Delineating urban areas with building density

Work in progress (first version)

Marie-Pierre de Bellefon
INSEE and Paris School of Economics

Pierre-Philippe Combes
University of Lyon,CNRS and Sciences Po

Gilles Duranton
University of Pennsylvania

Laurent Gobillon
Paris School of Economics - CNRS

Clément Gorin
University of Lyon

Objectives

Develop a novel statistical / dartboard approach to

1. Delineate metropolitan areas
2. Define and describe the internal geography of these areas (centres, subcentres, etc)
3. Test for differences between our baseline map, variants, or other maps produced by statistical institutes or other researchers (one- and two-sided tests)

We implement our approach using unique building-level data for France

Why this matters

- Urban research obviously requires defining its object
- Inappropriately defined units may bias results (MAUP)
- It may also alter perceptions and policies
- Across the world, official delineations of cities and metropolitan areas can be problematic or missing (eg much of Latin America, Asia, and most of Africa)

The challenges that must be overcome

1. Minimise the number and impact of arbitrary choices
2. Provide a statistical grounding for comparisons
3. Retain some computational feasibility
4. Find appropriate high-resolution data

Data on buildings

We use data from BD TOPO 2014 which is a 3D description of buildings in France with metric precision provided by the French Geographic Institute (IGN)

- It is constructed from satellite images and cadastral information
- It contains information on buildings (including footprint, height, and use), roads and train network
- We divide the French territory into 200m x 200m pixels
- For each pixel, we compute the "building density" as the volume of builtup space (whatever the building type)
- In the computation, a building is fully attributed to the pixel that includes the largest share of the building area

Grenoble: Google Map

Grenoble: Google Map, BD TOPO and pixel limits

Rural area close to Grenoble: Google Map

Rural area close to Grenoble: G. Map, BD TOPO, pixels

Buildable pixels

We use two datasets to determine whether pixels are buildable:

- Coverage by water (sea, lakes, rivers): BD Carthage 2006
- Data on elevation: BD Alti 2015

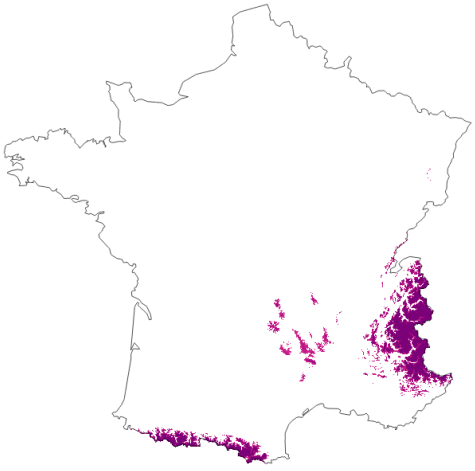
We consider 3 criteria to determine the buildability of a pixel: proportion of area covered with water, slope and elevation

We compute the distribution of each of these characteristics for built pixels and determine its 99^{th} centile:

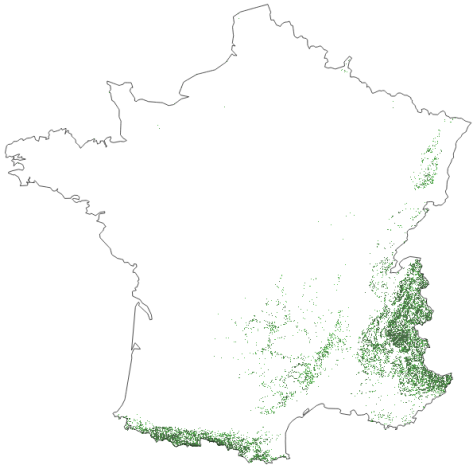
- Proportion of water: 42.4%
- Elevation: 1213 meters
- Slope: 21%

We keep only pixels which characteristics are below these values (8% deleted)

