

High-Res Landsat-8 satellite images for human density prediction

Youcef Kacer

www.github.com/ykacer

2017, January 25th

1 Presentation

2 Landsat-8 imagery

- Earth Covering
- Image Georeferencement

3 Importing Data

- Image query
- Image bands

4 Importing labels

- Importing densities
- Categorize densities to Classification

5 Vegetation index extraction

- NDVI extraction
- NDVI evolution
- Data to Machine Learning

6 Supervised Classification

- K-Nearest Neighbors
- Neural Network
- Support-Vector Machine (Gaussian Kernel)
- Convolutional Neural Network

7 Testing classification

- Data
- K-Nearest Neighbors
- Neural Network

1 Presentation

2 Landsat-8 imagery

- Earth Covering
- Image Georeferencement

3 Importing Data

- Image query
- Image bands

4 Importing labels

- Importing densities
- Categorize densities to Classification

5 Vegetation index extraction

- NDVI extraction
- NDVI evolution
- Data to Machine Learning

6 Supervised Classification

- K-Nearest Neighbors
- Neural Network
- Support-Vector Machine (Gaussian Kernel)
- Convolutional Neural Network

7 Testing classification

- Data
- K-Nearest Neighbors
- Neural Network

- Population census is expensive using conventional methods
 - 180 millions euros in France ([officials,1999](#))
- How High-resolution satellites images could explain human density ?
- How to transform HR satellite images to explain human density ?

1 Presentation

2 Landsat-8 imagery

- Earth Covering
- Image Georeferencement

3 Importing Data

- Image query
- Image bands

4 Importing labels

- Importing densities
- Categorize densities to Classification

5 Vegetation index extraction

- NDVI extraction
- NDVI evolution
- Data to Machine Learning

6 Supervised Classification

- K-Nearest Neighbors
- Neural Network
- Support-Vector Machine (Gaussian Kernel)
- Convolutional Neural Network

7 Testing classification

- Data
- K-Nearest Neighbors
- Neural Network

- Total earth covering defined by (path,row) grid pattern and achieved every 16 days



Landsat-8 grid covering (path, row) for Virginia (USA)

1 Presentation

2 Landsat-8 imagery

- Earth Covering
- Image Georeferencement

3 Importing Data

- Image query
- Image bands

4 Importing labels

- Importing densities
- Categorize densities to Classification

5 Vegetation index extraction

- NDVI extraction
- NDVI evolution
- Data to Machine Learning

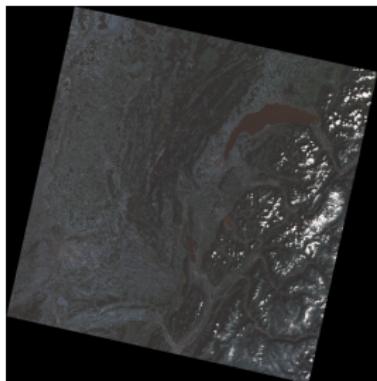
6 Supervised Classification

- K-Nearest Neighbors
- Neural Network
- Support-Vector Machine (Gaussian Kernel)
- Convolutional Neural Network

7 Testing classification

- Data
- K-Nearest Neighbors
- Neural Network

- Landsat-8 images are georeferenced which means each pixel has (x,y) meter coordinates in a certain Projection Coordinates System (ex : *UTM, Lambert IV, Lambert 93, Web Mercator,...*)



Landsat-8 Eastern-France image
path=196, row=028
georeferenced in *UTM* system
(image containing city
Thonon-les-Bains)

UTM corner coordinates

- upper-left (,)
- upper-right (,)
- bottom-right (,)
- bottom-left (,)
- center (,)

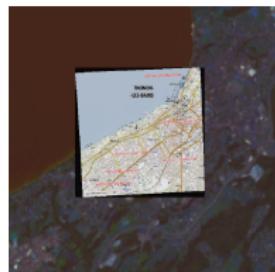
- Landast-8 georeferencement can be checked comparing with another georeferenced source like *IGN* using a *SIG* (open source *QGIS*).



