# Improved method for population disaggregation based on European land monitoring services

**Konštantín Rosina /** Slovak Academy of Sciences, Institute of Geography, Slovakia
**Pavol Hurbánek** / Catholic University in Ruzomberok, Slovakia

**EFGS Conference 2013**

**23 – 25 October 2012, Sofia, Bulgaria**

# Motivation: why to study disaggregation methods?

(if bottom-up methods are being widely implemented)

- They allow for the production of:
  - International to global  grids (e.g. LandScan)
  - National grids in countries where bottom-up approaches are not applied yet
  - Retrospective grids for years when bottom-up approaches have not yet been applied
  - Grids of alternative temporal population density conceptualizations (ambient, daytime)
- Further downscaling of coarse resolution bottom-up data (e.g. 1km to 100m)
- No confidentiality issues

# Disaggregation methods

- Two main categories:
  - Simple – no ancillary spatial data
  - Dasymetric – ancillary spatial data are used
    - Mapping agencies data (buildings, roads, DTM)
    - Open source data (e.g. OpenStreetMap - OSM)
    - In situ sampled data (e.g. LUCAS)
    - Earth observation imagery (incl. night time lights)
    - Land use/cover maps
      - CORINE Land Cover (CLC)
      - Imperviousness Layer (IL)

} Free of charge
EU-wide coverage
consistent accuracy

# Limitations of top-down approach

- Resulting values are “only” estimates and are always subject to error
- Appropriate ancillary data for disaggregation are not always available
- ”Ground truth” data needed for model calibration and/or accuracy assessment are not always available
- Most disaggregated datasets deal only with population density, other demographic/social variables are more challenging
- Accuracy of the estimate largely depends on the size of the source zones (which can vary substantially across the study area – inherent spatial heterogeneity of accuracy)

# Dasymetric disaggregation methods of the available EU grids

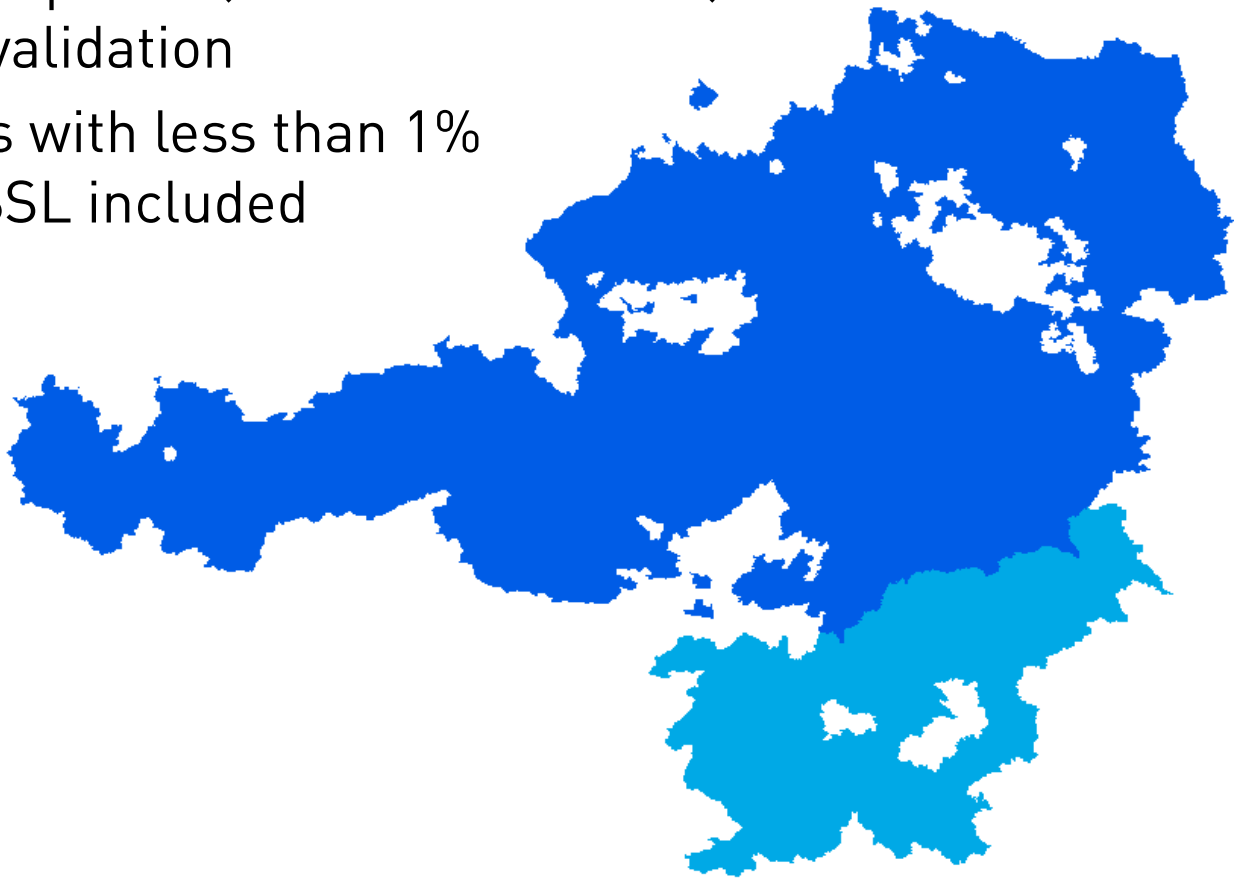
- JRC grid 2000, 100m pixel (F. J. Gallego)
  - CLC iterative
  - CLC-LUCAS simple
  - CLC-LUCAS logit regression
  - CLC expectation-maximization
  - CLC limiting variable
- JRC grid 2006, 100m pixel (F. Batista et al.)
  - CLC2006 refined by IL, SRTM water bodies, UrbanAtlas and TeleAtlas (commercial data)
- AIT grid 2006, 1km pixel (K. Steinocher et al.)
  - CLC-IL-OSM linear disaggregation

# Objectives

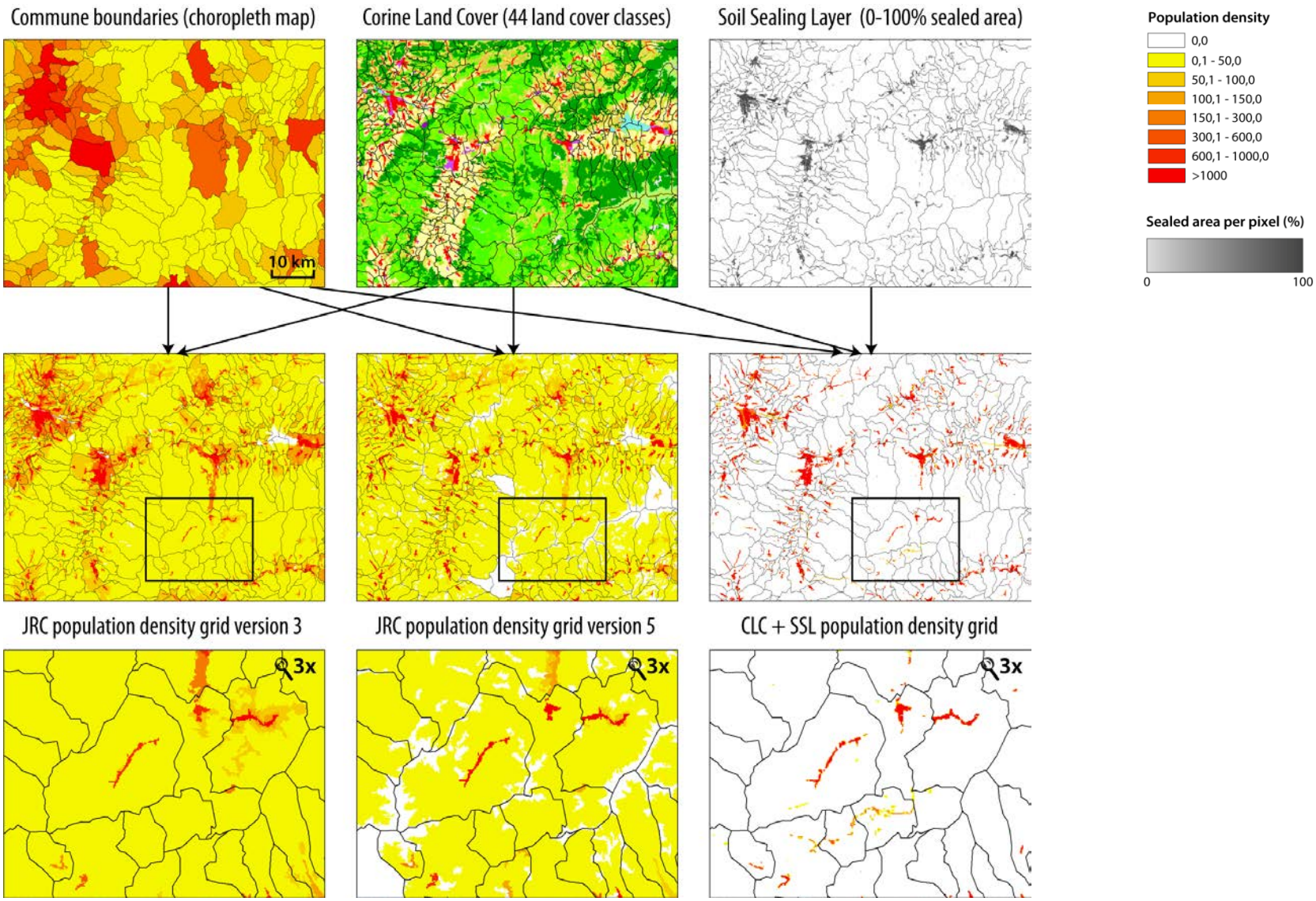
- Disaggregation of 2006 population data from commune level into 100m grid
- Improve the Gallego's iterative method in order to produce better estimates than the ones currently available (comparison possible with AIT grid 2006 at 1km and JRC grid 2006 at 100m or coarser)
- How to increase accuracy of the population estimates:
  - Input data quality is more important than algorithmic detail (Martin, Tate &Langford, 2000)
  - Increase quality of ancillary data or employ multiple ancillary datasets
  - Improve the method's algorithm

# Study area and validation data

- Austria – 1km bottom-up data available (GEOSTAT 1A grid)
- Slovenia – 1km bottom-up data available (GEOSTAT 1A grid) + 100m bottom-up data (Statistics Slovenia) => multi-scale validation
- Only communes with less than 1% cloud cover in SSL included



# Improvement of ancillary data – IL incorporated



# Improvement of the method's algorithm

- Population density coefficients for land cover classes are tuned iteratively in two repeated steps:
  - estimation from a group (subset) of communes to individual communes
  - modification of the coefficient values to reduce the overall error of the estimate (direction and strength of correlation between deviation of the last estimate and relative dominance of a land cover class in a commune tells how to modify coefficient value of the respective land cover class)
- When the overall error becomes stable, the fine-tuned coefficients are used to estimate from commune level to target grid cells
- We suggest that meaningful design of the commune subsets can improve the accuracy

# Approaches to the design of commune subsets

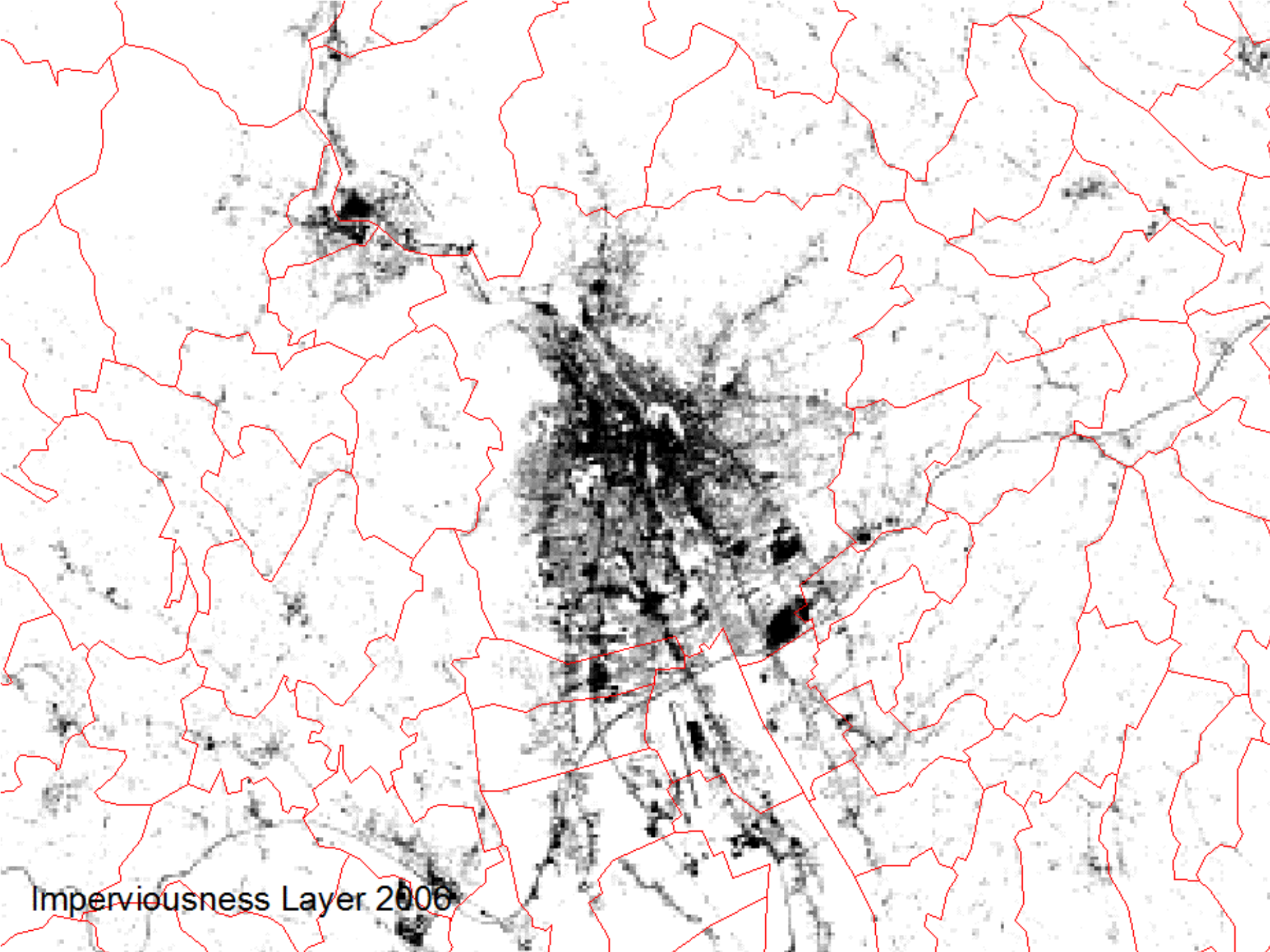
- Gallego, Peedell 2001: Conventional regions (e.g. NUTS2, spatially contiguous and non-overlapping subsets of communes)
- Gallego et al. 2011: Stratified conventional regions - spatially non-contiguous and non-overlapping subsets; stratification into densely, intermediate and sparsely populated communes (based on overall population density)

