Session #3: Qgis Georeferencing & data entry

GIS workshop
TSE, 2022/2023

Eric Mermet, eric.mermet@tse-fr.eu

Objectives of the TP

- Download data

- Understanding georeferencing
    - image coding
    - Raster data
    - Repositories

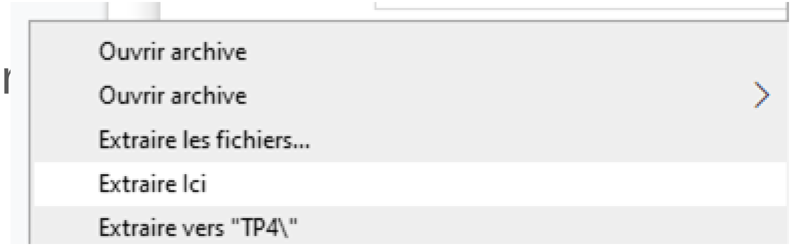
- Perform georeferencing
    - Checkpoints
    - Warp algorithm
    - Evaluate the error

For this TP, it will be very useful for you to use a mouse with wheel!

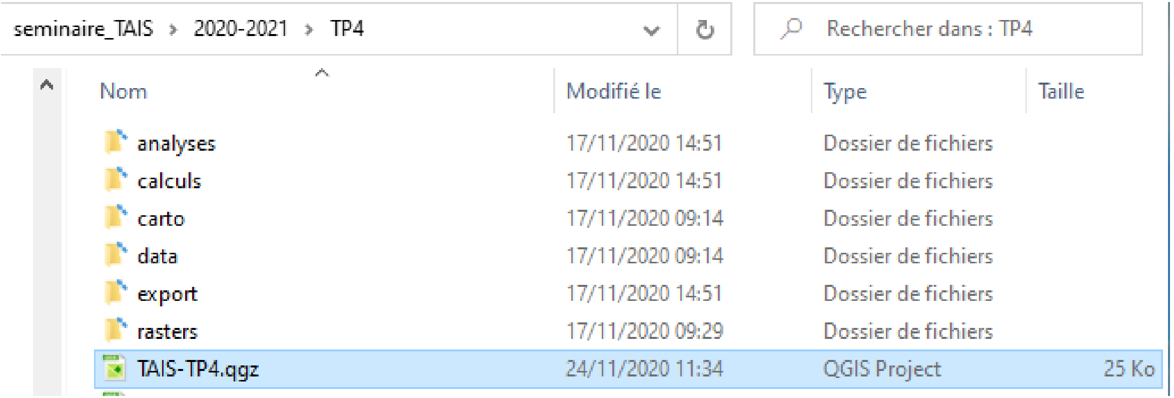
Download TP data

1. Data to download

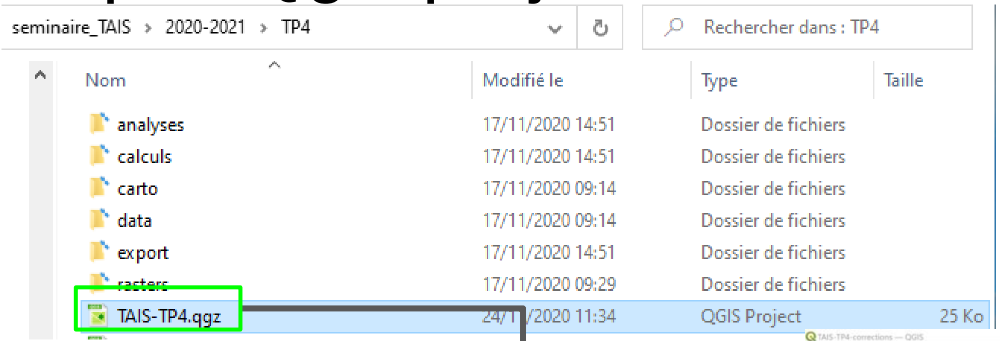
2. Unzip the directory to a working director



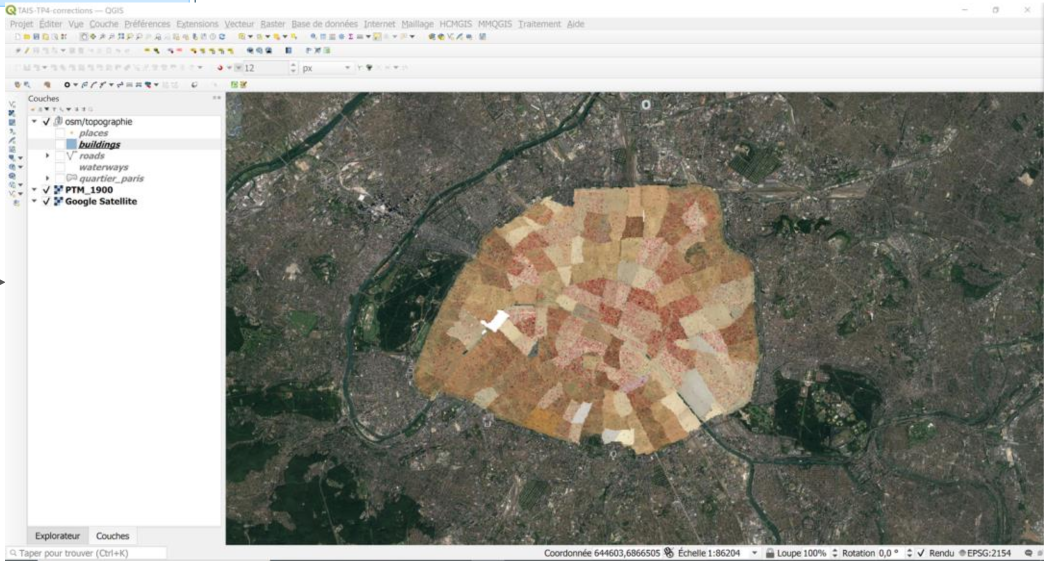
3. Open TP4 File



Open Qgis project file



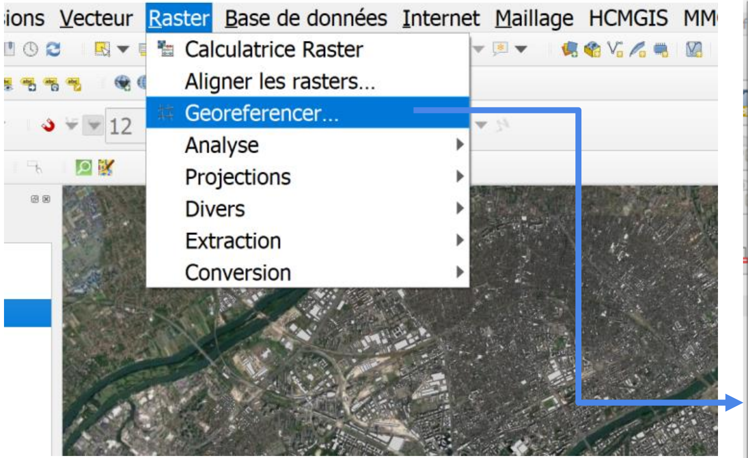
Open Qgis project file
TAIS-TP4.qgz

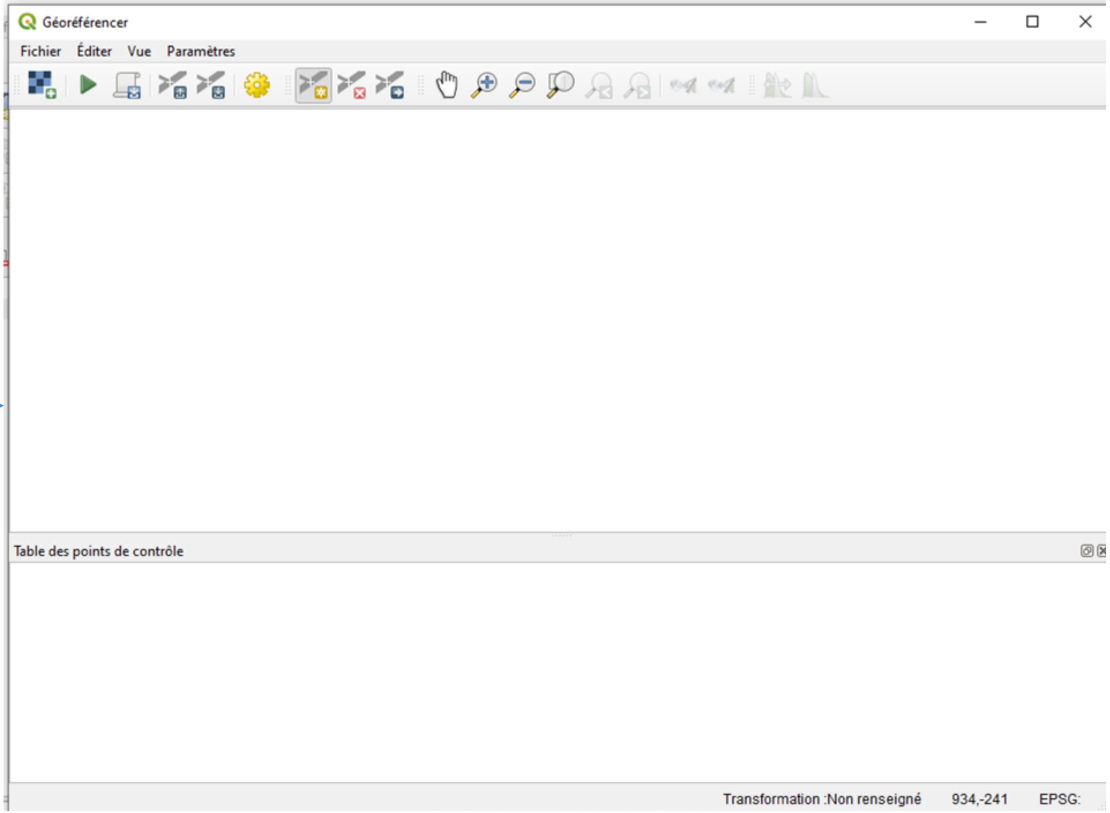


Georeferencing tool: first, adding an extension (optional, for some versions of QGIS)

Go to the "Extension"->"Install/manage extensions" menu, then find the "GDAL georeferencer" extension and check the box.

