

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

Youcef Kacer

[www.github.com/ykacer](https://www.github.com/ykacer)

2017, January 25<sup>th</sup>

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

- Neural Network

7 Testing classification

- Data
- 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

- Neural Network

## 7 Testing classification

- Data
- 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

- Neural Network

7 Testing classification

- Data
- 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

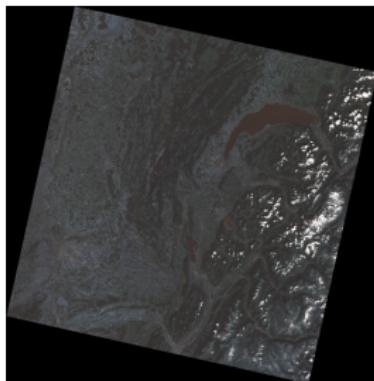
## 6 Supervised Classification

- Neural Network

## 7 Testing classification

- Data
- 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)

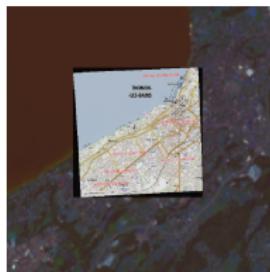
- Landast-8 georeferencement can be checked comparing with another georeferenced source like *IGN* using a *SIG* (open source *QGIS*).



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

Search Criteria    Data Sets    Additional Criteria    Results

4. Search Results

If you selected more than one data set to search, use the dropdown to see the search results for each specific data set.

Show Results Controls

Data Set [Click here to export your results](#)

LD-OILTIRS

1. [Landsat-8 - Province 1 - NLCD - Land](#)

Entity ID: C098602091273.GN08  
Cloud Cover: 11% NLCD 2.0  
Acquisition Date: 30 SEP 13  
Pedro200  
Row25

2. [Landsat-8 - Province 1 - NLCD - Land](#)

Entity ID: C098602091271.GN08  
Cloud Cover: 11% NLCD 2.0  
Acquisition Date: 01 SEP 13  
Pedro200  
Row24

3. [Landsat-8 - Province 1 - NLCD - Land](#)

Entity ID: C098602091271.GN08  
Cloud Cover: 11% NLCD 2.0  
Acquisition Date: 21 SEP 13  
Pedro193  
Row25

4. [Landsat-8 - Province 1 - NLCD - Land](#)

Entity ID: C098602091273.GN08  
Cloud Cover: 11% NLCD 2.0  
Acquisition Date: 13 SEP 13

5. [Landsat-8 - Province 1 - NLCD - Land](#)

Entity ID: C098602091273.GN08  
Cloud Cover: 11% NLCD 2.0  
Acquisition Date: 13 SEP 13

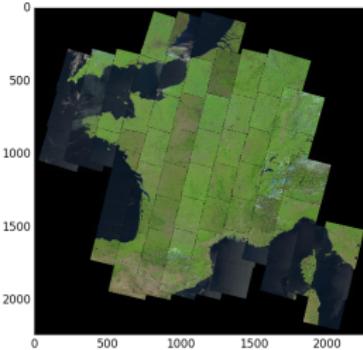
6. [Landsat-8 - Province 1 - NLCD - Land](#)

Entity ID: C098602091273.GN08  
Cloud Cover: 11% NLCD 2.0  
Acquisition Date: 13 SEP 13

[View more details](#) | [Submit Standing Request](#)

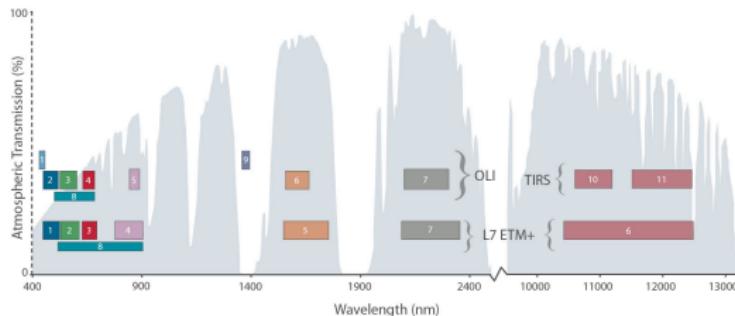
This on-line Google map is not for download; it is to be used as a guide for reference and search purposes only.