Non-buildable areas

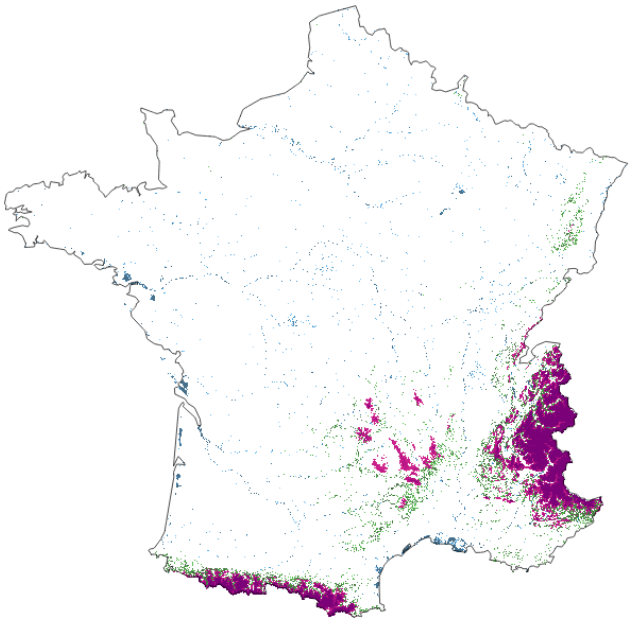


Elevation



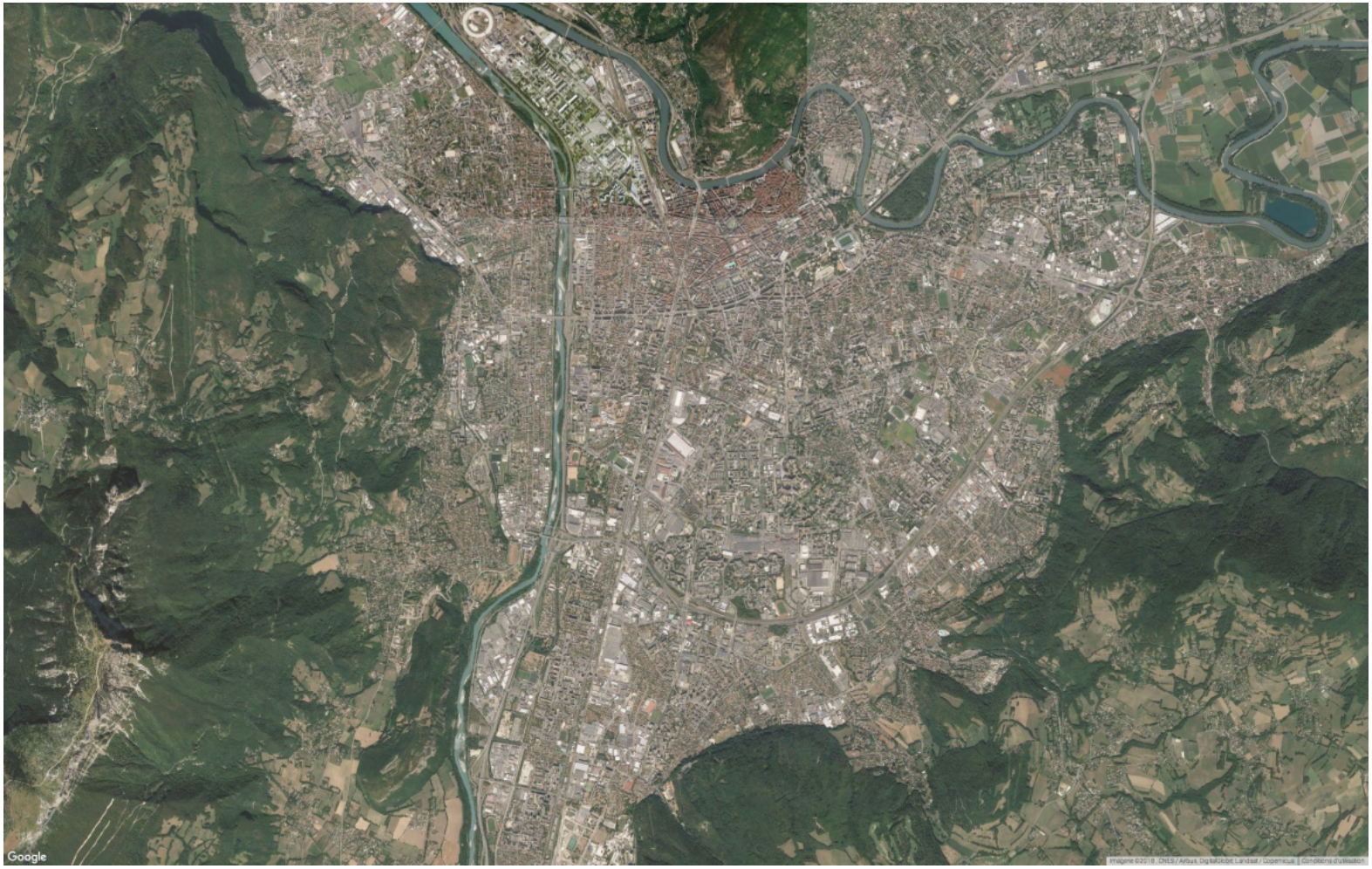
Slope

Water

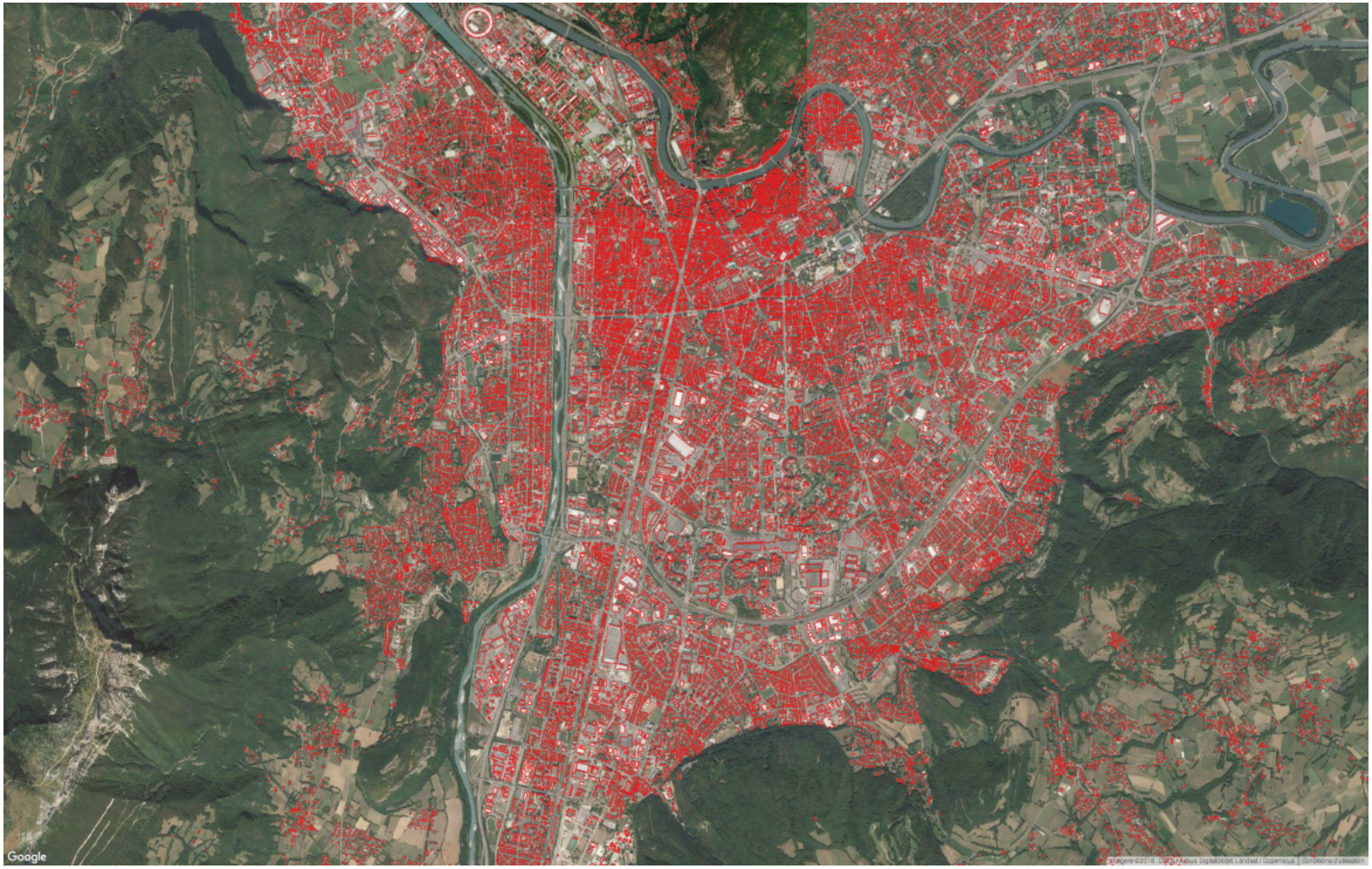


All criteria

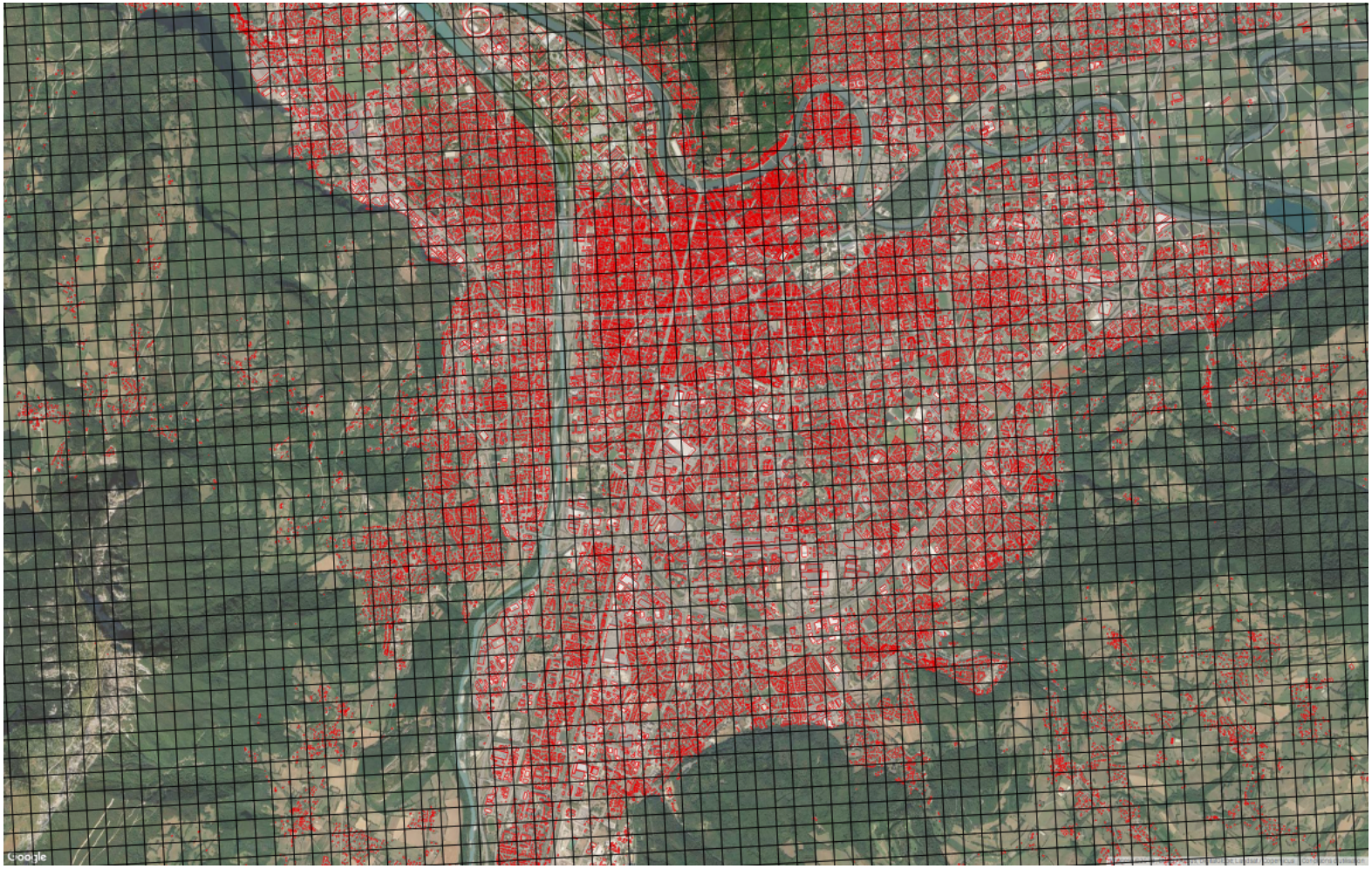
Greater Grenoble: Google Map



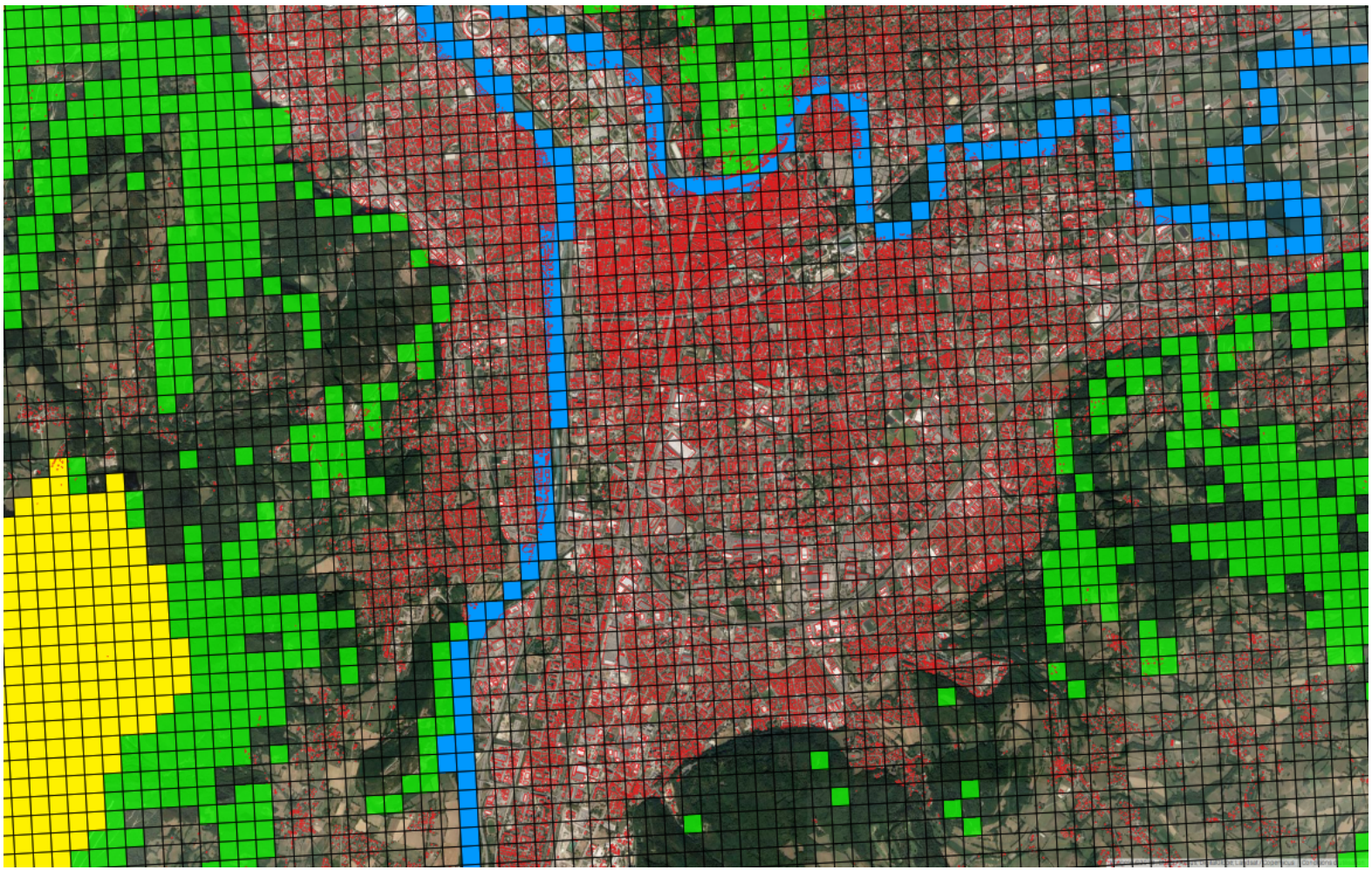
Greater Grenoble: Google Map and BD TOPO



Greater Grenoble: Google Map, BD TOPO and pixels

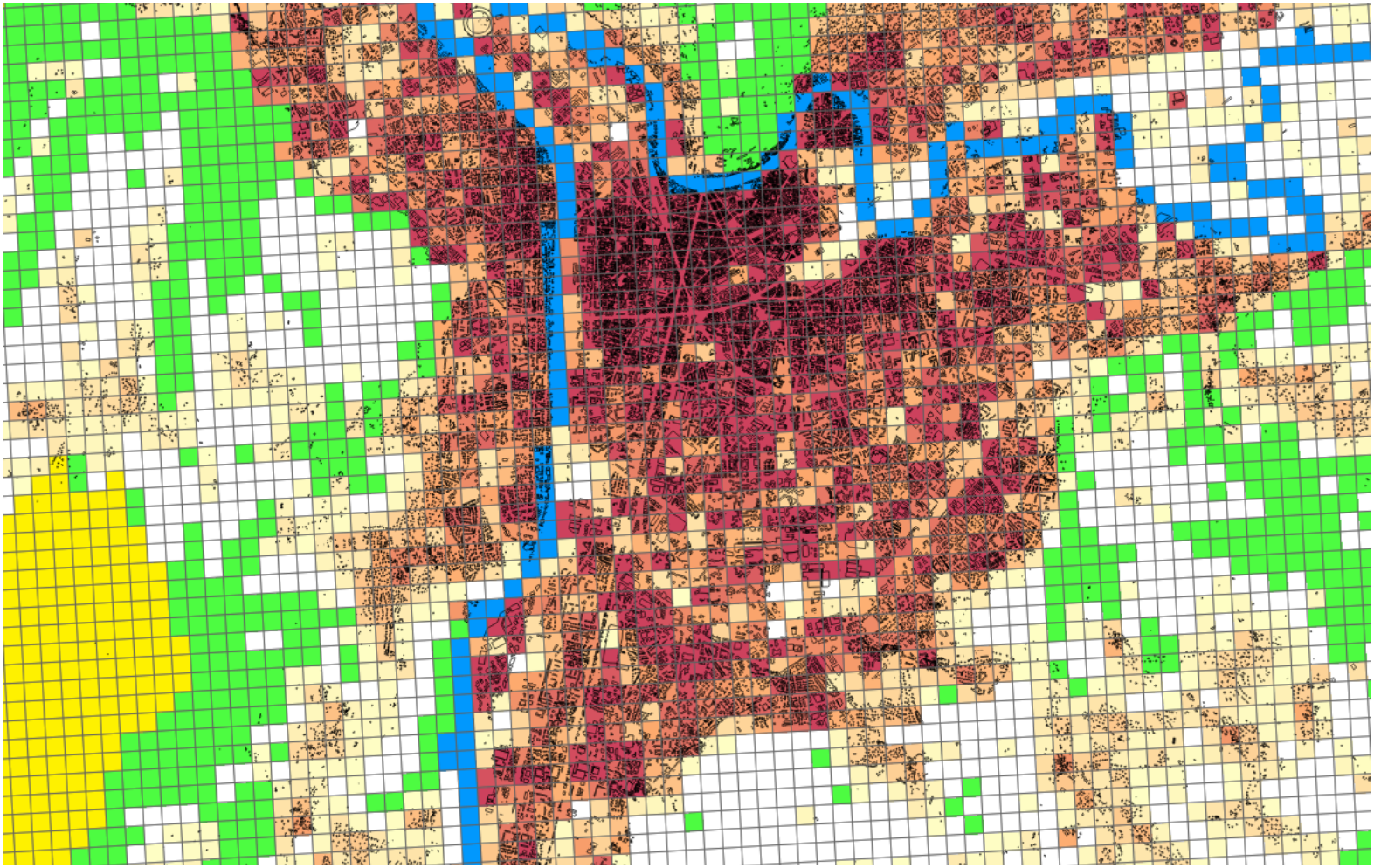


Greater Grenoble: G. Map, BD TOPO, non-buildable pixels



In red: buildings; in blue: water; in green: slope; in yellow: elevation

Greater Grenoble: Buildable and non-buildable pixels



In red: building density; in white: buildable without building; in blue: water; in green: slope; in yellow: elevation

Population data

We have population in 2010 for 200m x 200m pixels

- Geolocalised fiscal data are collected for income and residential tax purposes (Localised Tax Revenues)
- They are provided by the fiscal administration and treated by the French Institute of Statistics (INSEE)
- Pixels constructed from BDTOPO are designed such that they coincide with population pixels

A dartboard approach

We have raw data on building density for 200m x 200m pixels

We smooth this density using a kernel

For our counterfactuals, we conduct the following procedure $S = 100$ times:

- We redistribute randomly buildings across buildable pixels
- We compute the corresponding counterfactual building density
- We smooth this density

We consider that a pixel is urban if the observed density is above the 95^{th} percentile of the distribution of counterfactual densities computed for that pixel

A city is a set of contiguous urban pixels

Building density smoothing

To avoid using too many pixels for smoothing, we use a kernel with weights zero after a given distance (bisquare kernel)

Smoothed building density for pixel j with coordinates (x_j, y_j) is given by:

$$\widehat{z}_j = \sum_i K_h(d_{ij}) z_i$$

where z_i is the building density for pixel i, $d_{ij} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}$ is the distance between pixels i and j, and K_h is a bisquare kernel with bandwidth h verifying:

$$K_h(d_{ij}) = \left[1 - \left(\frac{d_{ij}}{h}\right)^2\right]^2 1\{d_{ij} < h\}$$

