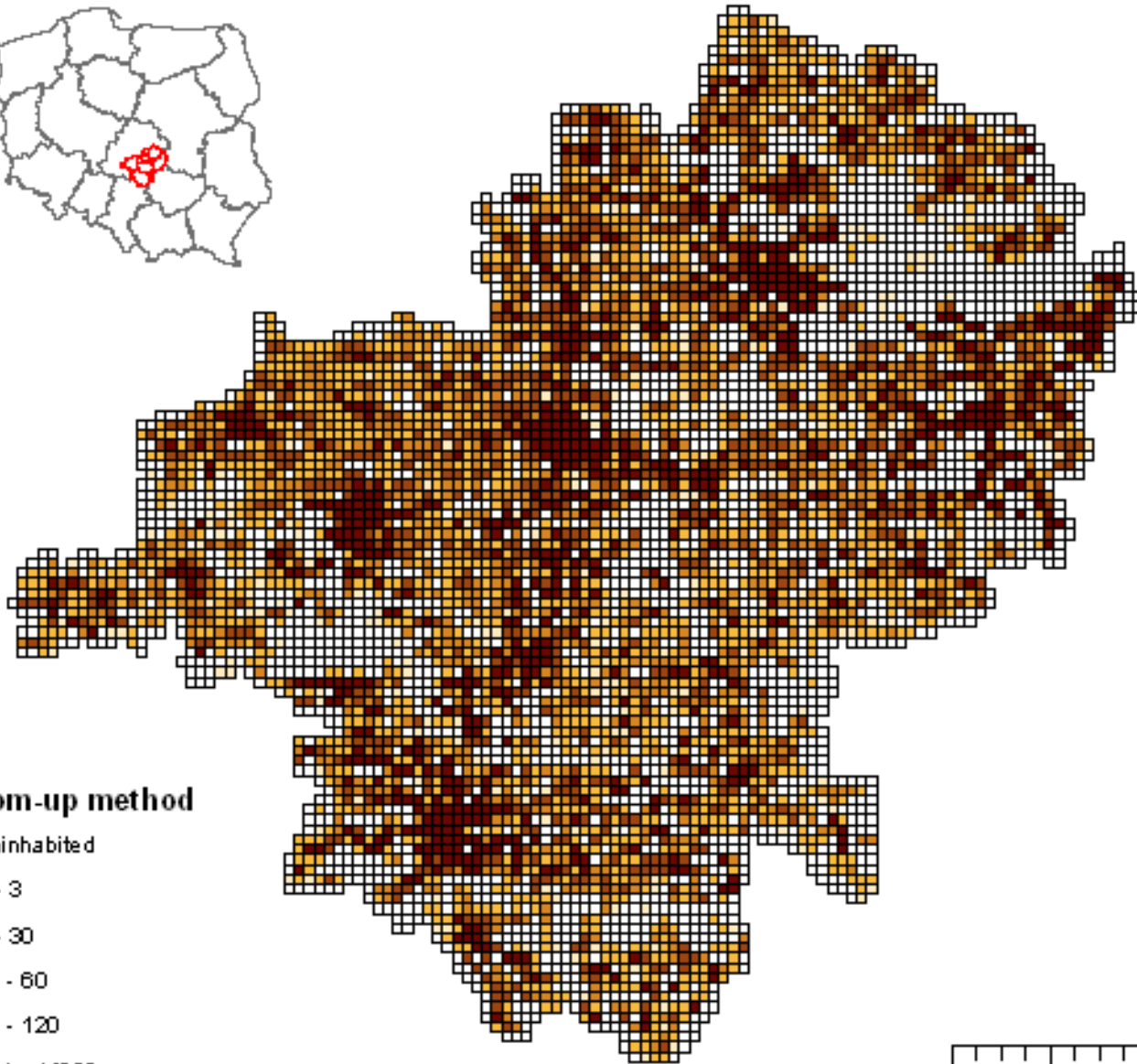
Test area

Subregion Piotrkowski was chosen for testing the bottom-up and top-down method .







