University of Southampton

Southampton Geospatial Summer School

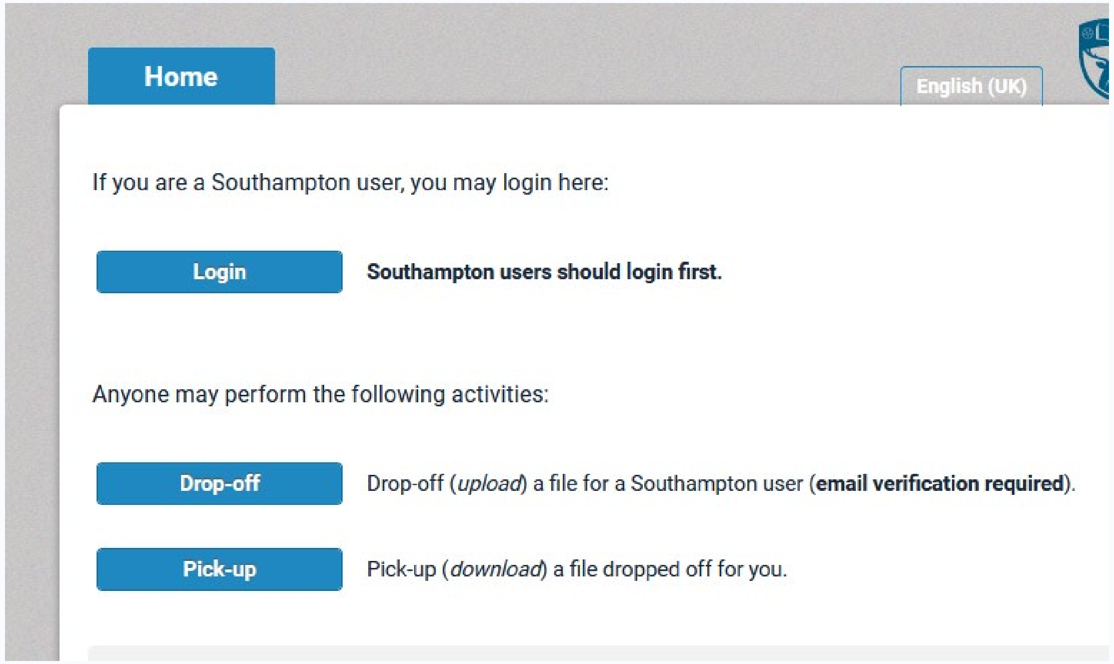
QGIS Image Analysis and Classification

Adapted from ‘Satellite Data Analysis and Machine Learning Classification with QGIS, Bratic G. and Brovelli M. A., Politecnico di Milano, 2021’

11/09/2024

Data Download

- Browse to safesend.soton.ac.uk
- Click **Pick-up**
- Enter in the following Claim ID and Passcode:

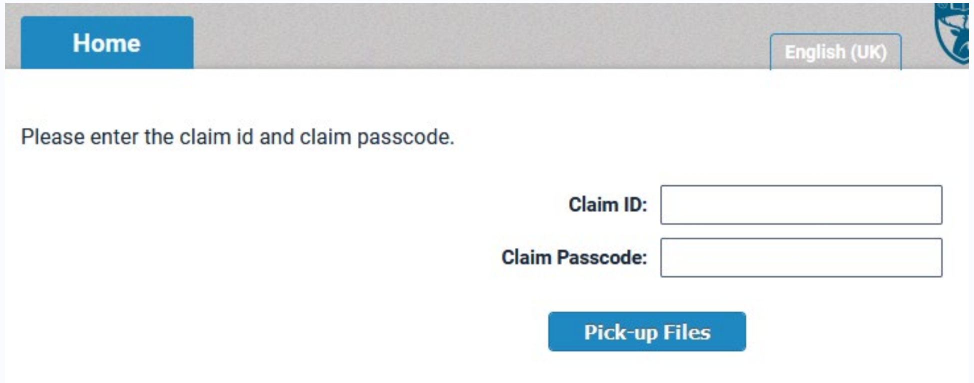


Claim ID:

Nc5NesADaJhycEDW

Passcode:

9rXukNrsEvBBnReU



QGIS Introduction

QGIS main graphical interface

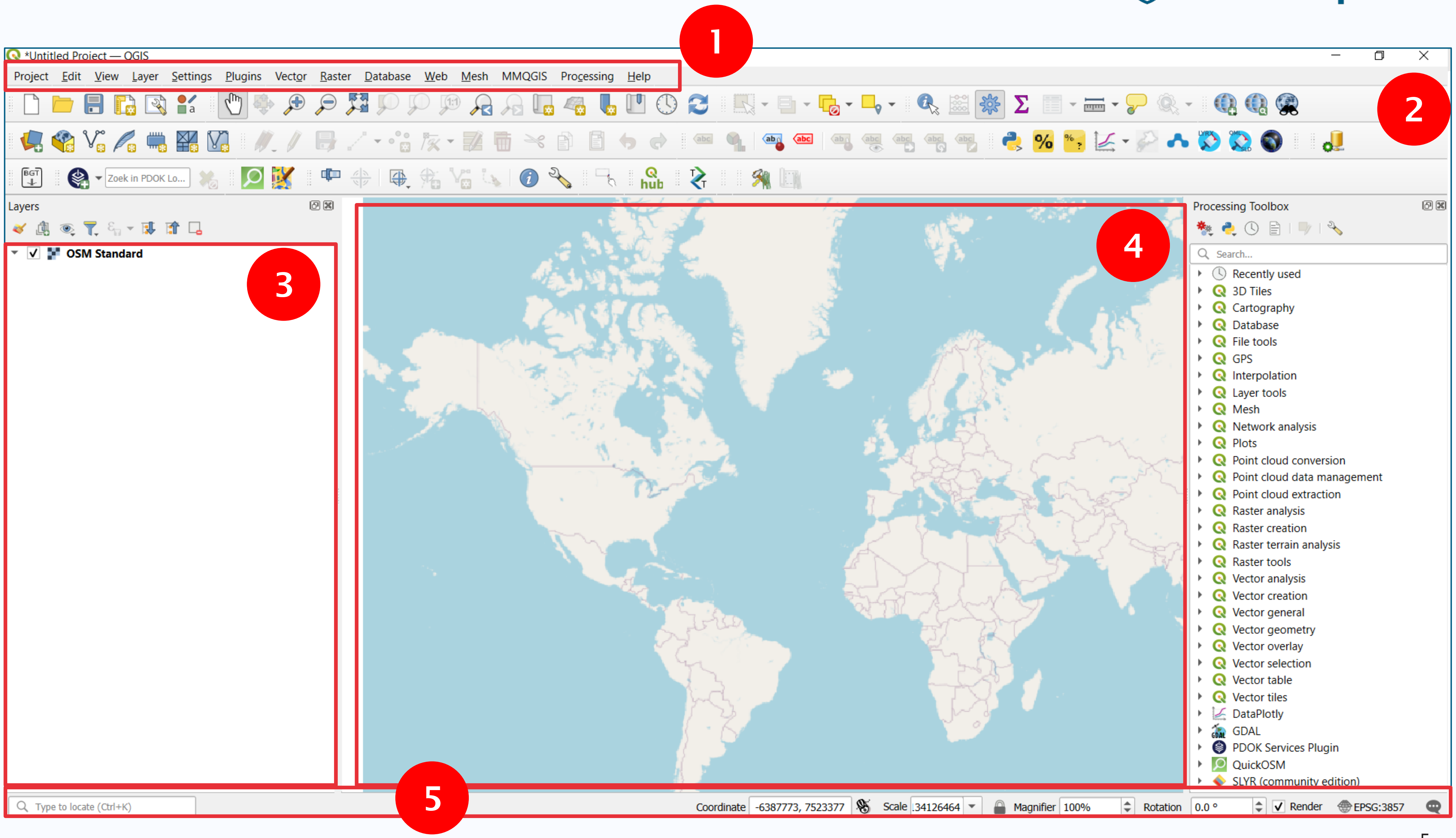
1. Menu: main features and functions

2. Toolbars: common features in a single click

3. Layer panel: list of the project layers

4. Map panel: geographic display of active layers

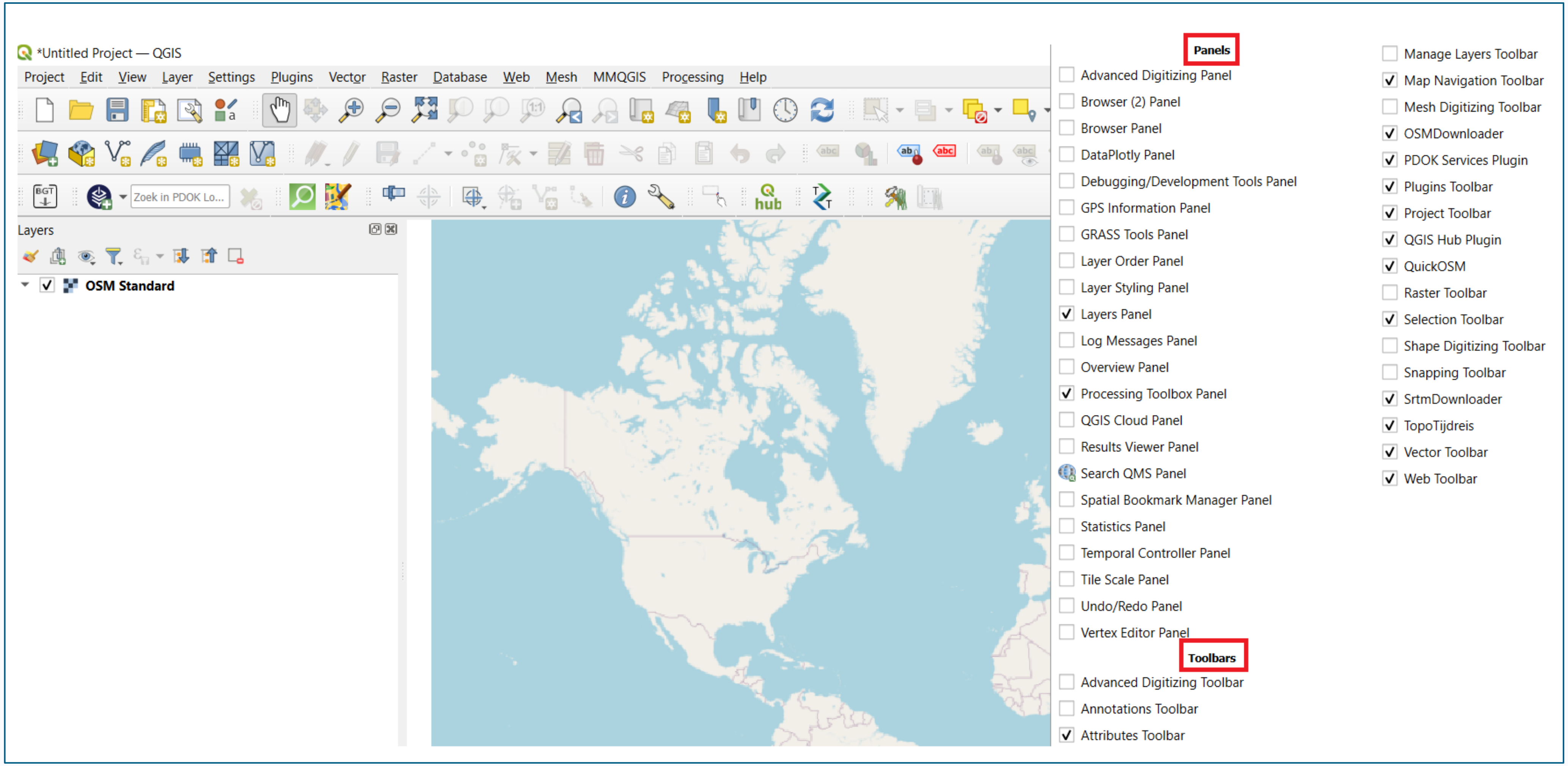
5. Status bar: current coordinates, scale, rotation and coordinate system



Display/hide panels and toolbars

Additional panels and toolbars can be activated or deactivated.

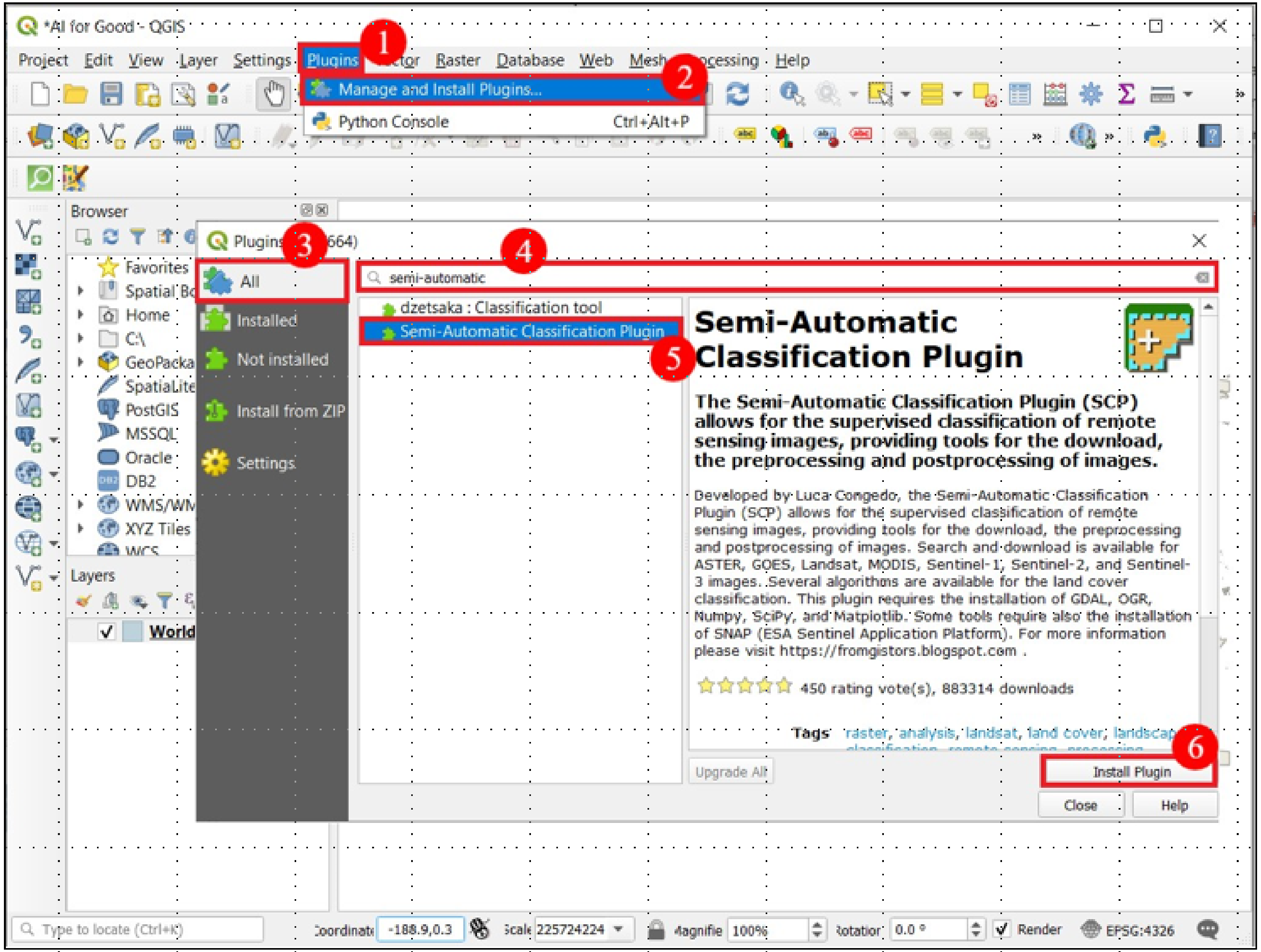
Right-click on an empty space (without icons) in the Toolbars or Menu bar section and you will get a full list of available panels and toolbars.



QGIS plugins

Install QGIS plugins

University of Southampton



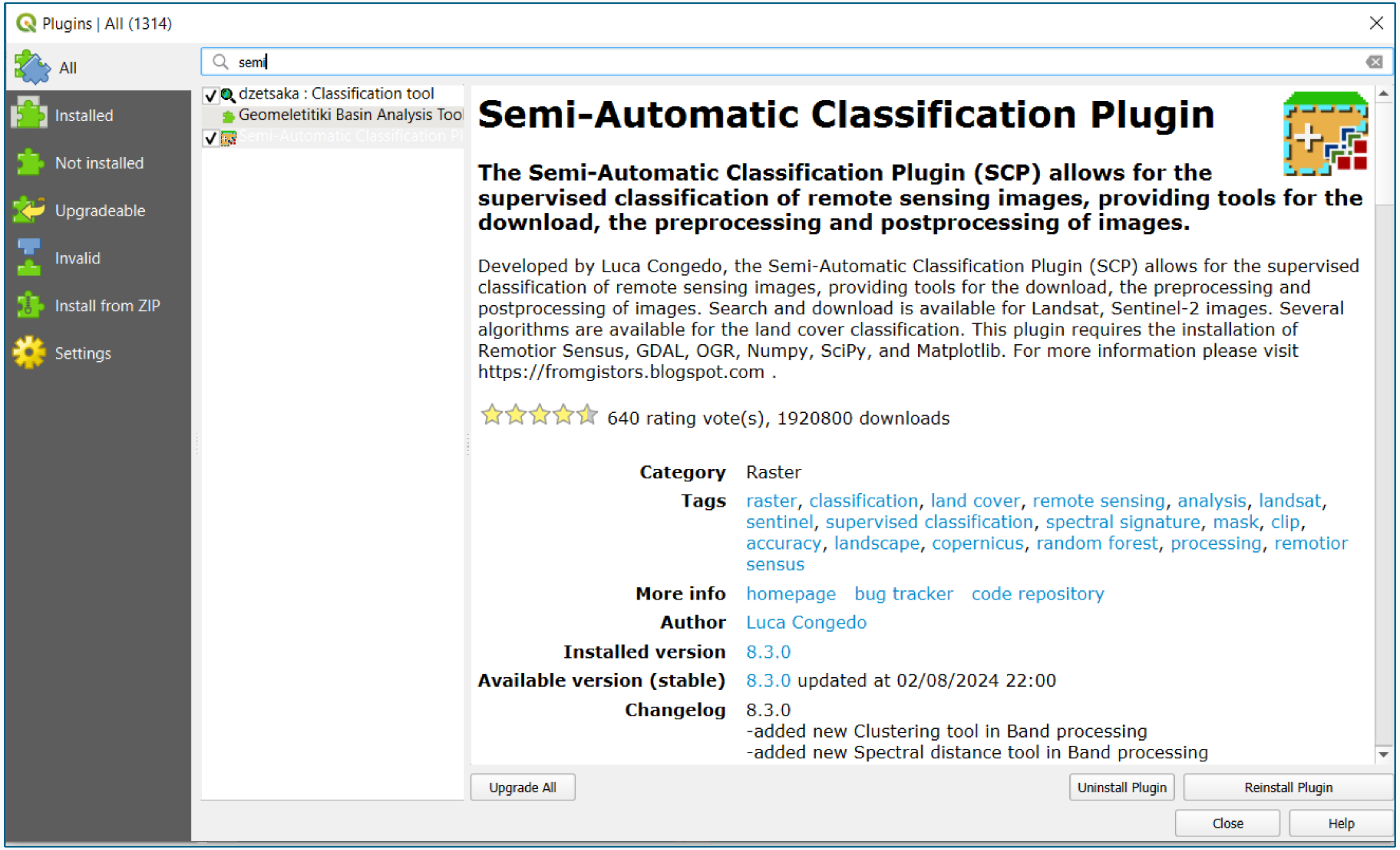
For installing a plugin

1. Go to the **Plugins** menu
2. Select **Manage and Install Plugins**
3. Go to the tab **All**
4. Type name of the plugin
5. Select the plugin in the list
6. Click on **Install Plugin**

- Please follow the procedure above to install following plugins by inserting their name in the search field (step 4):

- *Semi-Automatic Classification Plugin*
- *dzetsaka: Classification tool*

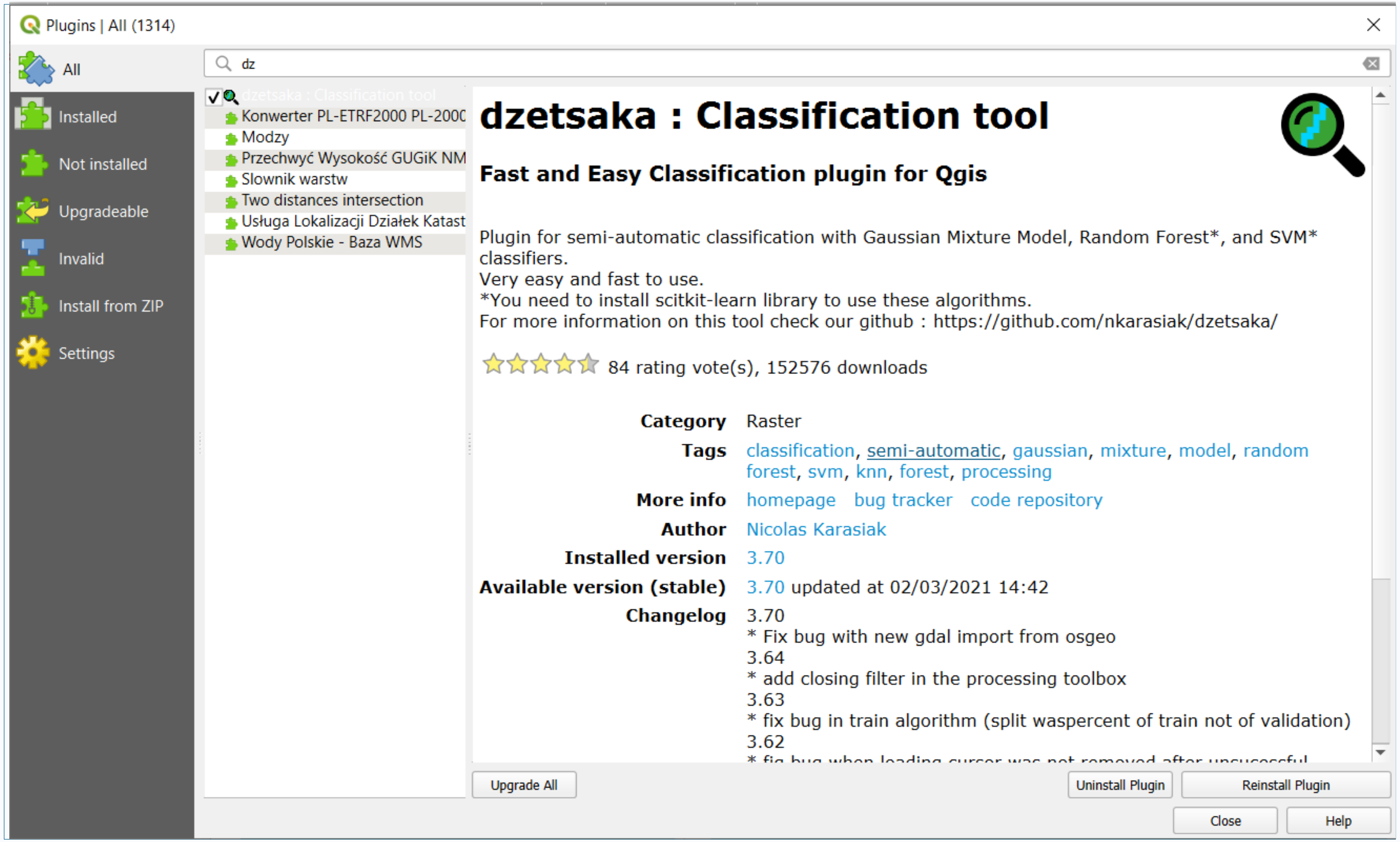
Install QGIS plugins



Semi-Automatic Classification Plugin (SCP) is used for download preprocessing, classification (supervised and unsupervised) and postprocessing of open satellite imagery (ASTER, Landsat, MODIS, Sentinel-2 and Sentinel-3).

We will use SCP plugin for downloading and preprocessing Sentinel-2 imagery as well as for algebraic operations with bands.

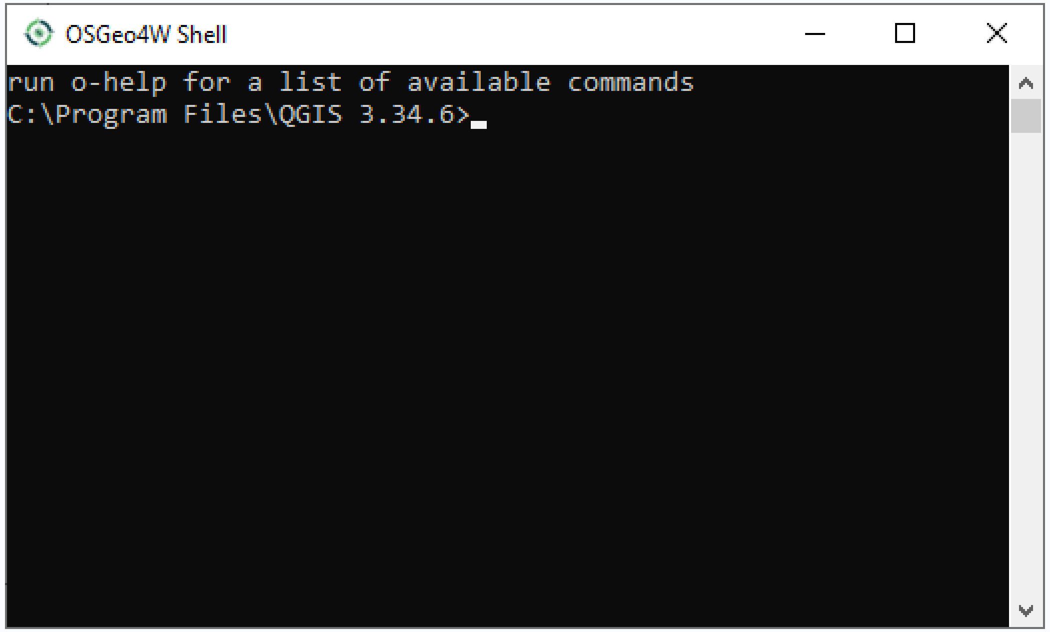
Install QGIS plugins



dzetsaka: Classification tool is a plugin that allows classification of satellite imagery with Gaussian Mixture Model, Random Forest, Support Vector Machine and K-Nearest Neighbors.

Installing Additional Python libraries

- Both the dzetsaka and Semi-Automatic Classification plugins require additional Python libraries installing to be able to function correctly.
    1. Open the OSGeo4W Shell from the Start menu
    2. Enter the following commands, pressing return after each:



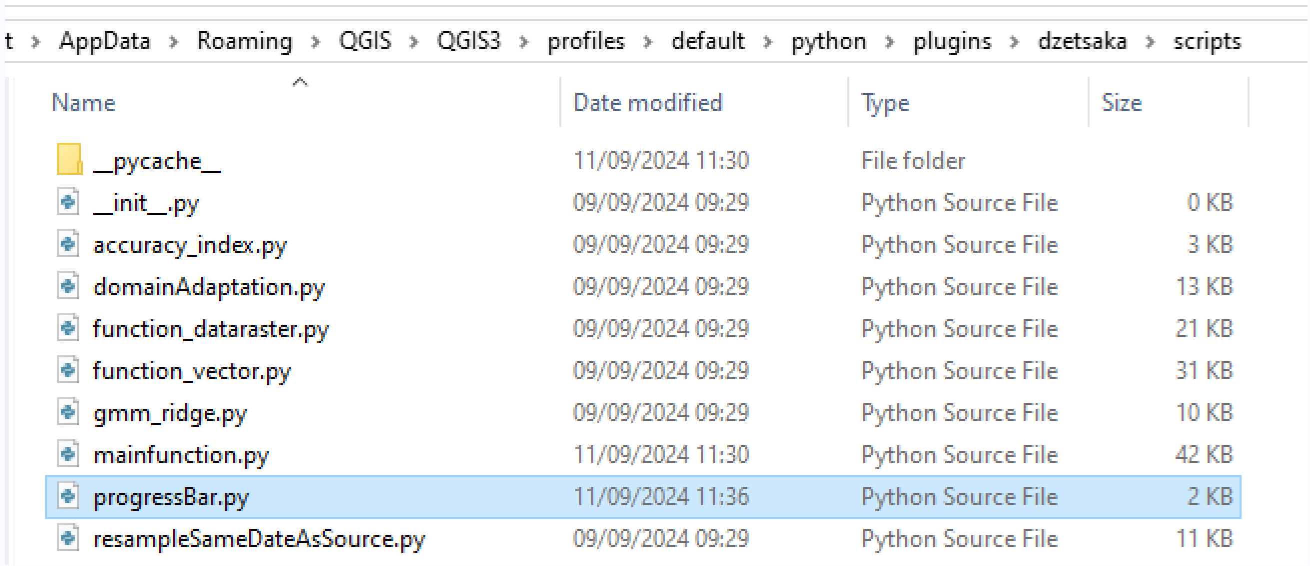

```
python3 -m pip install scikit-learn

python3 -m pip install --upgrade remotior-sensus
```

A Bugfix!