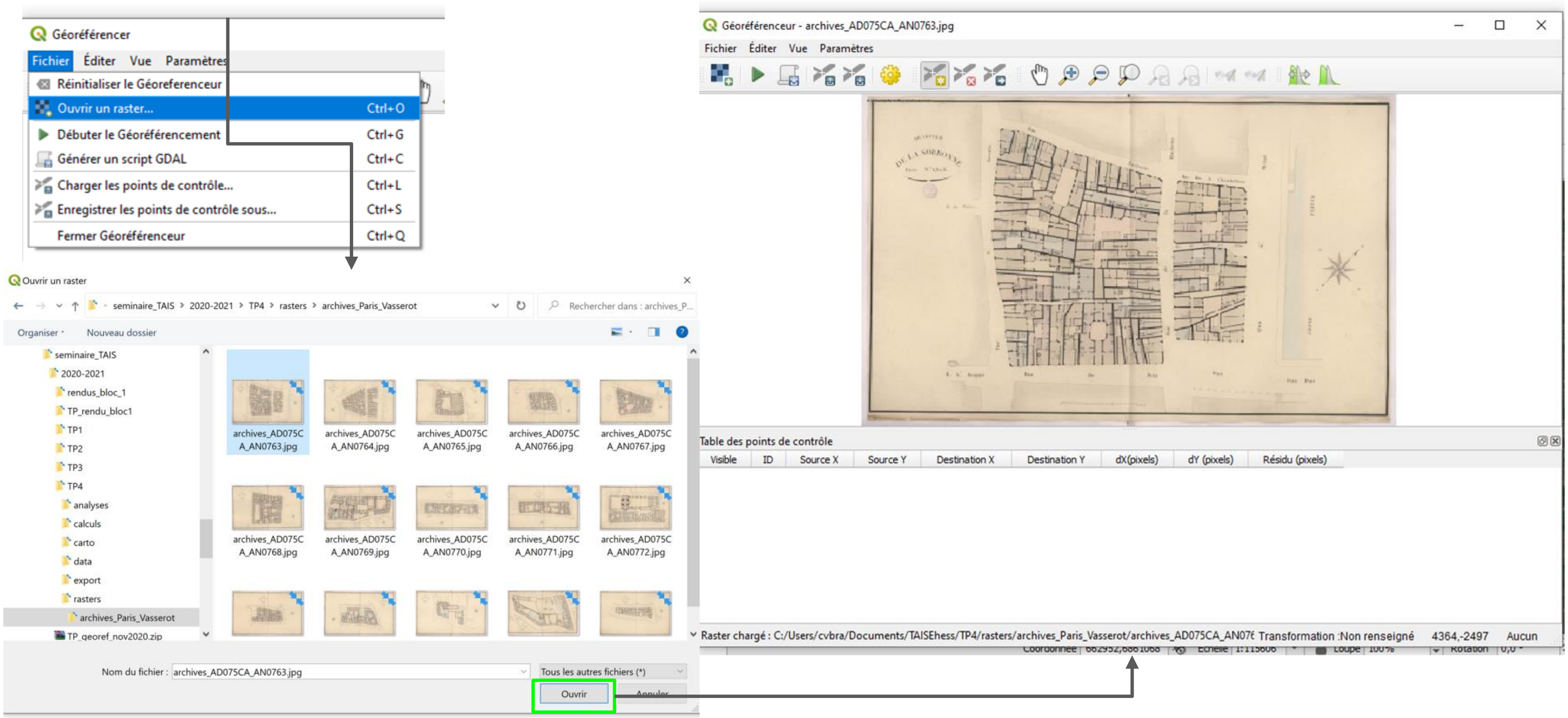
Georeferencing tool





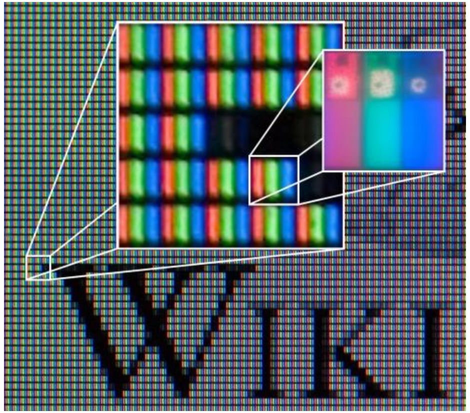
Open the georeferencer:
Raster/Georeviser menu...

Uploading a file

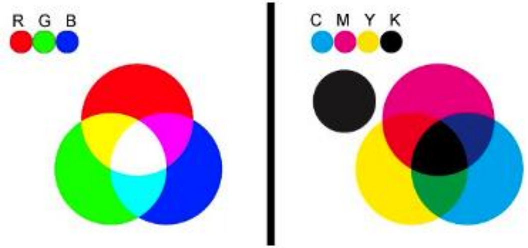


Open TP4\rasters\archives_Paris_Vasserot\archives_AD075CA_AN0763.jpg file

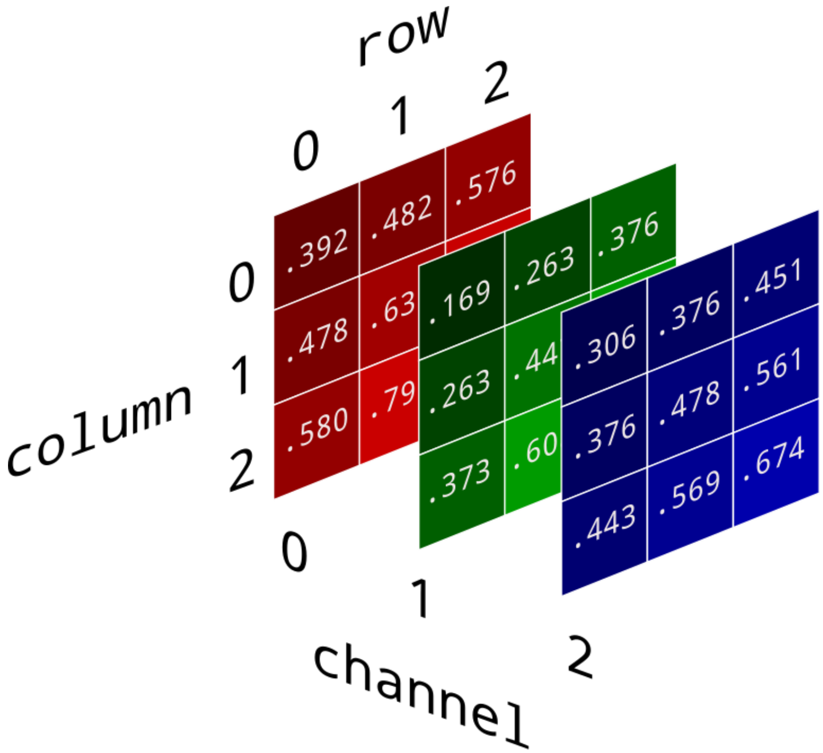
Some basics on image coding (1)



Vue physique d’une image affichée sur
un écran type LED

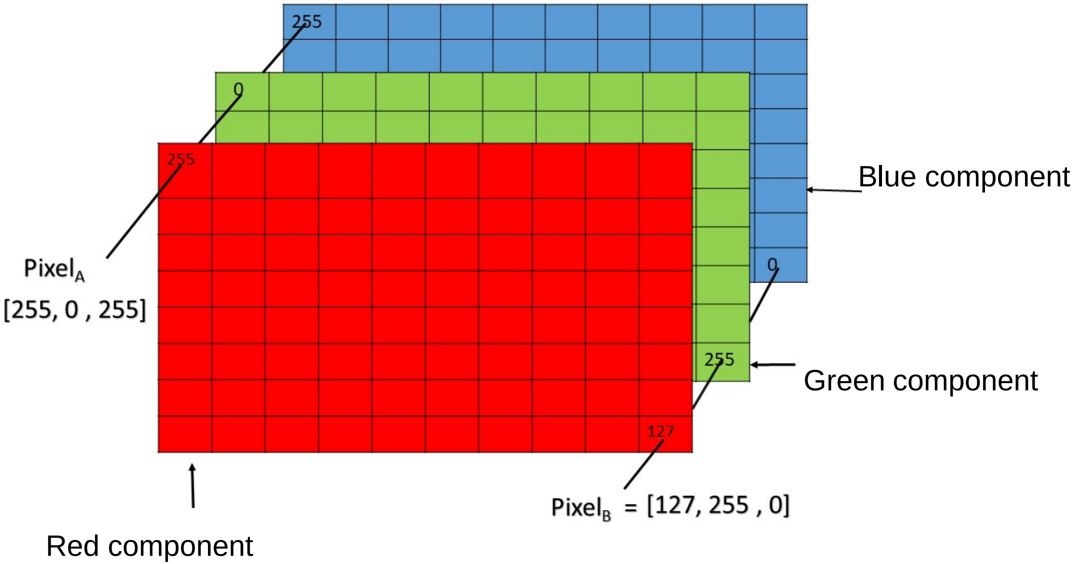


Color space to represent and encode colors visible to the human eye



Data structure: 3 overlapping matrices

Some basics on image coding (2)



Minimum: 0 -> no color
Maximum: here 255 -> color maximum

Thus:

one pixel is [0,0,0] = black
one pixel is [255,255,255] = white

Some basics on image coding (3)

What is 255?
This is the decimal value of the binary number: **11111111**

In computer science, information is coded with 0s (false) or 1s (true).
This binary digit or bit is the building block of the binary numeral system

In base 10: 0~9, tens are based on powers of 10, 10 = 101, 100 = 10^2

In base 2: 0~1, the "tens" are based on powers of 2, 10 = 21, 100 = 2^2
This is the binary system.

So : $_{B}11111111 = 2^7 + 2^6 + 2^5 + 2^4 + 2^3 + 2^2 + 2^1 + 2^0$

$= 128 + 64 + 32 + 16 + 8 + 4 + 2 + 1 = 255$

An image with 3 bands of color on 8 bits (= 1 byte = 1 byte)
is therefore on 24 bits, 3 bytes, or 256 × 256 × 256 possibilities, or about 16.7 million colors.

Some basics on image coding (4)

$_B11111111$ $= 2^7 + 2^6 + 2^5 + 2^4 + 2^3 + 2^2 + 2^1 + 2^0$

$= 128 + 64 + 32 + 16 + 8 + 4 + 2 + 1 = 255$

To "simplify" the writing of number in binary, we switch to base 16 or hexadecimal base.
In base 16, tens are based on powers of 16, 10 = 161, 100 = 16^2

And since 16 = 24, then we can group 4 bits on a digit,
but we need to be able to count up to 15 : 0~9,A~F : 0 1 2 3 4 5 6 7 8 9 A B C D E F
Puis $_{Hexa}$10 = 0x**1**0 = 16

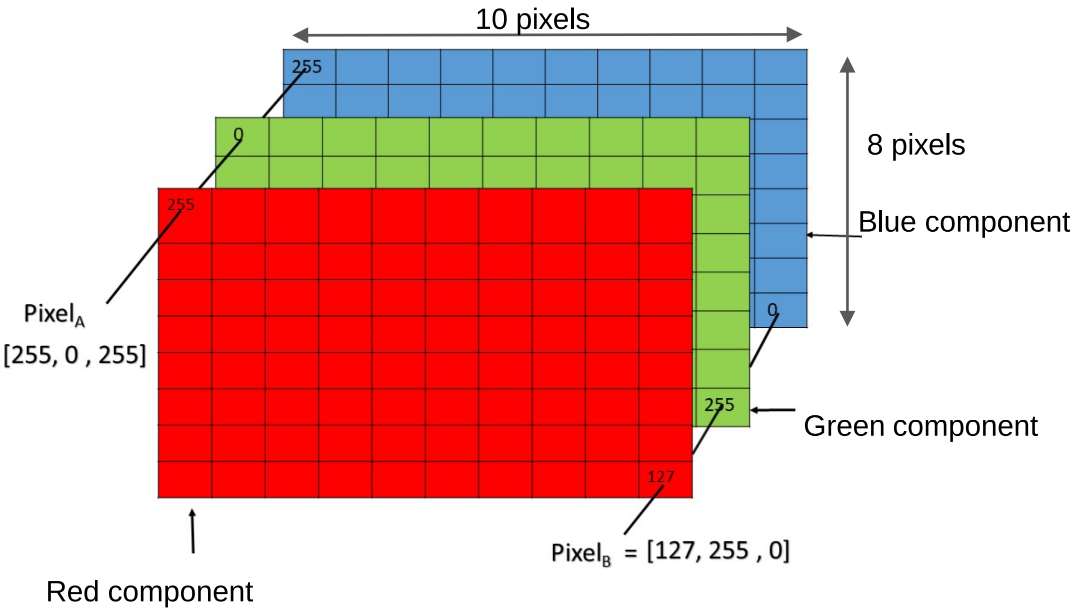
10 11 12 13 14 15

So : $_B$1111 1111 = 15 x 16^1 + 15 x 16^0 = 0xFF

0x F F,
since 15 in decimal is F in hexadecimal

Thus, the color code 0xDC0021
(Red / Green / Blue) R V B
Red = 0xDC = 13x161 + 12x160 = 220
Green = 0x00 = 0 and Blue = 0x21 = 2x16 +1 = 33