Municipalities:

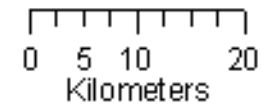
- Cities – (4)
- Rural communes – (42)
- Urban section in a rural-urban commune – (7)
- Rural section in a rural-urban commune – (7)





bottom-up method

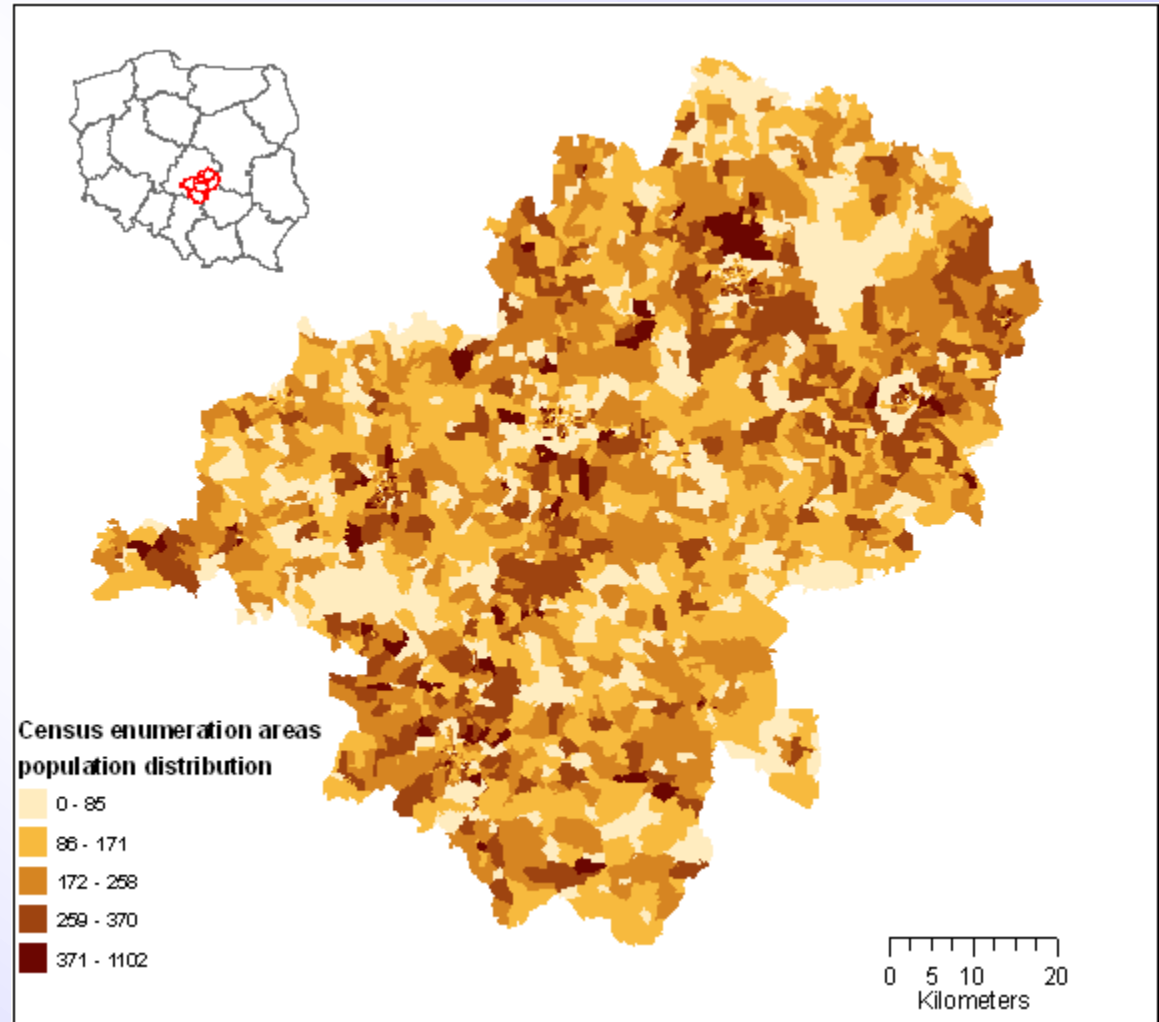
-  uninhabited
-  1 - 3
-  4 - 30
-  31 - 60
-  61 - 120
-  121 - 14029