- The dzetsaka plugin appears to have a bug, probably introduced by an upgrade to the Python version that QGIS runs.
- Choose: *Settings > User Profiles > Open Active Profile Folder*
- Browse to *python\plugins\dzetsaka\scripts*
- Open progressBar.py in Notepad
- Change line 38 from: `prgBar.setMaximum(inMaxStep)`

  To: `prgBar.setMaximum(int(inMaxStep))`
- Save and restart QGIS

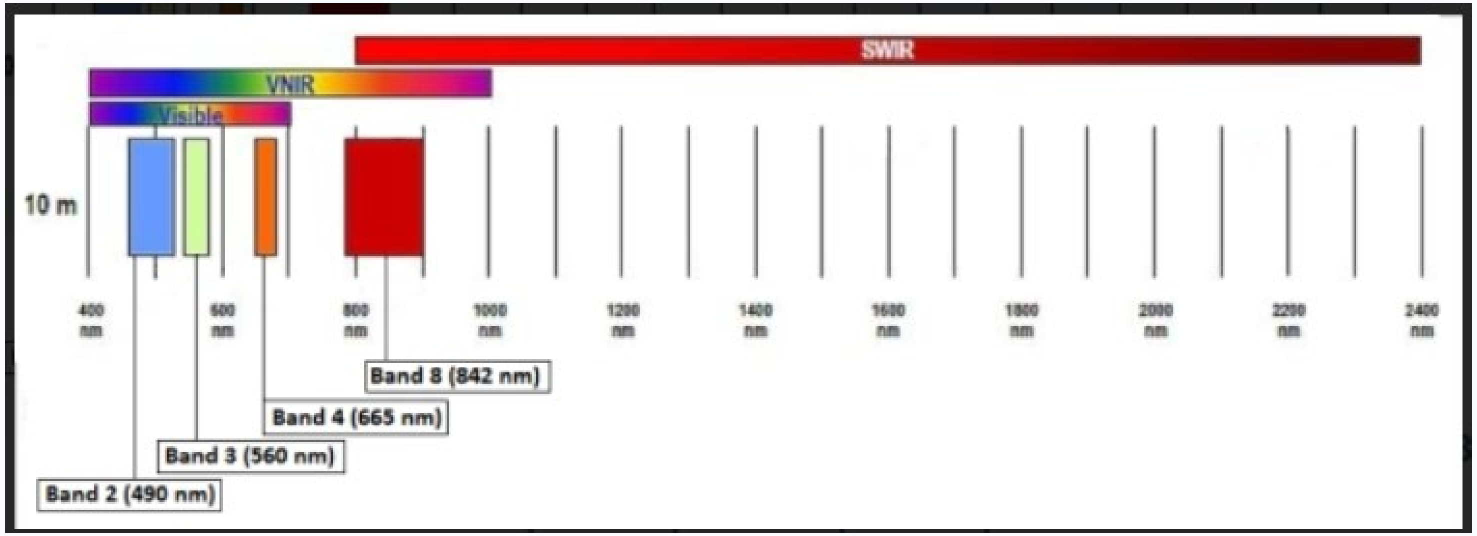


About satellite imagery

Multispectral satellite imagery

University of Southampton

- ✓ Satellite imagery is data that comes from sensors on board satellites
- ✓ Multispectral sensors capture electromagnetic (EM) radiation reflected by the Earth surface.
- ✓ A multispectral image consists of bands. Each band is a raster whose values represent the EM radiation at a certain range of wavelength. The multispectral imagery captures visible EM radiation (red, green and blue), but also the short-wave infrared (SWIR) and visible and near-infrared radiation (VNIR).
- ✓ In the example of the Sentinel-2 satellite imagery – Band 2, Band 3, and Band 4 are capturing the visible EM radiation, while band 8 represents the non-visible near infrared EM radiation.
- ✓ From the figure it can be seen also that not every band has the same range of values (i.e., range of values in Band 2 is larger than in Band 3 or Band 4)



Source: https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/spatial

Multispectral satellite imagery

The SENTINEL-2 Multispectral Instrument (MSI) samples 13 spectral bands: four bands at 10 meters, six bands at 20 meters and three bands at 60 meters of spatial resolution.

Satellite imagery, including Sentinel-2, is a valuable source of information for GIS due to frequent updates (revisit every 5 days).

Sentinel-2 Bands	Central Wavelength (µm)	Resolution (m)
Band 1 - Coastal aerosol	0.443	60
Band 2 - Blue	0.490	10
Band 3 - Green	0.560	10
Band 4 - Red	0.665	10
Band 5 - Vegetation Red Edge	0.705	20
Band 6 - Vegetation Red Edge	0.740	20
Band 7 - Vegetation Red Edge	0.783	20
Band 8 - NIR	0.842	10
Band 8A - Vegetation Red Edge	0.865	20
Band 9 - Water vapour	0.945	60
Band 10 - SWIR - Cirrus	1.375	60
Band 11 - SWIR	1.610	20
Band 12 - SWIR	2.190	20

Source: https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/overview

Load auxiliary vector data

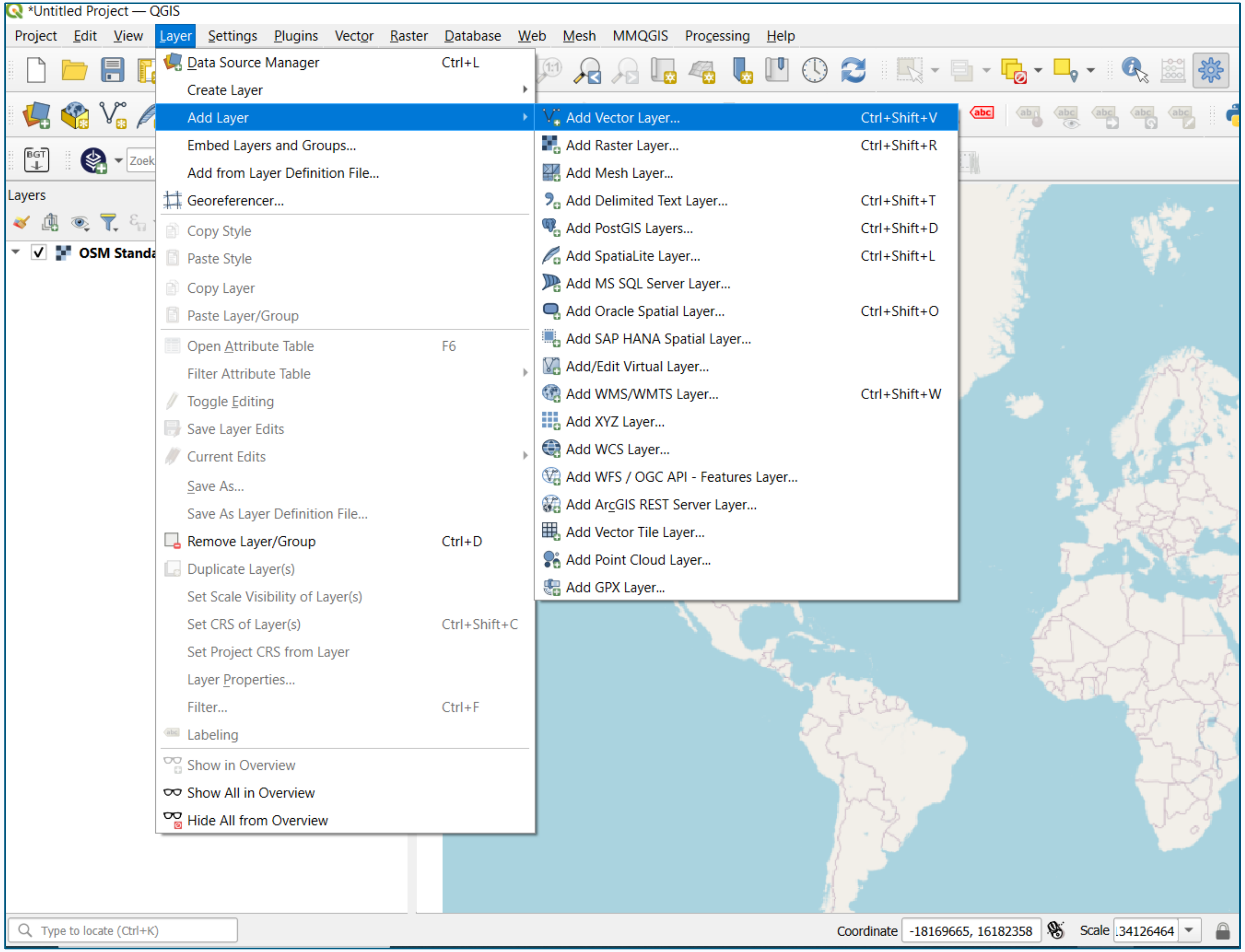
Area of interest

University of Southampton

Vector *working_area.shp* is the vector file which delineates a rectangular area of interest. The area of interest includes *Al Hashimiyya* village in northern *Jordan*. It was selected arbitrarily.

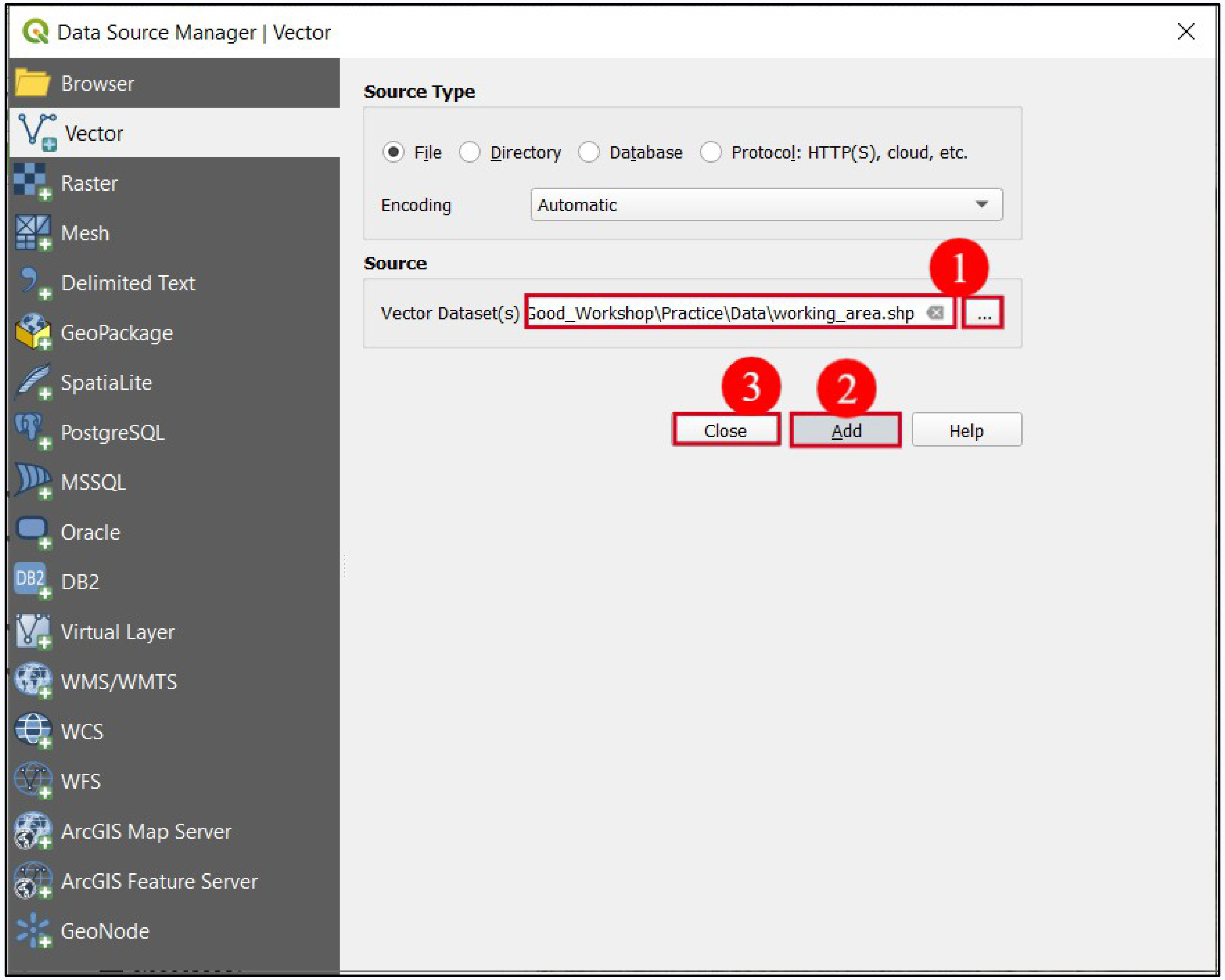
Load the vector by selecting:

1. **Layer** menu
2. **Add Layer**
3. **Add Vector Layer**



Area of interest

1. Navigate to the folder with data and select working_area.shp
2. Add Layer
3. Close Data Source Manager

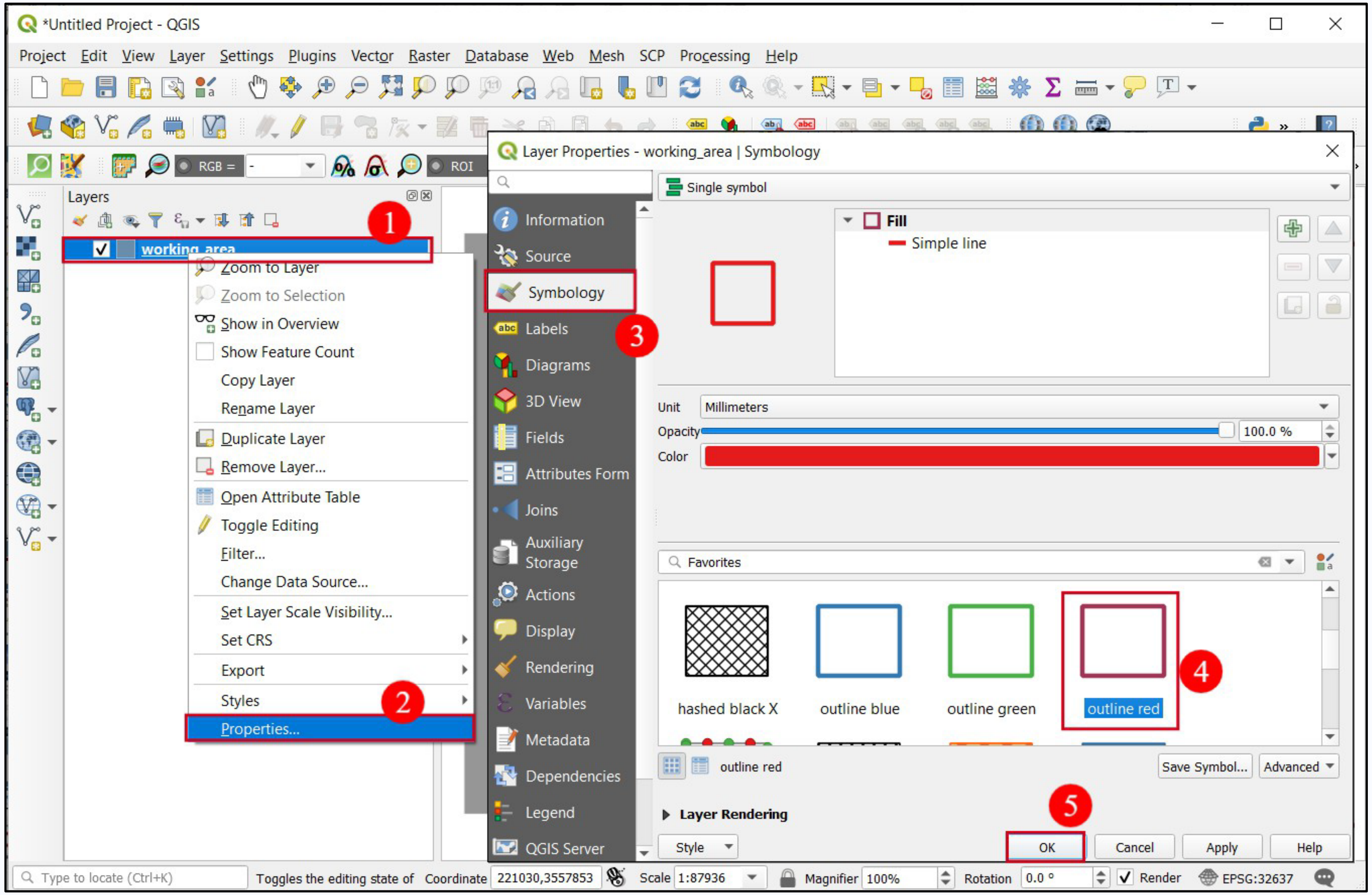


Area of interest - Symbology

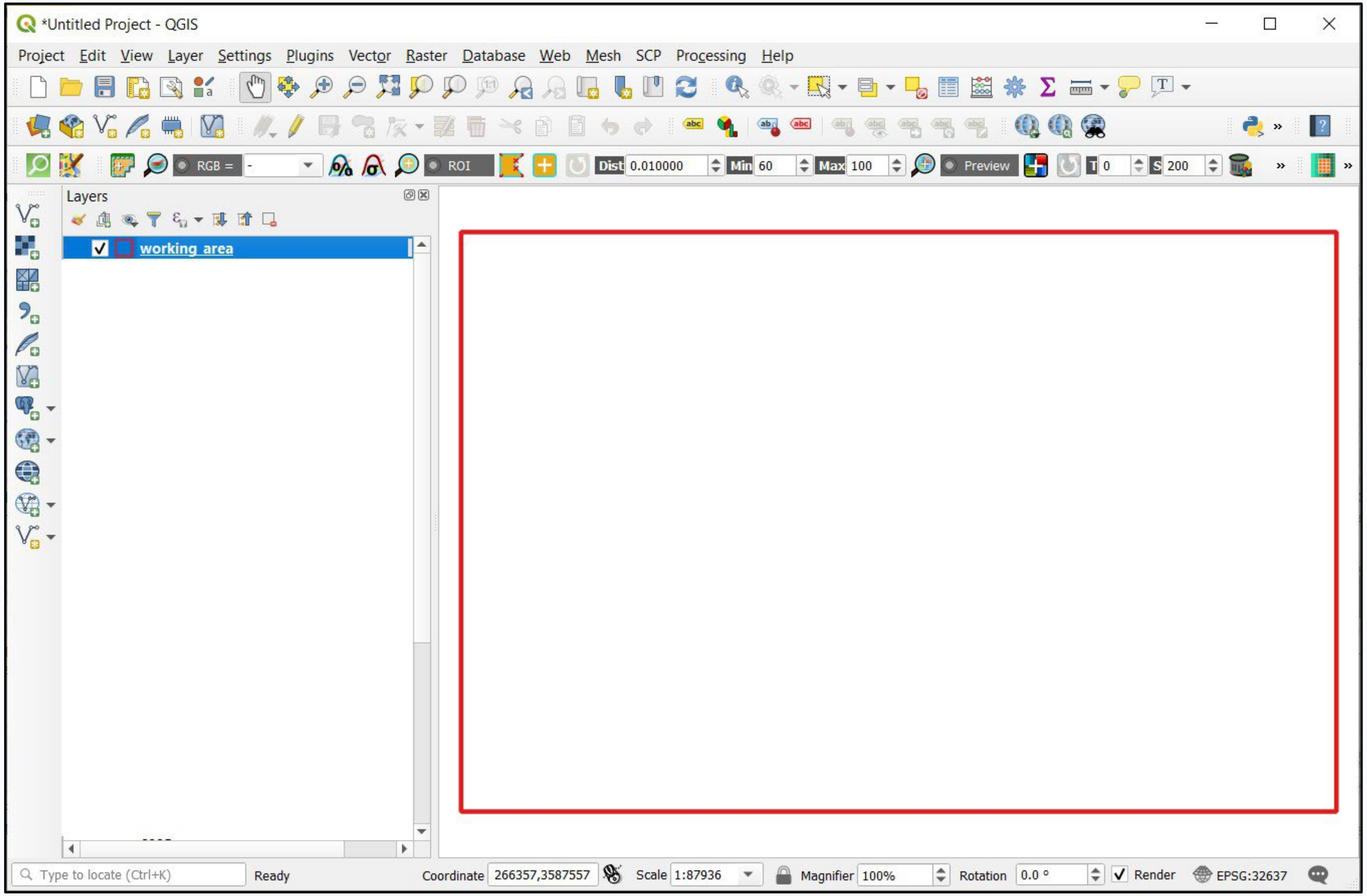
Apply vector visualization so that vector boundaries are in red while the polygon is transparent. This kind of visualization is suitable when we are interested only in vector boundaries, as it is the case in this example.

To apply this visualization:

1. Right-click on vector *working_area* in the Layers Panel
2. Select layer **Properties**
3. Go to **Symbology** tab
4. Select ***outline red*** predefined symbology
5. Click on **OK** to apply the symbology



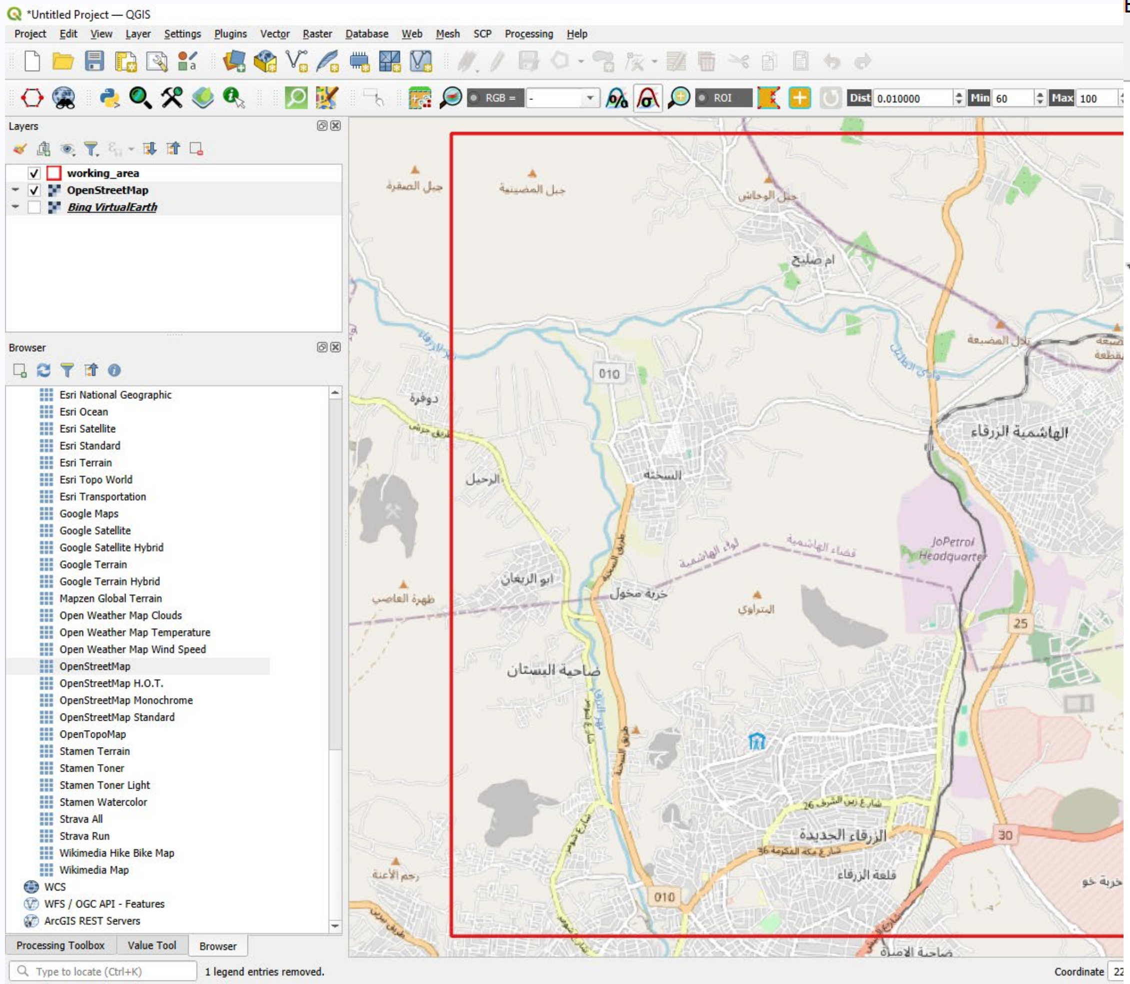
Area of interest

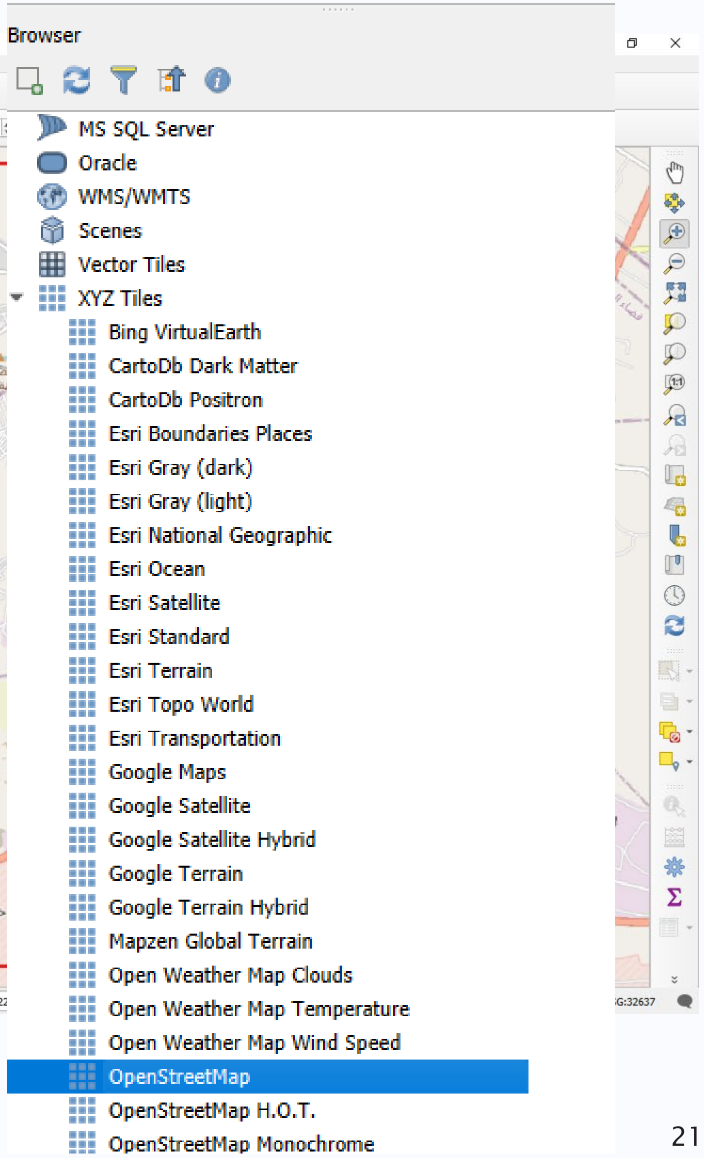


Area of interest – Adding Context

University of Southampton

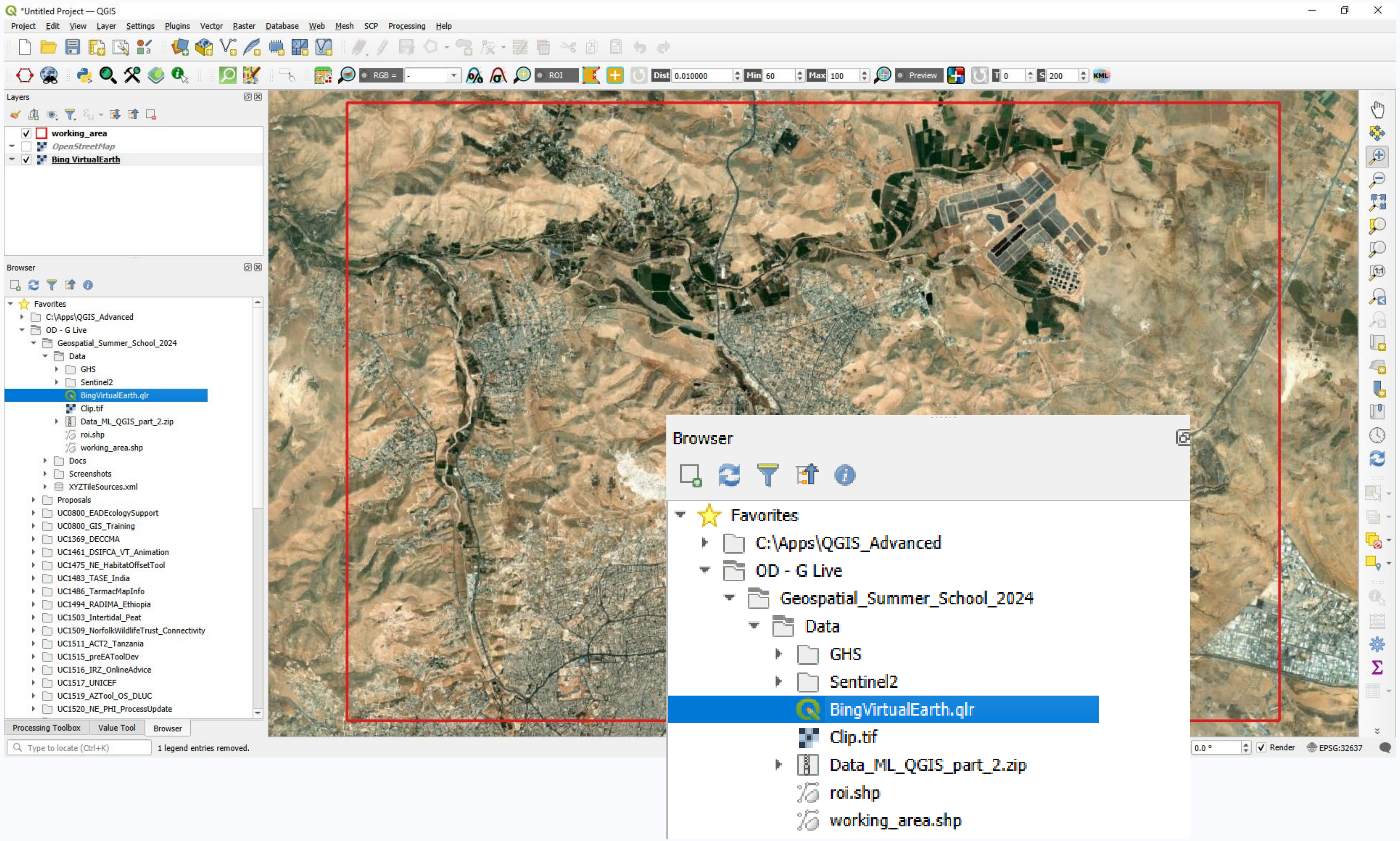
- Open the Browser panel and find the 'XYZ tiles' entry
- Expand this and drag the OpenStreetMap entry on to the map