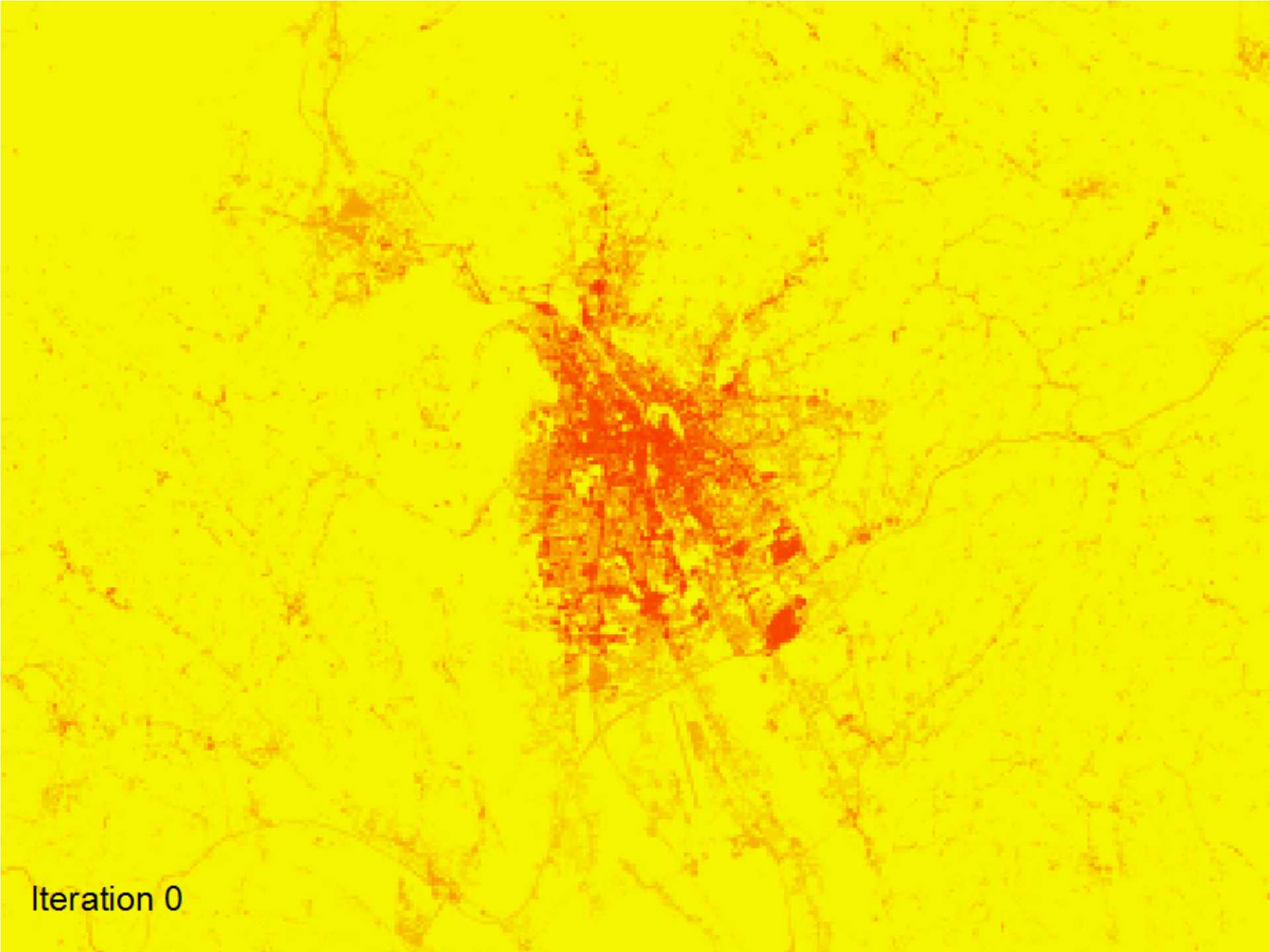
**Disadvandages :**

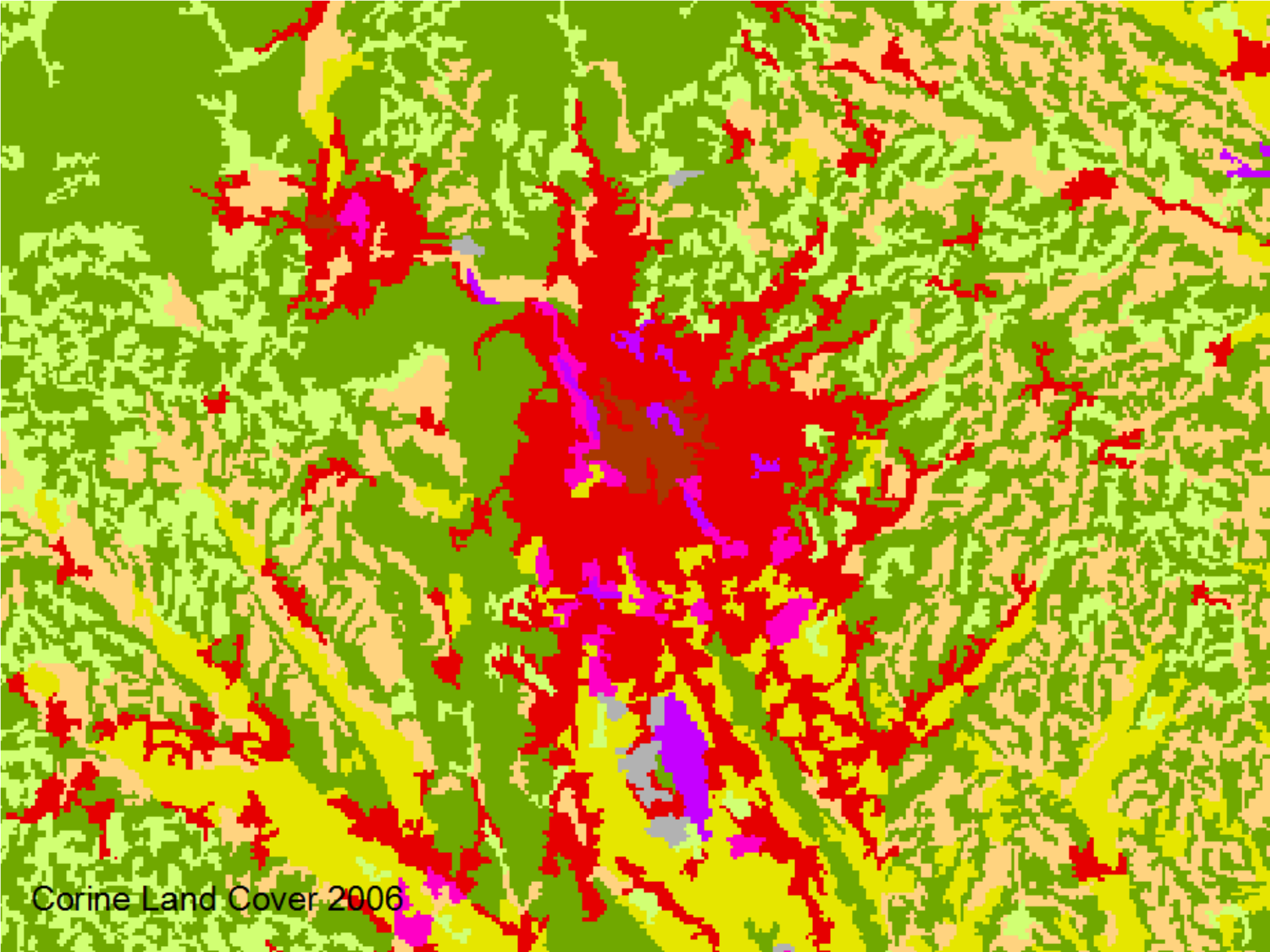
- Limitation of subsets by arbitrary NUTS boundaries
- Population density per commune area is partly function of commune boundaries (Modifiable Areal Unit Problem) – not the best variable for stratification
- All the communes in a region / stratum have the same set of pop. density coefficients (too restrictive assumption)

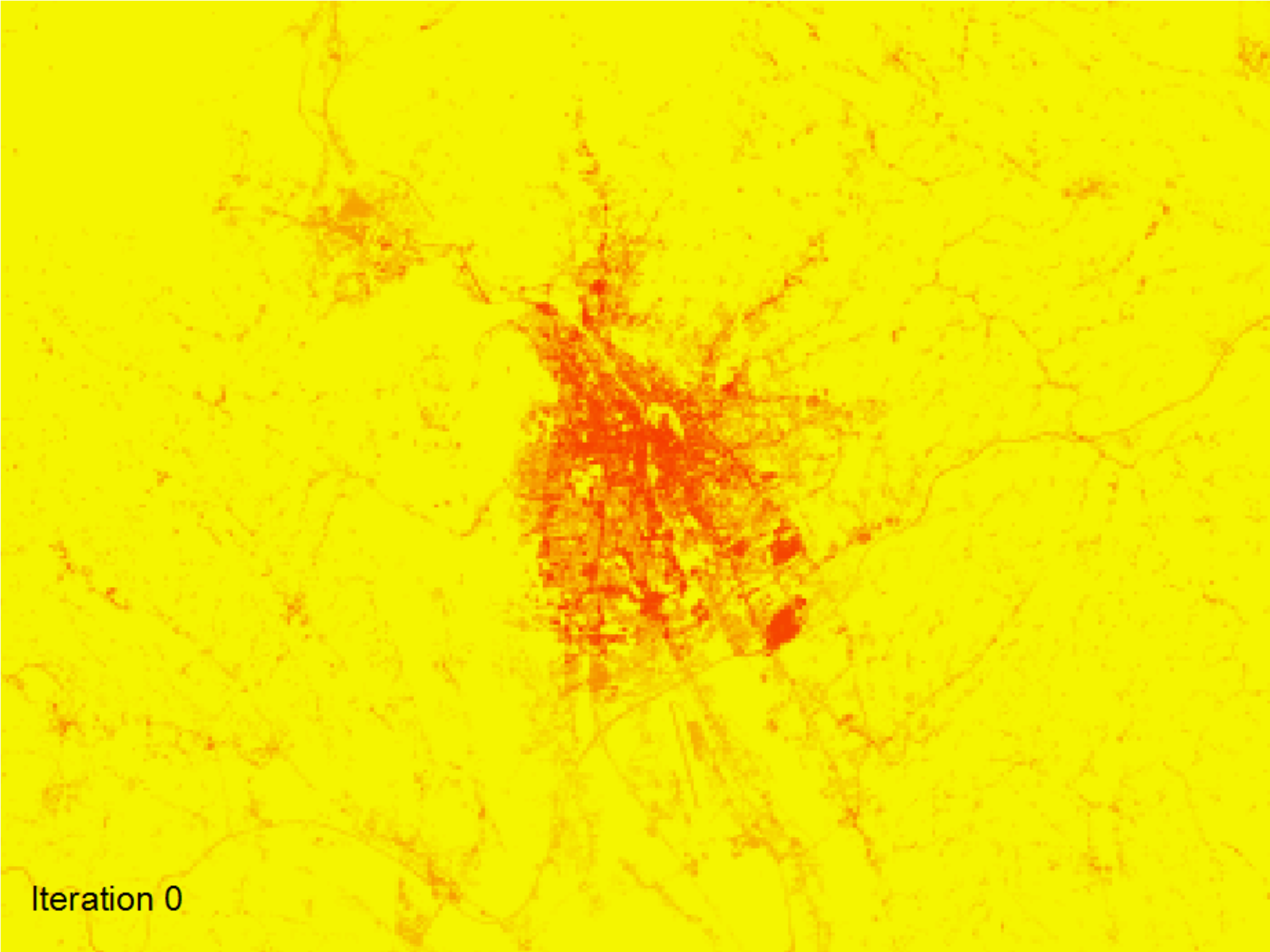
# New approach to design of the commune subsets