Top-down method - assumption

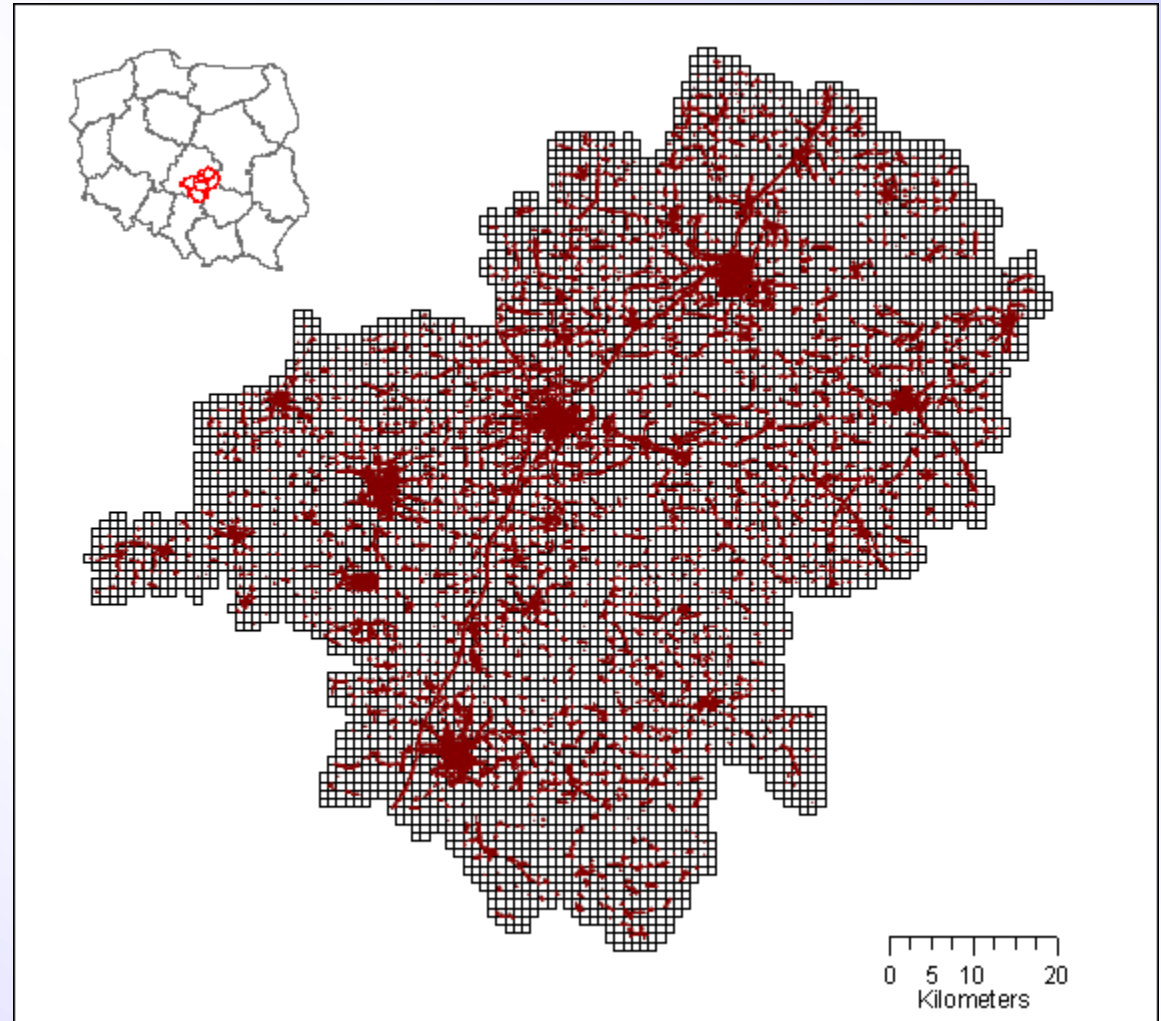
1. Using the smallest statistical regions in order to minimize error of the spatial distribution.