Area of interest – Adding Context

- Using the Browser panel, find the location of the downloaded datasets.
- Drag the Bing VirtualEarth file on to the map

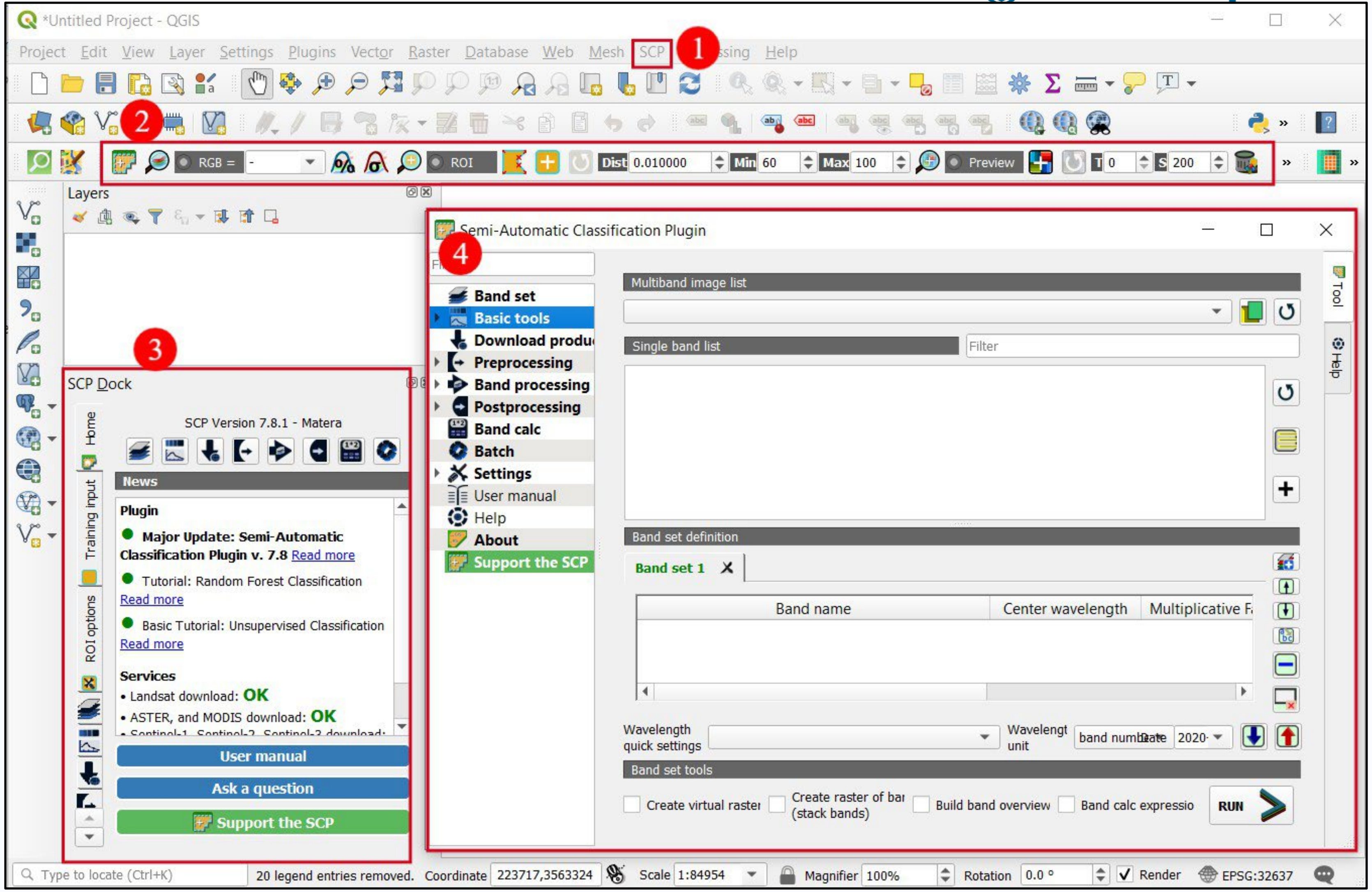


The SCP plugin

The functionalities of Semi - Automatic Classification Plugin (SCP) can be accessed from:

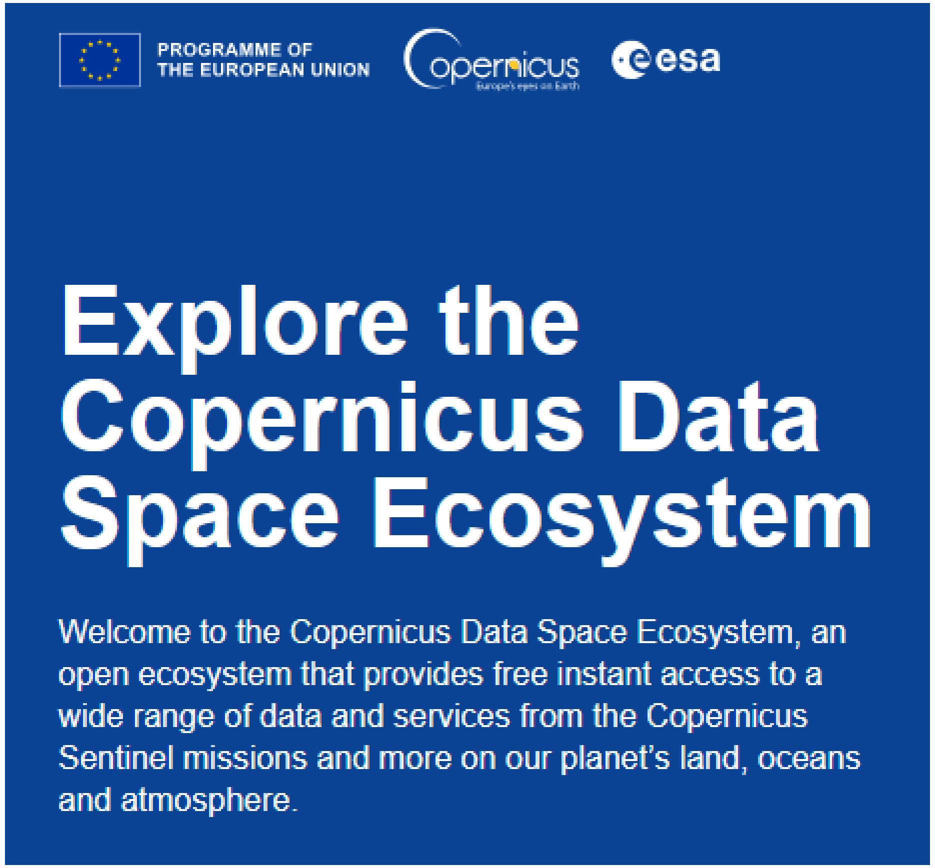
1. SCP Menu Bar
2. SCP Working Toolbar
3. SCP Dock Panel

When a functionality is selected it can be parametrized and executed from the SCP Command Panel (4).



Sentinel 2 data

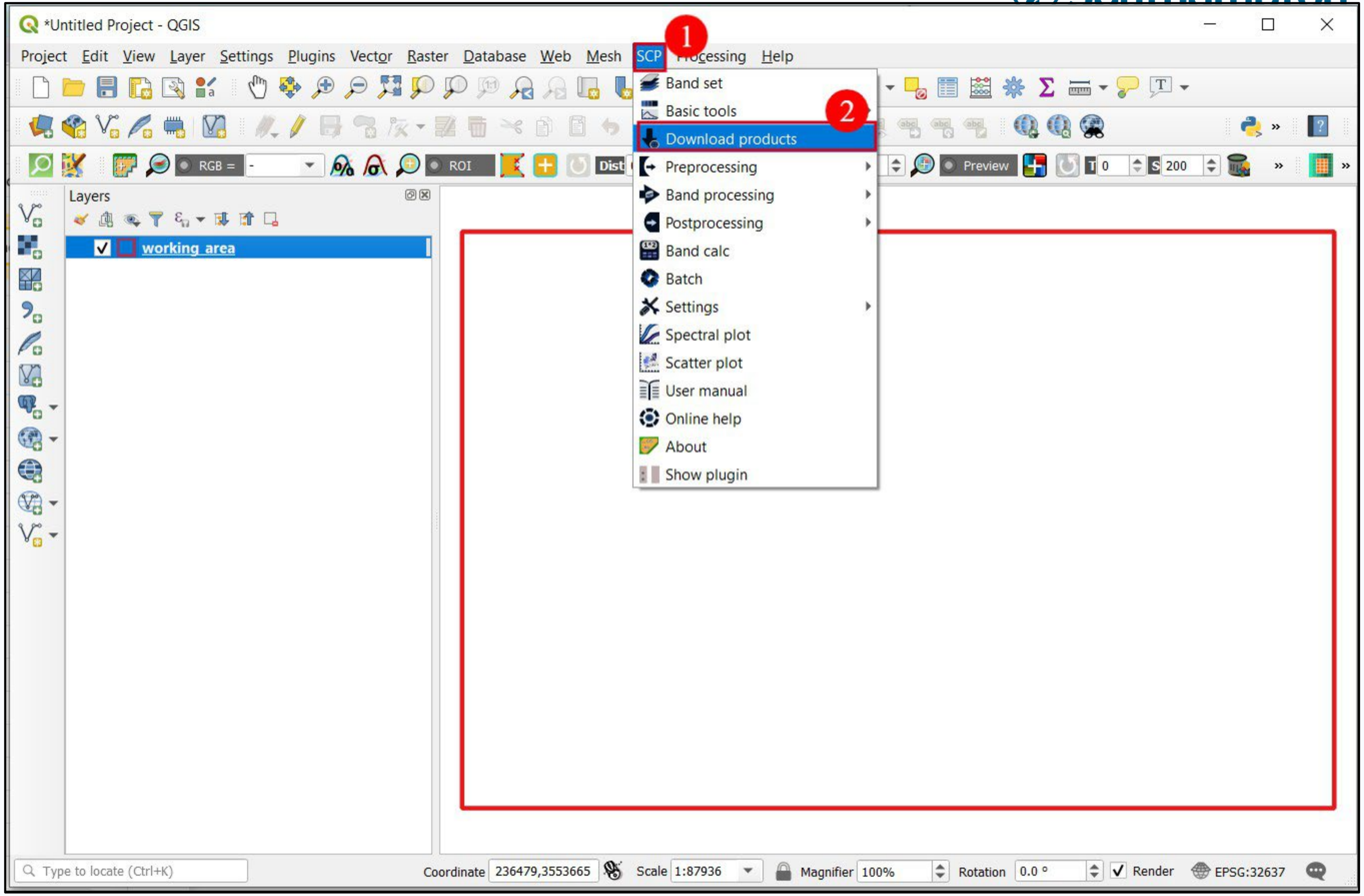
- We are going to query and browse the data catalogue of the ESA Sentinel-2 mission. This is enabled by the Copernicus Data Space Ecosystem which provides developers and data users with a set of API to access the imagery catalogue
- The API is implemented within the SCP Plugin by enabling the access to Sentinel-2 imagery data (and many others) directly from QGIS



SCP plugin – Image download

To visualize the options for imagery download

1. Go to the **SCP menu**
2. Select **Download products**



On the Login data tab, it is necessary to insert credentials for specific data provider.

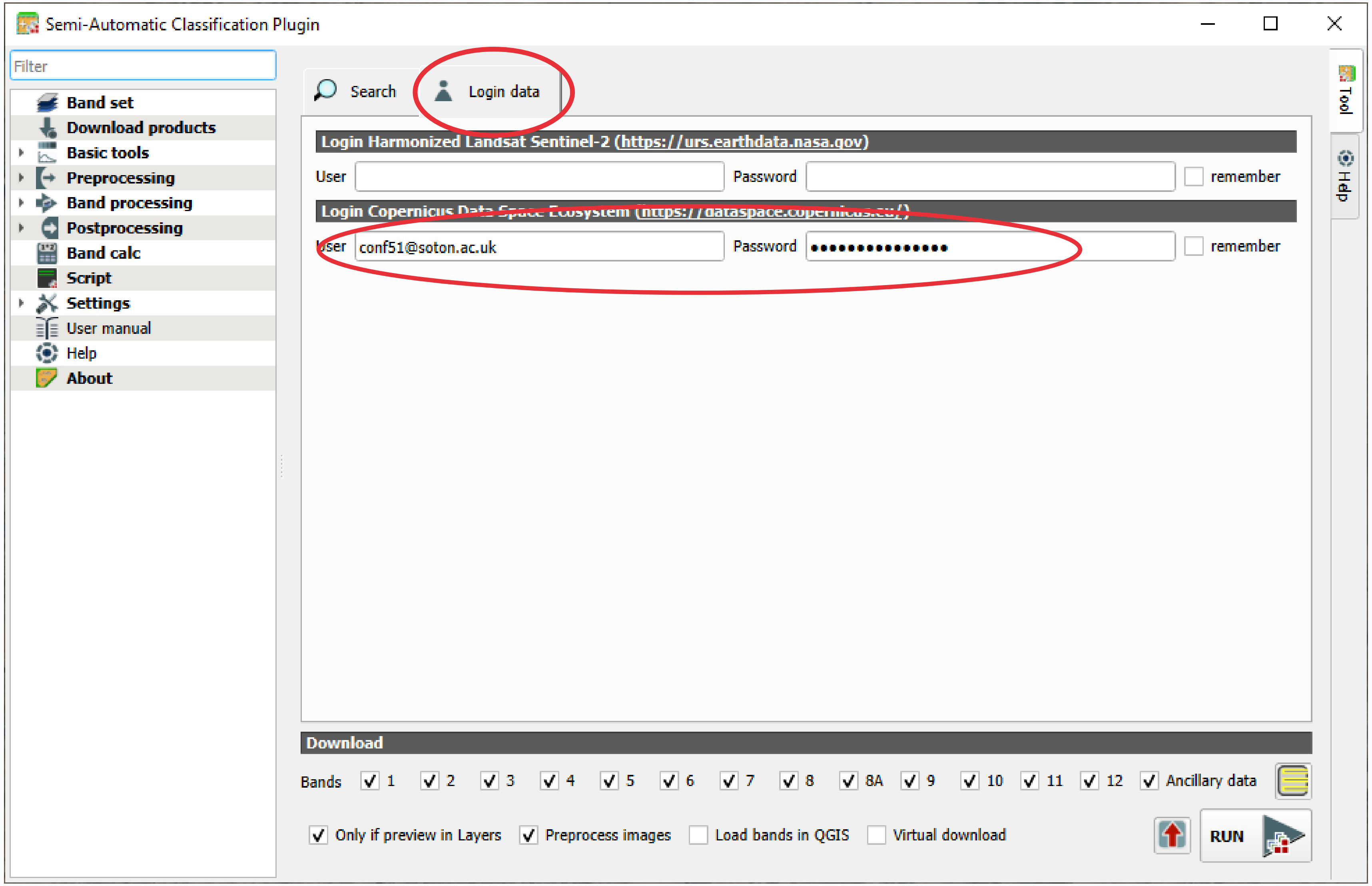
If you have credentials insert them in the User and Password fields, otherwise use the ones below:

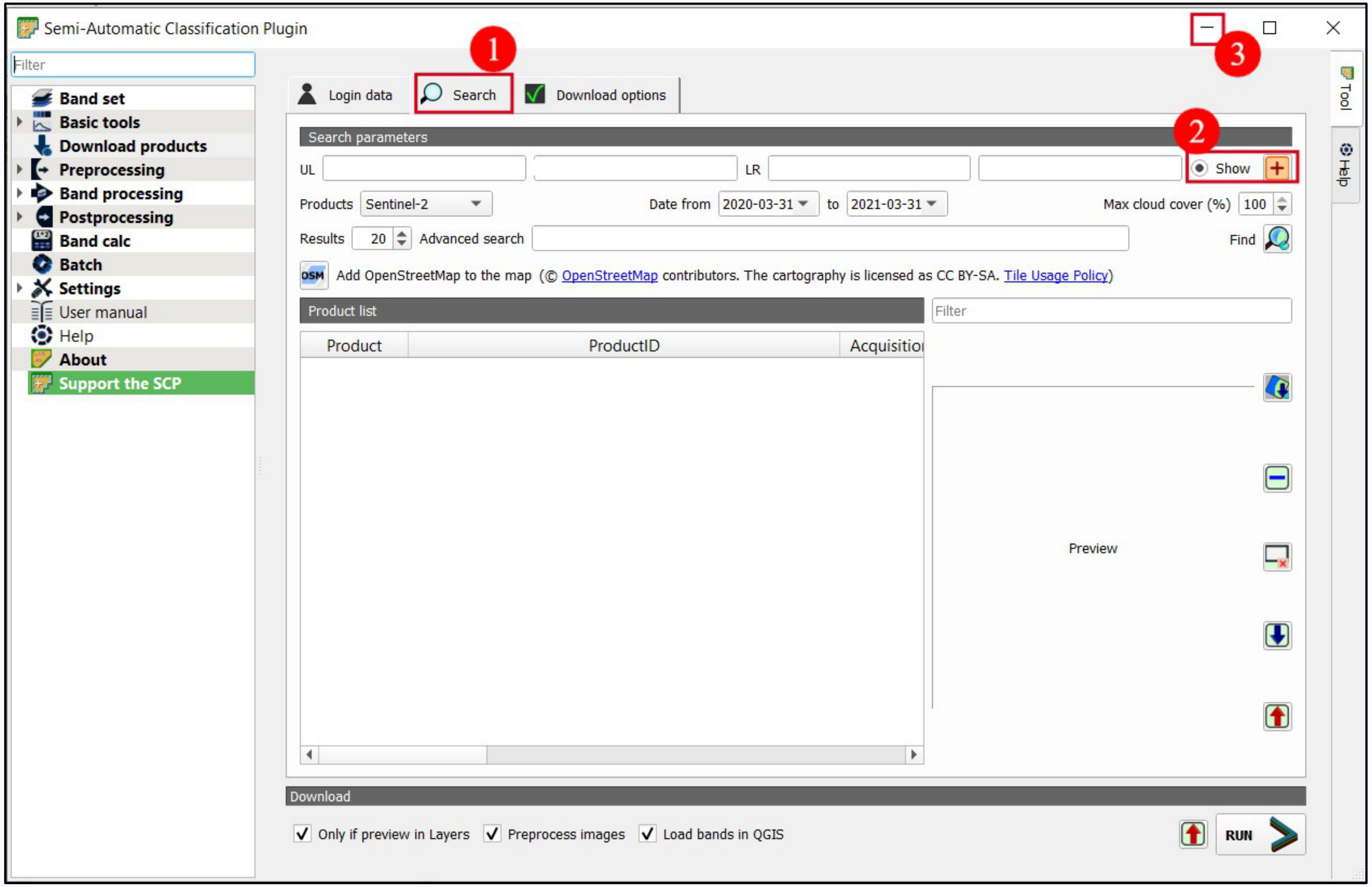
User:

conf51@soton.ac.uk

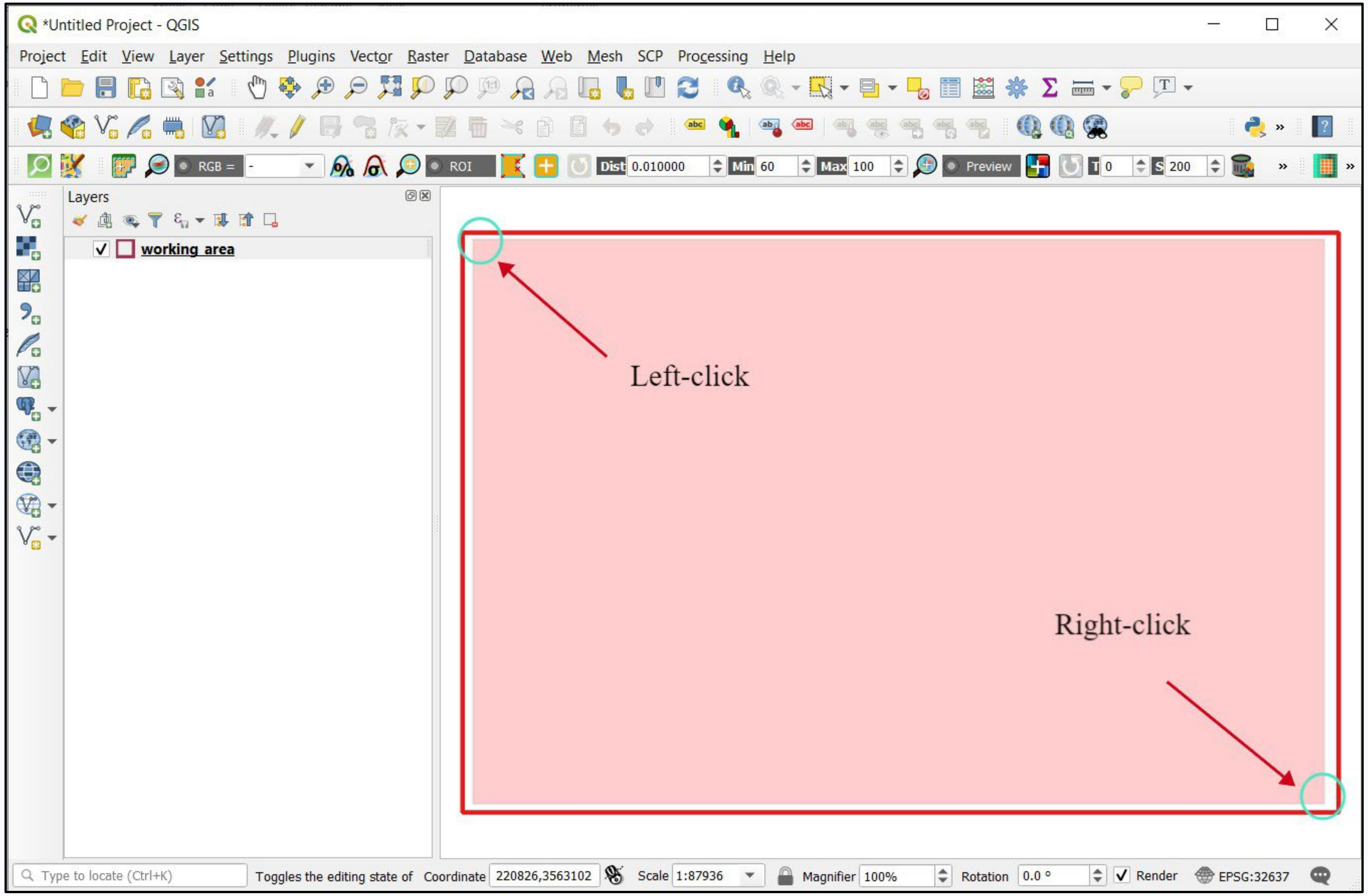
Password:

QGIS_Rocks_2024



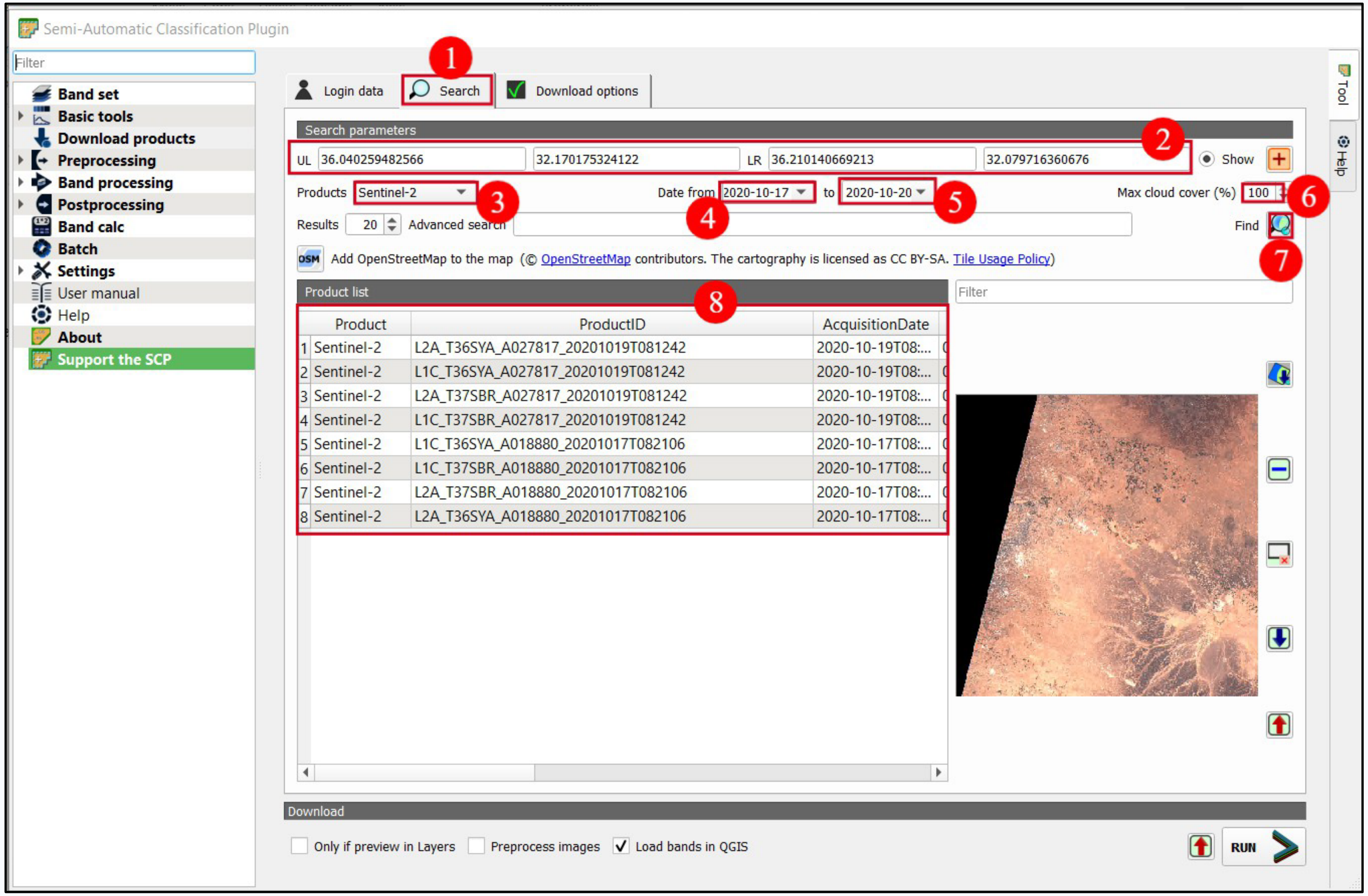


1. On the Search tab specify the parameters of the imagery to download.
2. For defining the area in which to search images we can click on and keep Show option active. This allows us to select the region in the Map panel.
3. Minimize the SCP Command Panel (3) to visualize Map panel and select the area