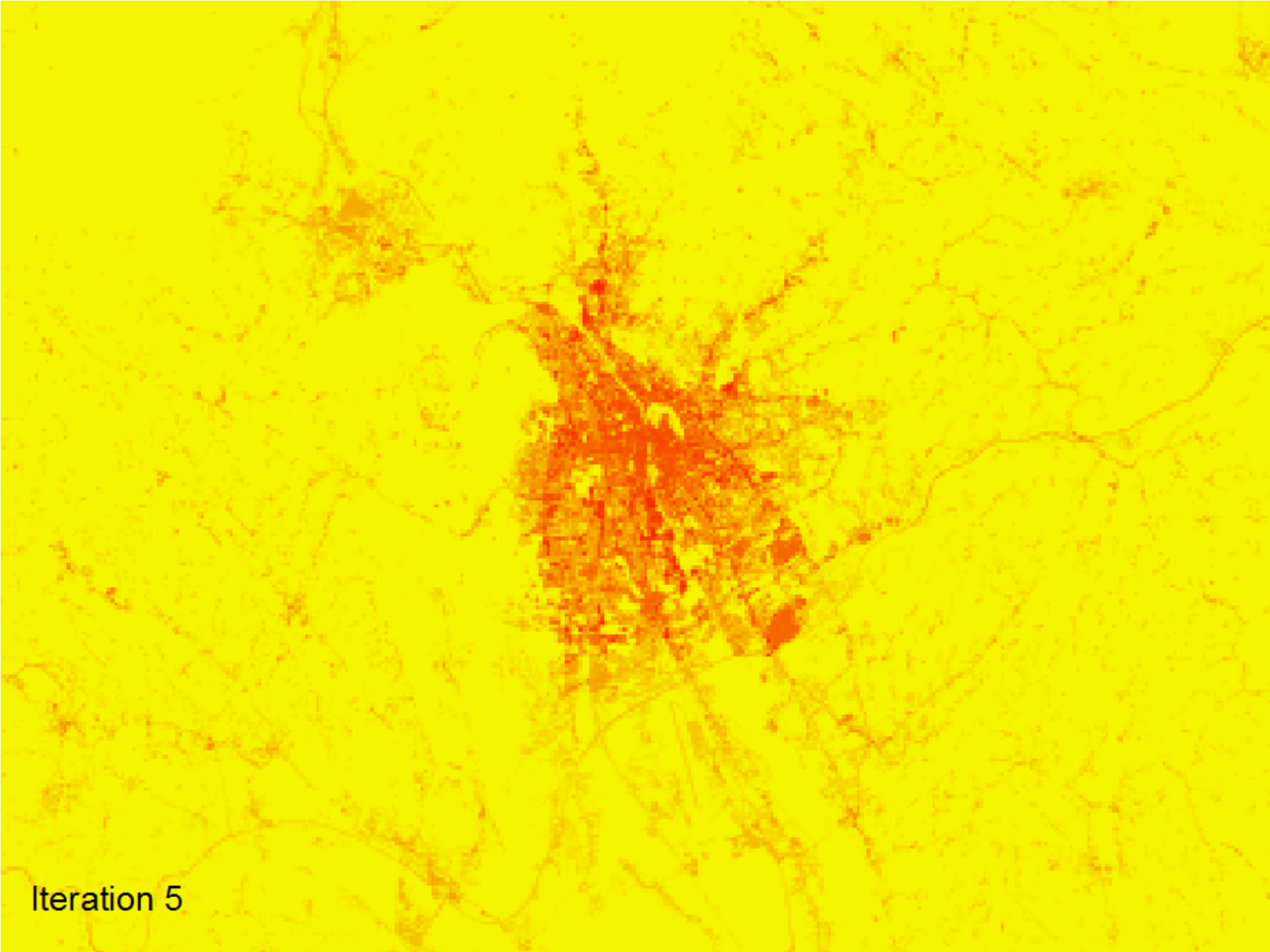
1. All the communes in the study area are in a pool of candidate communes for any subset
2. Each commune in the study area has its own unique subset and thus a unique set of population density coefficients
3. Each commune can be part of multiple subsets
4. Subsets can be non-contiguous and can overlap each other
5. A subset of a commune *c* is defined as the *m* nearest communes to the commune *c* in an *n*-dimensional space, where various commune characteristics obtained automatically from source and ancillary data can be used as a dimension (population size, population density per impervious area, share of impervious area, XY coordinates, share of individual CORINE classes etc.)