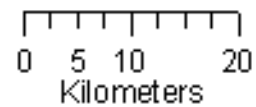
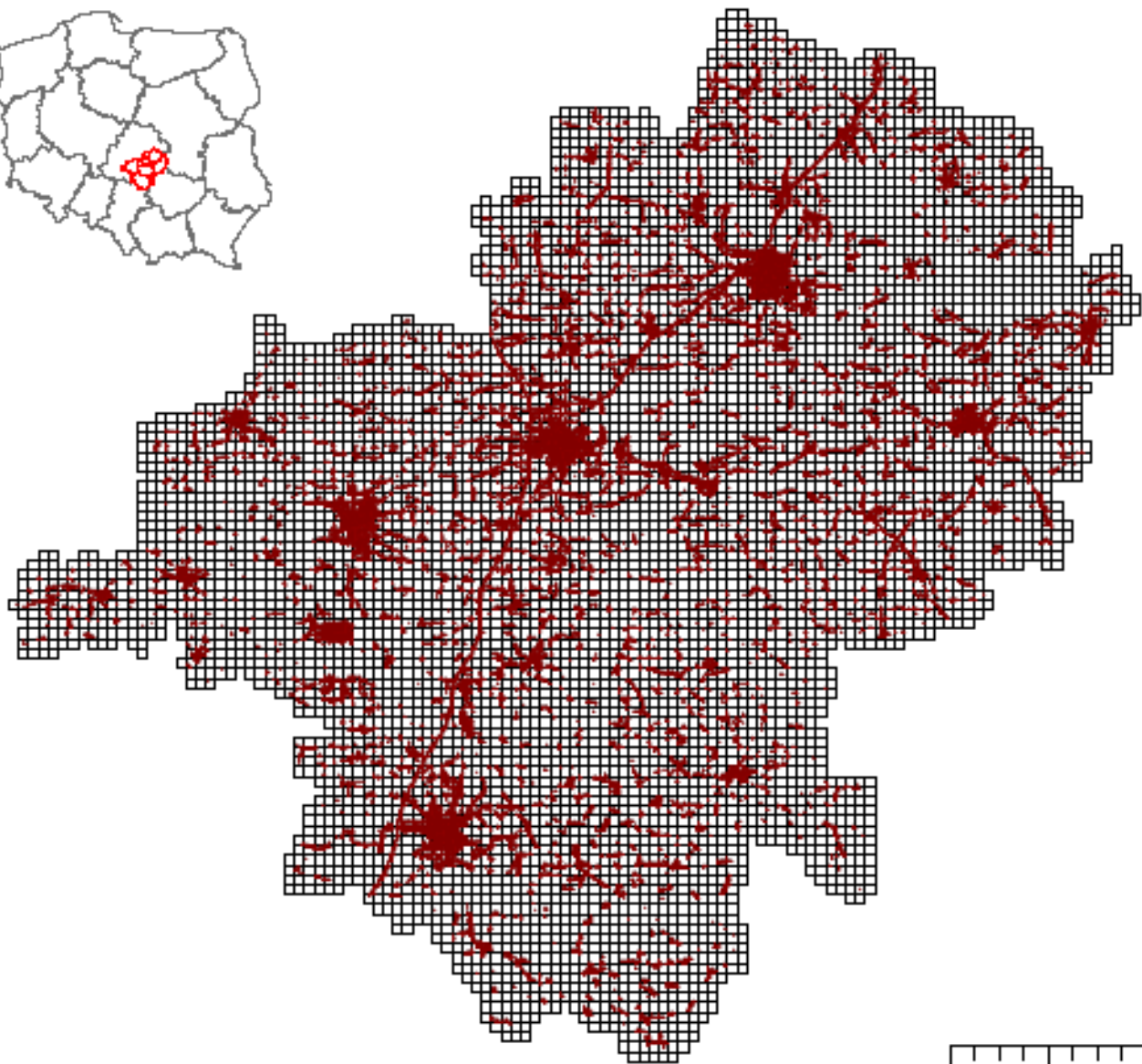