IGN image georeferenced
in *Lambert 93* system



Then IGN image is
transformed to be
georeferenced in *UTM*
system

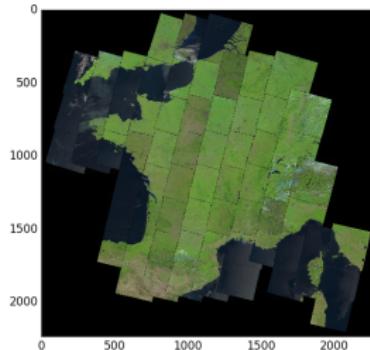


Superposition of IGN
and Landsat-8 images
both georeferenced in
UTM system

- Query images from *U.S geological Survey* website with criterias :
 - cloud covering $\leq 20\%$
 - day acquisition
 - between May, 2013 and September, 2013

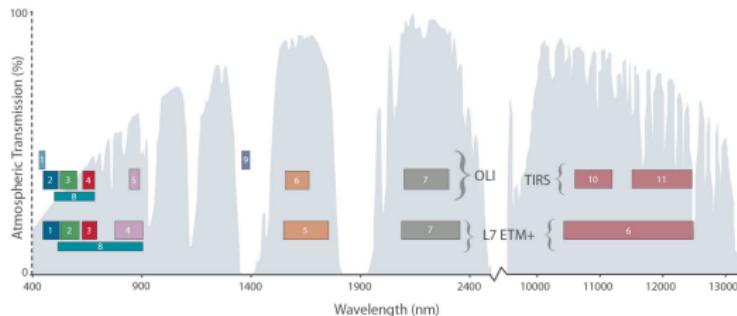
The screenshot shows the USGS Earth Explorer interface. On the left, there's a sidebar with 'Search Criteria' and 'Data Sets'. The 'Search Criteria' section has dropdowns for 'Input' (set to 'BBOX') and 'Extent' (set to 'Lat/Lon'). Below these are three entries for datasets 1, 2, and 3, each with a thumbnail, title, coordinates, and acquisition date. The 'Data Sets' section lists 'LD-OLITRS' and shows a preview of the dataset. A large map of Europe is centered on France, with a red polygon drawn over it. The map includes labels for countries like Belgium, Germany, and Switzerland, and cities like Paris and Lyon.

Polygon selection on *USGS* website



68 resulting datasets georeferenced in *UTM* system

- Landsat-8 dataset is composed of
 - 11 bands (OLI/TRS sensors) + 1 quality band (cloudyness of each pixel)
 - Possible combination of bands to extract information (bands 4 and 5 for vegetation presence)



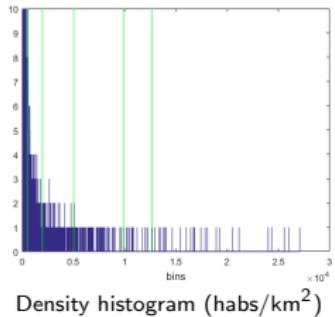
Landsat-8 bands (OLI/TIRS sensors)

- Take cities surfaces and densities from 2013 official census (INSEE)
- Take cities latitude and longitude from Google geolocator (Python API Geopy)

name	latitude (degrees)	longitude (degrees)	surface (km ²)	density (habs/km ²)
Ozan	46.391534	4.915265	6.6	98.3
Cormoranche-sur-Saône	46.240532	4.830863	9	118.9
Paris	48.856614	2.352222	105.4	21153.9
Lyon	45.764043	4.835659	47.87	10117.0
Tours	47.394144	0.68484	34.67	3888.2
Besançon	47.237829	6.024054	65.05	1797.9
...

34190 cities (instances)

- Categorize densities applying clustering (Otsu multi-thresholding)

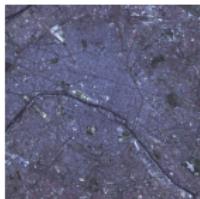


catégorie 1 : density between 0 and 500 habs/km²
 catégorie 2 : density between 500 and 2000 habs/km²
 catégorie 3 : density between 2000 and 5000 habs/km²
 catégorie 4 : density between 5000 and 10000 habs/km²
 catégorie 5 : density between 10000 and 13000 habs/km²
 catégorie 6 : density greater than 13000 habs/km²

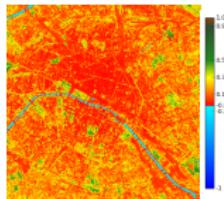
city	latitude (degrees)	longitude (degrees)	surface (km ²)	density (habs/km ²)	density (category)
Ozan	46.391534	4.915265	6.6	98.3	1
Cormoranche-sur-saone	46.240532	4.830863	9	118.9	1
Paris	48.856614	2.352222	105.4	21153.9	6
Lyon	45.764043	4.835659	47.87	10117.0	5
Tours	47.394144	0.68484	34.67	3888.2	3
Besancon	47.237829	6.024054	65.05	1797.9	1
...

