A Manual on
Prediction of Future Land Cover Changes using QGIS MOLUSCE Plugin

Innovative Tools to Support the Design and Monitoring of NBS and GHG Emission Reduction Projects (LoA/RAP/2023/48)

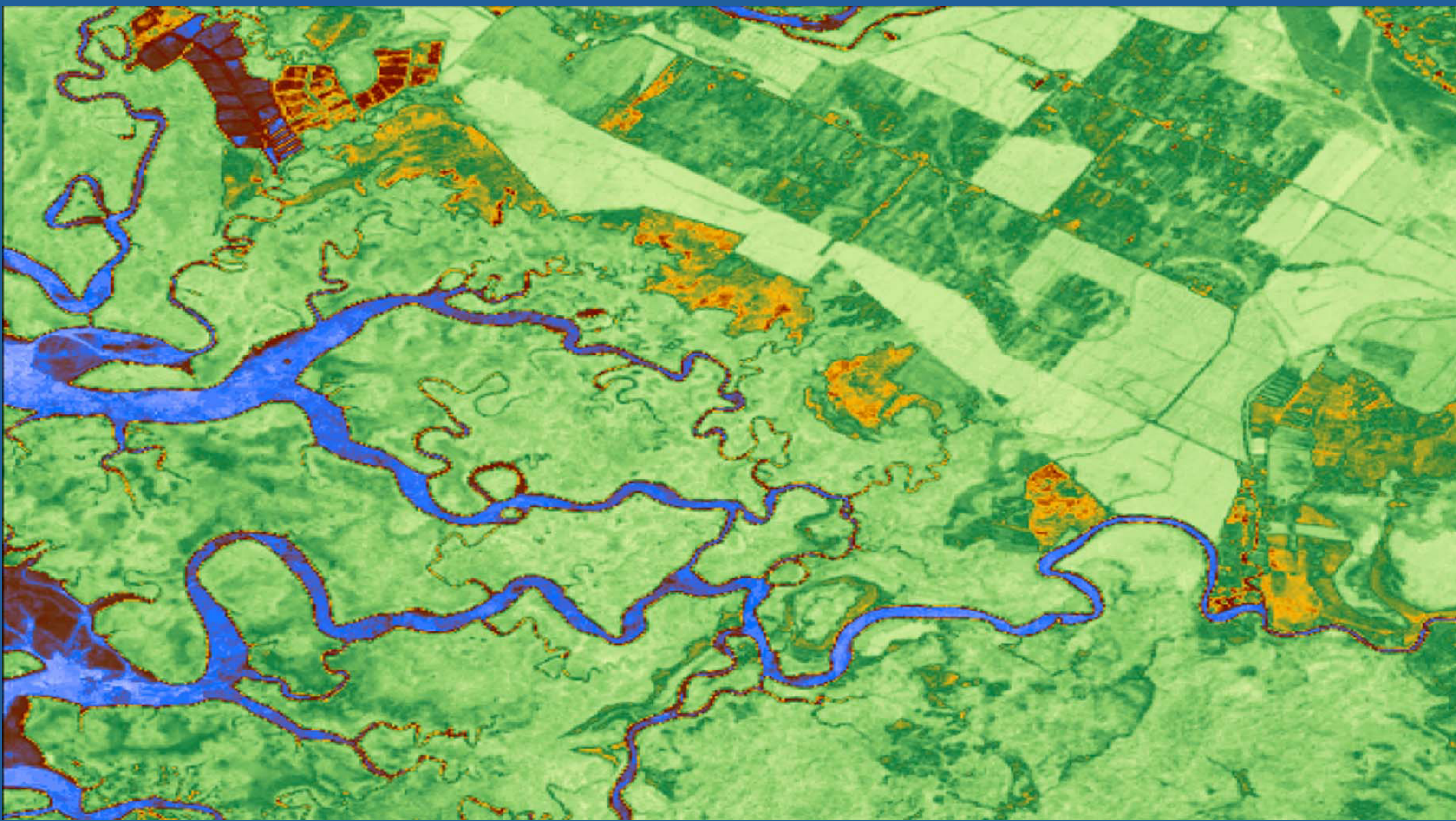


Photo Credit: NASA

PREPARED BY:

Asian Institute of Technology,

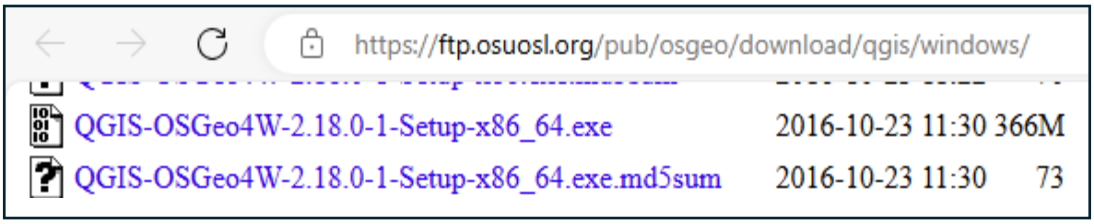
Pathum Thani, Thailand

2024

Table of Contents

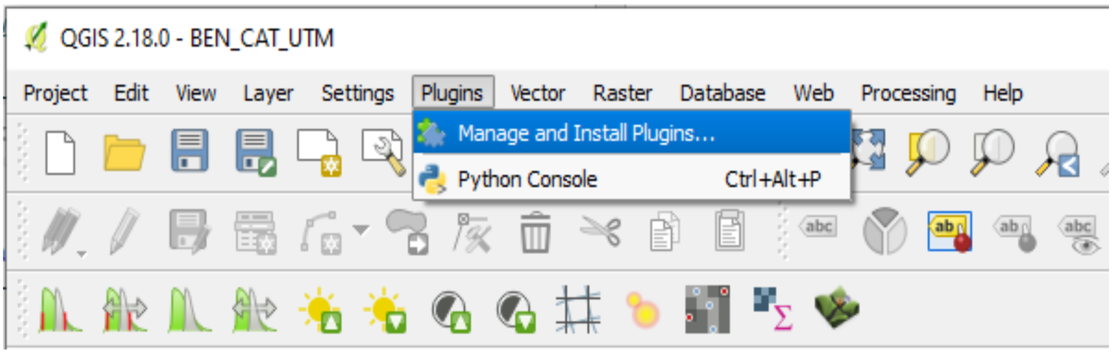
1. QGIS Installation

- Download QGIS setup file for windows from this website. The MOLUSCE plugin is supported only in QGIS version 2. Thus, QGIS version 2.18.0 is recommended.

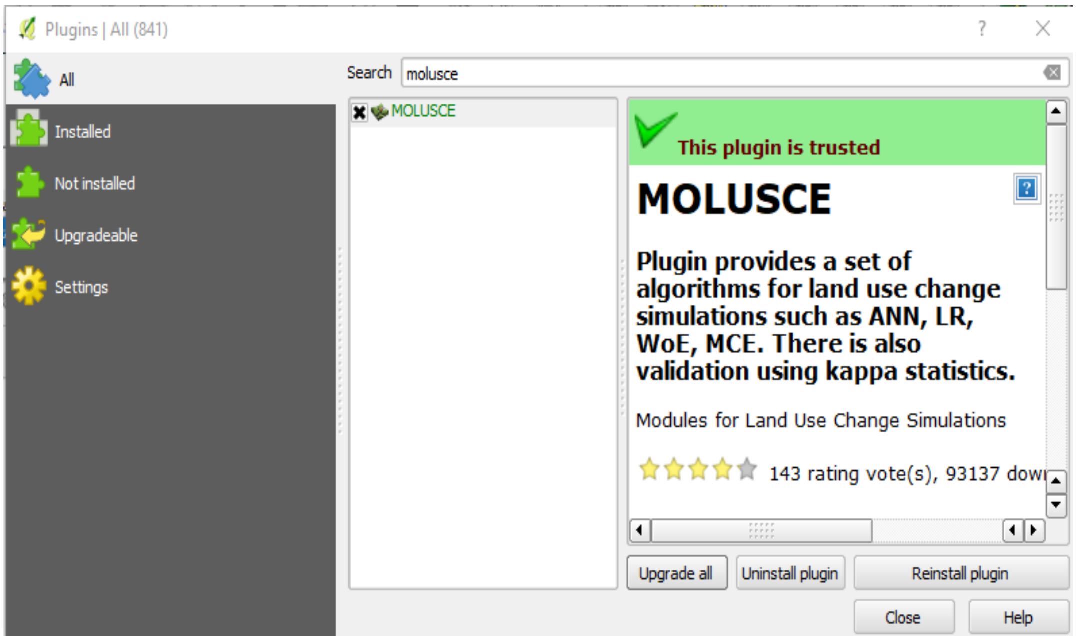


2. MOLUSCE Plugin Installation

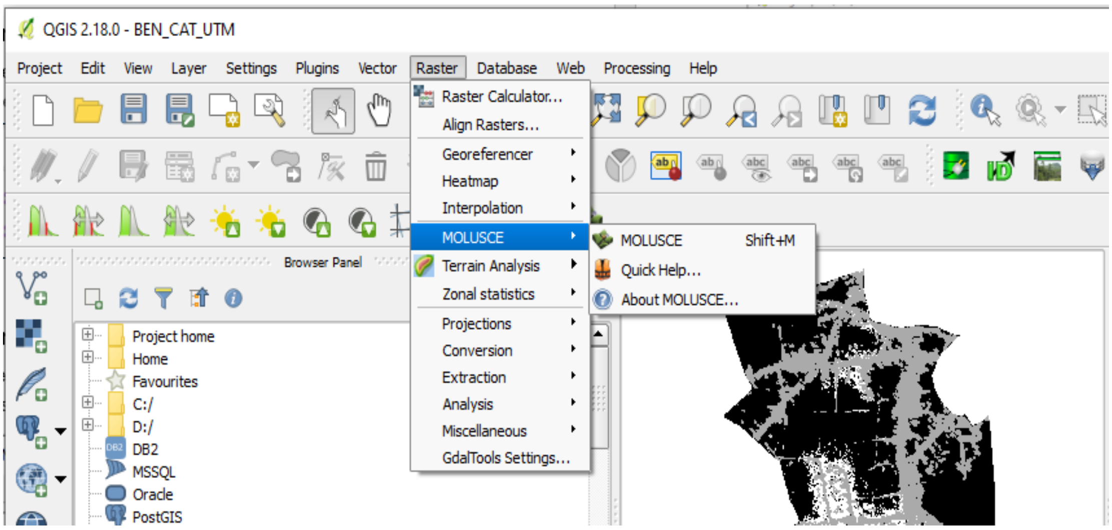
- Open QGIS application and create new empty project. In the menu bar, click the ***Plugins*** tab and click on ***Manage and Install Plugins*** option to open ***Plugins*** window as shown in Figure below.



- Search plugin by writing MOLUSCE keyword in search bar. Then, select MOLUSCE and click ***Install plugin*** button to install the plugin.



- After the successful installation of plugin, you would be able to see the ***MOLUSCE*** option listed in the Raster menu.