Some basics on image coding (5)

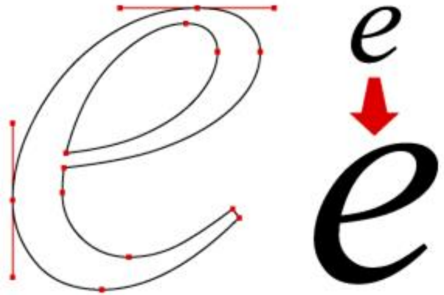


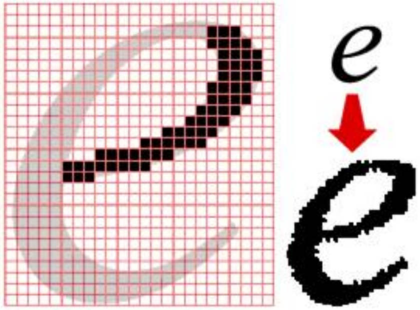
The image above has a resolution of: 10x8 on 3 bands, and each pixel is 8-bit encoded:

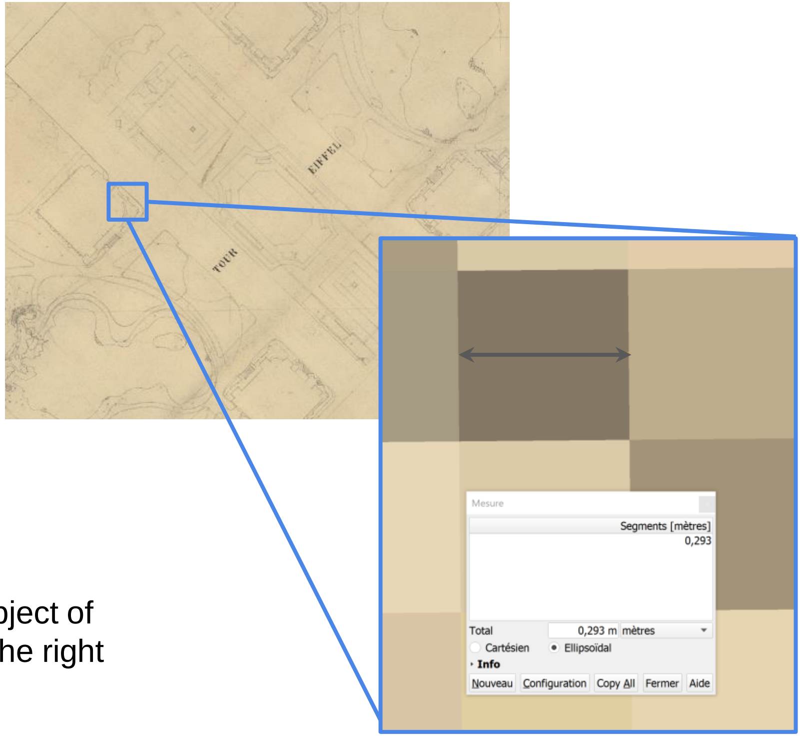
Image weight:
10x8x3(tapes)x8(bits) = 1920 bits
10x8x3(tapes) = 240 bytes (or bytes)

Rasters Data Reminder

In its simplest form, a raster consists of a matrix of cells (or pixels) organized into rows and columns (grid) in which each cell contains a value representing information encoded in a color (RGB, grayscale): digital aerial photographs, satellite images, digital images, scanned maps, temperature maps, elevation.





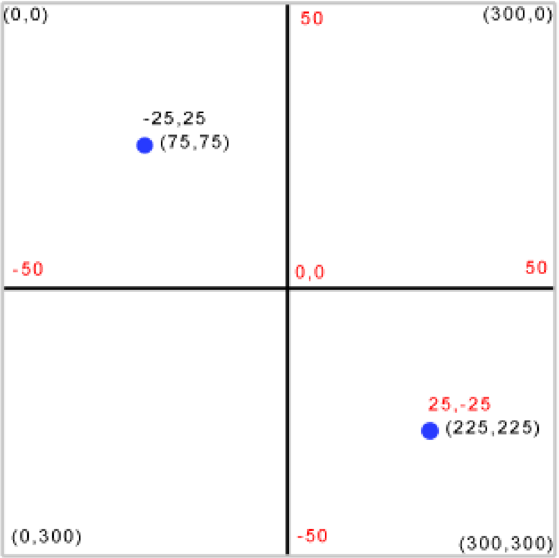


Vector Information:
- geometric primitives
- coordinates (x,y,[z])

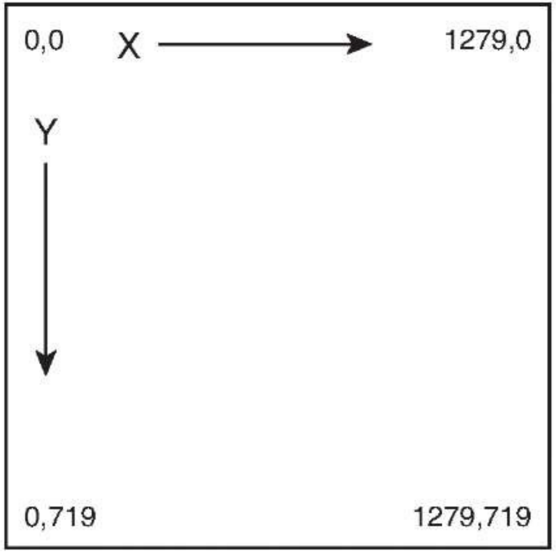
Raster information:
- Matrices
- coordinates (x,y,[z])

A pixel therefore represents an object of reality at a certain resolution, on the right a pixel represents 30 cm

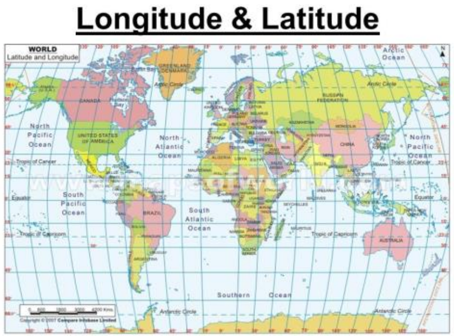
Computer coordinates to cartographic coordinates



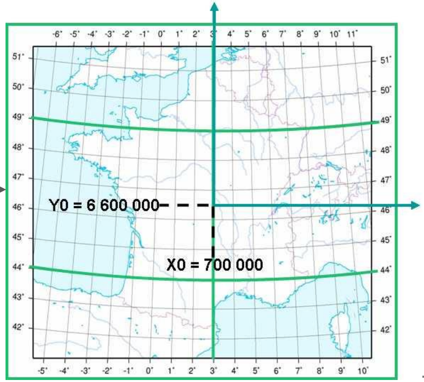
In mathematics: orthonormed reference frame



In computer science: the origin is located in the top left corner



A mathematical transformation is needed!

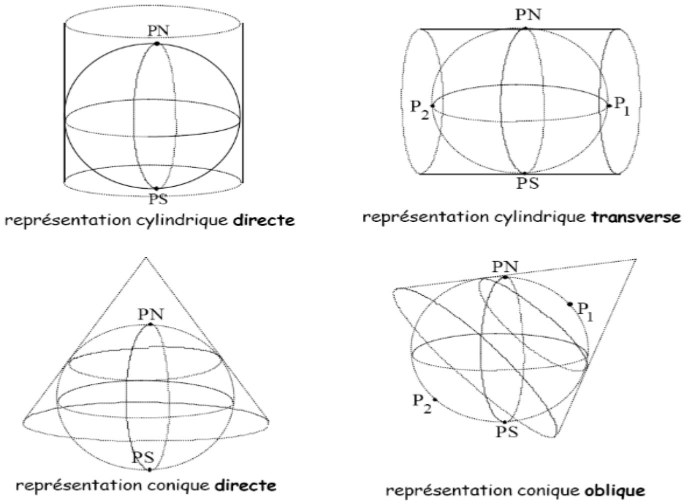


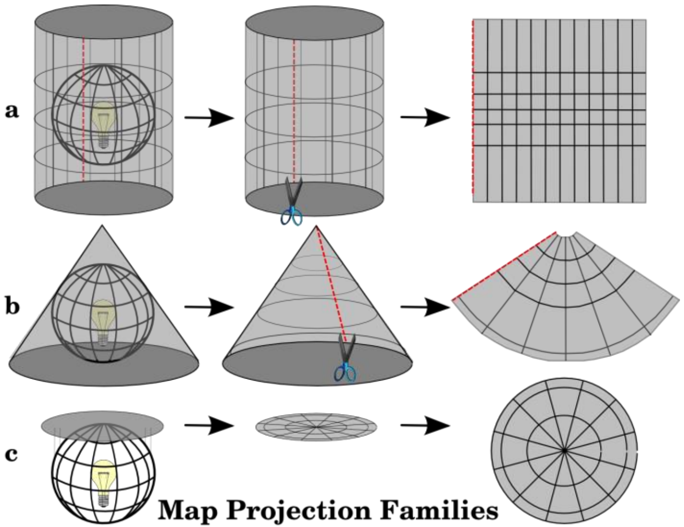
Definitions

A projection is a geometric transformation that makes it possible to move from geographical coordinates (latitude, longitude expressed in degrees) to cartographic coordinates (planes, expressed in meters)

Georeferencing consists of applying a coordinate system to images that initially have a computer reference frame in order to scale them (in a given projection system).