- Compute Normalized Difference Vegetation Indice using bands 4 (Red) and 5 (Near-Infra-Red) for each dataset :
- Values between -1 and 1
- ≤ 0 for water, snow and cloud
- = 0 for ground without vegetation
- ≥ 0 for ground vegetation

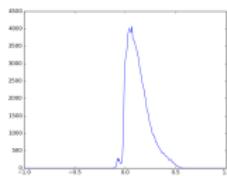
$$NDVI = \frac{NIR - R}{NIR + R}$$



RGB (Paris, May 2013)

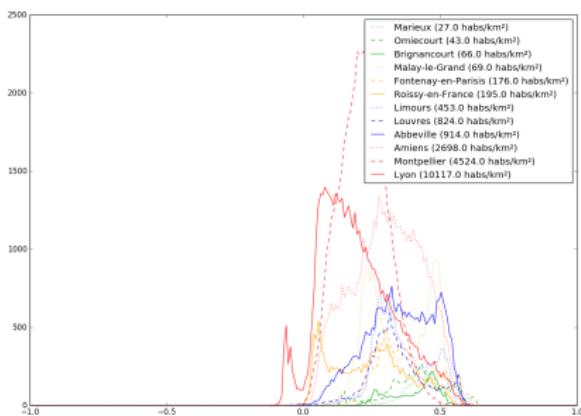


NDVI (Paris, May 2013)

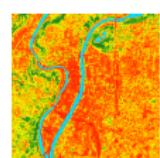
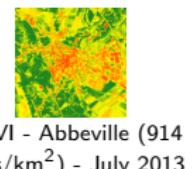
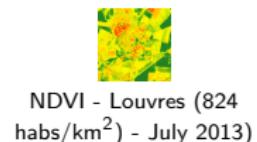


NDVI (1024 bins)-histogram
(Paris, May 2013)

- NDVI histogram could explain human density
 - low ndvi mode for high density (poor vegetation)
 - high ndvi mode for low density (rich vegetation)



NDVI (1024 bins)-histograms (French cities, July 2013)



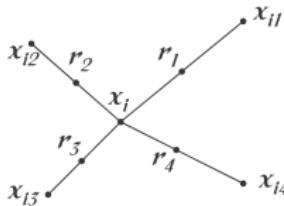
- Explanatory variables : NDVI (1024 bins)-histograms of the 34190 cities
- Predictable variable : density of the 34190 cities for regression
- Predictable variable : density category of the 34190 cities for classification

city	bin-1	bin-2	...	bin-511	bin-512	...	bin-1023	bin-1024	density (hab./km ²)	density (category)
Ozan	0	0	...	1	5	...	0	0	93.8	1
Cormoranche-sur-Saône	0	0	...	1	4	...	0	0	118.9	1
Paris	0	0	...	1953	1815	...	0	0	21153.9	6
Lyon	0	0	...	1099	1032	...	0	0	10117.0	5
Tours	0	0	...	268	238	...	0	0	3888.2	3
Besancon	0	0	...	97	122	...	0	0	1797.9	1
...

Very imbalanced data for classification !

category	number of samples
1	32533
2	1252
3	288
4	78
5	15
6	24

- Overcome imbalanced data by
 - Oversampling minority classes using SMOTE technique (Synthetic Minority Oversampling TECnique)



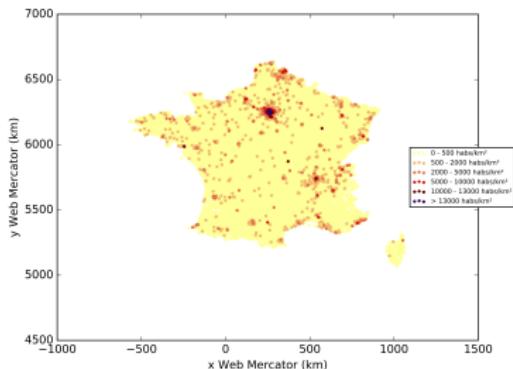
- Run training on obtained balanced data.