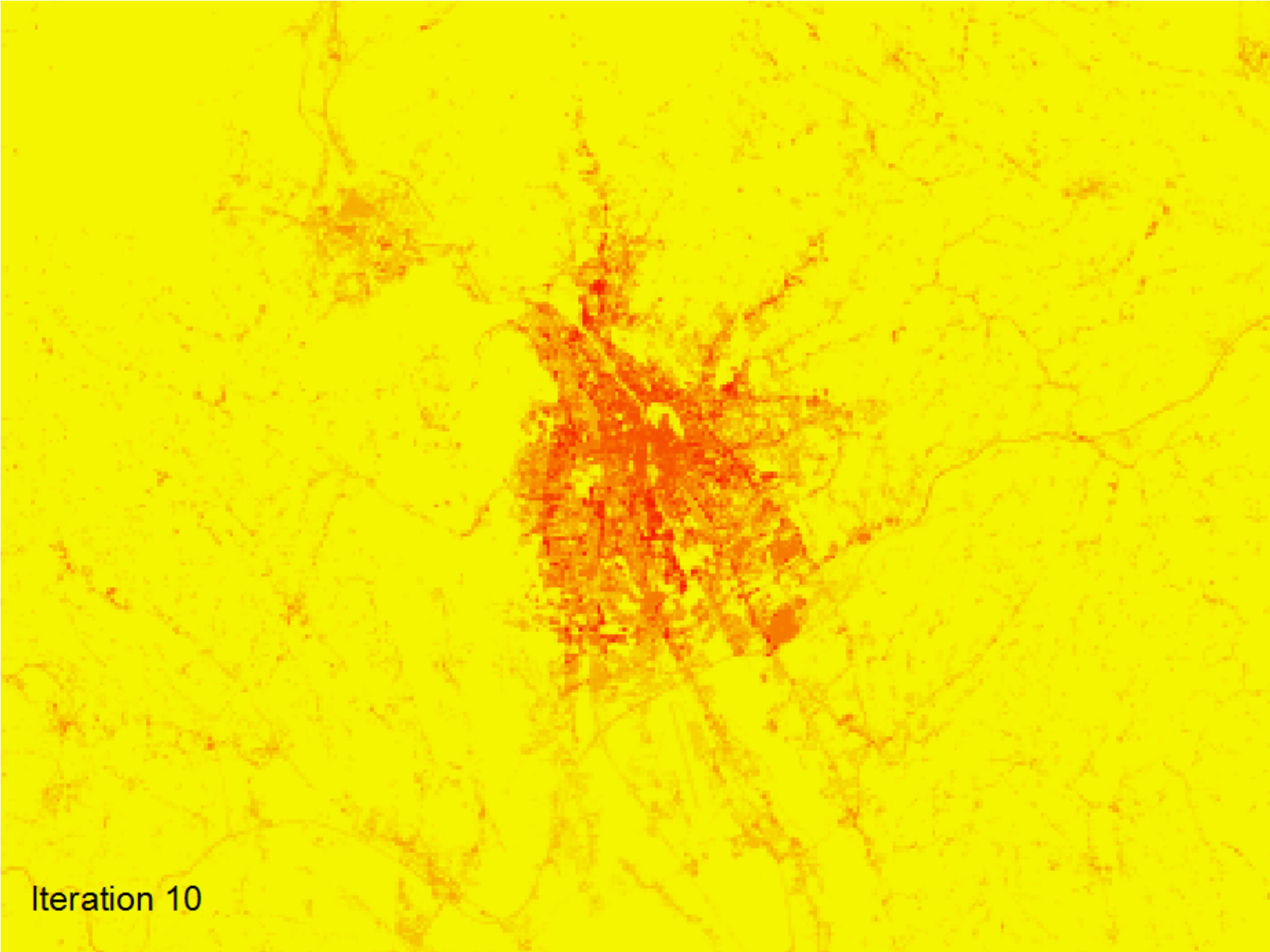

Imperviousness Layer 2006

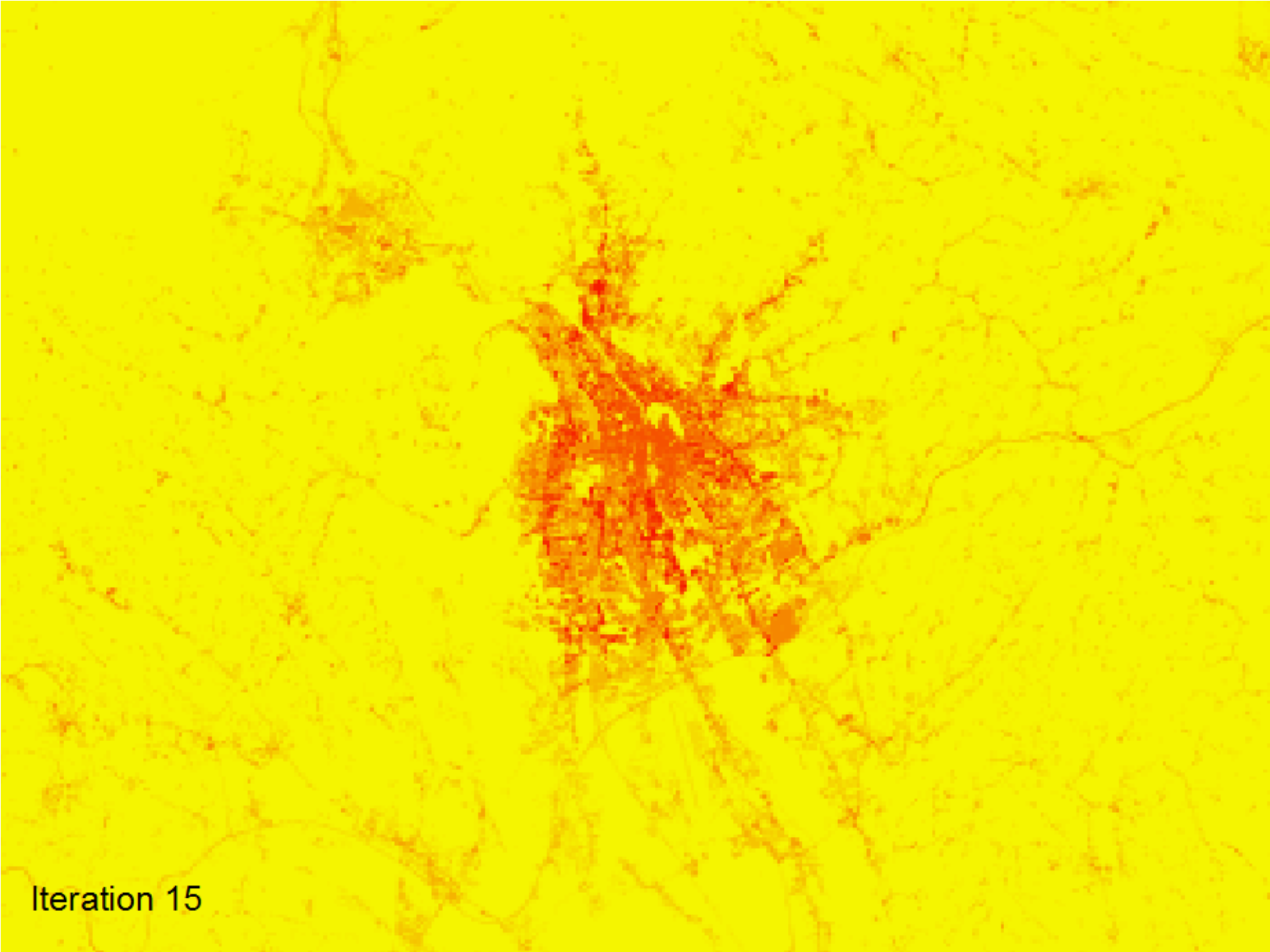

Iteration 0

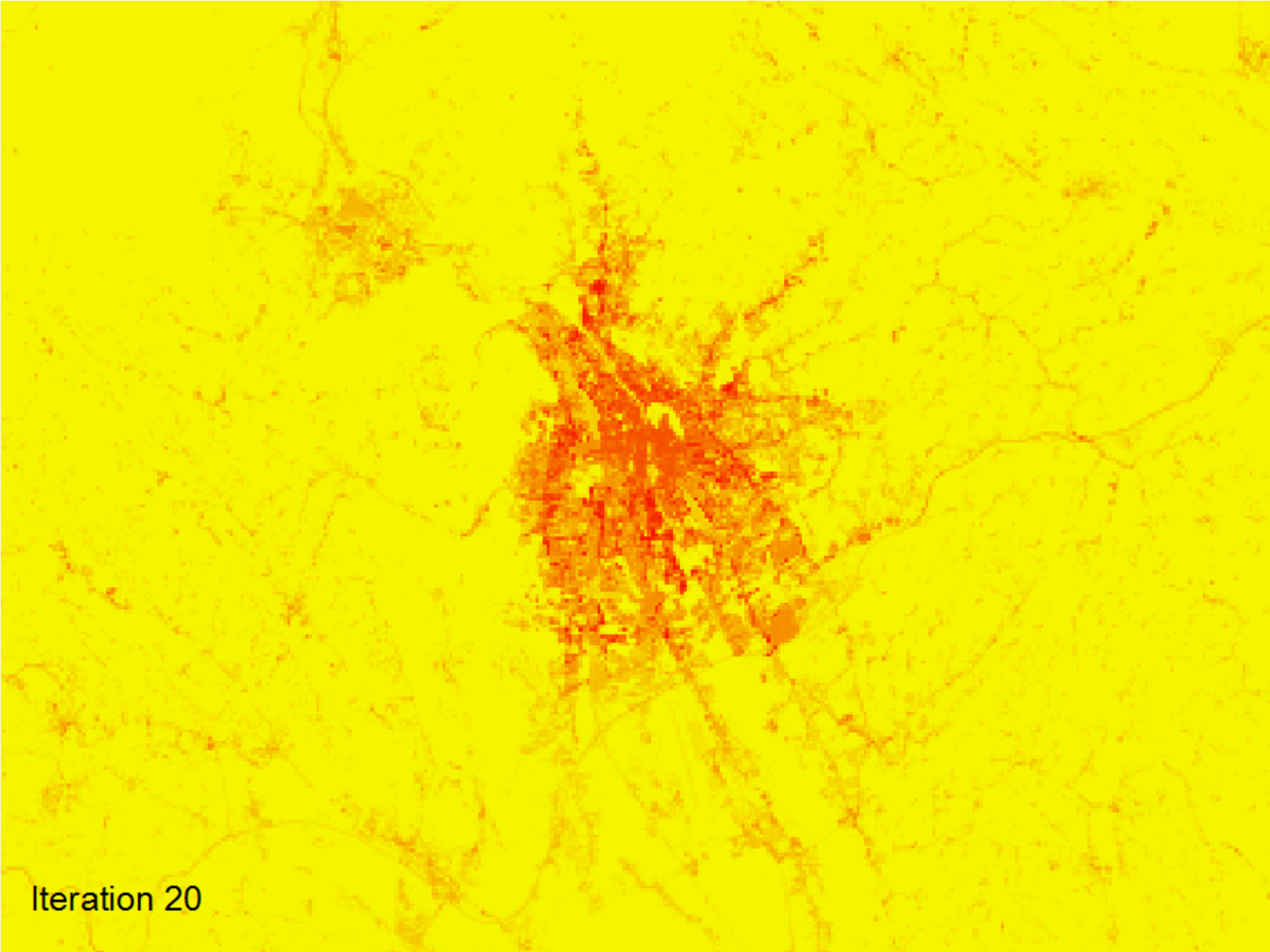

Corine Land Cover 2006

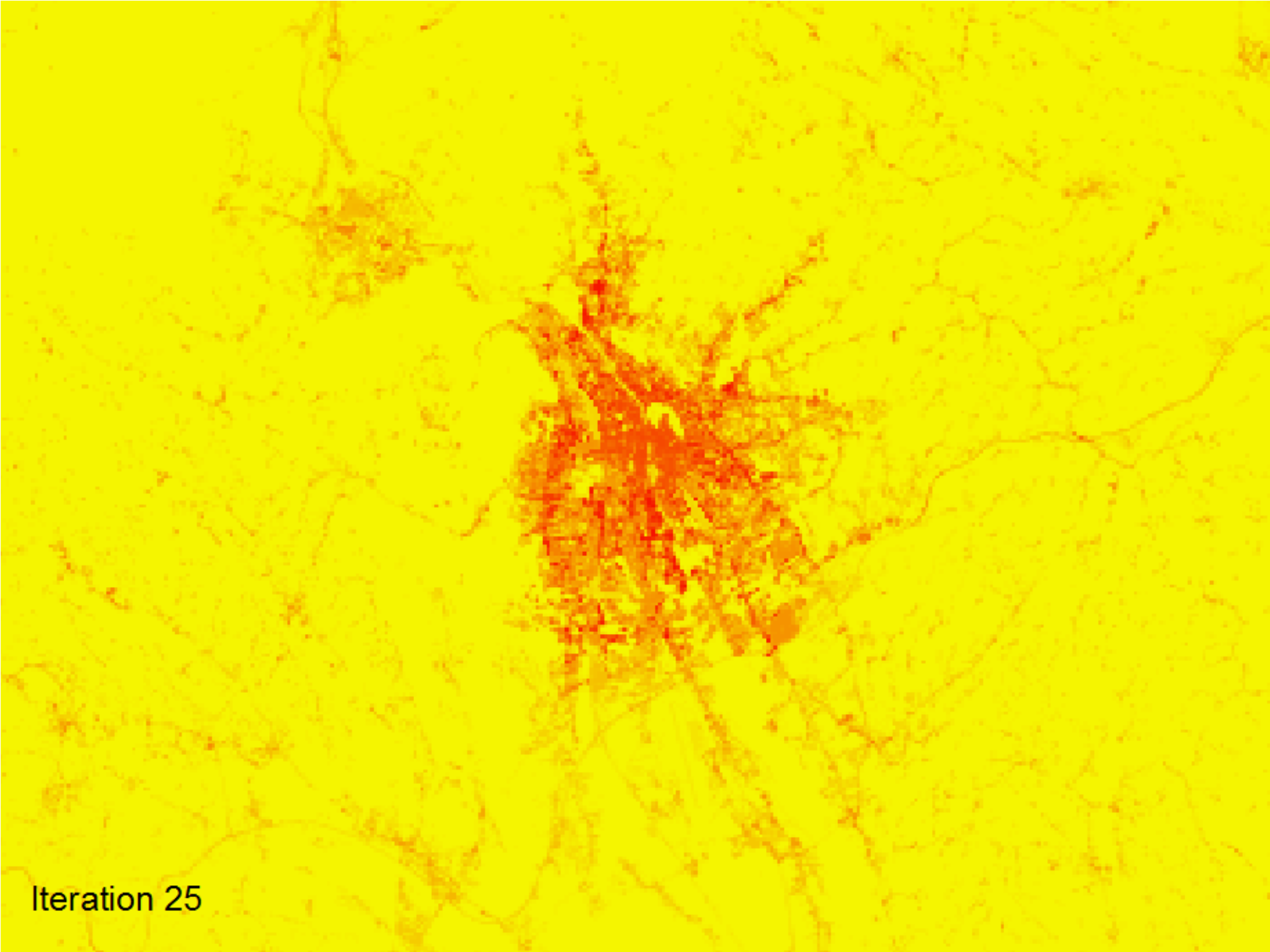

Iteration 0

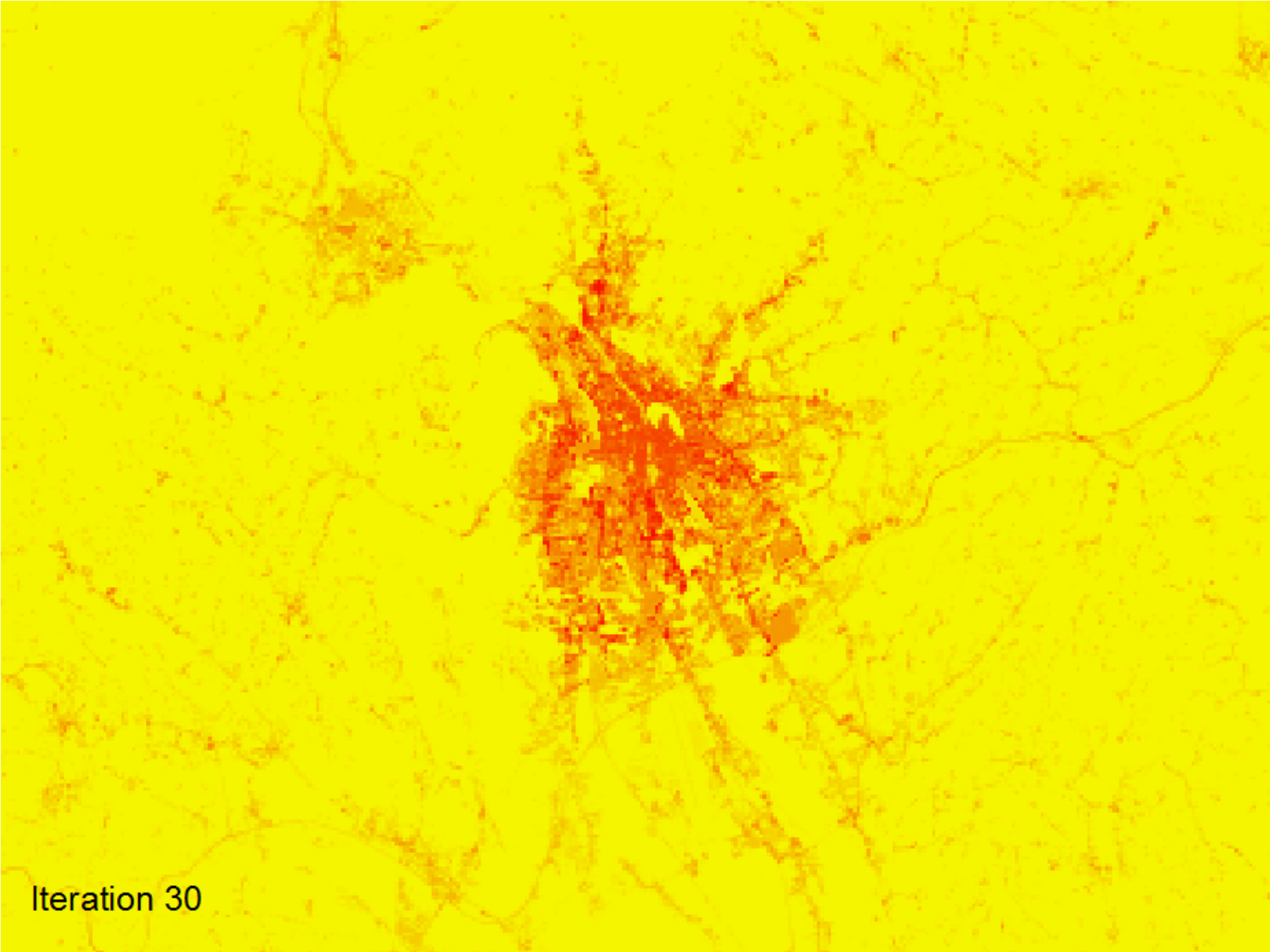

Iteration 5

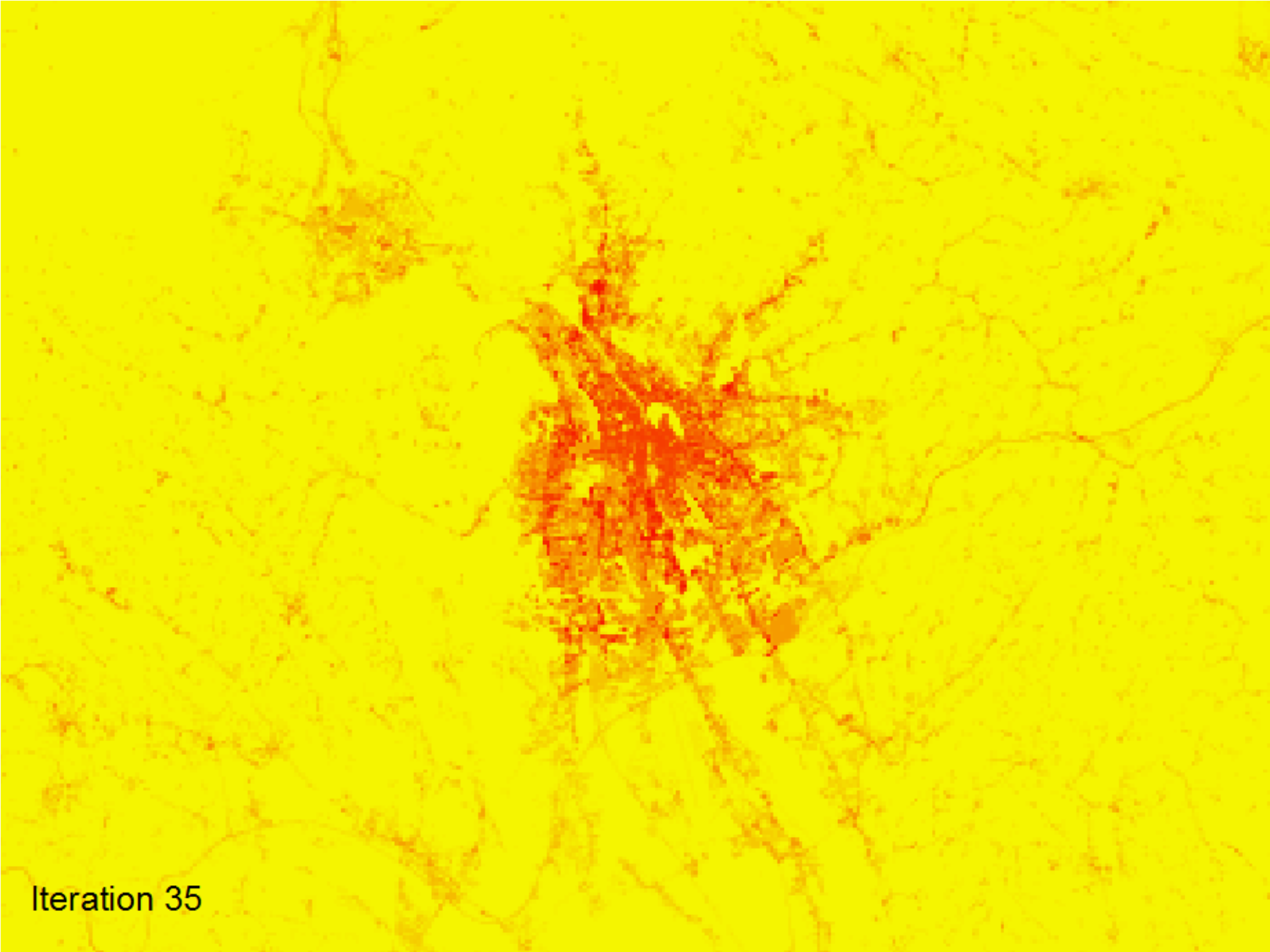

Iteration 10

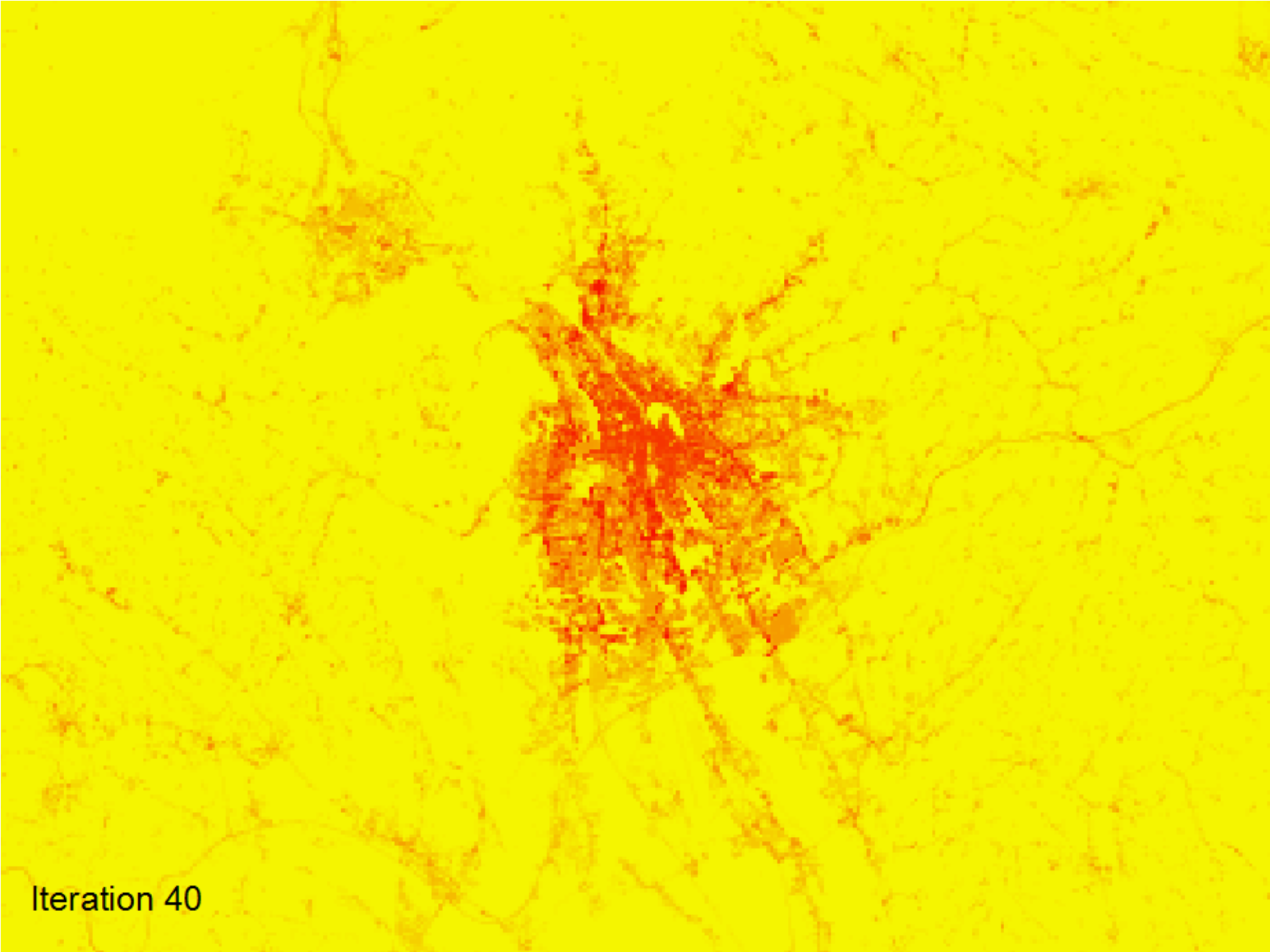

Iteration 15

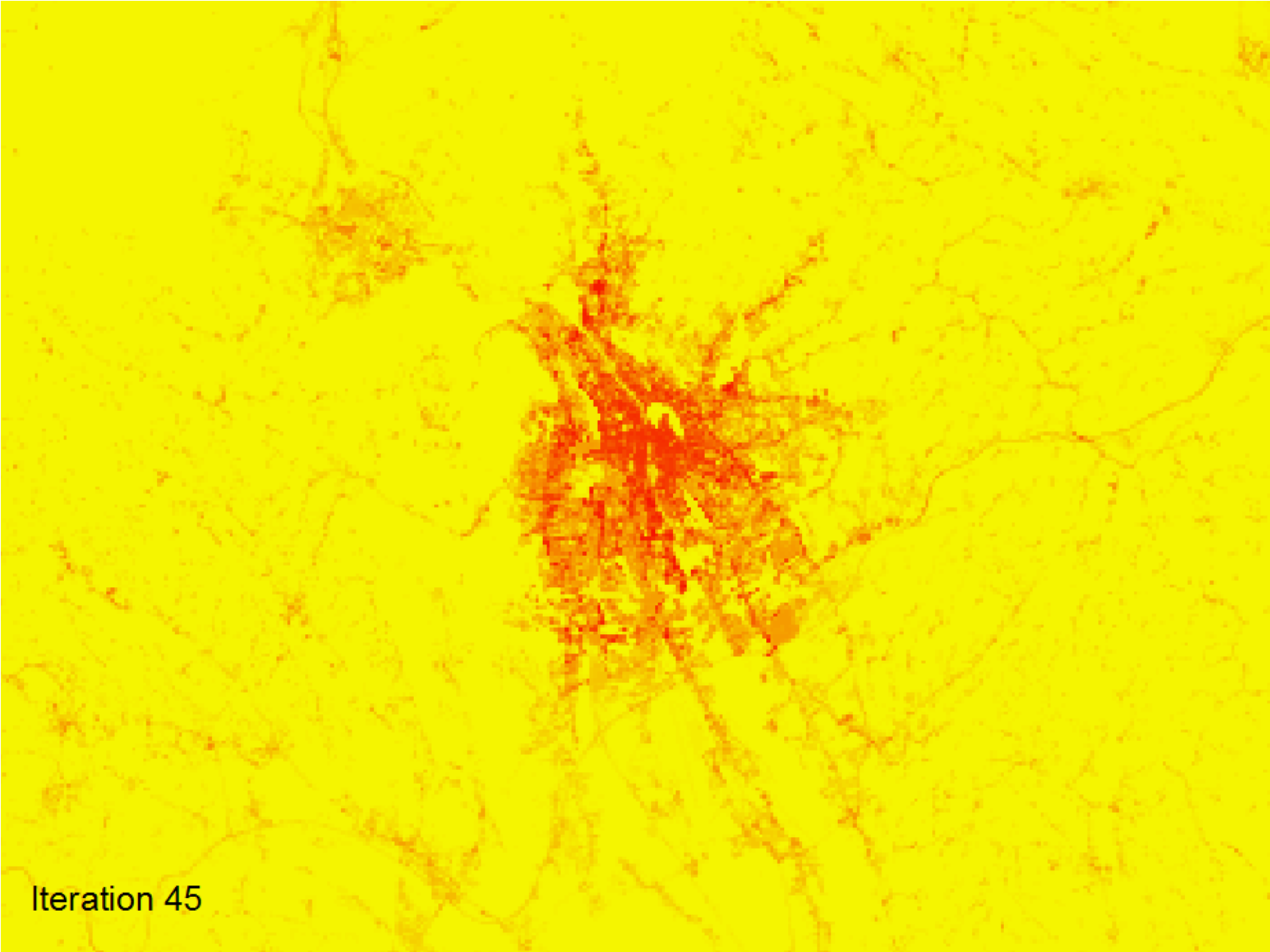