3. Study Area Selection

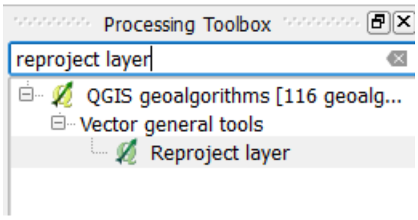
a. First of all, select the area where you would like to carry out case study. The extent of study area could be at national, regional, municipal, ward or any spatial unit depending upon requirements.

b. In this case study, Ben Cat district of Vietnam was chosen. The district level administrative layer in shapefile format was downloaded from OpenDevelopment Mekong data portal (link).

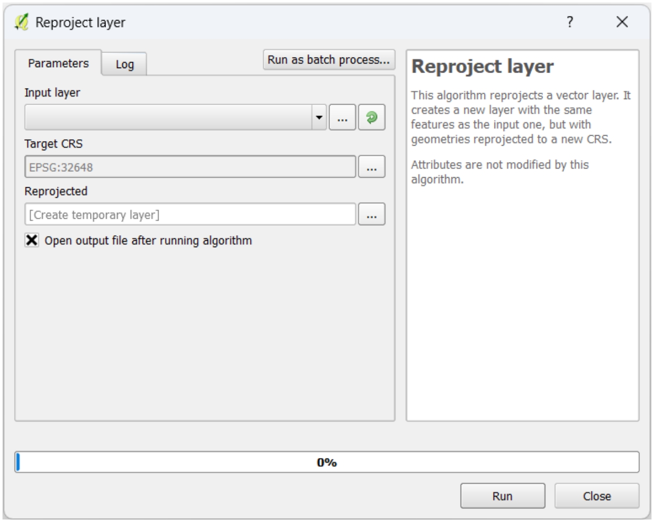
c. Layer was loaded in QGIS and only the BenCat district was exported.

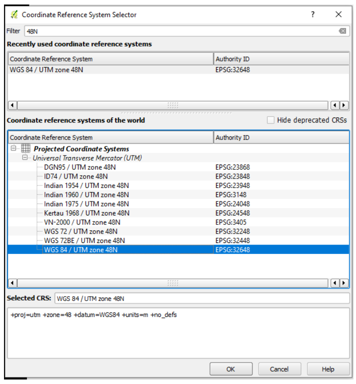
d. As projected coordinate system is preferred, it is essential to know to the UTM zone of the selected study area. To find the UTM zone, use the link1 or link2.

e. The Ben Cat district lies in WGS UTM zone 48N (EPSG: 32648). Hence, the layer was projected using **Reproject Layer** tool.
   - Find the UTM zone using this link1 or link2
   - Find and open the "Reproject Layer" tool from processing toolbox



   - Select the layer you want to reproject as an **Input layer**. Find and set Target CRS as per the required (WGS84/ UTM ZONE 48N in this case study)





- Run and save the reprojected layer

4. Dataset Preparation

4.1 Land Cover Dataset

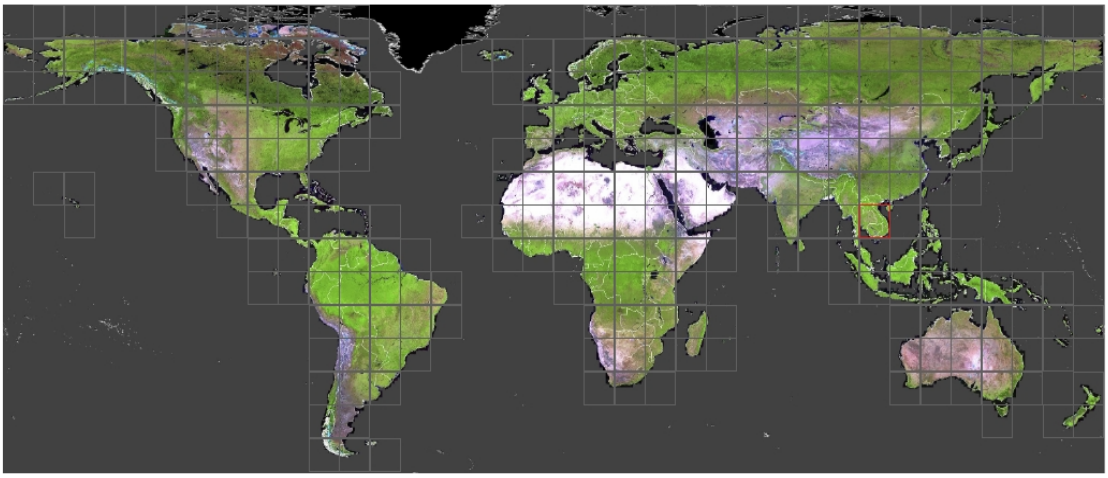
a. First of all, select the years you need/want to map the land cover. If you want to prepare land cover map from scratch, start by collecting high-resolution satellite or aerial imagery and ground truth data. Preprocess the data by correcting for radiometric and geometric distortions and remove clouds if necessary. Annotate the imagery by defining land cover classes and either manually digitizing or using semi-automatic methods with machine learning tools. Validate the dataset through comparison with ground truth and accuracy assessments, then refine as needed.

   Most of the countries or researchers have mapped the land cover in different levels which are publicly available. You can also find global land cover map datasets in different sources. Find the datasets and download the dataset and analyze its quality and reliability for the selected study area. While selecting the years, there should be at least three consecutive years for land cover prediction and should be of equal interval.

b. For the Ben Cat district case study, data were downloaded from Global Land Analysis and Discovery website (link). This dataset can be found in 10*10 degree tiles and has a spatial resolution of 0.00025° per pixel (~30 meters at the equator). The annual land cover and land use maps of years 2000, 2005, 2010, 2015 and 2020 can be found.

   - First select the grid, where the area of interest falls. Since Ben Cat district lies in 20N 100E grid, the grid was selected.
   - After the grid selection, the data download links will be listed below.

- Click the links to download the data for respective years.



https://storage.googleapis.com/earthenginepartners-hansen/GLCLU2000-2020/v2/2000/20N_100E.tif

https://storage.googleapis.com/earthenginepartners-hansen/GLCLU2000-2020/v2/2005/20N_100E.tif

https://storage.googleapis.com/earthenginepartners-hansen/GLCLU2000-2020/v2/2010/20N_100E.tif

https://storage.googleapis.com/earthenginepartners-hansen/GLCLU2000-2020/v2/2015/20N_100E.tif

https://storage.googleapis.com/earthenginepartners-hansen/GLCLU2000-2020/v2/2020/20N_100E.tif

https://storage.googleapis.com/earthenginepartners-hansen/GLCLU2000-2020/v2/2000-2020change/20N_100E.tif

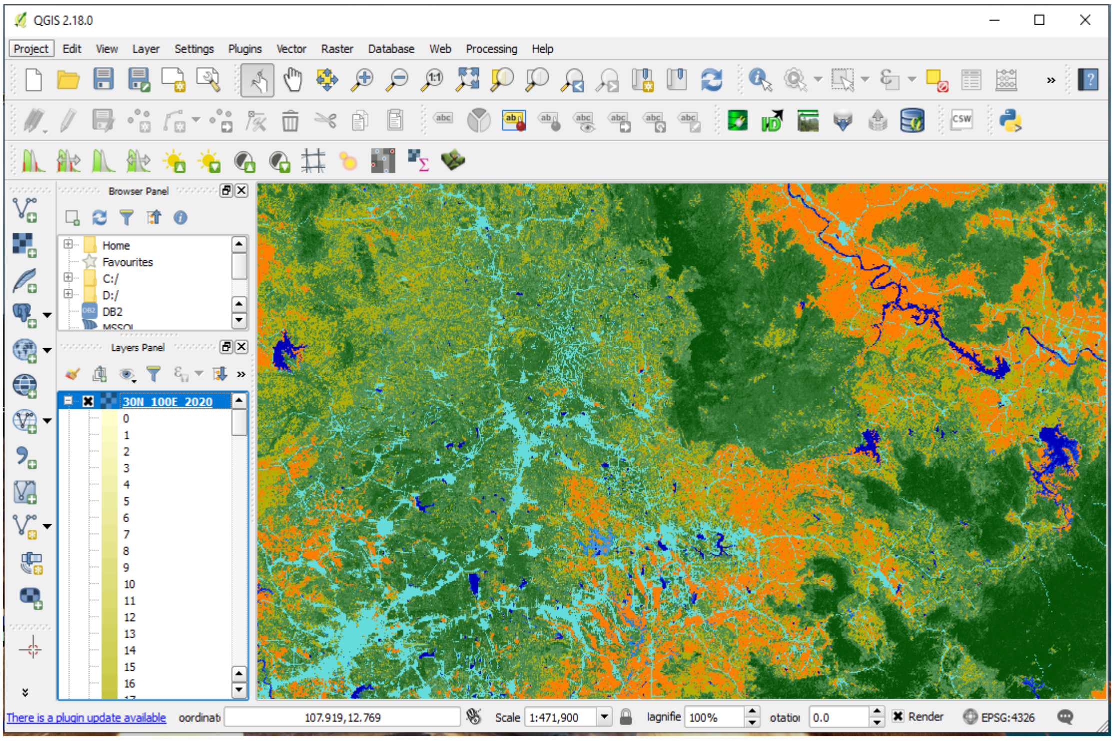
c. Import downloaded data into QGIS, you will see the gradient legend in layer panel.

Refer to this legend on which land cover mapping is based on.

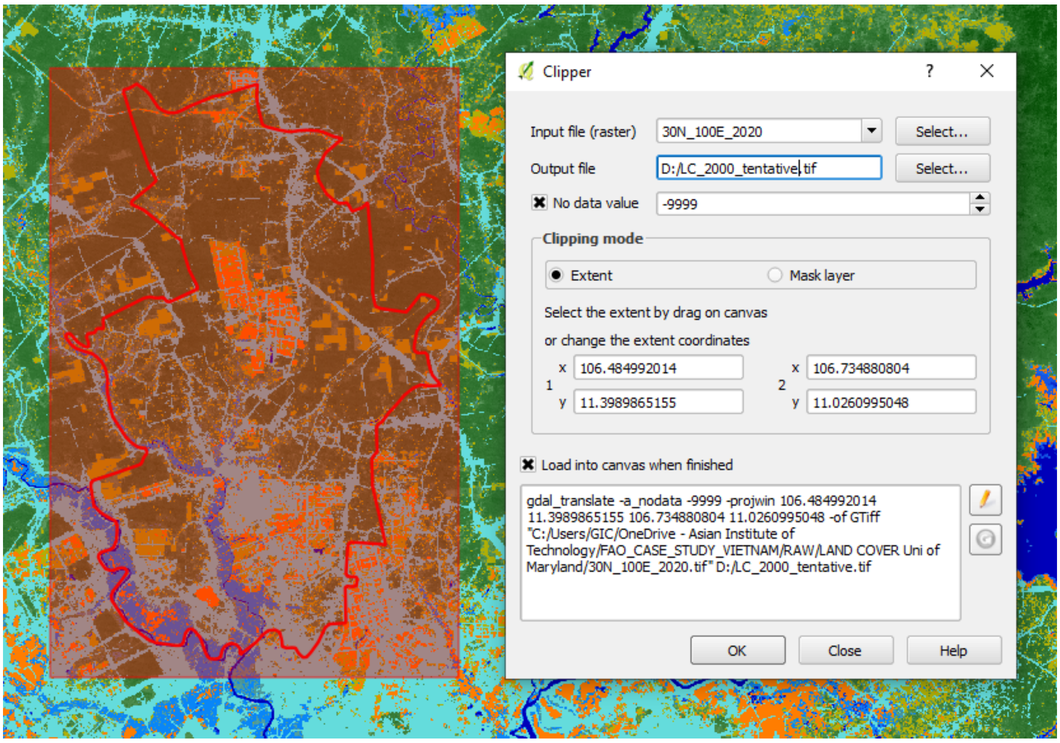
Since we are not working on these all these classes, reclassification is required. In this case study, land cover classes were reclassified as follows:

Pixel value	New Pixel value	Land cover class
0-48, 100-148	1	Forest
200-207	2	Water Bodies
250	3	Built-up
244	4	Cropland

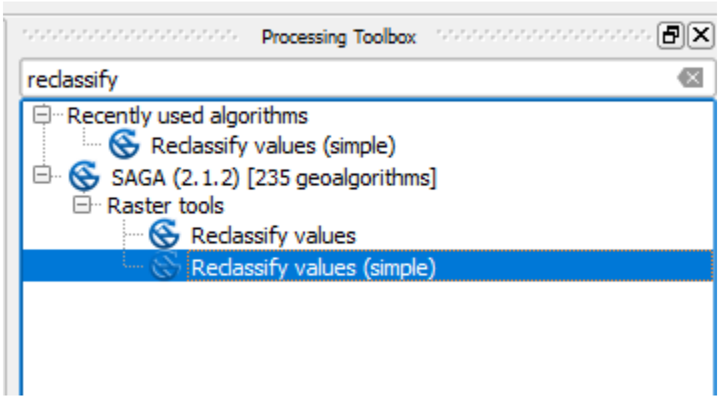
Since snow/ice and ocean land cover are not in the study area, they are not mentioned in above reclassification table.

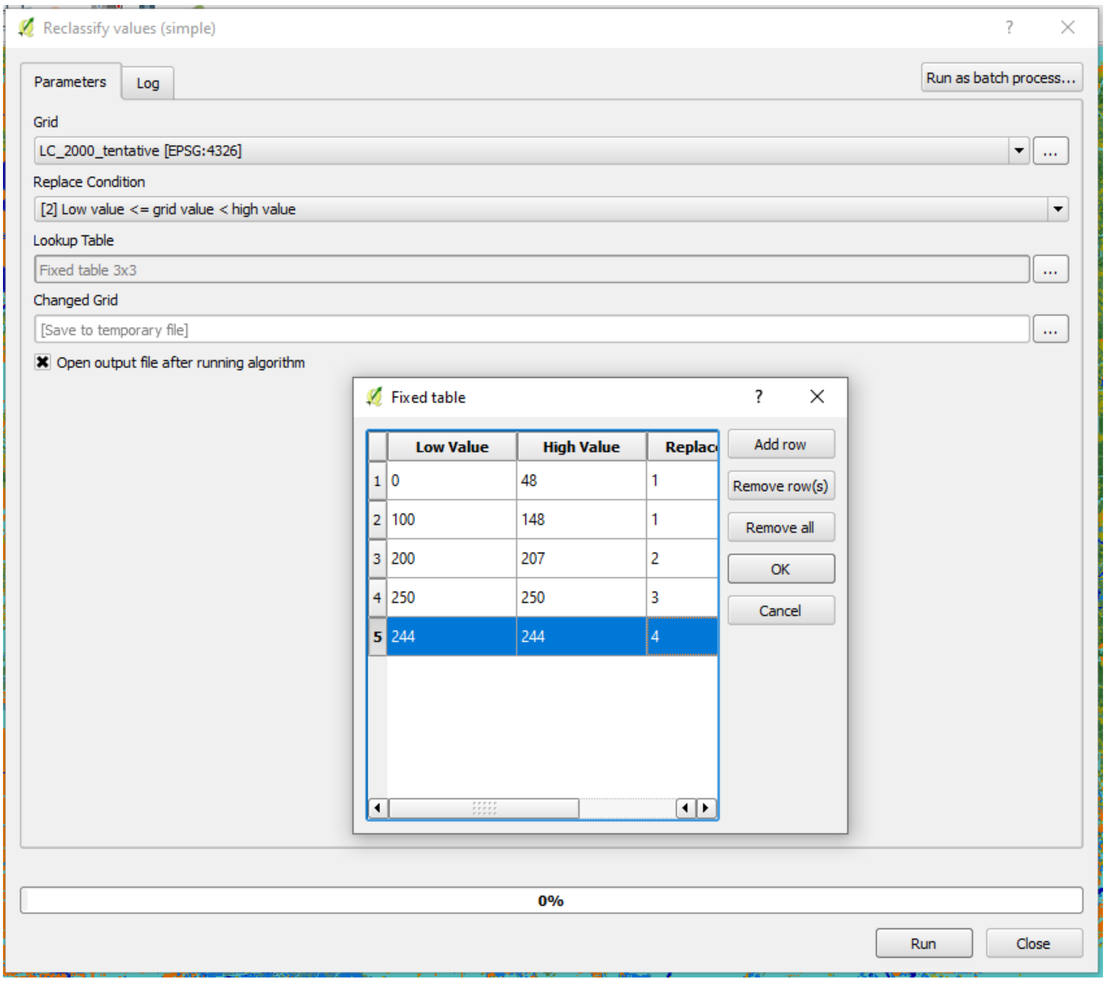


d. First clip the tentative study area using Clipper tool available in extraction option of Raster menu.



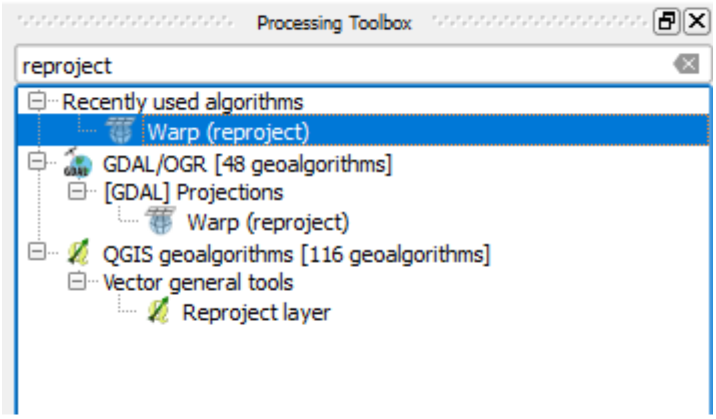
e. For the reclassification in QGIS, open ***Reclassify values (simple)*** tool.
   - Select the layer you want to reclassify as grid
   - Set the replace condition
   - Define lookup table as shown in figure below
   - Run the reclassification tool.





f. To work in projected coordinate system, open the Warp (reproject) tool to project the dataset

- Find the UTM zone using this link1 or link2.



- Select the layer you want to reproject as an input layer.
- Select the destination CRS as required (WGS UTM zone 48N in this case study).
- In this case study, the spatial resolution of dataset is 0.00025 degree.

  1 degree= 111.11 km

  0.00025 degree=0.00025111.11*1000 m

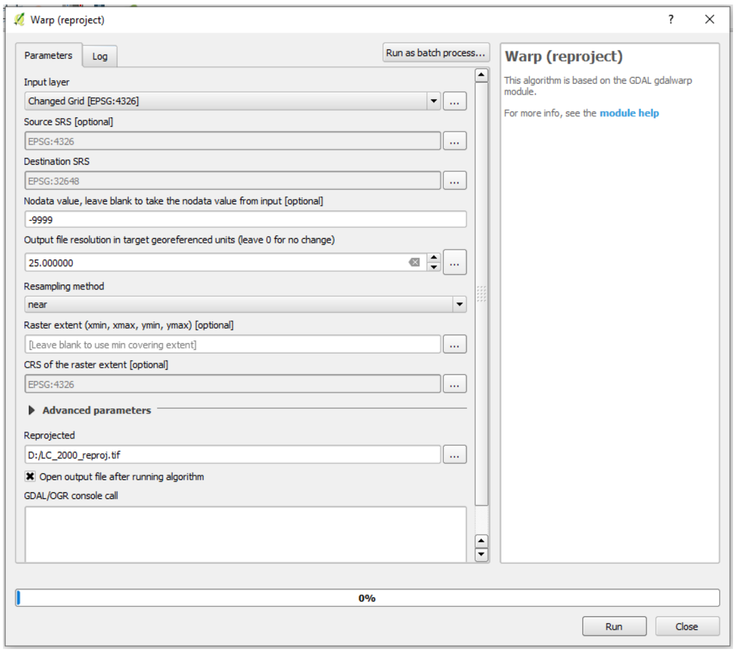
  0.00025 degree=27.75m

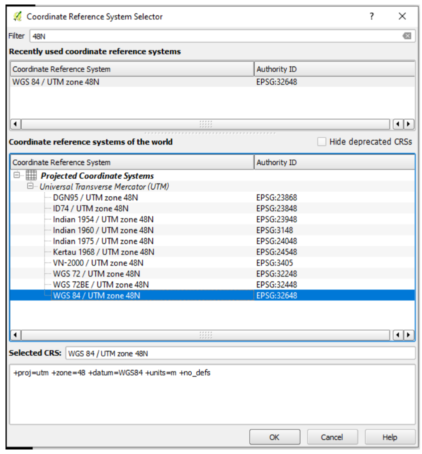
  Therefore, output resolution is set as 25m is this case study and "nearest neighbourhood" is selected as resampling method.

  You can also set output resolution as 30m or 27.75m but it should be same when to prepare driving factors datasets as well.