Top-down method - assumption

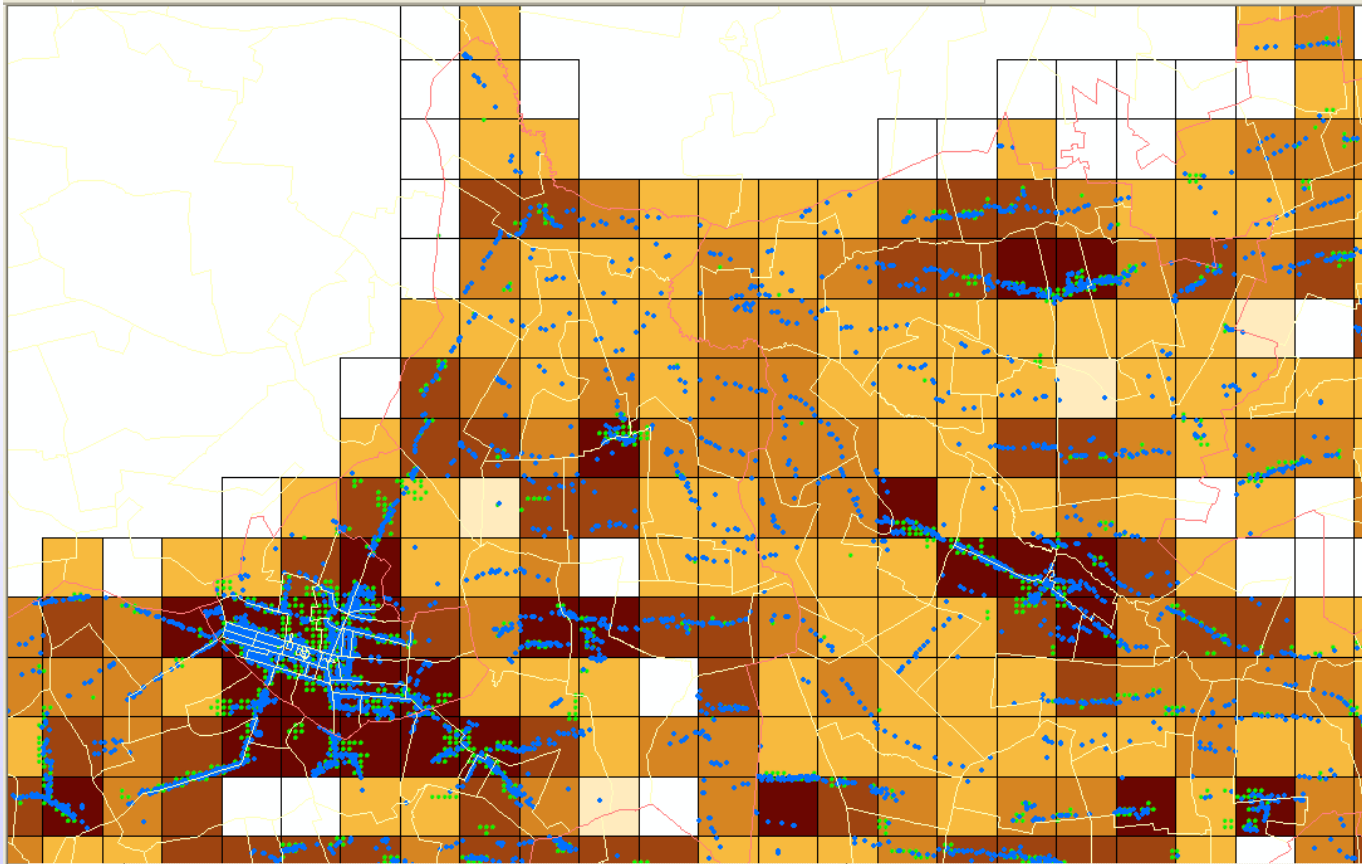
2. Using the SSL to increase the accuracy of address points distribution.