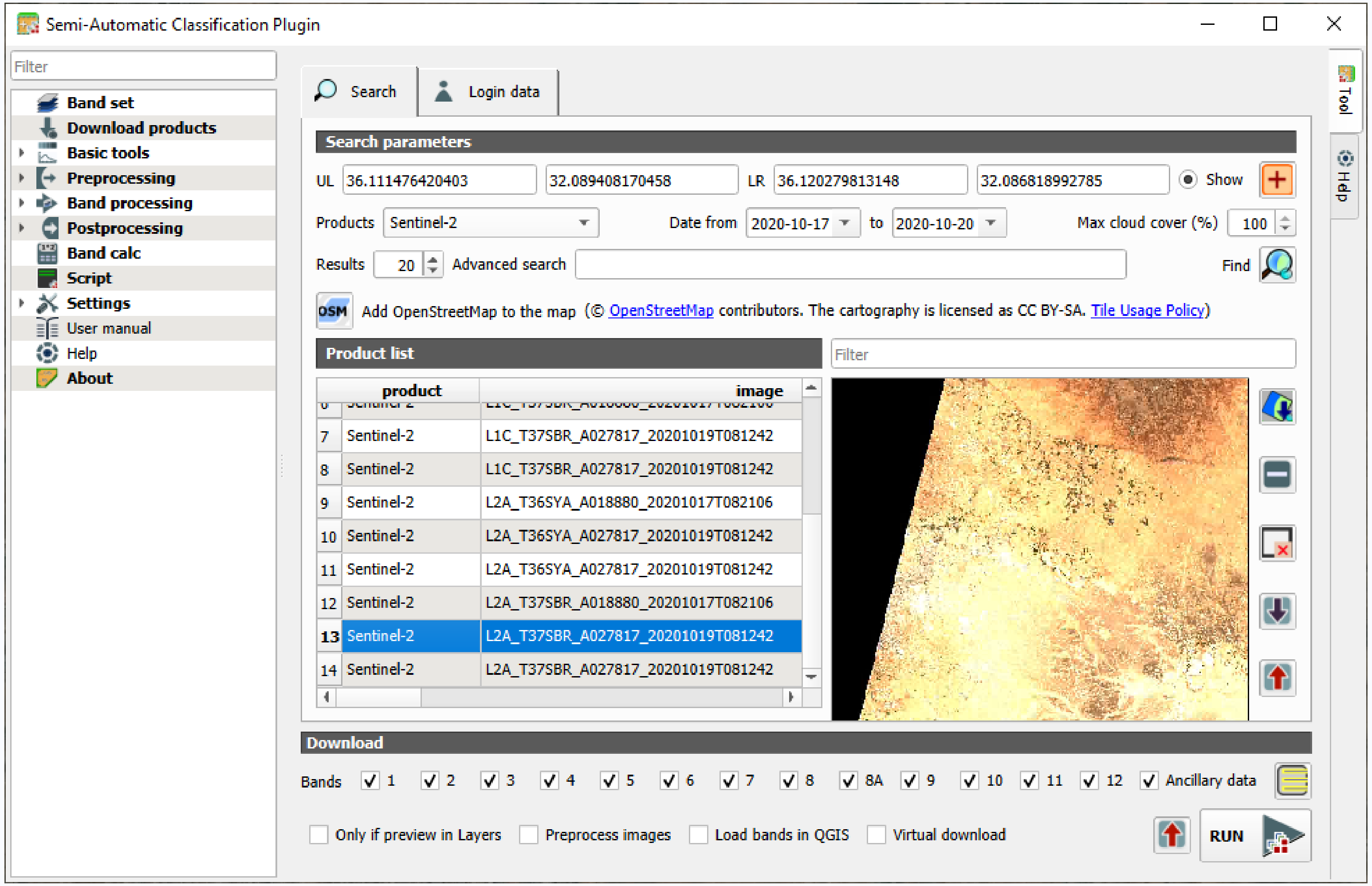


- In the Map Panel try to reproduce rectangle of working_area.shp
- Use **left-click** of the mouse to select **upper-left corner** of the area
- Use **right-click** of the mouse to define **lower-right corner** of the area
- The area selected will be filled with light red color
- The selected area does not have to be precisely as *working_area.shp*

Download satellite imagery using SCP (5)



1. When you unminimize SCP Command Panel go back to the **Search** tab
2. You should see coordinates for **Upper-Left (UL) and Lower-Right (LR)** corners of the area of interest inserted automatically
3. Specify Sentinel-2 for **Products**
4. Specify **initial** acquisition **date** to 2020-10-17
5. Specify **final** acquisition **date** to 2020-10-20
6. Leave **Max cloud coverage** at 100%
7. Hit **Find** button
8. After some time, you will get list of imagery with parameters corresponding to search parameters defined



As it can be seen, there are 14 images meeting the criteria

We are going to use: L2A_T37SBR_A027817_20201019T081242 which #13 in the list on the figure. Remove all the images except the one we want to download:

2. Hold CTRL and select images to be removed
3. Click on **Remove** button

IMPORTANT NOTE

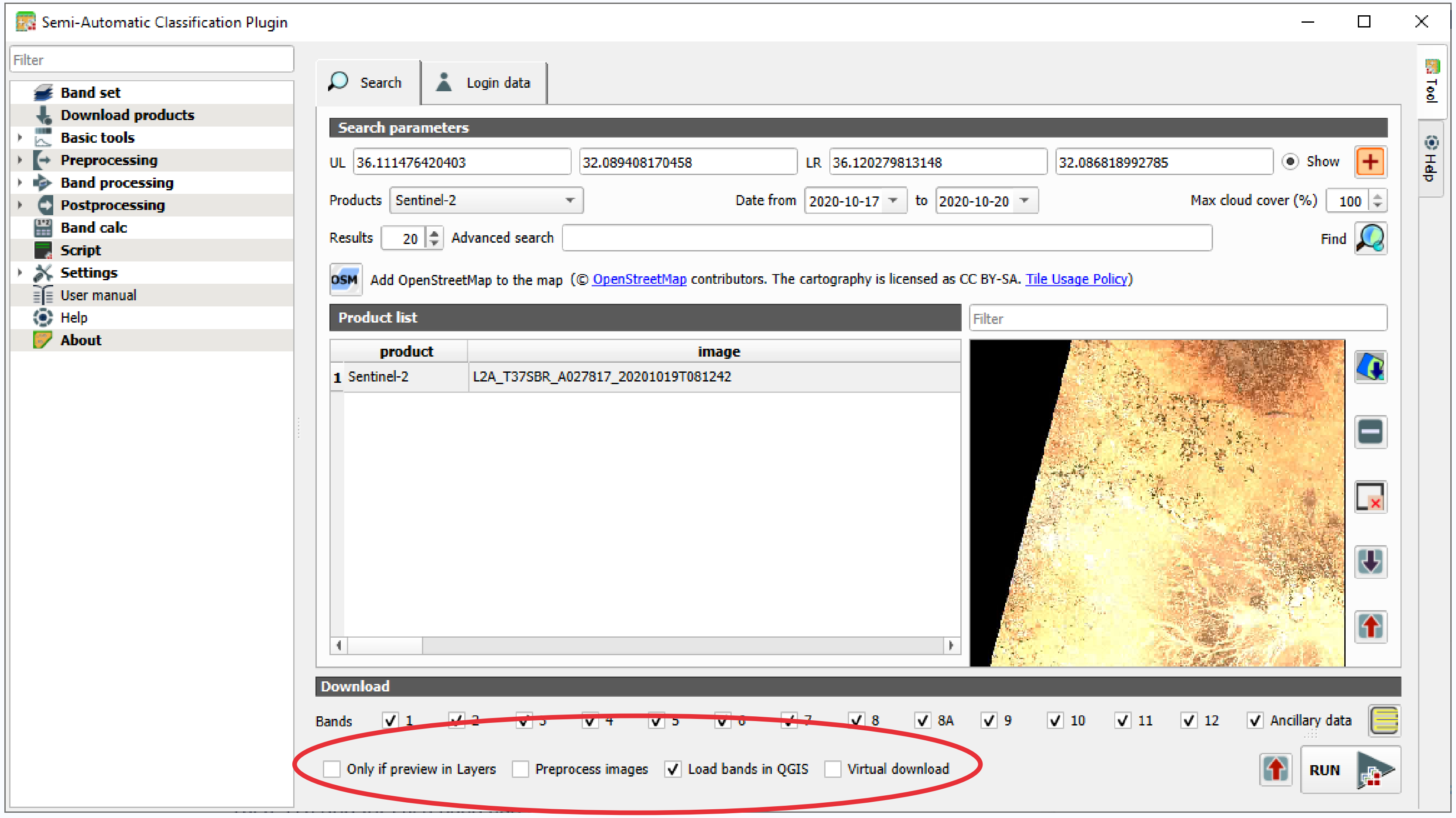
Please pay attention to the fact that, in general, when the level 2A image is available, **this image must be chosen**. The level 2A image in fact is corrected by the atmospheric reflectance, on the opposite the level 1C is not.

Level-2A products are systematically generated at the ground segment **over Europe since March 2018**, and the production was extended to **global in December 2018**.

Level-2A generation can also be performed by the user through the official SNAP tool for atmospheric correction (see http://step.esa.int) using as input the associated Level-1C product. This correction is the suggested one.

University of Southampton

1. Before starting download there are some of more options to be defined.
2. Unselect ***Only if preview in Layers*** to download images even if we do not have their preview downloaded.
3. Unselect ***Preprocess image*** because preprocessing is one of the following steps
4. Select ***Load bands in QGIS*** to load downloaded bands after download is complete
5. **Run** the download



The Blue Peter Moment

To save time during the workshop we will load the data that have been previously downloaded and provided to you in the Sentinel2 folder.

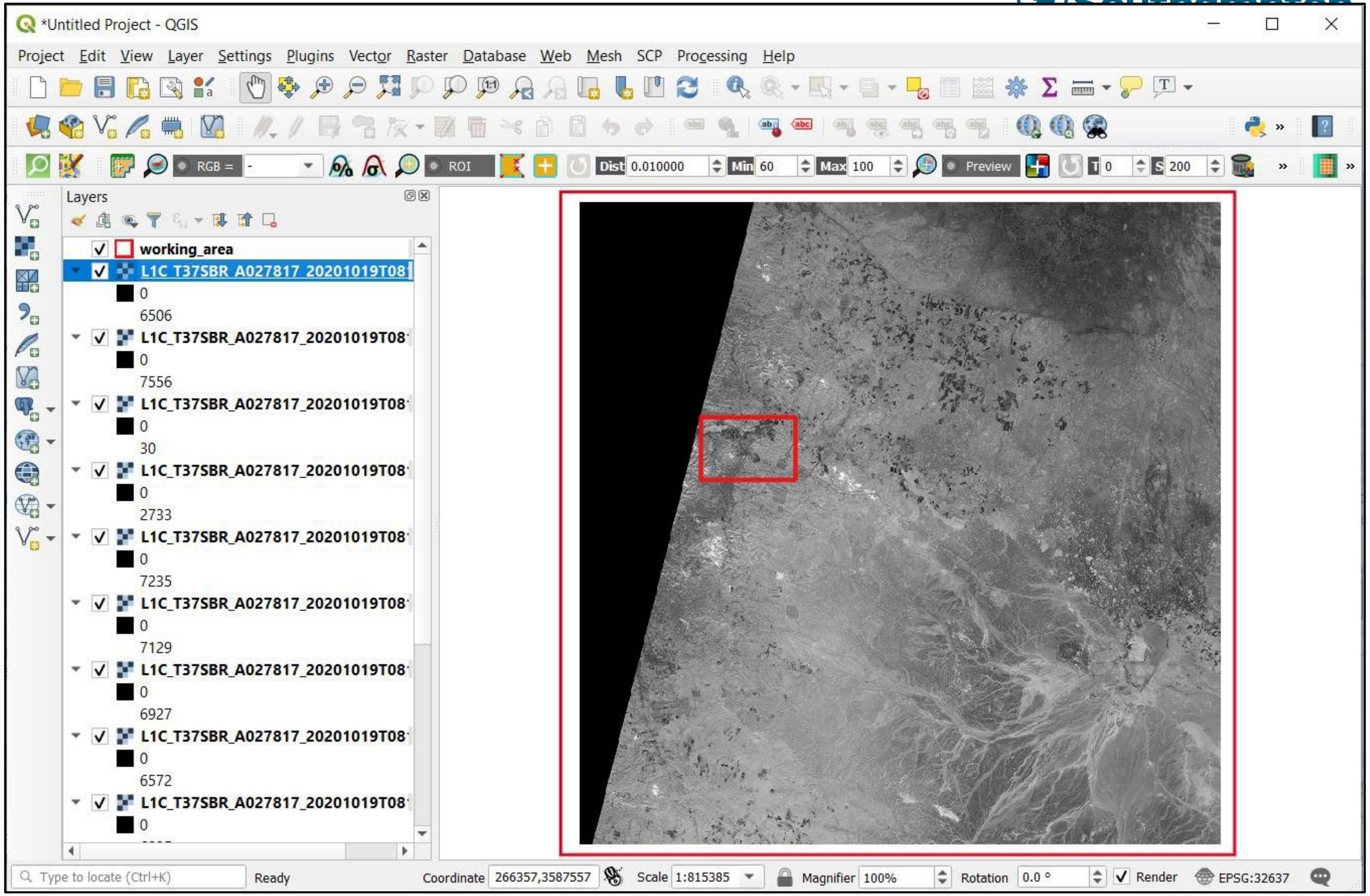
Load this into QGIS using either of the two methods described earlier:

1. Find the folder in the Browser panel and drag the image files into the map
2. Use *Layer > Add Layer > Add Raster Layer…* to open the images using a file dialog.

You should load all the .jp2 files in the folder, each one corresponds to a different band from the Sentinel 2 MSI

After the download the image bands will be automatically imported in your QGIS.

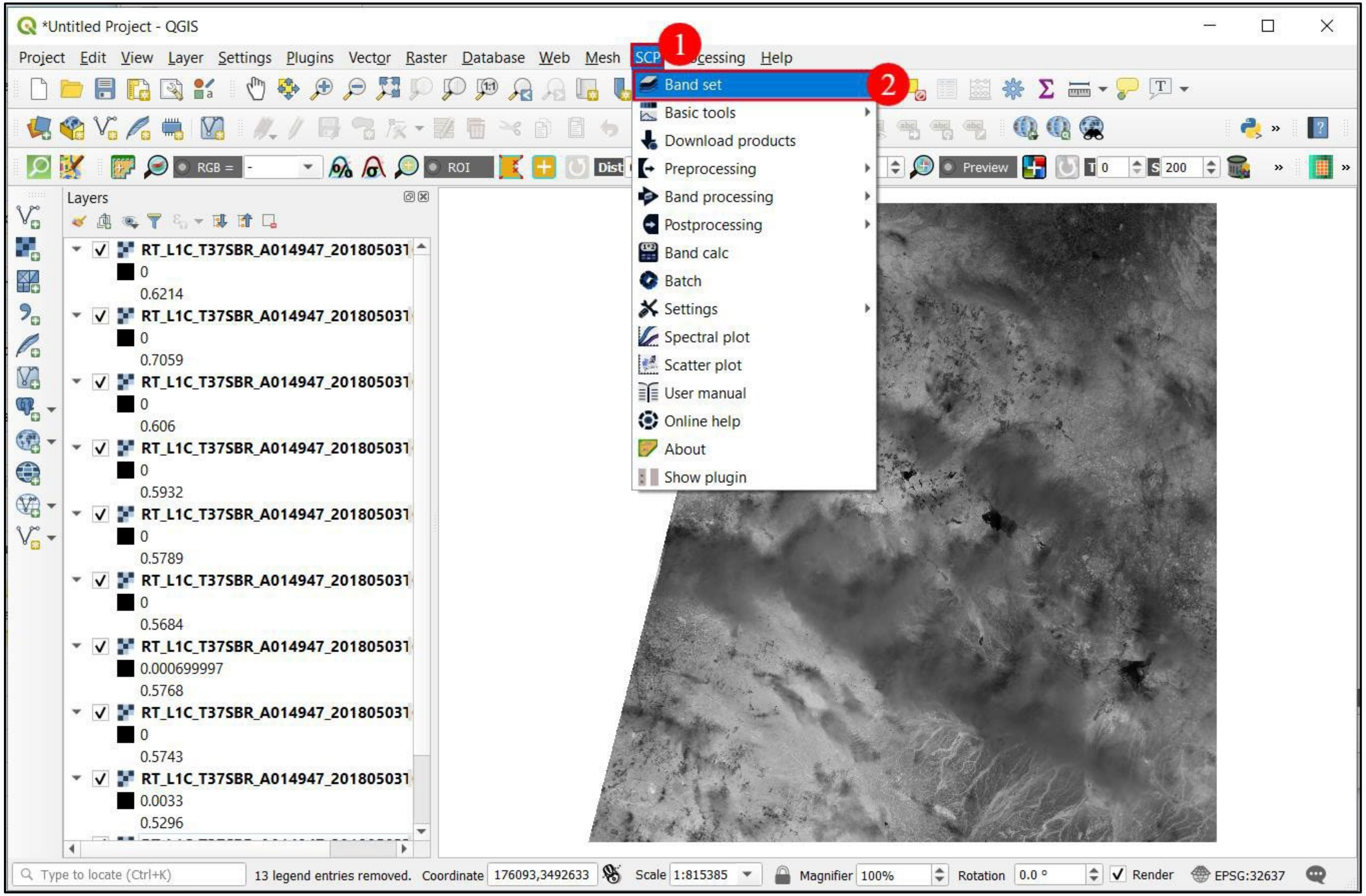
Bands appear as grayscale rasters.



SCP plugin – Band set

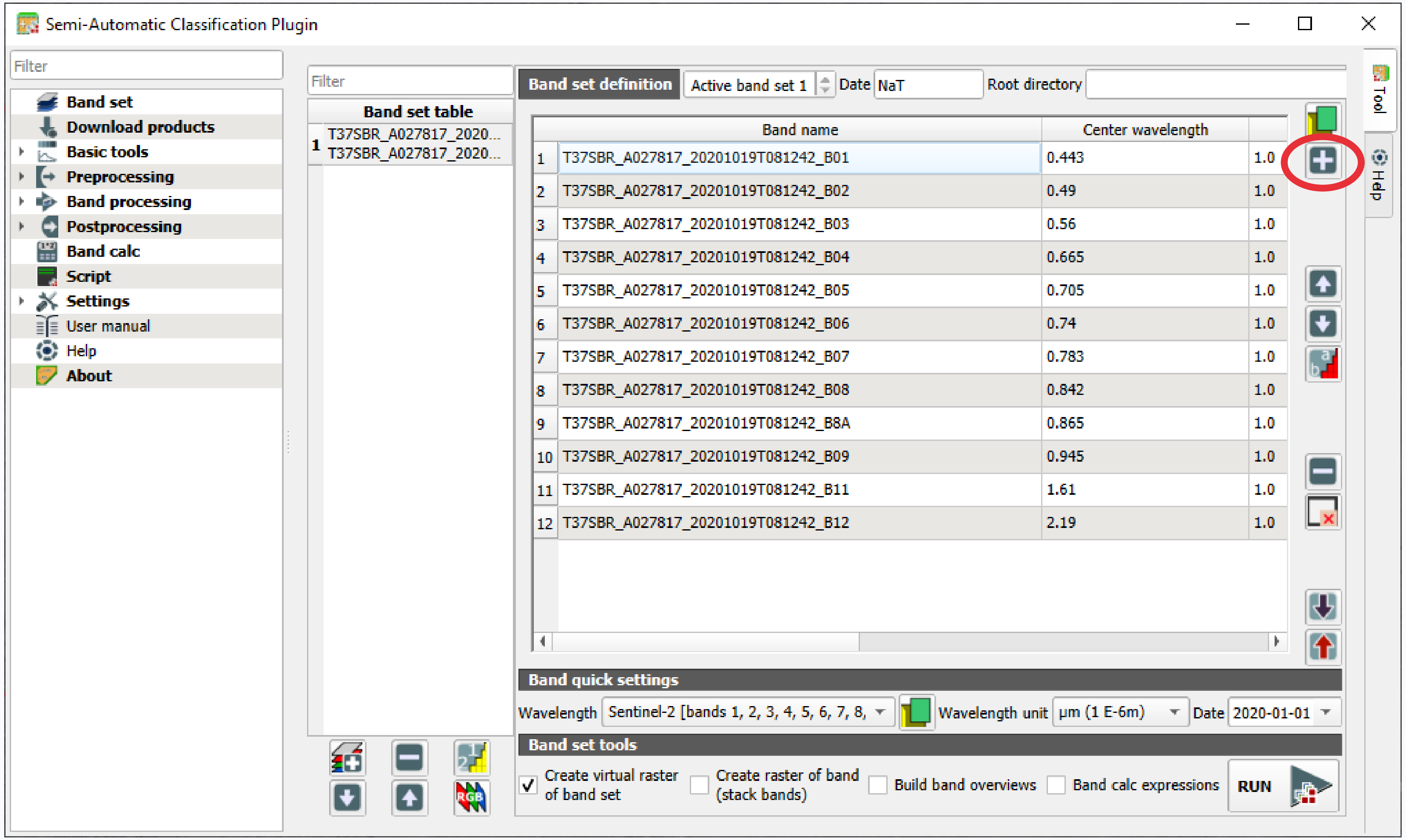
Band set is creation of composite raster (multispectral imagery). This means that all the bands that are currently separate files will be merged into one raster file.

- Go to *SCP > Band Set*



Choose which bands are going to be in the final multispectral image. In this example we will use all the bands.

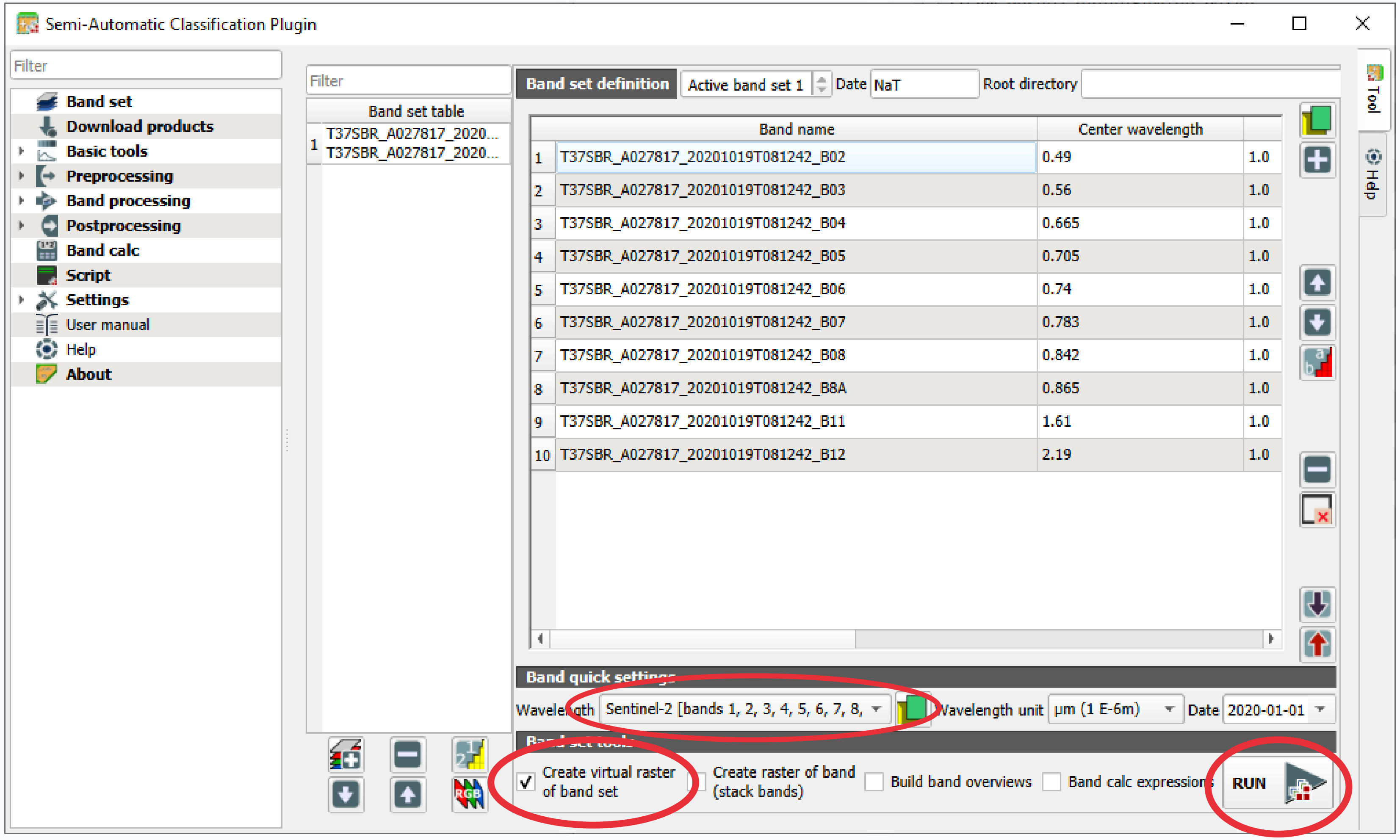
- Click the ‘Add band’ button to load layers from QGIS into the list
- Select all the Sentinel 2 band layers apart from B01 and B09 as these have lower resolutions than the other bands



Satellite images preprocessing – SCP – Band set (3)

University of Southampton

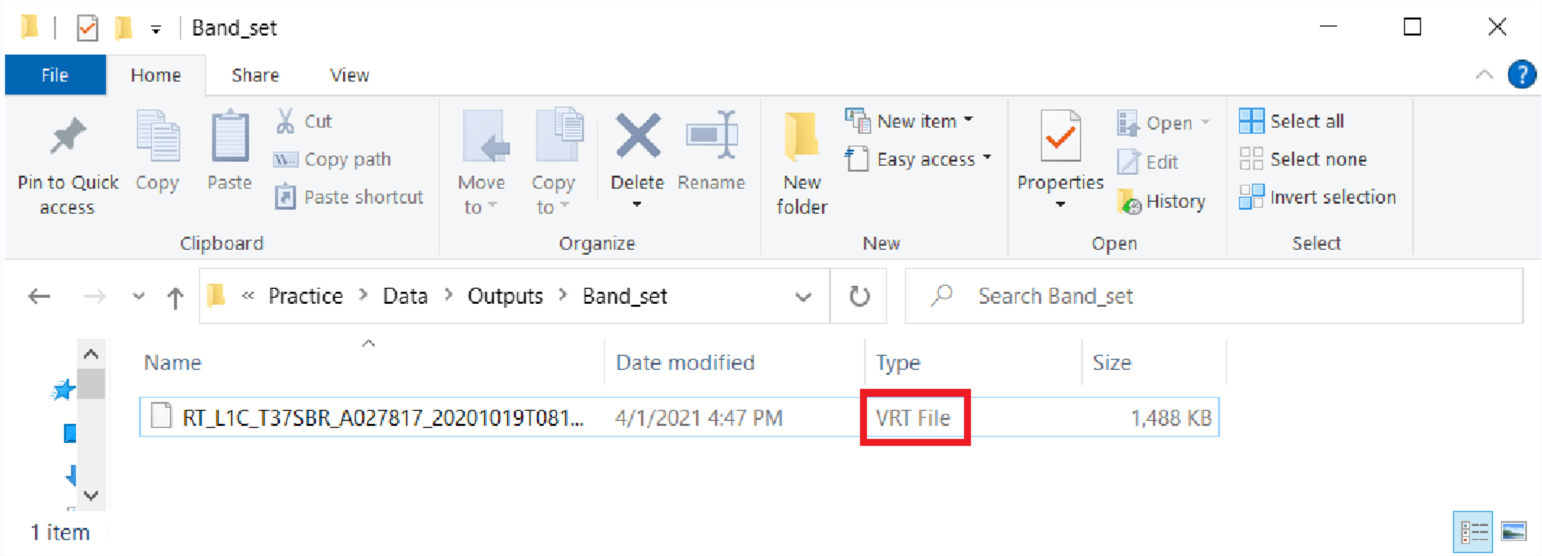
- Order bands in the way they should be ordered in the output multispectral image. In this example they should be ordered according to the center of the band wavelength.
- Select Sentinel-2 bands for the Wavelength quick settings so that the bands are ordered according to the center wavelength
- Tick ‘Create virtual raster of band set’
- Click on Run and select output folder



Virtual rasters

A clever way of saving “by-product” raster layers is the Virtual Raster format (.vrt).

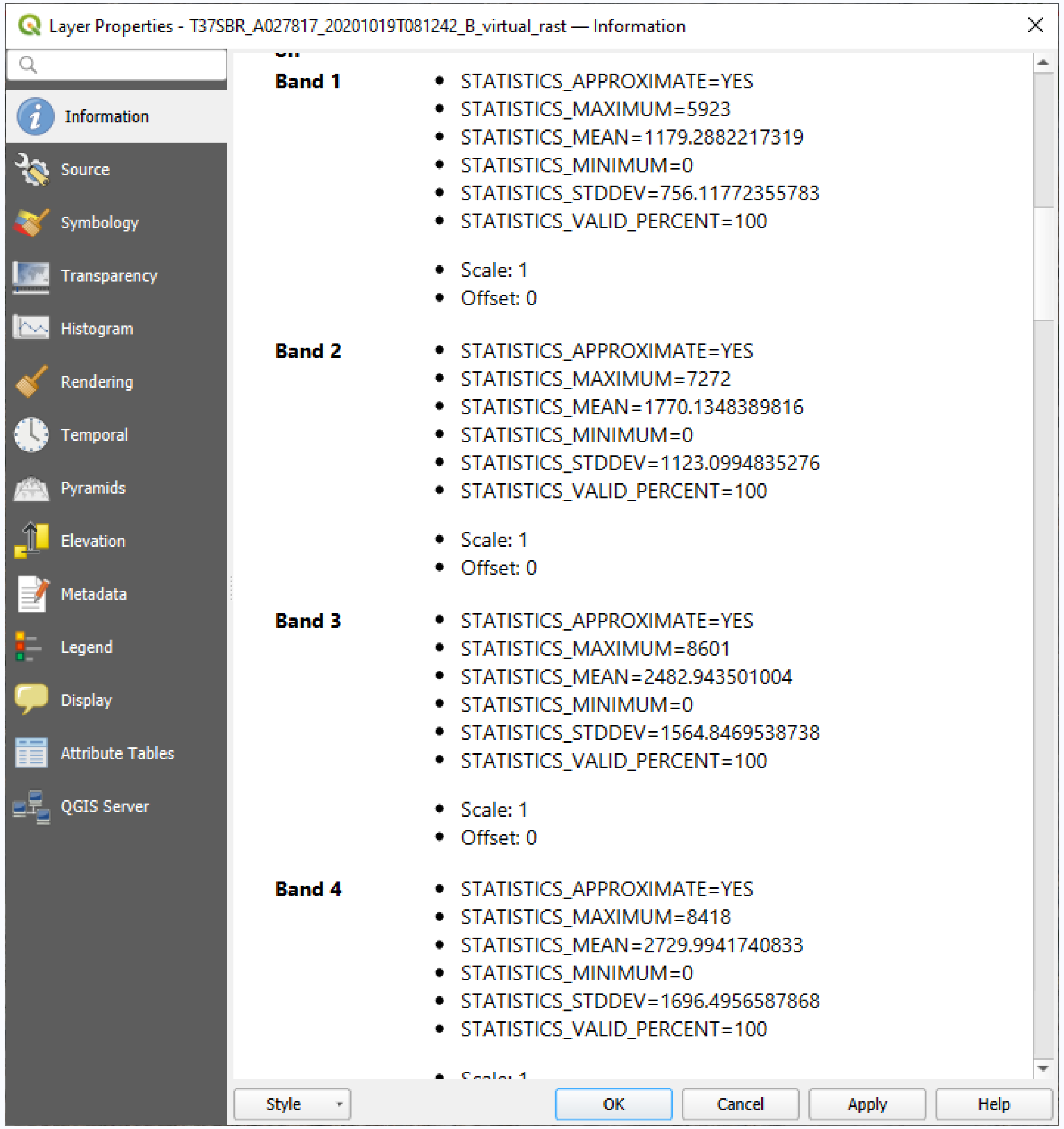
It consists of a text file pointing to multiple raster datasets which can be read and manipulated as a single layer. Virtual rasters are quick to create and take up almost no space.