Let's $x_i, x_{i1}, x_{i2}, x_{i3}$ and x_{i4} be points of minority class i and y_{jk} the points of majority class j , choose each new sample r_j for class i that maximizes $\sum_{k=1}^{n_j} (y_{jk} - r_j)^2$

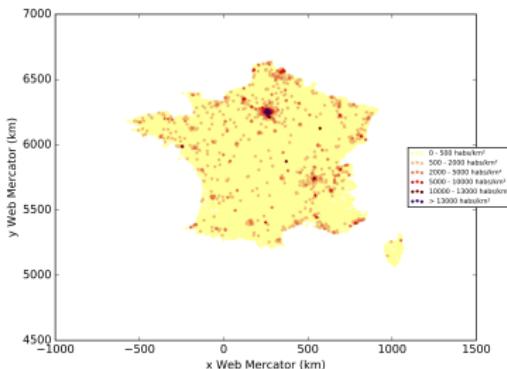
- Best cross-validation (stratified 3-folds) of 94.74% obtained for number of neighbours $k = 5$.

Presentation
 Landsat-8 imagery
 Importing Data
 Importing labels
 Vegetation index extraction
Supervised Classification
 Testing classification

K-Nearest Neighbors
 Neural Network
 Support-Vector Machine (Gaussian Kernel)
 Convolutional Neural Network



France density category - ground truth



France density category - prediction

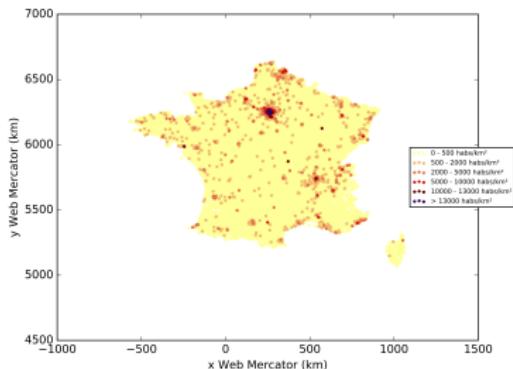
Confusion matrix after refitting						
	1	2	3	4	5	6
1	32533	0	0	0	0	0
2	0	1252	0	0	0	0
3	0	0	288	0	0	0
4	0	0	0	78	0	0
5	0	0	0	0	15	0
6	0	0	0	0	0	24

category	mean score
1	100%
2	100%
3	100%
4	100%
5	100%
6	100%
mean	100%

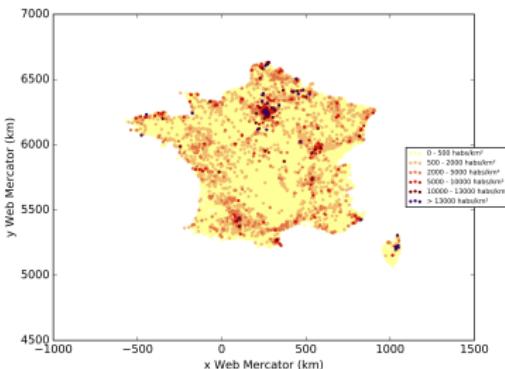
- Overcome imbalanced data by
 - Majoring error by weights to penalize error on minority class, during back-propagation : $w_i = \frac{n_1}{n_i}$.
- Best cross-validation (stratified 3-folds) of 92.26% with SGD learning rate of 0.001 and penalization of 0.001.