Polygon selection on *USGS* website



68 resulting datasets georeferenced in  
Web Mercator 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)



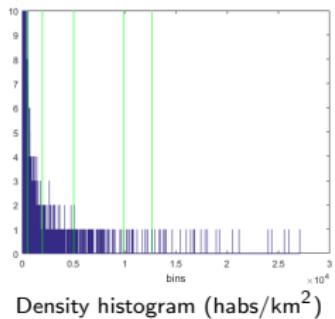
Landsat-8 bands (OLI/TIRS sensors)

- 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 <sup>2</sup> )	density (habs/km <sup>2</sup> )
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)

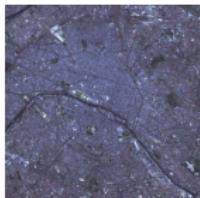


catégorie 1 : density between 0 and 500 habs/km<sup>2</sup>  
 catégorie 2 : density between 500 and 2000 habs/km<sup>2</sup>  
 catégorie 3 : density between 2000 and 5000 habs/km<sup>2</sup>  
 catégorie 4 : density between 5000 and 10000 habs/km<sup>2</sup>  
 catégorie 5 : density between 10000 and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

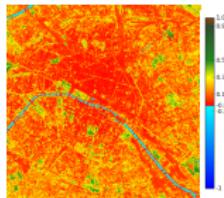
city	latitude (degrees)	longitude (degrees)	surface (km <sup>2</sup> )	density (habs/km <sup>2</sup> )	density (category)
Ozan	46.391534	4.915265	6.6	98.3	1
Cormoranche-sur-saone	46.240532	4.830863	9	118.9	1
Paris	48.856614	2.352222	105.4	21153.9	6
Lyon	45.764043	4.835659	47.87	10117.0	5
Tours	47.394144	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
- $\leq 0$  for water, snow and cloud
- = 0 for ground without vegetation
- $\geq 0$  for ground vegetation

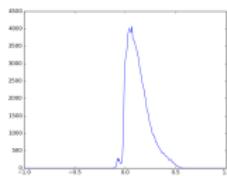
$$NDVI = \frac{NIR - R}{NIR + R}$$



RGB (Paris, May 2013)

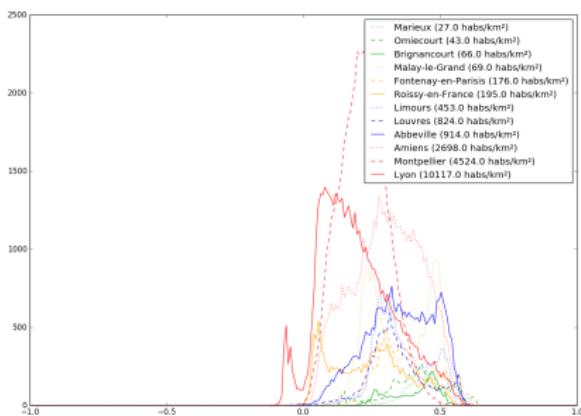


NDVI (Paris, May 2013)

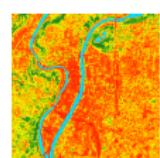
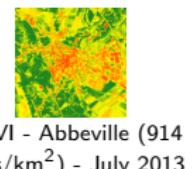
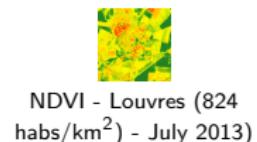


NDVI (1024 bins)-histogram  
(Paris, May 2013)

- 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)



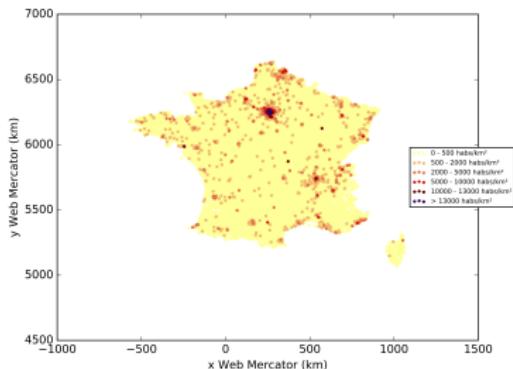
- Explanatory variables : NDVI (1024 bins)-histograms of the 34190 cities
- Predictable variable : density of the 34190 cities for regression
- Predictable variable : density category of the 34190 cities for classification

city	bin-1	bin-2	...	bin-511	bin-512	...	bin-1023	bin-1024	density (hab./km <sup>2</sup> )	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
...	...	...	...	...	...	...	...	...	...	...