Bandwidth choice (1/2)

Bandwidth choice:

- Large bandwidth $\Longrightarrow$ use of many observations; data too smoothed to identify local variations
- Small bandwidth $\Longrightarrow$ use of few local observations; hard to define homogenous areas with high density

Generalised cross-validation criterium: bandwidth chosen to minimise the difference between observed and smoothed density (excluding the observation at hand)

Bandwidth choice (2/2)

Bandwidth using all pixels in France leads to very large values ($>$10km) $\Longrightarrow$ over-smoothing

Consequently, we apply the following procedure:

- We split France in 100km x 100km tiles with partial overlap
- We determine a bandwidth for each tile using the cross-validation criterium
- We compute the median of bandwidths across tiles

This approach gives us a reasonable bandwidth of 1.97km

Determination of urban cores

We can also apply a dartboard approach to determine urban cores

We conduct the following procedure $S = 100$ times:

- We redistribute randomly buildings across urban pixels for the whole France
- We build the counterfactual building density for each urban pixel
- We smooth this density using only urban pixels

A pixel is part of an urban core if its observed density is above the 95^{th} percentile of the distribution of counterfactual densities computed for that pixel

Descriptive statistics on built area

Built area	Min	25^{th}	Med	75^{th}	95^{th}	Max	Std
Raw (m^2)	0	0	0	0	2,155	579,352	1,484
Raw (%)	0	0	0	0	5.39	1,448	3.71
Raw (m^3)	0	0	0	0	12,417	14,483,810	27,982
Smoothed (m^2)	0	69	171	343	1,377	23,104	808
Smoothed (%)	0	0.17	0.43	0.86	3.44	57.8	2.02
Smoothed (m^3)	0	394	984	2,032	9,011	456,697	8,056

Observation unit: 200m x 200m pixels

Before smoothing:

- 76% of pixels are not built at all
- At 95^{th} percentile, only 5.4% of a pixel is builtup

After smoothing:

- 3.3% of pixels are not built at all
- At 95^{th} percentile, only 3.3% of a pixel is builtup

Cities

7,223 cities that represent 11% of all pixels
(or 12% of buildable pixels)

Total urban population is 75% of population of mainland France

Many small cities (population at 95^{th} percentile is 10,562)

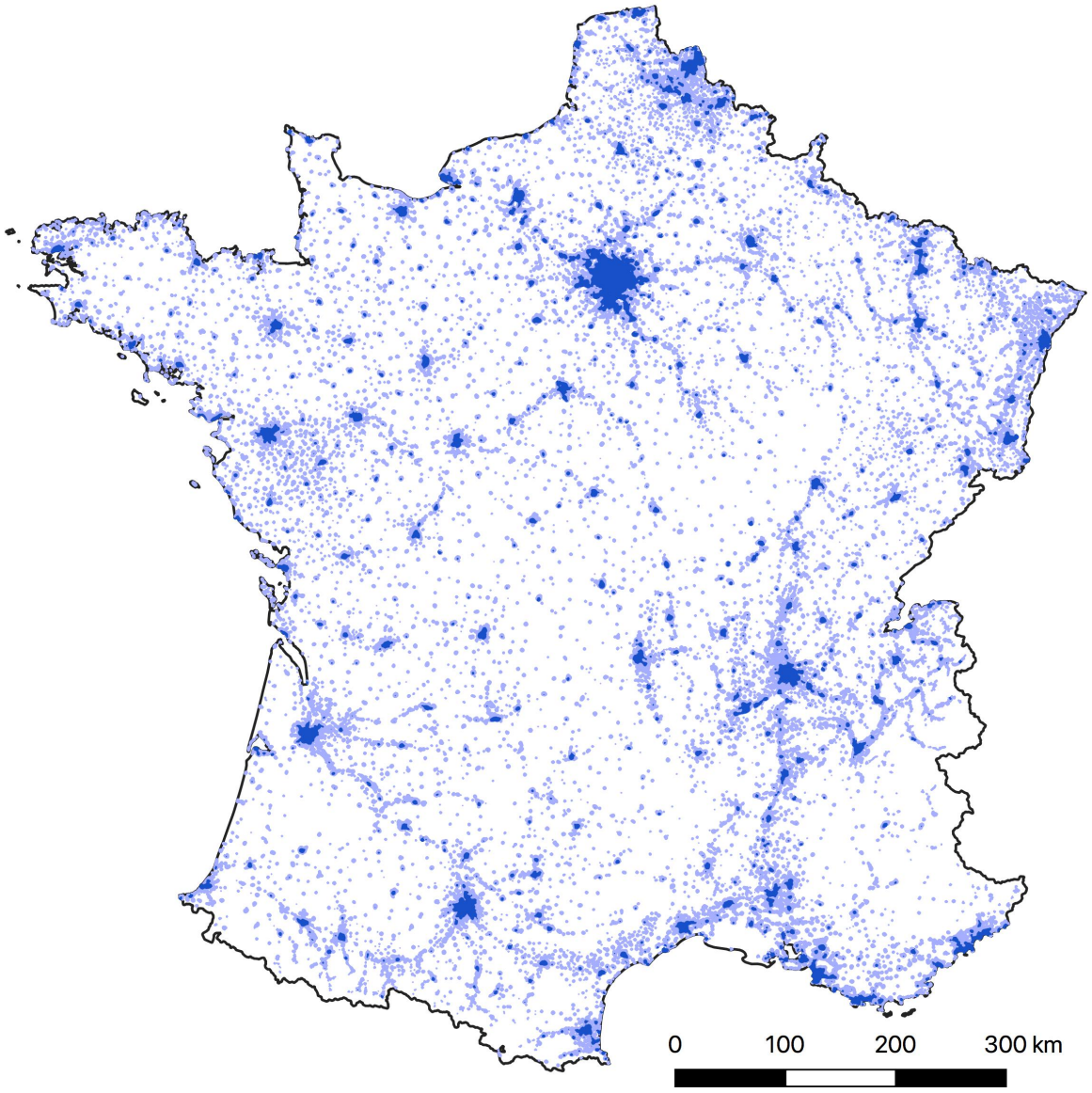
There are 695 cities with a core and 6,528 without a core

Average population for cities with (resp. without) a core is 60,127 (resp. 1,134)

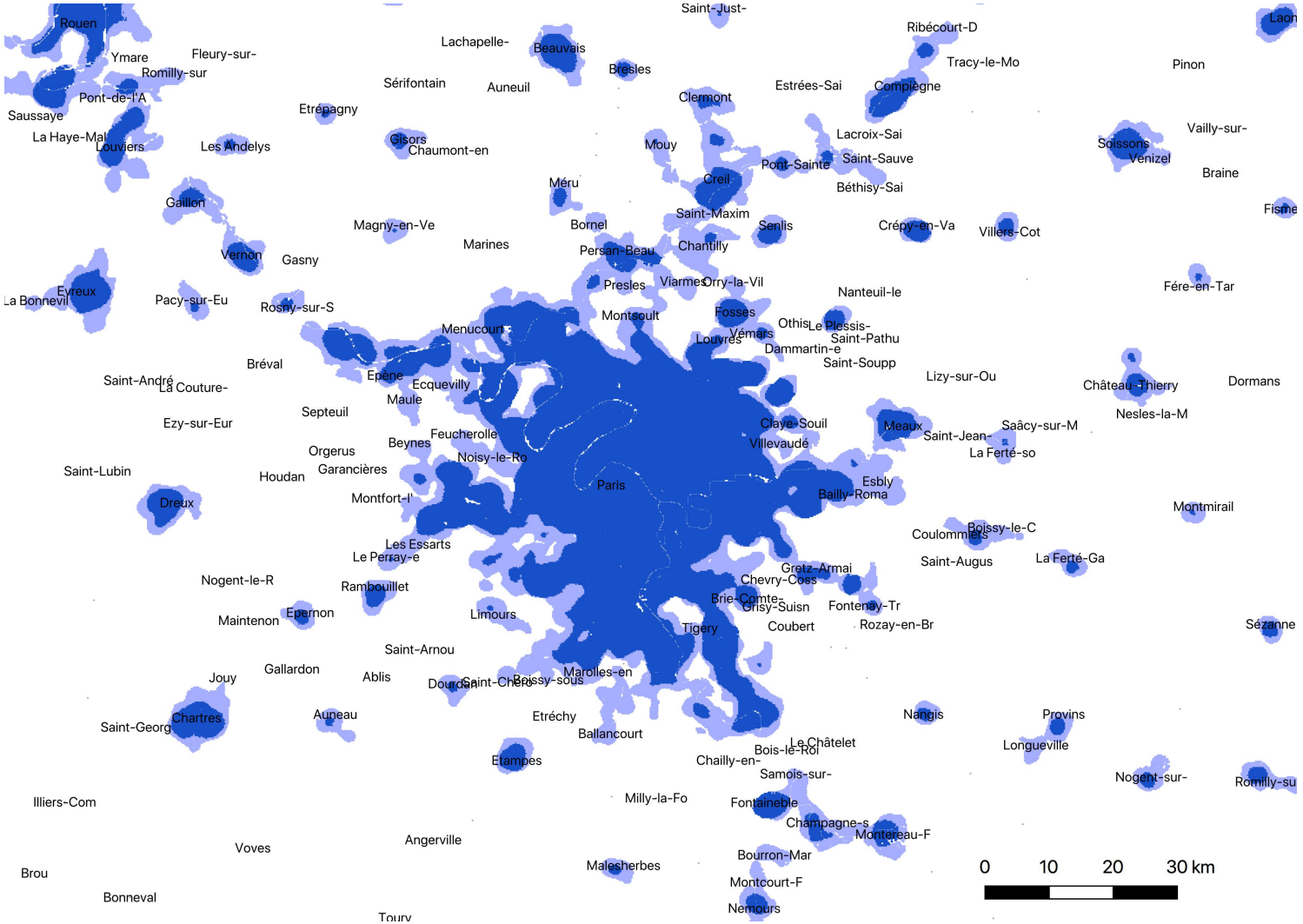
Cities with (resp. without) core host 64% (resp. 11%) of the French population

Cities with (resp. without) core occupy 7.7% (resp. 4.1%) of the French territory