Presentation
 Landsat-8 imagery
 Importing Data
 Importing labels
 Vegetation index extraction
Supervised Classification
 Testing classification

K-Nearest Neighbors
 Neural Network
 Support-Vector Machine (Gaussian Kernel)
 Convolutional Neural Network



France density category - ground truth



France density category - prediction

Confusion matrix after refitting						
	1	2	3	4	5	6
1	29743	2281	367	95	24	23
2	207	853	164	23	2	3
3	2	36	236	9	2	3
4	0	0	2	76	0	0
5	0	0	0	0	15	0
6	0	0	0	0	0	24

category	mean score
1	91.42%
2	68.13%
3	91.95%
4	97.64%
5	100%
6	100%
mean	89.82%

- Overcome imbalanced data by increasing regularization parameter C_i for minority classes i : $C_i = 10000/n_i$;
- Cross-validated (stratified 5-folds) gamma coefficient ($\gamma = 0.001$)

Presentation
Landsat-8 imagery
Importing Data
Importing labels
Vegetation index extraction
Supervised Classification
Testing classification

K-Nearest Neighbors
Neural Network
Support-Vector Machine (Gaussian Kernel)
Convolutional Neural Network

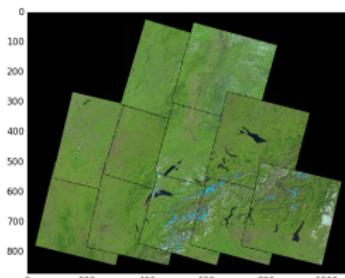
- Take ground truth densities and surfaces for Switzerland (2013), Belgium (2015) and Netherlands (2014).

Switzerland (2013)	
category	number of samples
1	1857
2	428
3	57
4	9
5	1
6	0
total	2352

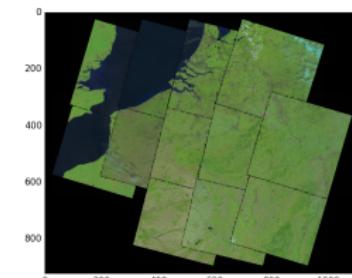
Belgium (2015)	
category	number of samples
1	406
2	154
3	14
4	7
5	1
6	7
total	589

Netherlands (2014)	
category	number of samples
1	234
2	121
3	31
4	1
5	0
6	0
total	388

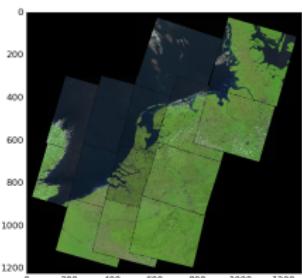
- Take longitudes and latitudes of each city using Google geolocator
- Take corresponding Landsat-8 datasets for each country in the corresponding year (between May and September, day acquisition, cloud covering $\leq 20\%$)



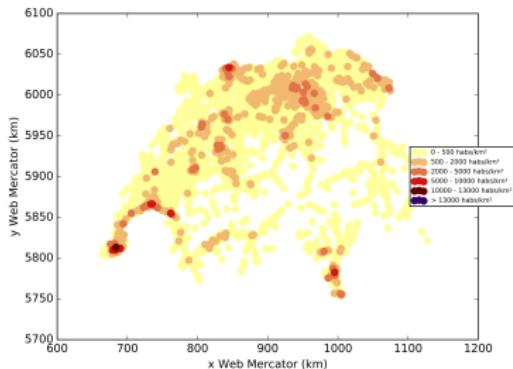
Switzerland datasets georeferenced in UTM system



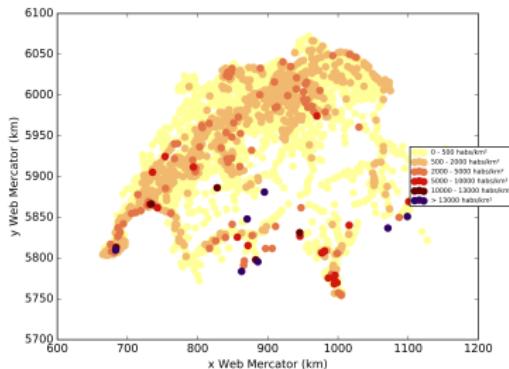
Belgium datasets georeferenced in UTM system



Netherlands datasets georeferenced in UTM system



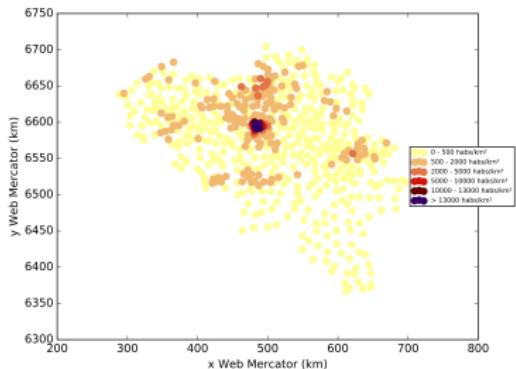
Switzerland density category - ground truth