Iteration 20

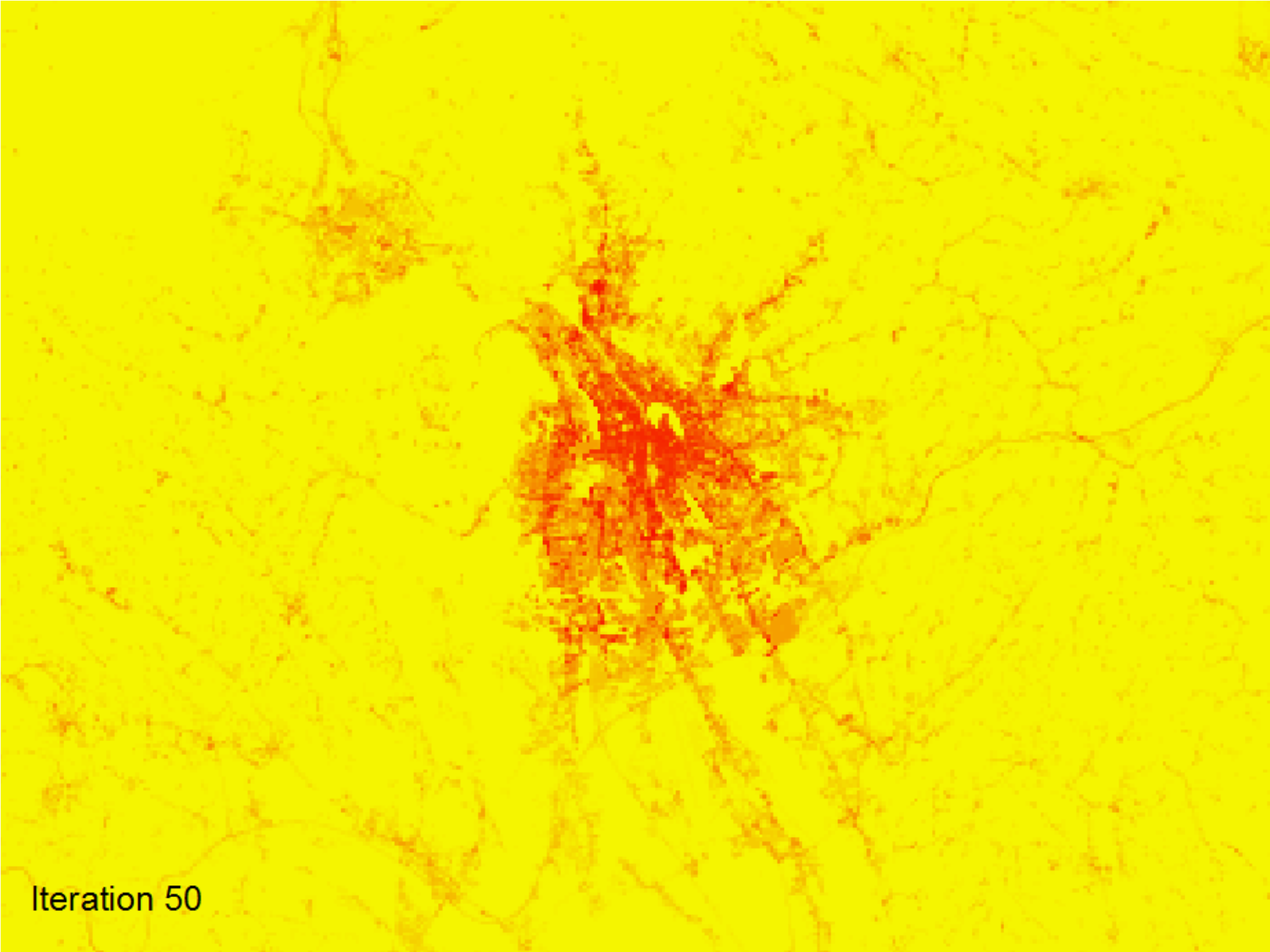

Iteration 25

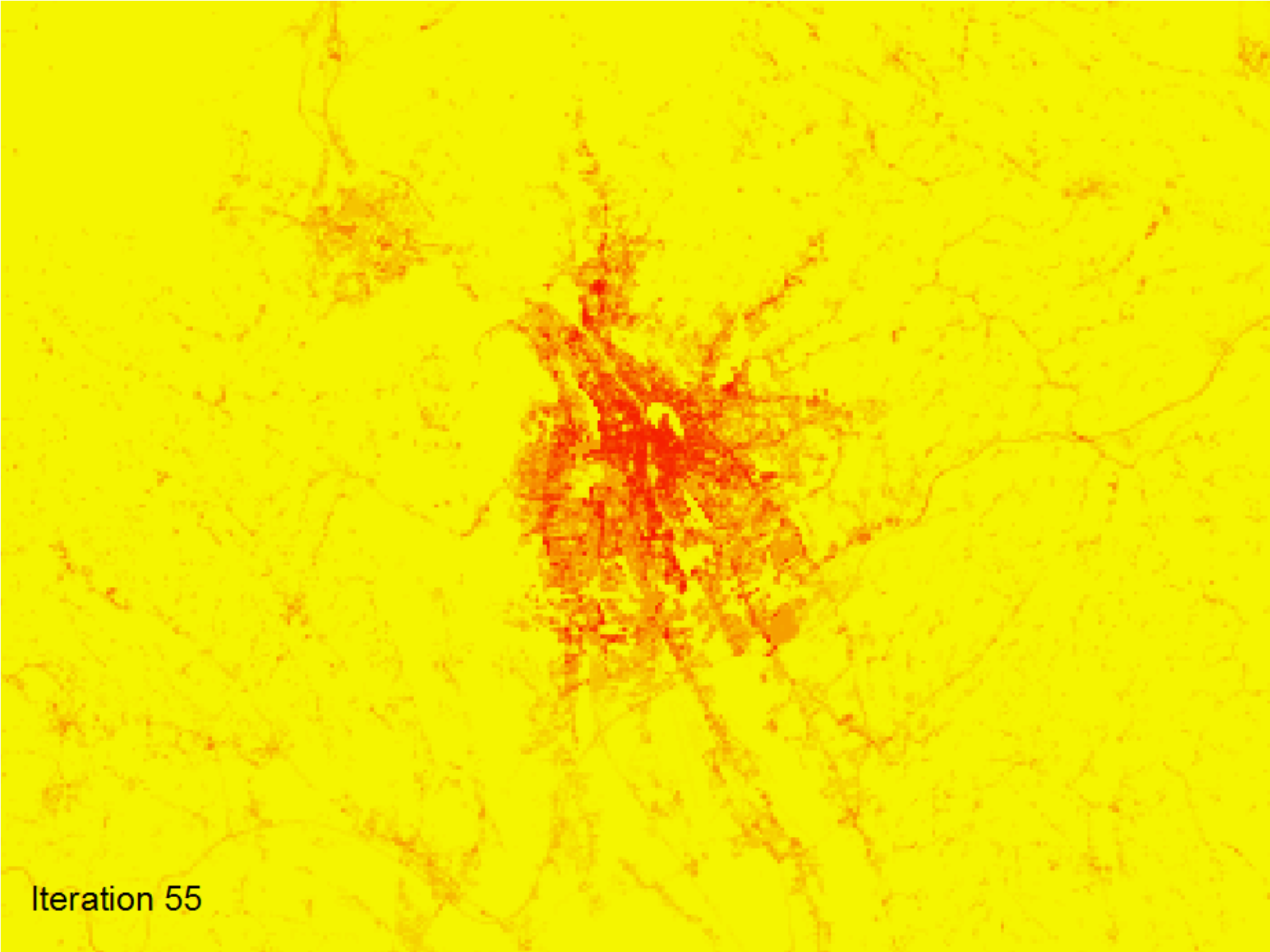

Iteration 30

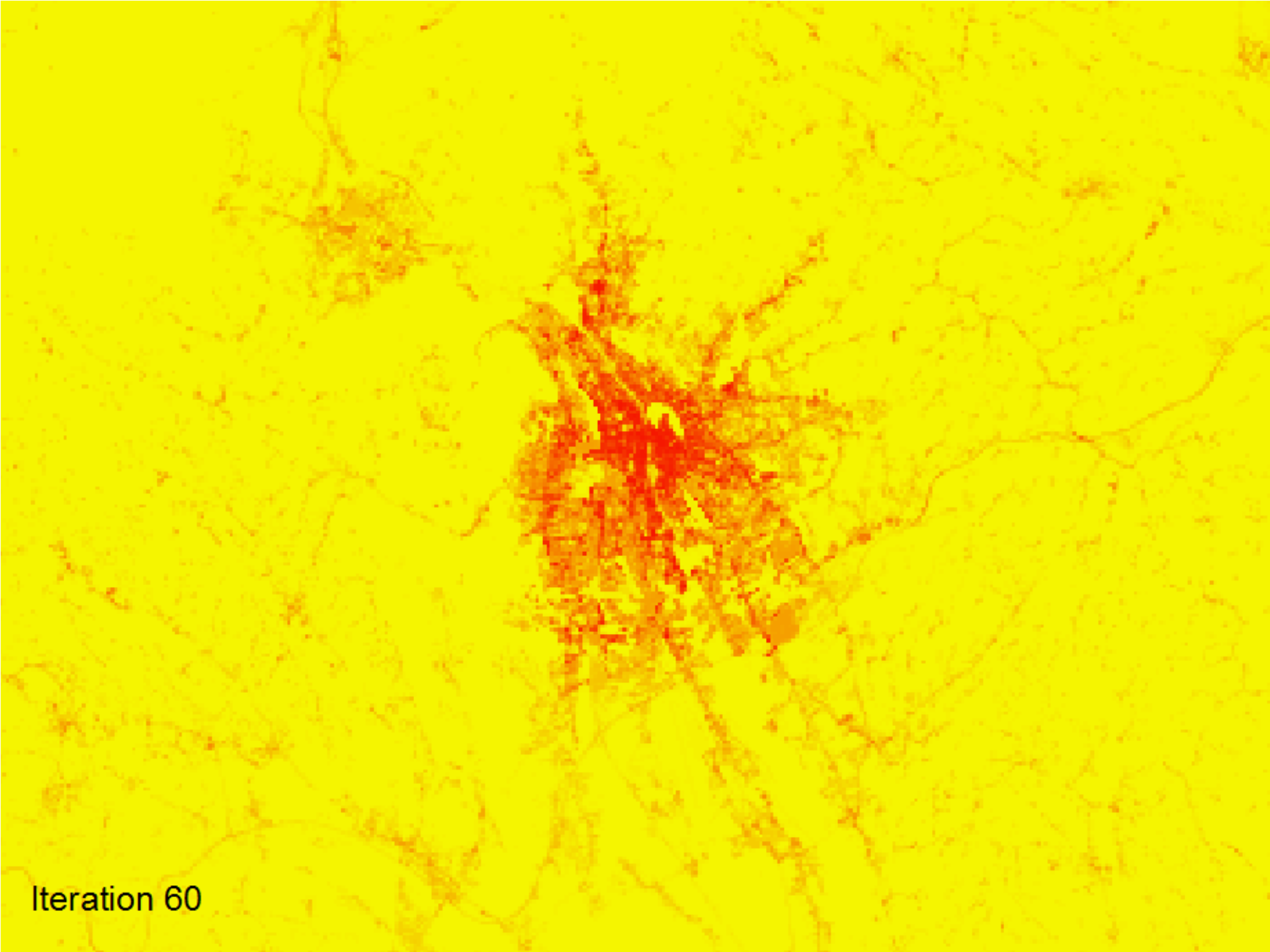

Iteration 35

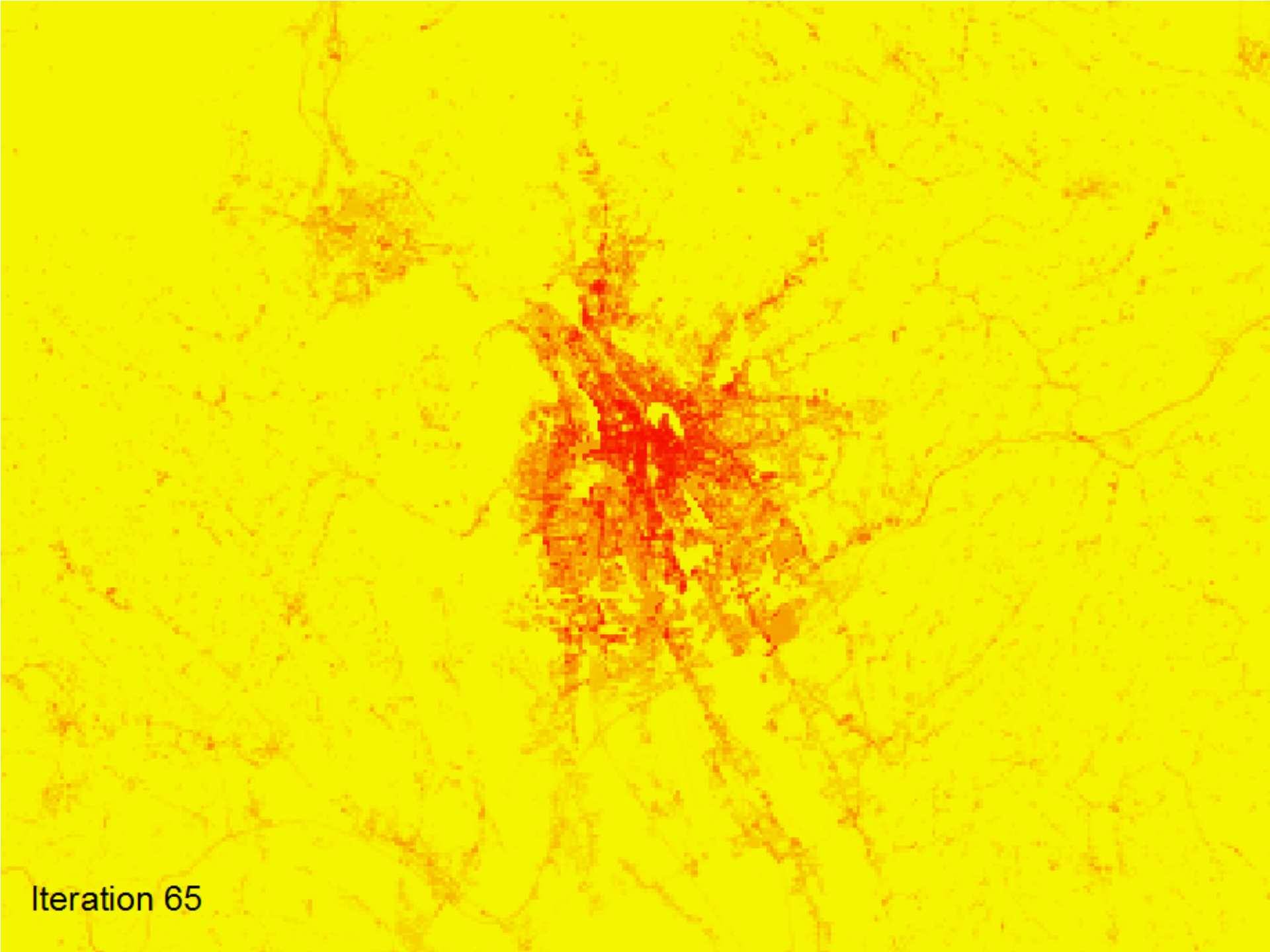

Iteration 40

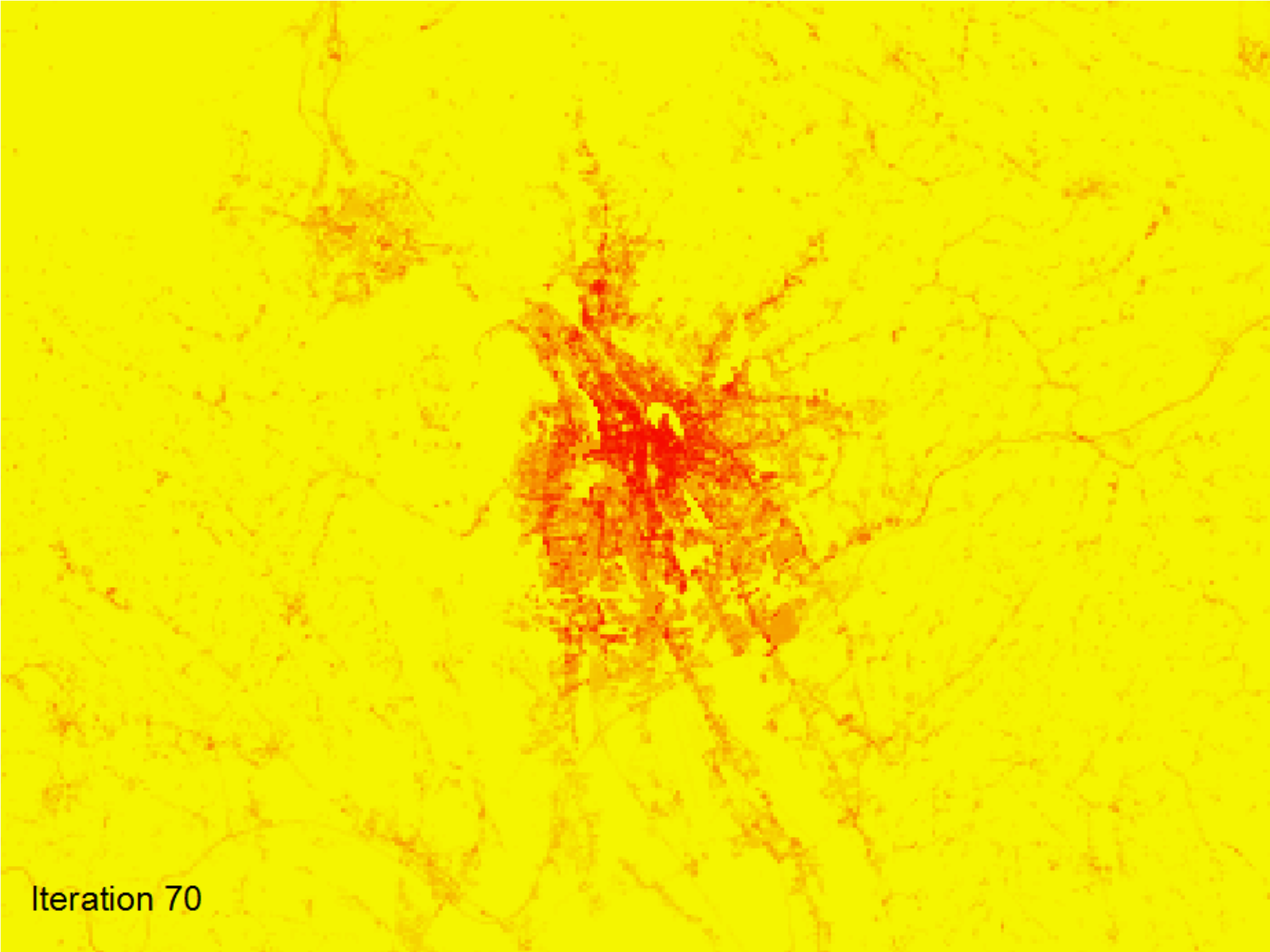

Iteration 45

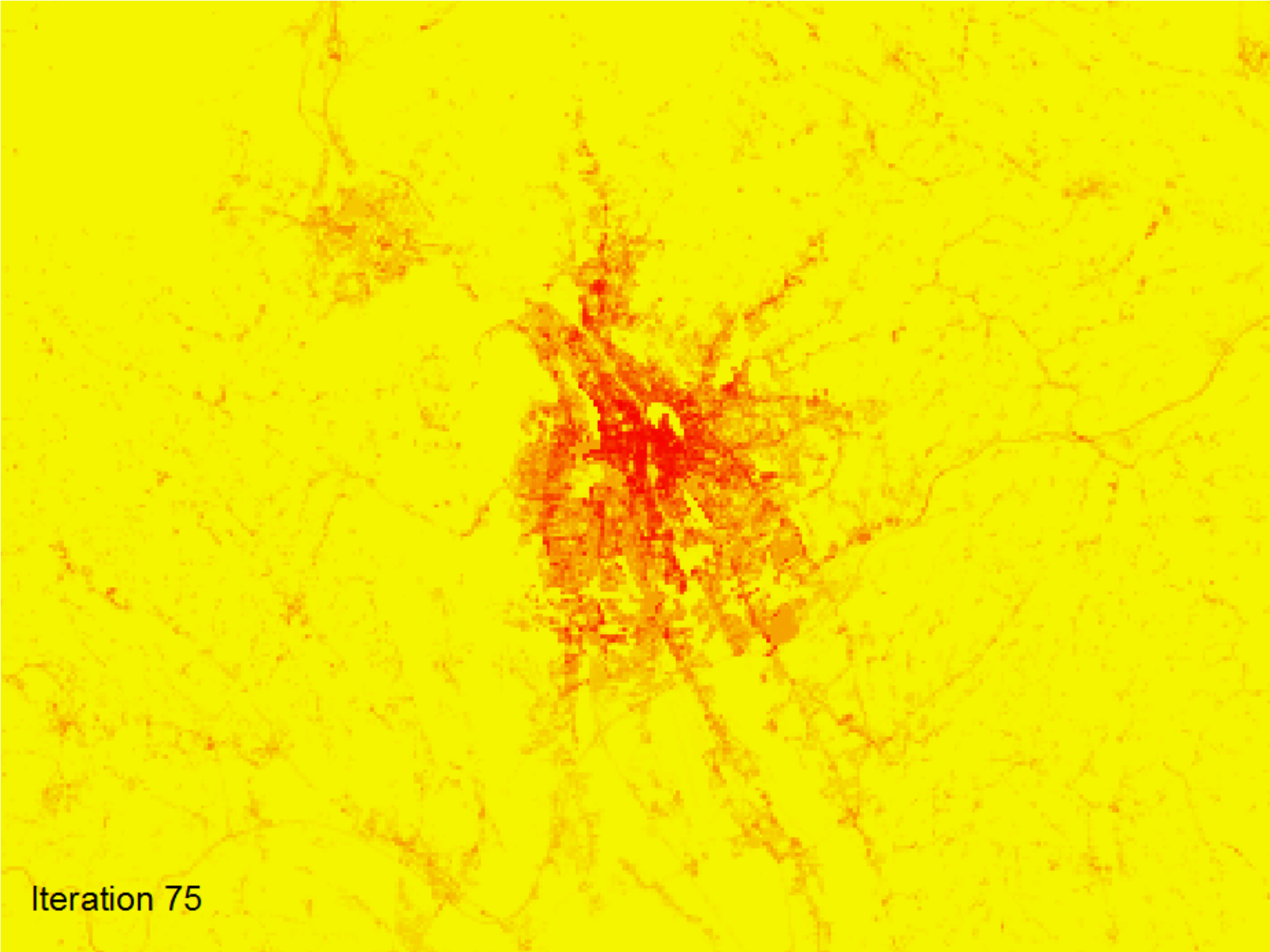

Iteration 50

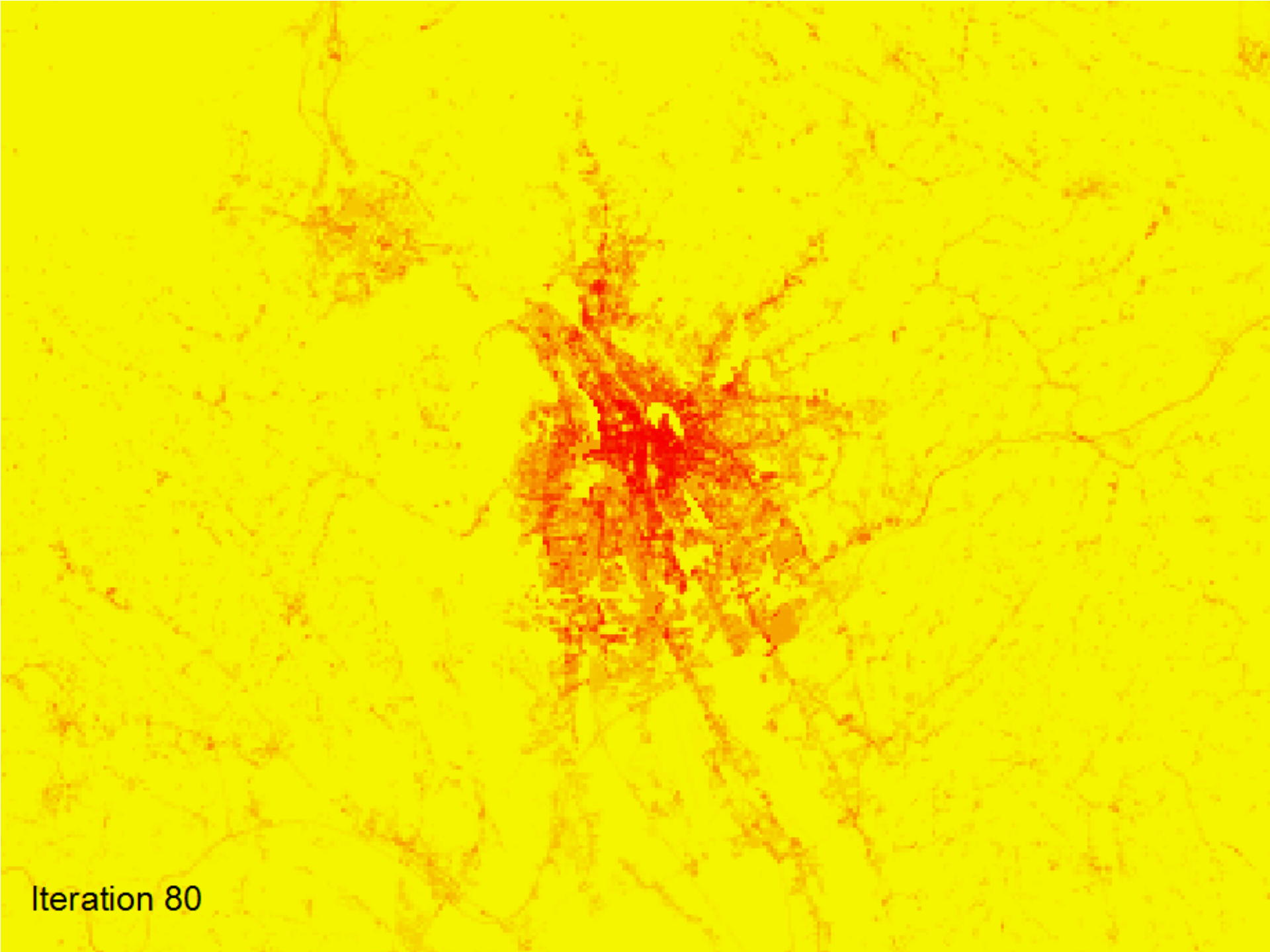

Iteration 55

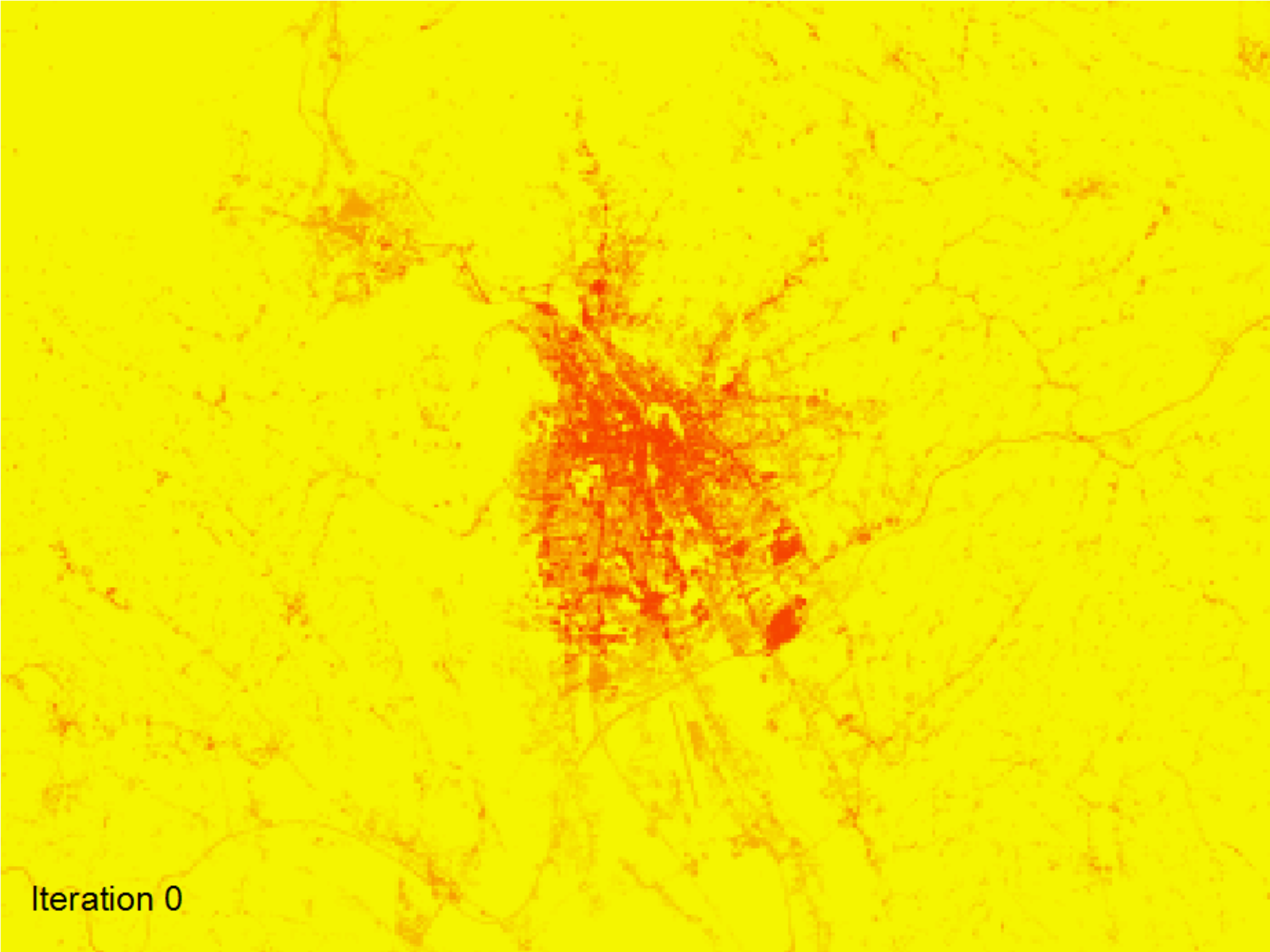