We can see that the virtual raster is not black and white as it was the case with bands. That is because it is composite raster now. This aspect will be tackled a bit later.

Open the virtual raster layer properties to examine characteristics of multispectral image created with the band set.

- Right-click on the virtual raster layer and select Properties.
- Remove the individual band raster layers from QGIS by right clicking on them and choosing *Remove Layer…*



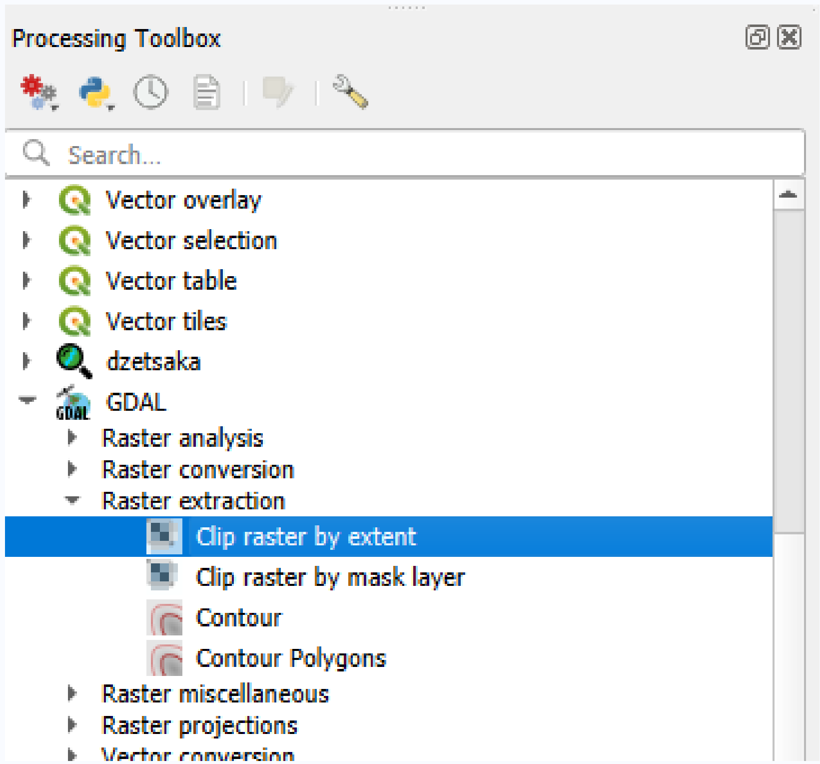
QGIS core - Clip raster/image

Satellite images preprocessing – Clip (1)

The composite raster is larger than the working area. To obtain a composite raster layer covering only the working area, thus preventing further processing on the full raster layer (ie. save time!), we can clip the full Virtual raster layer on a region of interest and save the results as a “real” layer (.tif).

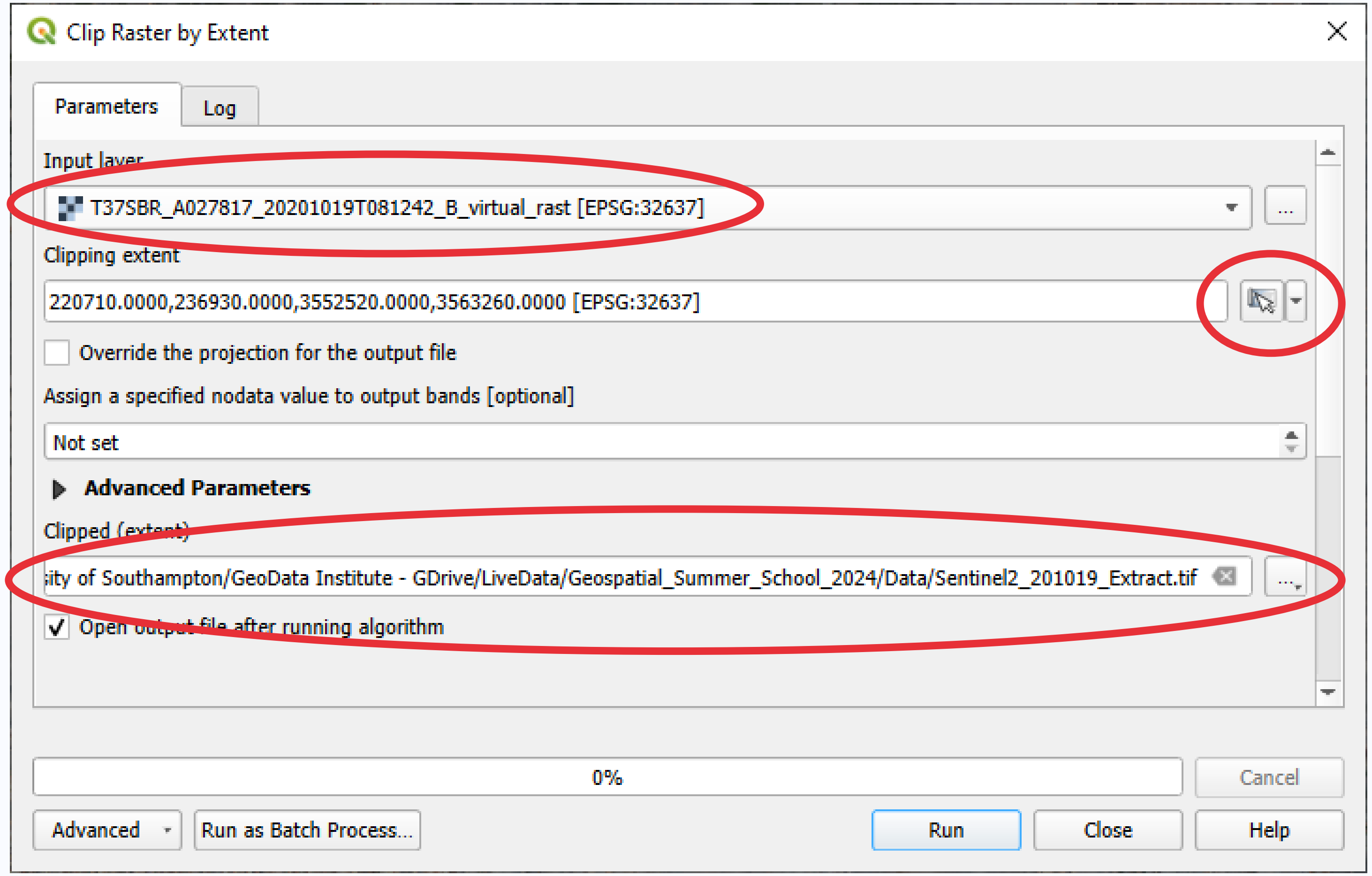
The Raster Clip tool can be accessed via two routes:

1. *Raster > Extraction > Clip Raster By Extent…*
2. In the Processing panel under *GDAL > Raster extraction > Clip raster by extent*

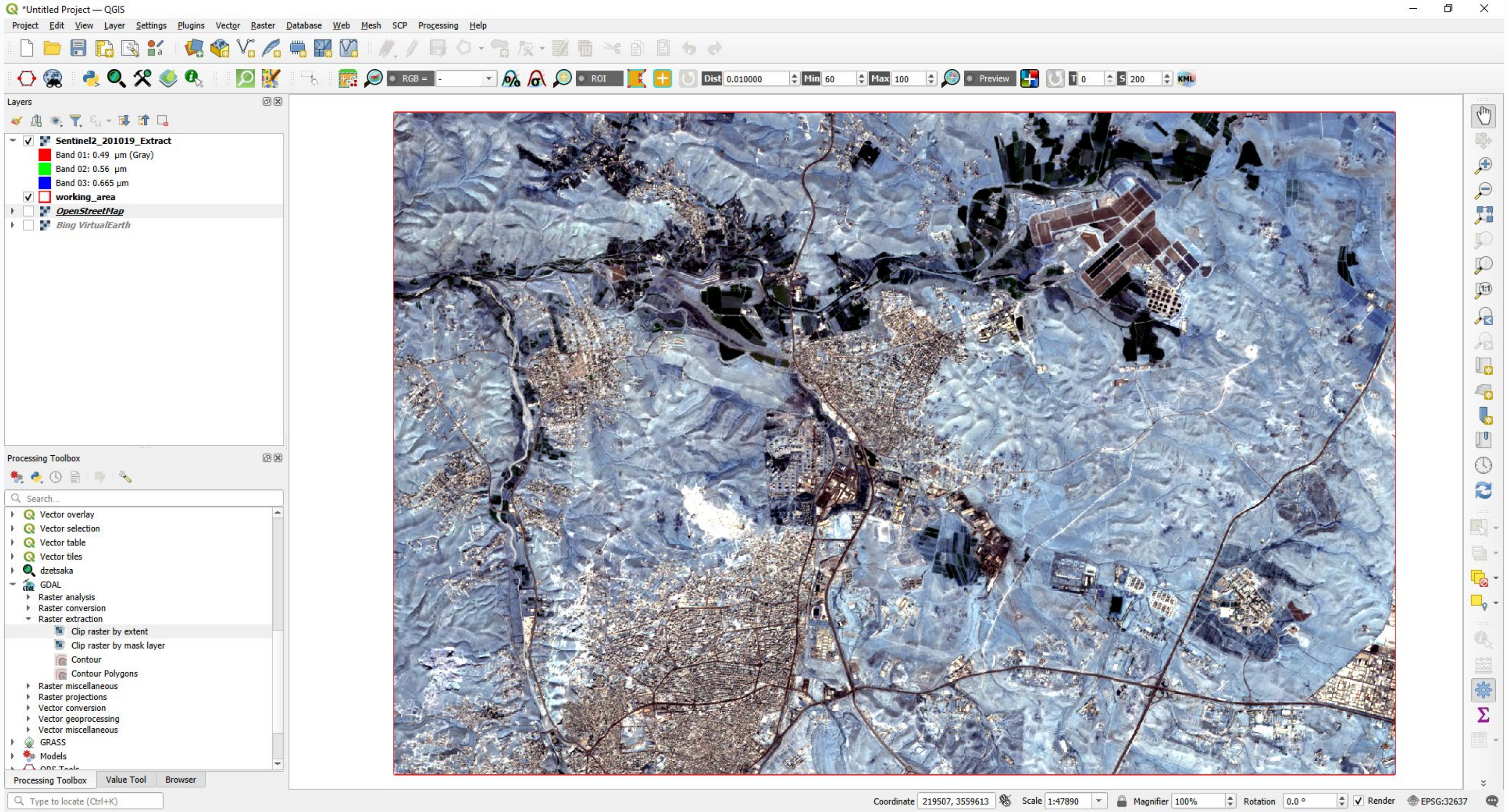


Define which raster to be clipped and what is the desired extent. In this case we will clip virtual raster of band set to the exact size of the working area polygon, and we will save output as a tif file.

1. Set the input to be the virtual raster layer.
2. Set the Clipping extent to be that of the *working_area* layer by using the dropdown at the end of the row
3. Define the output file folder and name in the Clipped (extent) field
4. Run the algorithm



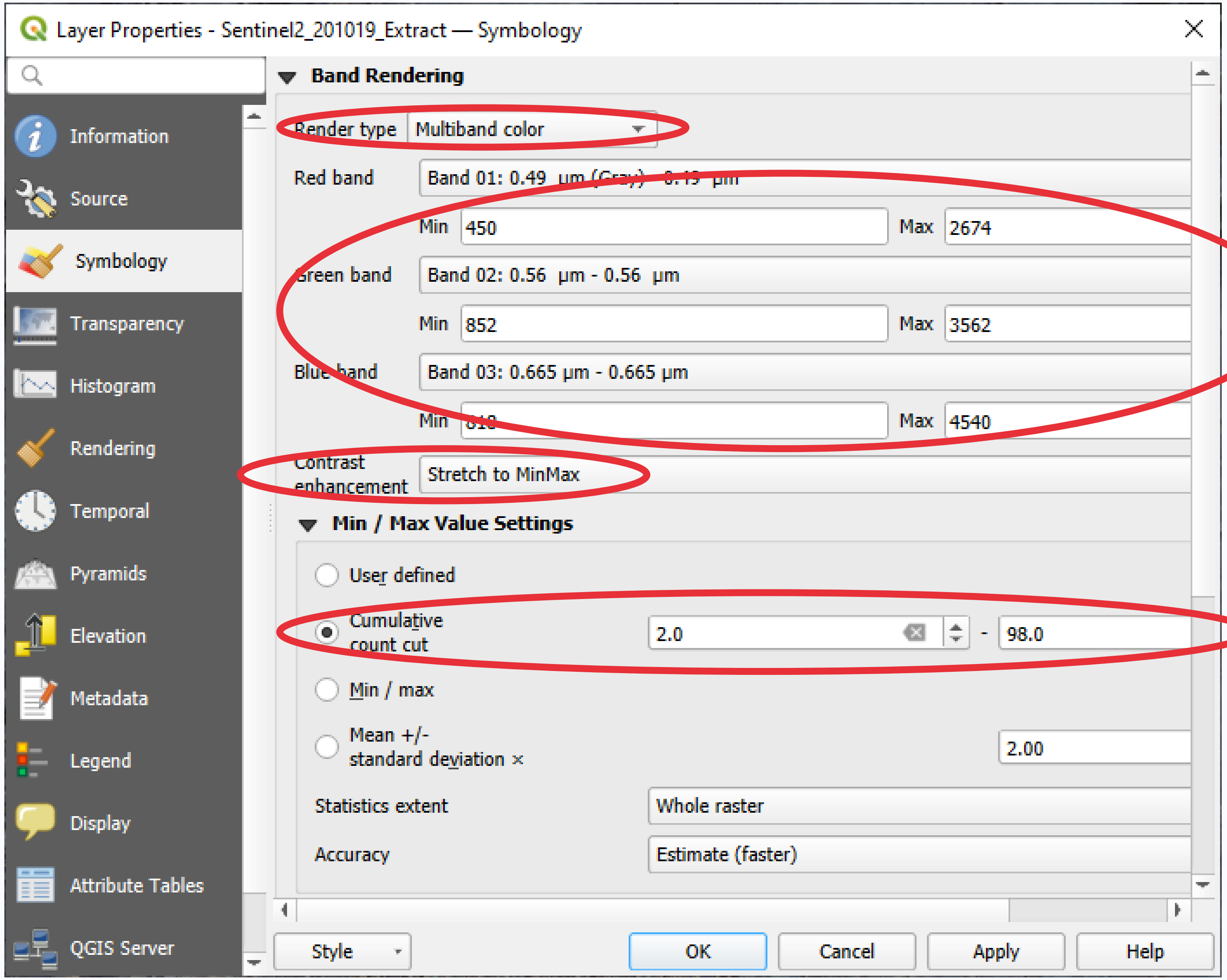
- The Clipped raster layer (.tif) is saved as a file and imported in your QGIS project.



Satellite images preprocessing - Clip (4)

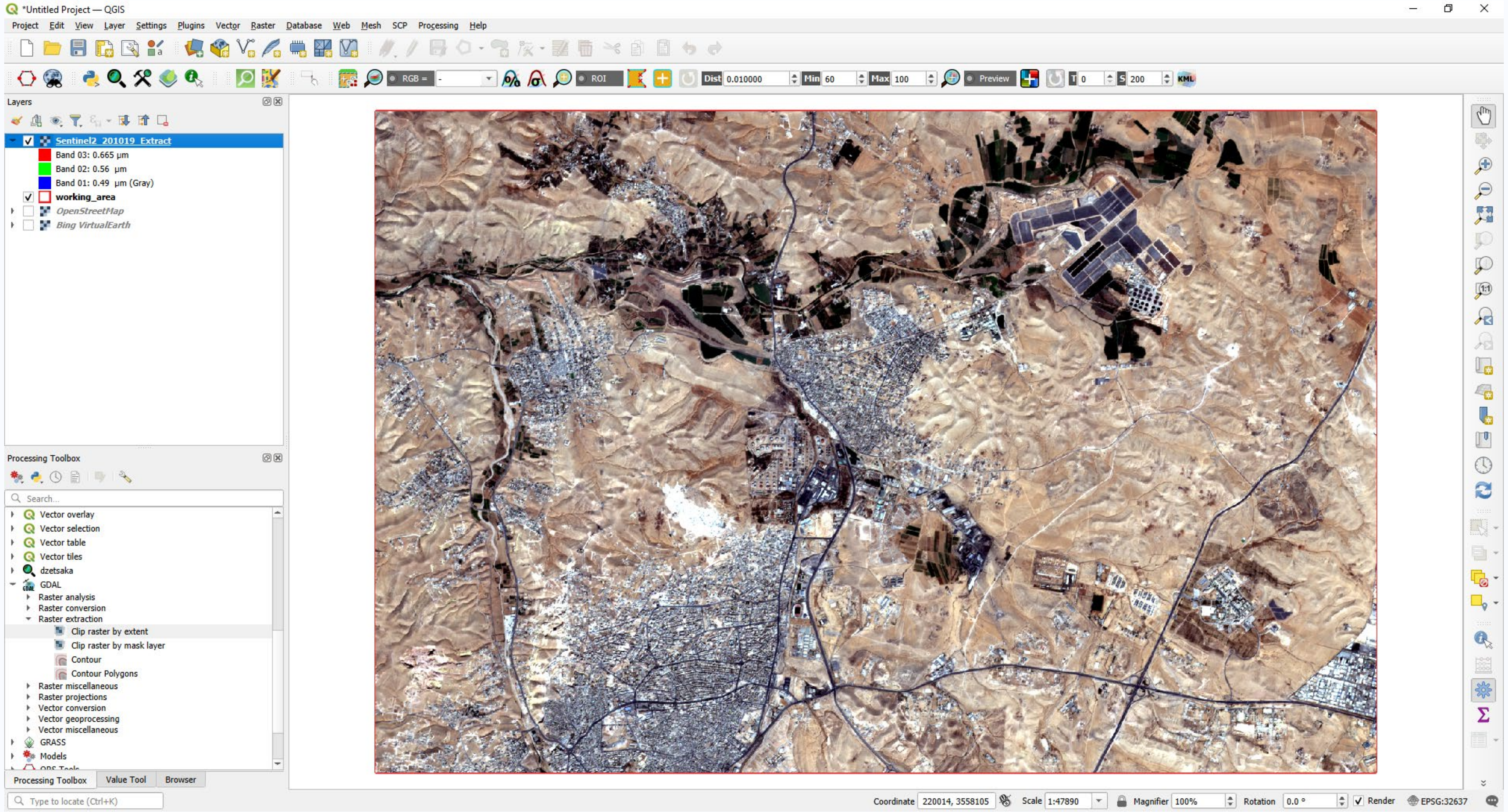
We are going to visualize raster in True colour / RGB. This means we are going to visualize it in the same colours in which we would see the scene if we looked at it with our eyes.

- Find the Symbology tab in the layer properties
- Change the Renderer type to ‘Multiband color’ if it isn’t already
- Assign the Red band 03, Green 02 and Blue 01.
- For maximum contrast, select ‘Stretch to MinMax’ and use a Cumulative count cut of 2 - 98% for the min / max value setting.



In the original Sentinel-2 image RGB is composed of bands 4, 3, 2.

As we removed band 1 when building the band set, QGIS renumbered the remaining bands,so now a combination of bands 3, 2, 1 will create the RGB image.

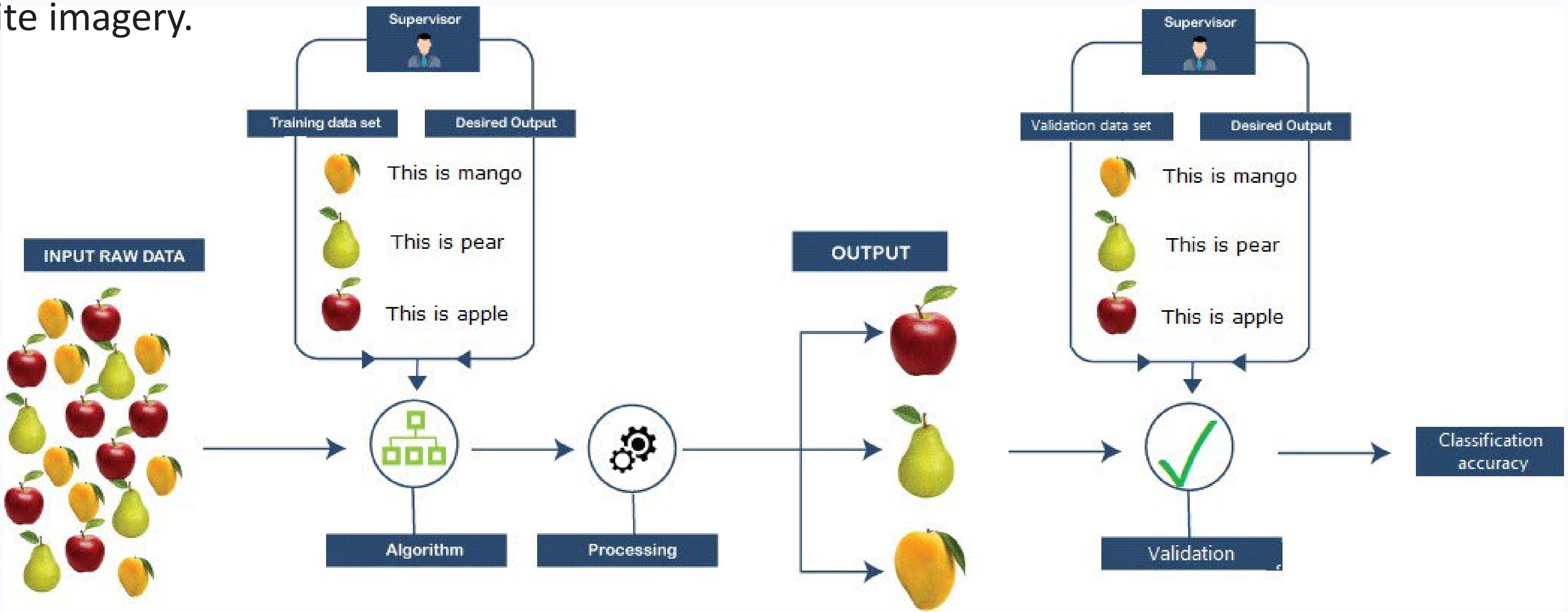


Introduction to supervised classification

Supervised classification takes advantage of the training set and classification algorithms to predict the class.

Training set is a set of areas or points in the region of interest for which the class is known (field survey, photointerpretation, etc.)

Classification algorithm uses the training set as input to "learn" to recognize similar values in the satellite imagery.



Simplified schema of classification: Classification of fruits

https://www.tutorialandexample.com/wp-content/uploads/2020/11/Supervised-Machine-Learning-1.png

Training set creation

Training set

The training set is a set of land cover class samples for each class expected in the classification output. Unique guidelines for training set extraction do not exist, and approach for doing so varies depending on:

- Classification algorithm
- Number of classes
- Desired accuracy
- Budget

Often sampling theory is the basis for estimating the suitable sample size that would result in appropriate characterization of spectral signatures.

In some other cases recommendation is to have a minimum of 10–30p samples per-class for training, where p is the number of bands used.

For every approach, the training samples must be correct, therefore we must have confident source of reference information:

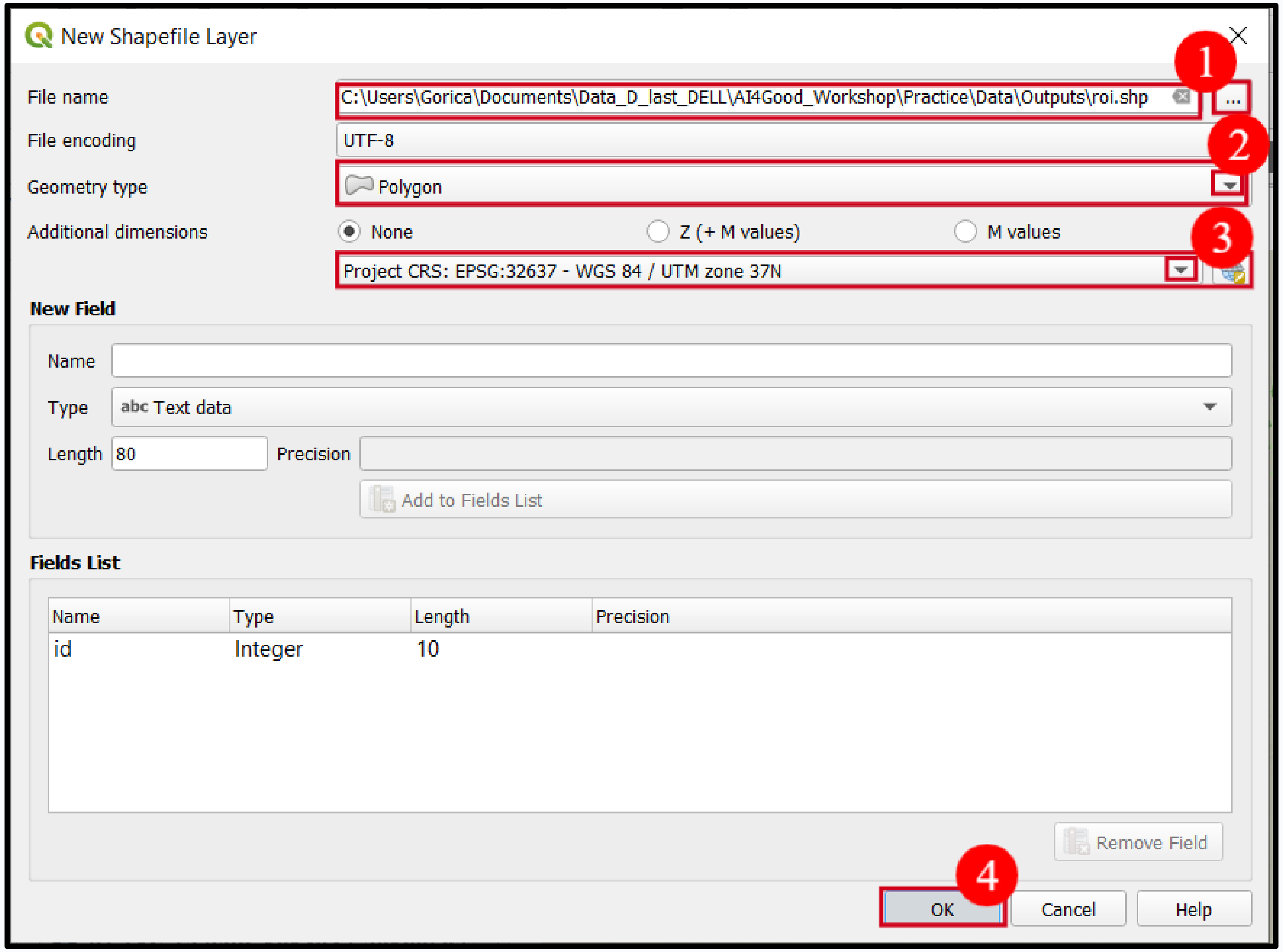
- in situ data collection
- photo-interpretation of very high-resolution satellite imagery

Supervised classification – Create training set (1)

To create a training set you need to create a new polygon shapefile which will contain an attribute field describing the membership of each polygon to a land cover class

To create a new shapefile go to *Layer > Create Layer > New Shapefile Layer*

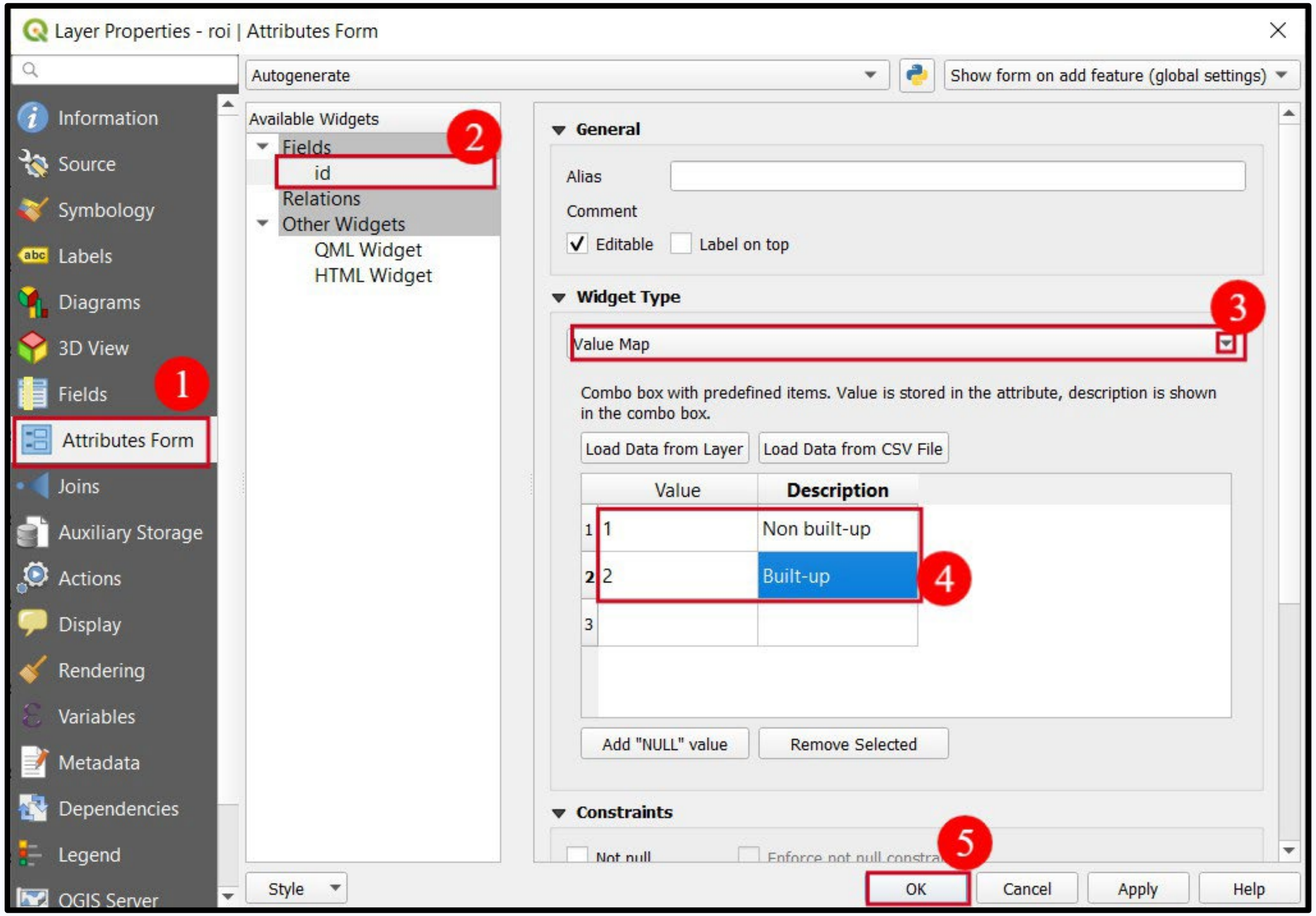
1. Define the File name (roi.shp) and output folder
2. Select the Geometry type as Polygon
3. Select CRS to be EPSG:32637 – WGS 84 / UTM zone 37N (the same as the CRS of the Sentinel-2 image)
4. Click on OK to create the new shapefile layer



The classes we are going to capture can be preconfigured for easier data entry through the layer properties:

1. Go to the Attributes Form tab
2. Select id
3. Select Value Map for Widget Type
4. Insert two Value and Description rows:

- 1: non-built-up
- 2: built-up



NOTE: the dzetsaka plugin does not accept value 0 as a class, therefore the (integer) values in the training set must be larger than 0.