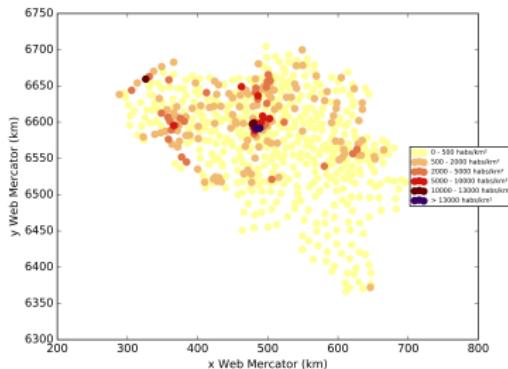
Switzerland density category - prediction

Confusion matrix						
	1	2	3	4	5	6
1	32533	0	0	0	0	0
2	0	1252	0	0	0	0
3	0	0	288	0	0	0
4	0	0	0	78	0	0
5	0	0	0	0	15	0
6	0	0	0	0	0	24

category	mean score
1	100%
2	100%
3	100%
4	100%
5	100%
6	100%
mean	100%

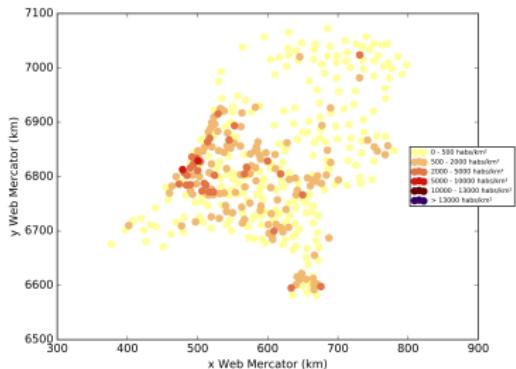


Belgium density category - ground truth

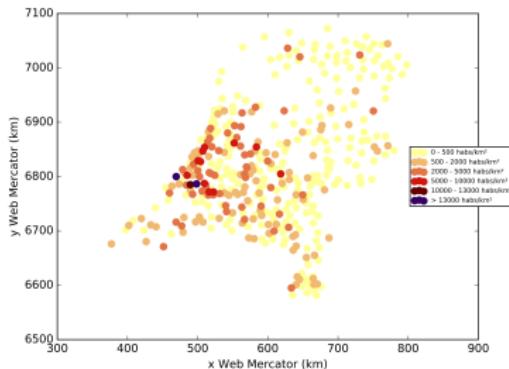


Belgium density category - prediction

Confusion matrix						
	1	2	3	4	5	6
1	32533	0	0	0	0	0
2	0	1252	0	0	0	0
3	0	0	288	0	0	0
4	0	0	0	78	0	0
5	0	0	0	0	15	0
6	0	0	0	0	0	24



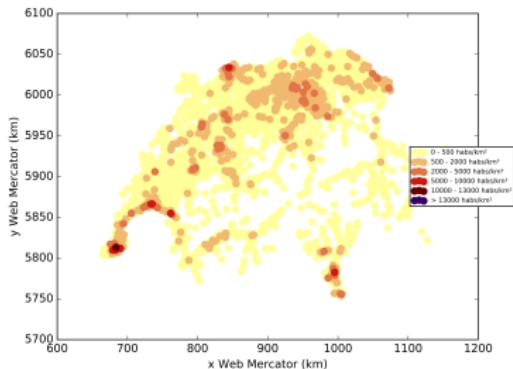
Netherlands density category - ground truth



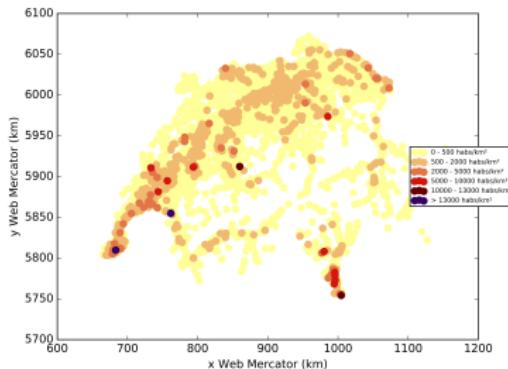
Netherlands density category - prediction