Iteration 60


Iteration 65


Iteration 70

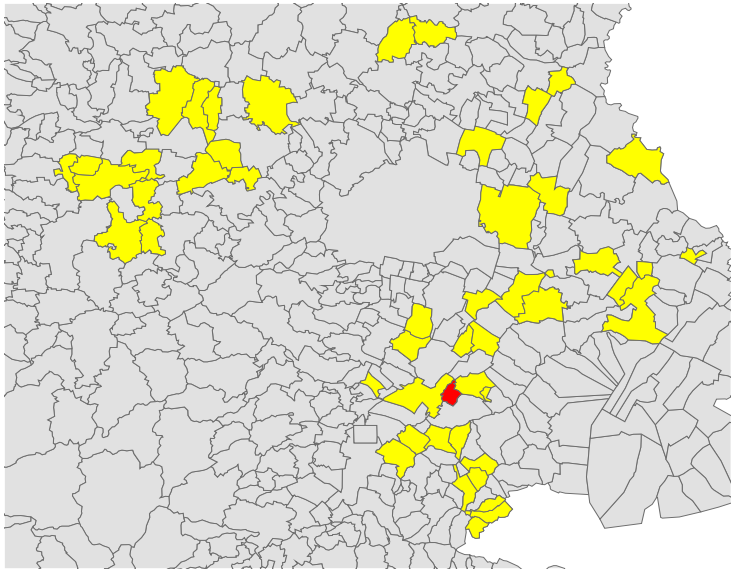

Iteration 75


Iteration 80


Iteration 0

# New questions raised

- What should be the size of the subsets (in terms of area and commune count)?
- What criteria should be used to cluster similar communes into subsets?
- Should be subset membership constrained by spatial distance?
- How does the extent of the study area (size of the pool of potential subset members) affect the accuracy of the estimate?
- How many iterations should be performed before the coefficients are applied



# Comparison of relative total absolute error

- Austria at 1km

| | |
|---|---|
| AIT Grid 2006 | 37,8% |
| JRC Grid 2006 | 36,6% |
| Presented method (CDEF_5000) | 39,4% |
| JRC Grid 2000 (CLC-iterative) | 57% |
| JRC Grid 2000 (CLC-limiting variable) | 53% |
| Communes (choropleth) | 112% |