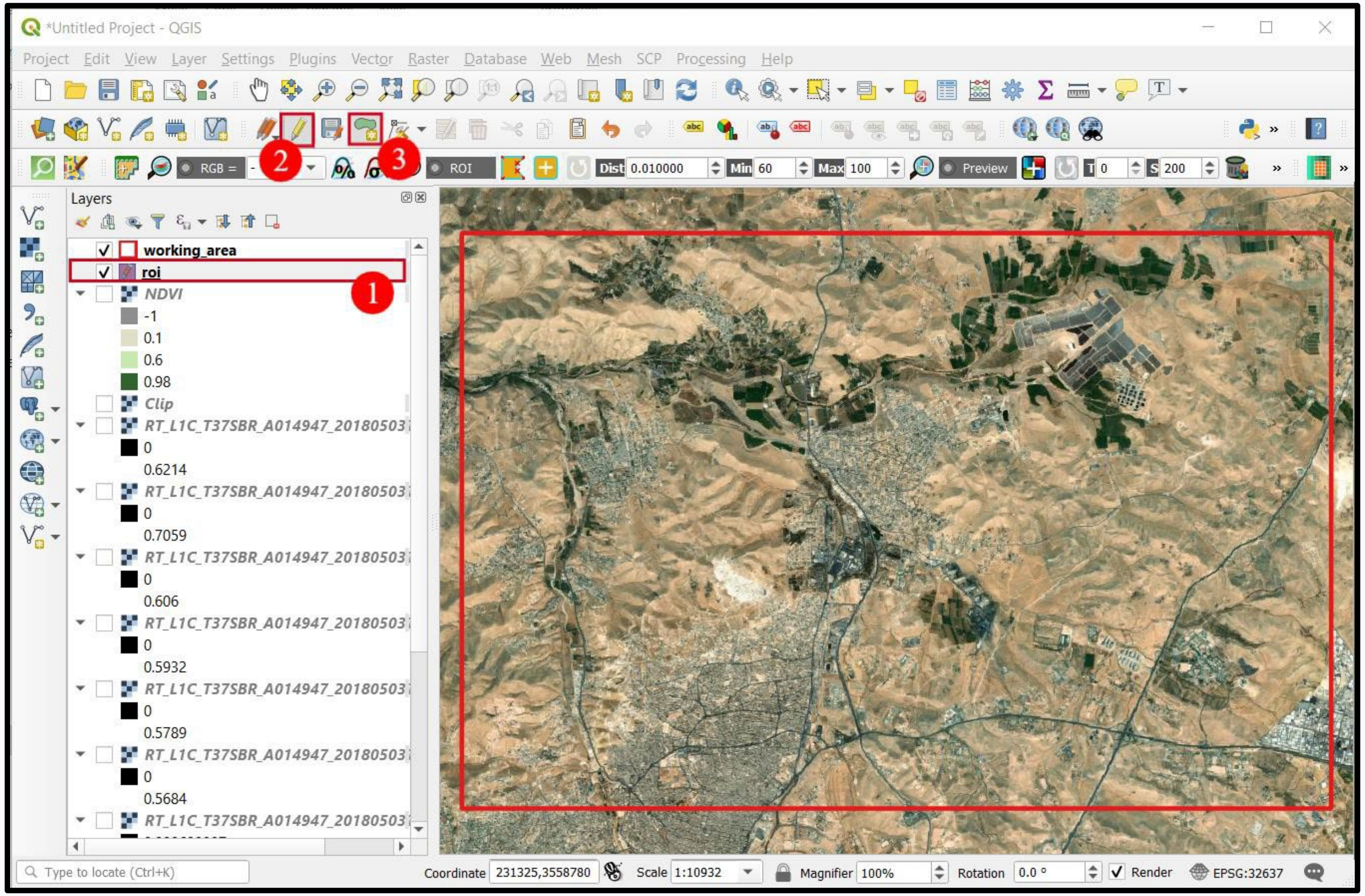
Supervised classification – Create training set – Start editing training vector

Now we are going to create our training set, digitising polygons from the Bing Virtual Earth aerial imagery.

- Load the Bing Virtual Earth layer from the data folder

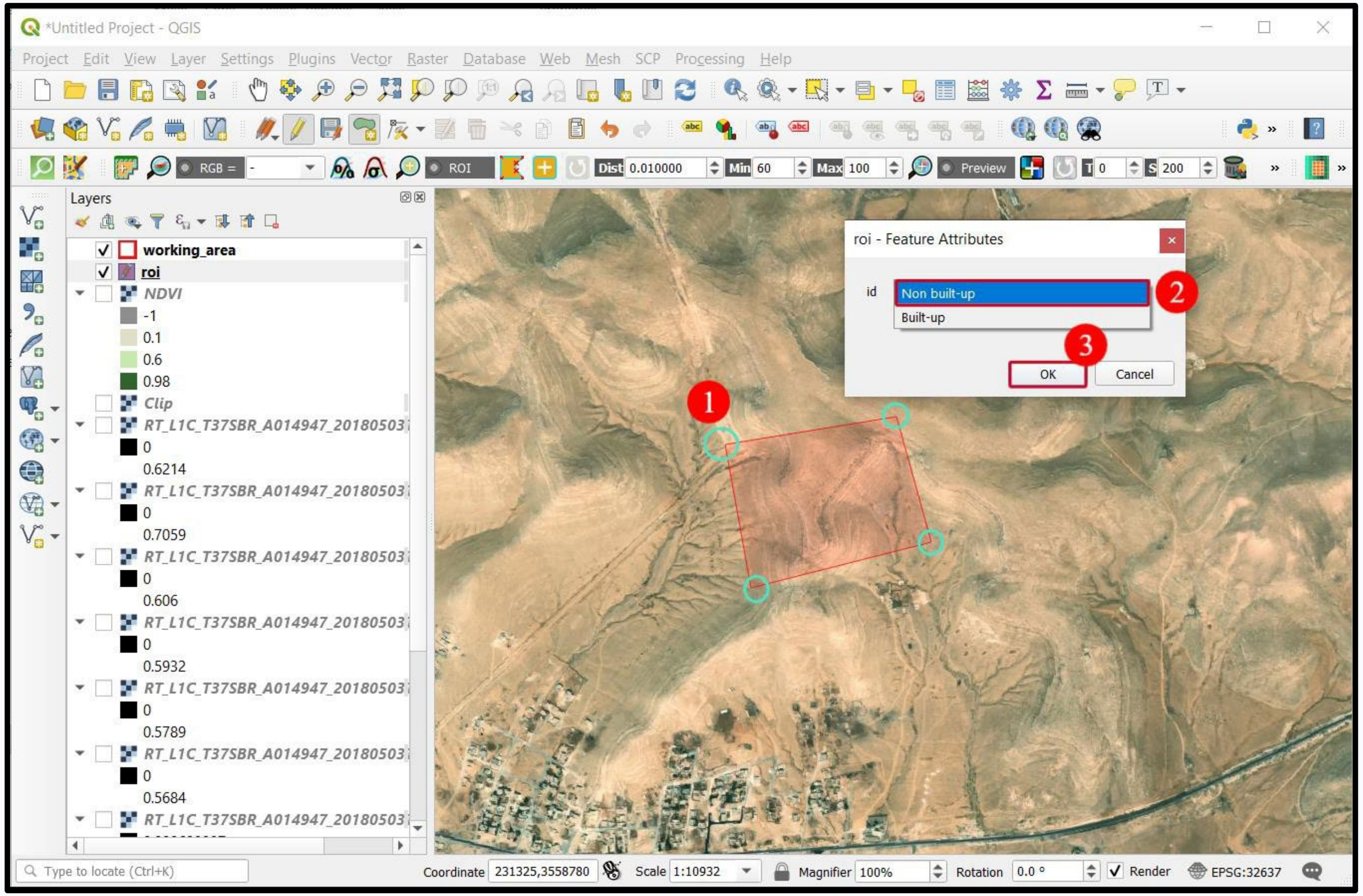
To start digitizing in QGIS, it is necessary to enter editing mode:

1. Select the training vector layer (e.g., roi.shp) in the Layers panel
2. Enter to the editing mode by clicking on the Toggle Editing tool
3. Select Add Polygon Feature tool.



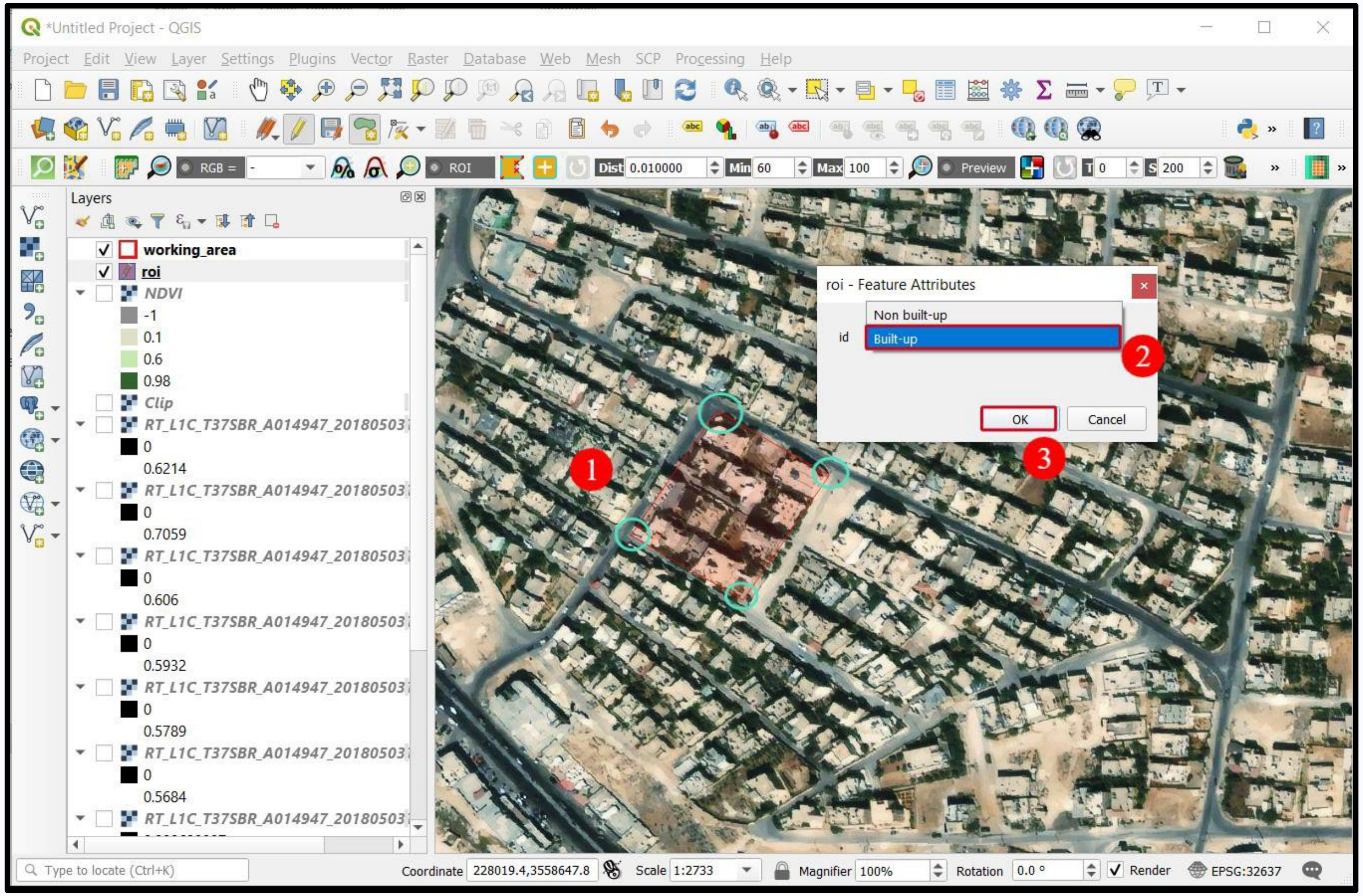
To create a polygon for the non built-up class, find an area without buildings, then:

1. In the Map left-click to create the vertices of the polygons. Right-click on the initial vertex to finish the polygon drawing.
2. Select ‘Non built-up’ from predefined values in the Attribute dialog that appears when you finish the polygon.
3. Click on OK to assign the value to the polygon feature



To create a polygon for the built-up class, find an area with buildings, then:

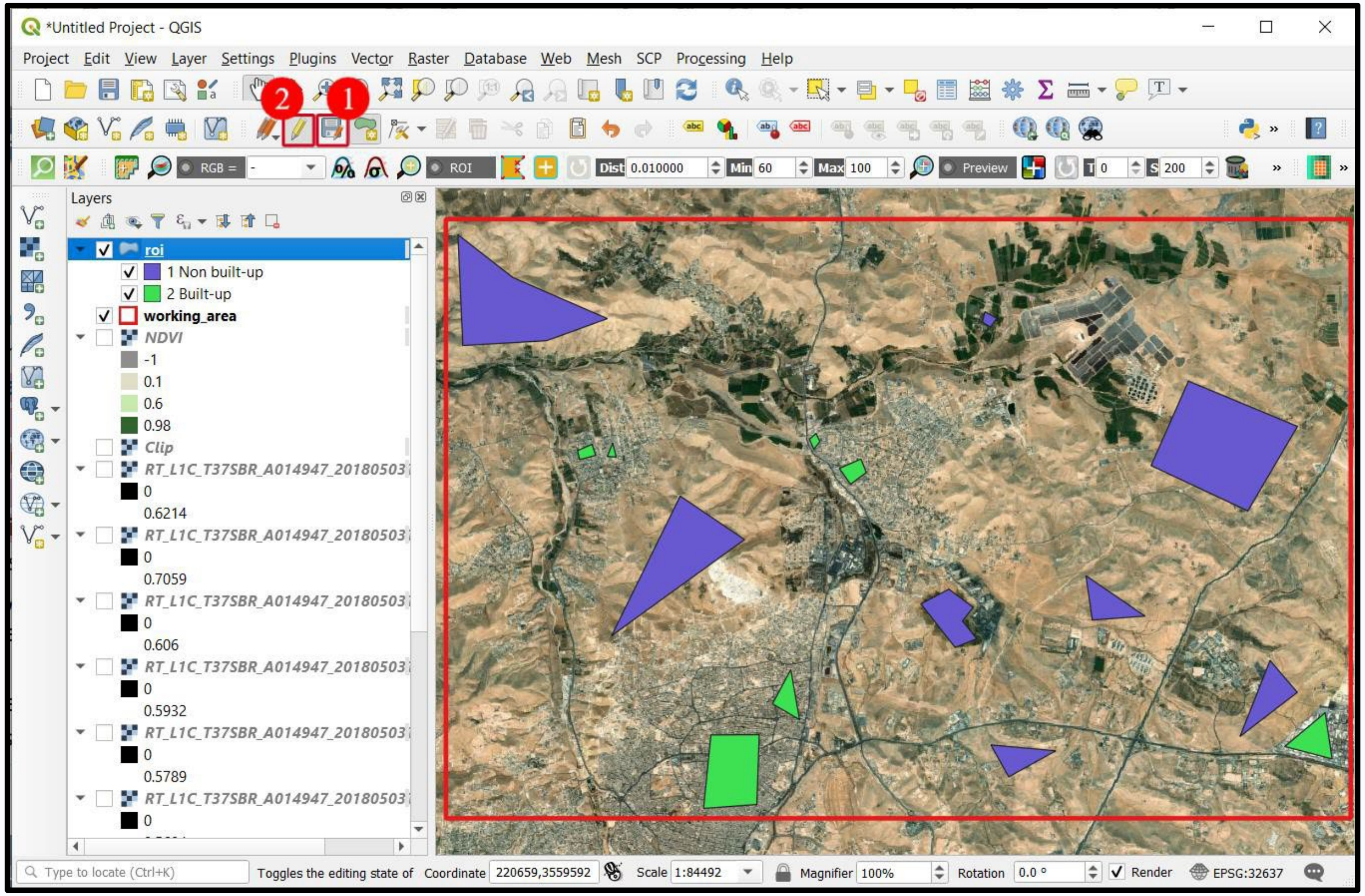
1. In the Map left-click to create the vertices of the polygons. Right-click on the initial vertex to finish the polygon drawing.
2. Select ‘built-up’ from predefined values in the Attribute dialog that appears when you finish the polygon.
3. Click on OK to assign the value to the polygon feature



Continue adding samples, trying to represent the breadth of membership in each class.

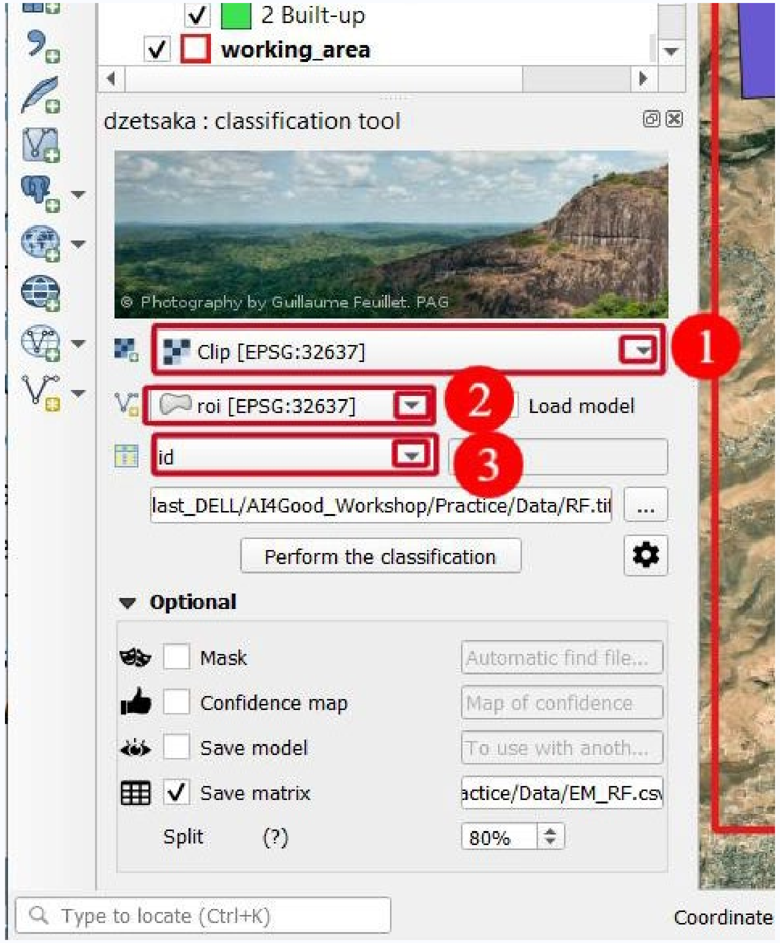
When enough samples are added, finish editing:

1. Click on the Save Layer Edits button to save all the polygon features added
2. Stop editing by clicking on the Toggle Editing button



Supervised classification

University of Southampton



First step in the classification is to specify input data.

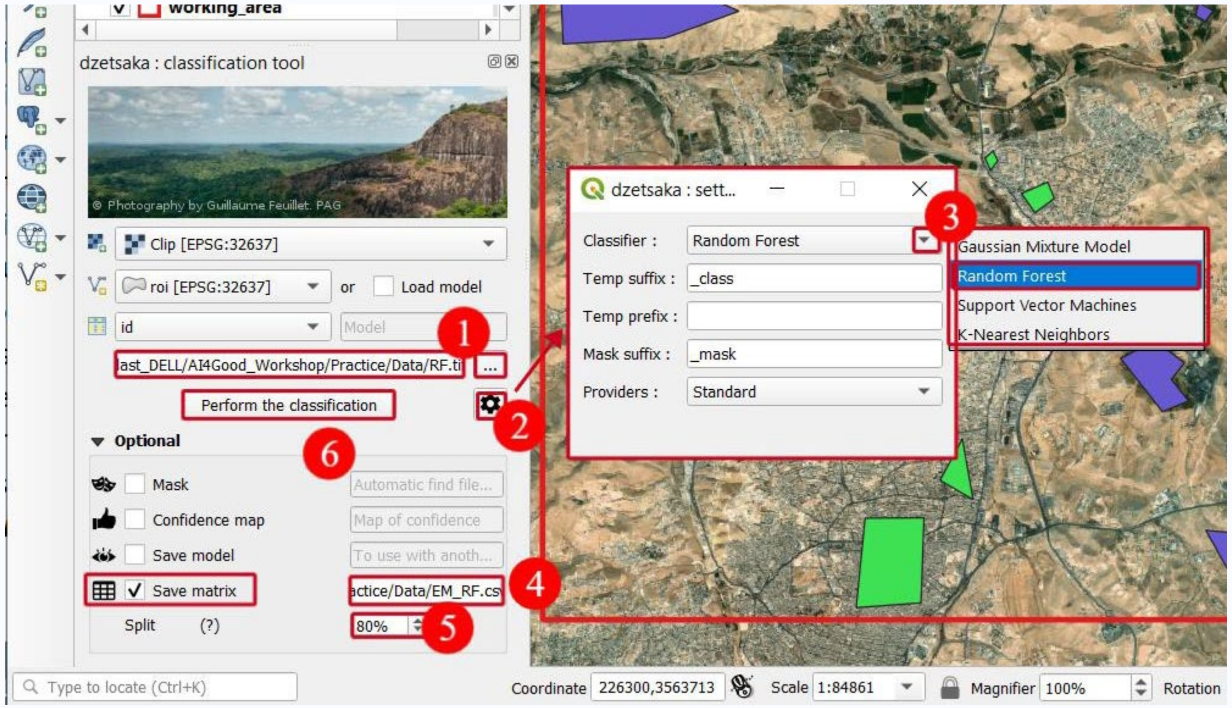
To activate the dzetsaka classification plugin choose *Plugins > dzetsaka > classification dock*

1. Specify the name of the image layer to be classified (e.g., Sentinel2_201019_Extract)
2. Specify the name of the training set vector layer (e.g., roi)
3. Specify the attribute of the training set layer that contains information regarding the classes (must be numerical, e.g., id)

University of Southampton

We must also specify outputs and classification settings:

1. Specify the name of the output
2. Open dzetsaka settings
3. Choose the Classifier (classification algorithm) (e.g., Random Forest)
4. In the Optional parameters flag Save matrix and specify the path where to save error matrix
5. Specify Split to 80 % so that 80% of the samples of the roi.shp file are used for training the algorithm and the rest for cross-validation
6. Perform the classification!



Repeat the procedure the for Gaussian Mixture Model and K- Nearest Neighbors by changing the Classifier (3), the output name (1) and the error matrix name (4).

Algorithm References

- Dzetsaka GitHub repository: https://github.com/nkarasiak/dzetsaka
- Scikit-learn - Random Forest:
- https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html#sklearn.ensemble. RandomForestClassifier
- Scikit-learn - K-Nearest Neighbors:
- https://scikit-learn.org/stable/modules/neighbors.html#classification
- Gaussian Mixture Model:
- https://doi.org/10.1109/JSTARS.2015.2441771

Validation

Using portion of training data for validation is just one approach for doing validation (internal validation – cross-validation). Often validation is conducted independently of training data set. Validation must ensure:

- Appropriate sample size - Enough samples with minimum cost:
  - Statistical – Calculate the number of samples based on binomial or normal approximation to the binomial distribution (Cochrane, 1977)*
  - Empirical – number of samples is driven by the available budget (not suggested)
- Sample Allocation:
  - Random – samples are distributed randomly in the area of interest
  - Stratified random sampling – split area of interest into strata and then select samples in each strata
    - Equal sample size per strata
    - Number of samples per strata adjusted according to the strata size
- High quality source of reference information:
  - In situ data collection
  - Photo-interpretation of higher resolution imagery

*https://hwbdocuments.env.nm.gov/Los%20Alamos%20National%20Labs/General/14447.pdf

In QGIS there is the AcATaMa plugin designed for the accuracy estimation of land cover maps. It supports :

- different steps in creation of training samples (e.g., sample size definition and sample allocation)
- a user-friendly interface for photo interpretation of samples and
- the calculation of multiple accuracy indexes

AcATaMa

AcATaMa is a Qgis plugin for Accuracy Assessment of Thematic Maps

Error matrix and accuracy indexes

Error matrix produced by dzetsaka

Classification with Random Forest

# Columns=prediction	Lines=reference (ground truth)
19309	15
26	3821

Error matrix interpretation

		Prediction/classification			
		1 Non built-up	2 Built-up	Sum reference	PA
Reference	1 Non built-up	19309	15	19324	99.9%
	2 Built-up	26	3821	3847	99.3%
	Sum prediction	19335	3836	23171	
	UA	99.9%	99.6%		99.8%

- **Producer's accuracy (PA)** of a class is probability that the class present on the ground is also captured by the classification in the thematic raster.
- **User's accuracy (UA)** of a class shows how often a user of classified map can expect to find the class on the ground
- **Overall accuracy (OA)** is the proportion of correctly classified pixels out of the total number of pixels.

$$PA_{Non\ built-up} = \frac{19309}{19324} * 100 = 99.9\%$$

$$PA_{Built-up} = \frac{3821}{3847} * 100 = 99.3\%$$

$$UA_{Non\ built-up} = \frac{19309}{19335} * 100 = 99.9\%$$

$$UA_{Built-up} = \frac{3821}{3836} * 100 = 99.6\%$$

$$OA = \frac{3821+19309}{23171} * 100 = 99.8\%$$

Inter-comparison – QGIS Modeller

Assessing the success of the classification – inter-comparison (1)

The classification success can be estimated by comparison with other maps with the same theme by computing the error matrix and the accuracy indexes like Overall accuracy, PA and UA. Instead of PA and UA, the Commission error (1-UA) and Omission error (1-PA) can also be used. Moreover, another index, the Kappa index, is often computed.

