

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

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

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

7 Testing classification

- 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

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

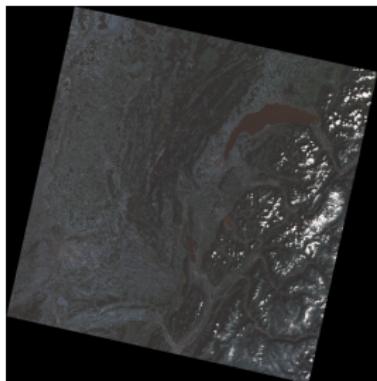
7 Testing classification

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



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

- 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)

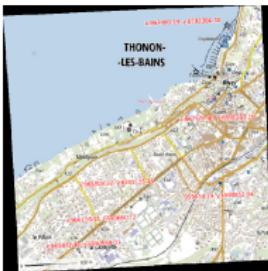
UTM corner coordinates

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

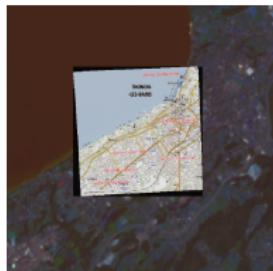
- Landast-8 georeferencing can be checked comparing with another georeferenced source like *IGN* using a *SIG* (open source *QGIS*).



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
 - cloud covering $\leq 20\%$
 - day acquisition
 - between May, 2013 and September, 2013

Search Criteria Summary (Show)

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 Result Controls

Data Set

LS-OLETRS

1. Early LS-C20130501T011711L000 Coordinates: 40.000000, 2.000000 Acquisition Date: 2013-05-13 Ped:200 Resolution: Rows25

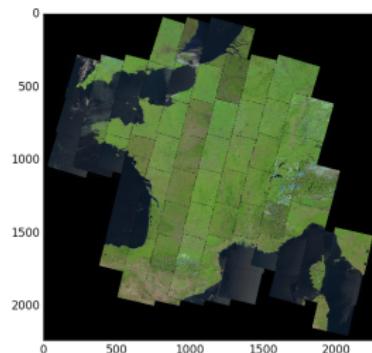
2. Early LS-C20130610T011711L000 Coordinates: 40.000000, 2.000000 Acquisition Date: 2013-06-13 Ped:200 Resolution: Rows25

3. Early LS-C20130701T011711L000 Coordinates: 40.000000, 2.000000 Acquisition Date: 2013-07-13 Ped:200 Resolution: Rows25

4. Early LS-C20130801T011711L000 Coordinates: 40.000000, 2.000000 Acquisition Date: 2013-08-13 Ped:200 Resolution: Rows25

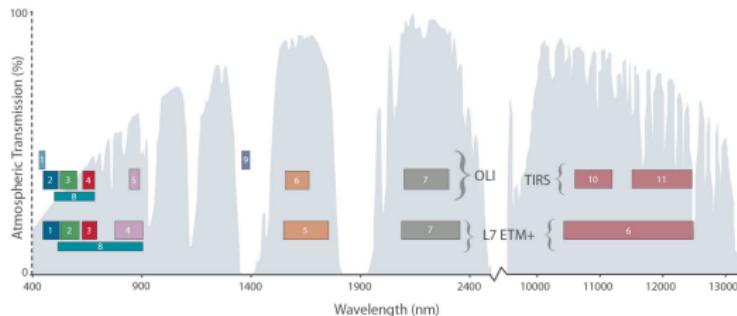
View More Details | Submit Standing Request

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)



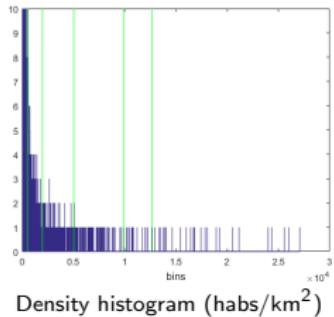
Landsat-8 bands (OLI/TIRS sensors)

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

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	10459.8
Tours	47.394144	0.68484	34.67	3888.2
Besancon	47.237829	6.024054	65.05	1797.9
...

34190 cities (instances)

- Categorize densities applying clustering (Otsu multi-thresholding)

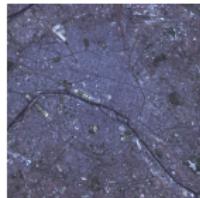


catégory 1 : density between 0 and 500
 catégory 2 : density between 500 and 2000
 catégory 3 : density between 2000 and 5000
 catégory 4 : density between 5000 and 10000
 catégory 5 : density between 10000 and 13000
 catégory 6 : density greater than 13000

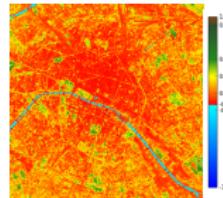
name	latitude (degrees)	longitude (degrees)	surface (km ²)	density (habs/km ²)	density (category)
Ozan	46.391534	4.915265	6.6	98.3	1
Cormoranche-sur-saone	46.240532	4.830863	9	118.9	1
Paris	48.856614	2.352222	105.4	21153.9	6
Lyon	45.764043	4.835659	47.87	10459.8	5
Tours	47.394144	0.68484	34.67	3888.2	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 ground vegetation

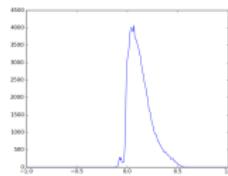
$$NDVI = \frac{NIR - R}{NIR + R}$$



RGB (Paris, May 2013)

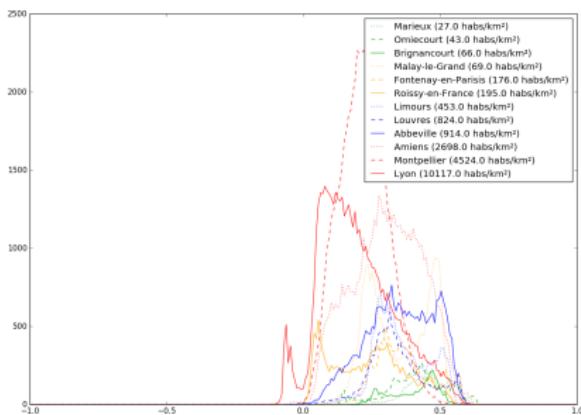


NDVI (Paris, May 2013)

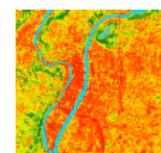
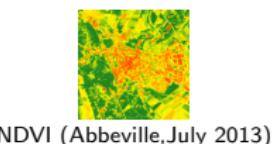


NDVI (1024 bins)-histogram
(Paris, May 2013)

- 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

nom	bin-1	bin-2	...	bin-511	bin-512	...	bin-1023	bin-1024	densité (hab./km ²)	density (category)
Ozan	0	0	...	1	5	...	0	0	93.0	1
Cormoranche-sur-Saône	0	0	...	1	4	...	0	0	107.0	1
Paris	0	0	...	1953	1815	...	0	0	21288.0	6
Lyon	0	0	...	1099	1032	...	0	0	460.0	5
Tours	0	0	...	268	238	...	0	0	3888.0	3
Besancon	0	0	...	97	122	...	0	0	1797.0	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 increasing regularization parameter C_i for minority classes $i : C_i = 10000/n_i$;
- Cross-validated (stratified 5-folds) gamma coefficient ($\gamma = 0.001$)

	1	2	3	4	5	6
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0
6	0	0	0	0	0	0

confusion matrix after refitting

best cross-validation score

93%

mean score per class
after refitting

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

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

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

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

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

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

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