**Urban heat island**

[Overview of Sentinel-2 Mission](https://sentiwiki.copernicus.eu/web/s2-mission)

Sentinel-2 is a European wide-swath, high-resolution, multi-spectral imaging mission. The full mission specification of the twin satellites flying in the same orbit but phased at 180°, is designed to give a high revisit frequency of 5 days at the Equator.

Each of the satellites in the Sentinel-2 mission carries a single payload: **the optical Multi-Spectral Instrument (MSI) that samples 13 spectral bands: four bands at 10 m, six bands at 20 m and three bands at 60 m spatial resolution.** The orbital swath width is 290 km.

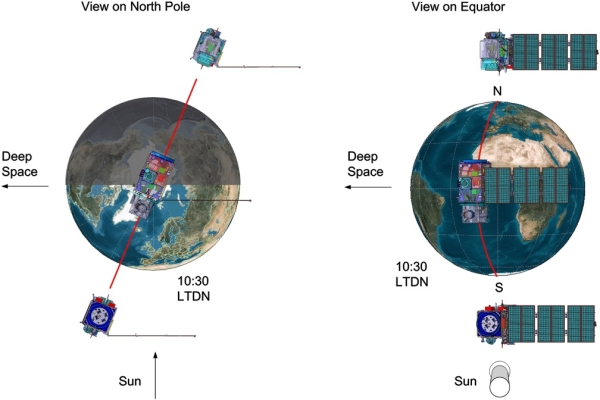
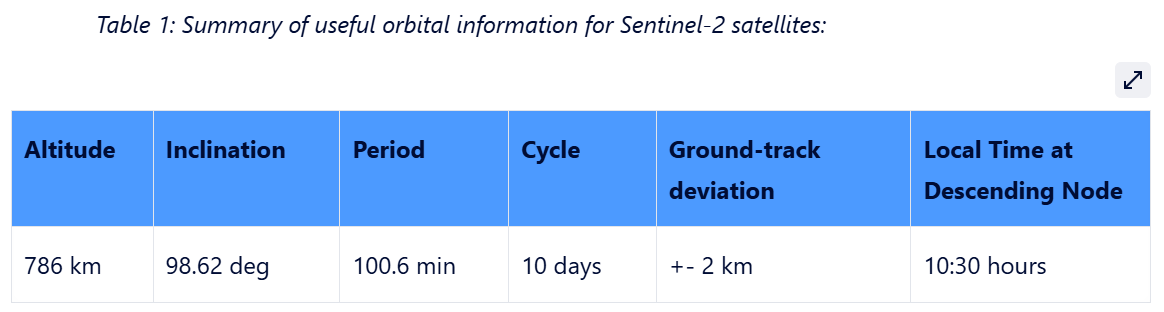


Figure 1: The Twin-Satellite Sentinel-2 Orbital Configuration [Credits: Astrium GmbH]

The Sentinel-2 twin satellites carry on the legacy of SPOT and LANDSAT by continuing to provide similar types of image data and contributing to ongoing multispectral observations. These satellites are used to support a variety of services and applications offered by Copernicus, including land management, agriculture, forestry, disaster control, humanitarian relief operations, risk mapping, and security concerns.

The Sentinel-2 mission consists of two identical satellites operating together, Sentinel-2B (launched in 2017) and Sentinel-2C (launched in 2024) that were launched using the European VEGA launcher. Each of these satellites weighs approximately 1.2 tonnes.



The following list of common bands can be loaded by the Open Data Cube (ODC)