Reminders: projections





Any projection induces distortions:
- along the tangent line: zero distortion
- the further away from it, the more distortion increases

A projection can keep...
- angles: conformal projection
- surfaces: equivalent projection
- Neither: aphylactic projection

but NEVER BOTH at the same time

The georeferencing process

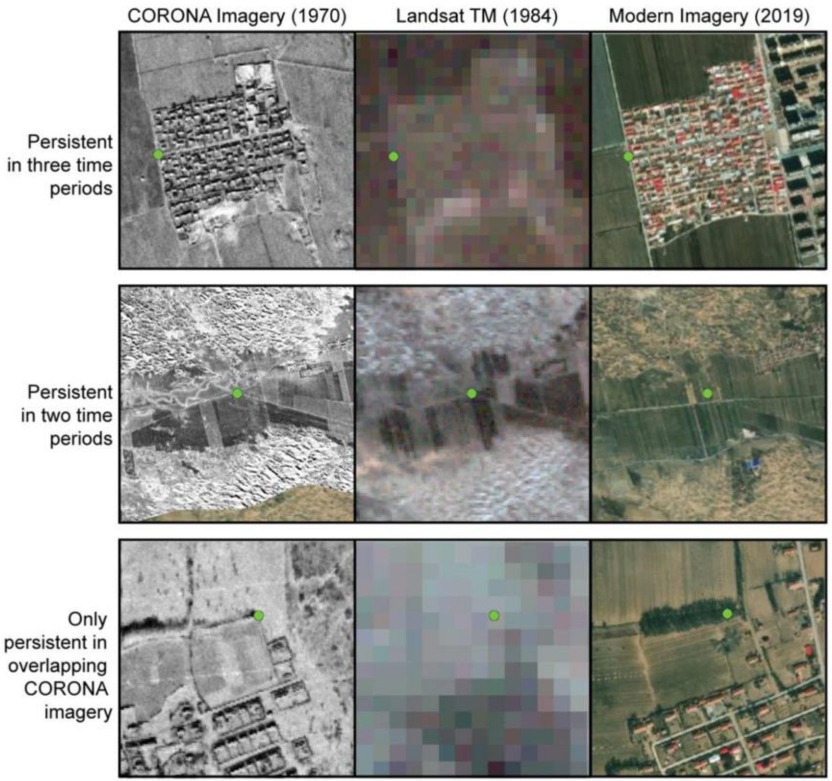
First step

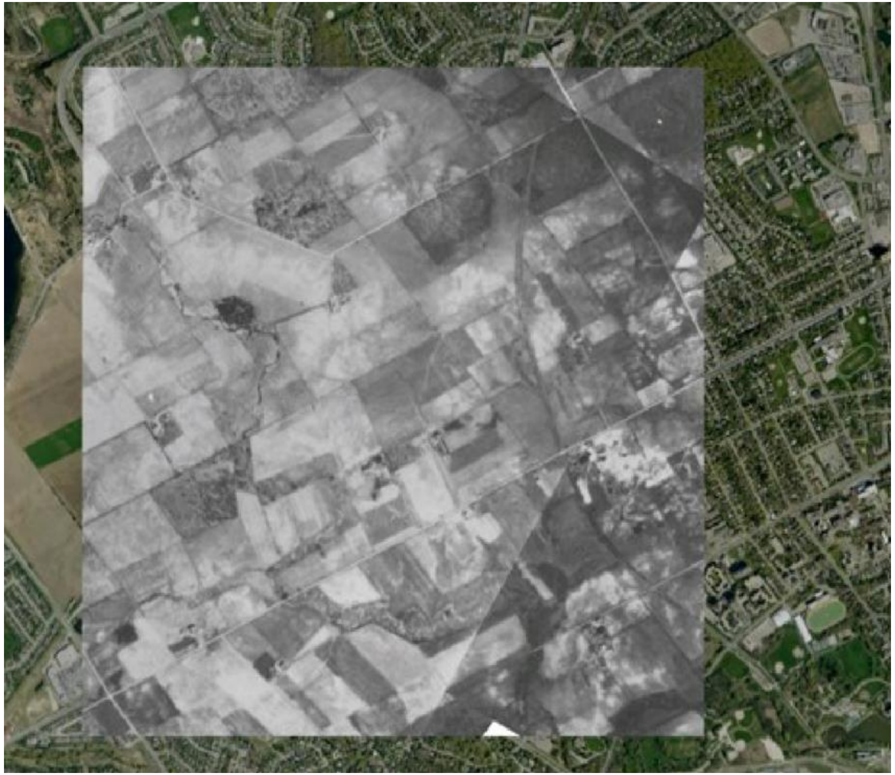
have established a data repository in a geographic or cartographic marker

4 steps:

1. Creating checkpoints
2. Transformation parameters
3. Interpretation of the erro
4. Resampling and export

Why georeference images?





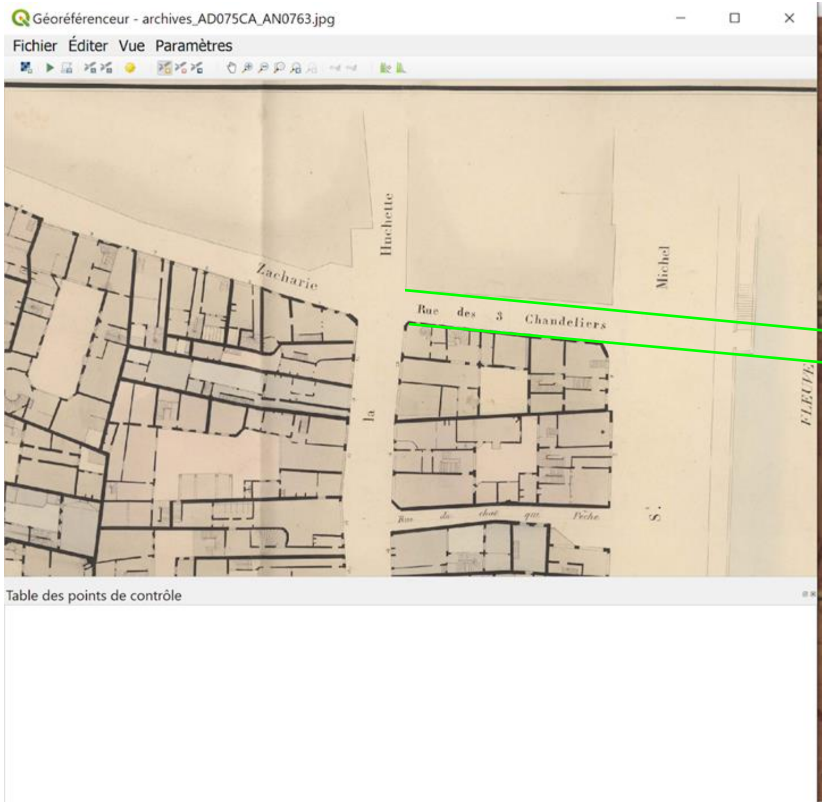
Analysis of the evolution of a territory, but images / data are in archives (not digitized, not in a spatial referentiel)

Environment, Persisency of shapes, vegetation, evolution of cities

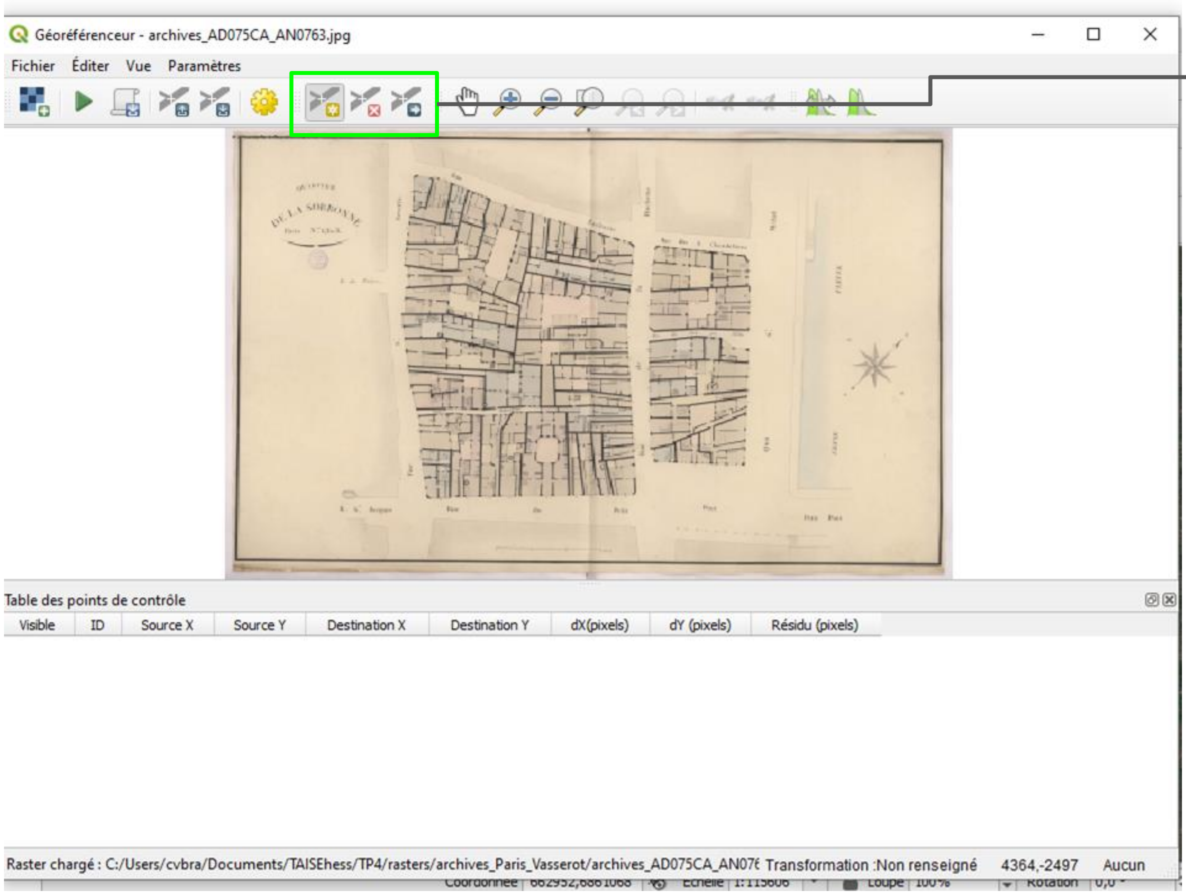
Find your marks in space

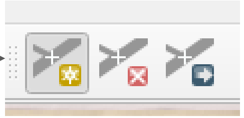
Look at the map to reference / map referential

Visual correspondence: exercising your eye



Creating checkpoints in Qgis

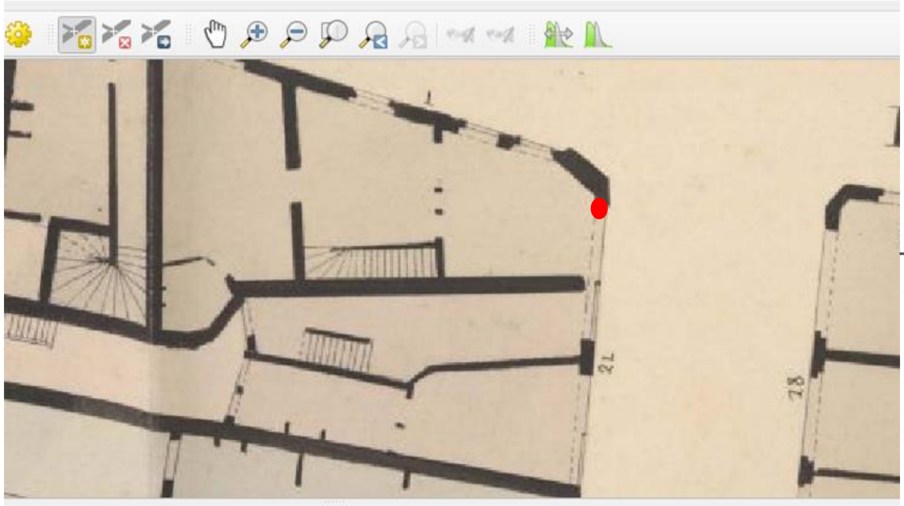




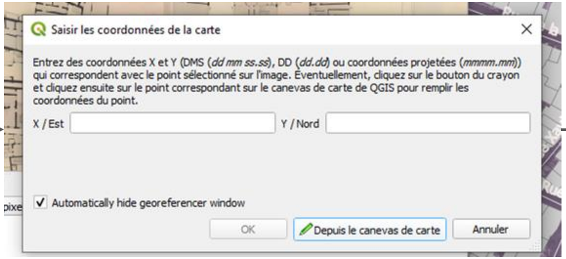
Tools buttons: add, remove, and move a control point