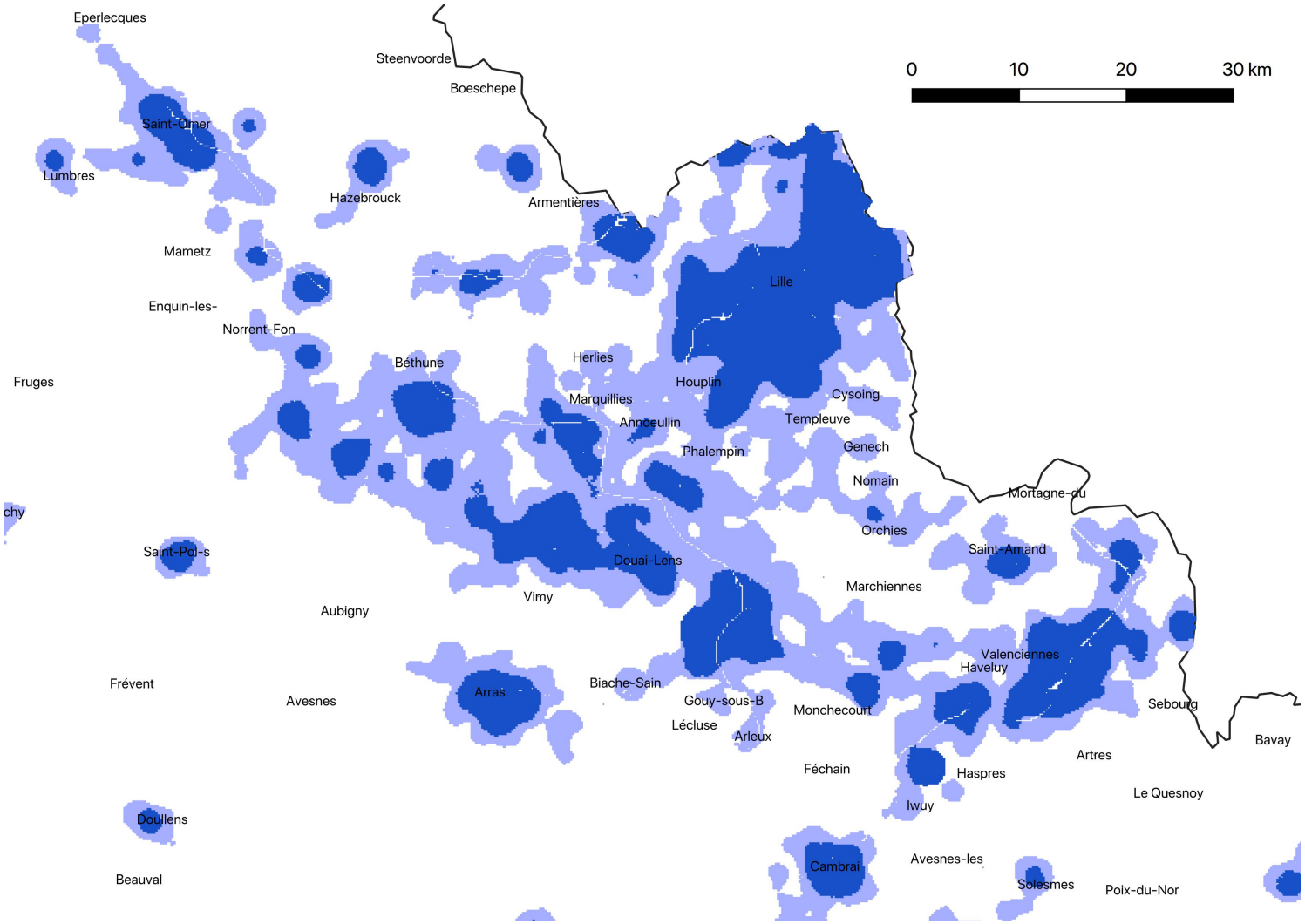
Cities in France



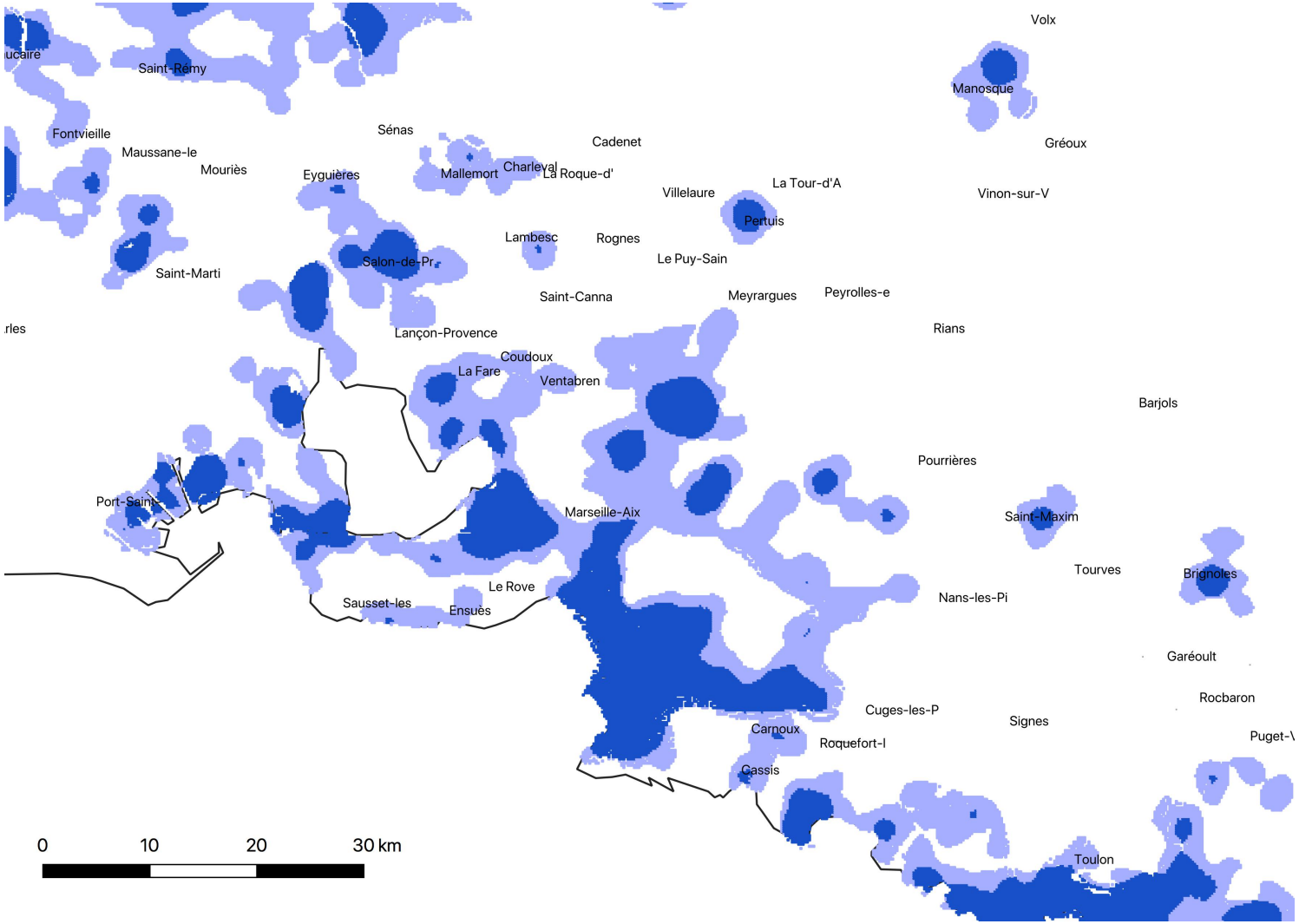
Paris and the Ile-de-France region



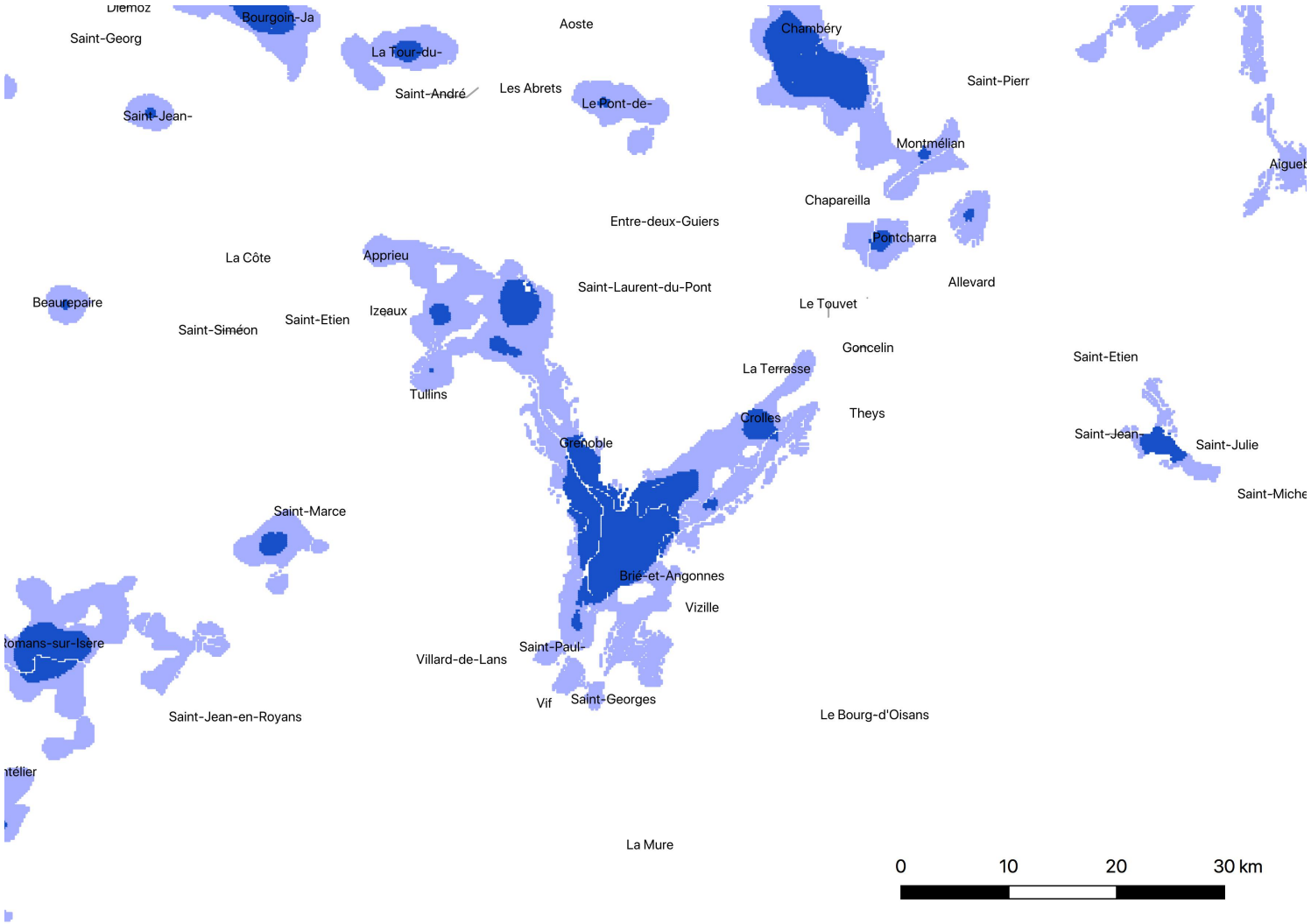
Lille and the North East



Marseille and the South East



Grenoble and the Alpine region



Alternative definition of cities

$\Longrightarrow$ Morphological zoning: Urban Units (UU)

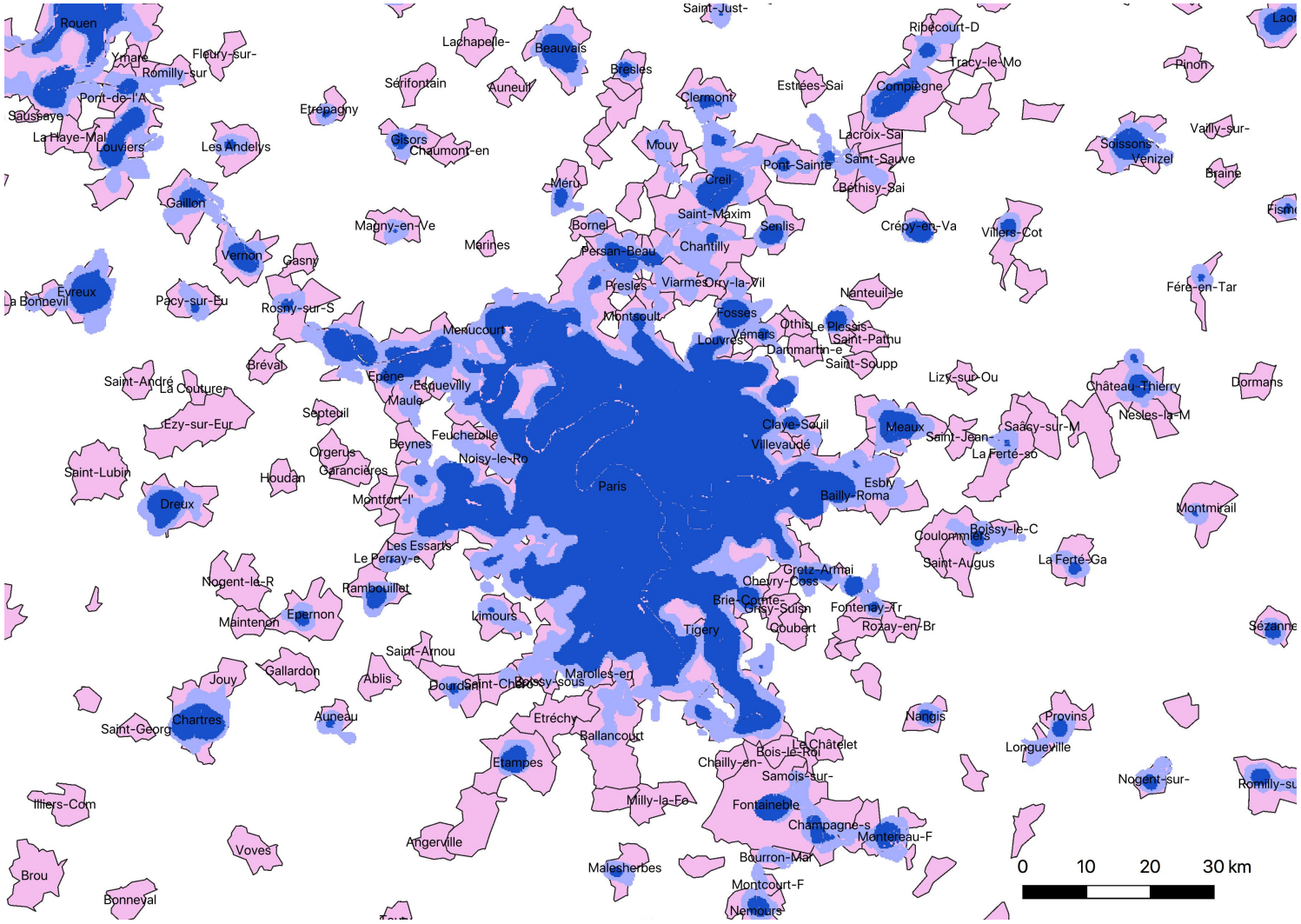
Group of municipalities characterized by a continuously builtup area (less than 200 meters between two constructions) and more than 2,000 inhabitants

There are 2,231 UU in 2010 (we obtained 7,223 cities)