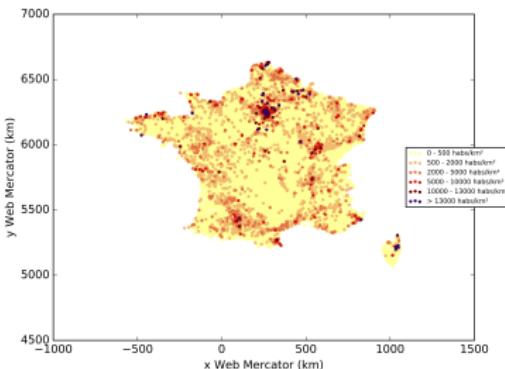
Very imbalanced data for classification !

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

- Overcome imbalanced data by
  - 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 one layer (1200 neurons), SGD learning rate of 0.001 and penalization of 0.001.



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	90.51%

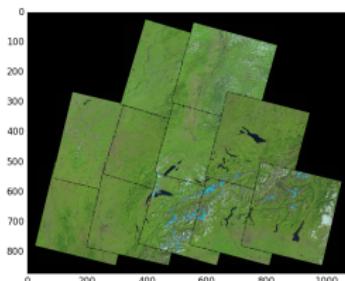
- Take ground truth densities and surfaces for Switzerland (2013), Belgium (2015) and Netherlands (2014).

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

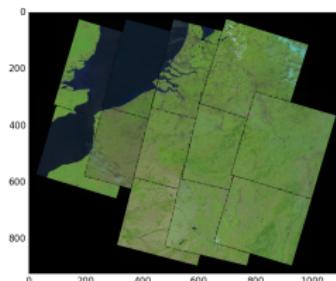
Belgium (2015)	
category	number of samples
1	406
2	154
3	14
4	7
5	1
6	7
total	589

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

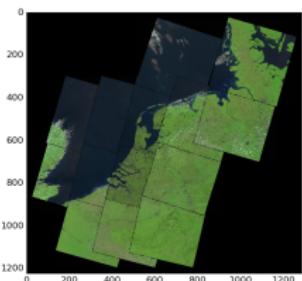
- Take longitudes and latitudes of each city using Google geolocator
- Take corresponding Landsat-8 datasets for each country in the corresponding year (between May and September, day acquisition, cloud covering  $\leq 20\%$ )



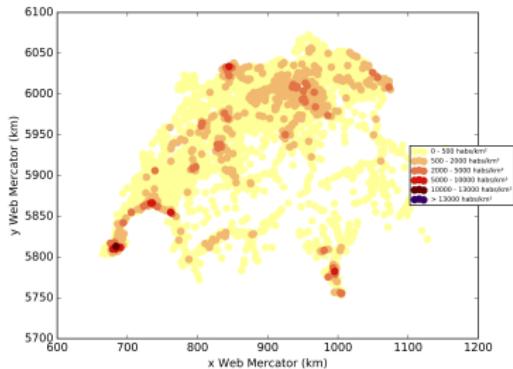
Switzerland datasets georeferenced in Web Mercator system



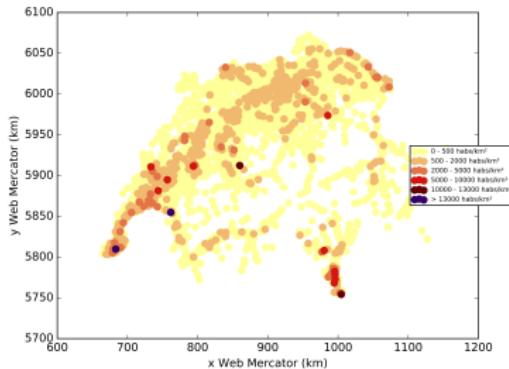
Belgium datasets georeferenced in Web Mercator system