B01 = Coastal Aerosol = 60m

B02 = Blue = 10m

B03 = Green = 10m

B04 = Red = 10m

B05 = Red Edge (704 nm) = 20m

B06 = Red Edge (740 nm) = 20m

B07 = Red Edge (780 nm) = 20m

B08 = NIR (833 nm) = 10m

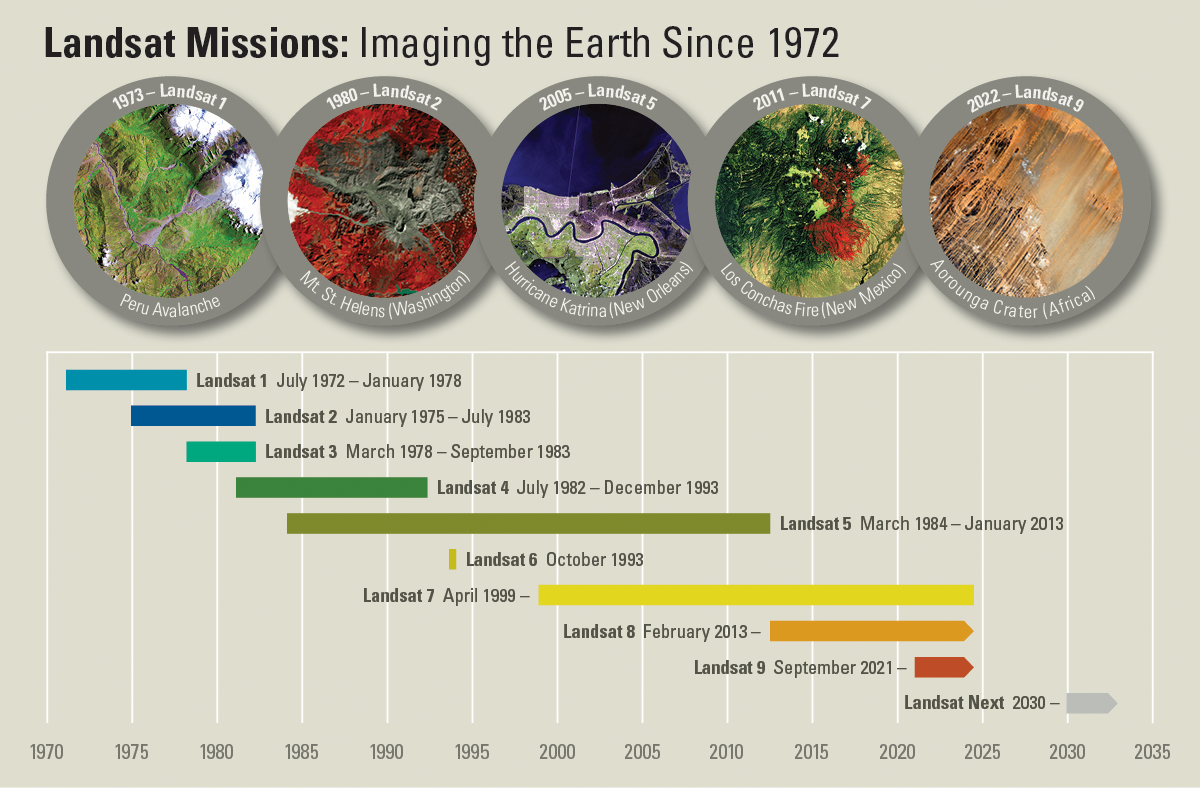
B8A = NIR (narrow 864 nm) = 20m

B11 = SWIR (1.6 um) = 20m

B12 = SWIR (2.2 um) = 20m

[Landsat Collection 2 | U.S. Geological Survey](https://www.usgs.gov/landsat-missions/landsat-collection-2)

The Landsat Missions are comprised of Earth-observing operational satellites that carry remote sensors to collect data and image our planet as a part of the U.S. Geological Survey (USGS) National Land Imaging (NLI) Program. Products generated from the imagery acquired by the sensors carried on the Landsat satellites are hosted at the USGS Earth Resources Observation and Science (EROS) Center in Sioux Falls, South Dakota.



Landsat Collection 2, the second major reprocessing effort on the Landsat archive, resulted in several data product improvements that applied advancements in data processing, algorithm development, and data access and distribution capabilities.

Landsat Collection 2 contains Level-1 data from Landsats 1-9, and Level-2 and Level-3 science products from Landsats 4-9.