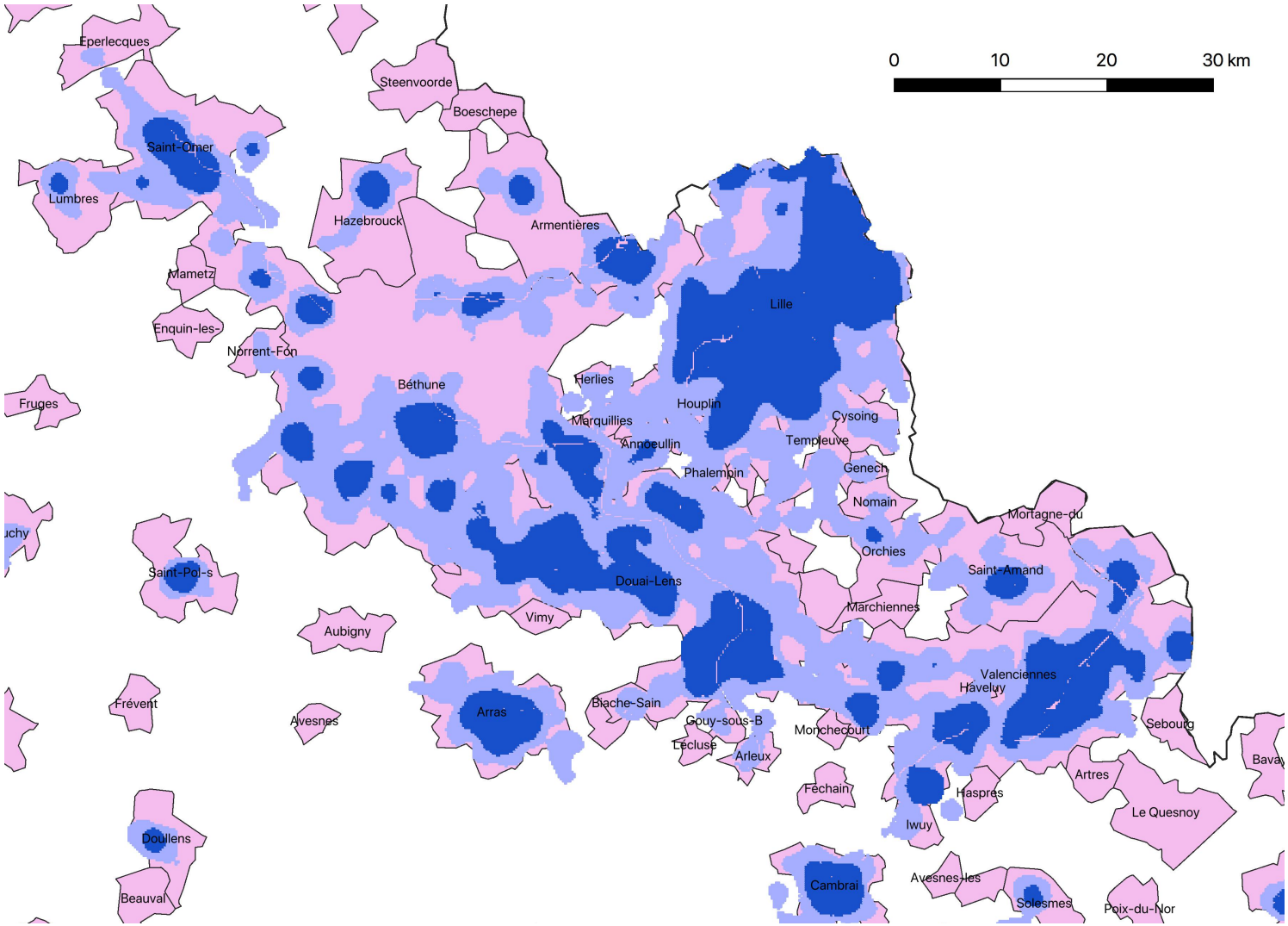
UU cover 22% of the territory (11% for our cities)

UU physically larger that cities with core but less populated

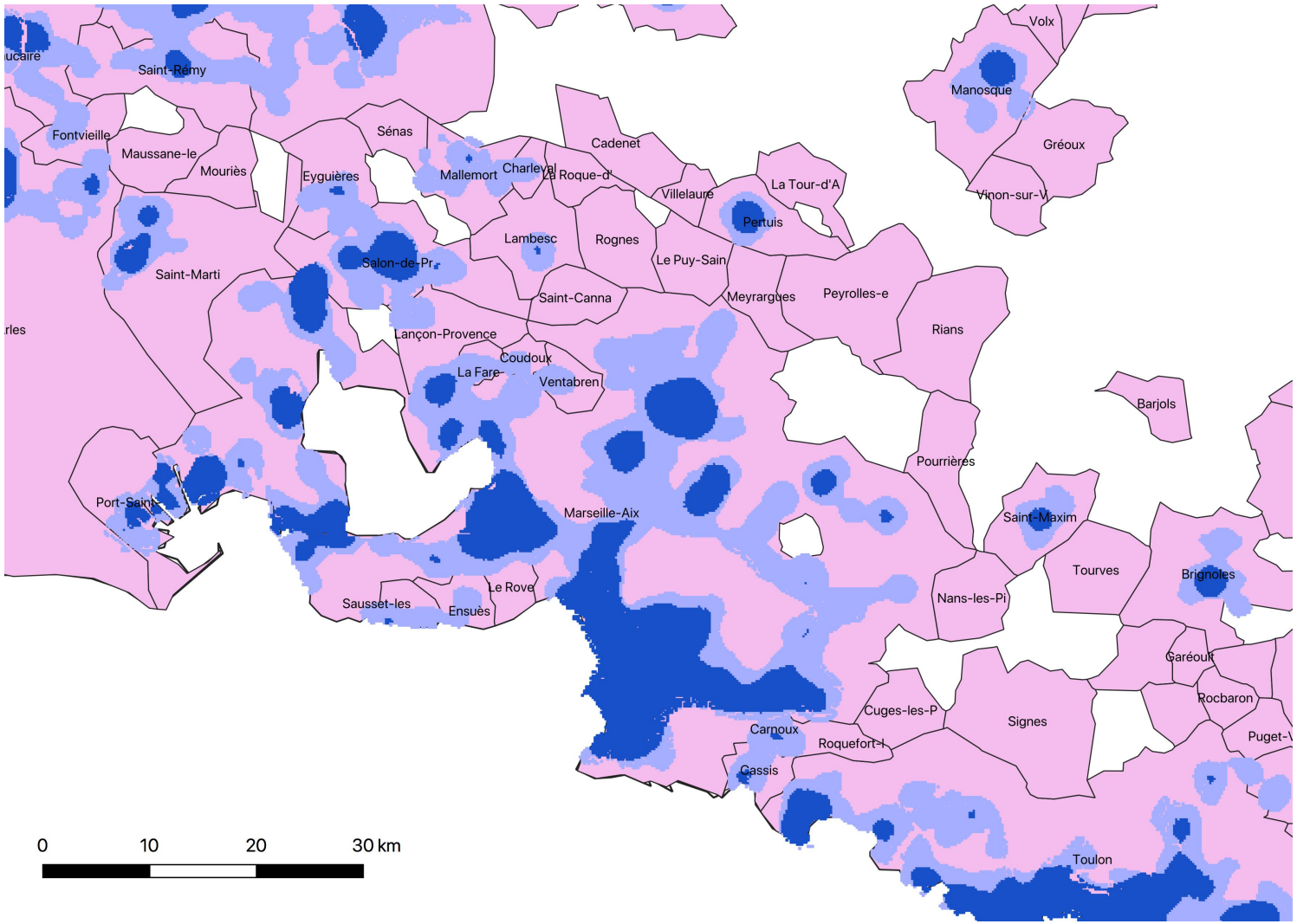
Paris and the Ile-de-France region: Cities and UU



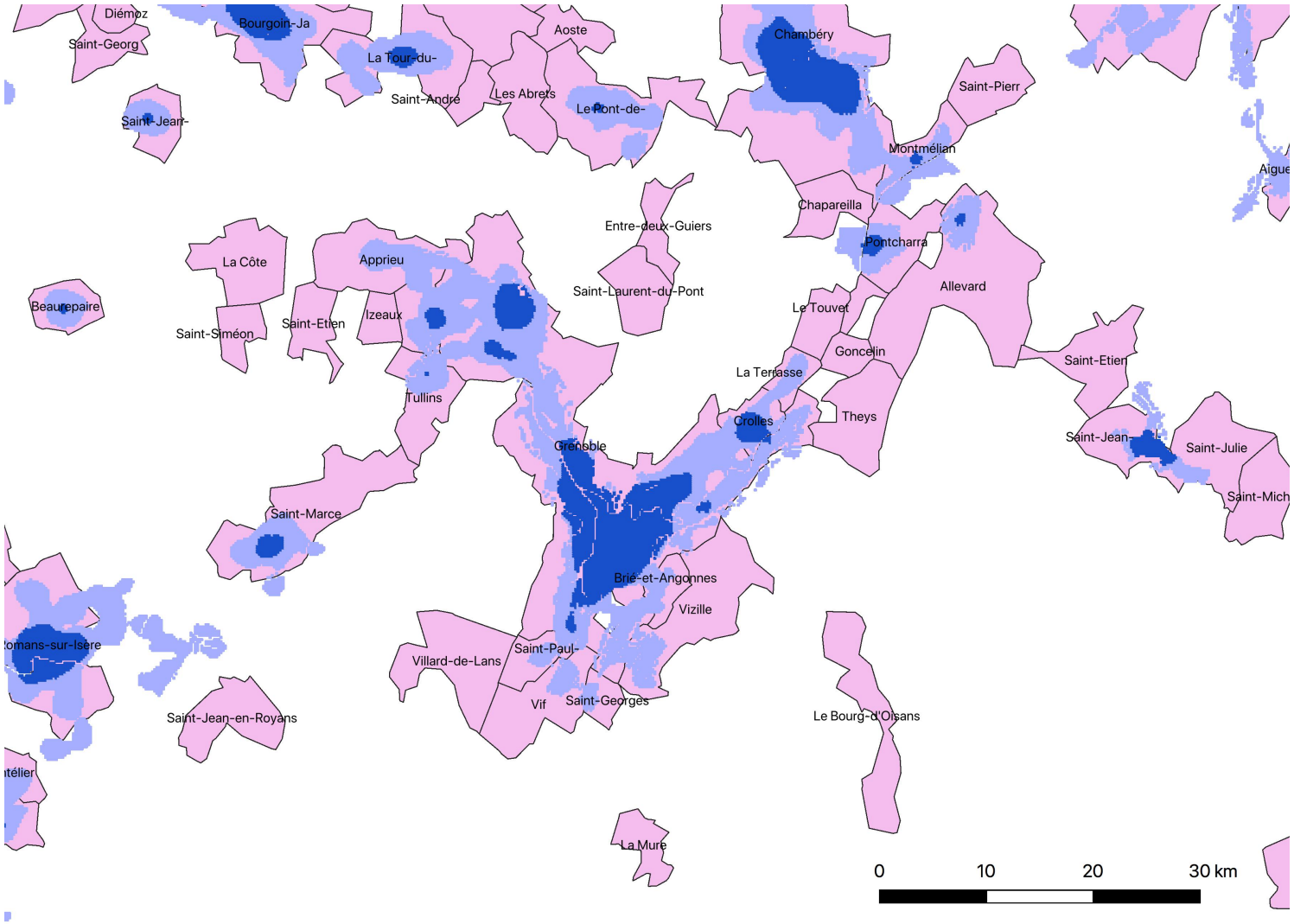
Lille and the North East: Cities and UU



Marseille and the South East: Cities and UU



Grenoble and the Alpine region: Cities and UU



Comparing maps

We compute Jaccard indices from pixels to compare two maps

Jaccard index: ratio of intersection between two sets divided by their union

- "Urban Jaccard index": urban vs. rural to compare extent of urbanisation
  (proportion of pixels that are urban in the two maps)
- "City Jaccard index": counting only pixels in the "same" city (proportion of pixels that are urban and in the "same" city in the two maps)
- Bootstrap procedure to compute the confidence intervals of our indices

Comparison of definitions

Comparison between the maps of our cities and UU:

- Urban Jaccard index: 0.319 (std. very small)
- Urban Jaccard index when restriction to cities with core: 0.298
- City Jaccard index: 0.177
- City Jaccard index when restriction to cities with core: 0.182

Ranking of cities with respect to population

Rank	City	Population	INSEE Population	INSEE rank	Jaccard index
1	Paris	10,932,881	10,730,549	1	0.66
2	Lille	2,197,967	1,037,834	4	0.23
3	Lyon	1,777,944	1,627,937	2	0.49
4	Aix-Marseille	1,442,734	1,570,325	3	0.44
5	Nice	1,024,679	956,189	6	0.65
6	Toulouse	875,595	938,284	5	0.65
7	Bordeaux	831,453	893,384	7	0.47
8	Strasbourg	692,009	451,522	13	0.22
9	Nantes	587,495	628,718	8	0.53
10	Grenoble	520,445	518,495	10	0.48