Netherlands datasets georeferenced in Web Mercator system



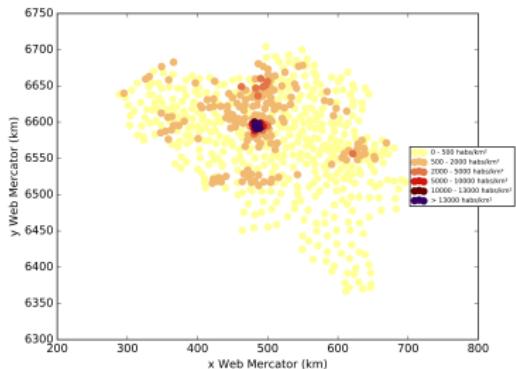
Switzerland density category - ground truth



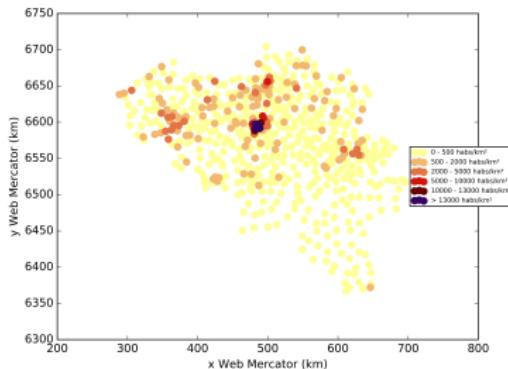
Switzerland density category - prediction

Confusion matrix						
	1	2	3	4	5	6
1	1601	244	7	4	1	0
2	158	249	18	2	1	0
3	2	36	16	3	0	0
4	1	0	5	1	0	2
5	1	0	0	0	0	0
6	0	0	0	0	0	0

category	mean score
1	86.21%
2	58.18%
3	28.07%
4	11.11%
5	0%
6	-
total	79.38%

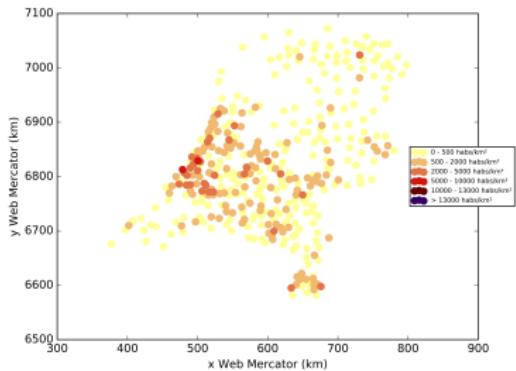


Belgium density category - ground truth

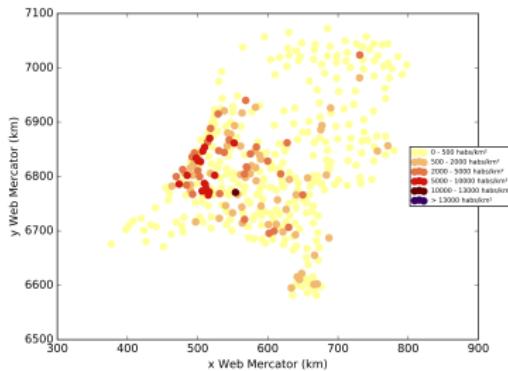


Belgium density category - prediction

Confusion matrix						
	1	2	3	4	5	6
1	363	41	2	0	0	0
2	73	60	19	2	0	0
3	1	6	6	1	0	0
4	1	0	2	2	2	0
5	0	0	1	0	0	0
6	0	2	0	1	0	4



Netherlands density category - ground truth



Netherlands density category - prediction

Confusion matrix						
	1	2	3	4	5	6
1	277	5	2	0	0	0
2	54	47	14	5	1	0
3	2	6	14	9	0	0
4	0	0	1	1	0	0
5	0	0	0	0	0	0
6	0	0	0	0	0	0