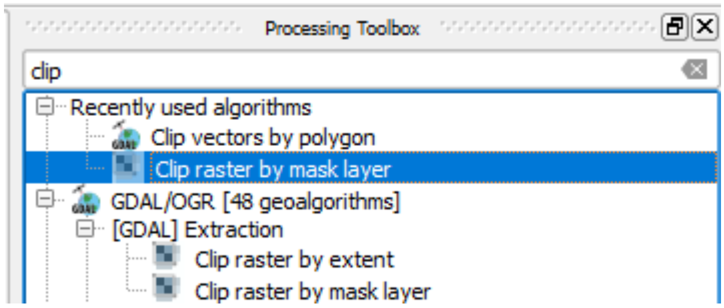
- Define the output file path in Reprojected

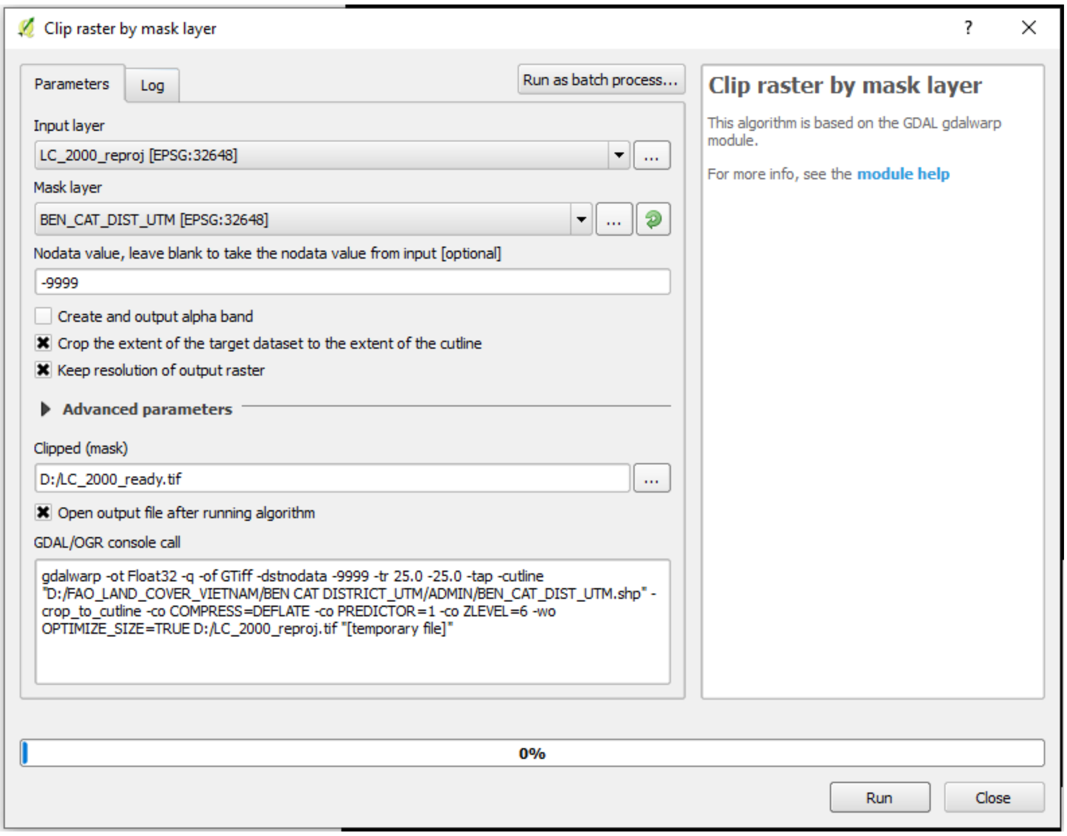
- Run the tool





g. Finally, clip the raster layer within the study area administrative boundary.





- Select the reprojected land cover layer as an input layer
- Select study area admin as a mask layer layer (Ben Cat District boundary data in this case study)
- Select the options crop the extent of the target dataset to the extent of the cutline and keep resolution of output raster.
- Define the output file path in Clipped (mask)
- Run the tool
- The final properties of land cover map of this case study are as follows:

  No of classes: 4

  Pixel size:25, -25

  Columns:1004

  Rows: 1559

  Coordinate Reference System (CRS): EPSG 32648, WGS 84/UTM ZONE 48N

h. Repeat the process to prepare the land cover datasets for other selected years.

4.2 Driving Factors

MOLUSCE Plugin only accepts the raster format datasets.

In this case study, population density, slope, hillshade, digital elevation model, distance from the road, distance from the rivers were considered as the driving factors based on the data availability. The preparation procedure involved in preparation of each dataset to feed into the MOLUSCE plugin is described below in detail.

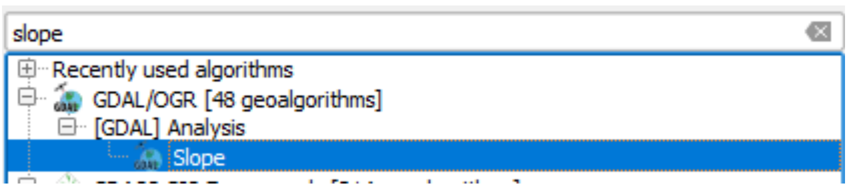
a) Digital Elevation Model and Hillshade
   - Clip the tentative study area using ***Clipper*** tool available in extraction option of Raster menu [refer to 4.1 (d)]
   - Reproject data into the selected projected coordinate system and set preferred spatial resolution which should be same as the land cover dataset. Select the bilinear convolution as a resampling method. [refer to 4.1 (f)]
   - Finally, clip the layer within the study area administrative boundary [refer to 4.1 (g)]

b) Population density
   - Clip the tentative study area using ***Clipper*** tool available in extraction option of Raster menu [refer to 4.1 (d)]
   - Reproject data into the selected projected coordinate system and set preferred spatial resolution which should be same as the land cover dataset. Select the cubic convolution as a resampling method. [refer to 4.1 (f)]
   - Finally, clip the raster layer within the study area administrative boundary [refer to 4.1 (g)]

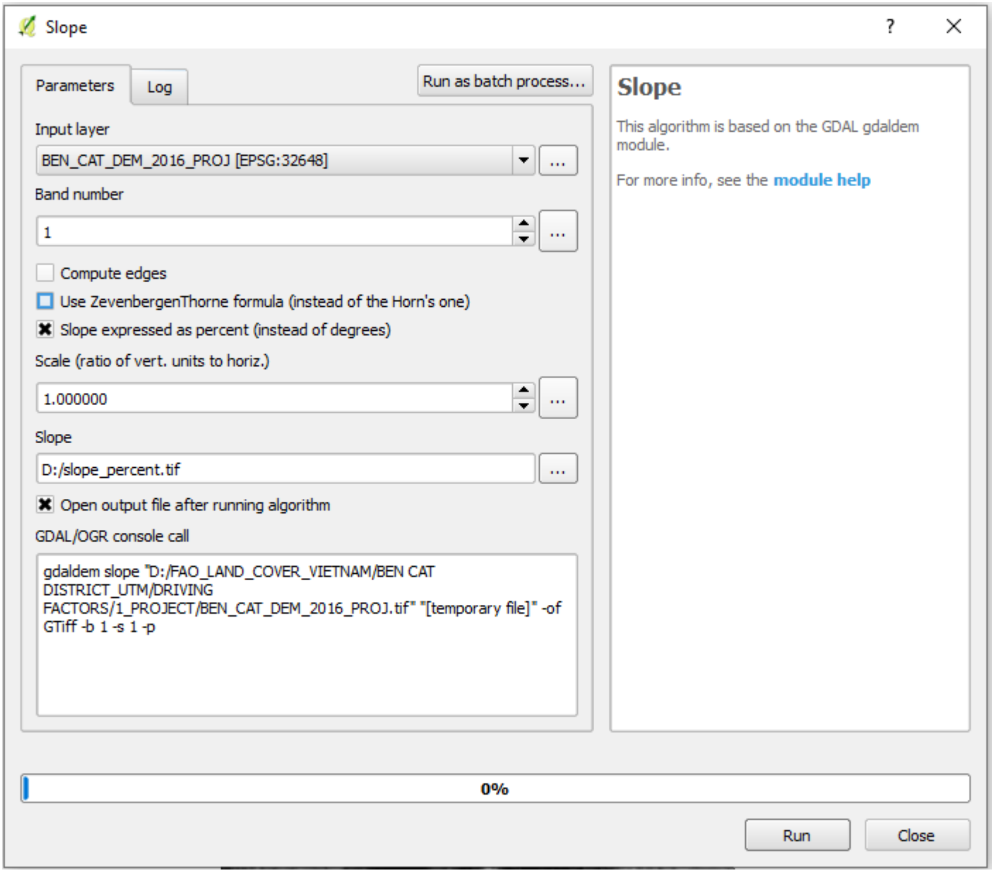
c) Slope

   Slope data is derived from the digital elevation dataset. Use the projected digital elevation data.

   - Open GDAL based slope analysis tool.



- Check the Slope expressed as percent (instead of degrees) option

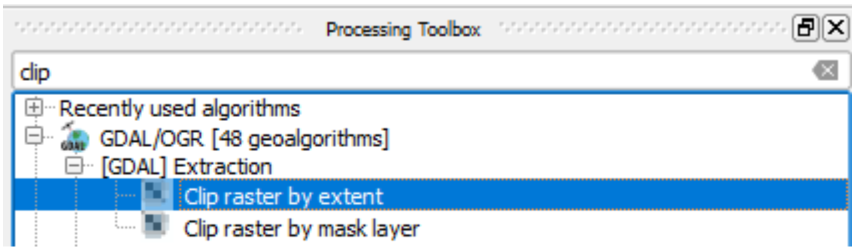


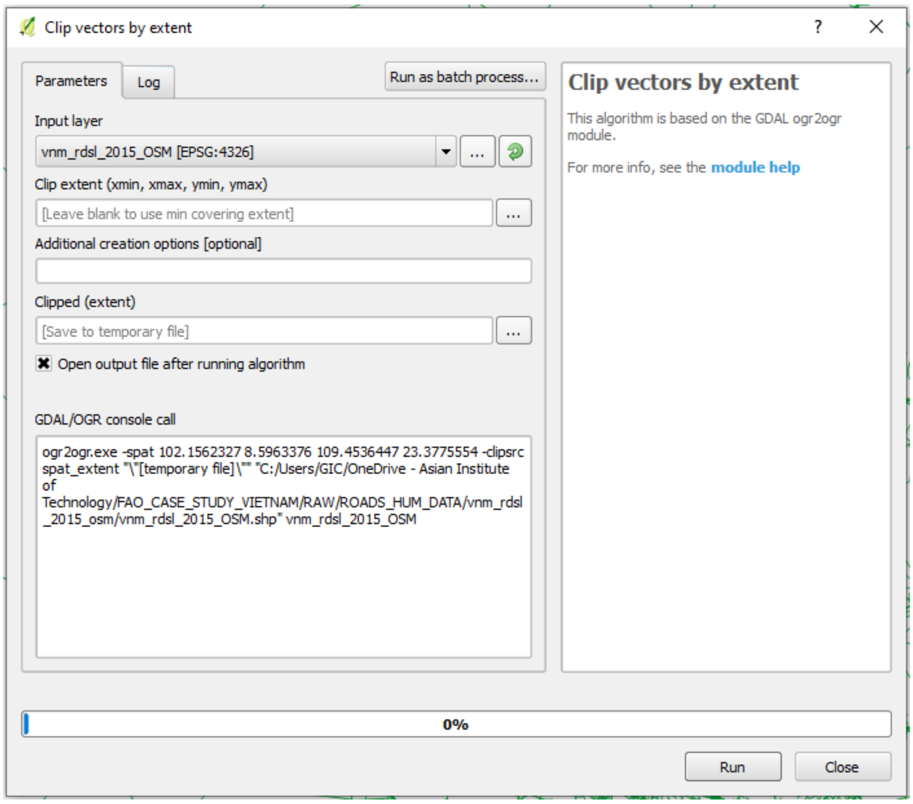
- Run the tool
- Finally, clip the slope layer within the study area administrative boundary [refer to 4.1 (g)]