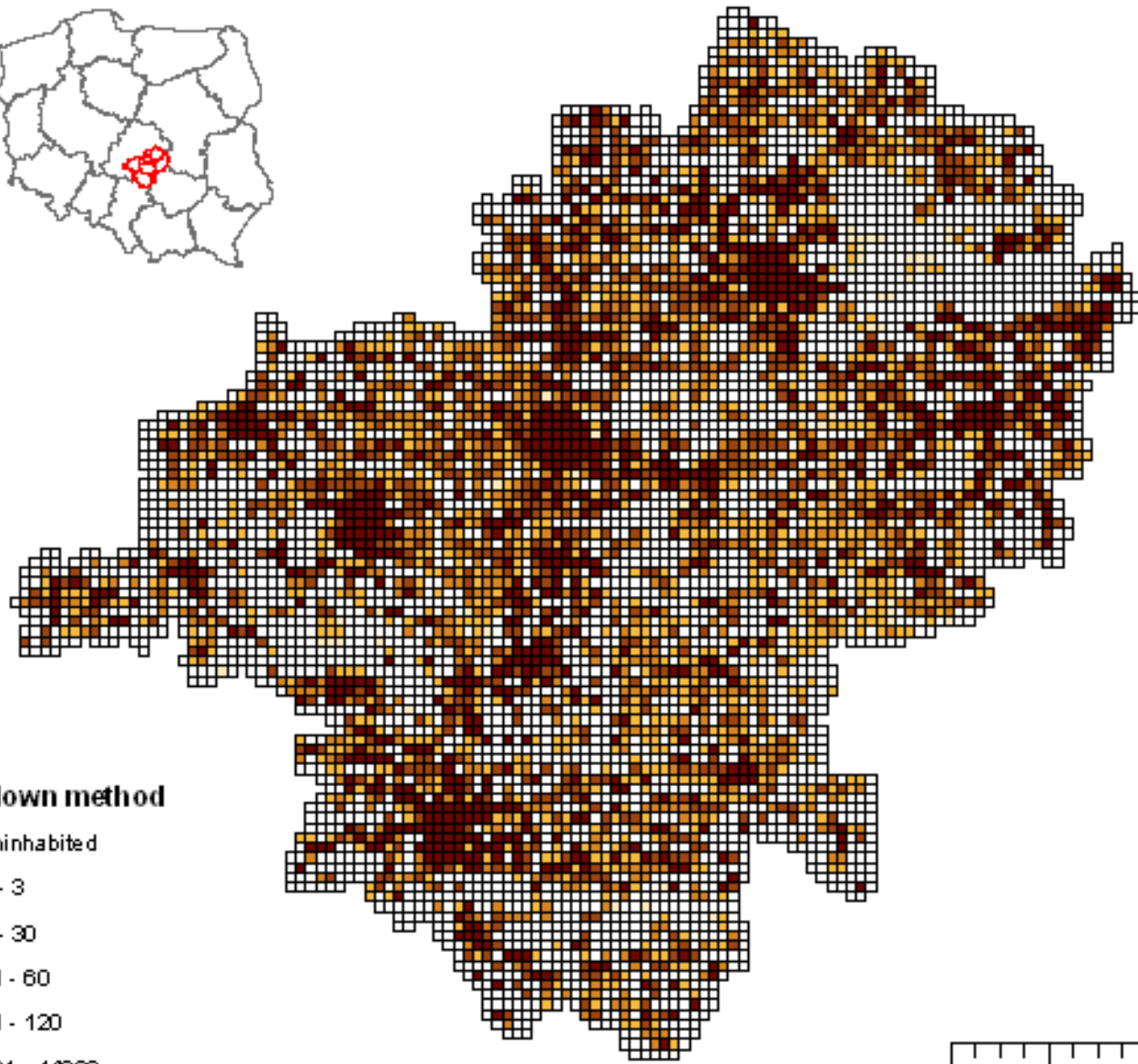












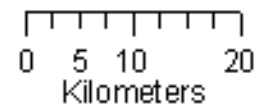
Top-down and bottom-up methods – in details





top-down method

-  uninhabited
-  1 - 3
-  4 - 30
-  31 - 60
-  61 - 120
-  121 - 14022





What can we do, in terms of Grid, to avoid ‘holes’ in the population?

- We can at best spread a population equally from over one or multiple districts, which are contain in a grid cel.
- SSL will not be treated as the only determinant of population assignment for grid but may be at best used as a „weight”.



Quality indicator - verifying the accuracy of the estimating population method.