Confusion matrix						
	1	2	3	4	5	6
1	32533	0	0	0	0	0
2	0	1252	0	0	0	0
3	0	0	288	0	0	0
4	0	0	0	78	0	0
5	0	0	0	0	15	0
6	0	0	0	0	0	24

category	mean score
1	100%
2	100%
3	100%
4	100%
5	100%
6	100%
mean	100%

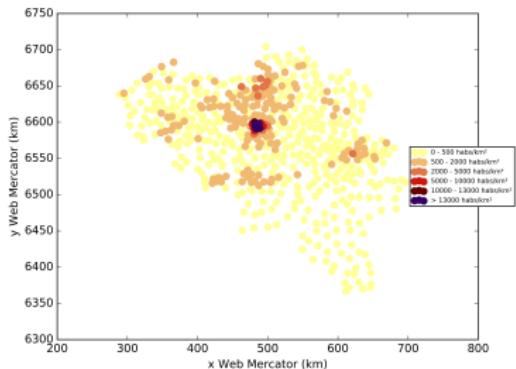


Switzerland density category - ground truth

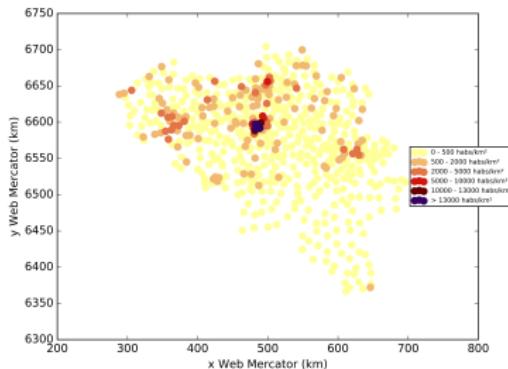


Switzerland density category - prediction

Confusion matrix						
	1	2	3	4	5	6
1	32533	0	0	0	0	0
2	0	1252	0	0	0	0
3	0	0	288	0	0	0
4	0	0	0	78	0	0
5	0	0	0	0	15	0
6	0	0	0	0	0	24



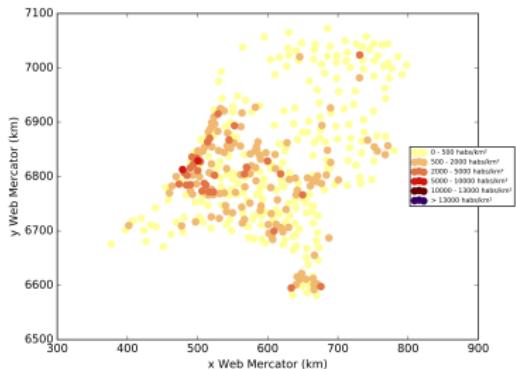
Belgium density category - ground truth



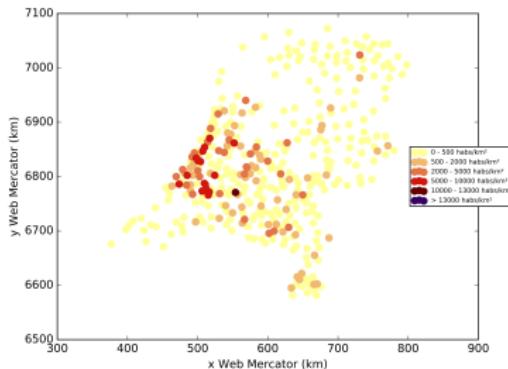
Belgium density category - prediction

Confusion matrix						
	1	2	3	4	5	6
1	32533	0	0	0	0	0
2	0	1252	0	0	0	0
3	0	0	288	0	0	0
4	0	0	0	78	0	0
5	0	0	0	0	15	0
6	0	0	0	0	0	24

category	mean score
1	100%
2	100%
3	100%
4	100%
5	100%
6	100%
mean	100%



Netherlands density category - ground truth



Netherlands density category - prediction

Confusion matrix						
	1	2	3	4	5	6
1	32533	0	0	0	0	0
2	0	1252	0	0	0	0
3	0	0	288	0	0	0
4	0	0	0	78	0	0
5	0	0	0	0	15	0
6	0	0	0	0	0	24