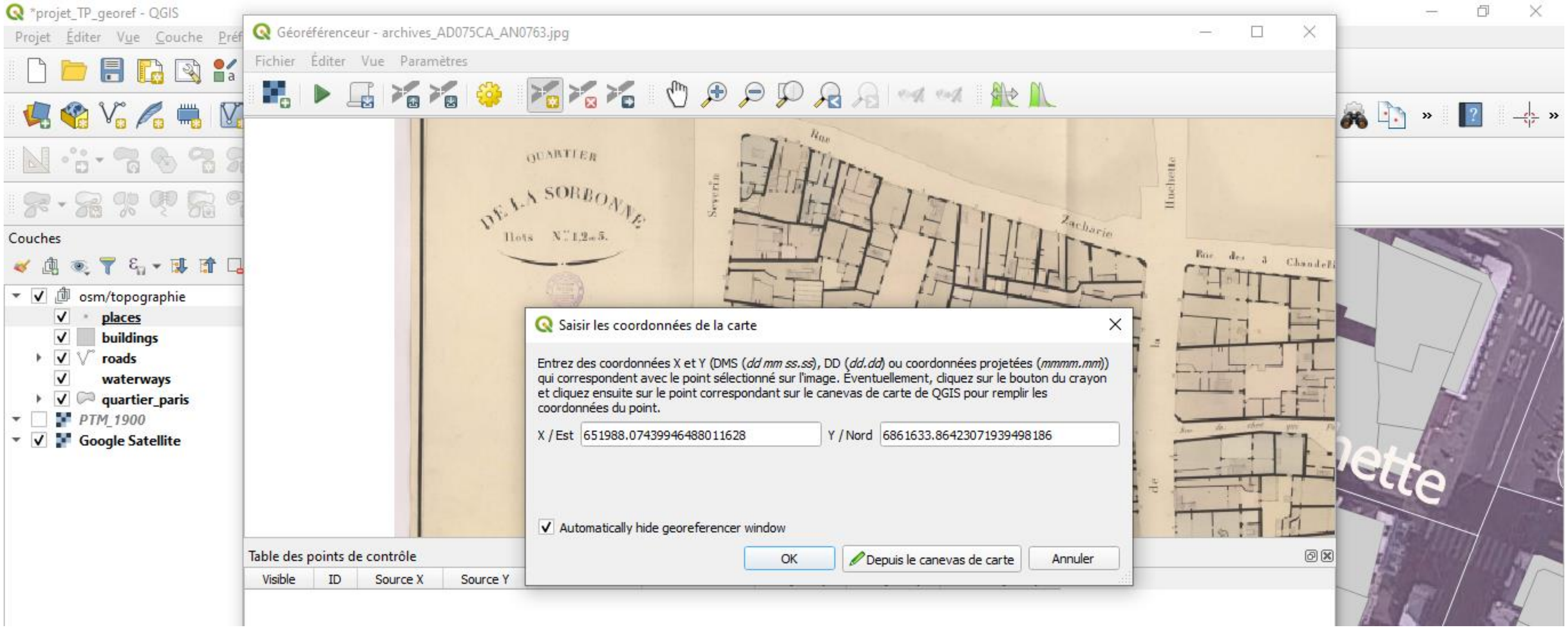
Checkpoint table

To create a control point, zoom in on a point (in red) in the image (intersection rue de la Huchette and rue Zacharie):



to associate it with a location, either by indicating the coordinates or by clicking on 'From the map canvas'.

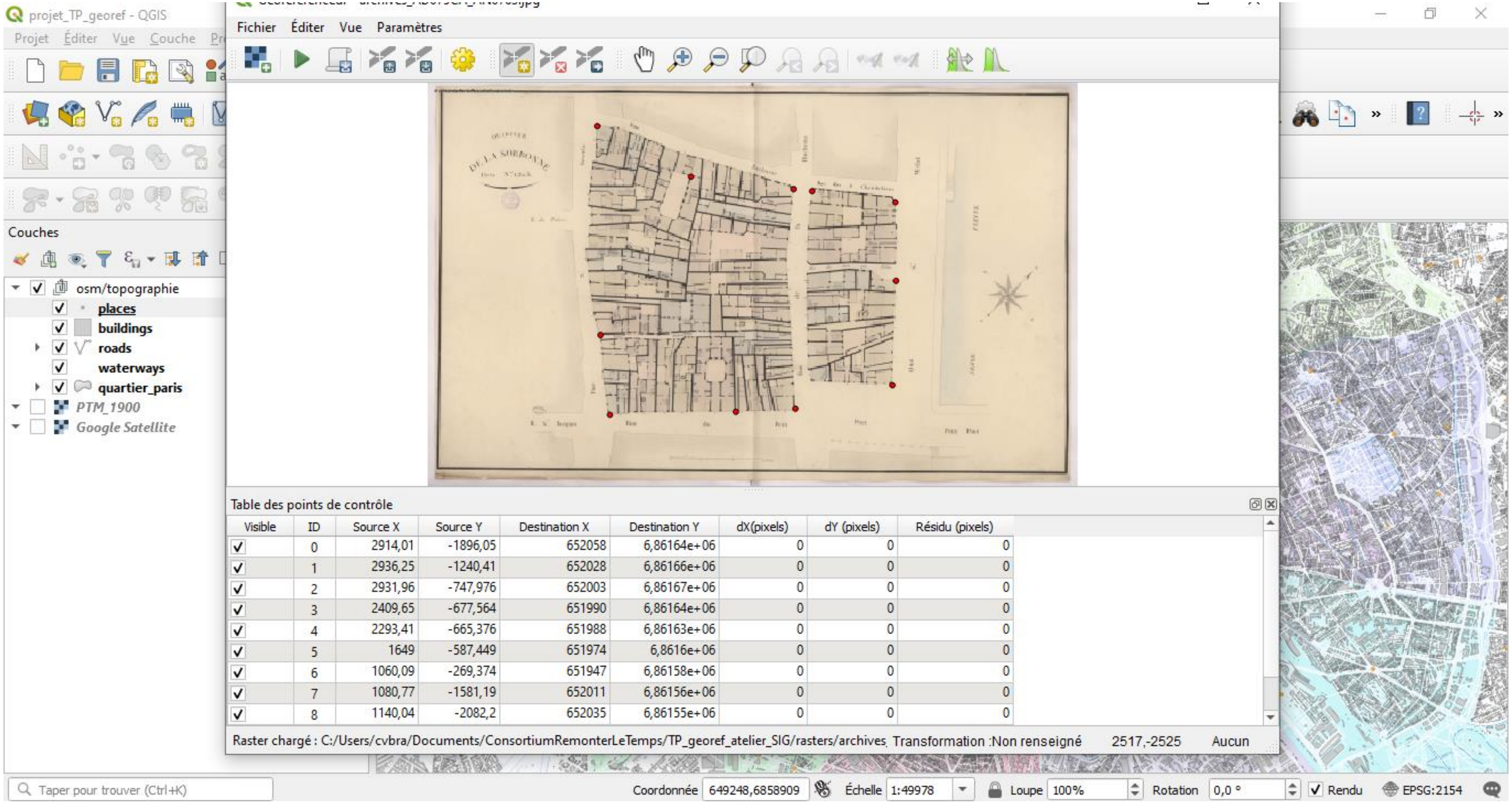




Coordinates are transferred and stored in the checkpoint table

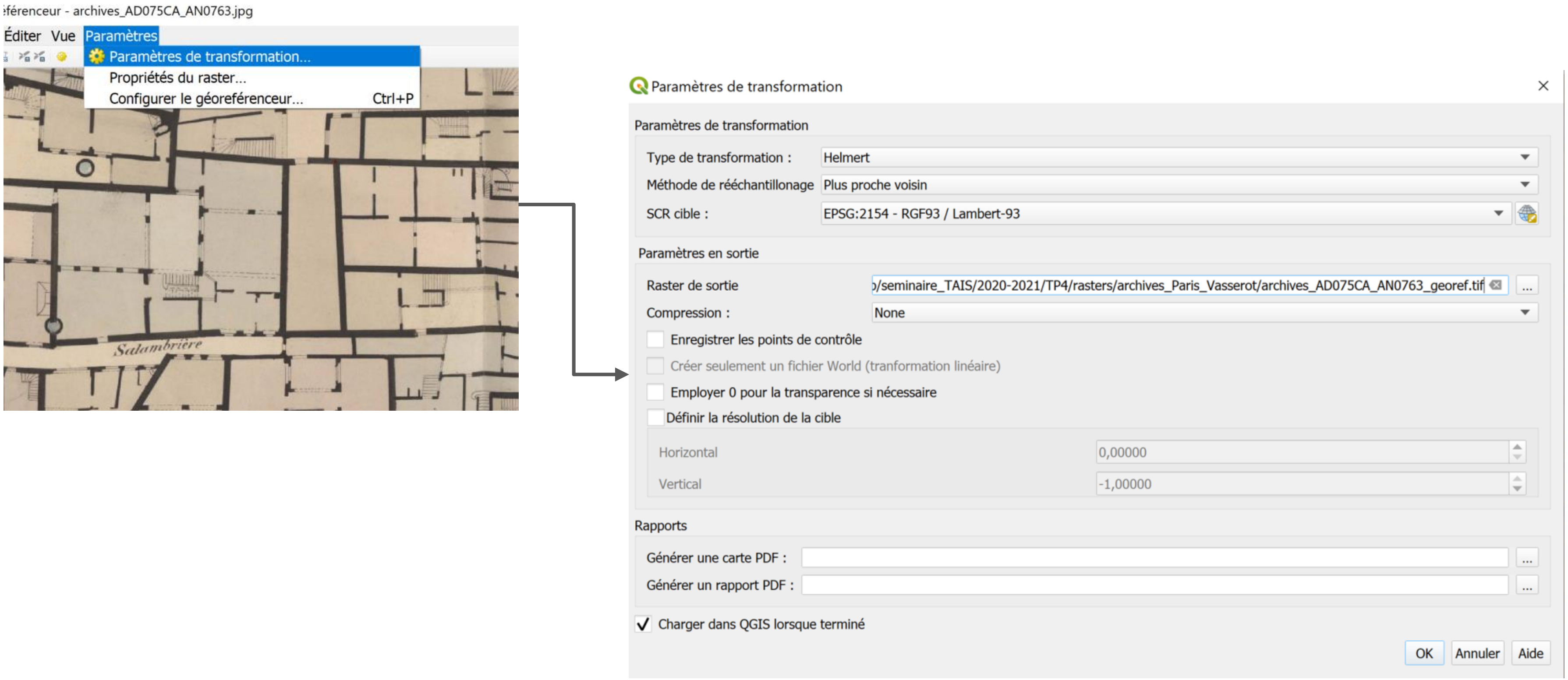
Table des points de contrôle

Visible	ID	Source X	Source Y	Destination X	Destination Y	dX(pixels)	dY (pixels)	Résidu (pixels)
✓	0	2292,47	-671,624	651988	6,86163e+06	0	0	0

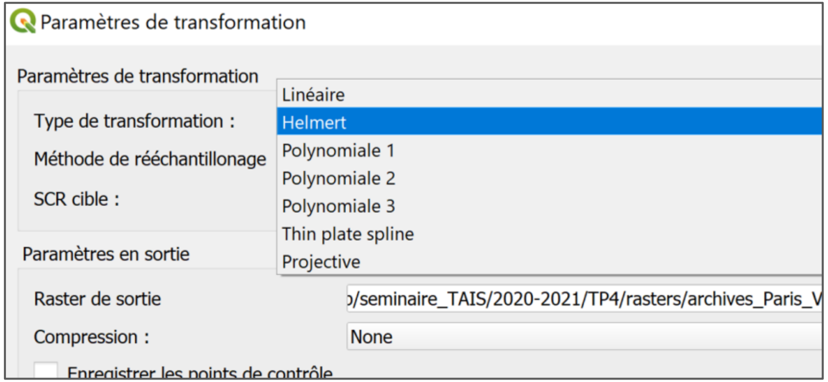


Create at least ten evenly distributed checkpoints using the same procedure.