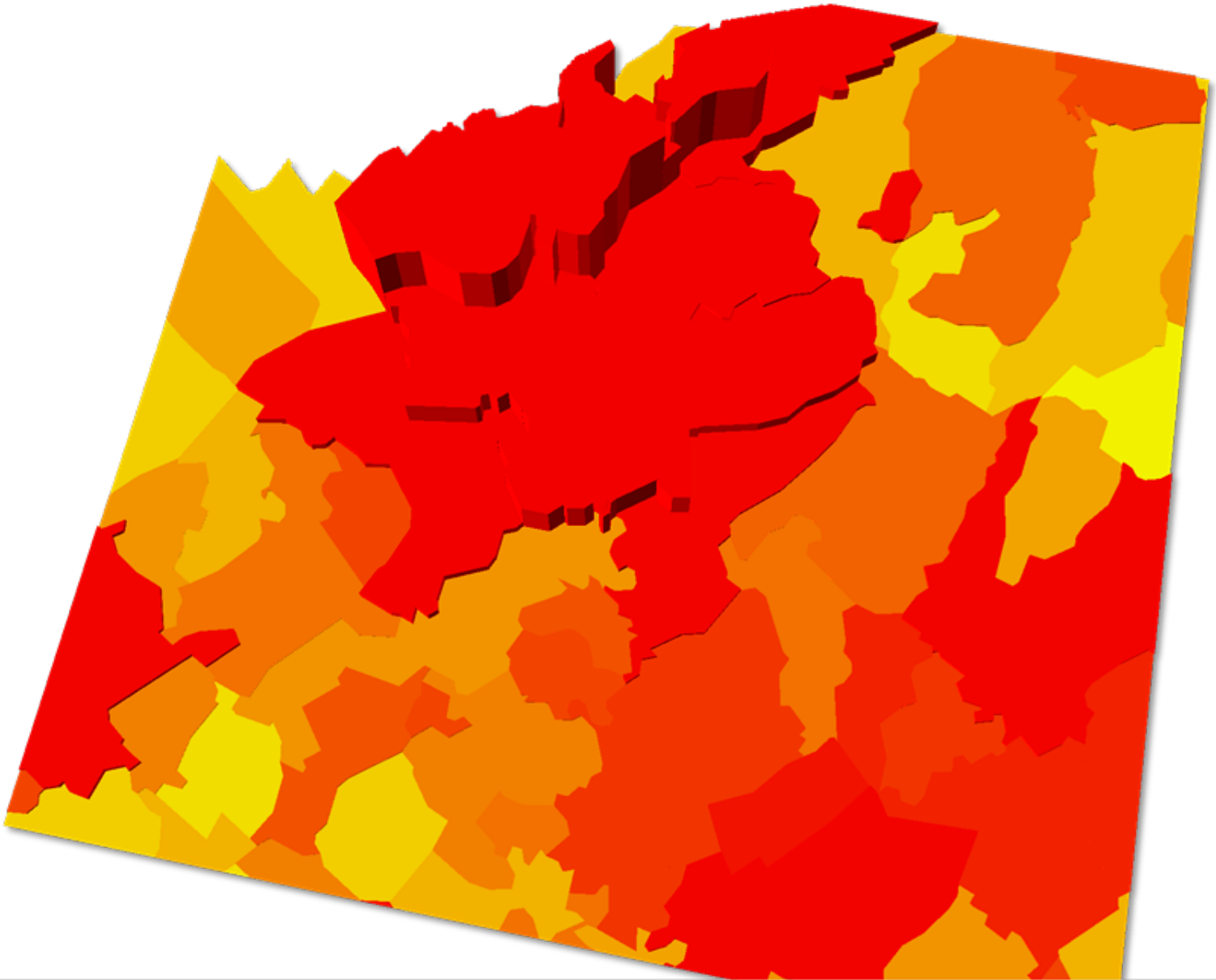
- Slovenia at 1km

| | |
|---|---|
| AIT Grid 2006 | 62,2% |
| JRC Grid 2006 | 65,1% |
| Presented method (ADEF_300) | 65,5% |

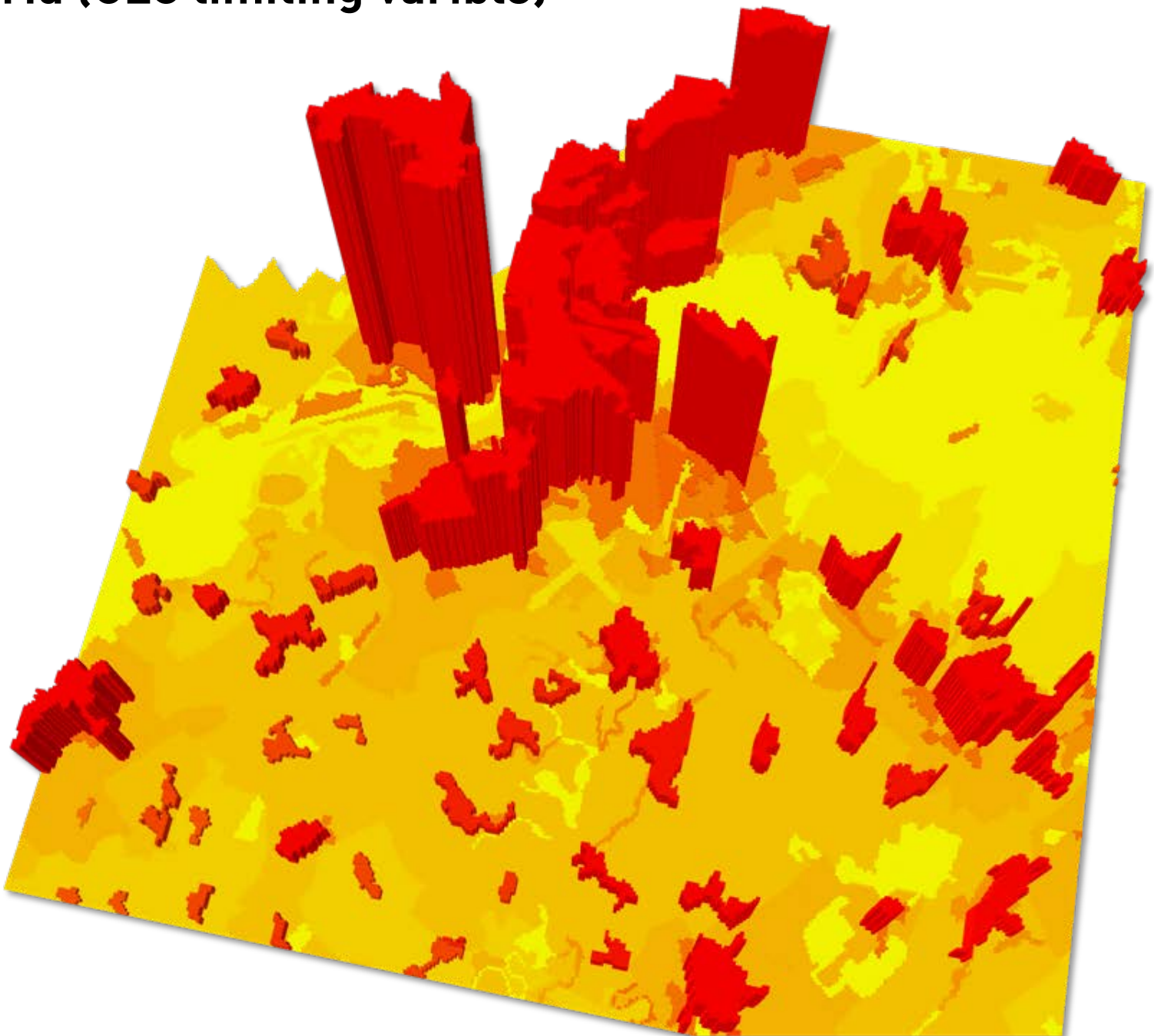
- Slovenia at 100m

| | |
|---|---|
| JRC Grid 2006 | 97,6% |
| Presented method (ACDEF_5000) | 97,0% |

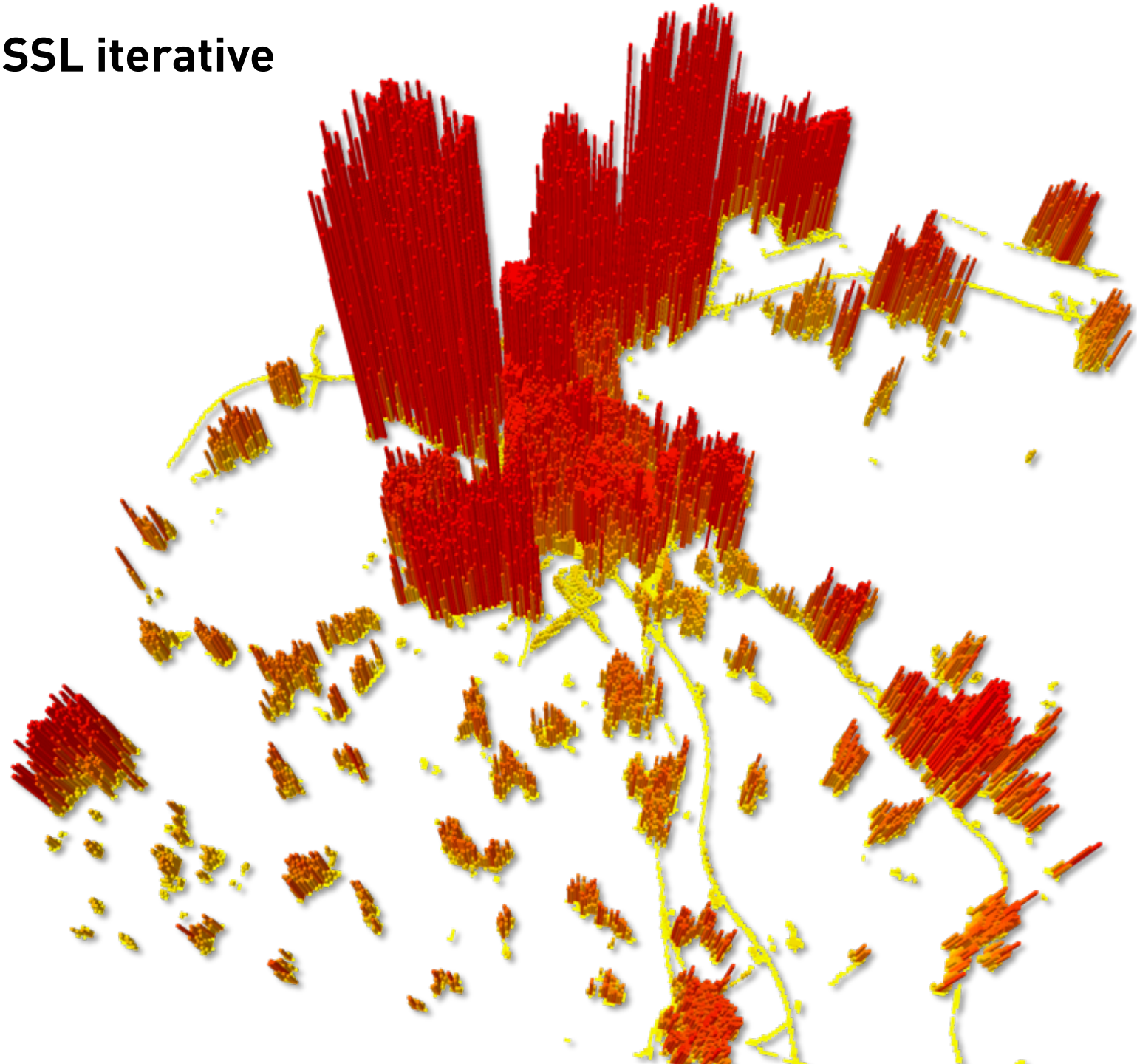
# Commune choropleth map



# JRC grid (CLC limiting varible)



# CLC-SSL iterative



# Conclusions

- The analysis confirmed that the use of more ancillary datasets results in better estimates: CLC > CLC-SSL > CLC-SSL-OSM; therefore, additional ancillary datasets (OpenStreetMap, etc.) should be employed
- Accuracy of the best performing model is comparable to other similar products despite the fact that - unlike the presented approach - they employ additional ancillary datasets (e.g. road network, Urban Atlas
- In Slovenia at 100m resolution our model preformed better compared to JRC grid 2006 (with fewer ancillary datasets)
- Accuracy of the estimate can be improved by meaningful design of the commune subsets
- There is no point in being confined to conventional regions (spatially contiguous and disjoint subsets)
- The best results were achieved using non-contiguous subsets created on the basis of other criteria such as commune population and share of built-up surface and population density per built-up area

# Thank you for your attention

konstantin.rosina@savba.sk
pavol.hurbanek@gmail.com

*This research was supported by the Slovak Scientific Grant Agency VEGA project no. 1/0275/13 "Production, verification and application of population and settlement spatial models based on European land monitoring services".*

# Literature


- BATISTA, F., GALLEGO, J., & LAVALLE, C. (2013). A high-resolution population grid map for Europe. Journal of Maps, (January), 37–41.
- GALLEGO, F. J. & PEEDELL, S., (2001). Using CORINE Land Cover to map population density. In *Towards Agrienvironmental indicators*. Topic report 6/2001 European Environment Agency, Copenhagen, 92-103.
- GALLEGO, F. J. (2009). Population density grid of EU-27+, version 5. European Environment Agency . Available at: http://www.eea.europa.eu/data-and-maps/data/population-density-disaggregated-with-corine-land-cover-2000-2
- GALLEGO, F. J. (2010). A population density grid of the European Union. *Population and Environment*, 31, 460–473.
- GALLEGO, F., BATISTA, F., ROCHA, C., & MUBAREKA, S. (2011). Disaggregating population density of the European Union with CORINE land cover. *International Journal of Geographical Information Science*, 25(12), 37–41.
- MARTIN, D., TATE, N. J., & LANGFORD, M. (2000). Refining population surface models: experiments with Northern Ireland census data. Transactions in GIS, 4(4), 343–360.
- ROSINA, K. &HURBÁNEK, P. (2012). An iterative method for dasymetric mapping of population density based on European land monitoring services / Iteratívna metóda dazymetrického mapovania hustoty zaľudnenia na báze európskych služieb pre monitoring krajiny. *Cartographic letters / Kartografické listy*, 20(2).
- ROSINA, K., HURBÁNEK, P. & ATKINSON, P. M. (2012). Spatial disaggregation of population data using maps of land cover and imperviousness / Priestorová dezagregácia populačných dát s využitím máp krajinnej pokrývky a nepriepustnosti povrchu. In *Symposium GIS Ostrava 2012 proceedings*. ISBN 978-80-248-2558-8.