**Landsat 8 Satellite Orbit Facts**

Orbits the Earth in a sun-synchronous, near-polar orbit (98.2 degrees inclination)

Achieved an altitude of 705 km (438 mi)

Completes one Earth orbit every 99 minutes

Has a 16-day repeat cycle with an equatorial crossing time of 10:00 a.m. +/- 15 minutes

Acquires about 740 scenes a day on the Worldwide Reference System-2 (WRS-2) path/row system, with a swath overlap (or sidelap) varying from 7 percent at the equator to a maximum of approximately 85 percent at extreme latitudes

Landsat 8 Instruments

Landsat 8 carries two sensors. The Operational Land Imager sensor is built by Ball Aerospace & Technologies Corporation. The Thermal Infrared Sensor is built by NASA Goddard Space Flight Center.

Operational Land Imager (OLI)

Nine spectral bands, including a pan band:

Band 1 Coastal Aerosol (0.43 - 0.45 µm) 30 m

Band 2 Blue (0.450 - 0.51 µm) 30 m

Band 3 Green (0.53 - 0.59 µm) 30 m

Band 4 Red (0.64 - 0.67 µm) 30 m

Band 5 Near-Infrared (0.85 - 0.88 µm) 30 m

Band 6 SWIR 1(1.57 - 1.65 µm) 30 m

Band 7 SWIR 2 (2.11 - 2.29 µm) 30 m

Band 8 Panchromatic (PAN) (0.50 - 0.68 µm) 15 m

Band 9 Cirrus (1.36 - 1.38 µm) 30 m

**Articles synthesis and ideas:**

**Towards Systematic Prediction of Urban Heat Islands: Grounding Measurements, Assessing Modeling Techniques**

**Jackson Voelkel \* and Vivek Shandas**

**10 June 2017**

The random forest technique best predicts temperature

Building data -> each pixel has the height of the building

gNDVI =

Topographic data of NYC :

Building Footprints from NYC Open Data

**En fait non ce fichier st mieux car buildings + leur elevation**<https://data.cityofnewyork.us/City-Government/Building-Footprints-Map-/3g6p-4u5s>

**More documentation on :**[**nyc-geo-metadata/Metadata/Metadata\_BuildingFootprints.md at main · CityOfNewYork/nyc-geo-metadata · GitHub**](https://github.com/CityOfNewYork/nyc-geo-metadata/blob/main/Metadata/Metadata_BuildingFootprints.md)

**HEIGHTROOF The height of the roof above the ground elevation, not height above sea level.**

**GEOMETRO : MULTIPOLYGON with coordinates of the buikding in ESPG 4326**

**"MULTIPOLYGON (((-73.96664570466969 40.62599676998366, -73.96684846176461 40.625977490862574, -73.96685938726297 40.62604419372411, -73.96661621040211 40.62606731716107, -73.96660638332114 40.626007324369795, -73.96664680403327 40.626003480977275, -73.96664570466969 40.62599676998366)))",,**

**Random forest Regression**

predicts continuous values by averaging the results of multiple decision trees, each trained on a random subset of the data

Regression algorithms

Linear Regression

* Too simple here

Decision Tree Regressor

r2\_score : 0.998088400925495

test : 0.7565291342961052

🡪 Overfit

Support Vector Regression

Lasso Regression

Random Forest

**Open session**:

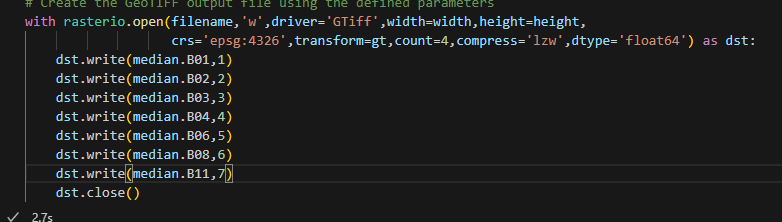
Sentinel 2 : water, green space, building level