Transformation parameters

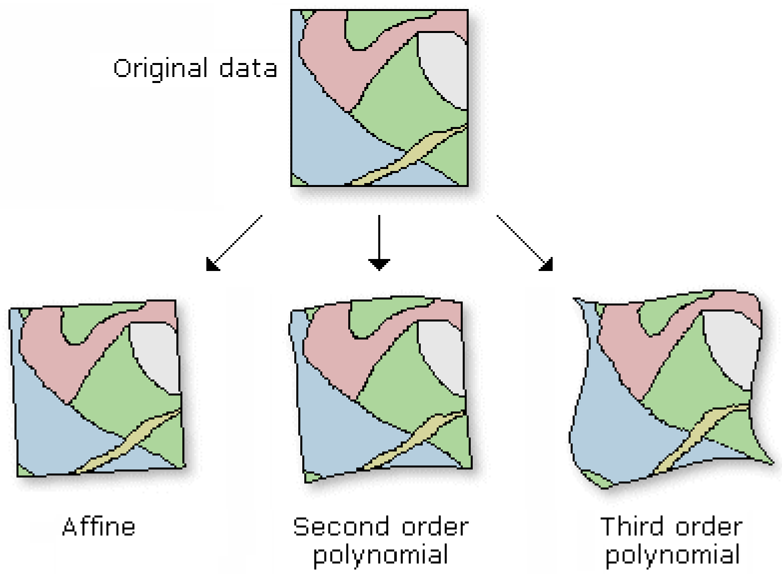


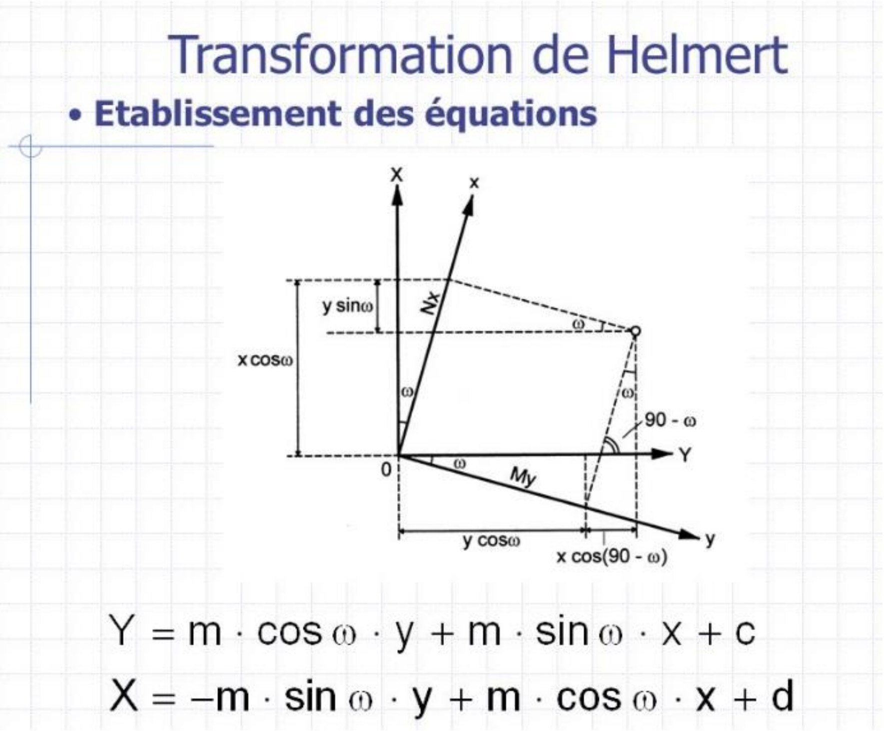
Transformation parameters



- The **Linear algorithm** is used to create a "world" file. It is different from other algorithms in that it does not transform the raster. This algorithm is unlikely to be sufficient to georeference scanned data.
- The **Helmert algorithm** applies simple translation, rotation and scaling.
- **Polynomial algorithms** of degrees 1 to 3 are among the most widely used algorithms for georeferencing and each differs in the degree of distortion it introduces to best match the source to the control points. The most widely used polynomial transformation is that of order two which allows some curves. The polynomial transformation of order one (also called affine transformation) preserves collinearity and allows only translation, rotation and scaling (like the Helmert transform).
- The **Thin Plate Spline** (TPS) algorithm is a more modern method that is able to introduce distortions on specific sectors of the image. It is very handy when low-quality sources are used.
- The **Projective algorithm** is a linear rotation and then a translation of coordinates

Transformation parameters





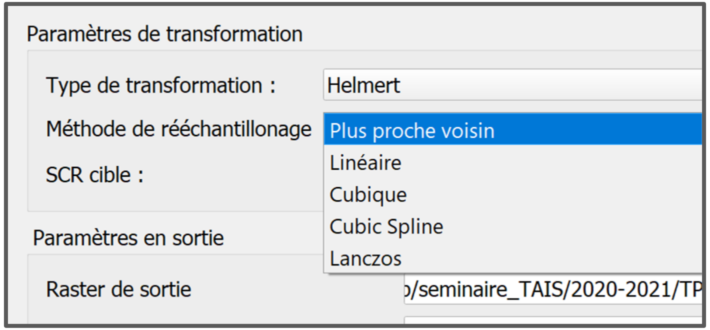
Resampling and export

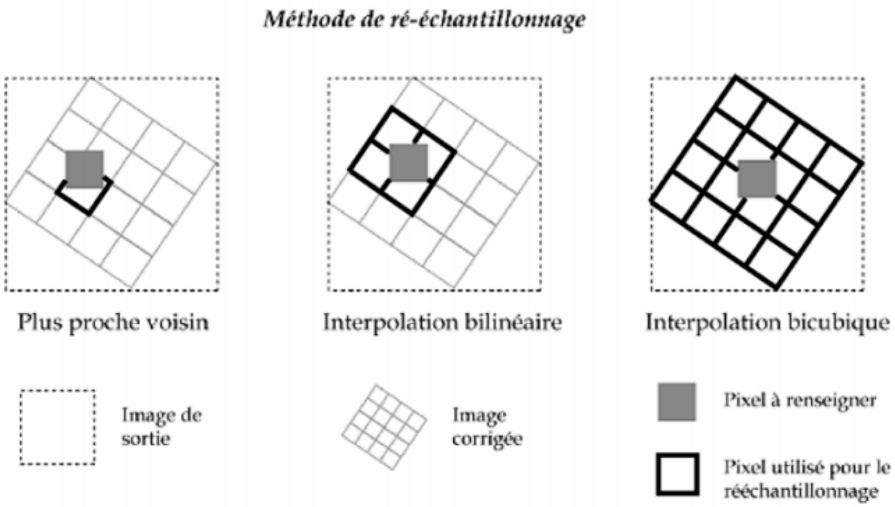
There are several resampling techniques:

- Nearest neighbor: Each pixel in the new image is assigned the value of the nearest pixel in the old image. This method is fast but causes an effect of stair steps on the diagonals. It is suitable for categorical or thematic data.

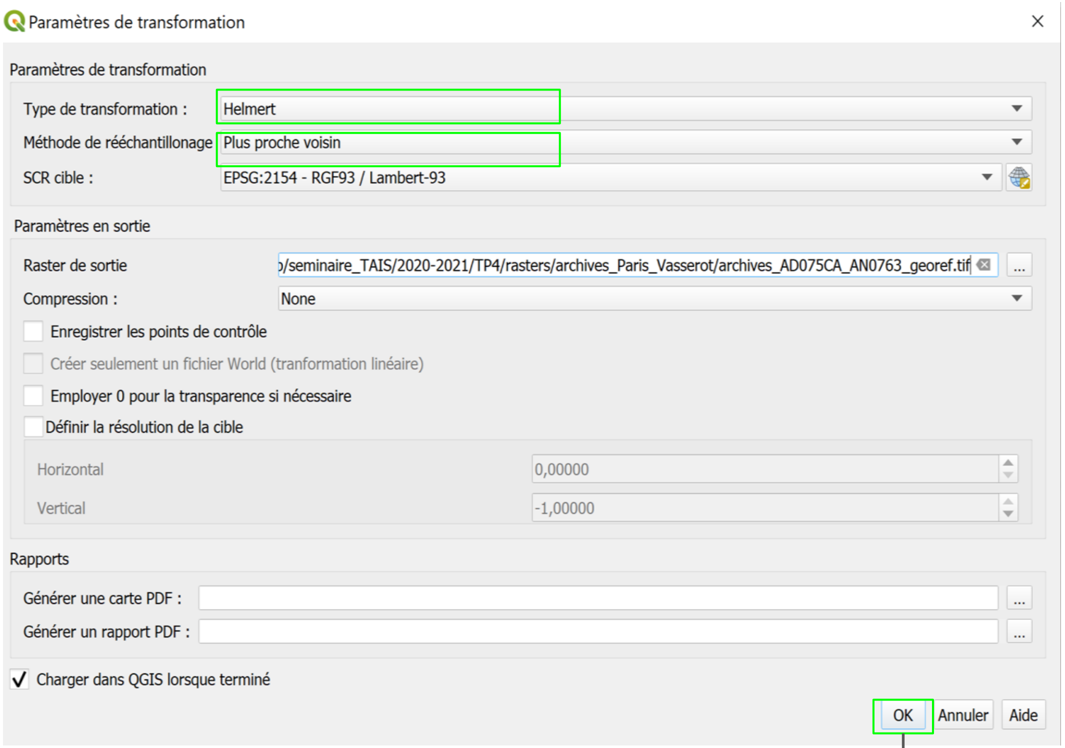
- Bilinear interpolation: assigns each pixel of the new image to the value of the nearest 4 neighbors (2x2) in the original image.

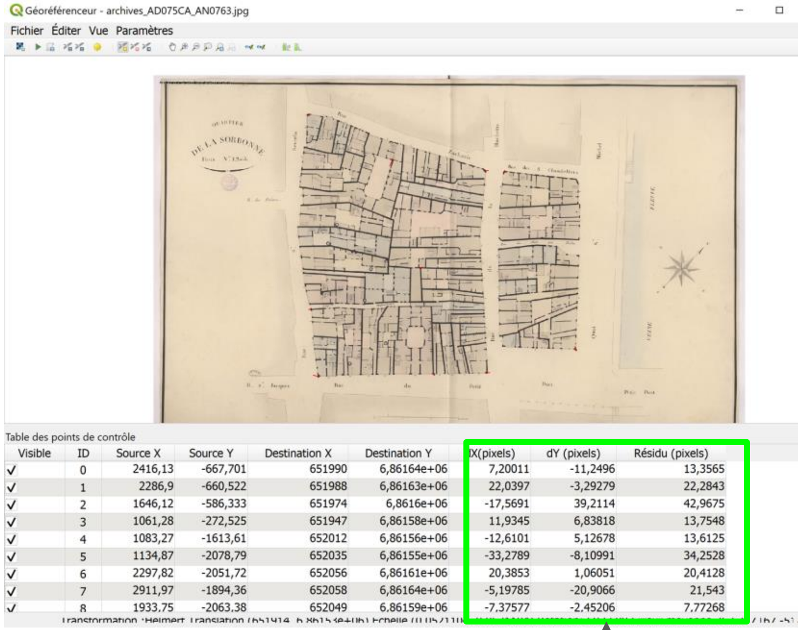
- Bicubic interpolation or cubic convolution: This assigns each pixel in the new image the value of the 16 nearest neighbors (4x4) in the original image.