Area in km^2; density: Number of inhabitants per km^2; INSEE population and rank are those of urban units; INSEE population is that of 2013 census; city named after municipality with the largest population included in the city

Building density defined as footprint

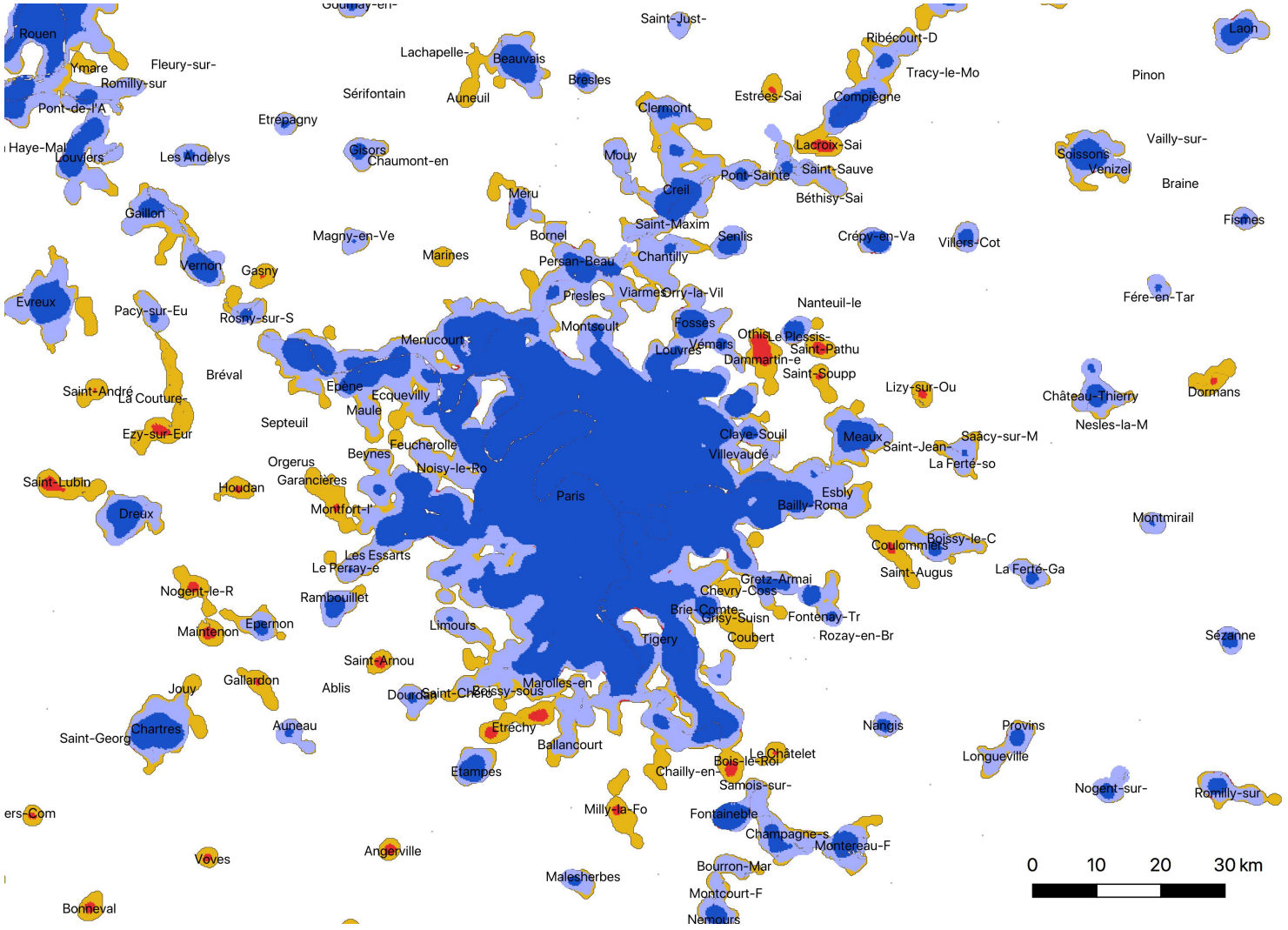
Cities cover 15% of pixels (instead of 11% with volume)

They host 80% of population (instead of 75% with volume)

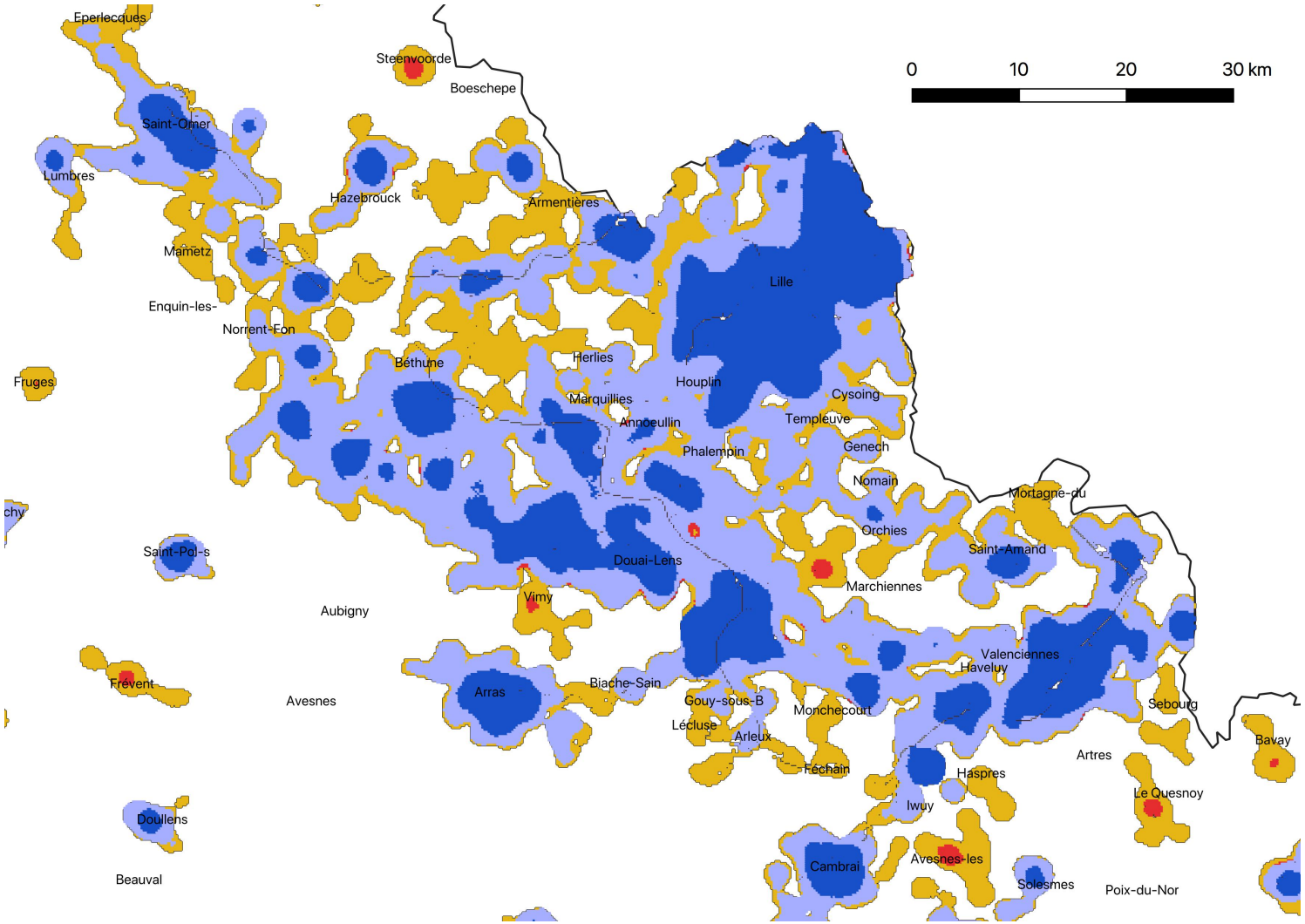
Not surprising because taller buildings tend to be located at the centre of cities

Areas with fewer and shorter buildings: Excess building density when using footprint but not when using volume

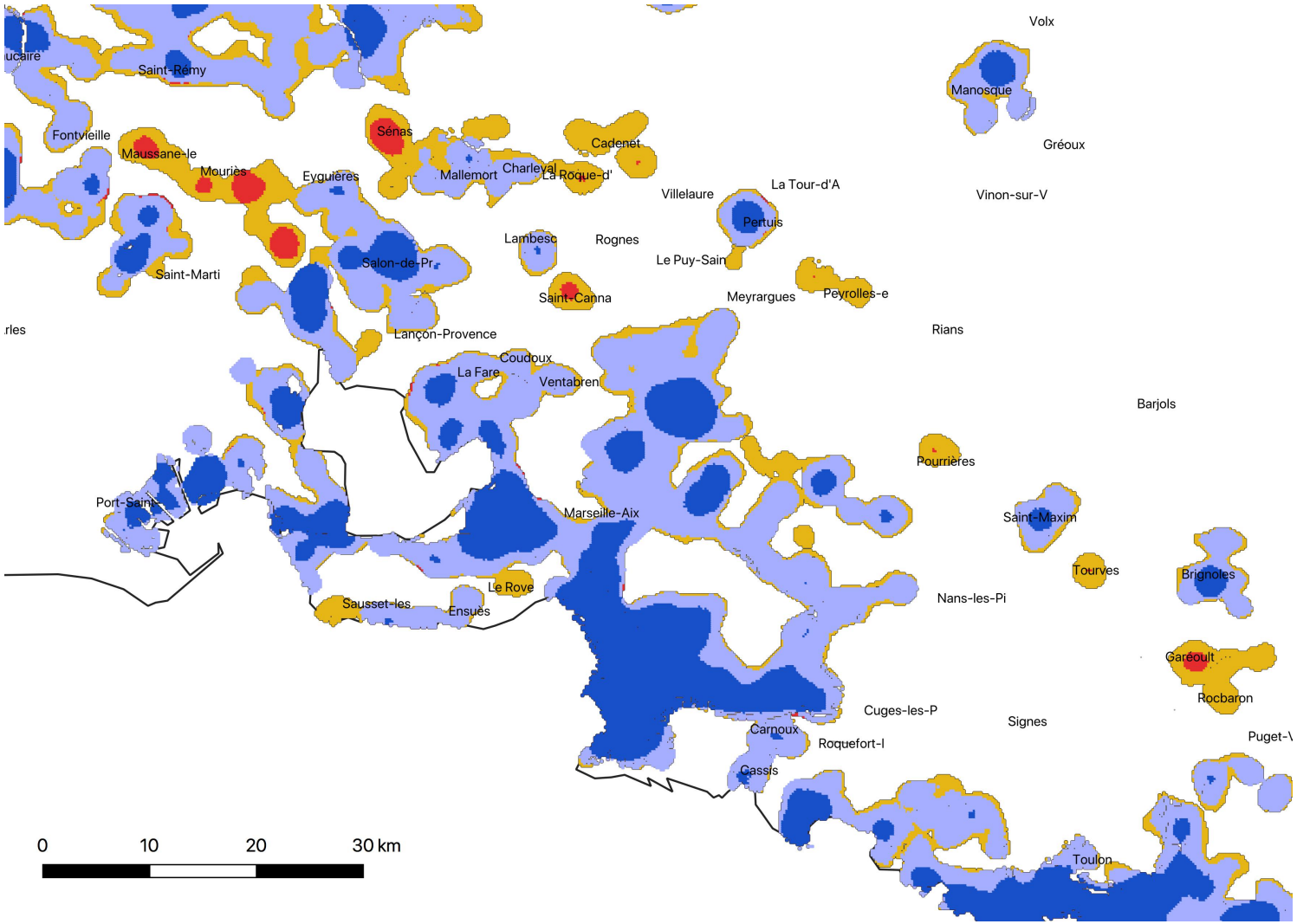
Paris and the Ile-de-France region: Volume vs. footprint