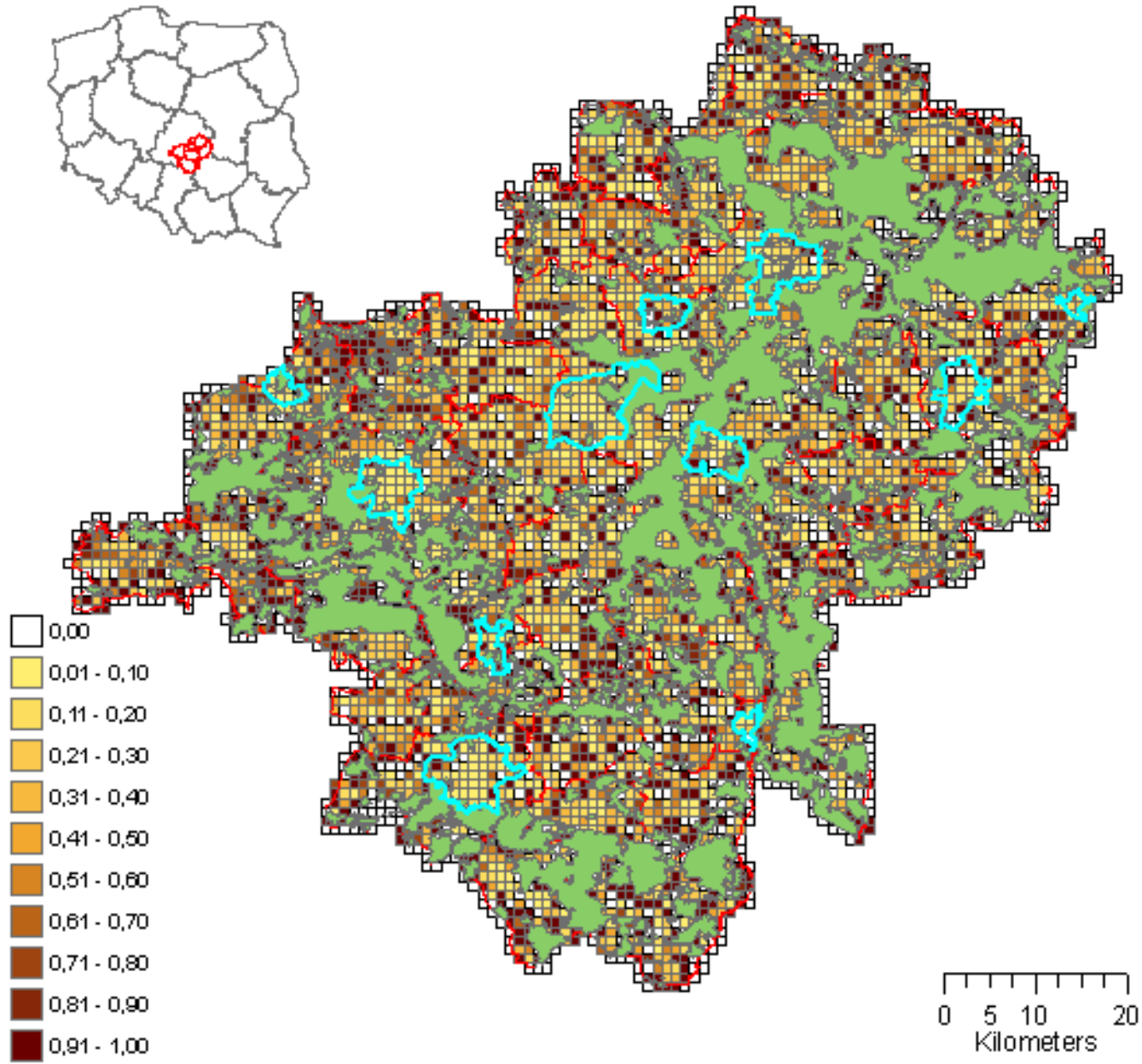
Having a set of reference data (the actual population from points of the address assigned for 1km grid squares) we can determine the error of tested top-down method.

$$\text{quality indicator} = \text{absolute value}(A_n - B_n) / \max(A_n, B_n)$$

where:

A_n - the value of the population in the n-th grid cell reference set of data A (bottom-up method)

B_n - the value of the population in the n-th grid cell with the values of the calculated B (top-down method)





Conclusions!

CLC data can be useful as a main factor in determining less populated areas. This factor helps you reduce noise in: rural communes and rural sections in a rural-urban commune.

In this way will allow us to achieve better results in the calculation of the estimated values.



Thank you for your attention.

Radoslaw Jablonski

Central Statistical Office of Poland

RJablonski@stat.gov.pl