Transformation parameters

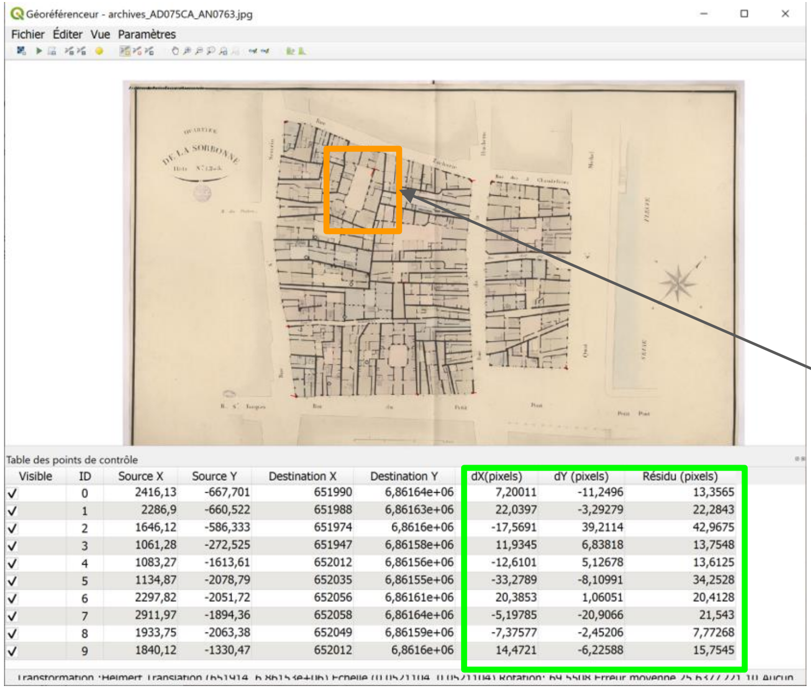




Visible	ID	Source X	Source Y	Destination X	Destination Y	dX(pixels)	dY (pixels)	Résidu (pixels)
✓	0	2416,13	-667,701	651990	6,86164e+06	7,20011	-11,2496	13,3565
✓	1	2286,9	-660,522	651988	6,86163e+06	22,0397	-3,29279	22,2843
✓	2	1646,12	-586,333	651974	6,8616e+06	-17,5691	39,2114	42,9675
✓	3	1061,28	-272,525	651947	6,86158e+06	11,9345	6,83818	13,7548
✓	4	1083,27	-1613,61	652012	6,86156e+06	-12,6101	5,12678	13,6125
✓	5	1134,87	-2078,79	652035	6,86155e+06	-33,2789	-8,10991	34,2528
✓	6	2297,82	-2051,72	652056	6,86161e+06	20,3853	1,06051	20,4128
✓	7	2911,97	-1894,36	652058	6,86164e+06	-5,19785	-20,9066	21,543
✓	8	1933,75	-2063,38	652049	6,86159e+06	-7,37577	-2,45206	7,77268

Clicking ok will calculate the residual error on the control points knowing the parameters of the transformation

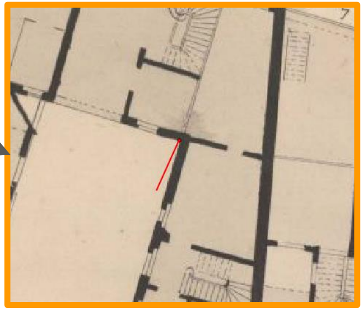
Interpretation of the error



Visible	ID	Source X	Source Y	Destination X	Destination Y	dX(pixels)	dY (pixels)	Résidu (pixels)
✓	0	2416,13	-667,701	651990	6,86164e+06	7,20011	-11,2496	13,3565
✓	1	2286,9	-660,522	651988	6,86163e+06	22,0397	-3,29279	22,2843
✓	2	1646,12	-586,333	651974	6,8616e+06	-17,5691	39,2114	42,9675
✓	3	1061,28	-272,525	651947	6,86158e+06	11,9345	6,83818	13,7548
✓	4	1083,27	-1613,61	652012	6,86156e+06	-12,6101	5,12678	13,6125
✓	5	1134,87	-2078,79	652035	6,86155e+06	-33,2789	-8,10991	34,2528
✓	6	2297,82	-2051,72	652056	6,86161e+06	20,3853	1,06051	20,4128
✓	7	2911,97	-1894,36	652058	6,86164e+06	-5,19785	-20,9066	21,543
✓	8	1933,75	-2063,38	652049	6,86159e+06	-7,37577	-2,45206	7,77268
✓	9	1840,12	-1330,47	652012	6,8616e+06	14,4721	-6,22588	15,7545

dx(pixels) and dy(pixels) is the error that corresponds to the difference between the final position of the origin point from that of the destination point, i.e. the actual location being searched.

The residue (pixels) is the norm of the vector v1 that represents the synthesis of dx and dy errors. This value is generally used to estimate the point-by-point accuracy of georeferencing. For good accuracy this value should preferably be less than 10.

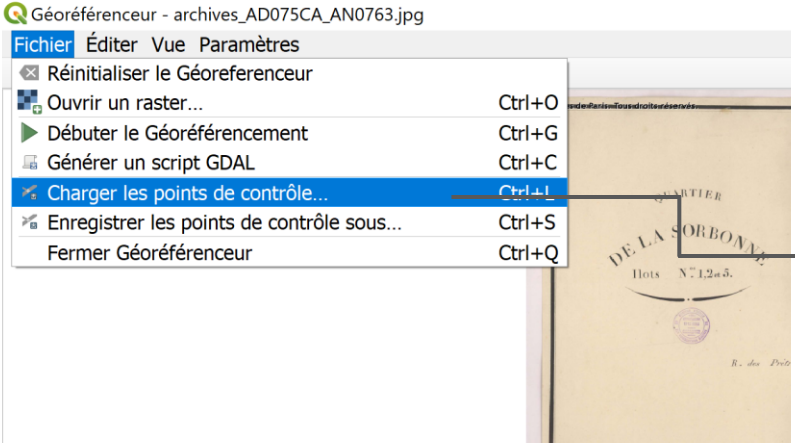


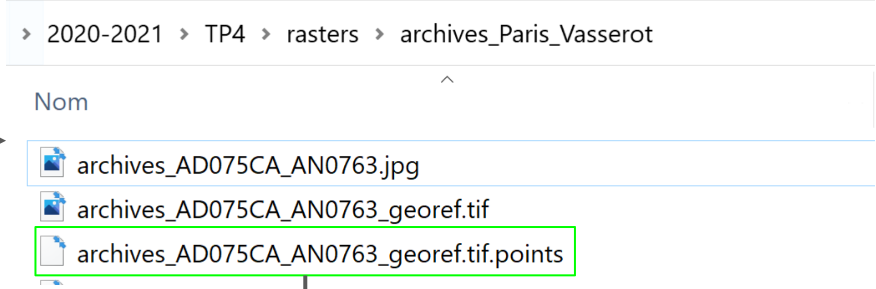
The average error gives a general assessment of the quality of georeferencing.

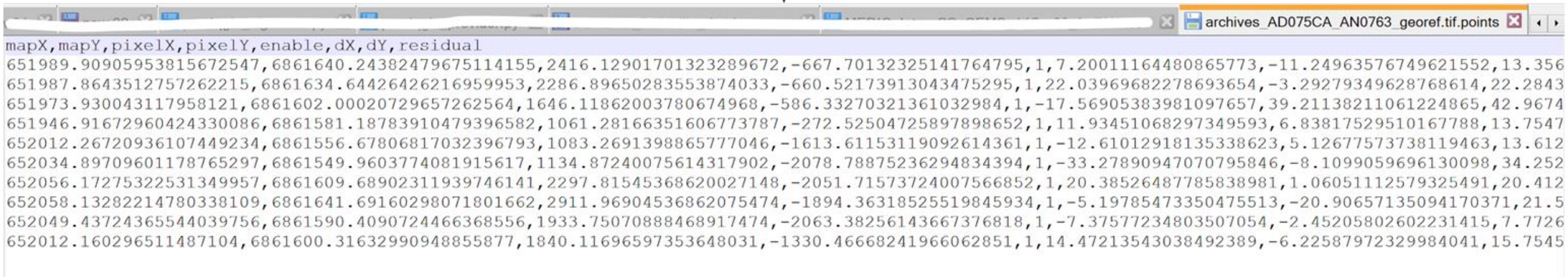
The calculation of errors depends on the type of transformation.