d) Distance from the Road and Distance from the River

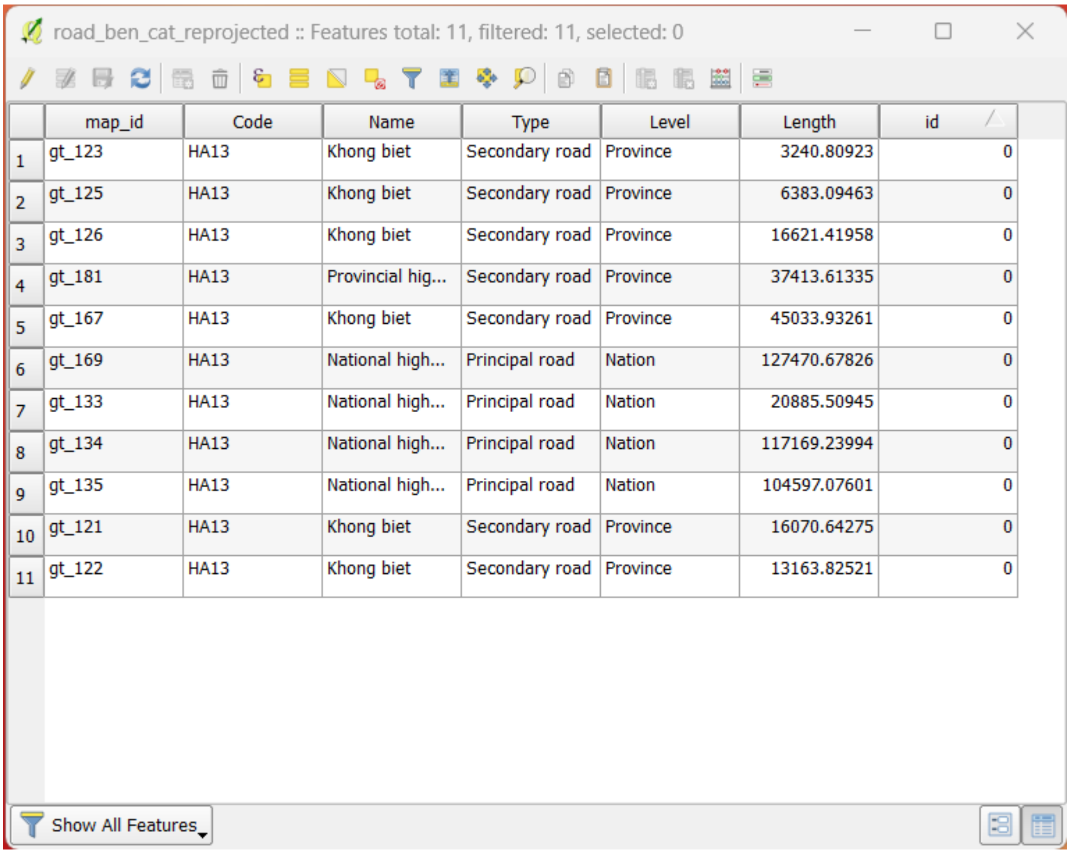
Road network and river network datasets are vector datasets which are mostly found in .shp or .geojson format.

- Clip the tentative study area using **Clip vectors by extent** geoprocessing tool. Click the Select extent on canvas option and draw tentative extent.



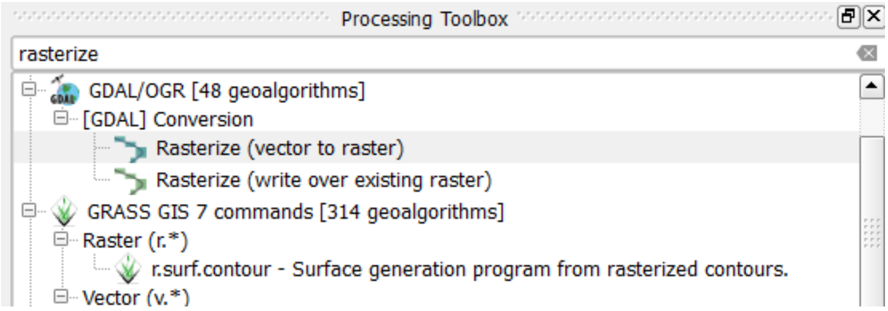


- Reproject data into the selected projected coordinate system [Refer 3 (e)].
- Open the attribute table window, toggle on the editing mode, add new field (Eg: 'id' in this case) and populate the field with value 0.

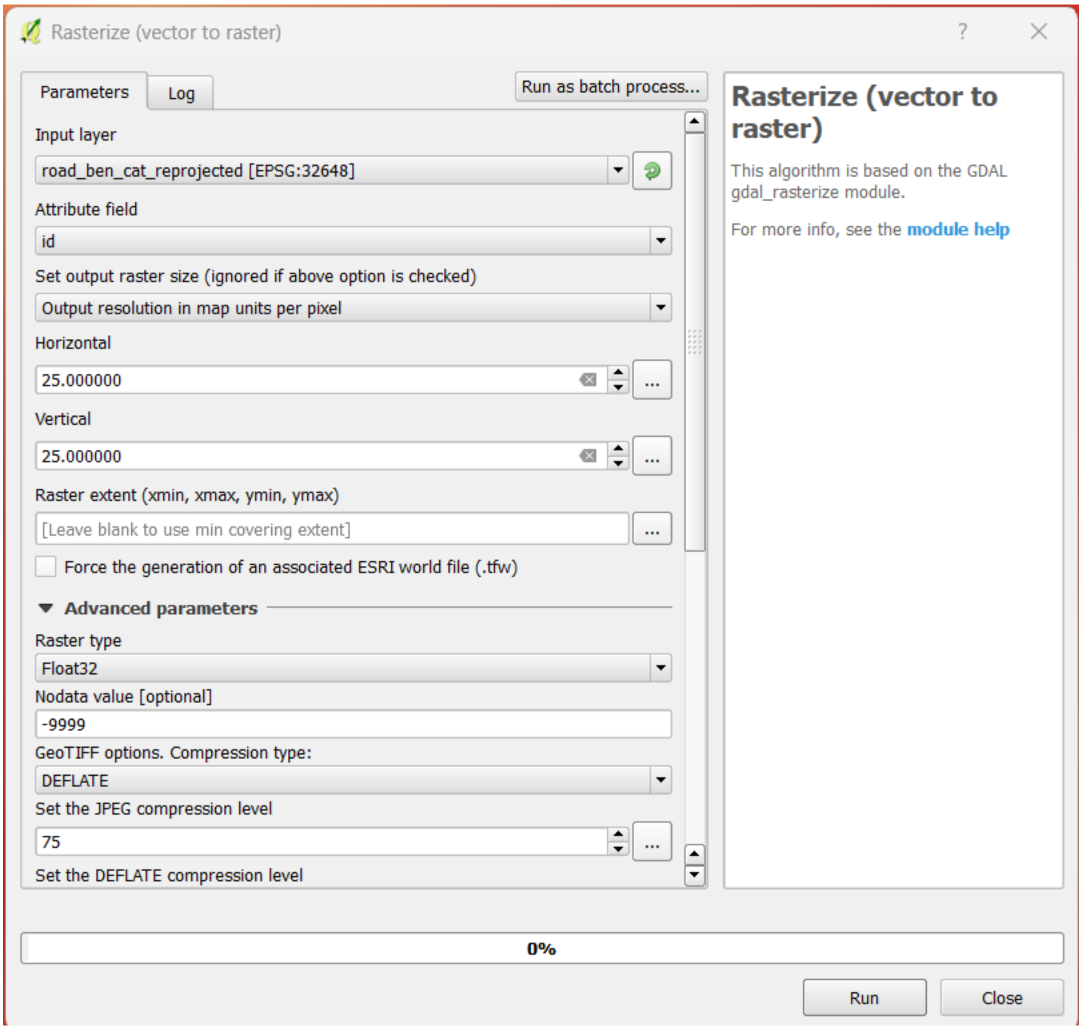


	map_id	Code	Name	Type	Level	Length	id
1	gt_123	HA13	Khong biet	Secondary road	Province	3240.80923	0
2	gt_125	HA13	Khong biet	Secondary road	Province	6383.09463	0
3	gt_126	HA13	Khong biet	Secondary road	Province	16621.41958	0
4	gt_181	HA13	Provincial hig...	Secondary road	Province	37413.61335	0
5	gt_167	HA13	Khong biet	Secondary road	Province	45033.93261	0
6	gt_169	HA13	National high...	Principal road	Nation	127470.67826	0
7	gt_133	HA13	National high...	Principal road	Nation	20885.50945	0
8	gt_134	HA13	National high...	Principal road	Nation	117169.23994	0
9	gt_135	HA13	National high...	Principal road	Nation	104597.07601	0
10	gt_121	HA13	Khong biet	Secondary road	Province	16070.64275	0
11	gt_122	HA13	Khong biet	Secondary road	Province	13163.82521	0

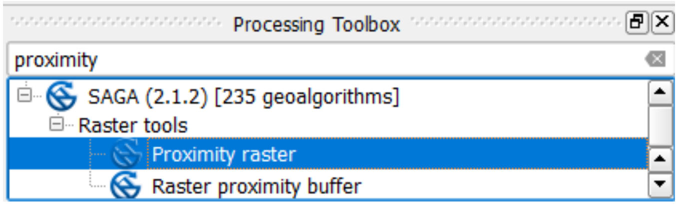
- Using the **Rasterize (vector to raster)** tool, the river and road datasets were converted into raster datasets.



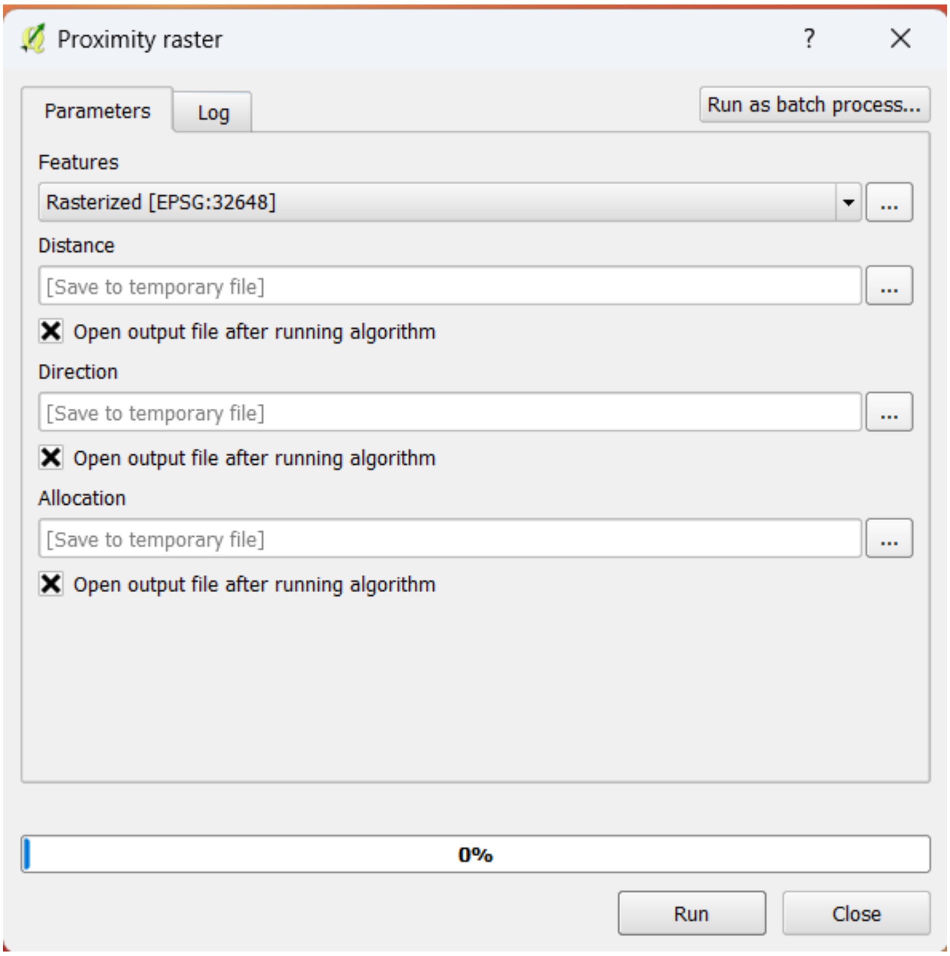
- Select the newly created field ('id' in this case) as an **Attribute field**. Set output resolution as per required (25*25 in this case). Define **Nodata value** in advanced parameters as -9999. Run the tool and save the rasterized dataset.