Sentinel-1

GHS-BUILT-S1

For the scope of this exercise, we will use [GHS-BUILT (Sentinel-1)] that contains an information layer on built-up presence as derived from Sentinel-1 image collections (2016). This dataset is produced by the Joint Research Center of the European Commission. The CRS of this dataset is WGS 84 / Pseudo-Mercator (EPSG:3857)

Assessing the success of the classification – inter-comparison (2)

In order to compare the data, two datasets must have:

- ✓ the same CRS
- ✓ the same resolution
- ✓ the same extent
- ✓ the same values for the same classes

Parameters	Classification output	GHS-BUILT S1
CRS	EPSG:32637	EPSG:3857
Resolution	10 m	20 m
Extent	1 tile (X: 1622 Y: 1074)	4 tiles (X: 1367 Y: 91)
Classes	0 - Non-built-up 1 - Built up	NULL - Non-built up 1 - Built-up

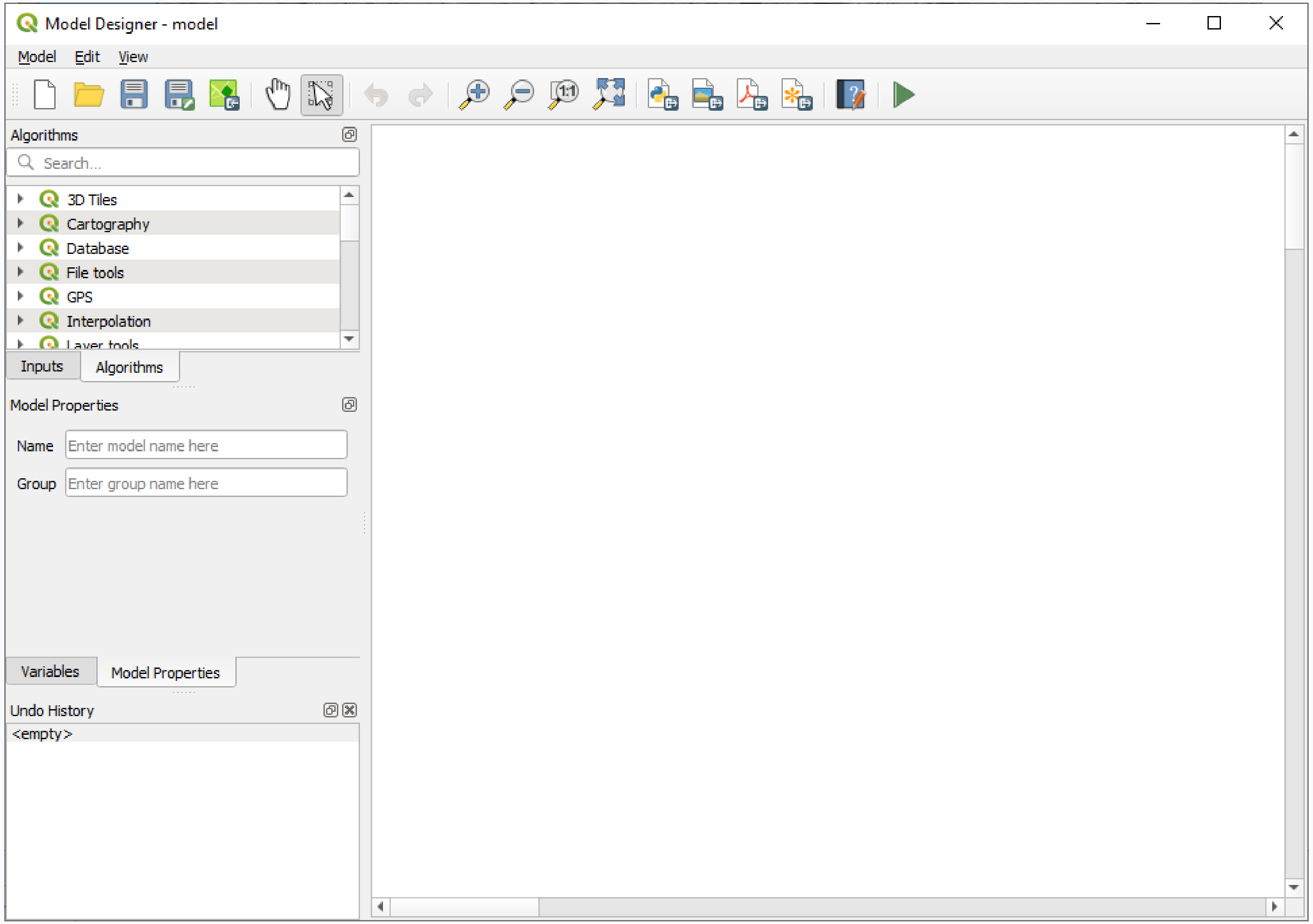
As you can see many properties of the two datasets need to be homogenized

University of Southampton

The Model Designer allows you to create complex models using a simple and easy-to-use interface. It is particularly useful for repetitive processing.

To activate the Graphical modeler

Choose: *Processing > Model Designer*



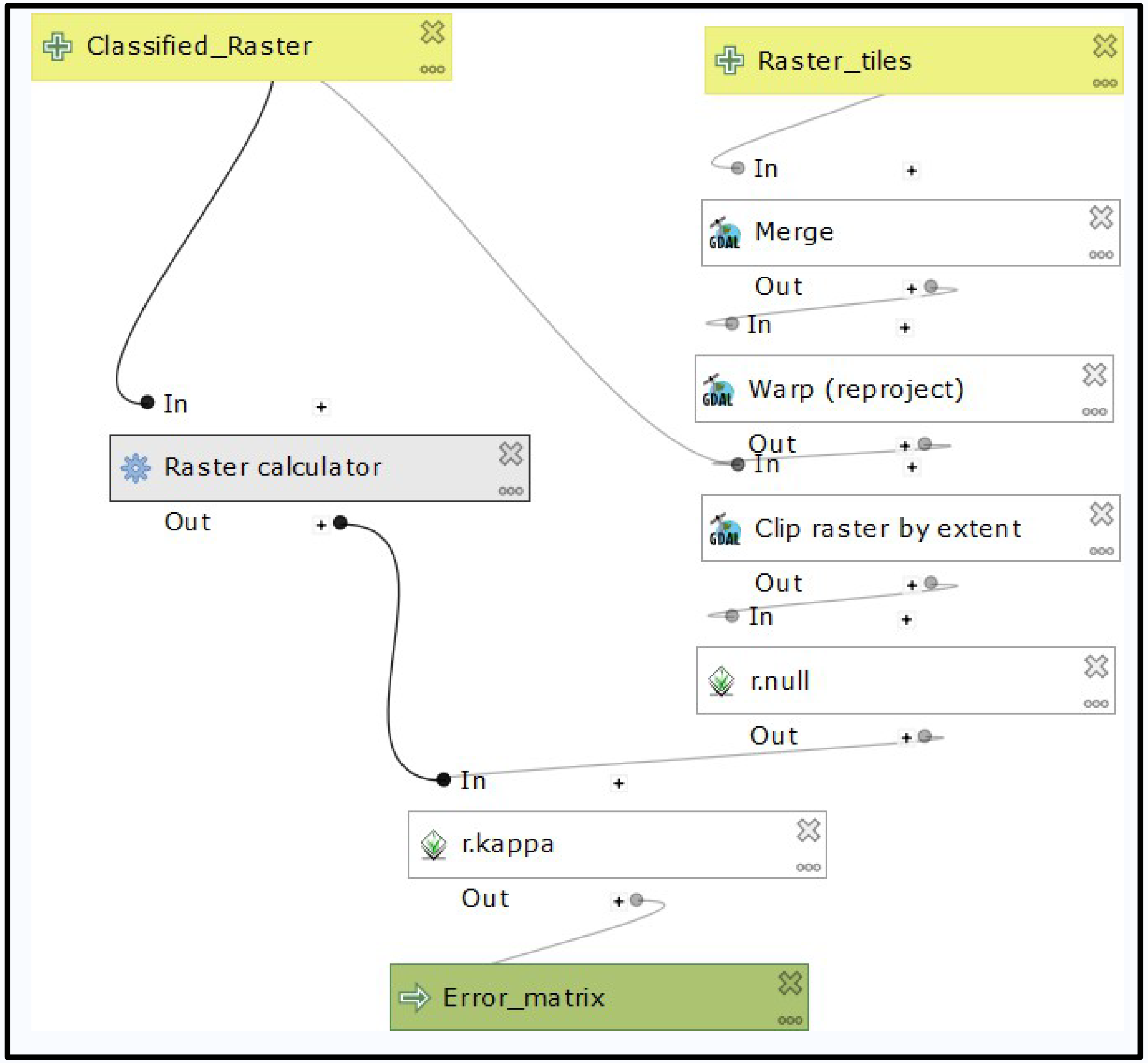
We can create a model to harmonise GHS BUILT S1 with the classified raster as shown on the figure on the right.

The classified raster requires only one preprocessing operation before the comparison.

For GHS BUILT S1 we will apply:

- Merge operation to merge 4 tiles
- Warp (reproject) to reproject it
- Clip raster by extent (the extent of the classified raster)
- r.null to adjust null values to 0 as in GHS the non built-up values are NULL by default.

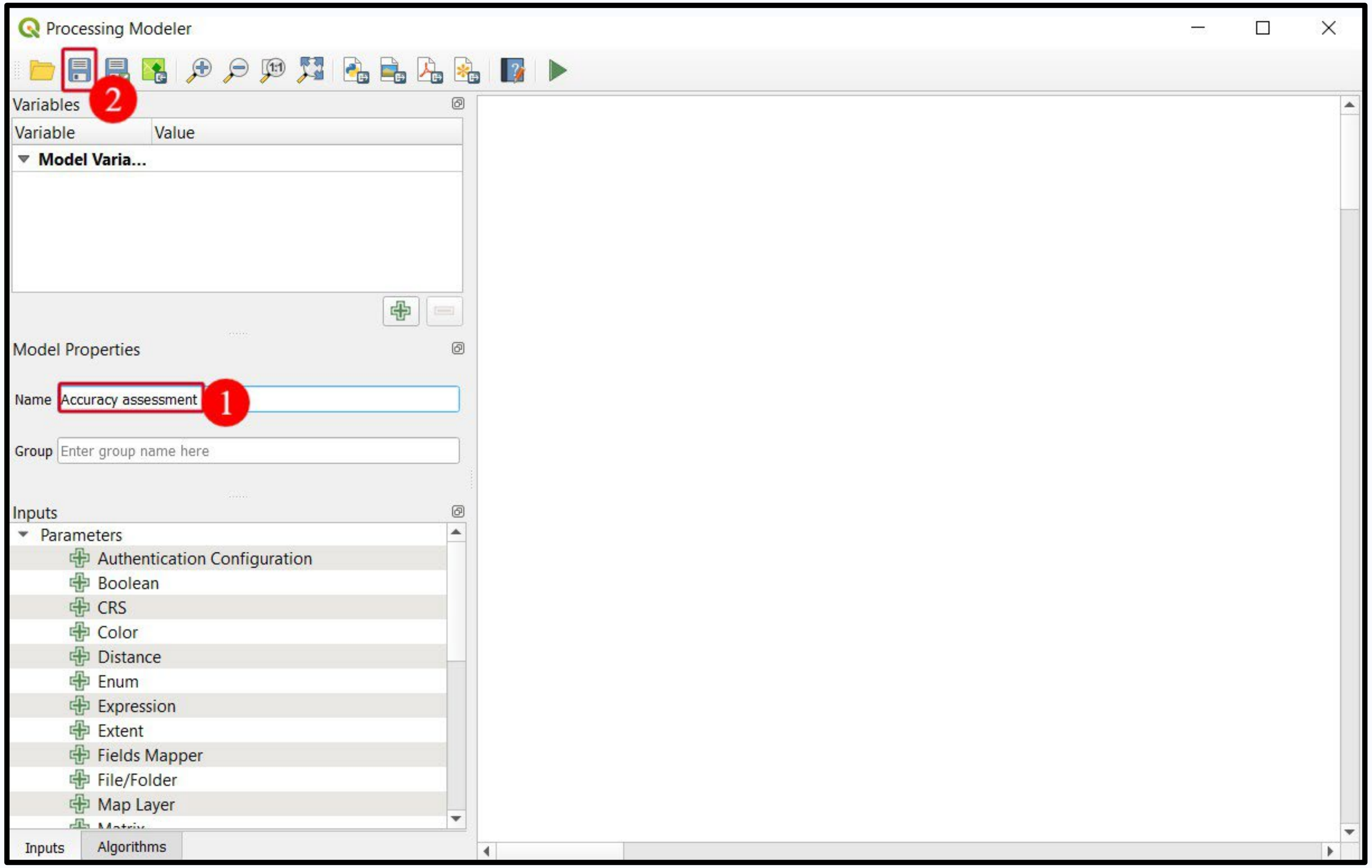
After the preprocessing r.kappa can be executed to compute the error matrix and the accuracy indexes. (r.kappa automatically adjusts the resolution of classified raster to the resolution of reference raster)



Graphical modeler – Define input data (1)

Start by saving the model

1. Specify the name of the model (e.g., Accuracy assessment)
2. Click on the save icon and save the file of the model by defining its destination path and its name (e.g., Accuracy assessment)



Models are saved in your user profile folder by default, though this can be configured.

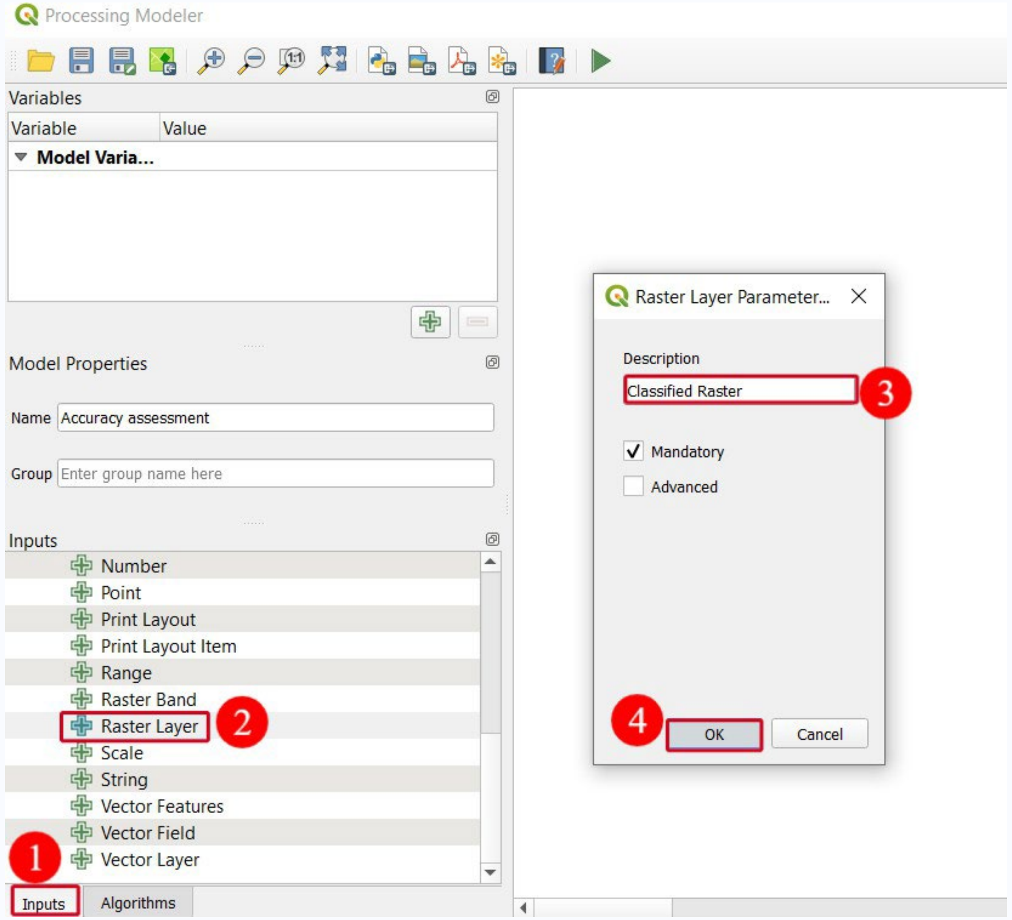
Graphical modeler – Define input data (2)

Input parameters should be the first thing to be defined in the model. When defining input parameters, we should keep in mind the expected data type that will be used in the model.

In our case the first input is the classification output raster we produced before.

1. Go to Inputs tab
2. Double click on Raster Layer
3. Insert the name you want to assign to this Raster Layer in the Description field (e.g., Classified Raster)
4. Click on OK

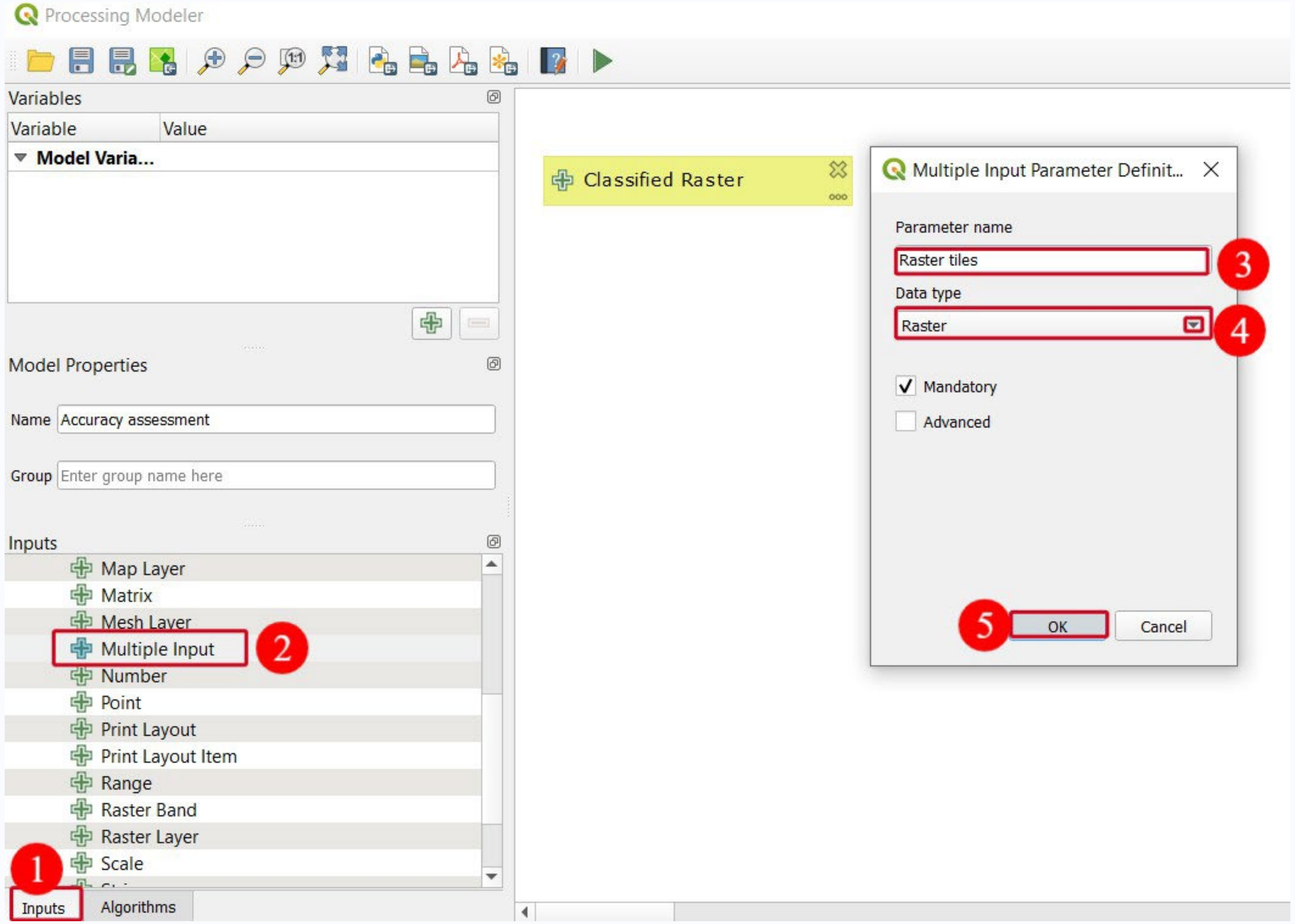
At this point only the type of input data is defined; the data to be processed will be specified as a parameter when running the model



Graphical modeler – Define input data (3)

The second input is the group of raster tiles with which we are going to compare the classification outputs we produced before.

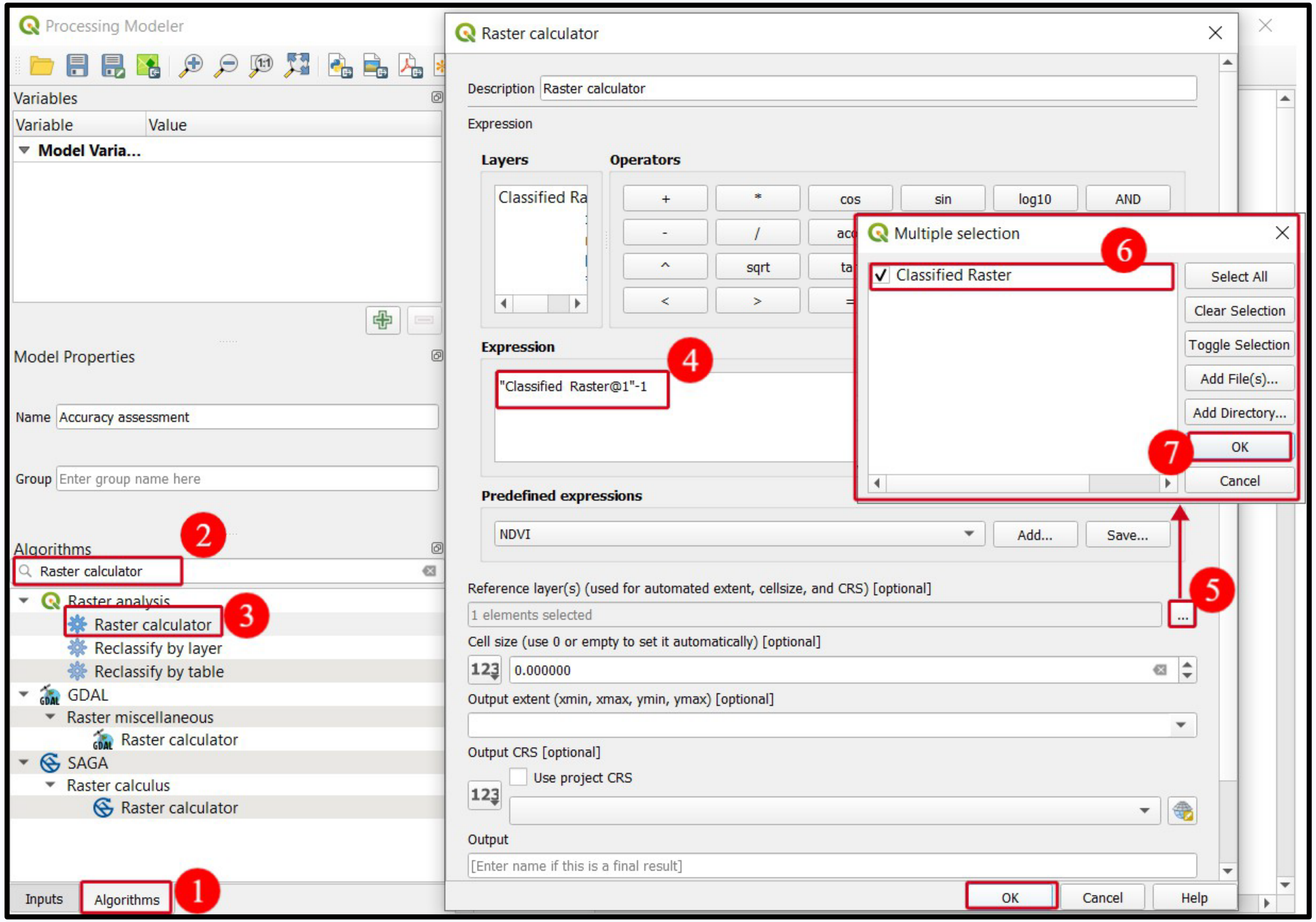
1. Go to Inputs tab
2. Double click on Multiple Input
3. Insert the inputs parameter name in the Parameter name field (e.g., Raster tiles)
4. Select Raster Data type
5. Click on OK



The first processing consists in changing the values of the classification output from 1 and 2 to 0 and 1, respectively.

It is a simple algebraic operation where we use the Raster Calculator to subtract 1 from the classified raster.

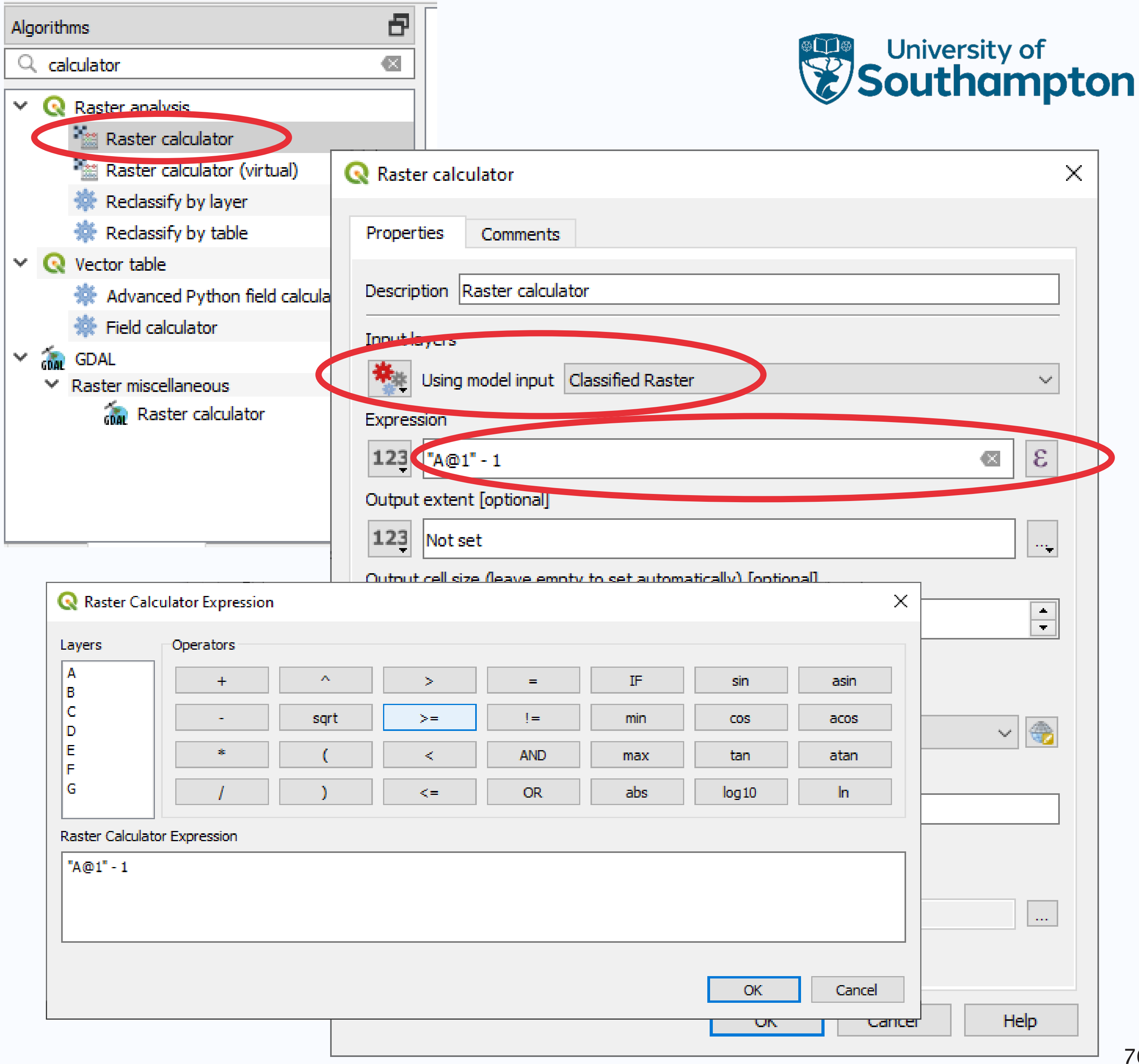
1. Go to Algorithms tab
2. Search for Raster calculator in the search bar
3. Double click on Raster Calculator in Raster analysis
4. Define Expression (e.g. “Classified Raster@1”-1)
5. Open Reference layer Multiple selection window
6. Select Classified Raster so that output CRS, extent, and cell size are adjusted according to this layer



The first processing consists in changing the values of the classification output from 1 and 2 to 0 and 1, respectively.

It is a simple algebraic operation where we use the Raster Calculator to subtract 1 from the classified raster.