1 $|v| = \sqrt{(dx^2 + dy^2)}$

Record checkpoints

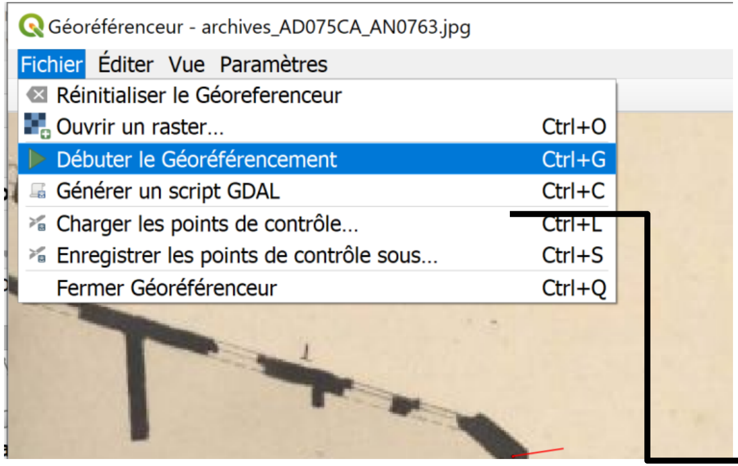


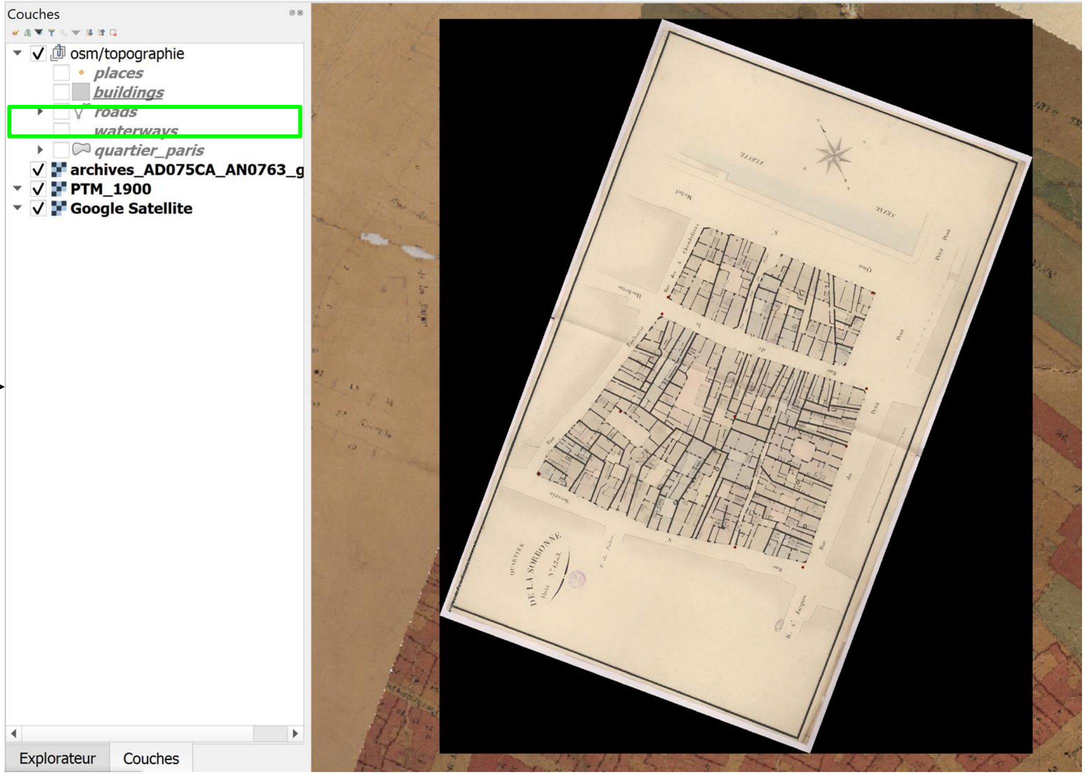



archives_AD075CA_AN0763_georef.tif.points

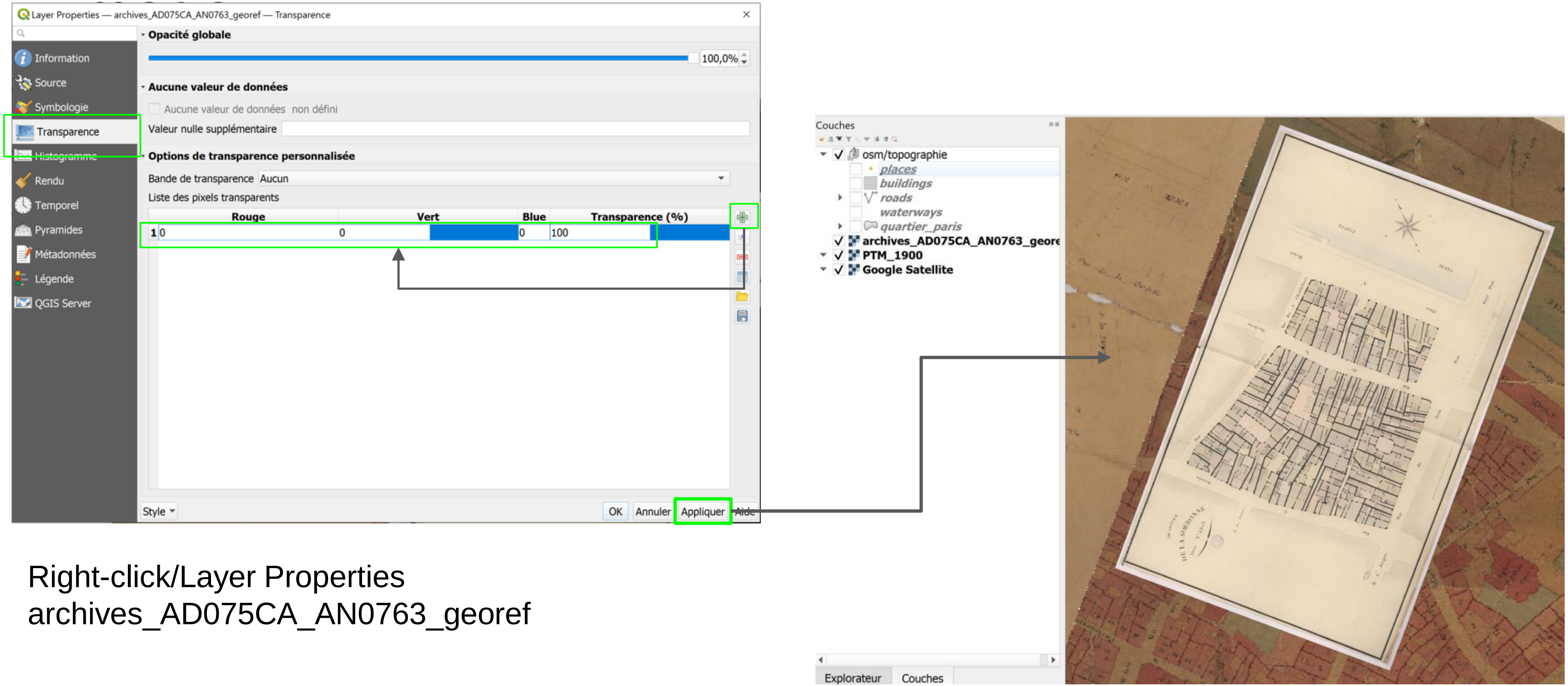
```
mapX,mapY,pixelX,pixelY,enable,dX,dY,residual
651989.90905953815672547,6861640.24382479675114155,2416.12901701323289672,-667.70132325141764795,1,7.20011164480865773,-11.24963576749621552,13.356
651987.8643512757262215,6861634.64426426216959953,2286.89650283553874033,-660.52173913043475295,1,22.03969682278693654,-3.29279349628768614,22.2843
651973.930043117958121,6861602.00020729657262564,1646.11862003780674968,-586.33270321361032984,1,-17.56905383981097657,39.21138211061224865,42.9674
651946.91672960424330086,6861581.18783910479396582,1061.28166351606773787,-272.52504725897898652,1,11.93451068297349593,6.83817529510167788,13.7547
652012.26720936107449234,6861556.67806817032396793,1083.2691398865777046,-1613.61153119092614361,1,-12.61012918135338623,5.12677573738119463,13.612
652034.89709601178765297,6861549.9603774081915617,1134.87240075614317902,-2078.78875236294834394,1,-33.27890947070795846,-8.1099059696130098,34.252
652056.17275322531349957,6861609.68902311939746141,2297.81545368620027148,-2051.71573724007566852,1,20.38526487785838981,1.06051112579325491,20.412
652058.13282214780338109,6861641.69160298071801662,2911.96904536862075474,-1894.36318525519845934,1,-5.19785473350475513,-20.90657135094170371,21.5
652049.43724365544039756,6861590.4090724466368556,1933.75070888468917474,-2063.38256143667376818,1,-7.37577234803507054,-2.45205802602231415,7.7726
652012.160296511487104,6861600.31632990948855877,1840.11696597353648031,-1330.46668241966062851,1,14.47213543038492389,-6.22587972329984041,15.7545
```

Georeferencing ... and outcome



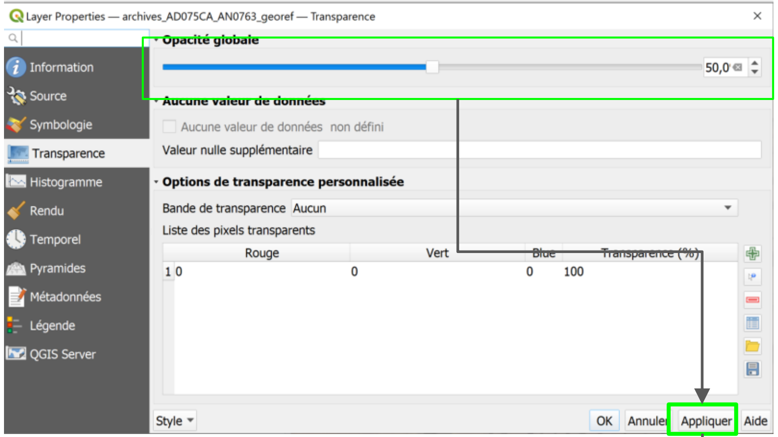


Eliminate the black outline on the georeferenced



Right-click/Layer Properties
archives_AD075CA_AN0763_georef

Visualization of the result by transparency



In the properties of the layer, change the transparency to 50%

Comparison of resampling methods

Nearest neighbour resampling

Resampling by cubic method

Comparison of transformations

Table des points de contrôle

Visible	ID	Source X	Source Y	Destination X	Destination Y	dX(pixels)	dY (pixels)	Résidu (pixels)
✓	0	2416,13	-667,701	651990	6,86164e+06	7,20011	-11,2496	13,3565
✓	1	2286,9	-660,522	651988	6,86163e+06	22,0397	-3,29279	22,2843
✓	2	1646,12	-586,333	651974	6,8616e+06	-17,5691	39,2114	42,9675
✓	3	1061,28	-272,525	651947	6,86158e+06	11,9345	6,83818	13,7548
✓	4	1083,27	-1613,61	652012	6,86156e+06	-12,6101	5,12678	13,6125
✓	5	1134,87	-2078,79	652035	6,86155e+06	-33,2789	-8,10991	34,2528
✓	6	2297,82	-2051,72	652056	6,86161e+06	20,3853	1,06051	20,4128
✓	7	2911,97	-1894,36	652058	6,86164e+06	-5,19785	-20,9066	21,543
✓	8	1933,75	-2063,38	652049	6,86159e+06	-7,37577	-2,45206	7,77268
✓	9	1840,12	-1330,47	652012	6,8616e+06	14,4721	-6,22588	15,7545

Table des points de contrôle

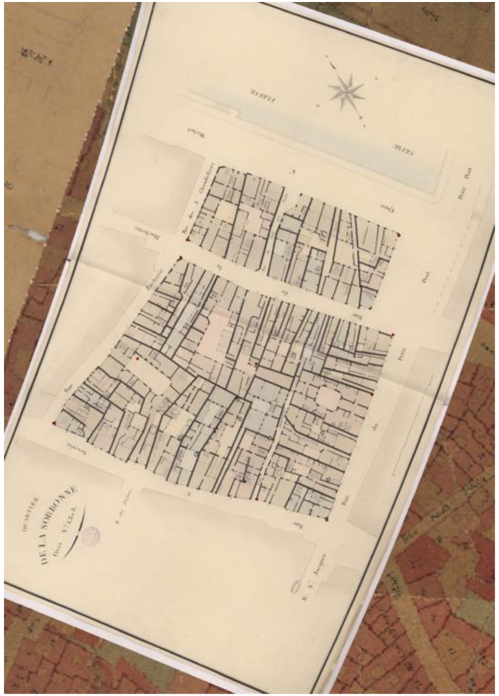
Visible	ID	Source X	Source Y	Destination X	Destination Y	dX(pixels)	dY (pixels)	Résidu (pixels)
✓	0	2416,13	-667,701	651990	6,86164e+06	-0,78172	-4,63357	4,69905
✓	1	2286,9	-660,522	651988	6,86163e+06	12,5842	-1,37888	12,6595
✓	2	1646,12	-586,333	651974	6,8616e+06	-27,1925	24,3814	36,5224
✓	3	1061,28	-272,525	651947	6,86158e+06	8,71983	-9,65264	13,008
✓	4	1083,27	-1613,61	652012	6,86156e+06	6,50587	4,24849	7,7702
✓	5	1134,87	-2078,79	652035	6,86155e+06	-5,37092	-0,0127509	5,37093
✓	6	2297,82	-2051,72	652056	6,86161e+06	14,7498	4,60178	15,451
✓	7	2911,97	-1894,36	652058	6,86164e+06	-9,36134	2,58321	9,71121
✓	8	1933,75	-2063,38	652049	6,86159e+06	-7,50333	-4,17756	8,5879
✓	9	1840,12	-1330,47	652012	6,8616e+06	7,65004	-15,9594	17,6982

Helmert:
Average error: 20.52
Standard deviation: 10.67

Polynomial 2:
Average error: 13.12
Standard deviation: 9.23

Helmert / Poly2 / Poly3 comparative results

Helmert



Polynomiale 2

Polynomiale 3 !!!!

A suivre… TP5

Créer des données :

- créer des nouvelles couches vectorielles
- saisir des données en se basant sur une couche raster
    - choix de modélisation
    - choix sémantiques et attributs
    - topologie