- Use raster tool named " Proximity raster" to generate the distance from the road/river layer.



- Select the rasterized layer generated from previous layer, run the tool and save the result.



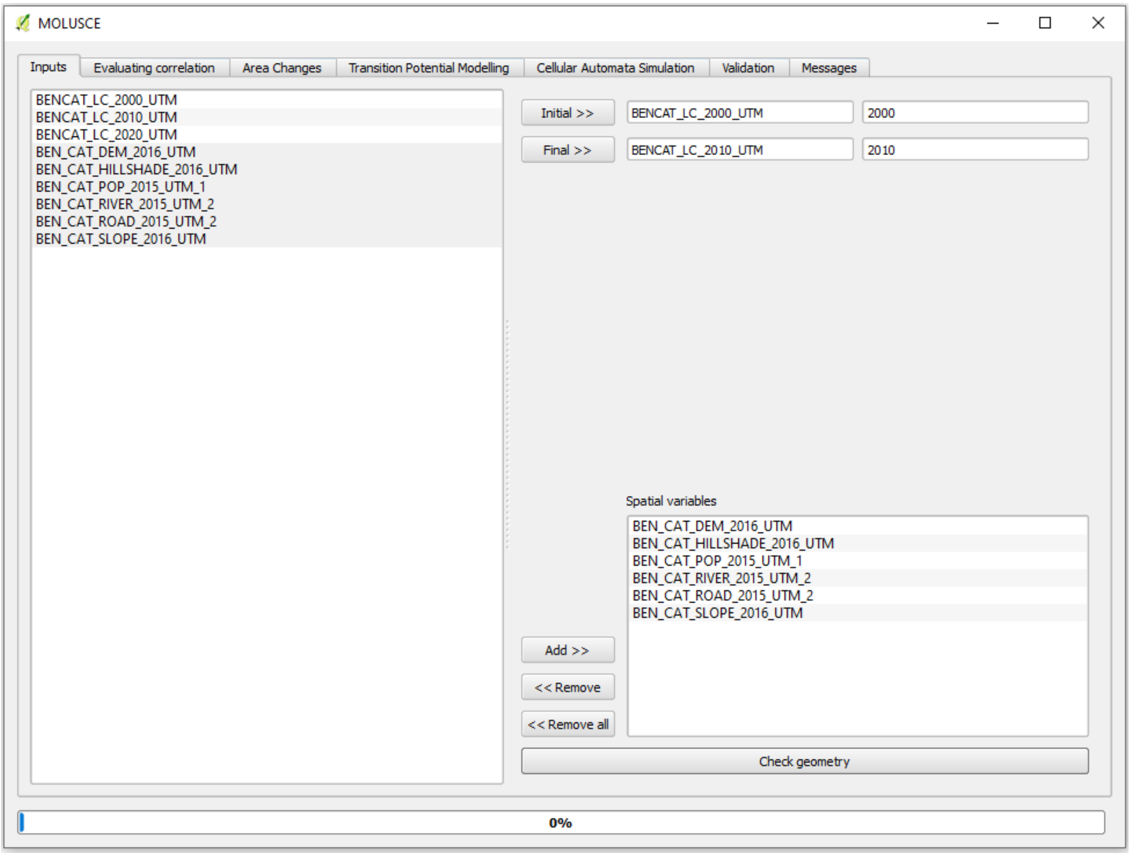
- Finally, clip the raster layer within the study area administrative boundary [refer to 4.1 (g)]

5. Steps in QGIS MOLUSCE Plugin

5.1 Step 1: Inputs

All the layers which have been added in QGIS project are listed in left panel in Inputs tab.

- Firstly, select the layer you want to set as an initial data and then click ***Initial>>*** button to set the initial data. The year box is filled automatically based on layer's name you have defined for selected layer, if not you can define/edit layer name as well as year. Similarly, select and set the final layer as well. In this case study, land cover of year 2000 and 2010 are set as initial and final layer respectively.
- In a similar way, select the layers you want to define as Spatial variables and set in spatial variables window by clicking ***Add>>*** button.
- Click on ***Check geometry*** button to check whether the inputs dataset fulfill all the requirement that includes same coordinate system, same resolution and same extent. If all the requirements are fulfilled, the prompt window is displayed with message "Geometries of raster's are matched" and all other remaining tabs will be activated.
- Click ***ok*** to close the prompt window.
- If the geometries of raster's do not match, error message will be displayed. Then, correct the dataset and try again.



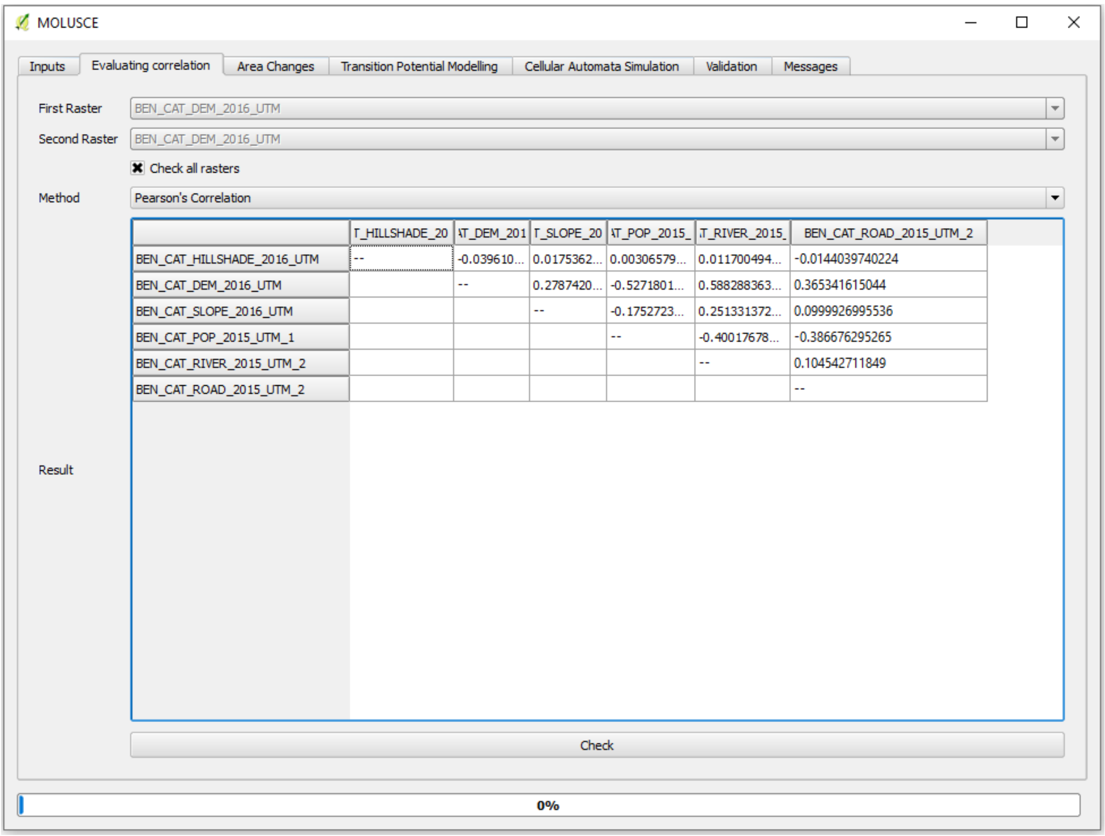
5.2 Step 2: Evaluating Correlation

This tab aims for evaluating the correlation between spatial variables that are defined in the previous input tab. The three methods that are available for calculating correlations among spatial variables are Pearson's correlation, Crammer's correlation and Joint Information Uncertainty**.**

Check all rasters option must be selected to calculate the correlation among all variables.

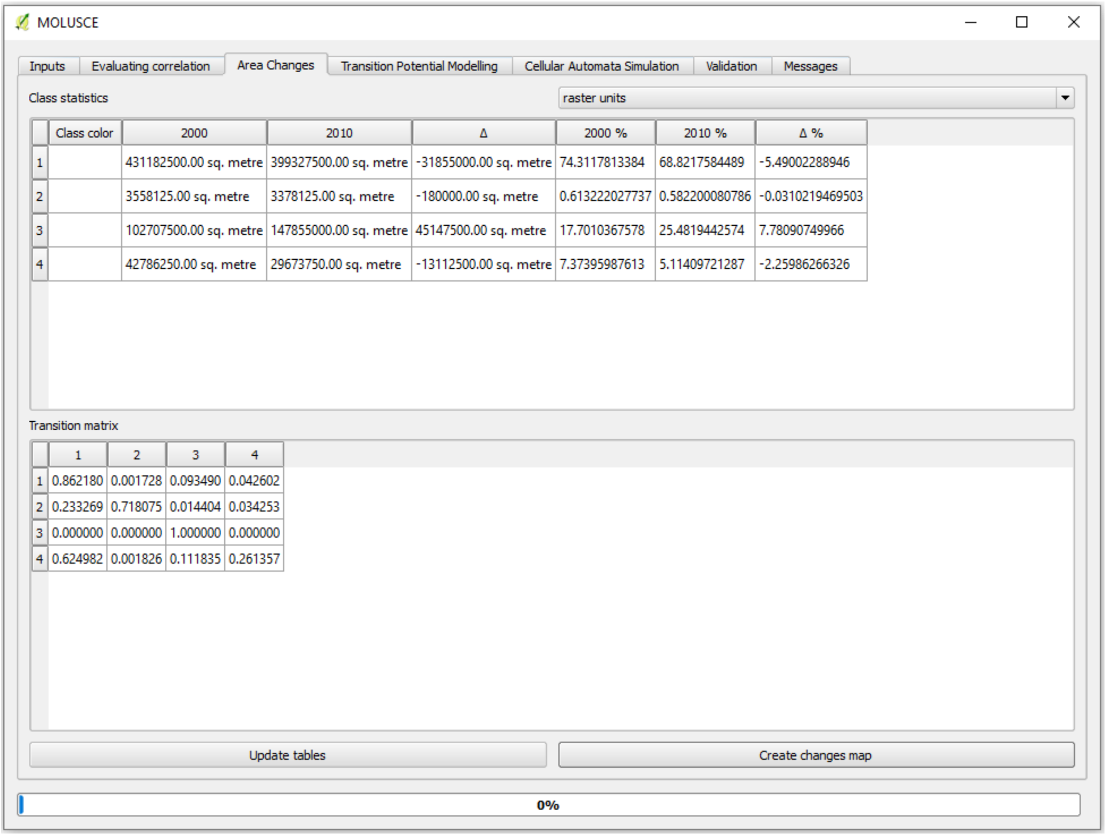
In our case, since all the spatial variables are continuous type, Pearson's correlation method has been selected.