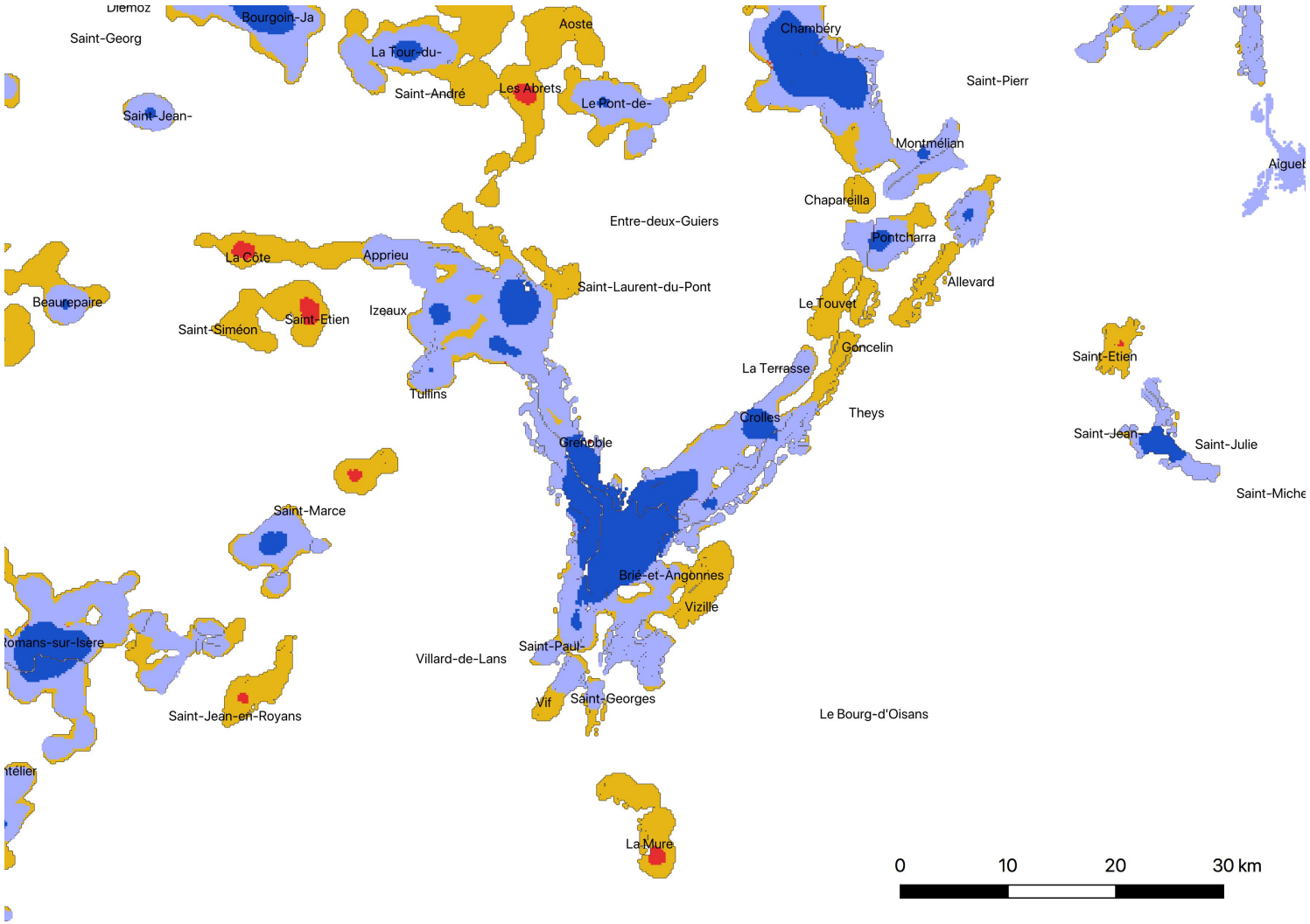
Lille and the North East: Volume vs. footprint



Marseille and the South East: Volume vs. footprint



Grenoble and the Alpine region: Volume vs. footprint



Conclusion

- We propose a dartboard approach using 200m x 200m pixels to determine cities and their cores in France
- We show that obtained cities are different from official urban units. In particular:

  - Obtained cities aggregate some municipalities for which there is building continuity but not the official definition (cf. Lille)

  - Official definition aggregates many municipalities containing no or barely any urban pixels

Going further

- Apply our approach to pixels in urban cores (and iterate)
- Detect centers and subcenters within cities by considering random distribution of buildings within cities
- Count centers and subcenters considering peaks of density
- Conduct traditional exercises using our definition of cities

  (Zipf law, real wage spatial disparities, estimation of agglomeration economies)

Research agenda

Evolution of city sizes over 250 years

Four series of maps:

- Cassini, initiated by Colbert for Louis XIV
  Started 1750s, France fully covered in 1789
- Etat-Major, for military purpose; 1825-1865
- 'Scan 50' photographs, aerial photographs, 1959
- BDTOPO 2014