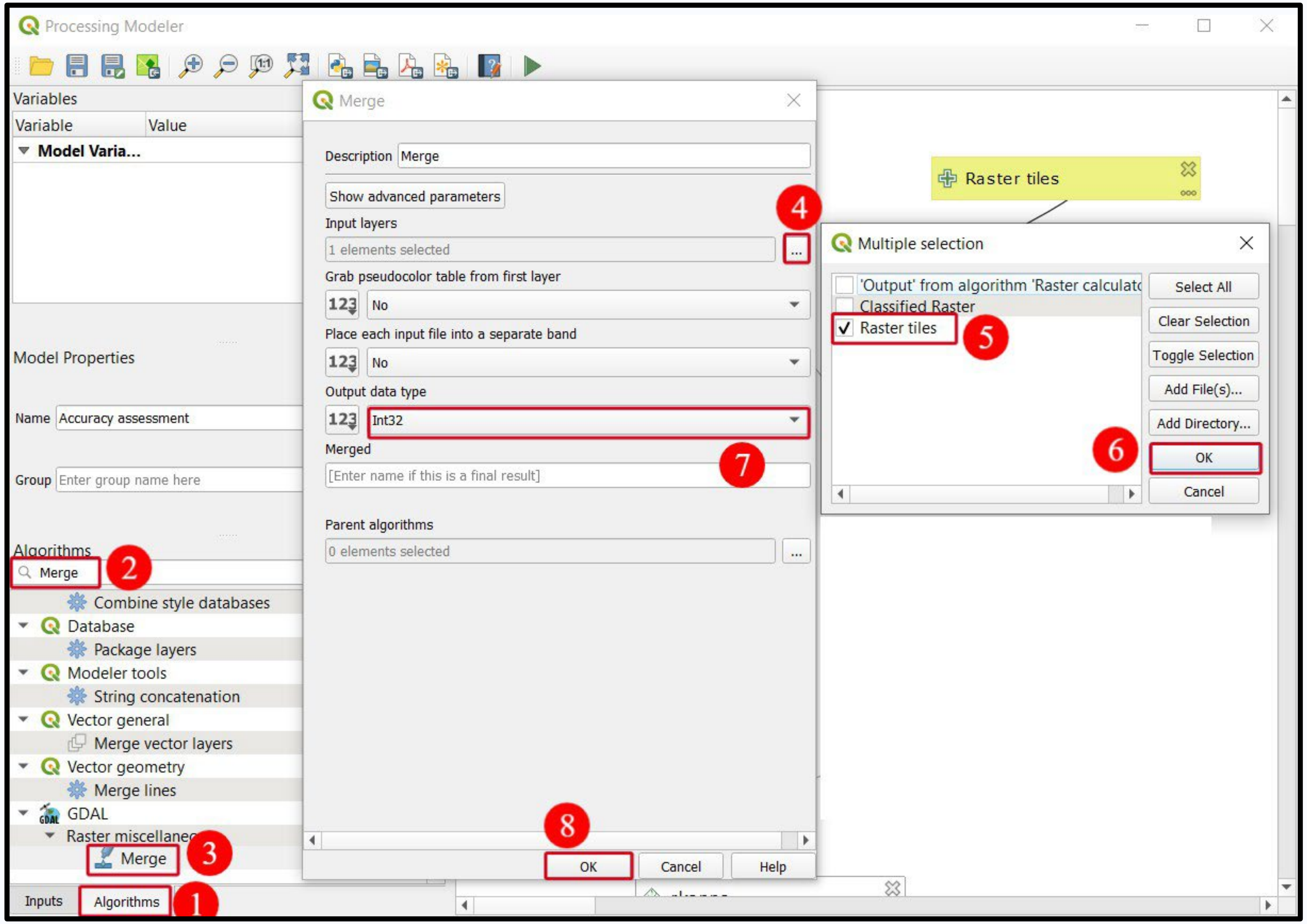
1. Search for Raster calculator in the Algorithms search bar and open
2. Choose ‘Model Input’ option for the Input layers parameter and select the *Classified Raster* input created in the previous step
3. For the expression, use the button on the right side to open the builder.
4. Layers A-G are the inputs, but only A is used in this case. Double click A to insert into the expression. ‘A@1’ denotes band 1 of raster A
5. Type or use the buttons to complete the expression and click Ok
6. Leave the remaining parameters unchanged and click Ok



University of Southampton

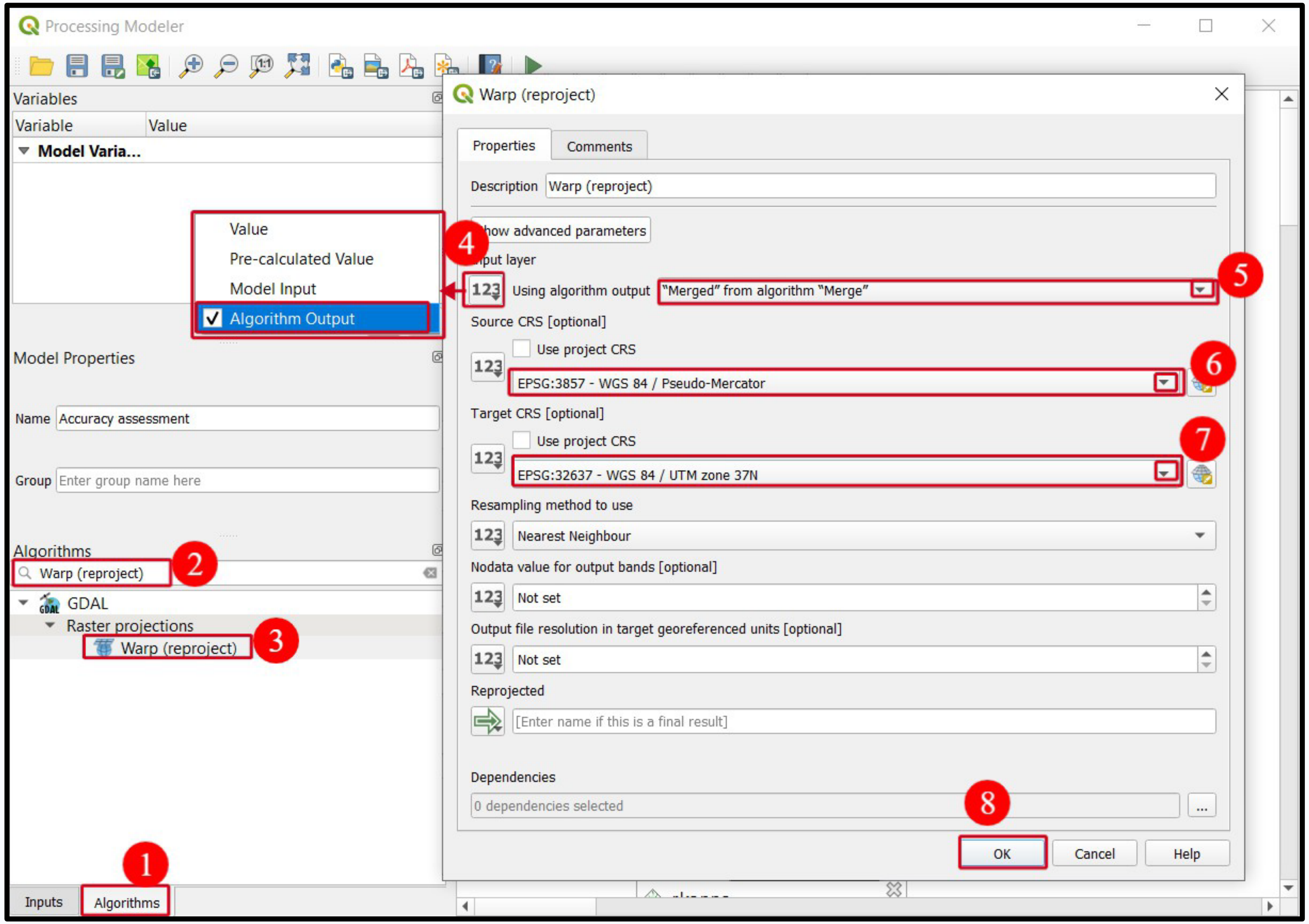
Now we process GHS BUILT S1. The first operation is to merge the 4 tiles of GHS BUILT S1 into a single tile.

1. Go to Algorithms tab
2. Search for Merge in the search bar
3. Double click on Merge in *GDAL > Raster miscellaneous*
4. For selecting Input layers open Multiple selection window
5. Select input layers (e.g., Raster tiles)
6. Click on OK to confirm selection
7. Select Output data type to be integer (e.g., Int32)
8. Click on OK



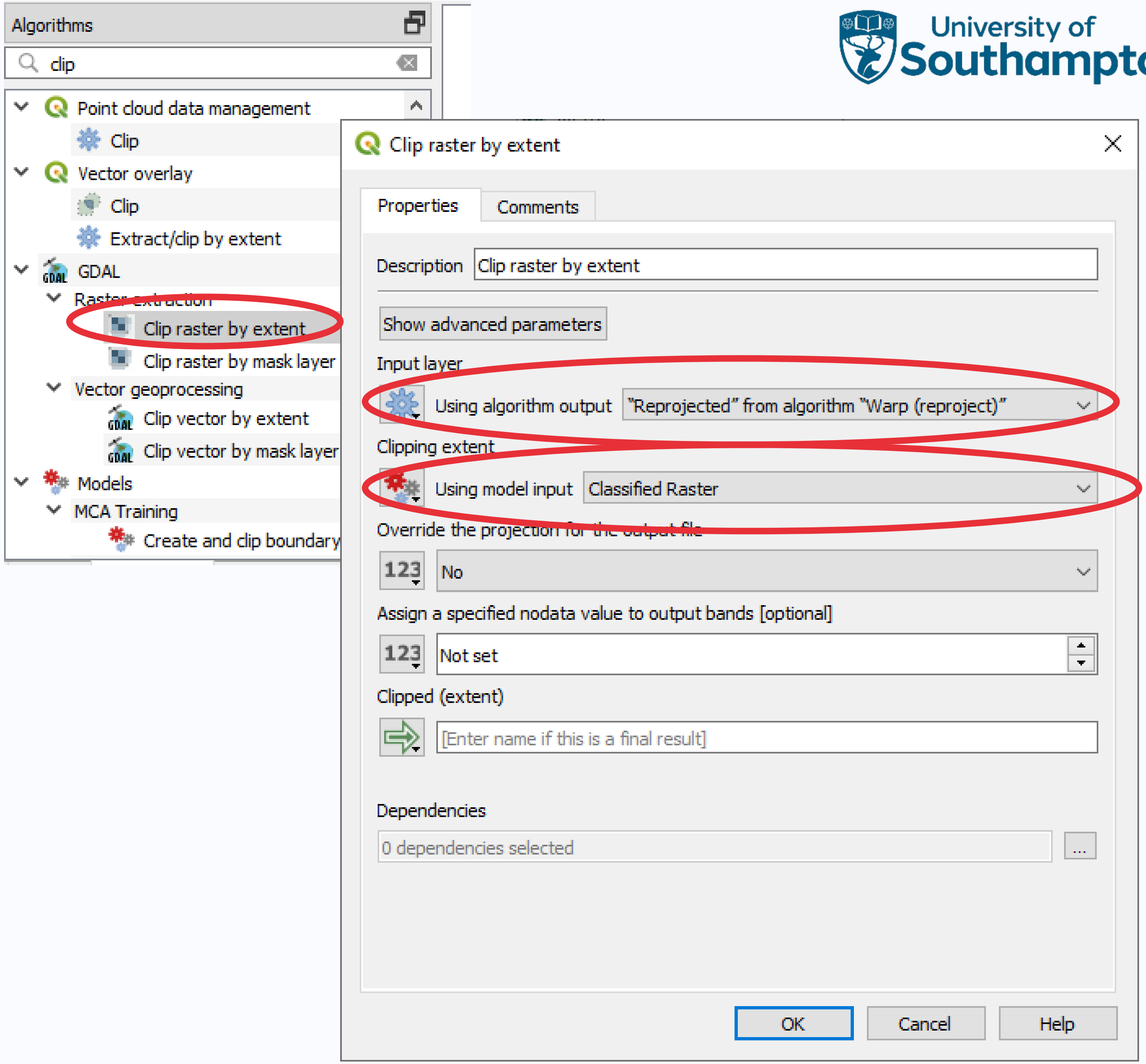
Next step is to reproject GHS-BUILT S1 to the CRS of classification output

2. Search for Warp (reproject) in the Algorithms search bar
3. Select Warp (reproject) in *GDAL > Raster projections*
4. Select Algorithm Output as a source of input data
5. Selecting Input layer from drop-down menu (e.g., outcome of merge operation denoted as ‘Merged’ from algorithm ‘Merge’)*
6. Select Source CRS (e.g., EPSG: 3857)
7. Select Target CRS (e.g., EPSG: 32637)
8. Click on OK



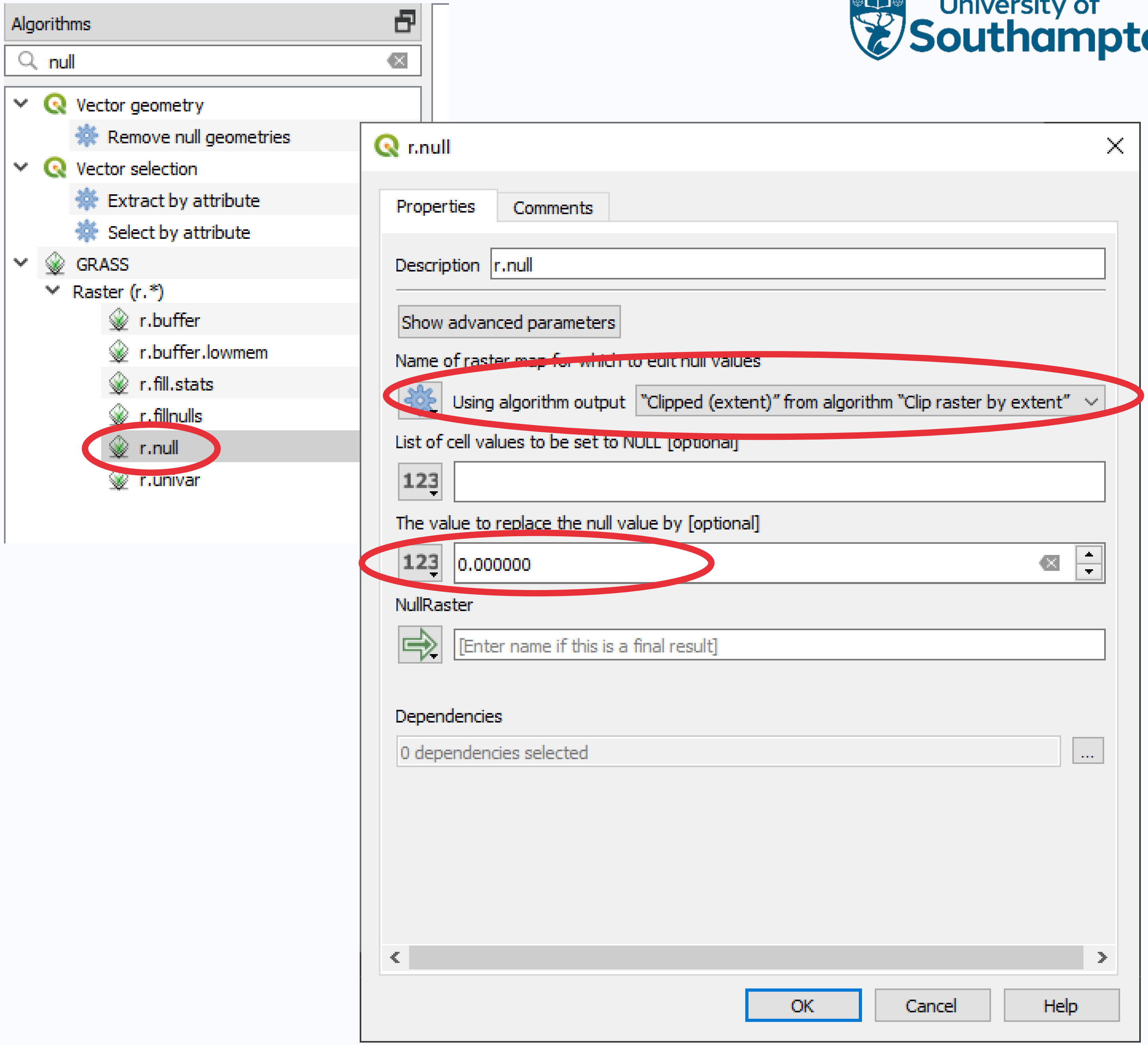
Now we clip GHS-BUILT S1 to the extent of the classification output

1. Search for Clip raster by extent in the Algorithms search bar
2. Select Clip raster by extent in *GDAL > Raster extraction*
3. Select 'Algorithm output' as the source for the Input layer and set to *'Reprojected' from algorithm 'Warp(reproject)'*
4. Select 'Model input' and the source for the Clipping extent and select *Classified Raster*
5. Click on OK



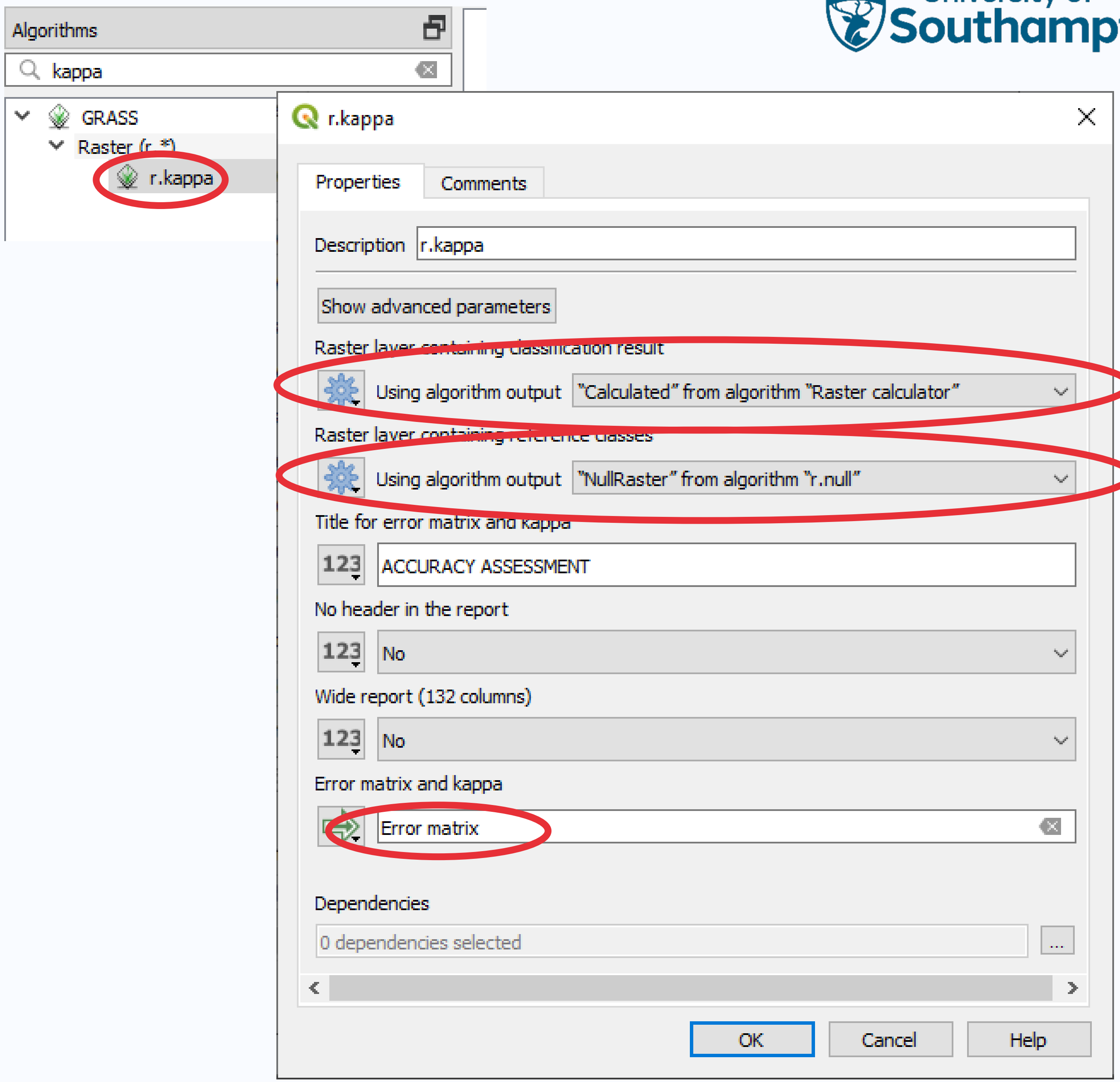
Finally, we need to replace NULL values of GHS BUILT S1 with 0 to match the classes of the classification output

1. Search for null in the Algorithms search bar
2. Select r.null in *GRASS > Raster (r.*)*
3. Select 'Algorithm output' as the source for the raster map and set to *'Clipped (extent) from…'*
4. Set the value to replace the null value by to 0.0
5. Click on OK



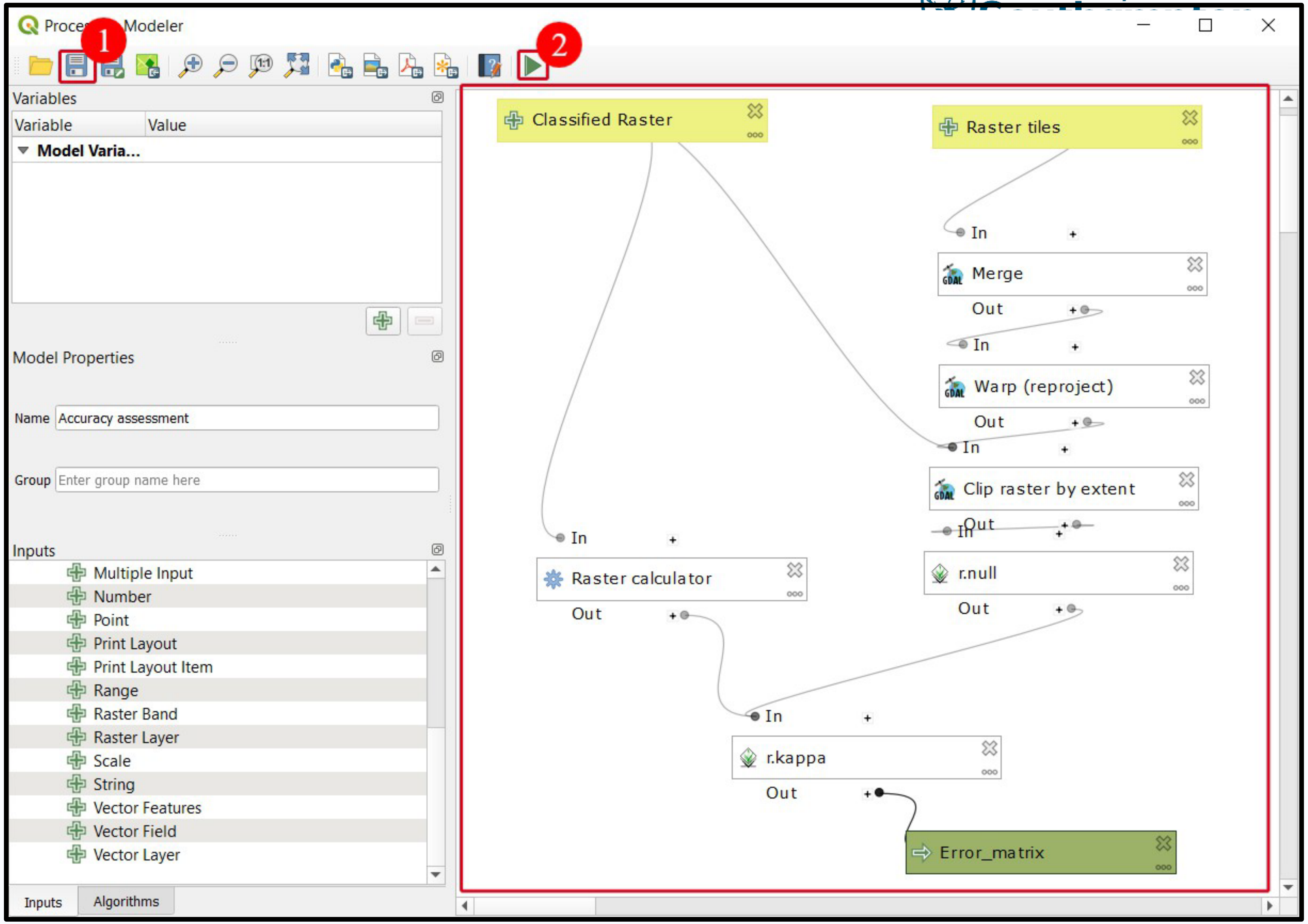
Now we are ready to generate the error matrix and accuracy indexes. For that, we use the r.kappa algorithm.

2. Search for r.kappa in the Algorithms search bar
3. Select r.kappa in *GRASS > Raster(r.*)*
4. Select ‘Algorithm output’ as the source for the classification result and set to *‘Calculated from…’*
5. Select ‘Algorithm output’ as the source for the reference classes and set to *‘‘NullRaster’ from…’*
6. As the result of this calculation will be a model output, give the ‘Error matrix and Kappa’ result a name (e.g., Error matrix)



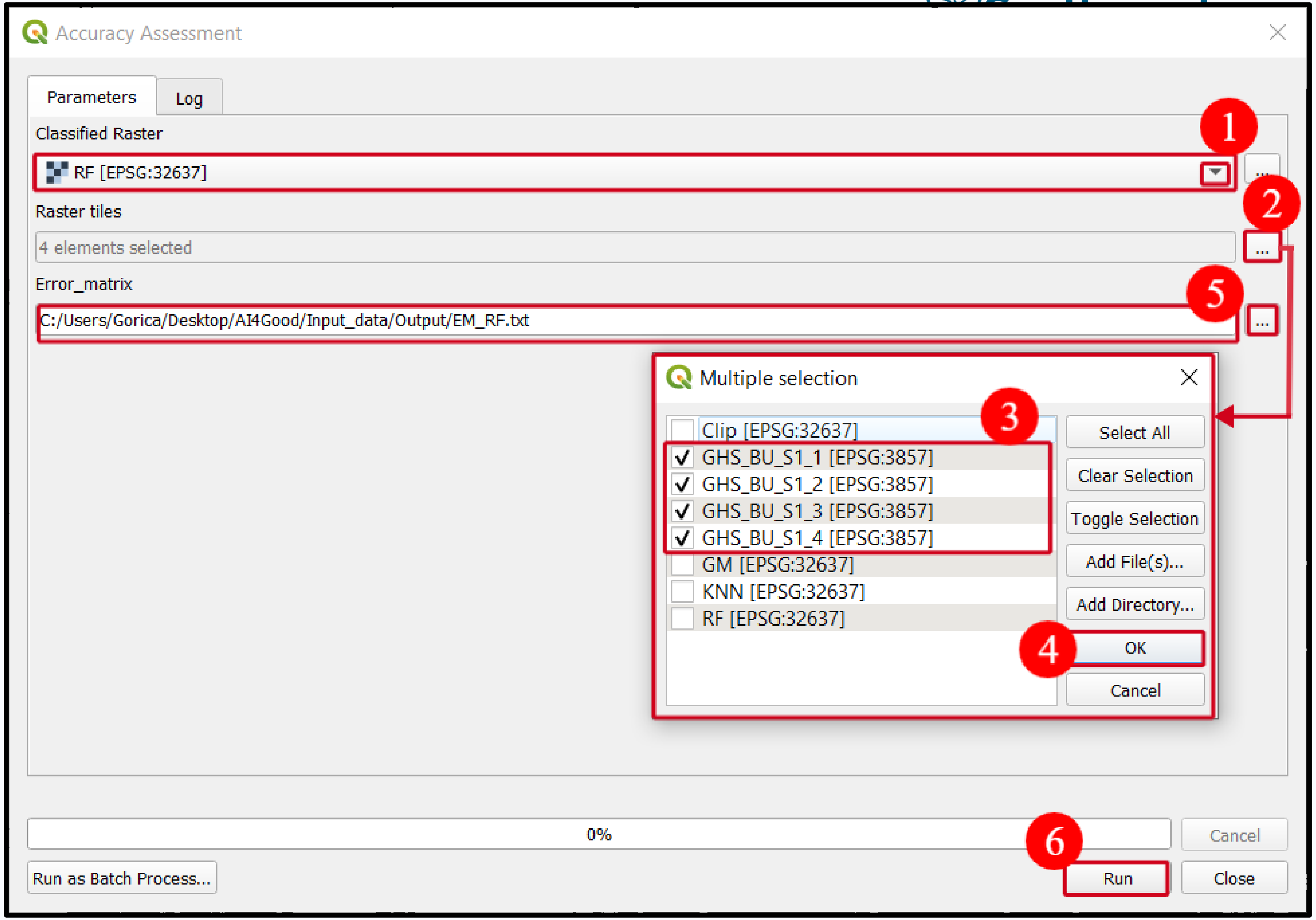
The model is ready!

- Click on Save icon to save last changes to the model
- Click on Run button to Run the model



The first step in running a model is to specify the input and output parameters:

1. From drop-down menu select one of your classification outputs
2. Open Multiple Selection for Raster Tiles
3. Select Raster Tiles (e.g., the 4 GHS_BU_S1_x)
4. Click on OK to confirm selection
5. Define the path and the filename for the output Error_matrix (e.g., EM_RF.txt)
6. Click on Run button to Run the model



University of

		Reference			
		1 Non built-up	2 Built-up	Sum prediction	UA
Prediction / classification	0 Non built-up	484810	53069	537879	90.13%
	1 Built-up	36818	90631	127449	71.11%
	Sum reference	521628	143700	665328	
	PA	92.94%	63.07%		86.49%

Result of r.kappa:

- **% Commission (error)=1-UA**
- **% Omission (error)=1-PA**
- **% Observed correct =Overall accuracy**
- **Kappa**

RF_r.kappa - Notepad

File Edit Format View Help

```
                        ACCURACY ASSESSMENT
LOCATION: temp_location                               Thu May 06 15:57:32 2021
MASK: none
MAPS: MAP1 = (untitled) (rast_6093f5c704ef44 in PERMANENT)
      MAP2 = (untitled) (rast_6093f5c704ef43 in PERMANENT)

Error Matrix (MAP1: reference, MAP2: classification)
Panel #1 of 1
                          MAP1
     cat#        0        1        Row Sum
 M    0        484810   53069      537879
 A    1        36818    90631      127449
Col Sum        521628   143700     665328


Cats    % Comission    % Omission    Estimated Kappa
0       9.866345       7.058287      0.543190
1       28.888418      36.930411     0.631533

Kappa           Kappa Variance
0.584040        0.000001

Obs Correct     Total Obs      % Observed Correct
575441          665328         86.489822

MAP1 Category Description
0:  (no description)
1:  (no description)

MAP2 Category Description
0:  (no description)
1:  (no description)
```

Ln 1, Col 1 | 100% | Unix (LF) | UTF-8