In general, if the correlation between two drivers (or variables) is greater than 0.75, it suggests a strong linear relationship between them. In such cases, one of the drivers may be neglected to avoid multicollinearity implications such as redundancy, overfitting, and difficulty in training the models.



5.3 Step 3: Area Changes

- Click on ***Update tables*** button to compute land use/cover area changes and transition probabilities table.The land use/cover area units can be expressed in raster unit, square kilometer (sq. km) and hectare (ha) as shown in below Figure.
- Generate and save the land use/cover change map by clicking ***the Create changes map*** button.



5.4 Step 4: Transition Potential Modeling

After successful completion of the previous steps, transition potential modelling is carried out. The modeling is carried out by defining samples, modeling method and parameters. Firstly, the available three sampling mode options are All, Random and Stratified and selection of sampling mode depends on the class's distribution of input dataset. In this case study, stratified sampling method is selected since precise class representation is crucial and input dataset classes distribution are also not balanced.

Secondly, there are four methods that encompass artificial neural networks (ANN), weights of evidence (WoE), logistic regression (LR), and multi-criteria evaluation for constructing potential transition maps. Each method has its own advantages and disadvantages. In this case study, the ANN method has been selected due to its capability to handle complex (non-linear) factors.

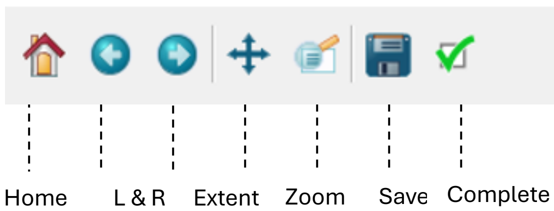
Thirdly, other miscellaneous parameters such as number of samples, neighborhood, learning rate, maximum iterations, hidden layers, and momentum are input.

After giving all the inputs, there is a button '**Train neural network**' which needs to be clicked to start the modelling process. There is also an option to '**Stop'** button which is used to terminate the model.

After modelling is completed, overall accuracy values appeared on the interface determines the quality of the modelling. In most cases, desirable overall accuracy might not be obtained in the first step. Hence, other miscellaneous parameters need to be adjusted until the desirable overall accuracy is obtained based on the hit and trial method.

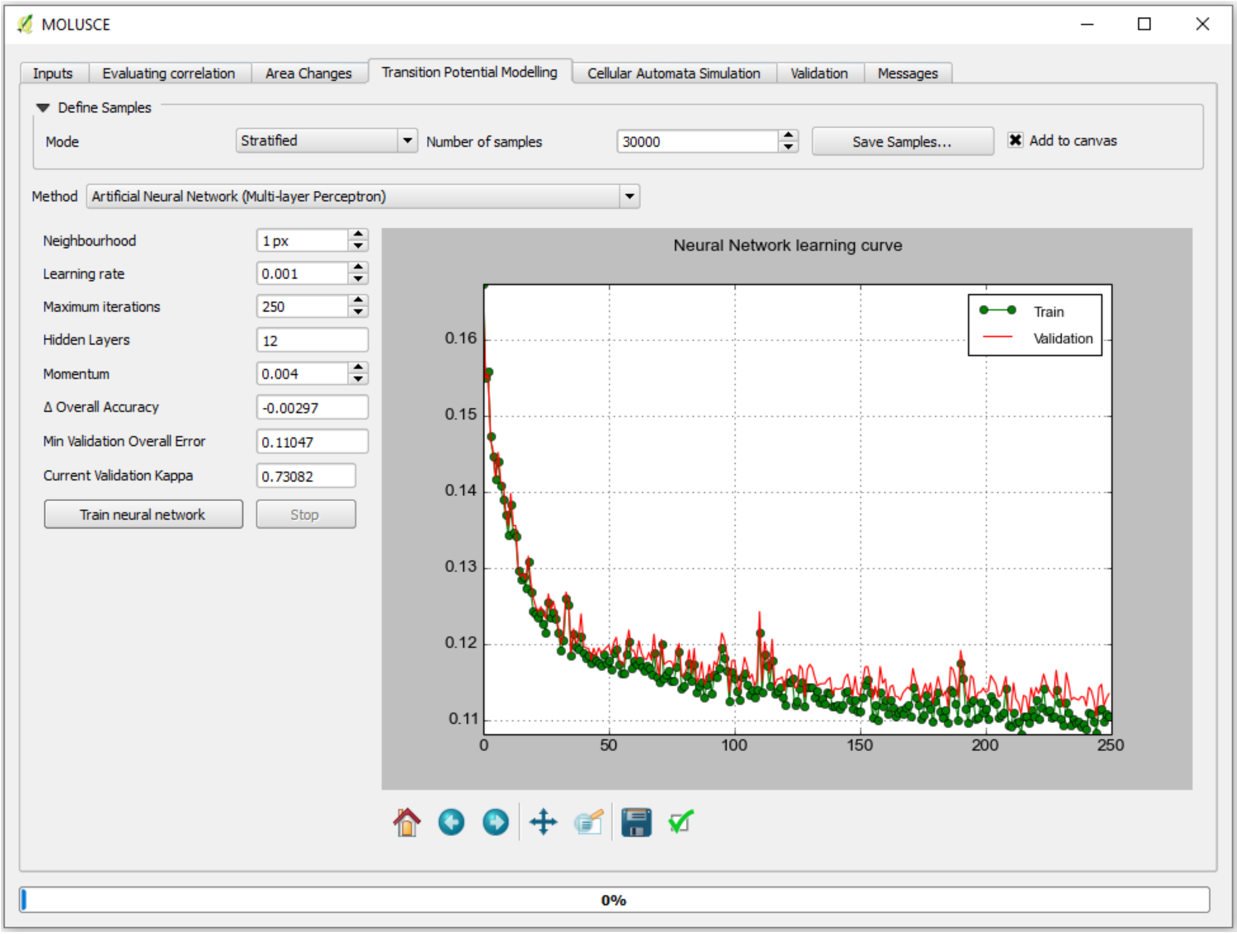
In this very crucial step, neural network learning curve (training vs validation datasets) need to be reviewed and diagnosed properly to prevent from the learning problems such as an underfit or overfit model, and unrepresentative training or validation datasets. [Reference: https://machinelearningmastery.com/learning-curves-for-diagnosing-machine-learning-model-performance/]

The input values used in this case study are shown in the Figure below. The interface also has different functions as shown below.



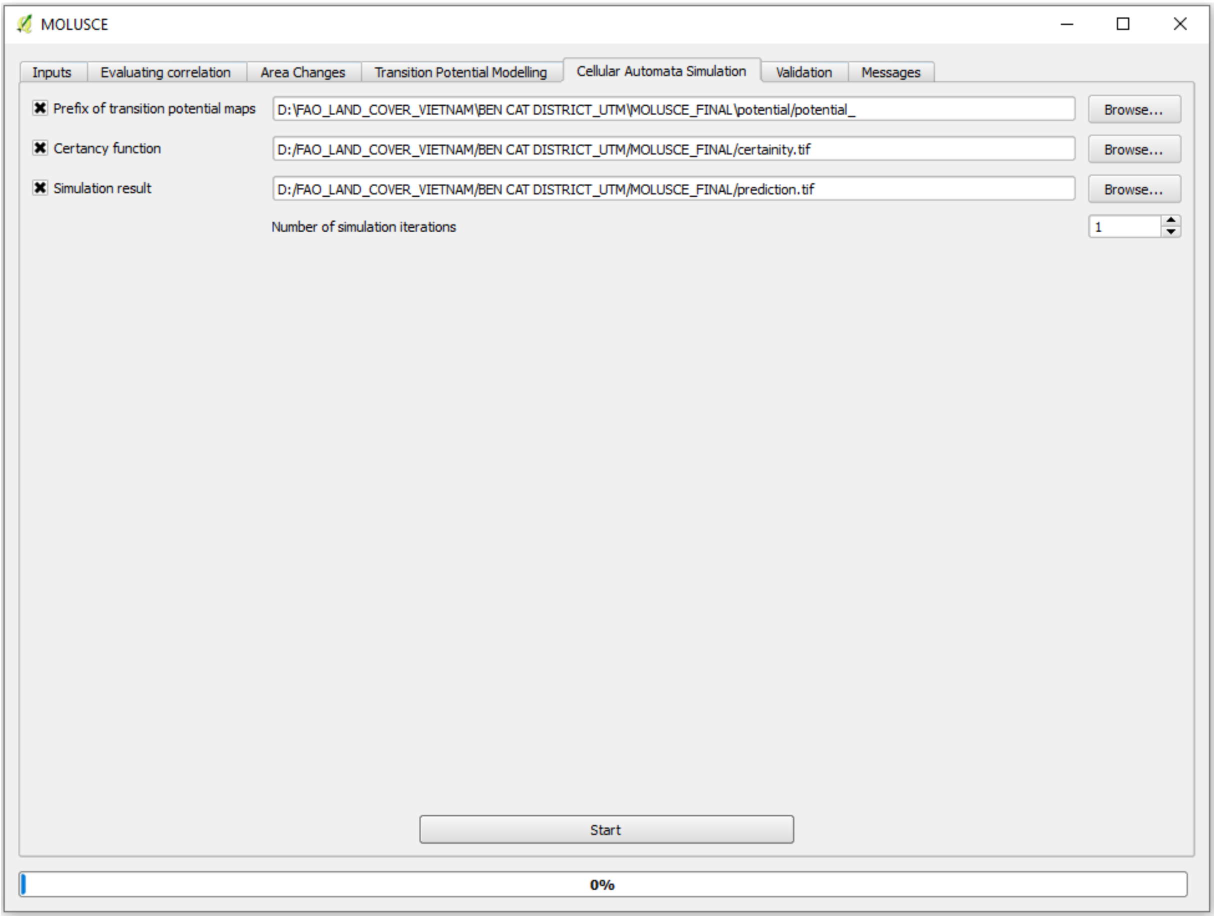
After successful completion of the training the model **click** on the save and save the graph into the designated folder.

In this case sudy, decrease in both training and validation loss to a point of stability with a minimal gap between the two final loss values can be witnessed in the learning curve which represents the well fitted model for this case study.



5.5 Step 5: Cellular Automata Simulation

This is the second last step. The main task that needs to be carried out is to save the file path for prefix of transition potential maps, certainty function, and simulation results as shown in below Figure. In step 1, the year difference between initial (2000) and final (2010) land cover map is 10. Thus, Number of simulation iterations defined as 1 in this step will simulates the land cover map of year 2020.