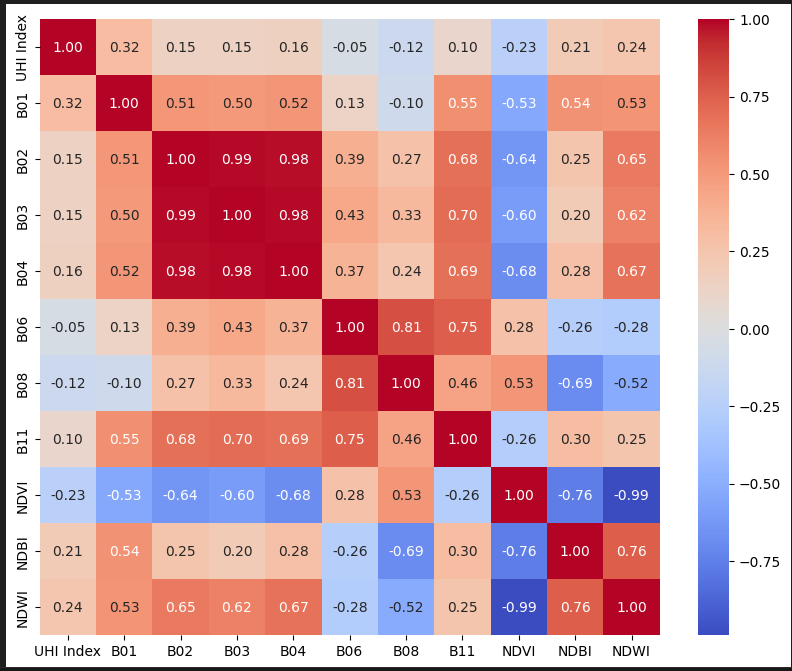
+ building footprint

Compuer : 4 core by 16go

R square score

S2\_sample\_median.tiff :





Scores

1. Initial :

Train : 0.9113116298079081

Test : 0.37068029346827613

Submission: 0 .5108

1. Ajout median data:

Train: 0.9160125448309037

Test: 0.3703900019838272

Submission: 0.519

1. Adding NDBI

Train : 0.9103208833374024

Test : 0.3259994501375395

Submission 0.4929

1. Adding the Buffer of 50 meters radius with only B1, B6 and NDVI

Train : r2\_score : 0.9402649781764467

test : 0.5735499214039286

Submission : 0.557

1. Adding LST data with all of the previous data

Train : r2\_score : 0.9498176020300231

test : 0.6441417859260559

Submission : 0.6232

1. Adding all the buffers

Train: 0.9825531182615634

Test : 0.890143460038042

Submission: 0.881

1. Adding the building data

Train: 0.9916342242218535

Test : 0.9404707702309756

Submission: 0.9394

Same avec 200 estimateurs

0.9920008520517642

test : 0.9413333345956438

Same avec un test size de 0.1

r2\_score : 0.9939148672055327

test : 0.9550613373886464

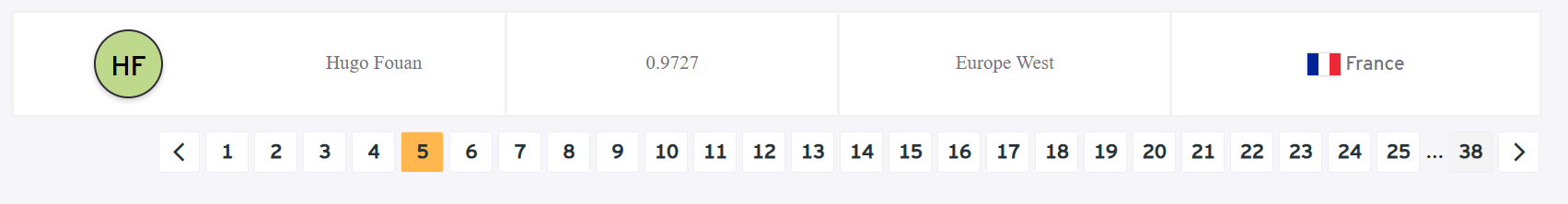
Submission: 0.9544

Final with all the data in training set:

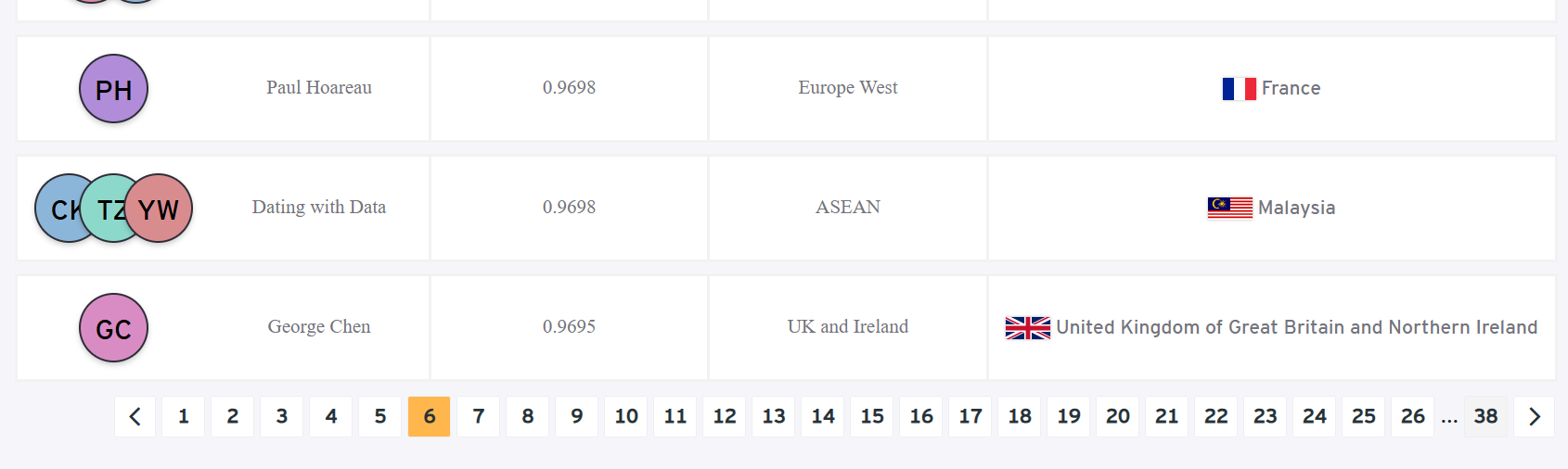
R2\_score: 0.9944609719276374

Submission: 0.9596

**Ranking : 93/380**







Test des different scaler:

StandardScaler()

r2\_score : 0.9825531182615634

test : 0.890143460038042

B03\_buffer\_1000 0.177969

B06\_buffer\_1000 0.054082

LST 0.050683

B11\_buffer\_1000 0.043970

NDBI\_buffer\_1000 0.020927

dtype: float64

MinMaxScaler()

r2\_score : 0.982431233621886

test : 0.890292986591134

B03\_buffer\_1000 0.177877

B06\_buffer\_1000 0.054298

LST 0.050719

B11\_buffer\_1000 0.044096

NDBI\_buffer\_1000 0.021093

dtype: float64

RobustScaler()

r2\_score : 0.9824907038989972

test : 0.8900386992197461

B03\_buffer\_1000 0.178004

B06\_buffer\_1000 0.054054

LST 0.050673

B11\_buffer\_1000 0.044081

NDBI\_buffer\_1000 0.021178

dtype: float64

MaxAbsScaler()

r2\_score : 0.9825276957299419

test : 0.8900534021638058

B03\_buffer\_1000 0.177902

B06\_buffer\_1000 0.054249

LST 0.050688

B11\_buffer\_1000 0.043817

NDBI\_buffer\_1000 0.021259

dtype: float64

Normalizer()

r2\_score : 0.9708677381927338

test : 0.816748692029292

NDWI\_buffer\_1000 0.081385

LST 0.052482

NDVI\_buffer\_1000 0.046263

NDBI\_buffer\_1000 0.034682

NDWI\_buffer\_200 0.031981

dtype: float64

QuantileTransformer()

r2\_score : 0.9824894230751577

test : 0.8911079146129939

B03\_buffer\_1000 0.178110

B06\_buffer\_1000 0.054210

LST 0.050785

B11\_buffer\_1000 0.044003

NDBI\_buffer\_1000 0.021142

dtype: float64

PowerTransformer()

r2\_score : 0.9825204736395793

test : 0.8909894347212852

B03\_buffer\_1000 0.177922

B06\_buffer\_1000 0.054046

LST 0.050650

B11\_buffer\_1000 0.043837

NDBI\_buffer\_1000 0.020821

dtype: float64

PCA indique que le mieux est a 14 composants

FINAL SCORE : 0.9596