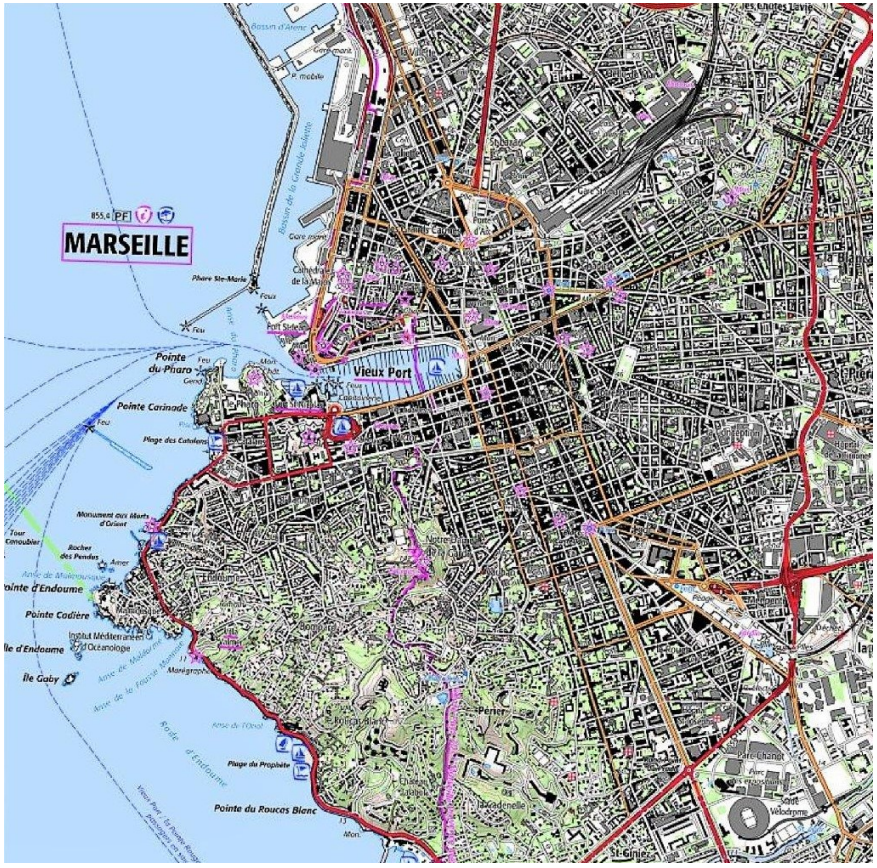
1750 and 1850: Marseille

Cassini, 1750

Etat-Major, 1850

1850 and 1950: Marseille

Scan50, 1950

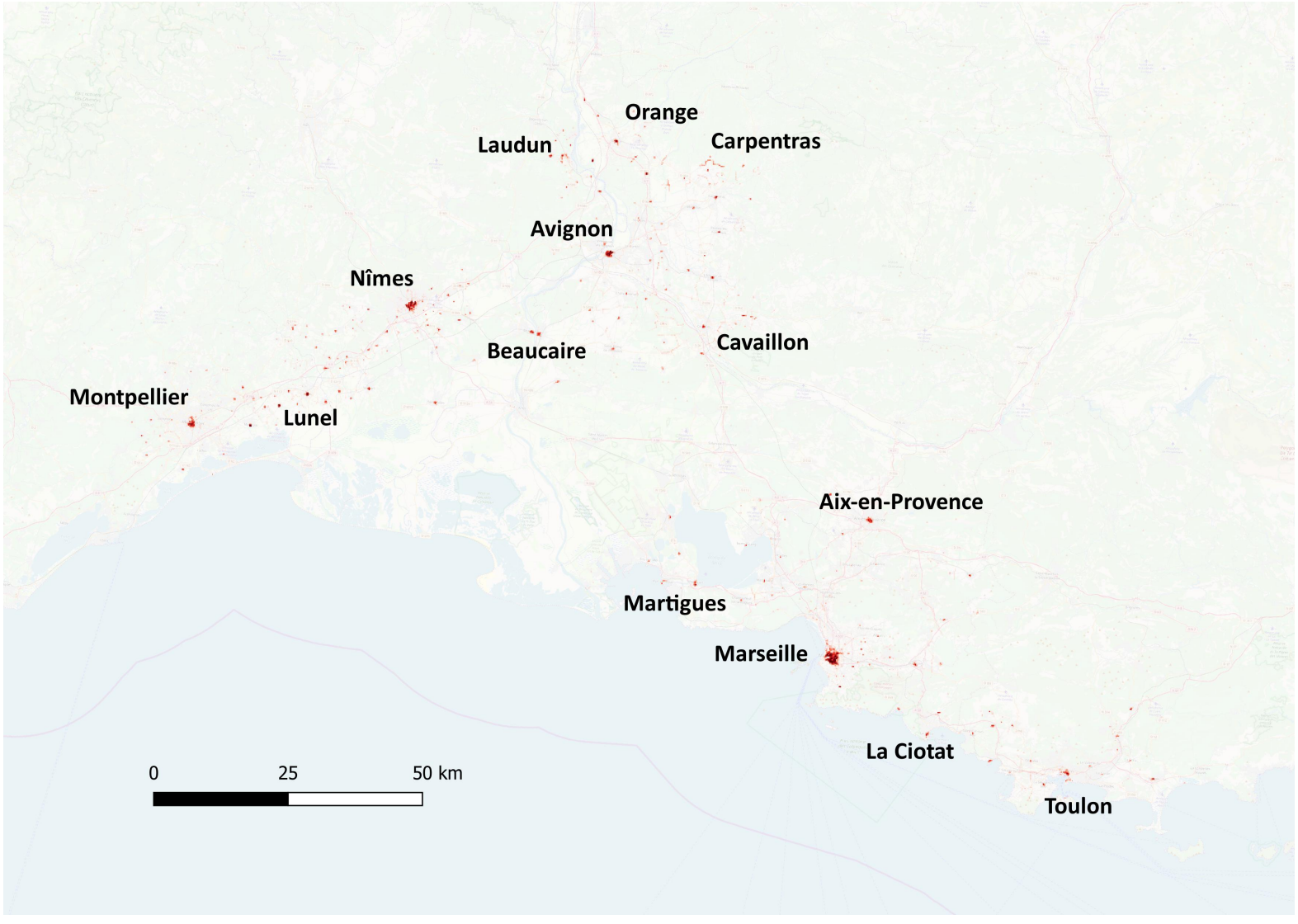


Top 25, Marseille, 2015

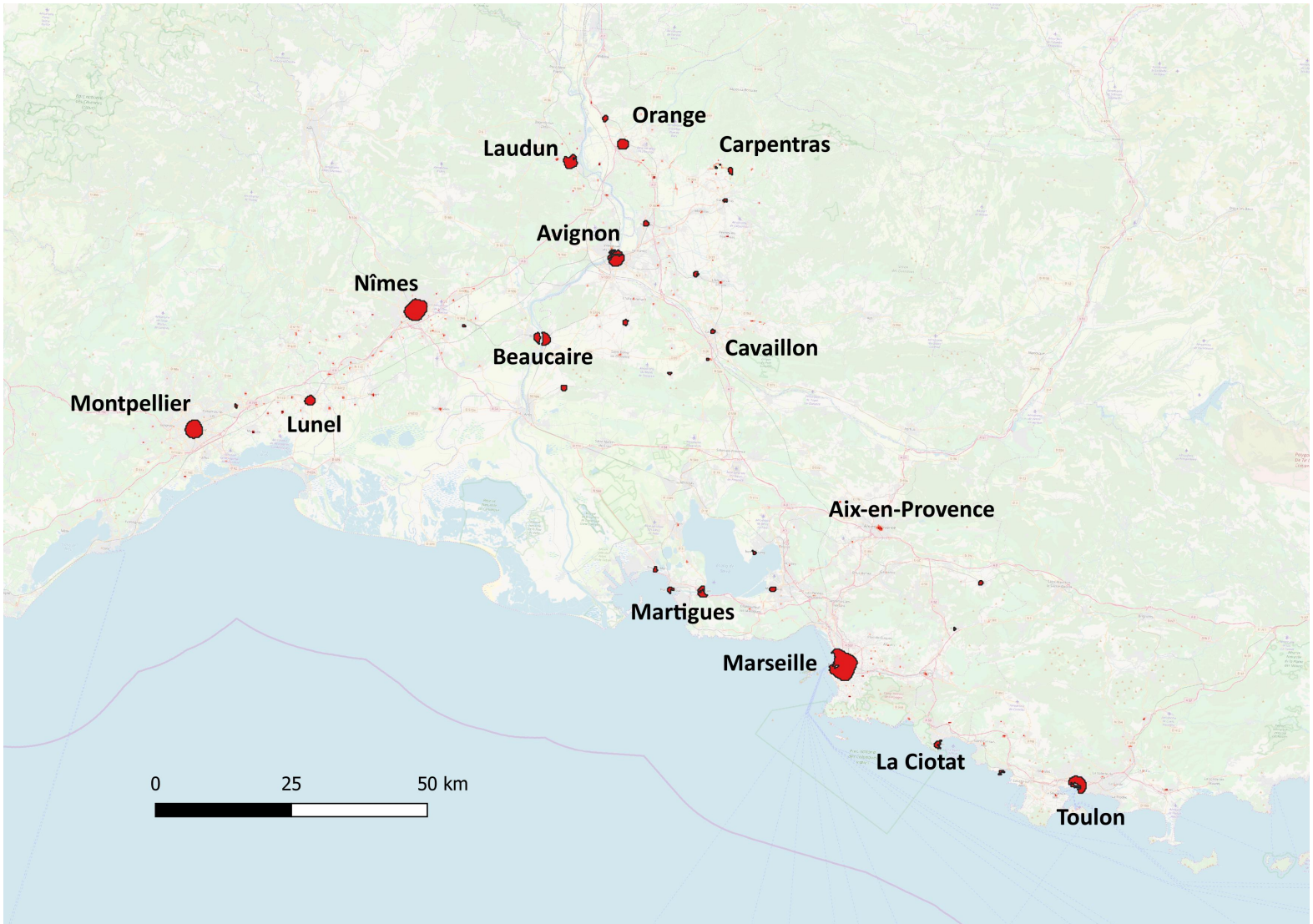
Land use recovery

Machine-learning approach (random forests) to mine historical maps

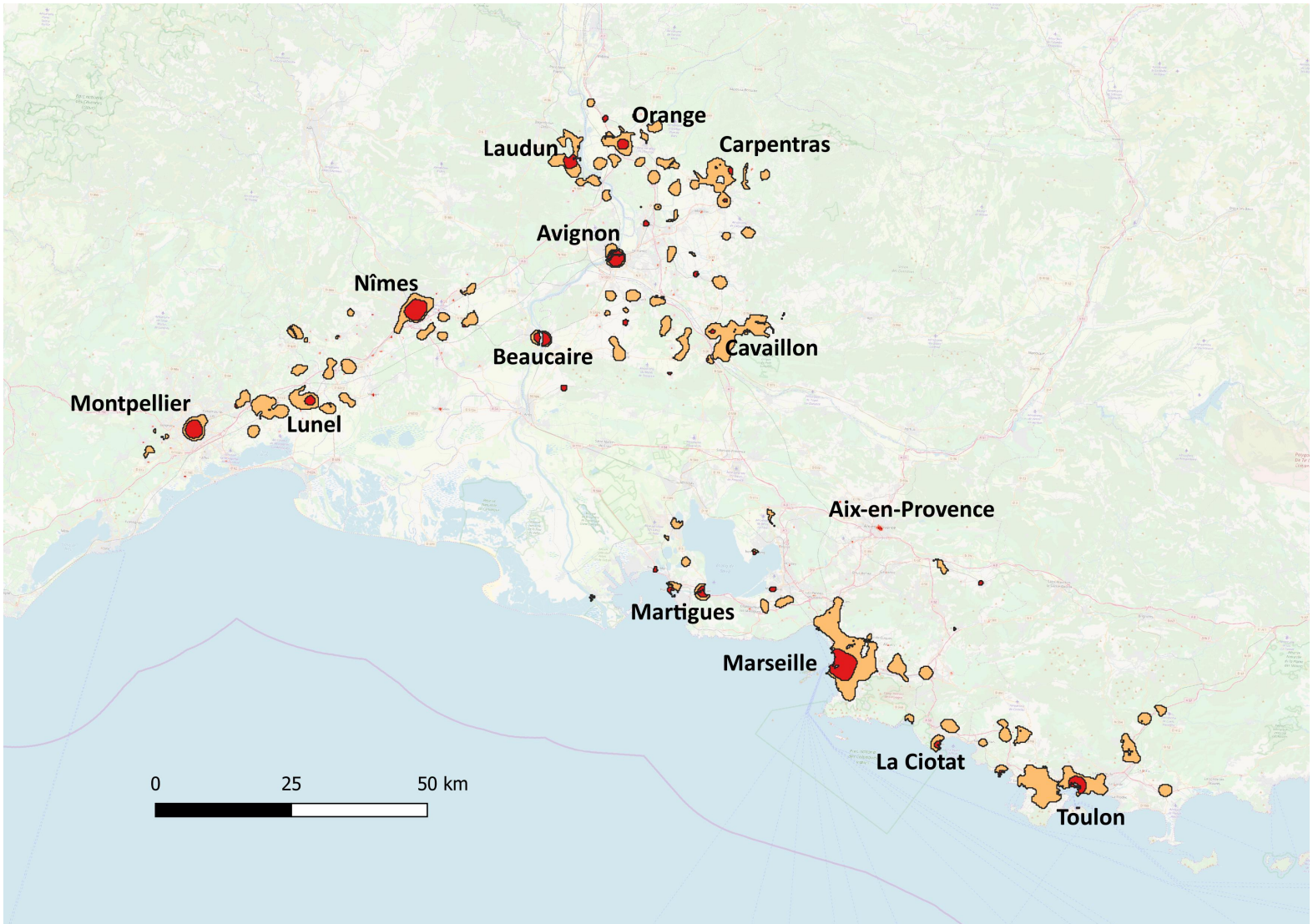
South-East 1850 building gridded data



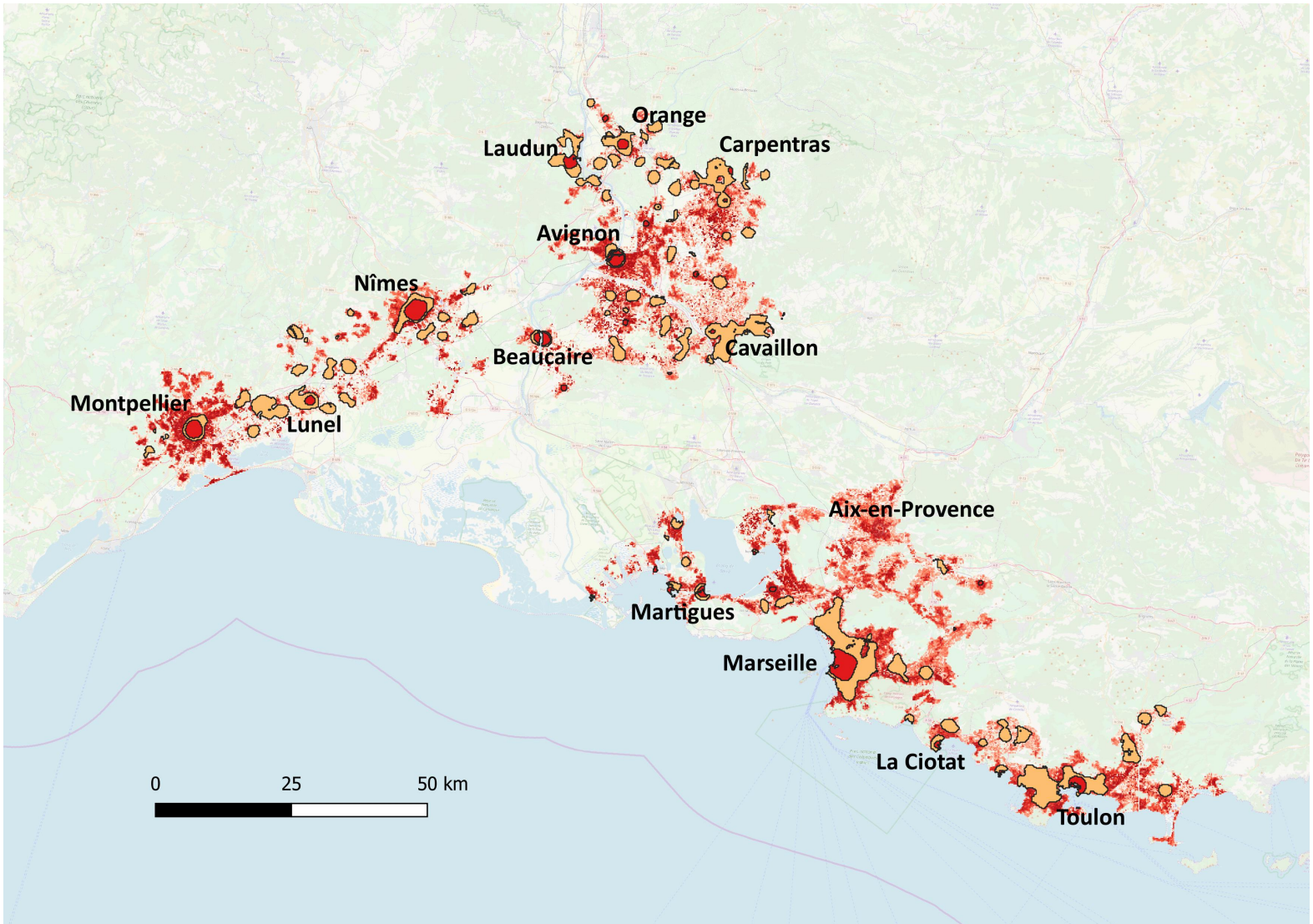
South-East 1850 building gridded data and city cores



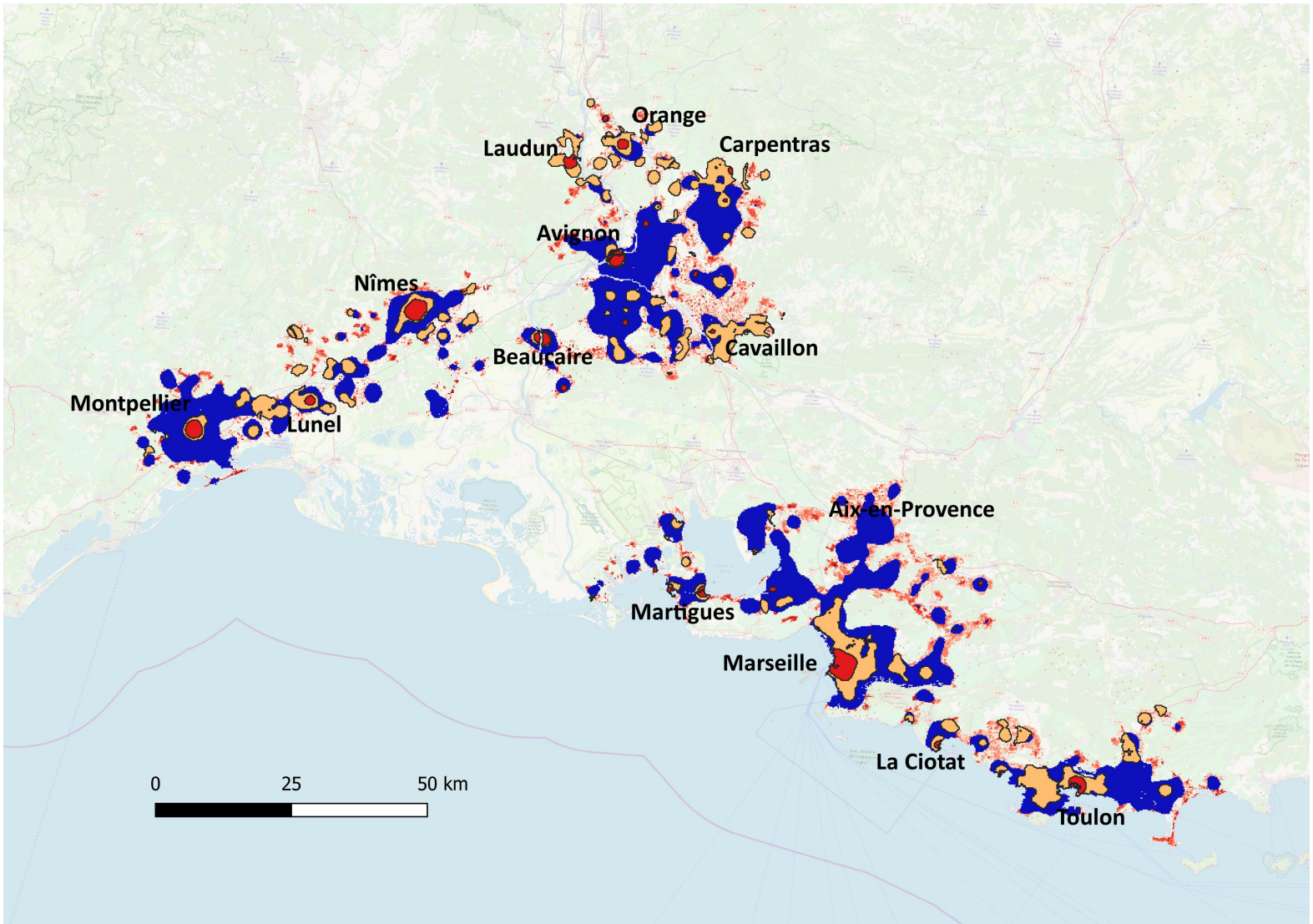
South-East 1850 cities



South-East 1850 cities and 2015 buildings



South-East 1850 cities and 2015 cores



South-East 1850 and 2015 five cities