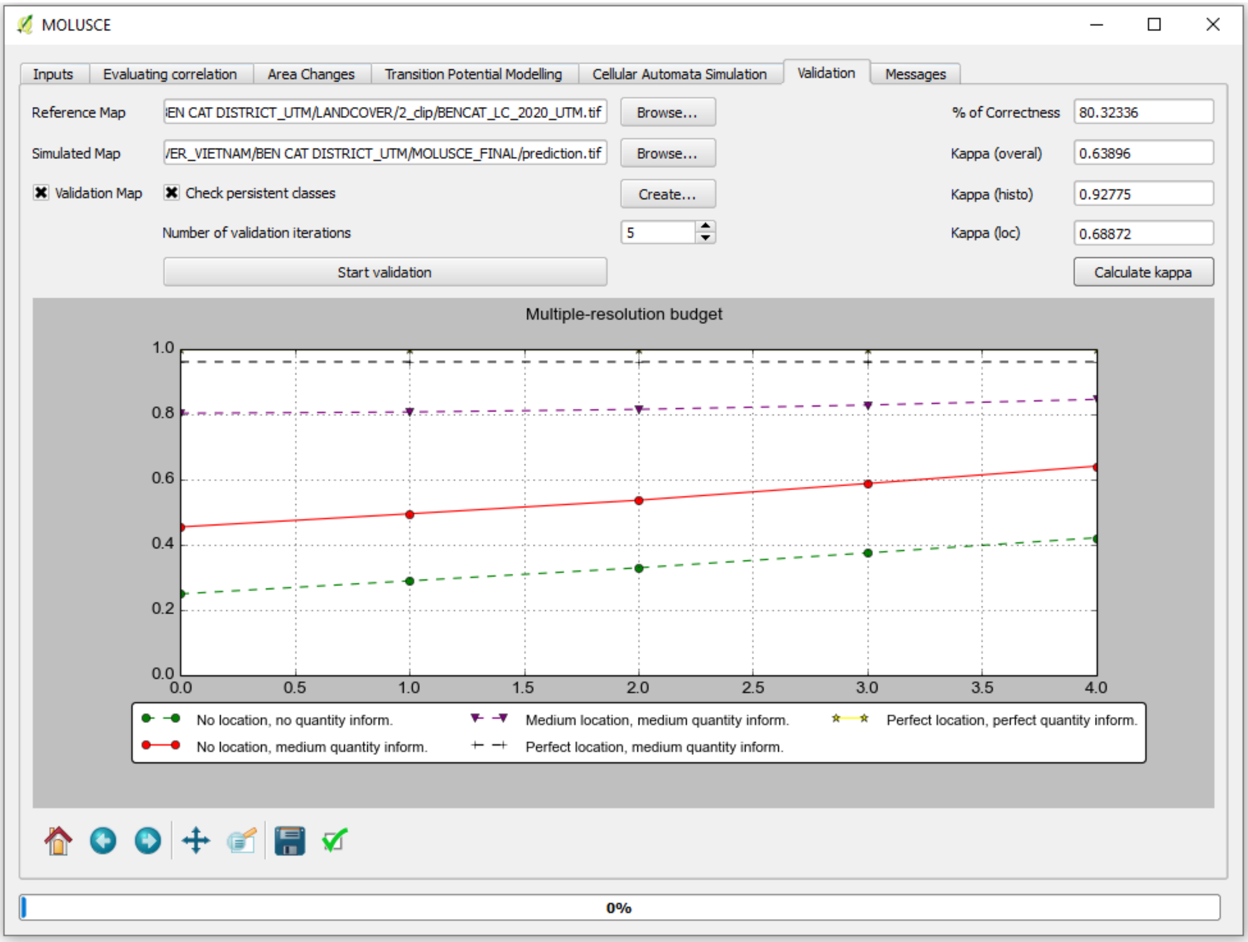
5.6 Step 6: Validation

This is the last step of the MOLUSCE. As shown in the Figure below there are many functions on validation steps.

Firstly, reference map needs to be imported. To do this, click on **Browse** and select the reference land cover which will act as validation data against the simulated map. To import the simulated map, adopt similar process using the browsing tool.

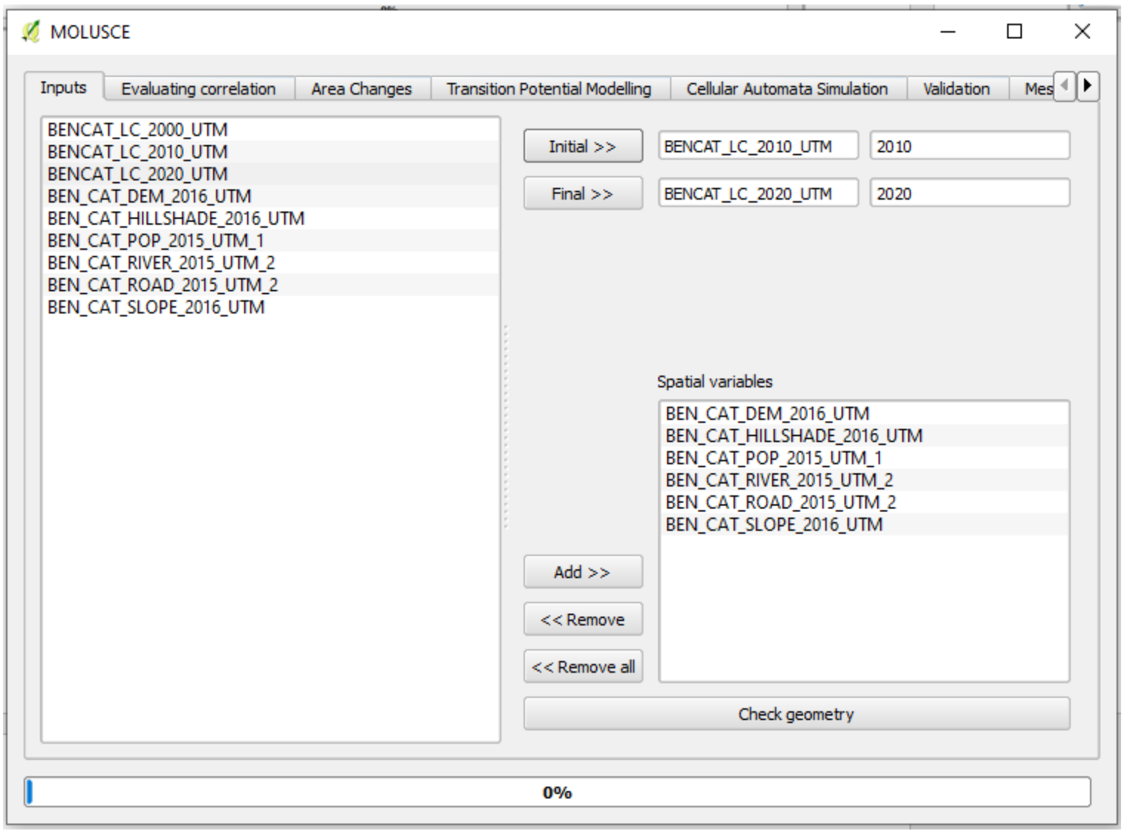
Ensure that you have checked the validation map and check persistent classes.

The number of validation iterations should be imported based on the results obtained on the % of correctness, kappa (overall), kappa (histo), and kappa (loc). If these values are satisfying (i.e. >0.6) the simulated map can be accepted else hit and trials method should be adopted by changing the parameters used in the transition potential modelling step.

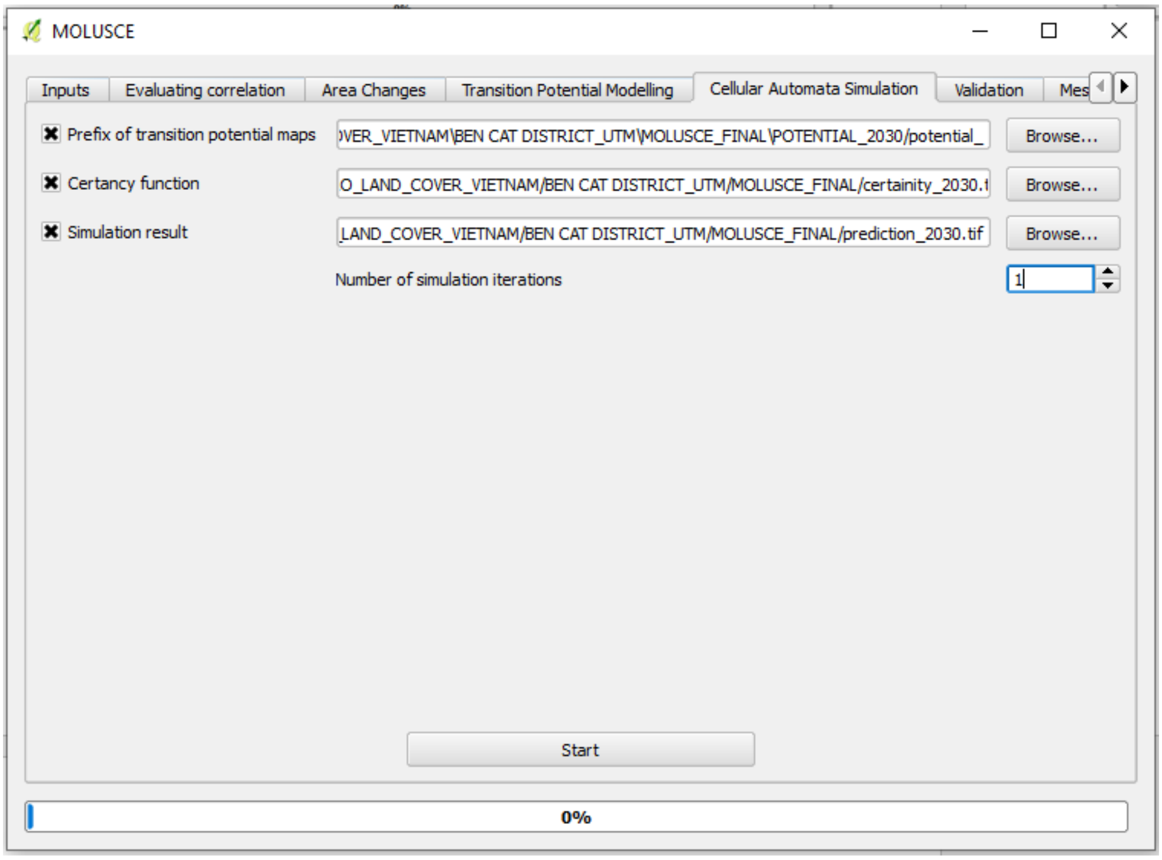


5.7 Step 7: Prediction

- Once acceptable kappa statistics are obtained in previous step, inputs need to be given to predict the future land cover map. As shown in the figure all the land cover maps and driving are given as input. The below figure shows the input for this case study. Land cover map of year 2010 and 2020 are set as initial and final respectively to predict the land cover for year 2030 and so on.



- Once the inputs are defined, proceed directly to step 5 (Cellular Automata Simulation). The number of simulation iterations must be entered in a specific way. In this case study, with an initial year of 2010 and a final year of 2020, setting the number of simulation iterations to 1 will predict outcomes for 2030, 2 will predict for 2040, 3 for 2050, and so forth. Therefore, each iteration represents one interval between the initial and final years.



- Simulate and save the predicted land cover map